IV

PLV Complete Genomes

Contents

IV-1	Introduction	**219**
IV-2	Sequences	**220**
IV-3	Alignments	**222**

IV-1 Introduction

Complete genomes of HIV-1, HIV-2, and SIV sequences are represented in the alignment. For the selection, one of each of the major subtypes and a few CRF s of the HIV-1 M group were chosen, plus one N group and 2 O group sequences. Most of the SIV-CPZ s were included, including new sequences from wild caught chimpanzees. Three HIV-2 A group and 2 HIV-2 B group sequences were included along with the single genomes sequenced to date for HIV-2 G and HIV-2 U. For the rest of the SIVs, all complete genomes were included. The principle in deciding which sequences to include in this alignment was to provide a representative sampling of all primate lentiviruses which have had complete or nearly complete genomes sequenced. Many HIV-1, HIV-2, SIVsmm and SIVmac genomes were not included, but they can be found in the HIV-1/SIVcpz and HIV-2/SIVsmm alignments. Also, many shorter sequences from PLVs are included in the PLV protein sequence alignments later in this compendium. Because of the complicated phylogeny, most likely muddled by recombination and many cross-species transmissions, we do not show a tree for the PLV alignment this year. Please refer to the Sequence Compendium of 2005 for a more comprehensive evaluation of the PLV phylogeny.

The alignment was based on the previous version published last year, produced by a hidden Markov model using HMMER and manual editing using the programs BioEdit and Se-Al. It is codon-aligned, meaning that the correct translation reading frame has been maintained as much as possible; in the case of overlapping frames, obviously one will not translate correctly. We have tried to "reset" the alignment so that each new gene starts in the first frame; this means some empty columns have been inserted. The HIV-2/SIVsmm *vpx* gene is postulated to be a duplication of the *vpr* gene and thus there may be two alternative alignments of this region of the genome, as there are for the duplicated stem-loops of the TAR element.

The annotation is based on known protein coding regions in HIV-1 and on annotations found in SIV sequence database entries. The protein cleavage sites that create Gag p17, Gag p24 and other mature peptides from the Gag and Gag-Pol precursor polyproteins have been experimentally determined for HIV-1 and at least one strain of HIV-2, but the study of analogous cleavages in SIV polyproteins have not been published. Two representative genomes have been translated: HIV-1 subtype B strain HXB2, and MAC239. The translations are provided as a visual aid for finding landmarks in the genomes.

IV-2 Sequences

Primate lentivirus (PLV) Sequences in the complete genome alignments.

Name	Accession	Country	Author	Reference
H1B.FR.83.HXB2	K03455	France	van Beveren, C.P.	*(in) Weiss,RL, Teich,N, Varmus,H and Coffin,J(Eds); RNA TUMOR VIRUSES, SECOND EDITION,* **2**, Vol 2: 1102-1123; Cold Spring Harbor Laboratory, Cold Spring Harbor (1985)
H1A1.UG.85.U455	M62320	Uganda	Oram, J.D.	*ARHR* **6**(9):1073-1078 (1990)
H1B.US.90.WEAU160	U21135	United States	Wei, X.	*Nature* **422**(6929):307-312 (2003)
H1C.ET.86.ETH2220	U46016	Ethiopia	Salminen, M.O.	*ARHR* **12**(14):1329-1339 (1996)
H1D.CD.84.84ZR085	U88822	D.R.C.	Gao, F.	*J Virol* **72**(7):5680-5698 (1998)
H1F1.BE.93.VI850	AF077336	Belgium	Laukkanen, T.	*Virology* **269**(1):95-104 (2000)
H1G.SE.93.SE6165	AF061642	Sweden	Carr, J.K.	*Virology* **247**(1):22-31 (1998)
H1H.CF.90.056	AF005496	C.A.R.	Gao, F.	*J Virol* **72**(7):5680-5698 (1998)
H1J.SE.93.SE7887	AF082394	Sweden	Laukkanen, T.	*ARHR* **15**(3):293-297 (1999)
H1K.CM.96.MP535	AJ249239	Cameroon	Triques, K.	*ARHR* **16**(2):139-151 (2000)
H101_AE.TH.90.CM240	U54771	Thailand	Carr, J.K.	*J Virol* **70**(9):5935-5943 (1996)
H102_AG.NG.x.IBNG	L39106	Nigeria	Howard, T.M.	*ARHR* **10**(12):1755-1757 (1994)
H1N.CM.95.YBF30	AJ006022	Cameroon	Simon, F.	*Nat Med* **4**(9):1032-1037 (1998)
H1O.BE.87.ANT70	L20587	Belgium	Vanden Haesevelde, M.	*J Virol* **68**(3):1586-1596 (1994)
H1O.CM.91.MVP5180	L20571	Cameroon	Gurtler, L.G.	*J Virol* **68**(3):1581-1585 (1994)
CPZ.CD.90.ANT	U42720	D.R.C.	Vanden Haesevelde, M.M.	*Virology* **221**(2):346-350 (1996)
CPZ.CM.01.SIVcpzCAM13	AY169968	Cameroon	Nerrienet, E.	*J Virol* **79**(2):1312-9 (2005)
CPZ.CM.05.SIVcpzEK505	DQ373065	Cameroon	Keele, B.F.	*Science* **313**(5786):523-526 (2006)
CPZ.CM.05.SIVcpzMB66	DQ373063	Cameroon	Keele, B.F.	*Science* **313**(5786):523-526 (2006)
CPZ.CM.98.CAM3	AF115393	Cameroon	Corbet, S.	*J Virol* **74**(1):529-534 (2000)
CPZ.GA.88.GAB2	AF382828	Gabon	Bibollet-Ruche, F.	*ARHR* **20**(12):1377-1381 (2004)
CPZ.CM.98.CAM5	AJ271369	Cameroon	Muller-Trutwin, M.C.	*J Med Primatol* **29**(3-4); 166-72 (2000)
CPZ.CM.05.SIVcpzMT145	DQ373066	Cameroon	Keele, B.F.	*Science* **313**(5786):523-526 (2006)
CPZ.TZ.00.TAN1	AF447763	Tanzania	Santiago, M.L.	*J Virol* **77**(3):2233-2242 (2003)
CPZ.CM.05.SIVcpzLB7	DQ373064	Cameroon	Keele, B.F.	*Science* **313**(5786):523-526 (2006)
CPZ.US.85.CPZUS	AF103818	United States	Gao, F.	*Nature* **397**(6718):436-441 (1999)
CPZ.GA.88.GAB1	X52154	Gabon	Huet, T.	*Nature* **345**(6273):356-359 (1990)
MON.CM.99.L1	AY340701	Cameroon	Courgnaud, V.	*J Virol* **77**(23):12523-12534 (2003)
MON.NG.x.NG1	AJ549283	Nigeria	Barlow, K.L.	*J Virol* **77**(12):6879-6888 (2003)
GSN.CM.99.CN166	AF468659	Cameroon	Courgnaud, V.	*J Virol* **76**(16):8298-8309 (2002)
GSN.CM.99.CN71	AF468658	Cameroon	Courgnaud, V.	*J Virol* **76**(16):8298-8309 (2002)
MUS_1.CM.01.1085	AY340700	Cameroon	Courgnaud, V.	*J Virol* **77**(23):12523-12534 (2003)
MUS_1.CM.01.CM1239	EF070330	Cameroon	Aghokeng, A.F.	*Virology* **360**(2):407-418 (2007)
MUS_2.CM.01.CM1246	EF070329	Cameroon	Aghokeng, A.F.	*Virology* **360**(2):407-418 (2007)
MUS_2.CM.01.CM2500	EF070331	Cameroon	Aghokeng, A.F.	*Virology* **360**(2):407-418 (2007)
RCM.NG.x.NG411	AF349680	Nigeria	Beer, B.E.	*J Virol* **75**(24):12014-12027 (2001)
RCM.GA.x.GAB1	AF382829	Gabon	Bibollet-Ruche, F.	*J Virol* **78**(14):7748-7762 (2004)
DEN.CD.x.CD1	AJ580407	D.R.C.	Dazza, M.C.	*J Virol* **79**(13):8560-8571 (2005)
MAC.US.x.239	M33262	United States	Kestler, H.	*Science* **248**(4959):1109-1112 (1990)

Name	Accession	Country	Author	Reference
H2A.GW.x.ALI	AF082339	Guinea-Bissau	Azevedo-Pereira, J.M.	Unpublished
H2A.DE.x.BEN	M30502	Germany	Kirchhoff, F.	*Virology* **177**(1):305-311 (1990)
H2A.SN.x.ST	M31113	Senegal	Kumar, P.	*J Virol* **64**(2):890-901 (1990)
H2B.GH.86.D205	X61240	Ghana	Dietrich, U.	*Nature* **342**(6252):948-950 (1989)
H2B.CI.x.EHO	U27200	Cote D'Ivoire	Rey-Cuille, M.A.	*Virology* **202**(1):471-476 (1994)
H2G.CI.92.ABT96	AF208027	Cote D'Ivoire	Brennan, C.A.	*ARHR* **13**(5):401-404 (1997)
H2U.FR.96.12034	AY530889	France	Damond, F.	*ARHR* **20**(6):666-672 (2004)
MAC.US.x.EMBL_3	Y00295	United States	Franchini, G.	*Nature* **328**(6130):539-543 (1987)
SMM.US.x.H9	M80194	United States	Courgnaud, V.	*J Virol* **66**(1):414-419 (1992)
SMM.US.x.PGM53	AF077017	United States	Novembre, F.J.	*J Virol* **72**(11):8841-8851 (1998)
SMM.US.x.SIVsmH635F_L3	DQ201172	United States	Kuwata, T.	*J Virol* **80**(3):1463-75 (2006)
SMM.SL.92.SL92B	AF334679	Sierra Leone	Chen, Z.	*J Virol* **70**(6):3617-3627 (1996)
STM.US.x.STM	M83293	United States	Novembre, F.J.	*Virology* **186**(2):783-787 (1992)
MNE.US.x.MNE027	U79412	United States	Kimata, J.T.	*J Virol* **72**(1):245-256 (1998)
DRL.x.x.FAO	AY159321		Hu, J.	*J Virol* **77**(8):4867-4880 (2003)
MND_2.x.x.5440	AY159322		Hu, J.	*J Virol* **77**(8):4867-4880 (2003)
MND_2.CM.98.CM16	AF367411	Cameroon	Takehisa, J.	*ARHR* **17**(12):1143-1154 (2001)
MND_2.GA.x.M14	AF328295	Gabon	Souquiere, S.	*J Virol* **75**(15):7086-7096 (2001)
OLC.CI.97.97CI12	FM165200	Cote D'Ivoire	Liegeois, F.	*J Virol* **83**(1):428-439 (2009)
GRV.ET.x.GRI_677	M66437	Ethiopia	Fomsgaard, A.	*Virology* **182**(1):397-402 (1991)
VER.KE.x.TYO1_patent	DJ048201	Kenya	Omori, T.	Patent: WO 2007049749-A 13 03-MAY-2007; DNAVEC CORPORATION
VER.KE.x.9063	L40990	Kenya	Hirsch, V.M.	*J Virol* **69**(2):955-967 (1995)
VER.KE.x.AGM155	M29975	Kenya	Johnson, P.R.	*J Virol* **64**(3):1086-1092 (1990)
VER.DE.x.AGM3	M30931	Germany	Baier, M.	*Virology* **176**(1):216-221 (1990)
TAN.UG.x.TAN1	U58991	Uganda	Soares, M.A.	*Virology* **228**(2):394-399 (1997)
SAB.SN.x.SAB1C	U04005	Senegal	Jin, M.J.	*EMBO J* **13**(12):2935-2947 (1994)
MND_1.GA.x.MNDGB1	M27470	Gabon	Tsujimoto, H.	*Nature* **341**(6242):539-541 (1989)
DEB.CM.99.CM40	AY523865	Cameroon	Bibollet-Ruche, F.	*J Virol* **78**(14):7748-7762 (2004)
DEB.CM.99.CM5	AY523866	Cameroon	Bibollet-Ruche, F.	*J Virol* **78**(14):7748-7762 (2004)
TAL.CM.01.8023	AM182197	Cameroon	Liegeois, F.	*Virology* **349**(1):55-65 (2006)
TAL.CM.00.266	AY655744	Cameroon	Liegeois, F.	*Virology* **349**(1):55-65 (2006)
SUN.GA.98.L14	AF131870	Gabon	Beer, B.E.	*J Virol* **73**(9):7734-7744 (1999)
WRC.CI.97.97CI14	AM745105	Cote D'Ivoire	Liegeois, F.	*J Virol* **83**(1):428-439 (2009)
WRC.CI.98.98CI04	AM713177	Cote D'Ivoire	Liegeois, F.	*J Virol* **83**(1):428-439 (2009)
WRC.GM.05.Pbt_05GM_X02	AM937062	Gambia	Locatelli, S.	*Virology* **376**(1):90-100 (2008)
LST.CD.88.524	AF188116	D.R.C.	Beer, B.E.	*J Virol* **74**(8):3892-3898 (2000)
LST.CD.88.447	AF188114	D.R.C.	Beer, B.E.	*J Virol* **74**(8):3892-3898 (2000)
LST.KE.x.lho7	AF075269	Kenya	Hirsch, V.M.	*J Virol* **73**(2):1036-1045 (1999)
SYK.KE.x.SYK173	L06042	Kenya	Hirsch, V.M.	*J Virol* **67**(3):1517-1528 (1993)
SYK.KE.x.KE51	AY523867	Kenya	Bibollet-Ruche, F.	*J Virol* **78**(14):7748-7762 (2004)
COL.CM.x.CGU1	AF301156	Cameroon	Courgnaud, V.	*J Virol* **75**(2):857-866 (2001)

```
                          5' LTR U3 start
H1B.FR.83.HXB2            .TGGAAGGGCTAATTCACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTAGCAGAACTACACACCA...GGGCCAGGG...ATCAGATATCCACTGACCTTTGGATGGTGCTACAAGCTAGTACCAGTTGAGCCAGA  163
H1A1.UG.85.U455           ................................................................................................................................................................  0
H1B.US.90.WEAU160         .-------------T--------AA-----------------------G-----------------T----------G---------------...---------...-CT-----C------TG-------------T--------------G--A-----  163
H1C.ET.86.ETH2220         ................................................................................................................................................................  0
H1D.CD.84.84ZR085         ................................................................................................................................................................  0
H1F1.BE.93.VI850          ................................................................................................................................................................  0
H1G.SE.93.SE6165          ................................................................................................................................................................  0
H1H.CF.90.056             ................................................................................................................................................................  0
H1J.SE.93.SE7887          ................................................................................................................................................................  0
H1K.CM.96.MP535           ................................................................................................................................................................  0
H101_AE.TH.90.CM240       ................................................................................................................................................................  0
H102_AG.NG.x.IBNG         ................................................................................................................................................................  0
H1N.CM.95.YBF30           ................................................................................................................................................................  0
H1O.BE.87.ANT70           C-------T-----T-------TAA---GC---A-----G-----T---G-G--TA---T--G--A-T----------G---------------...--A-----A...-C---G-TC--------A---------TG-TT--A-----------GTCAGA---  164
H1O.CM.91.MVP5180         C----T---T-----T-------TAAG---GC---A-----G-----C-----A--T-----T--G--A-T----------G----TGT--------G...--A-----A...CCT----TC--------A---------TG-TT--A--G--------GTCAG----  164
CPZ.CD.90.ANT             ................................................................................................................................................................  0
CPZ.CM.01.SIVcpzCAM13     ................................................................................................................................................................  0
CPZ.CM.05.SIVcpzEK505     ................................................................................................................................................................  0
CPZ.CM.05.SIVcpzMB66      ................................................................................................................................................................  0
CPZ.CM.98.CAM3            ................................................................................................................................................................  0
CPZ.GA.88.GAB2            ................................................................................................................................................................  0
CPZ.CM.98.CAM5            ................................................................................................................................................................  0
CPZ.CM.05.SIVcpzMT145     ................................................................................................................................................................  0
CPZ.TZ.00.TAN1            ................................................................................................................................................................  0
CPZ.CM.05.SIVcpzLB7       ................................................................................................................................................................  0
CPZ.US.85.CPZUS           .--------T--G--T-T---AG-A----------G-----A-----C---G-T----T--------AAT--------C-G------T-----C---...-----T-A...G-A--------C--A--A-A---GC------TT---T----C--TC-CACAGA---  163
CPZ.GA.88.GAB1            .--------T--G--T-----AGGA----------G--------C--C---G----T------------T--------C-G------------A--...--A-----A...-CA----TC-----CTGT------------T---A-----G--CC-GACAGAG--  163
MON.CM.99.L1              ................................................................................................................................................................  0
MON.NG.x.NG1              ................................................................................................................................................................  0
GSN.CM.99.CN166           ................................................................................................................................................................  0
GSN.CM.99.CN71            ................................................................................................................................................................  0
MUS_1.CM.01.1085          ................................................................................................................................................................  0
MUS_1.CM.01.CM1239        ................................................................................................................................................................  0
MUS_2.CM.01.CM1246        ................................................................................................................................................................  0
MUS_2.CM.01.CM2500        ................................................................................................................................................................  0
RCM.NG.x.NG411            ................................................................................................................................................................  0
RCM.GA.x.GAB1             ................................................................................................................................................................  0
DEN.CD.x.CD1              ................................................................................................................................................................  0
                          5' LTR U3 start
MAC.US.x.239              .--------A-TTA-T--AGTGC-A-------TAGA---T-A--CA-A-ACT-AG-AA-GGA-G-----AT-A-A--A----G----G-T-----CT--...--A-----A...--T-----C---AA---A-----C---CTA-GG--AT----C--T--AA-T.....  158
H2A.GW.x.ALI              .--------A-GT--T--AGTG-GA-------TAGA---T-A--CA-A-ACT-AG-AA-GGA-G----GATAA-T--A----G---------T--T-AT...---------A...--A--G--C--GA--TT------G---CTG-GG--------------A--T.....  158
H2A.DE.x.BEN              .--C-----A-GT--T--AGTAGGA-G-----TAGA-----A--CA-A-ACC-AG-AA-AGAGG----GATAA-A--A----G------T--T--T-AT...---------A...G-A--G--C---A--TA---C--G---CTG-GG---------T----A--A.....  158
H2A.SN.x.ST               ................................................................................................................................................................  0
H2B.GH.86.D205            .--------A-GTA-T-TAGTG-GA-------CAGA--AT-A--CACA-ATT-TG-GA-TGA-G-----ATTG-GT---GC-G---A-----T--T-AT...-----------...--A--GC----CAAATA------T---CTG-GG-----G--------A---...-T  160
H2B.CI.x.EHO              .--------A-TT-CT-TAGTG-GA----G--TA-A--A--A--CACA-ACT-AG-AA-TGA-G-----ATTG-GT---GA-G-------------AT...---------A...G-A--------CAA-TT------C---CTA-GG--------G---A-AA-T.....  158
H2G.CI.92.ABT96           ................................................................................................................................................................  0
H2U.FR.96.12034           ................................................................................................................................................................  0
MAC.US.x.EMBL_3           ................................................................................................................................................................  0
SMM.US.x.H9               ................................................................................................................................................................  0
SMM.US.x.PGM53            .....................................................................--A---G--A--T-----T--...---------A...---------C--TA--TT------C---CTA-GG--AT-G--C-----AAGT.....  84
SMM.US.x.SIVsmH635F_L3    .--------A-TTA-T--AATG-GAA------TAGA--AT-A---A---AC--GG-AA-GGA-G---AATAA-A--A----G--A--T------TT-...---------A...-CT-----C--TA--TA------G---CT--GG--AT----C-----A--T.....  158
SMM.SL.92.SL92B           ................................................................................................................................................................  0
STM.US.x.STM              .--------A-TTA-T--AGTG-.........................................................................................................................................  22
MNE.US.x.MNE027           ................................................................................................................................................................  0
DRL.x.x.FAO               ................................................................................................................................................................  0
MND_2.x.x.5440            ................................................................................................................................................................  0
MND_2.CM.98.CM16          ................................................................................................................................................................  0
MND_2.GA.x.M14            ................................................................................................................................................................  0
OLC.CI.97.97CI12          ..........................................................AA-AGG-A-C-TC---GCAACAGAGGAGGTA-AAGAG--CACGT-A--GAGA--TAA---GCCA--G----TA-GGG-T--TAGGAG---G--A.....  94
GRV.ET.x.GRI_677          .----T---A--TA-T----TG--A-----G--A-G-----GA--T---ATGC--TGA--GAGTGG--AATAA-AGA-----G---AGCT---T-----...--C--G---...--A--G--C--GAGAGT------C-TC----TT--------C-----G--C.....  158
VER.KE.x.TYO1_patent      .----T---A-TTA-T-----G-TA-G---A-TA-G-----GA------ATGCTCTTA-TGA-TGG--GATAA-TGA-----G-A-TGC--GGT-GAAG...--A-----A...--A----TC--TAAATG------G-TC----TT--------G-----G--C.....  158
VER.KE.x.9063             .----T---A-TTA-T-----G--A-G--GA-TAGA-----CA-C----ATGCTCTTA-TGA-TGG--AAT-A-TGA-----G-A-CGC--GGT--A--...--A-----C...G-A--------C-GCTG------C-TT----T-----------G--A---.....  158
VER.KE.x.AGM155           .----T---A-TTA-T-----G-TA-G---A-TA-G------A------ATGC-CT-A-TGA-TGG--AAT-A-TGA-----G-A-CGCA-GGT--AA-...--A--T---...--A-----C--GAG-TG------C-TC----T-----------G----CC.....  158
VER.DE.x.AGM3             ................................................................................................................................................................  0
TAN.UG.x.TAN1             .----G--CA--TA-TGG--A-C-AAG---G-GC-A-----GA----C-ATGC-CTAA-TGA-TGG---ATTA-AGA-----G---AGCA--TT-C--T...---------C...-C---GA-G--C-GATG---C--C-TT----TTG--T----G--G--A--T.....  158
SAB.SN.x.SAB1C            .--------A-TTA-T--AGTG--A-G---A-GA-G---T-G-----C-ATGCTCTTA-TGAGTGG---ATAG-AGA--GA-G------T--T---GAT...---------T...-C---G--C---AA-TG--------------TT--A--G-----T--G--C.....  158
MND_1.GA.x.MNDGB1         ................................................................................................................................................................  0
DEB.CM.99.CM40            ................................................................................................................................................................  0
DEB.CM.99.CM5             ................................................................................................................................................................  0
TAL.CM.01.8023            ................................................................................................................................................................  0
TAL.CM.00.266             ................................................................................................................................................................  0
SUN.GA.98.L14             .---------CTGGT-----AG-ACC---G----A---T-G---T---ATGCAG-AA-TGAGTGG----T-A-AA---GC-G---AG-T-----CAA-...-----T--A...G-G---------AA-G-------C---CTA-GG-------C------CACC...AT  160
WRC.CI.97.97CI14          .----G-ATA--T--TGG--AGCTA-----G----A--AG-GA-C--C-ATGCAC--A-TGA-TGG--GCTGA-AGA--GC-G---AGCT------GA-...--A------...--A----GG---AA---A-----G-------GG--AT-G--T-----AC-C.....  158
WRC.CI.98.98CI04          .----GTATA--T--TGG--T-CTA-----G----A--AG-GA-C--T-ATGC--T-A-TGA-TGG--ACTGA-AGAA-G--G----GCA------GAG...--A-----C...--AC-C-GG---AA---A-----C-----T-GG--AT----G--T--G--A.....  158
WRC.GM.05.Pbt_05GM_X02    ................................................................................................................................................................  0
LST.CD.88.524             ................................................................................................................................................................  0
LST.CD.88.447             ................................................................................................................................................................  0
LST.KE.x.lho7             .-----------TA-TGG-----GA-------G--C---T-AAT---C-ACTGTG-GA-TGAGTGG--TCT-A-TGGA--C-TCAT----------GAT...--A-----A...-CA-----C--T-----A----------TG-GGC--T---A------G-CT...TG  160
SYK.KE.x.SYK173           .....................................................................................................---A...-C-C-G--C--GT-A-T-----------TA-GGG-A-----T---AA----.....  55
SYK.KE.x.KE51             ................................................................................................................................................................  0
COL.CM.x.CGU1             ................................................................................................................................................................  0
```

```
H1B.FR.83.HXB2        GAAGTTAGAAGAA...........GCCAACAAAGGAGAGAACACCAGCTTGTTACACCCTGTGAGC............CTGCATGGAATGGATGACCCGGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCATTTCATCACATGGCCCGAGAGCTGCATC  309
H1A1.UG.85.U455       ........................................................................................................................................  0
H1B.US.90.WEAU160     ----G-------------------T--TG------------A---------------A-----------------------G----------AT----A------A---AT-----A----------AAA--------------G-A----------------  309
H1C.ET.86.ETH2220     ........................................................................................................................................  0
H1D.CD.84.84ZR085     ........................................................................................................................................  0
H1F1.BE.93.VI850      ........................................................................................................................................  0
H1G.SE.93.SE6165      ........................................................................................................................................  0
H1H.CF.90.056         ........................................................................................................................................  0
H1J.SE.93.SE7887      ........................................................................................................................................  0
H1K.CM.96.MP535       ........................................................................................................................................  0
H101_AE.TH.90.CM240   ........................................................................................................................................  0
H102_AG.NG.x.IBNG     ........................................................................................................................................  0
H1N.CM.95.YBF30       ........................................................................................................................................  0
H1O.BE.87.ANT70       AG--GC---AG-........CTA-GAG-T-C-T-T----GGG-T-ATC-CC-G--T--A-CAT-T............GCC-----T-T--A--TA-AC-T-A----A-AC-GAT-----A------T--ATCT---GCAACACC--TG-T--TATGATAACT--C-  313
H1O.CM.91.MVP5180     AG--GC---GAG-........CTG-GT--T-C-AAT--AG-TG-T--TC-TC----T--A-CTT-T............AAT-----GCT--G--TG-AC-CG-G--GA-AC--A-A---CA-----T--ATCAT---GC--AACA--T--A----TGC-AAA---C-  313
CPZ.CD.90.ANT         ........................................................................................................................................  0
CPZ.CM.01.SIVcpzCAM13 ........................................................................................................................................  0
CPZ.CM.05.SIVcpzEK505 ........................................................................................................................................  0
CPZ.CM.05.SIVcpzMB66  ........................................................................................................................................  0
CPZ.CM.98.CAM3        ........................................................................................................................................  0
CPZ.GA.88.GAB2        ........................................................................................................................................  0
CPZ.CM.98.CAM5        ........................................................................................................................................  0
CPZ.CM.05.SIVcpzMT145 ........................................................................................................................................  0
CPZ.TZ.00.TAN1        ........................................................................................................................................  0
CPZ.CM.05.SIVcpzLB7   ........................................................................................................................................  0
CPZ.US.85.CPZUS       -G-AG----GCGG...............-----T---------C--A--TAC--C-G----CA--T--CAGCATGGAATGGAAG---A-CAT...........G-C-----C---ATC---CA--------CTGAA--G--TCGGAGA----GA--TAA-----------  309
CPZ.GA.88.GAB1        -C--G-----C--...............-----TG-------T----A-T--C----G--T--CA-AT-TCAGCATGGGATGGAAG---AGGAC............-A---G--C--G-TC---C-C---------A-G--G--C--AAGA--T--T---A----A-AA----  309
MON.CM.99.L1          ........................................................................................................................................  0
MON.NG.x.NG1          ........................................................................................................................................  0
GSN.CM.99.CN166       ........................................................................................................................................  0
GSN.CM.99.CN71        ........................................................................................................................................  0
MUS_1.CM.01.1085      ........................................................................................................................................  0
MUS_1.CM.01.CM1239    ........................................................................................................................................  0
MUS_2.CM.01.CM1246    ........................................................................................................................................  0
MUS_2.CM.01.CM2500    ........................................................................................................................................  0
RCM.NG.x.NG411        ........................................................................................................................................  0
RCM.GA.x.GAB1         ........................................................................................................................................  0
DEN.CD.x.CD1          ........................................................................................................................................  0

MAC.US.x.239          .GTA-C---T--G............--AC-GG-G-AT---G-GCATTAT--AA-G--T--A-CTCAAACTTCCCAGTGGGATG-CCCTTG-............G----G--TC---CA----A------TCCAACT--G--C-ACAC-T-TGA---ATAT-TTAGAT-C-  303
H2A.GW.x.ALI          .GTCCC-C-----............-GGG-GG-CACT----CTCA-T--C--C-------A--ACAAACAAGCAGGCATGATG-CACGCAT...........G-G--GACA----TT-----A------CCTAAG--G--TCA-G--T---AA---TTTATT--A--C-  303
H2A.DE.x.BEN          .CTC-C-C----G............--AG-GG---AT---GC--A-T----AG-------A-CACAAACAAGCAGACATGATG---AGCAT...........G-G--GACA----T----CA------TC-ATG--G--C-A-A-CT---A----TTCACT---T-C-  303
H2A.SN.x.ST           ........................................................................................................................................  0
H2B.GH.86.D205        -CCAGC--CGACC...........CGAG-GG-G-AG--A-C-CATT--C-AA-G----G-CACAGATCTCCTCATGGGATG-CATCCAT...........G-G--GACTC-TATC---CA------TTC--T---G----A-G--T-TG----TTTCA-TAG-TT--  306
H2B.CI.x.EHO          .-T-A---C----............C-AG-GG-C-AG--A-C-CATT-TC--G-G--T--A-CACAGACCTCCGCATGGGATG-CCCCCAC...........GA---GACCC-T-TC---CA------TC--T--------A-G-CT-TG-----TTCAGCAG-TTC-  303
H2G.CI.92.ABT96       ........................................................................................................................................  0
H2U.FR.96.12034       ........................................................................................................................................  0
MAC.US.x.EMBL_3       ........................................................................................................................................  0
SMM.US.x.H9           ........................................................................................................................................  0
SMM.US.x.PGM53        .GTC-C---T---............--TC-AG---AC----CACATTA---AG-G--T--A-CTCAGATATCCCAGTGGGATG--CCCTG...........G----G--TC-G-CA----A------TTCA-AAT----C-ACAGAT-TGA---TTTTATTAGA--C-  229
SMM.US.x.SIVsmH635F_L3 .GTC-C---T---............--C-GG---AC----CACATT--C--G-G--T--A-CACAGACTCATCAGTGGGATG-CCCCTG...........G----G--AC-G-CA----A------TCCAGAAT-G--T-A-AGCT-T-A---ATTTATTAA-T-C-  303
SMM.SL.92.SL92B       ........................................................................................................................................  0
STM.US.x.STM          ........................................................................................................................................  22
MNE.US.x.MNE027       ........................................................................................................................................  0
DRL.x.x.FAO           ........................................................................................................................................  0
MND_2.x.x.5440        ........................................................................................................................................  0
MND_2.CM.98.CM16      ........................................................................................................................................  0
MND_2.GA.x.M14        ........................................................................................................................................  0
OLC.CI.97.97CI12      .CCA--GCCT--G...........A--CTTG---AT----GGGATGC------G--T--A-TCAGACCCCTCAAACA-CTG-G-AGT-A-GG...............C-GG-T---GTC--CA-TCCA-ATAG----..................A-TAC--C--  218
GRV.ET.x.GRI_677      .CT-CAT--G--G...........--ACG---CT-T----GACA-T-TC--A-G--T--A-CACAGATGGGG......GAAG--CCTGAT-GAATAGATC-TG-------C--G-TC----A-------CCGAAGT-G--G..................T-GA-T-C-  291
VER.KE.x.TYO1_patent  .TTACAT--G---.............--AC-A-C-T-T--A-GACATT--C-AG-C--T--A-C-CAGATGGGA......GAAG--CC-GAT-G-AT-AGCC-TG----GA-C--G-T-----A-----TCCTATGT-G---..................ATA-A-T-CG  291
VER.KE.x.9063         .-T-CAT----G............--AG-G-CCT-T----GACATTA----GCG--T----CACAAATAAAA......GA-G-CCCCGAT-GCAT-AGTC-TG----GACC--G-TC----A-------CCTATGT-----..................-T--A-T--G  291
VER.KE.x.AGM155       .CT-CAT--G---...........--AG-A-C-T-T--A-GGCATT-----G-------A-CACAACTGCAT......GAAG-CCCTGAT-G-ATAAATC-TG-----A-A--G-CA----A------TCCAATGT-G--T..................-TT-A-T-CG  291
VER.DE.x.AGM3         ........................................................................................................................................  0
TAN.UG.x.TAN1         .GTC-CCC-G---...........--C-GG-T-AG...-GACA-T-T---C-C-------CCCAGATTGAA......TG-G-AA-CGACCC-TGG.......A---GACA--G-T-----AA---A-TCCA-T-------..................-TT-A-T-CA  282
SAB.SN.x.SAB1C        .CT--C---G--G...........--T--G--TA-T------CA-T--C--C-G--T----CTCAAGTGGCC......TATG-G-ATGAC-CCTGG.......A---GACCC-G-T-----A-------CCA-TA------..................-T-GACT--G  285
MND_1.GA.x.MNDGB1     ........................................................................................................................................  0
DEB.CM.99.CM40        ........................................................................................................................................  0
DEB.CM.99.CM5         ........................................................................................................................................  0
TAL.CM.01.8023        ........................................................................................................................................  0
TAL.CM.00.266         ........................................................................................................................................  0
SUN.GA.98.L14         TG-TGA---CAGG...........-ATCCT--TCATCCTTGTCAAGC-C-AC-G--TT-CAGCCAGCAAGGGGTTAATGAAG-CCCTTG...........G----GAGAC-TATC----CA-----TCC-ACT--G--C..................T-TGATTT-A  288
WRC.CI.97.97CI14      .TTCAGT------TCAGATAATACA-A-CCAT-CTAT----GA-AT-AG--AC-G--T--A-CAT-TGCTGAG...CCAGA-G--CCCTG...........G-----CATC-G-TC----A-------CCTAAG--G--C......TGTGATTTTGT--CAGGAAGA-  306
WRC.CI.98.98CI04      .TTTAGT-----GTCTAATGACAGCT-TC-GT-CTAC-----G-A--AGC-------T--A-GCT--CAAGAA...TCAGA-G--CCTTG...........G-----CAC--G-TC----A-------CCT--A-----C......TGTGACTTTGTG-CAAGTAGA-  306
WRC.GM.05.Pbt_05GM_X02 ........................................................................................................................................  0
LST.CD.88.524         ........................................................................................................................................  0
LST.CD.88.447         ........................................................................................................................................  0
LST.KE.x.lho7         TG-TGAGT-TA-............-ATCC-TC--AT----CACAAT--C-T--G--TT-AAGCCAGCTGGGGGTCCTGGAAG-CCCCTG...........G----GAG-C-TATC---CAC--CA-TCCAATGT-G---..................-T-GACTT-A  288
SYK.KE.x.SYK173       .-TAGAG-G-TGC...........TTGG-GT-T-AG-----T--ACT-C-CC-C--T--A-CCTCT...GGGACAGGGGA-TTC-A-CAT...........G-G--GAACC-CAT-----AC--CA--CC--A---G--C..................T-TACA-CAG  179
SYK.KE.x.KE51         ........................................................................................................................................  0
COL.CM.x.CGU1         ........................................................................................................................................  0
```

```
                                                                                          TCF-1 alpha
                             |-----------------------------------------------------------------------------------------------------|
H1B.FR.83.HXB2               CGGAGTACTTCAAGAAC...........................................................................................TGCTGACA.......................................................  334
H1A1.UG.85.U455              ...........................................................................................................................................................................    0
H1B.US.90.WEAU160            ----------T--AG--...........................................................................................--------.......................................................  334
H1C.ET.86.ETH2220            ...........................................................................................................................................................................    0
H1D.CD.84.84ZR085            ...........................................................................................................................................................................    0
H1F1.BE.93.VI850             ...........................................................................................................................................................................    0
H1G.SE.93.SE6165             ...........................................................................................................................................................................    0
H1H.CF.90.056                ...........................................................................................................................................................................    0
H1J.SE.93.SE7887             ...........................................................................................................................................................................    0
H1K.CM.96.MP535              ...........................................................................................................................................................................    0
H101_AE.TH.90.CM240          ...........................................................................................................................................................................    0
H102_AG.NG.x.IBNG            ...........................................................................................................................................................................    0
H1N.CM.95.YBF30              ...........................................................................................................................................................................    0
H1O.BE.87.ANT70              -A---CT----C----G...........................................................................................GA--A-A-ACTGCTG................................................  345
H1O.CM.91.MVP5180            -A---CT----CCC--G..............................................................................................-A--TGACACTGC...............................................  343
CPZ.CD.90.ANT                ...........................................................................................................................................................................    0
CPZ.CM.01.SIVcpzCAM13        ...........................................................................................................................................................................    0
CPZ.CM.05.SIVcpzEK505        ...........................................................................................................................................................................    0
CPZ.CM.05.SIVcpzMB66         ...........................................................................................................................................................................    0
CPZ.CM.98.CAM3               ...........................................................................................................................................................................    0
CPZ.GA.88.GAB2               ...........................................................................................................................................................................    0
CPZ.CM.98.CAM5               ...........................................................................................................................................................................    0
CPZ.CM.05.SIVcpzMT145        ...........................................................................................................................................................................    0
CPZ.TZ.00.TAN1               ...........................................................................................................................................................................    0
CPZ.CM.05.SIVcpzLB7          ...........................................................................................................................................................................    0
CPZ.US.85.CPZUS              -A---------CG----..............................................................................................---TG.......................................................  331
CPZ.GA.88.GAB1               ---------A---AG--..............................................................................................----T.......................................................  331
MON.CM.99.L1                 ...........................................................................................................................................................................    0
MON.NG.x.NG1                 ...........................................................................................................................................................................    0
GSN.CM.99.CN166              ...........................................................................................................................................................................    0
GSN.CM.99.CN71               ...........................................................................................................................................................................    0
MUS_1.CM.01.1085             ...........................................................................................................................................................................    0
MUS_1.CM.01.CM1239           ...........................................................................................................................................................................    0
MUS_2.CM.01.CM1246           ...........................................................................................................................................................................    0
MUS_2.CM.01.CM2500           ...........................................................................................................................................................................    0
RCM.NG.x.NG411               ...........................................................................................................................................................................    0
RCM.GA.x.GAB1                ...........................................................................................................................................................................    0
DEN.CD.x.CD1                 ...........................................................................................................................................................................    0
                                      TCF-1 alpha
                             |---------------------|
MAC.US.x.239                 -A--AG-G--TGGA-G-AAG...TCAGGCCTGTCAGAGGAAGAG...GTTAGAAGAAGGCTAACCGCAAGAGGCCTTCTTAACATGGCTGACAAGAAGGAAACTCGC..---..........................................................  407
H2A.GW.x.ALI                 -A---G-A--TGG-T--AAG...TCAGGCCTGCCAGAAGATGAG...TGGAAGGCAAGACTGAAAGCAAGAGGGATACCATTTAGTAAGAACAGGAACAGC........---TTTGGTCA.................................................  409
H2A.DE.x.BEN                 -A--AG-G--TGG-C--AAG...TCAGGATTGCCAGAGAAAGAA...TGGAAGGCAAAACTGAAAGCAAGAGGGATACCATATAGTGAA....................-A---GGAACAACCATACTTGGTCA...................................  411
H2A.SN.x.ST                  ...........................................................................................................................................................................    0
H2B.GH.86.D205               -A--AG-G---GG-T--CAG...TCAGGGCTACCAGAGGAGGAG...TGGAAGGCTAGACTAAAAGCAAGAGGGATACCTACAGAT........................-AGGCAAGAGACAGCAGCATAAACAGGAACT.............................  417
H2B.CI.x.EHO                 -A---G-G--TGG-T-TCAG...TCAGGAATGCCAGAGAAAGAG...TGGAAGGCTAAACTGAGAGCAAGAGGAATACCTACAGAG........................-AGACAGGAAATAACAGCACAAAGAGGAACT.............................  414
H2G.CI.92.ABT96              ...........................................................................................................................................................................    0
H2U.FR.96.12034              ...........................................................................................................................................................................    0
MAC.US.x.EMBL_3              ...........................................................................................................................................................................    0
SMM.US.x.H9                  ...........................................................................................................................................................................    0
SMM.US.x.PGM53               -A--AG-G--TGGT-GTAAG...TCAGGCTTGTCAGGGGAAGAG...GTAAAGAGAAGGCTAACCGCAAGAGGCCTTTTAAAAATGGCTGACAAGAAGGAAACGAGC..---..........................................................  333
SMM.US.x.SIVsmH635F_L3       -A--AG-G--TGGT-GTAAG...TCAGGCTTGTCAGAGGAAGAG...GTAAAGAGAAGGCTAACCGCAAGAGGCATTTATAAAATGGCTGACAAGAGGGAAACAAGC..---..........................................................  407
SMM.SL.92.SL92B              ...........................................................................................................................................................................    0
STM.US.x.STM                 ....................................................GAGAAGGCTAACCGCAAGAGGCCTTCTAAAGATGGCTGACAAGAAGGAAACAAGC..---..........................................................   80
MNE.US.x.MNE027              ...........................................................................................................................................................................    0
DRL.x.x.FAO                  ...........................................................................................................................................................................    0
MND_2.x.x.5440               ...........................................................................................................................................................................    0
MND_2.CM.98.CM16             ...........................................................................................................................................................................    0
MND_2.GA.x.M14               ...........................................................................................................................................................................    0
OLC.CI.97.97CI12             TTA----AAGG...-CATGTTCAGACTTGCATAGGATG............TGGTTTGTGTGGTTAAGC...........................................-AGA-GACTTGGTTGCCTGGGAGACGGGGGTTGG....ACATGGGCGGGCT.........  318
GRV.ET.x.GRI_677             GCCC-......G-C-TGTTTAAGGACATGCACGAACAT.........GCAAAGCGC.......................................................-AGTGTCAG...................................................  341
VER.KE.x.TYO1_patent         ACCCCA-TCGGG--T--TTTACTGACATGCATGGGCTG.........GTGAAGAGGAAG....................................................-AG-CAGACCGCAAGCCTGCGGTTAGA.................................  368
VER.KE.x.9063                ACCCAA--AGAC--T-TTTTGAAGACATGCATGCAATA.........GTGAAGAGGAAG....................................................-AG-TAACCGCAGGCTTGTGGTTAAG..................................  367
VER.KE.x.AGM155              ACCCC-CAAGGG--T--TTTACAGACTTATATTCAACA.........GTTGGTACAGGAAAC.................................................-AG-CGACCACAGGCTTGCGGTTTCC..................................  370
VER.DE.x.AGM3                ...........................................................................................................................................................................    0
TAN.UG.x.TAN1                ACCC-......G-C-G-TTTAAGGACATGCATGGTCTT.........GTTAAGAGGAAG....................................................-AG-CTAACCGCAGGCTTGCGGTTAAGCTAACTAGGCCGTTGCCTAGGAGATGACATAAA  386
SAB.SN.x.SAB1C               TT-CT-GGAGGCT-C--CCAGAGCAGGTGCCGTCAGCA.........CAAGGC..........................................................-A--CACAAAACCACATCCTATG.....................................  352
MND_1.GA.x.MNDGB1            ...........................................................................................................................................................................    0
DEB.CM.99.CM40               ...........................................................................................................................................................................    0
DEB.CM.99.CM5                ...........................................................................................................................................................................    0
TAL.CM.01.8023               ...........................................................................................................................................................................    0
TAL.CM.00.266                ...........................................................................................................................................................................    0
SUN.GA.98.L14                G--CCATACAG---C--CCTGAAGAGTTCAAGCATGTT.........ACTAGTCTGCAGTGGGAGGTG...........................................-AG-TTAACCGCAAACCGCATCCTCTTGCATCGCCT.........................  383
WRC.CI.97.97CI14             TCCCTGGACAGTCA-T-GCCACTGGGAAAGAGGCCATTGAG...AGTCTCTATAAGAACAAGAAGCAGTGCAAG.....................................-AGAGACATGTTGCAGTAGGAAGTGGCTTTGCGGTTAAGCTACAT...............  422
WRC.CI.98.98CI04             TCCCTGGACAT-TA-TAGCCCAAGGAGAAGATGCTATTGAGACCCTCCATAAAGAGAAGCTGAAGCGGCACAAG.....................................-AGAGACATGTTGCAGTAGGATGTGGTTTTGTGGTTACGCTACAT...............  425
WRC.GM.05.Pbt_05GM_X02       ...........................................................................................................................................................................    0
LST.CD.88.524                ...........................................................................................................................................................................    0
LST.CD.88.447                ...........................................................................................................................................................................    0
LST.KE.x.lho7                TA-CCCT-AAA---C-GCCTGCGAAGATACAAAATACT.....GCTTTCGCTTTTGACTGCAAGAGGAAG.........................................-AG-CAAACCGCAAACCGCAGCACATCCTCTTGCAGCCCGGTTGCTAA............  401
SYK.KE.x.SYK173              GCTG-G-GA-GGCCCG-TTGCAGTTAGAGAGACAAACA.........GGAAAACCACAAGAGCTGCAGTCAGCACTTAGTAAGAACATCTCC..................-AGG-GACTCCATGGTGACAAGCTCGG.................................  289
SYK.KE.x.KE51                ...........................................................................................................................................................................    0
COL.CM.x.CGU1                ...........................................................................................................................................................................    0
```

```
                                                     NF-κ-B-II                        NF-κ-B-I            Sp1-III         Sp1-II     Sp1-I                  TATA Box
H1B.FR.83.HXB2              ....................TCGAGCTTGCT...ACA...AGGGACTTTCCG......................CTGGGGACT...TTCCAG............GG.AGGCGTGG.CCTGGGCGGGACTGGGGAGTGGCGA.......GCCCT..CAGATCCTGCATAT  429
H1A1.UG.85.U455             .........................................................................................................................................................................  0
H1B.US.90.WEAU160           ....................-----T--T--...---...-----------A......................---------...----G-............--.-------A.T------------A-----------.......-----..-----G--------  429
H1C.ET.86.ETH2220           .........................................................................................................................................................................  0
H1D.CD.84.84ZR085           .........................................................................................................................................................................  0
H1F1.BE.93.VI850            .........................................................................................................................................................................  0
H1G.SE.93.SE6165            .........................................................................................................................................................................  0
H1H.CF.90.056               .........................................................................................................................................................................  0
H1J.SE.93.SE7887            .........................................................................................................................................................................  0
H1K.CM.96.MP535             .........................................................................................................................................................................  0
H101_AE.TH.90.CM240         .........................................................................................................................................................................  0
H102_AG.NG.x.IBNG           .........................................................................................................................................................................  0
H1N.CM.95.YBF30             .........................................................................................................................................................................  0
H1O.BE.87.ANT70             ACCTGAAGATTGCTGACACTG-...................--A-------A..GCAAAGACTGCTGACACTGC-------..------............TG--A-G-ACAGG------TT-G---AGTG-C........TAA----.-----AG--------  464
H1O.CM.91.MVP5180           .........................................----------A.......GACTGCTGACACTGC-------..------.............C-TG--A-G-ATAAG------TT-G---AGTG-C........TAA----.-----G--------  437
CPZ.CD.90.ANT               .........................................................................................................................................................................  0
CPZ.CM.01.SIVcpzCAM13       .........................................................................................................................................................................  0
CPZ.CM.05.SIVcpzEK505       .........................................................................................................................................................................  0
CPZ.CM.05.SIVcpzMB66        .........................................................................................................................................................................  0
CPZ.CM.98.CAM3              .........................................................................................................................................................................  0
CPZ.GA.88.GAB2              .........................................................................................................................................................................  0
CPZ.CM.98.CAM5              .........GCGCCTGCGCAG-A--AAC----G..---A..---------TA.A.......................------..-----A..............-GG--A-GCTCA..--------.ACA---C----TTTC.....A----.----AG--------  105
CPZ.CM.05.SIVcpzMT145       .........................................................................................................................................................................  0
CPZ.TZ.00.TAN1              .........................................................................................................................................................................  0
CPZ.CM.05.SIVcpzLB7         .........................................................................................................................................................................  0
CPZ.US.85.CPZUS             GCTGTAACCGCGCAGGCGC.AATA-AAC----G..-..CTG.A--------TAA....GGGACTTTCCAA.......-----G..----.............AAGG--GTG--T-A.------A.ACA---C----TTTT.....A----.--A-GA-------A  456
CPZ.GA.88.GAB1              TTAG.ACTGGCGCATGCGC.ACAAGAAC----G..--TCTGC----------AA.....................------..----G-............--AGA-GTG-TCGGGA------GT-----T-----TTT.....------.----.AG--------  449
MON.CM.99.L1                .........................................................................................................................................................................  0
MON.NG.x.NG1                .........................................................................................................................................................................  0
GSN.CM.99.CN166             .........................................................................................................................................................................  0
GSN.CM.99.CN71              .........................................................................................................................................................................  0
MUS_1.CM.01.1085            .........................................................................................................................................................................  0
MUS_1.CM.01.CM1239          .........................................................................................................................................................................  0
MUS_2.CM.01.CM1246          .........................................................................................................................................................................  0
MUS_2.CM.01.CM2500          .........................................................................................................................................................................  0
RCM.NG.x.NG411              .........................................................................................................................................................................  0
RCM.GA.x.GAB1               .........................................................................................................................................................................  0
DEN.CD.x.CD1                .........................................................................................................................................................................  0
                                                     NF-κ-B-II                               NF-κ-B-I                                                                     TATA Box
MAC.US.x.239                ...................................-ACAGC-----------A....................CAA--GGA...-GTT-CGGG..GAGGTACT--GGA-GAGCCG.G-C--GAACG.......................---ACTTTCT-GA--T---A  492
H2A.GW.x.ALI                GGGCAGGAAGTAACTACTG..AAA-C.AGCTGA..GACTGC-----------A....................GAA--GGC...-GTA-C............CAGG--AAG-AC.A----A--AG----T-G-GAA-..............-----CAT-CTC.---T---A  525
H2A.DE.x.BEN                AGGCAGGAAGTAGCTA.CTAAGAA-C.AGCTGA..GGCTGC-----------A....................GAA--GGC...-GTA-C............CAAG--A-G-AC.A----A--AG----T-G-GAA-..............-----CAT-CT-A---T---A  529
H2A.SN.x.ST                 .........................................................................................................................................................................  0
H2B.GH.86.D205              AGCTGACACTGCACAA.GAAGGAA-CTAGCAGA..CACTGC-----------A....................AAA-AGGC...-GTA-C............A.TG--A-GA-C.AA---A---GT---TTG-GAA-.............----ATTA-ATCT---T---A  535
H2B.CI.x.EHO                AGCTAACACTGCATAGAGAAGGAA-CTAGCTGA..TACTGC---.-------A....................GAA--GGC...-GTA-C............A-TG--A-G-AC.A----A--AG-A---AG-.AA-.............----ATCA---CT---T---A  532
H2G.CI.92.ABT96             .........................................................................................................................................................................  0
H2U.FR.96.12034             .........................................................................................................................GAA-..............-----CAA-TGAT---T---A  25
MAC.US.x.EMBL_3             .........................................................................................................................................................................  0
SMM.US.x.H9                 .........................................................................................................................................................................  0
SMM.US.x.PGM53              ..................................GACAGC-----------A....................CAAA-GGGA..-GTT-T.............--GGA-GTACTG..G--AG-AACTG-TT-GACC...............----ACTT-TT-T---T---A  418
SMM.US.x.SIVsmH635F_L3      ..................................GACAGC-----------A..CAAGGGACTTTCCA.....CAA--GGC...-GT--TGGG..GAGGTACT--GGA-GAGCTG.GCT--.AACG.........................---ACTT-TTCT---T---A  505
SMM.SL.92.SL92B             .........................................................................................................................................................................  0
STM.US.x.STM                ..................................GACAGC-----------A....................CAA--GGC...-GTA-............CA--GGA-G..TAC.TG--AG-A-CTG-T--G-.AA-.............-----TTT-CT-T---T---A  165
MNE.US.x.MNE027             .........................................................................................................................................................................  0
DRL.x.x.FAO                 .........................................................................................................................................................................  0
MND_2.x.x.5440              .........................................................................................................................................................................  0
MND_2.CM.98.CM16            .......................................................................................-GTCGGGTGGGGGAGTG-TC-AC-CAACCTGGCT--ATATAA-CA-CTC-CTTTGCGCTTGAAGTGAGTCTC---CTG-G-G  82
MND_2.GA.x.M14              ....................................................................................................................................................................G-G-G  5
OLC.CI.97.97CI12            A................................T..CCTT--TTG-----G-TTTTGCCC........................................................................................................T---A  349
GRV.ET.x.GRI_677            ................................-ACTTTGCGGTT-----------..CCAGGGACTTTCCACA.....-T--GTGG.................ATCG-AGGC-GTACAGGGGC--TACTGGGA-T-GC-T.............T-CCCTCAGAG-------A  444
VER.KE.x.TYO1_patent        ACATCACCATGGAGATGACAT-AA-AAC----GAC.......----------AGCGAAGGGACTTTCCA.................................A--CG--ACAT-.GGC--T.CC-GGGA-T-GC-T.............TACCCTCAGAA-------A  481
VER.KE.x.9063               CCGTTGCCAGGGAGATGACAT-T--AAC----GAC-G.......----------A...AGGGACTTTCCA.................................G--CG---CAT-.GGC--T.AC-GGGA-T-GT-T.............TACCCTCAGAG-------A  478
VER.KE.x.AGM155             TGGTTGCCTAGGAGATGACAT-AAGAAC----GAC........----------A.GCACGGGACTTTCCA.................................A--CG--ACAT-.GGC--T.AC-GGGA-T-GC-T.............TACCCTCAGAG-------A  482
VER.DE.x.AGM3               .........................................................................................................................................................................  0
TAN.UG.x.TAN1               GAAGTGC.........................-GA..TCAGC-----------A.......................................G--TG-TGCGA-.GGC--TACTTGGGA-T-GC-T..............-CCCTCAGAG-------A  464
SAB.SN.x.SAB1C              GAGTTGTCATGGTGATGACAT-AAGAAC----GAC..TGAG-----------A.......................GGG...-GGAGA............CTGG-CG--ACTGGGA-T...........-GC-T.............G-CCCGT-C-G--CA----  454
MND_1.GA.x.MNDGB1            ....................................................................................................................................................GGAGTCTCTACTACAGA  17
DEB.CM.99.CM40              .........................................................................................................................................................................  0
DEB.CM.99.CM5               .........................................................................................................................................................................  0
TAL.CM.01.8023              .........................................................................................................................................................................  0
TAL.CM.00.266               .........................................................................................................................................................................  0
SUN.GA.98.L14               AGGCAACGGGGCTAGC.GCAT......GC--GC..TAGAGT------------...................................................GACAG--AGGG--A-GC-T...............ACC--CT-A--------  463
WRC.CI.97.97CI14            CCTGAGACTGTGTACCAACGGA-TTAGCAA-CATGCT-ATG-ACTGGGACATTGGGAGGAGACT.GGGAGGTGC--T-T.......................................................GGTTT-G-CCT---A  515
WRC.CI.98.98CI04            CCTGGACCTGTGTACCAACGAA-TTAGCAA-CATGCT-ATGGACTGGGAC-T-G.GTGGAGACTTGGGAGGAGA--T-T......................................................GGTTT-G-CCT---A  518
WRC.GM.05.Pbt_05GM_X02      .........................................................................................................................................................................  0
LST.CD.88.524               .........................................................................................................................................................................  0
LST.CD.88.447               .........................................................................................................................................................................  0
LST.KE.x.lho7               GGCAACCGGGCTAGCGCATGCG-A-TGGCTTGCTG--GAGC-----------A................GGACGGGC---GGAGG.--T....................................GCC--CT-G--------  490
SYK.KE.x.SYK173             CCCAGGGGA..............................-----C-GGGC...................GGGG-AA-G--...CA--CT.............-CTGC...............................................---A  335
SYK.KE.x.KE51               .........................................................................................................................................................................  0
COL.CM.x.CGU1               .........................................................................................................................................................................  0
```

```
                                                                                  +1 mRNA start site
                                              TAR element start                                                                                 TAR element end
                             TATA Box                                                         5' LTR U3 end 5' LTR R repeat begin
H1B.FR.83.HXB2               AAGCAGCTGCT..TTTTGCCTGTACT........................................................GGGTCTCTCTGGTTAGACCAGA.TCTGAGCCTGGGAG..CTCTCT..GGCTA.ACTA..GGG.AACCCACTGCTTA.....AGCCT..  525
H1A1.UG.85.U455              ..........................................................................................................................................................................    0
H1B.US.90.WEAU160            -----------..-------------........................................................----------------------.--------------..------..-----.G---..---.-------------.....-----..  525
H1C.ET.86.ETH2220            ..........................................................................................................................................................................    0
H1D.CD.84.84ZR085            ....................................................................................................................---..------..---CG.G---..---.-------------.....-----..   39
H1F1.BE.93.VI850             ..........................................................................................................................................................................    0
H1G.SE.93.SE6165             ..........................................................................................................................................................................    0
H1H.CF.90.056                ..........................................................................................................................................................................    0
H1J.SE.93.SE7887             ..........................................................................................................................................................................    0
H1K.CM.96.MP535              ..........................................................................................................................................................................    0
H101_AE.TH.90.CM240          ..................................................................................----------T---------G.--.----C-----.------.------.G-A-..---.------------A.........------..   72
H102_AG.NG.x.IBNG            ......................................................................................................----.------.------.G-GG..A--.-------------.......------..   54
H1N.CM.95.YBF30              .............--C-CG--TGTACT.......................................................----------T-C-G-------.-TA-----------.------.------.G-.---.------------------.......------..   85
H1O.BE.87.ANT70              -------C---..--C---T-----C........................................................------GG-TAGAG-----G.-------C-----.---C--.---CT.CTAG..CT-.-----G-------.....-CG--..  560
H1O.CM.91.MVP5180            -----------..--CC--T---.ACC.......................................................-----TAG-TAGAG-----G.-------C-----.---C--.---CT.CTAG..CT-.-----G-------.....-CG--..  533
CPZ.CD.90.ANT                ..........................................................................................................................................................................    0
CPZ.CM.01.SIVcpzCAM13        ......................................................................................------.-CAC-G-------.-------C-----.-------..-C-AGTG....A--.-------------.......------..   67
CPZ.CM.05.SIVcpzEK505        ......................................................................................--------T----------.-TA-----------.------.-----GCA...A--.-------------.......------..   69
CPZ.CM.05.SIVcpzMB66         ......................................................................................--------T----------T---------.----.------.-----GCT.-.A--.-------------.......------..   70
CPZ.CM.98.CAM3               ..........................................................................................................................................................................    0
CPZ.GA.88.GAB2               ..........................................................................................................................................................................    0
CPZ.CM.98.CAM5               -------C---..--C---T-----G........................................................T--------T--G-------.-TA-----------.-----.--.GCC-GC-..---.------------------.......------..  200
CPZ.CM.05.SIVcpzMT145        ......................................................................................--------T-G--------.---------------.-----.-----ACA.-..---.------------.......------..   70
CPZ.TZ.00.TAN1               ...........................GCTCTTGCCTAATCTG.......................................................----.--------------.------.--TAGTGGCTGGCTAG-GA--G-------.....-CG--..   75
CPZ.CM.05.SIVcpzLB7          ......................................................................................--------T----------T---------.----.------.-----GCT.-.A--.------------A.........------..   71
CPZ.US.85.CPZUS              -------C---..C-C---T------........................................................--------TC-C-AGA.-----.-AA-----C-----.------.------GTG--..---.------------------.......------..  550
CPZ.GA.88.GAB1               -----------..--C----------........................................................--------T-CAC-G-------.-T---------------.------.A.-C-GTGA-..---.-----------------.......------..  544
MON.CM.99.L1                 ........................................................................................................AG-G-CT-G-CTCCA.CAG-GGC.CT-GG-GC-CTTG-A-CT-GGT-AA--GGCCTGGG----CTT   64
MON.NG.x.NG1                 ..........................................................................................................................................................................    0
GSN.CM.99.CN166              ...........................................................................................................TGCTT-G-CTTTCACAG-GGC..............CTGGGAG-C--TG..GAGCTAG-TAAAG   47
GSN.CM.99.CN71               ...........................................................................................................TGCCT--GTCTCCACAG-GGC..............CTGGGAG-C--TG..GAGCTAG-TAGAG   47
MUS_1.CM.01.1085             ...........................................................................................................AGT-CTAGCTCC-.CAG-GGC..............CTGGGAG-C--CG..GAACTAG-TA-AG   46
MUS_1.CM.01.CM1239           ..........................................................................................................CAGTTCTAGCTCC-.CAG-GGC..............CTGGGAG-C--CG..GAACTAG-TAAAG   47
MUS_2.CM.01.CM1246           ...........................................................................................................TCTTCTG-CTCC-.CAG-GGC..............CTGGGGG-C--CA..GAACTAG-TAGAG   46
MUS_2.CM.01.CM2500           .......................................................................................................-TC-TCTCCTG-CTCC-.CAG-GGC..............CTGGGAG-C--CG..GAGCTG--TAGAG   50
RCM.NG.x.NG411               ..........................................................................................................................................................................    0
RCM.GA.x.GAB1                ..........................................................................................................................................................................    0
DEN.CD.x.CD1                 .............................................................................................G-AC-T-CTTGCGAGTCTCT--A-C-GCCTGGGAGCCCT-AG-GG-CTTA-C...-G-GG-CTGGG-GCCCTCTAAGGT   76

                                                         +1 mRNA start site
                                                   TAR element start
                             TATA Box         5' LTR U3 end 5' LTR R repeat begin
MAC.US.x.239                 -TATCA----A.T--CGCT-----T-C...AGTCGCTCTGCGGAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCT.CCAGCAC-A..G-A---.---G-CTGGGTGTTC----CT--.A-----A.............CCAGC--TTGGCCG..GTGCTGG--AGAG  637
H2A.GW.x.ALI                 -T-T-C-C---.GC--GCAT----T-C...AGTCGCTCTGCGGAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCT.CCAGCAC-A..G-A---.---G-CTGGGTGTTC----CT--.A-----A.............CCAGTG-TTGGCCG..GCACTGG--AGAC  670
H2A.DE.x.BEN                 -T-T-C-C---.TC--GCAT----T-C...AGTCGCTCTGCGGAGAGGCTGGCAGATCGAGCCCTGAGAGGTTCTCT.CCAGCAC-A..G-A---.---G-CTGGGTGTTC----CTG-.A-----A.............CCAGT--TTGGCCG..GTACTGG--AGAC  674
H2A.SN.x.ST                  ..............................AGTCGCTCTGCGGAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCT.CCAGCAC-A..G-A---.---G-CTGGGTGTTC----CT--.A-----A.............CCAGTG-TTGGCCG..GCACTGG--AGAC  119
H2B.GH.86.D205               -T-T-C-C---.TC-CGCAT----T-C...AGTCGCTCTGCGGAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCT.CCAGCAC-A..G-A---.---G-CTGGGTGTTC----CT--.A-----A.............CCAGC--TAGGCCA..GTGTTGG--AGAC  680
H2B.CI.x.EHO                 -T-T-C-C---.T-G-GCAT----T-C...AGTCGCTCTGCGGAGAGGCTGGCAGATCGAGCCCTGGGAGGTTCTCT.CCAGCAC-A..G-A---.---G-CTGGGTGTTC----CT--.A-----A.............CCAGC--TTGGCCG..GTGCTGG--AGAC  677
H2G.CI.92.ABT96              ................................................................................................................---CT--.A-----A.............CCAGCG-TTGGCCG..GTGCTGG--AGAG   41
H2U.FR.96.12034              -T-T-C-C---CT--.GCAT----T-C...AGTCGCTCTGCGGAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCT.CCAGCAC-G..G-A---.---G-CTGGGTGTTC----CT--.A-----A.............CCAGT--TTGGCCG..GTACTGG--AGAG  170
MAC.US.x.EMBL_3              ..........................C...AGTCGCTCTGGC.AGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCT.CCAGCAC-A..G-A---.---G-CTGGGTGTTC----CT--.A-----A.............CCAGC--TTGGCCA..GTGCTGG.-AGAG  118
SMM.US.x.H9                  ..........................C...AGTCACTCTGCGGAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCT.CCAGCAC-A..G-A---.---G-CTGGGTGTTC----CTG-.A-----A.............CCAGC--TTGGCCG..GTGCTGG--AGAG  120
SMM.US.x.PGM53               -TA--A----A.T--CGCT-----T-C...AGTCGCTCTGGGCAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCT.CCAGCAC-A..G-A--C.---G-CTGGGTGTTC----CT--.A-----A.............CCAGC--TTGGCCG..GTGCTGG--AGAG  563
SMM.US.x.SIVsmH635F_L3       -TA--A----A.T--CGCT-----G-C...AGTCGCTCTGCGGAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCT.CCAGCAC-A..G-A---.---G-CTGGGTGTTC----CT--.A-----A.............CCAGC--TTGGCCG..GTGCTGG--AGAG  650
SMM.SL.92.SL92B              ...............................GTCGCTCTACGGAGAGGCTGCCAGA.TAGAGCCCCGAGAGGTTCTCT.GCAGCAC-TA.GTA---.--G-CTGGGTGTTCT--ACT-A.A-----A............ACCAGC-A-AGGCCA..TTGCTGG-TAGAC  119
STM.US.x.STM                 -TA--A-C---.TA-AGCT-----T-C...AGTCACTCTGCGGAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCT.CCAGCAC-A..G-A---.---G-CTGAGTGTTC----CT--.A-----A.............CCAGC--TTGGCCA..GTGCTGG-TAGAG  310
MNE.US.x.MNE027              ..............................AGTCGCTCTGCGGAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCT.CCAGCAC-A..G-A---.---G-CTGGGTGTTC----CT--.A-----A.............CCAGC--TTGGCCG..GTGCTGT--AGAG  119
DRL.x.x.FAO                  ..........................................................................................................................................................................    0
MND_2.x.x.5440               ..........................................................................................................................................................................    0
MND_2.CM.98.CM16             GCTACCAGAT-.GAGCCTGGGTGTTCTCTGGTGAGTCTT.GGAAGGAGTGCCTGCTTGTAGCCCTGGGTGGTTCGCAGGCCCCT-G--TGTA-C-CT-GG---CTCG-C--GTGTTCCTGAAGAGGCTCAT-GGGGGCTCCCTTGCTTGGT--TGG--GACCTCTAGCAG  250
MND_2.GA.x.M14               GCTACCAGAT-GAGCC--GG-TGTTCTCTGGTGAGTCCT.TGAAGGAGTGCCTGCTTGTAGCCCTGGGCGGTTCGCAGGCCCCT-G--TGTA-C-CT-GGT--CTCG-C--GTGTTCT-GAAAG.G-CTC----AGTG-GCACTCTTG-TTGG..--TTGGTAGA---CT  171
OLC.CI.97.97CI12             ---TCTG----..C-GC--T---TAGA...GGTCT.TGCTCGGAGA..CCAGCC.............................................T-GA-------A-T-CTT--GGCAGCCTGCTTTG-GC-CCTGGGTGTGCTTGGCAGGTGGTTAAAG  461
GRV.ET.x.GRI_677             ------A----.CGC-G--T----ACTC..AGTCTCTTACTAGGAG....ACCAGCTAGAGCC.TGGGT.GTTCGC.................................---TTA.G-CTAAC.CC-G-.TGGCCA.CC-GGGGTA-GGA--..C...........  551
VER.KE.x.TYO1_patent         ------A----CGC-GGC.T----ACTC..AGTCTCTTACTAGGAGA......................................C-A-C--.--G-CTGGGTGTTCG----TTA.G-CTAAC.CT-G-.TGGCCA.CCAGGGGTA-GGA--..C...........  588
VER.KE.x.9063                ------A----.CGC-G--T----ACTC..AGTCTCTTACTAGGAGA......................................C-A-C--.--G-CTGGGTGTTCG----TTA.G-CTAAC.CT-G-.TGGCCA.CCAGGGGTA-GGA--..C...........  585
VER.KE.x.AGM155              ------A----.CGC-G--T----ACTC..AGTCTCTTACTAGGAGA......................................C-A-C--.--G-CTGGGTGTTCG----TTA.G-CTAAC.CT-G-.TGGCCA.CCAGGGGTA-GGA--..C...........  589
VER.DE.x.AGM3                ..........................C...AGTCTCTTACTAGGAGA......................................C-A-C--.--G-CTGGGTGTTCG----TTA.G-CTAAC.CT-G-.TGGCCA.CCAGGGGTA-GGA--..C...........   81
TAN.UG.x.TAN1                ------A----.CGC-G--T----ACTC..AGTCTCTTACTAGGAGA......................................C-A-C--.--G-CTGGGAGTTCG----TGAGC--AA-C.T--TC.TGGCCATCCAGGGGTA-GA--C-.C...........  574
SAB.SN.x.SAB1C               -----------.--GC-------AAAC..AGTCTCTGCTTGGAG.G.CTGCCAGATTGAGCC.TGGGT.GTTCTCTGGTAA-TC-GAAC-A-C--.--G-CTGGGTGTTCG-----GAAGGCAAGC.CTAAG-GCACTCT--TG-GGT-TTG-TA..GACCCTCG-CTG  612
MND_1.GA.x.MNDGB1            GG...CTAAGGGT-G-AT-TCTG-G.CAGATCCCCTTAGAGCAAGGACCAGAGTCCT.............GAGTGACTGGGTCT-AGCAC--CA.CTCGGGGCTGATCACCT-GA--TAGTGGAA--CCT..-GCTTG..............................  135
DEB.CM.99.CM40               ........................................................................CGGGAGTCGCGCCTTG-GAGTCCTGGT--GT---GCGGCCTGGGA-CCCTTG.---ACT--GATAG-CT.A-CTCC-G-GG-CTGGG-GCCCTGAG--GG   98
DEB.CM.99.CM5                ........................................................................CGGGAGTCGCGCCTAG-GAGTCCTGGTAGGT---GCGGCCTGGGA-CCCTTGC--GCTA-TGACT--CT.A-CTCC-GTGG-CTGGG-GCCCTGAG--GG   99
TAL.CM.01.8023               ..........................................................................................................................................................................    0
TAL.CM.00.266                ...................................................................................................................T-TCTG-AGAG-GGTCAC-GCG-CCT---AG---TTGGC-ACCGGTAA---G-CC   55
SUN.GA.98.L14                -----------.C--CGC.----T-GGG..AGTCTCTGCTCAGAGG....CTAGGCGACGACCCT.......................GAGA-G-CTCA-CT.AGAGCT-G--T-TC-.CTC--GA.............GT-AGC--TTG---A..GCACCTCA--CGA  584
WRC.CI.97.97CI14             ---T--A----..C-G-AG--A-CT............................................GGTTGCAGTTTTCTGCCTGA--CTAGGTGCC-GC-TAGT-GTC--AC-TGGCCT--CA----GAGT-C--GAGTGGTG-TCCC--TTGGCT-AG-GACTC--GTAGG  645
WRC.CI.98.98CI04             ---T--A----..C-G-AG--A-C.............................................CTTTGCAGTCTTCTGCCTGA--CTAGGTGCC-GC-TAGT-GTC--AC-TGGCCT--CT----GAGT-C--GAGTGGTG-TCCC--TTGGCT-AG-GACTC--GTAGG  647
WRC.GM.05.Pbt_05GM_X02       ..........................................................................................................................................................................    0
LST.CD.88.524                ..........................................................................................................................................................................    0
LST.CD.88.447                ..........................................................................................................................................................................    0
LST.KE.x.lho7                -----------..C-CC--T---T-GG...GAG...................................TCCTT.GCTCT--GGA-CAGCAG-C-CT-CTGGGT.----ACACTG-.T-AGCTA.A-CAGTC-GCCCT---AGG-TACACTTA..GCACCCA---GG  610
SYK.KE.x.SYK173              ----C-A----..C-G-AG--A-C-GCGG.AGTTGCAACTAGCTCTCTCCAGCGGCCTGGGAGCCCTGAGAGAC............................................TGG-AG.A-CGGCCCGGGAGCCCTTG--AGT......AAAGGCTAG-GA  448
SYK.KE.x.KE51                ..........................CGA..AGTTGCAGCTAGCTCTCCTCAGCGGCCTGGGAGCCC.TAGGAGAC............................................TGG-AG.A-CGGCCTGGGAGCCCTTG--AGT.....-AAAGGCTAG-AA   90
COL.CM.x.CGU1                .............................TGGAGTTCTGACTCTT..........................................G-CT-GCT-CG-CTGAG-CCT-AG---CTTGC-AGCTA.TCAGCCT.TGCCT-A-TCTGAGTAGCT-G.....CTTGG-TG   90
```

```
                                                                                                                                                                          Extensive secondary structure
                                                                                                       Poly-A signal          5' LTR R repeat end  5' LTR U5 start
H1B.FR.83.HXB2              ................................................................................................C.AATAAAGC..TTGCCTTG..AGT....GCTTCA.AGTA...GTGTGTGCCCGTCTG.TT.GT........GT  576
H1A1.UG.85.U455             ......................................................................................................----..--------..---....----G-.----...---------------.--AT-.........-   46
H1B.US.90.WEAU160           ................................................................................................-.--------..--------..---....------.----...---------------.--.--........--  576
H1C.ET.86.ETH2220           ..........................................................................................................................................................................    0
H1D.CD.84.84ZR085           ................................................................................................-.--------..--------..---....----T-.----...---------------.--TT.........--   90
H1F1.BE.93.VI850            ..........................................................................................................................................................................    0
H1G.SE.93.SE6165            ..........................................................................................................................................................................    0
H1H.CF.90.056               ..........................................................................................................................................................................    0
H1J.SE.93.SE7887            ..........................................................................................................................................................................    0
H1K.CM.96.MP535             ..........................................................................................................................................................................    0
H101_AE.TH.90.CM240         ................................................................................................-.--------..--------..---....----A-.---G...---------------.-G.T-........AG  123
H102_AG.NG.x.IBNG           ................................................................................................-.--------..--------..---....------.----...---------------.A-T--........--  106
H1N.CM.95.YBF30             ................................................................................................-.--------..--------..---......--AA-..---G...----------A--CA.--C-........--  135
H1O.BE.87.ANT70             ................................................................................................-.--------..--------..---GAGAAGCA...............--------T-A----.--CAACC........  610
H1O.CM.91.MVP5180           ................................................................................................-.--------..--------..---GAGAAGCA...............--------T-A----.--CAACC........  583
CPZ.CD.90.ANT               ..........................................................................................................................................................................    0
CPZ.CM.01.SIVcpzCAM13       ................................................................................................-.--------..--------..---....------.----...----------TA----.--CA........--  118
CPZ.CM.05.SIVcpzEK505       ................................................................................................-.--------..--------..---....----.-.----...-----------A--CA.--C-........--  119
CPZ.CM.05.SIVcpzMB66        ................................................................................................-.--------..--------..---....AC-T-.-----..--------T-A--CA.--GT........--  121
CPZ.CM.98.CAM3              ..........................................................................................................................................................................    0
CPZ.GA.88.GAB2              ..........................................................................................................................................................................    0
CPZ.CM.98.CAM5              ................................................................................................-.--------..--------..---....AC-TG.--C-..---------TA--CA.--CA........--  251
CPZ.CM.05.SIVcpzMT145       ................................................................................................-.--------..--------..---......--G--.-----..-------------A.--CA........--  121
CPZ.TZ.00.TAN1              ................................................................................................-.---------.C----G..--A....-TG-T..-AC-..----------A-T-CA-ACCGCGT.....C-  129
CPZ.CM.05.SIVcpzLB7         ................................................................................................-.--------..--------..---......--G--.-----..---------A-TCA.C-TT........--  122
CPZ.US.85.CPZUS             ................................................................................................-.--------..--------..---......--T-.-----..---------TA--C.--CA........--  600
CPZ.GA.88.GAB1              ................................................................................................-.--------..--------..---...GTA--TG.--C-..---------TA---A.--CA.........-A  596
MON.CM.99.L1                G.....................CCTAGAGAGTTGGTGAAGCTGACCCCGGCTCGCTTG..CT.T........TAGTA...AGGCCTTACAGCC..-.-------...........................................................................  130
MON.NG.x.NG1                ..........................................................................................................................................................................    0
GSN.CM.99.CN166             CGGCCTGGGAGCCCTTT.....GCTAG..........AAGACGGTAGAATTTCGCTTG..CT.T...GCCTAAGGCT...TATGACT..........-------.CT-CCAAAG-..-TAA.CGAG-CTGCTC-C....AT-CATTGG-.TC-AGGTCGGGCTTG....  168
GSN.CM.99.CN71              CGGCCTGGGAGCCCTTGC...CTAGAG...........GACGGTAGATTCTCGGCTTG..CT.T...GCCCTAGGGC...TTTAGTGGCT....-.-------.CT-C-AAAG-..TACAGCAAG-CTGCTC-C....AT-CATTGG---G-GA-C-GGCTTGG....  176
MUS_1.CM.01.1085            CGGCCTGGGAGCCCTTG.....CCTAG...........GAGTACTGATTCTCGGCTTG..CT.T..TGCTTCGAAGC...TTATTG.AAGCCCA..-------.G......................................................................  123
MUS_1.CM.01.CM1239          CGGCCTGAGAGCCCTTG.....CCTAG...........GAGAACTGGTTCTCGGCTTG..CT.T..TGCTTCGAAGC...TTATTG..AGCCCA..-------TAG-A-AACC.AAG-CAAGTCT..............................................  144
MUS_2.CM.01.CM1246          CGGCCTGAGAGCCCTTG.....CCTAG........GGGTTGAGACCCCGGCTCGCTTG..CT.T...GCAGAAA...GGCTTTTACATAGCC...-------TCC---AAAC.AACC-GCGA-TC-A...............----TCA-.C-GT--GGCGTCTAAG  172
MUS_2.CM.01.CM2500          CGGCCTGGGAGCCCTTG.....CTCAG...........AAGGAATAGACCCCGGCTTG..CT.T...GCTTACTTCAGGCTTTATC..AGCC...-.-------.CT---AAC-.AACT-GCGA-TC-A...............----TCA-.CA-T-GCGGCGC....  169
RCM.NG.x.NG411              ..........................................................................................................................................................................    0
RCM.GA.x.GAB1               ..........................................................................................................................................................................    0
DEN.CD.x.CD1                ATTG...............................CAACAAGTACCCAGACAAGCTTG..CTCG.....TGCTGCTTTTGATGACCATA...........--------.TCA-TTAA.................C-GGCATCT---CTTT-CTCCTCATTAGTGGGTTCC.  178
                     TAR element end                                                                   Poly-A signal      5' LTR R repeat end 5' LTR U5 start
MAC.US.x.239                T..GACTC.CACGCTTG.....................................CTTG..CT.T....AAAG........CCCTCTT........-----------.-.---A-T..TTAGAAGTAAG-T.-.GT..------T---A---.C-CCTAGCCGCCGCC-  731
H2A.GW.x.ALI                ...GGCTC.CACGCTTG.....................................CTTG..CT.T....AAAG........ACCTCTT........-----------.-.---AAT...TAGAAGCAGGTT.-AAG..------T---A---.C-CCTAGTCGCCGCC-  762
H2A.DE.x.BEN                ...GGCTC.CACGCTTG.....................................CTTG..CT.T....AAAG........ACCTCTT........-----------.-.---A.-..TTAGAAGCAAGTT.-AGT..------T---A---.C-CCTAGTCGCCGCC-  767
H2A.SN.x.ST                 ...GGCTC.CACGCTTG.....................................CTTG..CT.T...AAAAG........ACCTCTT........-----------.-.---AGT...TAGAAGCAAGTT.-AGT..-----CT---A---.C-CCTAGTCGCCGCC-  212
H2B.GH.86.D205              ...GGCTC.CACGCTTG.....................................CTTG..CT.T....AAAA........CCCTCTT........------.-.-.---A.A..TTAGAAGCAAGT..-AGT..------T---A---.C-CCTAGTCGC.GCC-  769
H2B.CI.x.EHO                ...GGCTC.CACGCTTG.....................................CTTG..CT.T....AAAA........CCCTCTT........------A-..-.-----.A..TTAGAAGCAAGAC.-AGT..------T---A---.C-CCTAGTCGCCGCC-  769
H2G.CI.92.ABT96             Y..GGCTC.CACGCTTG.....................................CTTG..CT.T..ACAAAG........ACCTCTT..........---------.-GC-AAAT...TAGAAGTAAG--AGTGT..------T---A---.C-CCTAGTCGCCGCC-  138
H2U.FR.96.12034             T..GACTC.CACGCTTG.....................................CTTG..CT.T....AAAA........CCCTCT.........----------.-.---AGT...TAGAAGCAAG-TA-.GT..------T---A---.C-CCTAGTCGCCGCC-  263
MAC.US.x.EMBL_3             T..GGCTC.CACGCTTG.....................................CTTG..CT.T....AAAG........ACCTCTT........----------.C-.---CATTT.TAGAAGTAAG-C.-.GT..----C-T---A---.C-CCTAGTCGCCGCC-  213
SMM.US.x.H9                 T..RGCTC.CACGCTTG.....................................CTTG..CT.T...AAAAG........ACCTCTT........----------.C-.---AGT...TAGAAGTAAG-C.-.GT..------T---A---.C-CCTAGMCGCCGCC-  214
SMM.US.x.PGM53              C..GGCTC.CACGCTTG.....................................CTTG..CT.T....AAAG........ACCTCTT........-----------.-.---GCT...TAGAAGTAAG.C.-AGT..------T---A---.C-CCTAGTCGCCGCC-  656
SMM.US.x.SIVsmH635F_L3      T..GGCTC.CACGCTTG.....................................CTTG..CT.T....AAAG........ACCTCTT........-----------.-.---A-T...TAGAAGTAAG--.-.GT..------T---A---.C-CCTAGTCGCCGCC-  743
SMM.SL.92.SL92B             ...GGCTC.CACGCTTG.....................................CTTG..CT.T..................GACTCT.........--------G.C-.--AACT...TAGAAGCAAGTC.-AG-..------T-T-A---.C-CCTAGTCGCCGCC-  208
STM.US.x.STM                T..GGCTC.CACGCTTG.....................................CTTG..CT.T....ATAT........ATCTCTT........-----------.-.---AAT..TTAGAAGTAAG.C.-AGT..------T---A---.C-CCTAGTCGCCGCC-  404
MNE.US.x.MNE027             T..GGCTC.CACGCTTG.....................................CTTG..CT.T....AAAG........ACCTCTT........-----------.-.---A-T...TAGAAGTAAG-C.-.GC..------T---A---.C-CCTAGTCGCCGCC-  212
DRL.x.x.FAO                 ..........................................................................................................................................................................    0
MND_2.x.x.5440              ..........................................................................................................................................................................    0
MND_2.CM.98.CM16            TCTCGG.......T.GGCCAAGAGGCTGTGGGATTGACTACCG...........CTTG..CT.T...GCTTT.......TAAGTGCT........-------AG.C--A--CAAGTTAAGAA.AAGCA-.GTG-...----ACT-G-C-A-.T--GTCTTTGGT....  367
MND_2.GA.x.M14              AGCAATCTC....TAGGCCAGGAGATTGTGGGTTAGACTACCG...........CTTG..CT.T....GCT......ATTGATGCT........--------G.C--A---GA..TTAGAA.TAGCA-.GTA-.......ACT-GTCCA-.T--GTCCTTGAT....  289
OLC.CI.97.97CI12            AGTAAGAACCTCAG........................................CTTTG.CT.GTGT.GAGAGA.........................-.------.....................GCTGAG-AA-AGAAGACTC-C---CT-G-GTGTT--AC-CATCCAGGT-  548
GRV.ET.x.GRI_677            ......................................................CTTGG.CT.TCATATAGCT.......................-.------C-...----TCGCT.TAGTCGCTA-ATTGGAGTCAA---CTCATT...............GCTGC-C  622
VER.KE.x.TYO1_patent        ......................................................CTTGG.CT.T..AGAAAG........................-T------.-.------GC..-T-AGAGCT-ATC.T-AG...TCAA---T--TCA-.TGAC-CCTCACTCTC-  666
VER.KE.x.9063               ......................................................CTTGG.CT.T..GGAAAG........................-T------A-.T------GC..-T-AGAGCT-ATC.T-AG...TCAA------TCA-.TGAC-CCTCACTC..A  663
VER.KE.x.AGM155             ......................................................CTTGG.CT.T..AGAAAG........................-T-------..C------GC..-T-AGAGCT-ATC.T-AG...TCAA------TCA-.TAAT-CCTCACTC..-  665
VER.DE.x.AGM3               ......................................................CTTGG.CT.T..AGAAAG........................-T------T..C--CG--GC..-T-AGAGCT-CTC.T-AG...TCAA------TCA-.TGAC-CCTCACTCTTG  160
TAN.UG.x.TAN1               ......................................................CTTG..CT.T...AGAA.........GGCTATTG.......-T-------.-.------GC..TTAAGAGT--ATC.T-AGCAA---CC.T-.AT-GC...GC-CCC.TTCCTC.  654
SAB.SN.x.SAB1C              GCGAC.TGGCCATTGCCAGTAGCAGAGACTCCG.....................CTTG..CT.T...GCTT.GA.......TTGCCT..........--------T.C-------A..GT-TACAAG-AA-.GCA-...-----------A-T-.A--CC-CAGGACAACC  728
MND_1.GA.x.MNDGB1            ......................................................CTTG..CTAT...................TGTCTT........--------TAAC-TAGAA...TTAGAGCAAG-G-.GTA-...----TAT--AT-G--CGCCTCTCTTCT....  211
DEB.CM.99.CM40              GG...............................TAGGCACTCCTCGGCACGCTTG..CT.T.....GCTACGCTTCTATAGCCCAG.......---------..TG--GTAA.................C-GCAA------CTG-CCTCATCATTGACTCCAGG-.  196
DEB.CM.99.CM5               GG..............................TTAGGTACTCCTCGGCACGCTTG..CT.T.....GCTGCATTTT.ATTGCCCAG.......---------..TG--GTAA.................C-GCAA------CTG-CCTCATCATTGACTCCAGG-.  197
TAL.CM.01.8023              ..........................................................................................................................................................................    0
TAL.CM.00.266               TGGGAGACCTTGCCGGTAAACGCTCAGTAGACATCCCGCTGCTTGCTTCGCGAAGCGCTTTAAACATAGCC...................---------..TG--TTCA.AAGAGCAGCAG-GTGTGCAGCCCA-C---TGGA-C-GCGGCG-GTGCTGGTCCC  196
SUN.GA.98.L14               G..CCTGA.TCAGCTTG..................GGGAGCTAGAGGCTC...CTTG..CT.T..TGCAGTACA...AG...CCTT........---------..CC------A..GT-TTACTGCAAG.CAA......----T-----G-TC-C.TCTTTCTCCTTC  696
WRC.CI.97.97CI14            AGGCTGTCACTCGCTTG.....................................CTTG..CT.A............GG.GTTT.GCAGTT.........-------AGTC-TTGACTTAGACAAGCAG-GAGTCAGTAGTC-TAT-TGGGCAAA-TAAG-AACCCTGGTTC  753
WRC.CI.98.98CI04            AGTCTGTC.CTCGCTTG.....................................CTTG...TTA............GGAGTTTTGCA.TT.........-------AGCC-TTTGCTTG-CAAGCA-TGAGTTT-GTAGTT-CAT-TA-A--ACAG-AG-AACCCTGGTTA  755
WRC.GM.05.Pbt_05GM_X02      ..........................................................................................................................................................................    0
LST.CD.88.524               ..........................................................................................................................................................................    0
LST.CD.88.447               ..........................................................................................................................................................................    0
LST.KE.x.lho7               GCCTGATCA..............GCCTAGGGAGC.TAAGGGCTCCTAG......CTTG..CT......ACTTTAAA...G....CCTT........-------AA.G----TCAT..T--TAGAAAG-AC.-AGC...TA----CT-ATAG-.AG-CTCTTTCTCCTTC  727
SYK.KE.x.SYK173             G.GTCCGGGTTTG.....GCCGCCCGGTCTTCGGCAGCTCTCCGTTG.......CTTG..CT.T..TGCTTACGCT..GATCAGCCCA.................--------T.G-G-TAAAC..GCAAGTAA-G--T.GC-C...TCA-C--TT-TG---GG-CACCTCCCGAA-G  581
SYK.KE.x.KE51               G..ACAAGGACTG.....ATCCACCTAGCCTCTACAATTCTCCGTTA.......CTTA..CT.T..TACTACGCG..CTGATCAGCC..........--------T.G-GTTAACT..-CAAGTAA-GAGT...GC...TCTCA-CTGTTCTGCCGGC--CTCCCAGAAG  220
COL.CM.x.CGU1               AAGGGCTTGAGGGACTC.......CAGTCCCCTCGCGCTAAGGGTAGATAGGTGCTTTGCAT.T...............AT.AAGAG........--------T.GA-TTTGAA..GC-TCAA--AG-G.TT-G...AGTAC-AGTTTCAC-TCCGCCTTCTC.....  214
```

```
                              Extensive secondary structure                                                 Lys tRNA primer binding site
                                                                                              5' LTR U5 end
H1B.FR.83.HXB2           GACTCT.GGTAACTAGAGATCCCTCAGACCCTTTTAGTCA.GTGTGG.................................AAAA.TCTCTAGCAG....TGGCGCCCGAACAGGG..ACCTGA............................................ 657
H1A1.UG.85.U455          ------.-----------------------AC-A---A-T.----A..................................----.----------...-----------------..--TC--.......................................... 126
H1B.US.90.WEAU160        ------.----------------------A---------.------..................................----.----------...-----------------..---C--.......................................... 657
H1C.ET.86.ETH2220        ................................................................................----.----------...-----------------.G------.......................................... 36
H1D.CD.84.84ZR085        ------.----------------------AGC-------C.-G---..................................----.----------...-----------------..------.......................................... 171
H1F1.BE.93.VI850         .................................................................................................................--..--T---.......................................... 8
H1G.SE.93.SE6165         ...........................................................................GGCTT----.----------...-----------------..--T---.......................................... 40
H1H.CF.90.056            .......................................................................................................................T---.......................................... 4
H1J.SE.93.SE7887         ....................................................................................................................................................................... 0
H1K.CM.96.MP535          ....................................................................................................................................................................... 0
H101_AE.TH.90.CM240      ------.-----------------------AC-C---A-T.-A--A..................................----.-----C----...-----------------.C--TC--.......................................... 204
H102_AG.NG.x.IBNG        ------.-----------------------AC-C---A-T.----A..................................----.----------...----------------A...-T---.......................................... 185
H1N.CM.95.YBF30          A-----.----C------------------A-C-AGAC.T.-A---A.................................----.----------...-----------------..--T---.......................................... 215
H1O.BE.87.ANT70          ....--.---GT---------------T-AC--AGAC.T.-AAGCAG.................................----.----------...-----------------..--T---.......................................... 687
H1O.CM.91.MVP5180        ....--.---GT---------------T-AC--AGAC.T.-AAGCAG.................................----.----------...-----------------..-.-GCG.......................................... 659
CPZ.CD.90.ANT            ................................................................................................................---..--T---.......................................... 9
CPZ.CM.01.SIVcpzCAM13    T-----.----T----------------TAAAA--TAGTC.CAG...................................G----.----------...-----------------..------.......................................... 198
CPZ.CM.05.SIVcpzEK505    A-----.----T----------------T.AA--CTA-TGAGTA....................................----.----------...-----------------..--T---.......................................... 198
CPZ.CM.05.SIVcpzMB66     ------.----T---------------TA..AC-GTAGTCTGTGC-GCC...............................----.----------...-----------------..--T---.......................................... 204
CPZ.CM.98.CAM3           ............................................................................................---...-----------------..--T---.......................................... 26
CPZ.GA.88.GAB2           .......................................................................................................................T---.......................................... 4
CPZ.CM.98.CAM5           ------.---------------------A....CA-T-AT.AA--A-.............................GTG.----.----------...-----------------..--T---.......................................... 331
CPZ.CM.05.SIVcpzMT145    ------.----T---------------A-.A-C--TATTGA-AGA...................................----.----------...-----------------..--T---.......................................... 201
CPZ.TZ.00.TAN1           -C-....CTGGGG---------------T..................................TTGTA..GTGGCTAAGTA.----.------C---...-----------------..--T---.......................................... 206
CPZ.CM.05.SIVcpzLB7      ------.----T---------------A-...--GT-GTCTGCGC-G.................................----.----------...-----------------..--T---.......................................... 202
CPZ.US.85.CPZUS          ------.---------------------A....C-CT-A-.AA--A.............................GGTG.----.----------...-----------------..--T---.......................................... 680
CPZ.GA.88.GAB1           C-----.---------------------T....-A-A-T-.TA--CA............................AGG..----.----------...-----------------..------.......................................... 676
MON.CM.99.L1             ...-GCTT-A--GAGACTTG-GAGTCTG----CA-CTGTG.CG-ATCAGGAGCGGGCTTGAGCCGGGGAG..GGGACAGCAG-C-.G-C--TCTCCA..G-----------------..--T---.......................................... 243
MON.NG.x.NG1             ....................................................................................................................................................................... 0
GSN.CM.99.CN166          ..............................................................AGTCCG.GGTGGACACC-G-..--CA-C----...-----------------..--T---.......................................... 219
GSN.CM.99.CN71           ..............................................................AGTCGC..GGTGGATACC-G-.--CA-C----...-----------------.G-T---.......................................... 227
MUS_1.CM.01.1085         .........TTG----AC-AAGG--AGT-TGC-CCTATT.CAT-CACACTGAGAGCCTATCTCAGCGTGGTGTGAGTCTGG---.CAC--C----...-----------------..--T---.......................................... 229
MUS_1.CM.01.CM1239       ..............................GC-CCTGTT.TAT-CACACTGAGAGCCTACCCCGGCGTGGTGTGGTTCCAG---.CAC--C----...-----------------..--T---.......................................... 229
MUS_2.CM.01.CM1246       AGTAGAC---TG-CG-CACAG--CG-.....................ACCTCGCGGG..................GGGA--T-.----CTC---...-----------------.G-T---.......................................... 248
MUS_2.CM.01.CM2500       CTAGAAGT--.....................................ACC...................GATCCAGGGCT-C-.-AC--CC---...-----------------..--T---.......................................... 228
RCM.NG.x.NG411           ..................................................................................................-----------------..--T---.......................................... 22
RCM.GA.x.GAB1            ..................................................................................................-----------------..--T---.......................................... 22
DEN.CD.x.CD1             ...............................................AGGAGGTTCGGGTTGTGAGCCTTACGCAGGGTGA----.C-C--C----...-----------------..--T---.......................................... 248
                                                                                               5' LTR U5 end     Lys tRNA primer binding site
MAC.US.x.239             -GTCAA.CTCGGTACTCA--AATAAG...AAGACCCTGGT.C---TAGGACCCTTTCTGCTTTGGGA..AACCGAAGCAGG----.--C-------A..T-------T---------..--T---.......................................... 843
H2A.GW.x.ALI             -GTCA-..TCGGTGTTCACCTG.AGTA--AAGACCCTGGT.C---TAGGACCCTTTCTGCTTTGGGA..AACGGAGGCAGG----.--C-------G..T-----------------..--T---.......................................... 875
H2A.DE.x.BEN             -GTCA-..TCGGTGTTCATCTGAGT-AC.AAGACCCTGGT.C---TAGGACCCTTCTCGCTTTGGGA..ATCCAAGGCAGG----.--C-------G..T-----------------..--T---.......................................... 880
H2A.SN.x.ST              -GTCA-..TCGGTGTTCATCTAAAGTA--AAGACCCTGGT.C---TAGGACCCTTTCTGCTTTGGGA..AACCAAGGCAGG----.--C-------G..T-----------------..--T---.......................................... 326
H2B.GH.86.D205           -GTCAC..TCGGTGCTCC-CTGA.....TAGACCCTGGT.C---TAGGACCCTTCTTGCTTTGGGA..AACCGAAGCGGG----.--C-------A..T-----------------..-----G.......................................... 877
H2B.CI.x.EHO             -GTCA-..TCGGTGCTCC-CTAG......AAACCCTGGT.C---TAGGACCCCTTCTGCTTTGGGA..AACCAAGGCAGG----.--C-------A..T-----------------..------.......................................... 876
H2G.CI.92.ABT96          -GTCAC.CTCGGTGTTC-CGTAGC.....AAGACCCTGGT.C---TAGGACCCTTTCTGCTTTGGRG..AACCAGAGCGGG----.--C-------A..T-----------------..--T---.......................................... 248
H2U.FR.96.12034          -GTCA-..TCGGAG-TTC-C-TTAT-.....AACCCTGGTAC---TAGGACCCTTCTTGCTTTGGGA..AACCAGAGCAGG----.--C-------G..T-----------------..--T---.......................................... 373
MAC.US.x.EMBL_3          -GTCAA.CTCGGTACTC..GGTAAT-AG.AAGACCCTGGT.C---TAGGACCCTTTCTGCTTTGGGAAAC..CGAAGCAGG----.--C-------TGAT----------------...--T---.......................................... 326
SMM.US.x.H9              -GTCA-.CTCGGTACTC--A-ATAAG...AAGSCCCTGGT.C---TAGGACYCTTTCTGCTTTGGGAAAC..CGAGGCAGG----.--C-------G..T-----------------..--T---.......................................... 326
SMM.US.x.PGM53           -GTCA-.CTCGGTACTC--A-ATAA--..AAGACCCTGGT.C---TAGGACCCTTTCTGCTTTGGGC..AACCGAGGCAGG----.--C-------A..T-----------------..--T---.......................................... 769
SMM.US.x.SIVsmH635F_L3   -GTCA-.CTCGGTACTC--CA-A-A-AG.AAGACCCTGGT.C---TAGGACCCTTTCTGCTTTGGGA..AACCGAGGCAGG----.--C-------G..T-----------------..--T---.......................................... 857
SMM.SL.92.SL92B          -GTCA-..TCGGGGTCCCGGATAA-....GAG.CCCTGGTAC---CAGAACCCTCTCACTCTTGGGGGCAACTCCTGAGTG----A-----G----T..T-----------------..--T---.......................................... 321
STM.US.x.STM             -GTCAA.CTCGGTACTC...........AAGACCCTGGT.C---TAGGACCCTTTCTGCTTTGGGC..TACTGAGGCAGA----.--C-------A..T-----------------..--T---.......................................... 507
MNE.US.x.MNE027          -GTCAA.CTCGGTACTC..AATAAT-AG.AAGACCCTGGT.C---TAGGACCCTTTCTGCTTTGGGA..AACCGAAGCAGG----.--C-------A..T-----------------..--T---.......................................... 324
DRL.x.x.FAO              ..................................................................................................-----------------..--T---.......................................... 22
MND_2.x.x.5440           ..................................................................................................-----------------..--T---.......................................... 22
MND_2.CM.98.CM16         .........AG-AACTCT.GGTTA-T-G.AGA-CCCTCAGATC-GT-GCAGAGATTTGATATCTAGTGGC..AGTAGAGAA----.--C-C----...-----------------..--T---.......................................... 471
MND_2.GA.x.M14           .........AG-AACTCT.GGTTA-T-G.AGA-CCCTCAGATC-GT-GCAGAGATTTGATATCTAGTGAG..AGTAGAGAA----.----C----...-----------------..--T---.......................................... 393
OLC.CI.97.97CI12         A..................................................................................--T.--C-A----...-----------------..--T---.......................................... 584
GRV.ET.x.GRI_677         CGAG-CTCTAG-GGT--ACCT-TCTTACTG........................GGTT..CTCCTG..TACCCAGGTGGG-G--ACTC-.-----...-----------------..--T---.......................................... 710
VER.KE.x.TYO1_patent     TGAA-G.--A-T-.TTCCT-A-TG..............................GGTTCTCTCTCT..GACCCAGGCGAG-G--.A---C----...-----------------..--T---.......................................... 748
VER.KE.x.9063            AG-AGG.--A-C-GTTCCT-A-TG..............................GGTTCTCTCTCT..GACCCAGGCGGG-G--.A---C----...-----------------..--T---.......................................... 746
VER.KE.x.AGM155          TGAA-G.--AG-AGTTCCT-A-TG..............................GGTTCTCTCTCA..AACCCAGGCGAG-G--.A---C----...-----------------..--T-............................................ 745
VER.DE.x.AGM3            A--GGG.TAA---.TTCCT-A-TG..............................GGTTCTCTCTCA...ACCCAGGCGAG-G--.A---C----...-----------------..--T---.......................................... 241
TAN.UG.x.TAN1            .-GGGAACCCTC.GTT-CTGG.................................GGTT...TCTTA..T..CCAGGCTGAG---.----C----...-----------------..--T---.......................................... 728
SAB.SN.x.SAB1C           CT....................................................GGTTACTAAGGA..TCCCTGATAGAA----.--CT-----...-----------------..--T---.......................................... 790
MND_1.GA.x.MNDGB1        ...................................AAACCTGT--TGTTCTCATTTAGAGAACAGAAGGACTTCTAGTTA-CCCTAGAAGCCTTTCAG.-------------.--..--T---.......................................... 295
DEB.CM.99.CM40           ................................................AATTCGGGTCGAGCGGGCTAGGCTCAAG.CG-G-G.GTC--------..T-----------------..--T---.......................................... 263
DEB.CM.99.CM5            ................................................GTCGGGAGCTCAAGCGGCTAGGCTCTGGGCAGG-G.AAC--------..T-----------------..--T---.......................................... 265
TAL.CM.01.8023           ....................................................................................................................................................................... 0
TAL.CM.00.266            -CGC-CTC-G............................................GTCGAGGGAACTTGGTGGGTGAAGA--TTTA-C-ACCAGC...-----------------..--T---GAACACCCAGGTTCTCTGTCCTGAGGGCCAAGCGGCTTCGGGGACC 314
SUN.GA.98.L14            T....A.AACCCTGTTC-............AC-GCTC-T-.ACTA--GAGCAATTACTGGCCGCCTA..GTGATCCGGGGCTG-C.GGC.------...-----------------..--A---.......................................... 792
WRC.CI.97.97CI14         -CT-GAG-T-G-TCTTCAG.AGGAGCTGAAGACCCC..............................................-A-----T----...-----------------..------.......................................... 822
WRC.CI.98.98CI04         -CT-GAGCT-GGACTTCAGAAGGAGCTGAAGACCCC..............................................-A-----T----...-----------------..--AA--.......................................... 825
WRC.GM.05.Pbt_05GM_X02   ....................................................................................................................................................................... 0
LST.CD.88.524            ....................................................................................................................................................................... 0
LST.CD.88.447            ....................................................................................................................................................................... 0
LST.KE.x.lho7            -C-CTG.-T-C-G-GATCTCT-ACT---GAGA--GGAG-C.T---.............................GATCCGGG-C-.GGG--TC---...-----------------..--T---.......................................... 813
SYK.KE.x.SYK173          TG---A.--G-G-GC-......................................................................-CTT---...--------.--------..--T---.......................................... 624
SYK.KE.x.KE51            .--G-AC--G--AC.......................................................................G-G-CTT---...-----------------..--T---.......................................... 265
COL.CM.x.CGU1            ..............................-GCCCT....C....AGAGCGGGGCCTTTTTGCTG.....TCCGGGAAGG---.GGC--T----...---------G-------..--T---.......................................... 286
```

```
H1B.FR.83.HXB2          ........................................................................................................ 657
H1A1.UG.85.U455         ........................................................................................................ 126
H1B.US.90.WEAU160       ........................................................................................................ 657
H1C.ET.86.ETH2220       ........................................................................................................ 36
H1D.CD.84.84ZR085       ........................................................................................................ 171
H1F1.BE.93.VI850        ........................................................................................................ 8
H1G.SE.93.SE6165        ...................................................................................AAGCGAAAGTTAACAGG 57
H1H.CF.90.056           ........................................................................................................ 4
H1J.SE.93.SE7887        ........................................................................................................ 0
H1K.CM.96.MP535         ........................................................................................................ 0
H101_AE.TH.90.CM240     ...................................................................................AAGCGAAAGTTAATAGG 221
H102_AG.NG.x.IBNG       ...................................................................................CGGTAATAGG 195
H1N.CM.95.YBF30         ........................................................................................................ 215
H1O.BE.87.ANT70         ........................................................................................................ 687
H1O.CM.91.MVP5180       ........................................................................................................ 659
CPZ.CD.90.ANT           ........................................................................................................ 9
CPZ.CM.01.SIVcpzCAM13   ........................................................................................................ 198
CPZ.CM.05.SIVcpzEK505   ........................................................................................................ 198
CPZ.CM.05.SIVcpzMB66    ........................................................................................................ 204
CPZ.CM.98.CAM3          ........................................................................................................ 26
CPZ.GA.88.GAB2          ........................................................................................................ 4
CPZ.CM.98.CAM5          ........................................................................................................ 331
CPZ.CM.05.SIVcpzMT145   ........................................................................................................ 201
CPZ.TZ.00.TAN1          ........................................................................................................ 206
CPZ.CM.05.SIVcpzLB7     ........................................................................................................ 202
CPZ.US.85.CPZUS         ........................................................................................................ 680
CPZ.GA.88.GAB1          ........................................................................................................ 676
MON.CM.99.L1            ...................................................................................GTAACGACCCAGCGACT 260
MON.NG.x.NG1            ........................................................................................................ 0
GSN.CM.99.CN166         ...............................................AACAGCCCAGGACTTCAAGCCCGAAGGCGGGACCCAAGGATTGCCCCT 267
GSN.CM.99.CN71          ...............................................AGCATCCCAGTGACTTCAAGCTCGCAGGCGGGACCCAAGGATTGCCCCT 276
MUS_1.CM.01.1085        ...............................................TTAAGCCCGCGAGGAGGTAGGCGGTGTGACGGAGAGAAGG...CCCC. 273
MUS_1.CM.01.CM1239      ...............................................TTAGCCCGCGAGGGGGTAGGCGGTGTGACAGAGAGAAGG...CCCC. 272
MUS_2.CM.01.CM1246      ...........CTACATATCAACCCGTTCGGAGGTTCGCGGCCGGCAGTCAACAGGCGGTGGAGCAGTCCCAGAAGGCTCCATCCCGAAGG..CGAGACGCCTGAG 341
MUS_2.CM.01.CM2500      ...........................CGAACGTCAGGTAGGATCGGTGAGCAGGCAGTAGAG.CCTCCCAGGGACTCTGCGCTGAAGGCGACAGGCGCCGAGAG 305
RCM.NG.x.NG411          ...................................................................................GAGTGACT 30
RCM.GA.x.GAB1           ...................................................................................GAGTGGCT 30
DEN.CD.x.CD1            ...............................................AAGAGAGTCCTGAAACAGGGACTTGAGTACCCAGAAGGCTCAGGCCCGA 297

MAC.US.x.239            ...................................................AGGAGAGTGAGAGA..CTCCT. 862
H2A.GW.x.ALI            ...................................................AGAAGACTGAGAA.G..TCTAG. 894
H2A.DE.x.BEN            ...................................................AGAGGACTGAGAA.G..CCCTG. 899
H2A.SN.x.ST             ...................................................AGAAGACTGAGAA.G..CCTTG. 345
H2B.GH.86.D205          ...................................................GAAGAGTGAGAG.T..CTTCAA 896
H2B.CI.x.EHO            ...................................................GGAAGAGTGAGAG.T..CTTCG. 895
H2G.CI.92.ABT96         ...................................................GAAGGAGTGAGAGT...CCTTT. 267
H2U.FR.96.12034         .....................................AGAGTGAGAGT...CTTCTG 390
MAC.US.x.EMBL_3         ...................................AGGAGAGTGAGAAGA.ACTCCT. 347
SMM.US.x.H9             ...................................AGGAGAGTGAGAG....CTCCT. 344
SMM.US.x.PGM53          ...................................AGGAGAGTGAGAG....CTCCT. 787
SMM.US.x.SIVsmH635F_L3  ...................................AGGAGGTGAGAG....CTCCT. 874
SMM.SL.92.SL92B         ...........................AGACG.......................... 326
STM.US.x.STM            ...................AGGAGAGTGAGAGA...CTCCT. 526
MNE.US.x.MNE027         ...................AGGAGAGTGAGAGA...CTCCT. 343
DRL.x.x.FAO             ...........AGACAGTCGACCCCACCTGTGAGGGACGAA.GGCGGCAGCCATCCGGACCGACCCACCCGGAAAAGAGAGTGC..TAAGAAGCCCCGACG 109
MND_2.x.x.5440          ...........CAAGGGTTCAGGTCGCACCCACCTGTAAGGGACGAGCGGCGGCAGC.AGTCGGACAGACCCACCCGGCG.AAAGAGTTAATTGAAGAGCCCCGACG 116
MND_2.CM.98.CM16        ...........CAAGGGTTCGAGTCATTCCAACCTGTAAGGGACGAAGGCGGCAGCCAGCCGGACCGACCCACCCGGCGAAGTGAGTTAACCAAG 555
MND_2.GA.x.M14          ...........CAAGGGTTCGAGTCGCTCCGACCTGTAAGGGACAAGCGGCGGCAGCCGGTCGGACAGACCCACCCGACGAAGAGAGTTAATCAAGGAGCCCCGACG 489
OLC.CI.97.97CI12        ...................................................................GAAGA 584
GRV.ET.x.GRI_677        ...................................................................GAAGA 715
VER.KE.x.TYO1_patent    ...................GTGA..GAGTGT 758
VER.KE.x.9063           ...................GGTCCGAGGACTT..GAGTGA 765
VER.KE.x.AGM155         ...................GAGTGA 751
VER.DE.x.AGM3           ...................CTTGAGTGA 250
TAN.UG.x.TAN1           ........................................ 728
SAB.SN.x.SAB1C          ...........CTAAAGACGTCCCGGCCGCAGAGTAGACCA 820
MND_1.GA.x.MNDGB1       ........................................ 295
DEB.CM.99.CM40          ...................GCTACCCAGAGCTCGGCCCGCAGGC 288
DEB.CM.99.CM5           ...................AGCATCCCAGTGTTTGAGCCCGAAGGC 292
TAL.CM.01.8023          ........................................ 0
TAL.CM.00.266           GCGGGTTTTGGGTCCCAAAGAAGAGAGAGAAACTCCGAGGGGTCGTGAGAGGACGACAGTTCCGGAGTCGAGCCCCGGATCGGCGAAGACGACGCGCGTCGCAAAGTGAAAGTAAGGTAAGGAGCCGATCGGTCACCCCGGGCGAGTAGCCGCCGTGAGCGAGTGCGGTG 484
SUN.GA.98.L14           ...................GGGAGAGAGAGTGACGAAGAGACGTGGAGTGAAACCTTCAGAAGAAGTGAGCAGCTGCTGAGAAGAAGGAAGGTTTCGACCTGGTGGAA 881
WRC.CI.97.97CI14        ...................GGGGCGTCTAGAGTGAAAGAAGACCGGCCGGCGA 856
WRC.CI.98.98CI04        ...................GGGGCGACTAGAGTGAAAGAAGACCGGCCGGCGA 859
WRC.GM.05.Pbt_05GM_X02  ........................................ 0
LST.CD.88.524           ........................................ 0
LST.CD.88.447           ........................................ 0
LST.KE.x.lho7           ...................AGGAGAAGAGGCGCTGGCAAGAACCTCTGGAGGGAGGAAAAGCGCGCAG........AAGAGGCCAGGAAGCTAGGACAGA 886
SYK.KE.x.SYK173         ........................................ 624
SYK.KE.x.KE51           ........................................ 265
COL.CM.x.CGU1           ...AAGAGGTCCCTGAGGGTCGCGGAAGAAGAAGAGGAGACTGATCGCCTCCTCAGAAGAGAGGTAAGTCGAAAAAGGGGACTTACGGGTGGAGTCAAGGAGGACTCTACGGTCTGCGCAAAAGTGA 410
```

```
                                                             Packaging loops begin
H1B.FR.83.HXB2            .......AAGCGAAAGGGAAA......CCAGA....GGAGC....TCTCTCGACGC..AGGACTCGGCTTGCTGAA...GCGCGCAC.GGCAAGAGGCGAGGGGCGGC.....GACTGGTGAGTACGCC...AAAAATTTTG.ACTAGCGGAGGC..............  776
H1A1.UG.85.U455           .......---------TT........-----.....-A--T...-----------.------------------G...-T--A---.A-----------A.-----....GACTG-TGAGTACGC-..TA---TT-----.------------.......CTAG....  246
H1B.US.90.WEAU160         .......--------------......-----.....-----...-----------.------------------G...------.------------------.-----....---------------G.---------------------------.......CTAG....  778
H1C.ET.86.ETH2220         .......---------T--G-......-----.....---A...-----------.------------------...-T--A-T-.-----------A.-----....----------------.AATTTTTA----.------------.......CTAG....  159
H1D.CD.84.84ZR085         .......---------TAG--......-----.....---A...-----------.------------------..A------T-.---------------A--....-------------TG.A---TT-----.------------...............  293
H1F1.BE.93.VI850          .......---------TAG--......-----.....-A--A...----------G.G-----------------..GTGCAA---.-----------A.-----....GACTG-TGAGTACGC-GAATTTTTT-----.------------...............  131
H1G.SE.93.SE6165          GACTCGA---------TT........-----.....-A--T...-----------.------------------G...-T--A---.A-----------A.-----....----------------.....----T-----.------------...............  179
H1H.CF.90.056             .......T--T-----TA--......-----.....-A--A...-----------.------------------...-T--A---.A-----------A.-----....----------------...-TTTTG---TG------------...............  123
H1J.SE.93.SE7887          ...................................T...-----------.------------------...-T--A---.-----------------G....----------------AAA--T-T-----.------------.......CTAG....  102
H1K.CM.96.MP535           ...........................................................................................................................................................................    0
H101_AE.TH.90.CM240       GACTCGA---------TT........-----.....-A--T...-----------.------------------G...-T--A---.A-----------A.-----....----------------....-----------.------------.......CTAG....  344
H102_AG.NG.x.IBNG         GACTCGA---------TT........-----.....-A--A...-----------A.------------------G...-T------.A-----------A.-----....GACTG-TGAGTACGC.-....--------.------------.......CTAG....  319
H1N.CM.95.YBF30           .........-A-----TAG--.......-G--.....--CTGAA..-----------.-----------C-T--.....-T--A---.A--G---------C-GC-GA.....-G--TGAGTACGC.........-----A.--G----T--..CC............  330
H1O.BE.87.ANT70           .........--T-----T----........-----...---AGAAAAC---C.-----AAC--G-------AGC-G-...-T--A-C-.-CT---------A--AACTC.ACA--GG-TGAGTA-...........------.--G----T--..CCA..........  805
H1O.CM.91.MVP5180         ......A---T-----T-G--.........-----...---AGAAAAC---C.-----AAC--G-------AGC-G-...-T--A-CT.-CT---------A--AACTCACAA--GG-TGAGTA...........------.--G----T--..CCA..........  780
CPZ.CD.90.ANT             ...............--AG........-T-C-.....-A-A....C--G-.-----C.---------TC-----C-..........GT.A---GA-----------A......-GG-TGAGTA.-C.TTCATTTTG-----..T-GCTA-T-A..GCA..........  116
CPZ.CM.01.SIVcpzCAM13     .......---------TA-G-.......--G-......--CTGAA..----C------A.C-----------T--.-...-T--A---.AA-------------.----G....ACTG-TGAGTACGC........--.T-----.-------T--..CTA..........  315
CPZ.CM.05.SIVcpzEK505     ..........---T-----TA-C........-G-......--CTGAA..--C----G---.-----------C-T--.-...-T--A---.A--G---------.----G.....--AA-------------.........-----A.--G---T--...............  310
CPZ.CM.05.SIVcpzMB66      .........--A------TA-G-......------....-A--T...-----------.------------------...-T--A---.-----------------A---G.....----------------....CT--T-----.------CG-T...............  321
CPZ.CM.98.CAM3            ......A-GTT-----T-G--.........-----....--CTGAC.---TC-----ACG---------------.-...-T--A---.A--G---------C-AACTC...CAGAG-TGAGTG............T------.-C-----T--..CTG..........  143
CPZ.GA.88.GAB2            .........---T-----CA---.......--G--GGAA--CTGAC.----C-----G.C-------A-------T-...-T--A--ACA--G---------C.G-TAC.......G-TGAGTACA.......TTTTC----.-----A-T--..CTA..........  125
CPZ.CM.98.CAM5            ...............---T----GAAGAA---------CTGAC.---TC-----A.C-----------A---.-...-T--A-G-.A-T----------C-ACT-C....TGAG-TGAGTATTT........-.T------.-C-----T--..CTG..........  449
CPZ.CM.05.SIVcpzMT145     ...............---T----GTAAAA--G-.......--CTGAA..-----------A.C---------------.-...-T--A-G-.A-----------C-GC-AC.TCAAGTG-TGAGTACGC..........-.-----------T--..CTA..........  322
CPZ.TZ.00.TAN1            ...............-AAGC-......GGGA-....C-C-G....C-C--G----.-----------.---CA.----AAT-..A-----------C-GACTC...C-GTG-TGAGTAC......A--TTT-G--.......T---T--..GCA..........  313
CPZ.CM.05.SIVcpzLB7       .........--A------TA--........-T-G...A-C--T...----------G.G-----------------...-T--A---.-------------------A---G....ACTG-TGAGTACGC.....CT--C-----.-----------.......TA..........  318
CPZ.US.85.CPZUS           .........--A------TAC--........-----.....--CTGAA..A--TC-------.G---------C----.-...-T--A---.A--G---------C-AACTC...CAGAG-TGAGTACA...........------.-C---A-T--..CTG..........  795
CPZ.GA.88.GAB1            ...............--T----GTAA..--G-.......--CTGAA..----C-----A.C---------------.-...-T--A---.A-----------C-GC-AC......TG-TGAGTACGC........--.T-----.-------T--..CTA..........  790
MON.CM.99.L1              CGAGCCCGCAG-CG-ACAG-TCCGCGAGA--CCTTGCCGAGTGCGC-CGA-A-A--TGC-C-GCAA-AGGTAA-G-GAGATCG-GCG.-T-CCCTA-G-T--CAGACTA....-CG---GAC-GGAA--TATCC--GAGGGAAGG-TC........CAGAGGTGCTCCAA  417
MON.NG.x.NG1              ...........................................................................................................................................................................    0
GSN.CM.99.CN166           GGCCGAGTG-TCCGG-CTCC-CGAAATTTGCTGCGCA-CC-CTTG...AAGAGAAGGT-A-GAGACCGGACGG-TCTA.-AAG-GTG.-CAG-CTTCTAG-CCAG-TG..GGTTG....AAGAAGG-TACCC-C-C-G..A-GGGAGAAACT-TG..............  409
GSN.CM.99.CN71            GGCCGAGTG-TCC.G-CTCC-C..AACTTGCTGCGCA-CA-CTAG...AAGAGAAGGT-A-GAGACCGGACGG-TCTG.-AAG-GTG.-CAG-CTTC-AG-CC-GT-G..GTT--....GAGAAGG-TACCC-C-C-GAGG-G-GA-A-T-T-A..............  416
MUS_1.CM.01.1085          ..AGCCG..AGTGCGT-CC--G..CCA--G--GAACAAGTGCTGCG-AGC-ATAAGGT---TGA-TTGGACGG-TCTGG-AAG-GTG.-CAG-CTTC-TCA-AT--AGCA...-G-ATAGAGAG-TC-TGCGGG-GGGGAGAA---C......................  410
MUS_1.CM.01.CM1239        ..AGCCG..AGTGCGT-CC--G.ACCA--G--GGACAAGTGCTGCG-AGC-A-GAGGT---TGG-TTGGACGG-TCTGG-AAG-GTG.-CAG-CTTC-TCA-AT--AGC....AGG-ACGAGAA--C-TGCGTG-G-GGAGAA---CC.....................  410
MUS_2.CM.01.CM1246        GATCGCCT---CG-GT-..GCCCGGCA-GGAGAAAAAAGAGCTGCG-AGCAA-A-AGGTAAGG-GTCGGGCGA-CCTA.-A-G-GTG.-CAG-CCTCTAG-T--GC-GGT...AGAC-AG---.GTC--CGCG-GG-GGAGAA--CGCT.................  482
MUS_2.CM.01.CM2500        AACGCCCC---CG-GT-...-GCACCGA--CGAGGAAATCGCTGCAGC.GT--A-AGGTAAGG-GTCGGGCGA-TCTA.-A-G-GTG.-CAG-CCTCTAG-C--GC-GT.....T-A-AGA--.GGA--GCG-G-G-GGAGAA--.........................  438
RCM.NG.x.NG411            GAGAGACCTC---CGC--GTCCGGACCAG-AGAGTGACT-GTAGAA-G-GGAC-CAGTCCAGG-AA-AGCAGCCTGGTG--T--T-A.AT-T--C-AGA-AA-TGA-GG....CTAGTAGAGCA-G--GACCCGGTCAGG-ACTGGGC........CTCTAGGGAAGGGT  187
RCM.GA.x.GAB1             GAGAGACCTC---GGCTA-GGTTCGGCG-G-CAGGTCACCGCGGGAG-GGAAC-TTGACCAGG-AA-AGCTGCCTGGTG--TTCA-A.-TGCTAGA-AA--T-AGCTAG....CGAA--AAGCAGG--AACCCGGTCCGG-ACTGGGC........CTCTAGGGAAGGAG  187
DEN.CD.x.CD1              AGGCGGTTG-AACGTACC-CCCTCTTA--GAGCAAG--C-AGGAAGACGCT-CGCAGCGCA-AGGTAAGGATCTCGCCG-A--A---.ACT-GTG-CA-TA-T-T--GG....CTGGAAGATAGCAAGGAGGGCT-TC-GGAA--A-A........GGGCTAGATAAGGG  454

MAC.US.x.239              GAGTACG..--TG-GT-A-GGCAGTAAGGGCGGCAG-A-C-AACCA-GACG--GTGCTCCT-TAAA-GCGCGG-TC..G-TA-C.....AG-C.G-CGTGA--AGC-GG....AGAG-AA---G.-CT-CGGTTGC...AG-T-AGT--.........AACACAAAAAAAGA 1006
H2A.GW.x.ALI              GAACACG..--TG-GT-A-GGCAGTAAGGGCGGCAG-A-CAAACCA-GACG--GTGCTCCT-GAAA-GCGCAG-CCAAG-TA-C..A.AAGGCCG-CGTGT--AGC-GG....AGTG.AA---G.-CT-CGGGTG-...AG-T-AGTA-.........CTACACCAAAA.T 1041
H2A.DE.x.BEN              GAACTCG..--TG-GT-A-GGCAGTAAGGGCGGCAG-A-CAAACCA-GACG--GTGCTCCT-GAAA-GCGCGG-CCGAG-TA-C..A.AAGGC.G-CGTGT--AGC-GG....AGTGAAA---G.-CT-CGGGTG-...AG-T-AGTA-.........CTACACCAAAAAC 1047
H2A.SN.x.ST               GAACACG..--TG-GT-A-GGCAGTAAGGGCGGCAG-A-CAAACCA-GACG--GTGCTCCT-GAAAAGCGCAG-CCGAG-TA-C..A.A-GGC.G-CGTGT--AGC-GG....AGTGAAA---G.-CT-CGGGTG-...AG-T-AGT--.........CTACACCAAATAC  493
H2B.GH.86.D205            GAGCACA..--TG-GT-A-GGCAGTAAGGGCGGCAG-A-T-AACCA-GACG--GAGCTCCTGTAAAAGCGCAG-CC..G-TA-C.....AGGC.--CGTGA--AGC-GG....AGGA-AA---G..CT-CGGG-GT..AAG-T-AGTA-.........TCCAAA....... 1033
H2B.CI.x.EHO              GGACACG..--TG-GT-A-GGCAGTAAGGGCGGCAG-A-T-AACCA-GACG--GTGCTCCTGTAAAAACGCAG-CC..G-TA-C.....AGGC.--CGTGA--AGC-GG....AGGA-AA---G..CT-CGGG-GT-AAAG-T-AGTA-.........TCCAAA....... 1034
H2G.CI.92.ABT96           GAACACG..--AG-GT-A-GGCAGTAAGGGCGGCAG-ATC-AACCA-GACG--GTGTTCCT-GAAA-GCGCGG-TC..G-TA-C.....AG-C.--CGTGA--AAC-GG....AGTG.AA---G.-CT-CGGGTGCGC.AG-T-AGT--.........ACCAAAAATTTGT  412
H2U.FR.96.12034           AAACTCG..--TG-GT-AGGGCAGTAAGGGCGGCAG-A-CAAACCA-GACG--GTGCTCCTGTAAA-GCGCAG-CT..G-TA-C.....AAGC.--CGTGA--AGC-GG....AGTG..A---G.TCT-CGGTTGC...AG-T-AGT--.........AACAAAAAATAGA  532
MAC.US.x.EMBL_3           GAGTACGG.C-TG-GT-A-GGCAGTAAGGGCGGCAG-A-C-AACCA-GACG--GTGCTCCT-TAAA-GCGCAG-TC..G-TA-C.....AG-C.G-CGTGA--AGC-GG....AGAG-AG---G.-CT-CGGTTGC...AG-T-AGT--.........AACACAAAAAAAGA  492
SMM.US.x.H9               GAGTACG..--TG-GT-A-GGCAGCAAGGGCGGCAG-A-CAAACCA-GACG--GTGCTCCT-TAAA-GCGCGG-CC..G-TA-C.....AGGC.G-CGTGA--AGC-GG....AGTC--A---G.-CT-CGGTTGC...AG-T-AGT--.........AACAAAAGAGTCA  488
SMM.US.x.PGM53            GAGTACG..--TG-GT-A-GGCAGTAAGGGCGGCAG-A-CAAACCA-GACG--GTGCTCCT-TAAA-GCGCGG-CC..G-TA-C.....AGGC.G-CGTGA--AGC-GG....AGTC--A---G.-CT-CGGTTGC...GG-T-AGT--.........AACAAGAAGAG..  929
SMM.US.x.SIVsmH635F_L3    GAGTACG..--TG-GT-A-GGCAGTAAGGGCGGCAG-A-CAAACCA-GACG--GAGCTCCT-GAAA-GCGCGG-CC..G-TA-C.....AGGC.G-CGTGA--AGCAGG...GAGTC--A---G.-CT-CGGTTGC...AG-T-AGT--.........AACAAAAA..... 1014
SMM.SL.92.SL92B           GACT.CG..--TG-GT-A-GGCAGTAAGGGCGGCAG-A-T-AACCG-GACG--G....TTTGA....-CC-G--CC..A-T-G--...T-AGC-G-AGT-CA--TAAGA...GAGGCCTGTG-G--AGTAAGGTCC-AAAG-T..GT-A.......ATATCCTACCTACG  468
STM.US.x.STM              GAGTACG..--TG-GT-A-GGCAGTAAGGGCGGCAG-A-CAAACCA-GACG--GTGCTCCTGTAAA-GCGCGG-C.............AAGC.G-CGTGA--AGC-GG....AGGC-AA---A..CT-CGGTTGC...AA-TGAGT--.........AACAAGTAGCCAG  663
MNE.US.x.MNE027           GAGTACG..--TG-GT-A-GGCAGTAAGGGCGGCAG-A-C-AACCA-GGCG--GTGCTCCT-GAAA-GCGCGG-TC..G-TA-C.....AG-C.G-CGTGA--AGC-GG....AGAG.AA---G.-CT-CGGTTGC...AG-T-AGT--.........AACACAAAAAAAGA  486
DRL.x.x.FAO               CGAGACCCGC-TCGGTAAGT-AAAAGGAA--TAAAATC--TGCAC-G-GAG-CGAAGGGTCTAGT--GCC-TGTG-GCA---GCT-G.-CA---GCAACTCC-TA---A....A-AA-AAAGAA-AAGGGAG--GT-AGGCCTGT-GC........TGATAAGTGAGGCA  266
MND_2.x.x.5440            CGCAGGACTCAA-GGTAAGC-GTGCACTGTA....ATTGATTGAGAGAGAGTC-AGGG-TCCGCTT-AGCAGGCG-GATCGCGAAGG.CAACCCCA-TAG-CAAAA-AA....A-GAA-G-GAAGTAAGGCCG-CGCAAAGTG-AAGT........AAGGGCTTCCTCCG  269
MND_2.CM.98.CM16          GAGCCCCG-CGCGC---ACTCAAGGTAAGC-GTGCACCGAACAGT-G-G-GTGT-AGTCCAGGGTCCGC-AGA-C-AGC-A-AC-G-.C-AGGC-AC-CCA-TAG-CAA....A-GGAAAAGAGGG-AAGTA-GGCCGAGGCA-AGT-........AAAAGAGGTCCTTG  712
MND_2.GA.x.M14            CGCAGGACTC.A-GGTAAGCGGTGCAC-.......GTATTGTGGTAG-GAGTC-.AGG-TCCGCTT-AGCAGGCG-GATCGCTAAGG.CAACCCCA-TAG-T.AAAAGA....A-AGA-G-GAAGTAAGGCCG--GCAAAGTG-AAGT........AAAAGGGCCTTCCG  636
OLC.CI.97.97CI12          ......................................................................................-AAA-G-TTCGAG.TC-CT-AAGAAGAAGTTGGTC-T-AGG--AGTG-C........CGGGGAACACGAGG  646
GRV.ET.x.GRI_677          GGCATCGGCA-CG-CC-CTG-GTTGCTGAGCGTCGGA---GGACGA---AG-TA-GGTGA--GC-TA-GA-T-TTTTGCTAC-TAGT.CAGCGAGAAA-GCTA-GCCGC........---A--G-C-CGG-T-CCATT-GTGGCAACCAACC........CAGTTGGACGAAGG  872
VER.KE.x.TYO1_patent      AGGCACGT-CA-CTGA-A-GGCGTCGGA-GCGAAGGAAGCGCGGGGTG-GACG--ACC-A--AGGA-AC-TGGTG-GTA-GCTT-T-.-AGTGCC--GA-AAA--TCGA....-C--A--T--AGGA-TAGG-G-GGCCG-AGC-GTA........ACTACTCTGGGCAA  915
VER.KE.x.9063             AGGCACGT-CA-CTGA-A-G-CGTCGGA-GCGTAGGAACCGCGGGGTG-GACGT-ACC..A-GAA--GC-TGGTG-GTG-GCTT-T-.-AGTGCC-.AA--AA--TCGA....-C--A--T--AGGA-TAGG--GGGCC.-AGC-ATA........ACTACTCTGGGCAA  918
VER.KE.x.AGM155           AGGCACGT-CA-CTGA-A-G-CGTCGGA-GCGAAGGAACCGCGGGGTG-GACGT-ACC..--GAA--GC-CGGTG-GTA-GCTT-T-.-AGTGCC--GA-AAA--TCGA....-C--A--T--AGGA-TAGG--GGGCCG-AGC-GTA........ACTACTCTGGGCAA  906
VER.DE.x.AGM3             AGGCACGT-CA-CTGA-A-G-CGTCGGA-GCGAAGGAAG--GCGGGG-GCGACGTGAC.CA-GAA--GC-TGGTG-GTA-GCTT-T-.-AGTGCC--GA-AAA--TCGA....-C--A--T--AGGA-TAGG--GGGCCG-AGC-ATA........ACTACTCTGGGCAA  406
TAN.UG.x.TAN1             .......GTCA-----ACTTCACTGAAGGG--GGAA--GCATCACCAGTCGAGTT-CCC----G---TAA-TC-GGACGCG-G-AGG.C-A-GAG-ATCC-CCAG-AAA....A-AGTACTGCC-G-GAGGC-GGGG-GG--C-GGTA........GATCTAGGTGGGCG  878
SAB.SN.x.SAB1C            GCCTGAGG-CTCCTTT-ACTGAGTAAGA-C-GACGGAAGAAAGGCGGA-CGTG-CGGAGT-CAGGTAAGA-GC-CCTGC-GA-C-GA.-A--CCCTCG-GCT-TGT-TG....TGT.....---A-GAGAGGA----GAGAA-TCAG-C........AGGGCTGACAAGGC  973
MND_1.GA.x.MNDGB1         ............................AG---GGCACTGA-ACTTGAGGCA--GCACTCC-C--G-AAGAAGCAGGTTGAA-GAG-G.T-G-CTG-T-TGAA-A--C-A....-GAG-TGAGTC-GTGGGACTG-CT--ACAAGAATT........AGTTGTACCCTAGT  425
DEB.CM.99.CM40            GGTGTGCGGAGTCTG-AA-CCCTGAGC-G-TCTCTTCCC--CGGA-.AGCT-CGCAGC-AAGG-AA--GGA-GCCGGAGAATACAGG.-C-GC--ATAAG--C-G-TGT....-TG.-TCTCTGCA-G-AGAGT-CCAC-GTGGAGCA........AGAGAGGCTCCTGG  443
DEB.CM.99.CM5             GGAGTGTGGAGTCTG-AA-CCCTGAGC-G-TCTCTTCCC-TCGAG-GAGCT-CGCAGC-TAGG-AA--GGA-GCCGGAGAATACAGG.-C-GC---TG-G-AC-G-TGT....-TG.ACCTCTGCT-G-AGAGTGTCAC-GTGGAGCA........AGAGG..CTCCTGG  446
TAL.CM.01.8023            ................................CAG--G-GAAGG--G--T--G-GACGC-CAGGCAGGG-TCTGCGGG-AAGT--G.C-AC-CTTC-AGACAAGCA-C....CGG-A-AA--GTGA-AGCTC-G-GGAGGAGCTC--........GGACCTGCGGGTGG  124
TAL.CM.00.266             CTAGGAAG--T-GC--A--GGCCTAGCTT-ACCCAG--G-GA.GGA-GGA--GA-GACTT-CAGGCAGGG-TCTGC.AG-AAGTTCGT-CGT-CCTC-AGACAAGCA-C....CGG-A-AA--G..AGACTGCTC-GGAGGAGCAC--........GGACCTGCGGGTGG  638
SUN.GA.98.L14             GGCGCCGGGAA-TT-A--T--GTGAGGG--CTTACCA-GCACTGTCAGGGGACA-GCC--A-AGTCCTAG----GGGGGAG-TC-CG.CTAGGTTA-TACT-TCAA-TA....-G-A-ACT----GTA-GTGG--T-AAA-AAGAAGT........GTGTGTATAACTCT 1038
WRC.CI.97.97CI14          GGGATTC-CT------AAGTCGTCC...AGA-.GGTAAGAGGGAGGACT--TGGT-CCT-TTGAGAC--CTAGA-GGAATTAGAGG-.TAGCCTT--AAC-TA-A---TGAGTAGT-CA-A--GGACTACTGGG-G-CAGGAG-AAG-ACC---GGGGAGGC....TATA 1017
WRC.CI.98.98CI04          GGGATTC-CT---G--A-GTCAGCCTC-AGA-AGGTAAG-AGGAGGGCT--TGGT-CCT-TTGAGAC--CTAGA-GGACTTAGAGG-.TAGCCTT-AGAC-CATA---TGAGTAGT-CA-A--G-GCTACTGGG-G-CAGGAG-.GTAACC---GGGGAAGC....TATA 1023
WRC.GM.05.Pbt_05GM_X02    ...........................................................................................................................................................................    0
LST.CD.88.524             ...........................................................................................................................................................................    0
LST.CD.88.447             ...........................................................................................................................................................................    0
LST.KE.x.lho7             GAAAGTCTGCTCGGG-CAGG-AGAGAAGACCTGCTT-AGA-GAGGATAAGAAGGA-TTGACTGCTA-AAG--G-C-GAA-GTAAGGA.-AGT-AGAC-TCACC-G-AGA....-TGGCAGC--GC--GGGGC-GCCT-G-GAG-AGT-........TGTAATTCTGATCT 1043
SYK.KE.x.SYK173           ...........CT-G-CC-GGGACCCGAG--GAGGT-AC--CAGTCAAAGG--GT-CGGATC-AA-AGGGATC-GCTGTC--TC-TA.---CGACC-GTGCT--T---T....CGTG-TGT-CACACGAGACG-GGTAAG--AT-C--........CGCCAGACTGCCCA  770
SYK.KE.x.KE51             ..............CT--GCTAGGGA--C--GAGA----ACGCCAG--AAAGG-ACCGCAGAGGAA-G---GC-CGAA-AAGAG-A.-CGGT-TCCGCCCTATGCCGA....TCGGC-ACC-G-TTGAGTAGGTG-G--CTCCGGC-........TCAGACTGCACGCG  407
COL.CM.x.CGU1             AAGTAAA-GTGCGCGCTCGTCGCGCAGT---GAAGT--CAGACGAGAGAG-ACA-AGAG---G-GCT-GCCAGAG-AGGAGAA-AG-.TCATTAG-AG-TCTTCA---A....-GGA-TA-C-CC-CTGGGC-GGTCGAGG-GGGCCT........GCAAGTGGAGTAAG  567
```

```
                                                   Packaging loops end      Gag and Gag-Pol start
H1B.FR.83.HXB2     ..............................................TAGAAGGAGAGAG.....ATGGGTGCGAGAGCGTCAGTA...TTAAGCGGGGGAGAATTAGATCGATGGGAAAAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAATATAAATTAAAACATATAGT  893
Gag                                                                                  M  G  A  R  A  S  V  .  L  S  G  G  E  L  D  R  W  E  K  I  R  L  R  P  G  G  K  K  K  Y  K  L  K  H  I  V
H1A1.UG.85.U455    ...............................................-AGGA---A.......-------------------------AA-A-------TC------G-------------------------C---------G-C-G------T---  360
H1B.US.90.WEAU160  ...............................................-AGGA-----.......-------------------G-----------------------A-G-----------------------T-----------C-----------  892
H1C.ET.86.ETH2220  ...............................................-AGGA-----.......-------------------A--.-------A--C-A-A--------GCC------------AA----------------------C-C----TGC-G-----CC----  273
H1D.CD.84.84ZR085  ...............................................................-------------------------------A---G---GCT-----------------------C-----------------G-----GC-------------  410
H1F1.BE.93.VI850   ...............................................................-------------------A--.-------------A--------GA-------------A--------G-------------G-------A-G------C--A-  248
H1G.SE.93.SE6165   ...............................G-----------------------------------------------.--.-C------A--------GCT-----------------G-------------G----TC-------A--------C----  296
H1H.CF.90.056      ...............................................................-----------------------------C---A--------GCT----G---------C-----------------------------GGC--------C----  240
H1J.SE.93.SE7887   ...............................................-AGGA-----.......-------------------A--.------T------A--------GAT-----------------G-------------GGA----G---C----  216
H1K.CM.96.MP535    ...............................................................-------------------------------A--------GC--------G---------C-----------------------C-G------C----  104
H101_AE.TH.90.CM240 ..............................................-AGGA-----.......-------------------.--.------T------A--------GC-----------------GC----------GA---------GGC-G------T----  458
H102_AG.NG.x.IBNG  ...............................................-AGGA-----.......-------------------.--.------T------A--------GC-----G-----------------------------------G-C--------C----  433
H1N.CM.95.YBF30    ..............................AGAAA...........GTAGGA-----G.......-------------------G..-----CA-----A--------A-------TC----TAT--G--A-------------------C-G-A-G-----T----  451
H1O.BE.87.ANT70    ........GACCTAGGGGAAGGGCGAAGTCTC..............-GG-AG--AG.......-----------T-----T--G...--G-CA--AA-CA----G---GC------C-----A------A------ATCT--A--G------G-C-------CT----  945
H1O.CM.91.MVP5180  ........GACCTAGGGGAAGGGCGAAGTCCC..............-GG-AG--AG.......-----------T--G...--G-CA---A-TA---G---GC------CG----A-------------ATCT--A--GGC----G-GGC-------T----  920
CPZ.CD.90.ANT      ........ATCCTAGGGGAAGGGTCGAAGTCTC..............-G-AC--GA-.......-A--.-G-G-----T--T...--G--G-A-AGA-GC------AC-------GT--CA--C-TC----C--T--C--------G--C-TGA-------C-G--  257
CPZ.CM.01.SIVcpzCAM13 ........GACCTAGAGGAAGGGCGAAGTCTC..............---GG---AG.......-----------------G...----CA----GA----G----C---------G----C-T-A--C-------G-------TGA-G------T----  455
CPZ.CM.05.SIVcpzEK505 ..........CCAGGA.......................G-AGGA-G--G.......-----------------.--.--CA--A---A--------A-------------TAT--G--A-------------------C-TGA-G-----CT----  432
CPZ.CM.05.SIVcpzMB66 ..........GG.CTAG.......................-AGGA-----G.......-----------------T..-----CA------A-------CA------G------TATC-G-----------------------TGA-G------T----  442
CPZ.CM.98.CAM3     ........GACCTAGGGGAAGGGCGAAGTCTC..............-GG-A----G.......-----------------.--.-----CA------CG------GCT-----------A--C-------------------TGA-G------T----  283
CPZ.GA.88.GAB2     ........GACCTAGAGGAAGGGCGAAGTCTC..............-GT-----G.......-----------------TA--.----C---C--GCG------GCC----G-----A--C-C--A----------G-----G---C-TGA-G----------  265
CPZ.CM.98.CAM5     ........GACCTAGAGGAAGGGCGAAGTCTC..............-GG-A----G.......-----------------.--.-----CG--A---CG---G---GCT----G------A--C----M----A-----------------TGA-G------T----  589
CPZ.CM.05.SIVcpzMT145 ........GACCTAGAGGAAGGGCGAAGTCTC..............-GG----AG.......-----------------A--.-----C-------AG------GCT-----------A--C----A-----------------C-TGA-G------T----  462
CPZ.TZ.00.TAN1     ........ACCCTAGAGGAAGGGCGAAGTCTCTAGG.........AC---G--A.......-----------------G..--G--G--A-ATA-GC-G---AC------TCC--AA--C-T-AAT-CA-A--C-G-------TT-A--------C----  456
CPZ.CM.05.SIVcpzLB7 ........GG.CTAG.......................-AGGA-----G.......-----------------A-C..-----CA--C---A-G------GAT--------G----C-------C-------GA---G-----TGA-G------T----  439
CPZ.US.85.CPZUS    ........GACCTAGGGGAAGGGCGAAGTCTC..............-GG-A----G.......-----------------T..C-G-CA------CG------GCT--------G-----C-T--A--T-------------G---TGA-G------T----  935
CPZ.GA.88.GAB1     ........GACCTAGAGGAAGGGCGAAGTCTC..............-GG-----AG.......-----------------T..C---CA------A---------C--------G-----C-T-----C-------GA---G-----TGA-G------T----  930
MON.CM.99.L1       GAGGGAGCATAAAATAAGGGCACG.....................................-----------------G----GCAC--G-CGATGC-T--T--TACTA-------CAAG-AT--G--GG-G---C-T--A--CG-C--T-----G--G--CCTGA-T--G--------  548
MON.NG.x.NG1       ..............................................................-----------------.--.--.--
```

```
H1B.FR.83.HXB2          ATGGGCAAGCAGGGAGCTAGAACGATTCGCAGTTAATCCTGGCCTGTTAGAAACATCAGAAGGCTGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAGAACTTAGATCATTATATAATACAGTAGCAACCCTCTATTGTGTGCATCAAAGGA 1063
Gag                      W  A  S  R  E  L  E  R  F  A  V  N  P  G  L  L  E  T  S  E  G  C  R  Q  I  L  G  Q  L  Q  P  S  L  Q  T  G  S  E  E  L  R  S  L  Y  N  T  V  A  T  L  Y  C  V  H  Q  R
H1A1.UG.85.U455         -----------------G--AA---A--C----C--------T--------G-------A--CAG------------AT-------G-T--C--------A--------------------------------GT-----------A---------  530
H1B.US.90.WEAU160       ----------------------------------C-------------------------------------------AT----------------------------------------------------TGT-------------------A-- 1062
H1C.ET.86.ETH2220       C-------A--------G--AA--T---C----C----A--T----T------C-------A------A-TAA-----------G-T-----------A---G------A-------T--------G----T----------A------A--  443
H1D.CD.84.84ZR085       -----------A-----G-------T---C----------T--------------------A------A--A------------G-T--A--------------------------A-------------------------A--A------  580
H1F1.BE.93.VI850        -----------------------T---C--G--------TC------------------CA-A----A-AA--------------------------------------G---AG------T------------GT-------A--A--------G  418
H1G.SE.93.SE6165        -----------------G--GA---T---C----C----A--T--------G------T--CA-------A--A----T-G-------T--C--------A---G--A--A-----------G-----------------------A----------  466
H1H.CF.90.056           -----------------G--A----T---C----C--C-----T---------C-----------CT---G---A-A-A---A---G--G-TA--A--------A--------A--------T---CT------GT---------C--A---G--AA-  410
H1J.SE.93.SE7887        -----------------G--CA---T---C----C--------TC----GT--G--A-------CA---------A-T------C-----G-T--C---------A--C-----A---A-------G--------------------C--A----G----  386
H1K.CM.96.MP535         -----------------------T---C-C--C--------T----G--A----------C-G------A-AAC---AA---G-----A---A--------------GA---A-------A-----GT-------T---A------A--  274
H101_AE.TH.90.CM240     -----------A--T----A--------C----C---A--T-T--------G------A---CA------A-A-A---T----GT--A-T--CA--------T-------------A--------T-------------------GG--C--A--C-------  628
H102_AG.NG.x.IBNG       -----------------G--A--------C----C--------T--------G---G--A--CA---C--A---A---T----T--G-T--CAG---------------------A--------T------A---------T-GG--C--------------  603
H1N.CM.95.YBF30         -----------------G--A------TTG---C--A--T--CA-G--C---G-G--C-----GCCA-GT----AAAT--AT--G----G-T--CA------G----------GC-C--T--------CG-TC-----GTT--T--------C---AGT----  621
H1O.BE.87.ANT70         -----------A-----G--A-------TG-------AG--AC------TG---G--TAA-GAGA-GC-G--ACAG---T--G-G---G-T--CA------G----CAGC--GCAG---C-C-GG--CG--A---T-GTG----GG-----T--CA-C--AT 1115
H1O.CM.91.MVP5180       -----------------G--A---A---TG--------T--A-------TG-------TAC-GAG---C-G--ACAG---T--G-G---G-T--CA------G----G--C--G-A---TC-C-GG--CG--A-----GTA---GG--C--T--CA-C--AT 1090
CPZ.CD.90.ANT           T--------ATC-----GC-G--T--T--GC-C-GCT-CTC---TC---------------T---GA-A-GGCTA-CCAT--AT-GAGC--T---A-AG-A-T-A---CCCT---A-A-T---T--G-T---C--CA-TTGTGTT--G-GG--C--A---A--G--G  427
CPZ.CM.01.SIVcpzCAM13   -----------A-----G--A-------TG---C--C--G--AA-G----GTG-------A---CTTA--T-G--ACAG---T--G-G---G-T--GA-A--C--C--G----GG--GC----T----T---C--TC-----GTGT-A-GG-----T--CAGT-AC-  625
CPZ.CM.05.SIVcpzEK505   -----------------G--A------GTG---C--A--T--CA-G--C---G-------T---GCC---T----CA---AT--G----G-T--CA-A----G-------G--GC-C--TC---T---C--CC-G---GTT--T--C-----T---A-T-AC-  602
CPZ.CM.05.SIVcpzMB66    -----------------G--A------TTG---C--CTC---CA-G-----A------------AG--GC-TT--CAG-----G-----G-T--CA-A----T------G---------T--G-T---C--CA---T-GT---T-GG--C---------G---  612
CPZ.CM.98.CAM3          -----------A-----G--CA-----TTG---C-----T--AA-G----GG-------A---GAG--GT--T-AA-G--AT--G----G-T--CA-------------GC--GC----GC-T-T------TC----GTG--T-GG-----T---A-G--A-  453
CPZ.GA.88.GAB2          -----------------G--A------TTG-G-C-----T--AA-G-----AA-----------TAT--GT--T--AA-------G-----G-T---A-A------------G-T-AC-G--C----T------TT-G--TGTGT-A-GG-----T---GCT--A-  435
CPZ.CM.98.CAM5          -----------------G--A---T--TTG---C-----T--AA-G--G---G-T-----A---GA-A-TC---AAA----T--G----G-T--CAG------------G-GCT-GC----GC---T---C--TC----GGTA--T-GG-----T---A-G--A-  759
CPZ.CM.05.SIVcpzMT145   -----------A-----G--A------TTG---C-----T--AA-G---ATG-------T---GAGA-GT--T--AA-------G-----G-T---A------G-------G---GC-C--T----T---C--TT-G---GT---T-GG-----A---A-G--A-  632
CPZ.TZ.00.TAN1          ------CG-A--C--A---C-G--T-----GA-G-----C--T--CA-G--G-ACGT---------CT-GA----CA-CCTC-----G--G--T--GG-AG-C-TT--T--TCC----A-C-TT--T--G-T------CA-CTGTGTA-----C--C--A--CGC-G-AG  626
CPZ.CM.05.SIVcpzLB7     -----------------G--GA----T--TTG---C--C--G--TA-G-----A---------GAG---T-GA-TAC----T--G-----G-T---A------G-----A-C----A---T----T---C--CA---TGGT----GG--C--------G--A-  609
CPZ.US.85.CPZUS         T----------------G--A-----T--TTG---C--A--T--AA-G-----G----C-----CTC---T--T-AAA----T--G-----G-T--AA-A-------G---GC--GC-----GC-C-T---C--CT----GGT---T-GG-----C--CAGT---G 1105
CPZ.GA.88.GAB1          -----------------G--A---------TG-G-C--C--G--AA-G---GTAAG----A----CTA--T-GT-ACA---AT--G-G--G-T--CA-A----C-------G---GC-G--C--G-T---C--TC-G---GTA--G-GG--CA-A---AGTGAC- 1100
MON.CM.99.L1            ------TGC--A------T--TA------GTC-CTCGGACTCG--CC-G--G---CAG--C-------AGA-G--TT---A-GTTA-T-TT--TCTA-AG-C----------C--GTCTA-C-A---TC-C-T-GG--TC-CTT-GGTAT-G-----CA-A---GCCG-T-  718
MON.NG.x.NG1            ...........................................................................................................................................................    0
GSN.CM.99.CN166         ------CTC--AA-----T--GA---AT--GC-GTCCGAC-CT--TC-T----ATCA---G--G---C---GG---T-A-A-ATTG-CTTC--CCTAGAA-CA-----G-------CT--C-A---TC-C-T-GGA-TC--GT--GTG---------A-C---GC-GAC-  725
GSN.CM.99.CN71          ------CTC--A------T--GA---A---GC-ATCAGGCA-T--TC-T---T-TAA------G---CA---GG--T---A-GTCACCTCC----TAGAA-CA-AT--G--C---TCT--C-A-AGCC-C---GGA-TC--GT--GTG---------A-C---GC-GAC-  731
MUS_1.CM.01.1085        ------TTC---------T--TA----T-G-C-CTC-GA-TC---T--G---T--AGG--T--G---CA-A-G--CA-A-AGGTCA-C-TC--C-TAGAA-CAG----G--C--GTCC--C-A----C-C-T-GGA-TC--GT--GTG----GG---A-C--CGC-GA-G  723
MUS_1.CM.01.CM1239      ------TTC--A------T---A-G--T-G-C-CTCGGACTC---T--G--G---AA---G--A---CA-A-G--CT-A-A-GTT--CAT-----TAGAA-CA-----G--C---TCT--C-A----C-C-T-GGA-TCACGT--GTG----GG---A-C---GC-GA-G  725
MUS_2.CM.01.CM1246      ------TTC---A-----T--G--G----GTC-CTCAGA-ACG--CC----G-GCAAG--C--A--CCA--GC--C--A-A-GTAA-CTGC--C-TGGAAG-A--T--GAGC--G-CG--C-A----C-C-T-GGA-TC--GT--GTA----GG---A-C---GC-GACG  794
MUS_2.CM.01.CM2500      ------TTC--AA--A--T--CA------G-C-CTCAGACTC---T--------TAAG----G--CCAG-GC--C--A-A-GT-A---T-----TGGAG-C---C--G--C--GTCT--C-A----C-C-T-GGA-TC--GT--GTA---------A-C--CGC-GA-G  753
RCM.NG.x.NG411          -------T-------A----GA---T-GCT-G-G-GA-ACTT---------GTG-------AATCCAGA----TA-A--GGTTA-T-TG--TCTAG---C------G--T--G-GCT-A-A---CC-C-T----TTGTGCTGCGTA--T-GG--C-----CA-GGAAG  485
RCM.GA.x.GAB1           -------T---AAA-A---A-TAA---T-GCT-G-G-GA-CATT-----------G--AC---A---GA-A-----T-A---GTC--G-TG--TCTAG---C------G----G-GG--A-A-AGCC-C-T----TTGTGCTGCGTA----GG--C--A--CA-GGAAG  485
DEN.CD.x.CD1            T---T---AG-A----A-G--G-----T-GCT-AGGAGAACAG---C-------C-AT-----A--CCA-ACT--CA-A-AGACCT--GCT----TAGAA-CACA---CAGT--GAGT--G-A-AGCC---TC-GA---TGCTGCGTTG---GG---TGC--CAGGG-AT  755

MAC.US.x.239            -------GCA-AT--AT----TA----T-G-T-AGCAGAAA-------G--G-ACAA------A---CA-A-------TTCGGTCT--GCT----TAG-G-CA-----C------A-TT-A-A-AGCC-T--------T--CTGCGT-A---GG--CA-T--CGC-GAAG 1326
Gag                      W  A  A  N  E  L  D  R  F  G  L  A  E  S  L  L  E  N  K  E  G  C  Q  K  I  L  S  V  L  A  P  L  V  P  T  G  S  E  N  L  K  S  L  Y  N  T  V  C  V  I  W  C  I  H  A  E
H2A.GW.x.ALI            G------GCG-AT--AT-G--CA------G-T-GGCAGAAA-------G--GT--AA------T--CCA-AG---T--TAA-GTTT--G------TAG-G-CA------------A-TT-A-A-AGCC-T-T------T--CTGCGTAG-T-GG--C-----CGC-GAAG 1369
H2A.DE.x.BEN            G------GCG-AT--AT-G--TAA-----G-T-GGCAGAGA-------G--GT--AA------T--CCA-A-G--T--CA--GTTT--G-T----TAG-A-CA-----G------A-TT-A-A-AGCC-T-T------C--CTGCGT-A-T-GG--CT----CGC-GAAG 1376
H2A.SN.x.ST             G------GCG-AT--AT-G--CA------G-T-GGCAGAGA-------G--GT--AA---G--T--CCA-A----T--TAC-GTTT--G-T----TAG-A-C------G------A-TT-A-A-AGCC-T-T------T--CTGCGT-A-T-GG---A-A--CGC-GAAG  821
H2B.GH.86.D205          -------GTG-AC--AT----CA----T-G-T-AGCAGAGA-------G---T--AA------A---CA-A----T--TAA-GTCT--GCT----TAG-A-CA-----G------A-TT-A-A-AGCC-T-T-----TC--CTGCGT-A-T-T---CC----CGC-GAAG 1352
H2B.CI.x.EHO            -------GTG-AT--AT----GA----T-G-T-GGCAGAGA---G---G-G-T--AA------A--CC-GA----T-G-AA-GTCT--GG-----TAG-A-CA-----T------A-TT-A-A-AGTC-T--------C--CTGCGT-A---T---CC----CGC-GAAG 1353
H2G.CI.92.ABT96         G------GCG-AC--AT-G--CA------G-T-AGCAGAGA-------G---T--AA------T--CCA-A----T--TTC-GTAT--GCT----TAG-A-CA-----T------A-TT-A-AGAGCC-T--------T--CTGCGT-A-T-GG---T----CGCCGAAC  739
H2U.FR.96.12034         G------GCG-AC--AT-G--CA----T-G-T-AGCAGAAA-------G--GG-CAA------A---CA-A----TT-ATCGGT-T--GCT----TAG---CA-----T------A-TT-A-AGAGCC-T--------T--CTGCGT-A-T-GG---T----CGC-GAAG  855
MAC.US.x.EMBL_3         -------GCA-AT--AT----TA----T-G-T-AGCAGAAA----C--G--G-ACAA------A---CA-A-------TTCGGTCT--GCT----TAG-G-CA-----C------A-TT-A-A-AGCC-T--------T--CTGCGT-A---GG--CA-T--CGC-GAAG  812
SMM.US.x.H9             -------GCA-AT--AT-G--CA----T-G-T-AGCAGAAA-------G--G-ACAA------T---CA-A----T--TTC-GTTT--GCT----TAG---C------T------A-TT-A-AGAGCC-T-----C--T--CCGCGT---T-GG--CA-T--CGC-GAAG  808
SMM.US.x.PGM53          -------GCA-AC--AT-G--CA----T-G-T-AGCAGAAA-------G--G--CAA------T---CA-A----TA-TTC-GTTT--GCT----TAG---C-G----T------A-TT-A-AGAGCC-T--------T--CTGCGT-A-T-GG--CA-T--CGC-GAAG 1252
SMM.US.x.SIVsmH635F_L3  -------GCA-AT--AT-G--TA----T-G-T-AGCAGAAA-------G--T-ACAA------T---CA-A----T--TTCGGTTT--GCT----TAG---C------T------A-TT-A-AGAGCC-T--------TA-CTGCGT-A-T-GG--CA-T--CGC-GAAG 1338
SMM.SL.92.SL92B         -------GC---A--AT----CA----T-G-TCGGCAGAAA-------G----GCAA------A---CA-AG----T-A-C-GTA---GCT----TAA-G-CA-----C--G---A-TT-A-A-AGC--G-T--G---T--CTGCGTAG-T-GG---C----CGC-GAA-  766
STM.US.x.STM            G------GCA-AC--AT-G--CA----T-G-T-GGCAGAAA-------G--GT--AA------G---CA-A----TA-AACGGTCT--G------TAG-G-CA-----T------A-TT-A-AGAGTC-T-T------T--CTGCGT-A-T-GG---A-C--CGC-GAAG  982
MNE.US.x.MNE027         -------GCA-AT--AT----TA----T-G-T-AGCAGACA-------G--G-ACAA------A---CA-A-------TTCGGCCT--GCT----TAG-G-CA-----C------A-TT-A-AGAGCC-T--------T--CTGCGT-A---GG--CA-T--CGC-GAAG  806
DRL.x.x.FAO             C----T-GCG-A---A-----T-------GCC-GC--GAGA-AT-A-------GTCAG-----G--CCA-A----TT-A---GTA--CTTT--TCTAG---CT-----G-----GA-TT-A-TT--TC-C-T------CTGCTGCGTAG-T-GG--C--T--CTCG-AAG  564
MND_2.x.x.5440          -----T-----A---A-----CA----T-GCC-AC--GAGAAGT---------GTCAG-----A--CGA-A----T--TA-CGTA--CTTT---CTAG---CT-----G------A-TT-A-TT--GC-G--C--C--CTGTTGTTG-A-T-GG--C--A---GCG-AAG  567
MND_2.CM.98.CM16        -----T-----A---A-----CA----T-GCC--C--GAAAAG----------GTAAG-----A--CGAGA----T--TA-CGTA--CTTT---CTAG---CT-----G------A-TT-A-TC--GC-G--C--C--CTGCTGCTG-G-A-GG--C--A---GCG-AAG 1010
MND_2.GA.x.M14          -----T-----A---A-----T-----T-GCC--C--GAAAAA--A---------TCAG-----A--CGAGA----T--TA-CGTA--CTTT---CTAG---CT-----G-----GA-T---CTTG-GC-G--C--C--CTGTTGTTG-A---GG--C--A---GCG-AAG  934
OLC.CI.97.97CI12        -------GTACAA---T-GA---ATC-ATGCA-AG----A-ATGATC---G---TAAG----------A-AG---TG-A-A-A-AT---TG--T-TGGAAG--GG----A-TC-GA-GT-G-A-AGT--GCTG--C-TCTGTAT-GTTA-T-TG---T----AG--G-CC  979
GRV.ET.x.GRI_677        -------G-A-AA--AA-G-------T-GGT-AC--GAGAAA--T---------AA----------CA-A---CA-A-A-GTTT--ACC--G-TGGAA-C-----C--C--G-GG--A-A-G-TC-G-T----TTGTGCTGCGT-A-T-GG--CA-T--CGC-GAAC 1170
VER.KE.x.TYO1_patent    -------G---A---A-G--G--C----GCC-CC--GAGA-GT----G--G---GAG--G--G----A-AG---CA-A-A-GTC--CT-C--CCTAGAA-CA--------G--G-GCT-A-A-AGTC-G-TC---CTT--GTGCGTA--A-----CT----CA-GGAAC 1212
VER.KE.x.9063           -------G-T-AA---A-G--G--C----GCC-CC--GA-A-GT----------GA---T--A----A-A-G--CA-A-A-GTC--TTTT--T-TAGAA-CA-----C--G--G-GCT-A-A-AGTC-G-TC---CTT--GTGCGTG--A-----CA-C--C-GCGAAC 1215
VER.KE.x.AGM155         -------G---A-A--A-G--C--C----GCC-CC--GAGAAGT-A--G--G---GAG-----T----A-A-G--CA-A-A-GTT--CTCT--CCTAGAA-CA-----G--G----G-A-G-A-AGTC-G------CTG--GTGCGTAT-GCT---C--C--C---GAA- 1204
VER.DE.x.AGM3           -------G---A---AA-G--G--C----GCC-CC--GAGA-AT-AC-----T--GA------A----AGA-G--CA-A-A-GTA--CT-C--GCTAGAA-CA-----G--G--G-GCT-A-A-AGTC-G-T---CCTT--GTGCGTAT-G-T---C--A--CA--GAT-  704
TAN.UG.x.TAN1           -------G---AAA-AA-G--G--C--T-GGC-CC-CGAGAAA--C----------AA---G------CA-A-G--CA-A-A-GTAT-GAGC--GCTAGAA-CA-------------G---A-A-AGTC-G-T----TTGTGCTGCGT-A-T-GG---A-T--CGC-GAAC 1175
SAB.SN.x.SAB1C          -----------A---A-----T-----TAGCC--TCAG-AAAT--C-----G---AAG----GGTCGTCA----T--TA-TGTA--C-TT--CCTAG---CT-----G-----GA-TT-A-TTG-TC-C-T----TT-TGCTGCGTGT-AGC----A-T--CGC-GAA- 1290
MND_1.GA.x.MNDGB1       T---TGT-AAG-T--AT----TA----T-GCT-ATCGGA-AAA--CC-T------CAGC----A--GA-A-----C--CTC-GTATGTTGG---TATA-G-CCA---------TA-T--A-A-G-T--GGTAGGG-----CTGTGTTG-AGCC--CA-A--CGC-G-T-  723
DEB.CM.99.CM40          C------G-T-A---A-----CA----T-GCC--GCAG-A-CAT-A-------ATAAG---------CAGG----TT--ACTAG-T----G--T-TAGAAGCA-----CAGT--GA-TT-G-A---TC-----GGC-T---CTGTGT-A---GGGCATGC---GCC-ACC  761
DEB.CM.99.CM5           C--------T-----A-----CA----T-GCC--GCAGGA-CAT-A--G-----CAA------A---CA-G-G--CT-AACTAGAT-------T-TAGAGGCAT---CAGT----GC----AG--CC----CGGC-T----TGTGT-G-T-GGGCATGC--CGCC-ACC  764
TAL.CM.01.8023          ------TTC--AA---A-G---A-G--T-GCC-G-G-GAC-C---T--G--G---CAA------G--CGTTA-G--CA-A-A-GCCA-TTCT--G-TAGAA-CA-AT--C--------G-A---AG--TC-C-T-GGA-TC--GTGTGTT--A--C--CA-T---GC-G-T-  425
TAL.CM.00.266           ------TTC--AA---A-G--GA-G--T-GCC-G-G-GAC-C---------G--CAA------G--CGTTA-G--CA-A-A-GCCA-TT-CATG-TAGAA-C--AT--C--------GGA-A-AG--TC-C-T-GGA-TC--GTGTGTT--A--C--CA-T---GCTG-T-  939
SUN.GA.98.L14           -------G-T-AA--AT----TA-G--T-G-C----CAAG-AA------AG----GTG---------CAGA-G--T---TC-GTCTGTTGG--T-TGTA-GCAT-T--T--TA-GA-C--G-AGG-C---GTAGG-------TGTGTG--TGCC---TGT------G-C- 1336
WRC.CI.97.97CI14        ----TT-GTG-A---AT----GAAG--GTGCT-AG--T-A-CAT-AA-CAGT---G-------GGTG-AGA-GC-TA--TC---T---GCT--T-TA--G-CTGAT-CC--TA-G---G-A-AGAGCA-T-GGGGA-TGT--TGTGTGT-AGCA---C-C---G-G--A- 1321
WRC.CI.98.98CI04        ----CT-GTG-A---------GAA---GTGTT-AG-CT--ACAT-AA---C----GA-AC---GAC--A-A--C-GA---A---TT--GC-----TAG---CAGAT------A-----G-G-A-AGCC-T-GGGGA-TGA-GTGTGTG--TGCA---T-----GC-C-T- 1327
WRC.GM.05.Pbt_05GM_X02  T---CT-GTA-AA---T----GAAG--ATGTC-AG--GGGAAAT--A--TCT-GCCA------AGTG-AGA--C-CTGT-AG--TA--GCT--T-TGG-A-CAGA---G---A--AC-A-A---AGTA-T-GGGGA-TGA--AGTGTG--AGCA--CT-AT--A-T-AC-  277
LST.CD.88.524           C-------CG-AA--A-----TA-G--T-GGC-AGGAG-AAA---TC--------G----G-------AGA-G--TT-----GTC-GTTGG--T-TATA--CT-----C--TA--A-T--G-A-G-TC--GTAGGG-----GTGTGT-A---G---CTGT--C-T-G-A-  274
LST.CD.88.447           G-------CA--A--AT----TA-G--T-G-T-AGGAG-ACA-T--C----G---G----T-----C-AGA----T--T---GTCTGTTGG---CTATACTCC-----C--GA-GA-TT-A-A-G--C-TGTAGGA--T---TGTGTG--G-G---CTGC---TT-G---  274
LST.KE.x.lho7           --------CG-AA--AT----CA----T-G-T-GGGAT-ACAA--C--G------G-------G----AGA----C--ATCTGTCTGCTGG--T-TATA-GCA--T--T--TA-GA-C--A-AGG-C---GTAGGG------TGTGT-A---GC--CTGT--C-T-G-AG 1349
SYK.KE.x.SYK173         T-------AG-A----T-G--CA------GCC--TCGGACCAA--CA-G--GT-TAA----------GA-A----T--CAC-GTC----TT----TAGAAGCT-AT--G------A-TT-A-A-AGCC-T-T-GGC-T-A-TT-GGT-G-A-GGGCC--T---GCC-AA- 1071
SYK.KE.x.KE51           T------G-A-A------G--CA------GCC-CTCAGACCAG--TC-T----GTAA------T---GA------CA-AAA-ACCA-CTT-----TAGAAA-ACAT--C------A-TT-A-A--------T-GGC-T-AC----GTAG---GGGCC--C--CGC-GAA-  708
COL.CM.x.CGU1           ----ATGT---CA---G--AGC--C-GT-TTT-A-TATT--AG------A-GT--G--AC---AGTGGCT------T---A-A--G-GACG----TAG-GG-C-------------TGT-G--G--CC-----GGA...TGTT-TGTGTGT--C--CC----CAG--AAT  874
```

```
H1B.FR.83.HXB2     TAGAGATAAAAGACACCAAGGAAGCTTTAGACAAGATAGAGGAAGAGCAA.........................................................AACAAAAGTAAGAAAAAAGCACAGCAAGCAGCAGCTGACACAGGA 1158
Gag                I  E  I  K  D  T  K  E  A  L  D  K  I  E  E  E  Q  _  _  _
```

```
                            Gag p17 Matrix end Gag p24 Capsid start
H1B.FR.83.HXB2          ...CACAGCAATCAGGTCAGCCAAAATTACCCTATAGTGCAGAACATCCAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAATGCATGGGTAAAAGTAGTAGAAGAGAAGGCTTTCAGCCCAGAAGTGATACCCATGTTTTCAGCATTATCAGAAGGAGC 1325
Gag                     _._ H  S  N  Q  V  S  Q  N  Y  P  I  V  Q  N  I  Q  G  Q  M  V  H  Q  A  I  S  P  R  T  L  N  A  W  V  K  V  V  E  E  K  A  F  S  P  E  V  I  P  M  F  S  A  L  S  E  G  A
H1A1.UG.85.U455         ...............AG---T----------C--------A--TGCA--A-----CCA----C-----T---------G--C--G----------G-------------C------------------A---------------------G---- 780
H1B.US.90.WEAU160       ...A----A--C-----------------------------C-A----A-----------------C-------------------------------------A-------A------------------A------------------------------------ 1324
H1C.ET.86.ETH2220       .........GGAA---A---T-------T-------------T--G------------------C-------G---------G--------------------G--A--------------G--A----------A----T------------- 699
H1D.CD.84.84ZR085       ...A------GC----------------T--------A-----C-A--A-----------C-----C-----------------G--C---------------A-------A---------A-------A---------------------------- 842
H1F1.BE.93.VI850        ...............-----T-----------------A-----TC-T-----A-----------C---T-TC-----------------------------G--GA--------------------T----------A---------------------------G-- 668
H1G.SE.93.SE6165        ...A-------C--A-------------T-------------TGCA--A------------C---------A-----------G------------------------------A-----C----T-------A---------------------------------- 728
H1H.CF.90.056           ...A-AGA---CA-------T-------T--------A----TGCT--A-----G-------C---------------G--C------------------------A-------T----------A-------------------------------- 672
H1J.SE.93.SE7887        ...G---A--G---------T-------T-------------TC-G--A-----CC-----C-----C------------------------------GA-------A--A---------------------------------------------- 648
H1K.CM.96.MP535         ......GA---AGG------T-----------------A-----TC-G---------------C-----C---------------------G--G---A---G-------------------A-----------A------------------ 524
H101_AE.TH.90.CM240     ...AG-----GCA-A--------------------------A--TGCA--A-----------C-------C-TT---------------G---------G--------------A----G---T-A--------A-----------C-------------G----- 890
H102_AG.NG.x.IBNG       ............AGCAG------------------------A--TGC-A-A---------AC-------T----G---------------------------G------A-------A----G---------------A-----------------G----- 856
H1N.CM.95.YBF30         GCAACTGAT-GCA-TA----TAGG-----T---C----C----CTGCT--A--A---------------C-GC-GA----C-----C-------T-----G-----GA----G--------C--T--T--------A-----A------ATG--C--G--------G-- 889
H1O.BE.87.ANT70         ACAAG-GCA-GG----CGG-T-----------------ATCA--TGCG-----A-----------------C--C--C--G---------------------G-C----------A--A--C--T-A---T---A-C--T--T-----CATG-----G-----G----- 1371
H1O.CM.91.MVP5180       ACAAG-CCT-GG---ACA--T-----------------AACA--TGCA-----A-------------A----C--C--C--G---------------------G-C-----------------C--T-A---T---A-T--T--T------ATG--------------G-- 1346
CPZ.CD.90.ANT           CGA--TCTT-G-GGC-AAG-GAG---------C--CA-AGT-G-TGCAGGA--AAT-GCAAGG------C-AC-GA----A-----C--------C--------GTGT-----------AAA-----AT--------C--C--T---------T--T-----------G-- 737
CPZ.CM.01.SIVcpzCAM13   AGTGCTTC-TC-GGCACT--TGGG---------------A-----TGCT--A-----G---------ACTC----C--C--------C----------------------G-----------------T-AT--T--G--A-----A-----------TC-G----------- 902
CPZ.CM.05.SIVcpzEK505   GCA-CTGAA-GCAGTA----TAGG--C--T---C-G--A--A--TGCT--A------------------C-GC-G-----C-----C-------T-----------G--------A-----C--T-AT---------A-----A------ATG--C--G--------G-- 873
CPZ.CM.05.SIVcpzMB66    .......-CT-GCACT-----TAGG--C-TT---------AGCA--TGCT---------G-----------AC----C--G-------------------------------------A-----A--T-----------A--T---------CATG--TC-G----------- 868
CPZ.CM.98.CAM3          CAG--AGA--GCGCT-----TAG---C-----AG----A-----TGCA--A--A--GC-A--G------C----G-----C--G--C---------T-----T------A-------A---AA---T-A------G--------A-----CATG--CC-G----------- 715
CPZ.GA.88.GAB2          TCAA---C--C-G--A-A---AG----------AG-G-------TGCA-----A--G---A------------G-----C---------G---C-----T----C---------A--A--C--T------------C--T--A-----CATG--T------------G-- 703
CPZ.CM.98.CAM5          CAG--GGAA-GCAGCA-A--TAG---C-----A-----A-----TGCA--A--A-----A------------G--------G--CC------T-----C--G---A-------A---AAC--T-A---------A-----A------ATG--CC-G--G--G----- 1018
CPZ.CM.05.SIVcpzMT145   GGTG-A-CACCATCA-CT---GCG--C-T----G----A-GA---GCT-----A--GT--ACG---A-TC-A--------C-----CC-------C----------C---------A--A-------AT-----G--C-----A-----CATG---------------G-- 864
CPZ.TZ.00.TAN1          ACAG-G-CACC-AGT-GG--TAG-CTA-----AG-GA-AACTG-TGCA-----AGTTGCAAGG------C-T--T-------------C-------C-------GG---A-------A--A-GG--T-AT---------A-----A-----C--------G--T--G----- 957
CPZ.CM.05.SIVcpzLB7     AGGG---ATG-CAGT-GT---AG---C-------------ACA--TGCA--A-----GCC------C------C-C-------------------C-----C--G-----------A-----A--T-------------A-----------CG----C---G----------- 871
CPZ.US.85.CPZUS         GGC...--T--CATT-GT---AGC---------AG--A-A------GCT-----A--G---------------A--G--T--C----------------C-----------CG-------A-----C--T-A---G-----C--C--G-----CATG--C------------G-- 1373
CPZ.GA.88.GAB1          AGTGC-TCTGC-GGCA-T--TGG---------CC-C--A--A--TGCT--A-----G-----G----------C-----C-----C--G--------------------G------G--A-----A--T-----T--G--A-----A-------------TC-------G----- 1386
MON.CM.99.L1            CCAACAG--GGCGTACC---TGG---C------G-C---AGA-CACAGGG---AGGCT-CCA---------AG-GGAG---C-GCTCC-C--GA--------C-G--CA-T--G------AAA--TGC----------G-GG--C-C--CCAG--TC----------CAT 989
MON.NG.x.NG1            ..................................................................................................................................................... 0
GSN.CM.99.CN166         GGGGGACAGGCGGTA--T--T-----C-------G-GA-TAG---TGC---A--A--GTATCA----------C-CAATAG-----T-C-T--GA-T-------GCA----G--G-----AAAG--TGCT-----GACAG-TG--T-----CAG--T------------CTG 1002
GSN.CM.99.CN71          GGGGGACAGGCGGTA--T--T--G--C-----C-----TAGA--TGCT-----A--GT-TCA---C------C-CAACAGC---GTCC-T--GA-T-------GCA----G-----A---AAG--TGCT-----GACAG-CG-TC-----CAG--TC--A-------CTG 1008
MUS_1.CM.01.1085        CAAGCGCAGGCAGCT...--TGCC--C-----AG-CA-AAG----GCA--A-----GT-CCA-----------C-CAATG------TA-----AA-------GTC-A----G--G--A---AAA--TGCTG-------TG-GG-A------CAG-----GG----G--G-- 973
MUS_1.CM.01.CM1239      ACAGCACAG-----AACT--TGCT--------C--CA-AAG----CCA-----A---T-TCA---C---T--C-TAATG------TC-----GA-------GTC-A----G-----A---AAA--TGCAG-------AG-GG-A------CA----C--G---------A- 975
MUS_2.CM.01.CM1246      GCAA-AGC--CCACCAGTG----G---------AG-CA-AAGA-CAGCA-----A--GTATCAG------T-TC-TAGC-----GCTA-----AA-C-------TC-AC-------G-----AAAA--TGC----------AG-GG--C----CCAG--CC------------TG 1056
MUS_2.CM.01.CM2500      TCAGGAGCAGGGGCAAG----------C-----AG-GA-TAG--CAGCT--A---ACCTATCA-------T--C-CAGC--C-----T-------GA-C-----GTCCAC-A-T----------AAA--TGC----------AG-GG-TT----CCA---C------------TG 1018
RCM.NG.x.NG411          GTGACATCT-G-GGACAA--AGG----------A-----AACA-TT-AT...CA---GCCT-A--------C-A-----T---------A--------T-------------------G--A---AAA--TG-AG-------AG-G--T---------G---C-G---------CTG 753
RCM.GA.x.GAB1           GCG-CACCT-G-GGACAA--AGG------T------A-TACT-TA-AT...CA---GCCT-AG---A-TC-T---------A--G---C---------C-----C--G--G-----G----AAAA---TCAG-------AGCG---------C--G---C-----------CTG 753
DEN.CD.x.CD1            TCCTTT--AG---CAAATTCT--------T--A--T--CAG--CAGCT-----A---TATCAG-----A---T--AGTA-C--GGTG-----AA-----A-T----AG-----G--C-GAAAA--TTCG-----GA-AG-G--AC-A-----G-TCC-CG-C-----GCT 1092
                            Gag p17 Matrix end Gag p24 Capsid start
MAC.US.x.239            GCA-CATCT-GCGGCAGAG-AGG---------AG--CAA--A-TAGGT...--TA-CTAT--C--C-T-C-AT--AGC--G-----A--------C---------T-GA----G--A---AAA--TG-AG-------AG-G--AGGA---CAG---C-G--------TTG 1594
Gag                      A  P  S  S  G  R  G  G  N  Y  P  V  Q  Q  I  G  .  G  N  Y  V  H  L  P  L  S  P  R  T  L  N  A  W  V  K  L  I  E  E  K  K  F  G  A  E  V  V  P  G  F  Q  A  L  S  E  G  C
H2A.GW.x.ALI            GCA-CACCT-G-G-ACAGG-GGG---C-T---CG--CAA--AGTAGC-...--CA-CTACACC---GT-C-GC-GAGC--CC----C---G----T---------T----------A---AAG---G-GG-------AG-G--AGGA---CAG---C-C---------CTG 1637
H2A.DE.x.BEN            GCA-CACCT-G-GG-AAA--AGG---C-----CG-GCAA--AGCGGGT...--CA-CTAT--C---GT-C-AC-GAGC--CC-----C-----------------T----G--G--A---AAG---G-GG-------AG-G--AGGA---CAG---C-C---------CTG 1644
H2A.SN.x.ST             GCA-CACCT-GCGG-AA-G-AGG---C-T---CG--CAA--AGTGGC-...--CA-CTATACC---GT-C-AC--AGT--CC----CC-------T----------C-------G--A---AAG---G-GG-------AG-G--AGGA---CAG---C-C---------CTG 1089
H2B.GH.86.D205          GCA-CACCT-GCGGC-GA.......-----T--AG-GCA---ACTGGCT...--CA-CTAC--C--C-T-C-GC--AGC--CC----C--------T--------GT-----------A---AAG---G-GG-------AG----AGGA---CAG---C------------TG 1605
H2B.CI.x.EHO            AAA-CA-CT-GC-GCC-A......GC---T--AG-GCA---A-TAGCT...--CA-TTATTCC--C-T-C-GC--AGC--CC----C--------C----------T------------------AAG---G-GG-------AG----AGGA---CAG---C-----------GTG 1606
H2G.CI.92.ABT96         GCA-CACCT-G-GGCAGAG-AGG---C-----AG-GCA---AGTAGG-...--CA-CTAT--T----T-C-AT--AGT---C-------------------------T-G---------A---AAG---G-GG--------G-G--AGGA---CAG--GC-G---------CTG 1007
H2U.FR.96.12034         GYT-CACCT-G-GGC-G---AGG---------GG-GCA-----TAGGT...--CA-TTAT-CC--C-TAC-TC--AGT--------AC-C-----C---------C-G-----G-----AAAA--TG-AG-------AG-G--AGGA---CAG---C-C-----G--CTG 1123
MAC.US.x.EMBL_3         GCA-CATCT-GCGGCAGAG-AGG---------AG--CAA--A-TAGGT...--TA-CTAT--C--C-T-C-AT--AGC--G-----A--------C----------T-GA----G--A---AAA--TG-AG-------AG-G--AGGA---CAG---C-G---------CTG 1080
SMM.US.x.H9             GCA-CACCT-G-GGCAGAG-AGG---------AG-GCA---AGTAGGT...--CA-TTATACC--C-TAC-TT--AGT--A-----A--------T----------T-G-------------AAAA--TG-GG----G--AG-G--AGGA---CA---GC-----------CTG 1076
SMM.US.x.PGM53          GCA-CACCT-G-GGCAGAG-AGG---------AG-GCA---A-TAGGT...--CA-TTACACC--C-TAC-TT--AGT--A-----A--------T----------T-G-------------AAAA--TG-GG----G--AG-G--AGGA---CA---GC-----------CTG 1520
SMM.US.x.SIVsmH635F_L3  GCA-CACCT-G-GGCAGAG-AGG---------AG-GCA---AGTAGGT...--CA-TTAT--C--C-TAC--T--AGT--A-----A--------T----------T-G-------------AAAA--TG-GG----G--AG-G--AGGA---CA---GC-----------CTG 1606
SMM.SL.92.SL92B         GCT-CACCT-G-GGA-GA.......--C--T--AG-GCA---AGTAGGA...AATA-TTAT--G--CACTC-AC-G--C--AC-G--AC-C-----G-----C---C-G---------A---AAA--TG-AG-------AG-G--AGGA---CAG---C-----------CTG 1028
STM.US.x.STM            GCA-CACCT-G-GGCAGAG-AGG---------AG--CA---AGTAGG-...--CA-TTAT--C--C-TAC-AT--AGC--A-----C--G--------------GT----G--G--A---AAA--TG-TG--------G-G--AGGA---CAG---C-----------CTG 1250
MNE.US.x.MNE027         GCA-CATCT-G-GGCAAA--AGG---------AG--CAA--A-TAGGT...--TA-CTATACC--C-TAC-AT--AGC--G-----A--------C----------T-GA----G------AAA--TG-AG-------AG-G--AGGA---CAG---C-G---------CTG 1074
DRL.x.x.FAO             GCGACATCT-GCGGCAGGTCAAG---C-----C--TCA-GTTGTT-AT...CAA--GGCA-----C-----T--T-----A-----C--------C-----G--G---A-T--G--A---AAG-----TG--------G-G--T-----C-TG--CC-G---------CTG 835
MND_2.x.x.5440          GCGACACCTGC-GTTCG-TCAA----------C--TCA-GTA-TA-AT...CAAAC-CCA----------GA--T-----A-----C--G---------------TGTA----G-----AAAG-------------A-AG-G--T------ATT--C--G---------TG 838
MND_2.CM.98.CM16        GTGACACCT-G-GGCCG-TCAA----------A---CA-ATA-TA-AT...CA-ACCCCA--C--------GA--T-----GC-C--GC-G-----G----------TGCA-------------AAG--T-----------A-AG-G---------CAT---C--G--T-----GTG 1281
MND_2.GA.x.M14          GCGACACCT-G-GGCCG-TCAA-------T-----TCA-ATA-TA-AT...CAAAC-CCA--G-----A-GA--T-----A----------------C-----C---TGTA-T--G--A---AAA-----T------A-AG----T-----CAT---CC-------------TG 1205
OLC.CI.97.97CI12        .......-AG--AG-A-AGCAGTT-GCCATG--A---A-CAGA--TGCT-----AATCT-T-AG---GT-C-GC-GG-TA-------CC-GC-AAGT-----------C-T-G-----C-G--G---T-AAT-----A-AG-G--TC-A------T--G--TG-T--G---TT 1193
GRV.ET.x.GRI_677        GCG-CACCTGG-GGC-AATCAAG---------AG----AA-TC-G-AT...AATGCCTG-------C---C-TT-G--T--GC-C--G---------G-----C---TGC--G--G--A--AAGG-GGG-AG-------AG-C---------CCA----C-C-----G---TG 1438
VER.KE.x.TYO1_patent    GCG-CACCTGG-GGC...--T--G----TT--AGCGCAA--AC-AGGA...AATGCCTG-------GTAC--T-G-----GC-C--C---------G-----------C-------G-----AAAA--TG-AG-----A-AG------------CA---CC------------CTG 1498
VER.KE.x.9063           GCG-CACCTGG-GGC...--T-------T---AGCGCAA--AC-AGGA...AATGCCTG-------GTTC-AC----C---C-C--A---------G-----------C-------G--A---AAA--TG-AG----GA-AG-G--A-----CCA----C-C---------GTG 1501
VER.KE.x.AGM155         GCGACACCTGG-GGC...-----------TT--CGC-CAA---C-AGGG...AATGC-TG---G---GTAC-AC-T-------C-C--CC-------------------C-------------AAAA--TG-GG-----A-AG---------CCA---CC-C---------CTG 1490
VER.DE.x.AGM3           ACAGTGCCACC-GGT-G---T-------T---AGC-CAA--AC-GGGA...AATGC-TG-A-----GT-C--T-G-----AC-C--C---------G-----------C-------G-----AAAA---G-AG-----A-AG-G---------CCAG--T-----------GTG 1002
TAN.UG.x.TAN1           ACA-CT--TGGCAGA...CATGGT---------G----A--AC-G-AT...AAT---TG-------ACTC-TT-G-----AC-T--CC-------------T---AC---T--G--A--AAGG---G-GG-----A-AG-----------CCA---CC-C---------GTG 1461
SAB.SN.x.SAB1C          GTGACACCT-G-GGT-G-TCG-G---------C--T--CAGTGTT-AT...AAT---TG---C--C---C-AT-G--C--A--G--AC-T-----T-----------GA-T--G--A---AAG-----TG----G--AG-----------AGC--CC--G----------- 1591
MND_1.GA.x.MNDGB1       ....................--TA-G----T----G--CAAAG-G-TGCAGCA--A--GTATCAGT--ACTC-A---AGT-----G-T-A--C-GA-----------AC---G-----A---AAG-GG-AA--G--G--C--C--TC-A--C---------GA----------- 952
DEB.CM.99.CM40          GCA-C-CAG-CAACACCATCAGTT--C---------CT-AAGAC-GGGG...CAA--GT-T--G---AT-C-AT--AGC--C--G-TAG----AA-C---A----T-CT-----G--A---AAA--TTCT-----GA-AG-T--TC-T---CA--TC--GG----G--CTG 1095
DEB.CM.99.CM5           CAG--A-AA-CAA-CCAACCAGCT----TT--A--CC--AGAC-AGGA...CA----TAT-------AT-C-TT--AGT-----G-T-G-G--AA-C---A----T-C---G--G--A---AAA--TTC------GA-AG-G---C-A---CA--T---GG-------GTG 1086
TAL.CM.01.8023          GTGACATCT-G-GGCCAT--TGGG--------C--TA-AAGA--TGCA--ACAA--GTATCAG----------T-G-----C---CTGC-G--AA-----A--GCCACTA--------A---AAA--TGC------GACAG-GG-GC-T--C--T---C--G---------TG 711
TAL.CM.00.266           GTGACATCT-G-GGCCAA-CAGG---------C---A-AAGA--TGCA---CA---GTATCA----------AC-G---------GCTGC----GA-----A-TGCCACTA----G-----AAAA--TGC---T---ACAG-GG-GC-C--C--G---C--G-G-----GTG 1213
SUN.GA.98.L14           ...................TCT--------TT--AG-GCA-AGAG-GGGA...CAAA-TTACA-T-------C-TC-C--G---------AG-GC-AA-------G--GA----G----------GA-GG-AA------ACAG-GG-A-------G----T---ATGAC------ 1559
WRC.CI.97.97CI14        ...................TCT--G---ATG--AG-GC--AGA-CAGCA...CAAGGGCCTCA-TGGG-CC--T-G-----------AA--C-AA-------G---CATA-T--G--T-G--GG--TG-AG-----AGTG-T--AT-C---G----TC-TA----CA-G-- 1529
WRC.CI.98.98CI04        ...................TCT-----CATG--AG--T--AG--CAGCT...CAAGG-CCTCA-TGGG-TC-TC-G--T--A-----CA--C-GA-T-----G---CACA---G--T-GA-GC--TG-AG-----TGTG-T--AT-C-----T--T--GA----TA---- 1547
WRC.GM.05.Pbt_05GM_X02  ...................TCAGTG---ATG--AG-CC-TAGAGGACCA...CAAGG-CCACA-TGGG--C-TC----C--G--G--CA-TC-AA-----------CATA--------T-G--GA--TG-AG-----TGTG-T--AT-C-----T--T--GA----TA---- 491
LST.CD.88.524           .......CCTGTG-CA-AA--TA------TT---T-GA-TAGAG-A-AT...CAAAG-TG---C---ACCC-TC--------C-----AA-TC-AA-------G---A----------T-G--GA-GG-AG------ACAG-GG-T------------TC-GA-----AA--- 491
LST.CD.88.447           ...G-A--GC-G-CTAAA--TGGC------T--AC-C--CAGAG-A-AT...CA-AG-TG-------CACAC-AT-G------------AA--C-AA-T----------A-----------T-GA-GA-GG-AG------ACAG-GG-A---------T--CC-CA------AA--- 497
LST.KE.x.lho7           ..........C-A---ACAG-GGG------T--CC-GA-AAGAG-A-AT...CA-AG-TG-------ACAC-TT----G--C------AA--C-AA-------G---A-----G------C-GA-G--GG-AA------ACAG--G-A-----------C---A------AA--- 1581
SYK.KE.x.SYK173         GCAGCA-CACCG-CACCT--AGG---C--T--AC-GC-CAGA--TCCT--AAAT--GTG-A-T---ACA-GAG-CC--GT---G---C----AA-------GG-G-CT--GA-CTCC--AAAA--TGATG--TCCA-AG-G--AT-A---CAG--TC--A---------TT 1405
SYK.KE.x.KE51           ATG-CATCT-G-GGC-G-TCAGG-----------C--A-TAGA---CCA--AAAT---TG-A-T---GTA-GGG-CAACA-C--G---------AA-C-------G-G-C---GA-CAGC--AAAA--TGATG-CTC-C--G-C--AC-A--CCA-ATT---A-------TTT 1033
COL.CM.x.CGU1           .......-AGG-AGCA-AG-AATTGG-CATG--A-----AACAGGGCCT...CAAGGGCCA--G-----AC--C-GAGC--C--G--G---GGA--C-----T--GTGT-----G...GGA-GAA-TGCT--TTCCT--GCT---------C-T---T-AT--TACG--G-- 1133
```

```
H1B.FR.83.HXB2              CACCCCACAAGATTTAAACACCATGCTAAACACAGTGGGGGGACATCAAGCAGCCATGCAAATGTTAAAAGAGACCATCAATGAGGAAGCTGCAGAATGGGATAGAGTGCATCCAGTGCAT...GCAGGGCCTATTGCACCAGGCCAGATGAGAGAACCAAGGGGAAGTG 1492
Gag                          T  P  Q  D  L  N  T  M  L  N  T  V  G  G  H  Q  A  A  M  Q  M  L  K  E  T  I  N  E  E  A  A  E  W  D  R  V  H  P  V  H  .  A  G  P  I  A  P  G  Q  M  R  E  P  R  G  S
H1A1.UG.85.U455             ------------------T-TG-----G--TGT------------C--G-----T------------------------T-----------------------G-----C--GT-A------------...------------C--------------------------------- 947
H1B.US.90.WEAU160           ---------------------------------------------------------------------------------------------------------------T-A------------...---------------------------------------------- 1491
H1C.ET.86.ETH2220           ------------------------------T-----------------------------------------C--------------G--------------C--GT-A------------...---------G--------------A-------C------------- 866
H1D.CD.84.84ZR085           ------------------------------------------------------------------------------------------------------C--GC-A-----------G...---------G------------------------------------ 1009
H1F1.BE.93.VI850            ---T--CAC---------------------T-----------------------------------------C-----------------------------C---T-A------------...---------GCCC-----------------G-----T-----G--- 835
H1G.SE.93.SE6165            ------------------TCT----T------------G-----------T---------C----G--T--T--------------------G-----C--GA--------CAA--G...---------T--C--------------A------------------ 895
H1H.CF.90.056               ------------C-----TG-T--------T-----------------------------G-----------T--A--------------------------C--G--A------------...------------C-----------A-------------------C- 839
H1J.SE.93.SE7887            ------G-----------T-------------A-A--------C--------T-----------------T--T--------------------------C--G--A--------A---...------------------------G--------------------- 815
H1K.CM.96.MP535             ---T---------C------------------------------------------------------T--T--------T-----------------C--GT-A--C---------...-----------CC-----------A----------G-----G---- 691
H101_AE.TH.90.CM240         ------------------T-TG--------T-T------------C--G-----A-----------------A--------------C--------------G--A--C-----A---...------------C-----------------G--------------- 1057
H102_AG.NG.x.IBNG           ------G-----------T-TG----------T------------C--G-----A-----G-----------T-----------------------------G--A--------A---...------------C-----------------G--------------- 1023
H1N.CM.95.YBF30             A--G--CTC----C----T--T---T----T-----A-----------G-----A-----G--C--G--G--AGT---------------A-----C--------GACA--------C-C-...-TG--A--AC-AC-C-----G--AC-------C--T--A------- 1056
H1O.BE.87.ANT70             A-TTT-CT-T---A-T--T--T--------TG-CA-A--A----------G---TT-A--G--C-----G--AGTA--------------A-T---G---------ACT--C---CCA-CG...-T------GT-GC-------G-----A---G------CA------- 1538
H1O.CM.91.MVP5180           TGT---CT-T---A-C--T--------G--TG-CA-A--------C----GG--TT-A---G----G--G--AGTA--------------A-------------------ACT------CCAGCA...ATG-----GT-AC-------G-----A---G------CA------- 1513
CPZ.CD.90.ANT               A--T--T--T-----------------T--TG---T-----AC-------G---------GG--C-------AGTA--------------A--T--G--------GT-A--C--CACT---...-----A--AG-ACAGG----A--AT-A---G------CA------- 904
CPZ.CM.01.SIVcpzCAM13       TCTA--TTTG---G-T-----A-----G--TG-T--A-----------G-G---------G-TC--G------GT---------A------------------C-GA-------T--T---...--C--C--C--A--T-----A---C-A---------C----T---- 1069
CPZ.CM.05.SIVcpzEK505       A--G--C-----C------------T----T-----A-----------G-----A-----G--C--G--G--AGTT--------------A-----G-----C--G-GA--------A---...ATG-----A--AC-------A--AG-A--G--G-----A------- 1040
CPZ.CM.05.SIVcpzMB66        ------T-----C-----------------------------------C--G-----A---------C------AGTA--------------------G-----C--GC-C--------T---...-----A--G--C--------A--A---------------------- 1035
CPZ.CM.98.CAM3              A--T---------G-T--T--T-----T--TG-TA-A--A--G--C--G-G---T-----GG-C--G--G--AGT---------------A---------------ACA--C--T------...-----T--AG-A--C-----A---C-A-----GG--C-A------- 882
CPZ.GA.88.GAB2              A--A--C------G-G-----A-----G--TG-TA-A--A-----C----G---T-----GG-AC---------GTA--------A-----A--G--G---------T-A--C---T-A---...-----A--AG-G--------T--A--------G-----A--C---- 870
CPZ.CM.98.CAM5              T--T--C--G---G-C--T--T-----T--TG-C--A--A--G--C--G-G---T-----GG-T-----G--AGT---------A-----A-----G--------GC-A--C--T------...--------A--A--C-----A--AC-------GG----A--G---- 1185
CPZ.CM.05.SIVcpzMT145       A--G--T-----CC-C--T--T--------T---A-A--A---------G-----------------TC-------AGTA--------T--G-----T--G-----C--GACA-----TACC---...-----------A--T-----A--A------G-------A--C---- 1031
CPZ.TZ.00.TAN1              A-----TT-T---C----T-GT-----C--TG-T--T---A-------------A-----------G--G--AGT---------------A-----G-----C----CA-----C-CT---...-----A--CCAGCA-G----GAT-C-A-----G--C-CA--GGCA- 1124
CPZ.CM.05.SIVcpzLB7         ------T-----------T-----------T-----A---------C--G-----------G--C----------GT---T-----------ACA------------GC-------TA-C---...ATG-----A--A-----------A---------------------- 1038
CPZ.US.85.CPZUS             G--T--C------G-C-----T-----C--TG-TA-A-----------G-G---A-----G-T------G---AG----------------A--T-----A------ACA-----TC-----...-----C-----A--------G--------------C---------- 1540
CPZ.GA.88.GAB1              -TTA--T--G---G-T---------T----TG----A-------------G-----------G--C----G--AGT---T-----------C--C--G-----C-CT-A------CACT---...--T--T--G--A--C-----A---C-A--G--G--------T---- 1553
MON.CM.99.L1                G-TT--CT-T--CA-C--TCAGC-A------G-CA-A--T-AT-----G-GG--AC-C-----TA-T--G--TGT---------------GCA---T-----CTT------C---CA--CG...C--CAA--GCAAC-CAAT-CAGG-C--G----C----C---GGCC- 1156
MON.NG.x.NG1                .......................................................................................................................................................................... 0
GSN.CM.99.CN166             T-TA--CT-T---A-G---CAA-----C--TG-TA-A--A-ATT------GG--AG-A--G--TA-T-----TGT---------AC-G--A-----G------CCTTT-A-----G--A-CA...----CA---CAAC-GGTG-CAGG-C-T-----C--T--C---GCA- 1169
GSN.CM.99.CN71              T-TA--TT-T---A-G---CAA-----C---G-T--T--A-ACT------G---AG-A--G--TA-T-----TGT---------AC-G--A-----G-------CTCT-A-----GCA--CC...--G-CA--CCAAC--GT--CAGG-C-T-----C--T--C---GCA- 1175
MUS_1.CM.01.1085            A-TA---T-T--CA-C--TCAAC----G--TG-CA-A--A-A------G-G---T--A-----TA-------TGTA--T-----AC----A------------CCTGT--A-C---CA--CC...CA-CA---GCAGC-CAAT-CAGGCC-C--GA-C----CA---TCA- 1140
MUS_1.CM.01.CM1239          A-TA--CT-T--CA-G--TCAGT-------TG-CA-A----AT-----G-GG--T--A--G--CA-------TGT---A-----C-----A-------------CTGC-A--C---CA--C-...CAGCA---GCAGC-CAAT-CAGGCC--C-GAGC---TCA---TCA- 1142
MUS_2.CM.01.CM1246          --T---CT-T---G----TCAG-----C---G-CA-A--A-AC--C--G-GG--A--A---C-AA-------TGTGG--------C-G--A-----CA-A--CGTTT-A-----CCAA-CA...A--CA---ACAAC--AAT-CAGG-C-A--GT-C---TCA--GGC-- 1223
MUS_2.CM.01.CM2500          T-TT--TT-T---G-G--TCAA-----T--TG-------A-AT--C----GT--AG----GC-AA-T--G--TGTGG--------C-G-----T--G-----CTTT-----C--CCAA-CA...---CA---CCAAC--AAT-CAGG-C-TC--T-T--CTCA---GCA- 1185
RCM.NG.x.NG411              --TA--CT-T---A----TCAG---T----TG-CA-A---A---------GT--T--A-----AG----G--AGTA--------T--G--A-----T-----CTTGAGA-----CAAT-CA......--A--AC-AC--G----T--AT-A--G-------CA-----C- 917
RCM.GA.x.GAB1               --TA--CT-T---A----TCAA--------TG-CA-A---A----C--G-GT--GC----G--AG----G--AGTG-------------A-----C------GCTAGA--------A-CA......--C--G--AC--G----G--AC-T----------CA------- 917
DEN.CD.x.CD1                G-TA--TT-T---A----TCAAC-C-----TG--A-T--A-----C----G----------G----------TGT---T-----A------CAT--T-----TTG-CT------G-CTA--TTACA-CA---GC--C------CAGCACAA--G--C-----A--TTCA- 1262

MAC.US.x.239                ------CT-T--CA-T--TCAG---T----TTGT-----A-AC--------G--T-----G--TA-C-G---T-TT--A--C-----G--------T------CTTGCA---C---CAA-CA......-CT--A...CA--A---A--AC-T--G--G--GTCA---TCA- 1755
Gag                          T  P  Y  D  I  N  Q  M  L  N  C  V  G  D  H  Q  A  A  M  Q  I  I  R  D  I  I  N  E  E  A  A  D  W  D  L  Q  H  P  Q  P  .  .  A  P  .  Q  Q  G  Q  L  R  E  P  S  G  S
H2A.GW.x.ALI                ---G--CT-T---A-T--TCAA-----T--TTGT-----C-AC--------------------AA-C-GG----TT--------A-----A------C------GTT-CA------CA-A-CA......--C--CT-AC--G----G--C--T-----------A--GTC-- 1801
H2A.DE.x.BEN                ---G--CT-T---A-T--TCAA-----T--TTGT-----C-AT--C--------T--------AA-C-G----TT-T-------A-----A------C-----TCGCA----C---A-A-CA......--C--CT-AC--G----A--C--C-----C-----A--GTC-- 1808
H2A.SN.x.ST                 ---G--CT-T---A-T--TCAA-----T--TTGT-----C-AC-----------T--------AA-C-GG-A-TT-T-------A-----A------T-----CGC-CAA--C---A-A-CA......--C--CT-AC--G-G--G---C-C--G--G---------TC-- 1253
H2B.GH.86.D205              ------TT-T---A----TCAG--------TTGT--A--A-A------G--------------TA-T-G---A-TA--------------A------C-----CCAGCAA--C--GTCA-CA......--C--A--GC-GG----A--AC-T--G--C-----A--GTCA- 1769
H2B.CI.x.EHO                ------TT-T---A-T--TCAG---T----TTGT--A--A-A------G--------------TA-T-GG---TA--T------------A------C-----CCA-CA-------TC--CA......--C--A--GC--G----G--AC-C--G--------A--GTCA- 1770
H2G.CI.92.ABT96             -CTA---T-T---A-C--TCAA--------TTGT--A--A-A---------------------TA-T-G--A-TA--T---------G--A------K-----CTTCAA-------CAA--A......-----GCCGC--G----A--AM-C--G-----G-CA---TCA- 1171
H2U.FR.96.12034             ---G--CT-T---A-T--TCAA---T----TGC-----A-A---------T--T---------TA-C-GG-A-TA--T------T-----A------T-----C-C-CA---C---AAC--A......--C--ACCAC--G----G--AC-T-----G-----A--TTC-- 1287
MAC.US.x.EMBL_3             ------CT-T--CA-T--TCAG---T----TTGT--A--A-AC--------G--T-----G--TA-C-G---T-TT--A--------G--------T------CTTGCA---C---CAA-CA......-CT--A...CA--A---A---C-T--G--G--GTCA---TCA- 1241
SMM.US.x.H9                 ---T--CT-Y---A-C--TCAA--------TTGT--A--A-A------G--------------TA-T-G--A-TT--A------A--G-----T--T-------TT-CAA--C--RCA--CA......--T--A--AC--G----A--AC-T-----G-----R---TCC- 1240
SMM.US.x.PGM53              ---T--CT-T--CA-C--TCAA--------TTGT--A--A-A------G--------------TA-T-G--A-TT--A------A---------T--T-------TTGCAA--C--GCA--CA......--T--A--AC--G----A--AC-C-----T-----A---TCA- 1684
SMM.US.x.SIVsmH635F_L3      ---T--CT-T--CA-C--TCAA--------TTGT--A--A-A------G--------------TA-T-G----TT--A------A---------C--T-------TT-CAA--C--GCA--CA......--T--AC-AC--G----G--AC-T-----G-----A---TCA- 1770
SMM.SL.92.SL92B             T--A--CT-T--CA-T--TCAG-----T--TTGT--A--T-A------G--------------TA-C-G--A-TA--T-----------A------C------CCT-CAA--C---AGAGGACAACAGCC-G-GCAAC-GG-T--AGG-C-A--G------TCA---TCA- 1198
STM.US.x.STM                T-----CT-T--CA-C--TCAAC--T----TTGT-------AG----------------------TA----------T--A--------------T--T------C-TGCA-------CAA-CA...C-G--T--GC-GC--G----G--AC-T-----------C--GTCA- 1417
MNE.US.x.MNE027             --T---CT-T--CA-T--TCAG---T----TTGT-----A-A------------T-----G--TA-C-G----A-TT--A--C-----G--------T------CCTGCA---C---CAA--A......-CT--A...CA--A---A---C-T--G--G--GTCA---TCA- 1235
DRL.x.x.FAO                 --T-AGCT-T--CA-C---CAG---T----TG-TA-T----AT--C----GT---C-C-----AG--------C-T---------------A--T--T-----------GAGA----TCAAGTA......--A---C-CC---A---GGT-C-T---A-C--CTCA--GTCA- 999
MND_2.x.x.5440              --T---GT-T--CC-C--TGG------C--TG-CA-T--A-AT-----G-G---TC-C-----AG-G--G--TGT-----------------------C-------CTTAGA--C--CCCTGTG......--A-----GC-C-A---AGT-C-A---A-C----CA----A-- 1002
MND_2.CM.98.CM16            -CT---CT-T---C-C--TGG------C--TG-TA-T--A-AC-----G-G---GC-T-----AG-G--G--TGT-----------------------C---A-CTTGAGA--C---CAAGTG......--T-----GC-C-A---TGTCT-----A-C----C---C---- 1445
MND_2.GA.x.M14              -CT----T-T--CC-C--TGGT---T----TG-TA-T----A------G-G---TC-C-----AG-G--G--TGT-----------------------C-------CTTAGA--C--CCAAGTA......--A---T-GC-T-A---GGT-C-----A-C----CA------- 1369
OLC.CI.97.97CI12            GCTG---G----GA-C--T---C--T-G--TGT---T----ACT------GG--T-----G-----G--G--AG-------------------A-A---G-AT--C--GCA-------CA--CA...AT------AT-GC-GGTG--A--A--TCC-CCG-------CT---- 1360
GRV.ET.x.GRI_677            TCT-T-CT-T---G----TCAG-----C--TGT-A-A--A-AC-----G-GG--AT-A-----TC-T--G--AGT---T-----A-----A------G------C--GACA--CAG-CCA-CA...--T--C--GT-AC--G----G---C-A-----C--G-CA--GTCA- 1605
VER.KE.x.TYO1_patent        ---A--CT-T--CA-T--TCAG-----T--TGTGC-A--A-AT-------GG--AT-A-----AG-G--------T--T-----A-----A--CC-G-------GT-ACA-C---C-A-CC......--A--CC-AC--G----A---C-C--G--C--TC-C--CTCA- 1665
VER.KE.x.9063               T-----CT-T--CA-C--TCAA-----T--TGTGT-A--A-AC-------GG--AC-A-----AG--------A-T---------------A--CC-G------GT-ACA--C---CCA-CA......--A---T-AC-------A---C-T-----C--T-----TTCA- 1668
VER.KE.x.AGM155             -------T-T--CA-C--TCAG-----T--TGTCT-A--A-AT-----G-GG---T-------AG-G-----A-TA--T------------A--CC-G------GT-ACC--C---CC--CG......--C--CT-GC--G-G--A---C-C--G--T--GG-----TCA- 1657
VER.DE.x.AGM3               ---A--CT-T--CA-C--TCAA-----T--TGTCC----A-AC-------GG--GC-A-----AG--------A-T---------------A--CC-G-------T--CT--C---CCA-CA......--A--AT-AC--G----A--AC-C-----C--T--A--CTC-- 1169
TAN.UG.x.TAN1               -TTAAGTT-T--CA-T--TCAG---T----TGTCA-A--A-AC-------GG--A-----G--TA-C--G--AGTA--------T-----A--CC-G-------T-ACA--C--CCCA-CA......--A--GC-AC--G----G---C-A--G--C-----A--TTCA- 1628
SAB.SN.x.SAB1C              T-T---TT-T---A-C--TCAA-----C--TG----A--A-A---C--G-GG--AC-A-----AG-G-----TGTT--------A-----G------T------CCTTAGA--C--GCC--CG...CA-CA---CCCA--T-AG--AGT-C-A--G--C---CAA--GTCA- 1758
MND_1.GA.x.MNDGB1           A-T-AGT--T-----G--T-T-----G--TG----A--A-AT-----G-G---A-----G-C---------TGTA--T------C---A-------------CT-ACA-----TCAA--A...CA-CCAG-ACAAC--GG---AGGAT-A--GACC--TTCA--CTC-- 1119
DEB.CM.99.CM40              T-----CT-T--CA-T--TGGTC-A--T--TG--A-T--A-AT-T---G-GG--T-----G--CA-T-----TGTA--T------------A------G-------TT-CAA--C---CAA--A...CAGCC-......C---AG--G---C-T--G------TCA---GCA- 1256
DEB.CM.99.CM5               T--T--CT-T---A-T--TGGGC-------TG-CA-A--A-AT-TA----GG--A----------TA-------TGTG--T------------A------G------CTT-CAA--C---CA---A...CAGCCA......C---A---T---C-T-----G---TCA---GC-- 1247
TAL.CM.01.8023              --TT--TT-T---A-C--TCAG-----T--TG----A--A-A------G-GG--------G--CA-T-----T-T---A------C-G-----C-----------CAGA----C--CCA-......AT---T--AT-AC--G----A---C----G--G----GCA-C---- 875
TAL.CM.00.266               --T---CT-T--CA-C--TCAA-----T--TG----A--A-AG-----G-GG--A-----G--CA----G--C-T---T------C-G--AT----G--------CAGA-------CCA-......AT---T--AC-AC--G----A---T-A--G-------GCA-G---- 1377
SUN.GA.98.L14               A-TT---G----CA-C--TGTG-----G--TG-CA-A---AG--------G---T-----G---A--------TCAT--AGTG-----AG-----------------C---A----C--TCA---G...CAGCCTG-ACAGC--GG---AGGCC-A--GACT--T-AT--T-CA- 1726
WRC.CI.97.97CI14            ATTAT-T--T---A----T---C--T------ACA-A--A-AT----------T--A--------AA-------C-TAG-A-----A----ATACA--G---------ACA-------CA-GCA...-----------A--------A--A------C-G----CA---TCA- 1696
WRC.CI.98.98CI04            TTTGG-T--T--CA-T------C--T----T-ATA-A--T-AT-----G--T-------------TA-------A-TAG-A----------AT-CA--C---------ACT--C---CAA-CA...--------A--A--------A--A-----GC-G----CA--GTCA- 1714
WRC.GM.05.Pbt_05GM_X02      ATTGA----T--CA-C--T--TC-AT-G--T-ATA-A--A-AT-----------------G--AG-T-----A-TAG-G-----A--ATAA-----------C---ACA------CCA-CA...-T---A---C-GC-C-----GGCT------A-T----CA---TCA- 658
LST.CD.88.524               -TTG---G----C-----TGT----T----TG-TA-A--A-AC-------GG--A-----G--AA-------TCAT--AGCG--A----GA--T------------CA------CCA---G...CA-CCAG-ACAAC--GG---AGG-T-A---ACT----AT-C--CA- 658
LST.CD.88.447               TTTG---G-T--C-----TGT----T----TG-CA-A--A-AC-------GG--A-----G--AA-T-----TCAT--TGTG--A----GG-----------C--GAAA-----GCAA--A...CA-CCTG-ACAAC--GG---AGG-C-T---ACC--T-AT-C--CA- 664
LST.KE.x.lho7               TTTG--TG-T--------TGTT---T----TG--A-T--A-AT-----G-GC-----------AA-------TCA---AGTG--------G-----G---------CA---C--GCAA--A...CAGCCAG--CAAC--GG---GGGTT-A---ACT--T-AT-C--CA- 1748
SYK.KE.x.SYK173             T-TT--TT-T---C-C--TGG------T--TG-T--A--A-AT-----G-G---------GG-AA-T-----C-T---------------GG--G-C------CTT-AA-----CCAA-CG...CA-CAA---GC-C---A--CAGGAT----G--T--CTCC-CCTCA- 1572
SYK.KE.x.KE51               T-TA--TT-T--CC-C--TGA------T--TG-CA-T--A-AT-------G---A------G--A-C-----TGTG--T------------GGT-T-------CCT-CAA--C---CAA-C-...CA-CA--AGCAGC-TGT--CAGG-C-A-----T--T--T-CCTCA- 1200
COL.CM.x.CGU1               A-TAG-TT-T---A-G---TTG-----C--T-TTT---ATACT-------G-TT-CCT--GG-C--G--G--TGAG--T---A-AA-G--A-A---G-AT--CCT-C-A-----G--C......CAGCA---ACAACAG-A---AGCAT-A--GC-G--T-CA-C-TCA- 1297
```

H1B.FR.83.HXB2	ACATAGCAGGAACTACTAGTACCCTTCAGGAACAAATAGGATGGATGACAAAT..............AATCCACCTATCCCAGTAGGAGAAATTTATAAAAGATGGATAATCCTGGGATTAAATAAAATAGTAAGAATGTATAGCCCTACCAGCATTCTGGACATAAGA	1647
Gag	D I A G T T S T L Q E Q I G W M T N . . . . . N P P I P V G E I Y K R W I I L G L N K I V R M Y S P T S I L D I R	
H1A1.UG.85.U455	------------------C---G---A-------------------GGC..............-----------------G----C--C---G---GTT-----T----------	1102
H1B.US.90.WEAU160	---------------------------------------C-------------..............--------C---------------C-------------------A-------------------------------------GT-------------------	1646
H1C.ET.86.ETH2220	------------A-------------------------C---------GGG..............--C-----G-T-----G----C--C---------------------GC------------------------------GT-------T---------A-	1021
H1D.CD.84.84ZR085	-T--T---GC..............-----------------------C--GT------T------C---	1164
H1F1.BE.93.VI850	-T-----T---------------------------CA---------GGGC..............--C-----G------G----C--C-------------C----A--------------------------T---GT------T---------A-	990
H1G.SE.93.SE6165	-T-----------------T----G-----------ACT--------CGGC..............--C---------------G--------------------------------------G-----------------------GT-------T-A---------	1050
H1H.CF.90.056	-T-----------------------G------------C---------GGC..............------G--------G----C--C---G----------------T---GT-----------------A-	994
H1J.SE.93.SE7887	-T-----------------------C-------------------GGC..............-----------------------G-------------------T------C--------------------------------GT---T---T----T------	970
H1K.CM.96.MP535	---------------C--C------------------C----------GC..............--C-----G------G--G----C----------------------T-----C-------------------------GT------T----------	846
H101_AE.TH.90.CM240	---------------------------A--------------------C..............--------C------G----C--C-------G---GTT-----T----------	1212
H102_AG.NG.x.IBNG	---------------------------A-------------------GC..............--------------G------C------------G----------------------G-----------GT-------T----------	1178
H1N.CM.95.YBF30	-T-----------A-----C-----GGCA-----GG-G-CT--------TGC-..............------T--G-T---------T------G-----------G------G-----C-G---T--G----------T---GT---------A--G--C-A-	1211
H1O.BE.87.ANT70	----T--T--G--A-----C----AG--A--G-----TCAC----CT--C-GGCCC............--C-A---------------C--C----G--A-------G-GT-A---C----C-----G----A-----C-----AGTG-----CT-A--T--T-AG	1696
H1O.CM.91.MVP5180	----T--T-----A-----C--A-AG--A--G-----TAT-----CT--T-GAGGG............GC-AACT---------------C--C----G--A-------G-G--A---C----C-----G----A-----C--T--AGTG-----CT-A--T--T--G	1671
CPZ.CD.90.ANT	-T--------G--A--A--C--AG-G-----G--G--GCA-------T---CACCT.....CAACAG---GG-GGAG---------G--C--C-----G--------C---A-------------GG-G--C--G-N------T--AGT---------A--G---A-	1068
CPZ.CM.01.SIVcpzCAM13	-T--------G-----C--C--AT-G-----G-----T------GCA---GCA..............--C-----C--------------C--C----G-------G----T--A--GC----C-G-G------A-------T-T--AGTA-----C--T--T----AG	1224
CPZ.CM.05.SIVcpzEK505	-T-----------A-----------GGC---G---G---C--------TGCC..............------T--G-T---------T------G-----------G--T--------------G--T---------------T---G-----------A-----C-A-	1195
CPZ.CM.05.SIVcpzMB66	-------------C--C--C---A-C--A--G--GG----G--------T-GC..............--C--T----T---------T--C----------------------TT---GC-------G-------A-------T----GT------A--A-------A-	1190
CPZ.CM.98.CAM3	-------T--G--------C--T--AGTA-----G--T-CT---------GCA..............--------CG-T-------------A--C-GG-----G-GG-T--A--GC-T--C---G-------G-----T---AGTT--T--CT-A-----C-AG	1037
CPZ.GA.88.GAB2	----T-----T-----------T-G--A--G---G---------------TCC..............--------AG----T-------G----C-G------G-GG-AT-A--GC-T-----GG-G-----------T-----GTT-----CT---------A-	1025
CPZ.CM.98.CAM5	-------T-----------C--T--A--A------G-T-CT---------GCA..............--------C----------G----C----G------G-GG-T--A--GC-C--C---G----G-A-----CT----CGT---T-----A--T--C-A-	1340
CPZ.CM.05.SIVcpzMT145	----------C--------CT----AG----G---G----G---------GCA..............--C--T----T----------T-------G---------G-------G------------T-A-----C--T---GTT------T-------T-A-	1186
CPZ.TZ.00.TAN1	-T--T-----G--C--------G--A--A------G--CTG---------CCCCACAGGCA......C-AGG-GGAG-G----------C--C--------G-------TT-A--------------T----C-------C------GTT------T---------A-	1288
CPZ.CM.05.SIVcpzLB7	-------------C-------TG-A--A-------------------T-GC..............--C--T----A---------------C---------------TA---GC-G----GG---C-------C-----GTT-----C---------AG	1193
CPZ.US.85.CPZUS	----T--------C--C--C---T-G--A------G-G--C--------GCA..............--C-----C--T---------C--A---G------G-GG---A--GC-T------G------A------CT-T--CGTTG-T--CT-A--T--C-AG	1695
CPZ.GA.88.GAB1	-T--------G--C--C--C--A--G-----------T--G----CA---GCA..............------T--C----------G--TG-----G------G----TT-A--G-----C---G-G--T---------T----AGTA--T--C--T-----C---	1708
MON.CM.99.L1	-T---------GTCT-GTCC--A-CC--AC-G--G----AG-----A--C-GGGCT............--CAAC--CG-T-AG--CTC---C--C--C-G--A----G-C---T----CC-GC-GCG-TGC--C-A------C-A----GT--A-----T-----T--C-A-	1314
MON.NG.x.NG1	....	0
GSN.CM.99.CN166	-----------GT----TCA--T-CCA-T--G--G----AG-----C----GGCAG............--C-A----G--AAT--GTC---T--A----G--AG-----C--TT----CC-GC-GCG-TGT----A--------A----GT--AT---T-------C-A-	1327
GSN.CM.99.CN71	-----------GTA--ATCA--T-C-A-T--G--G----AG-----T----GGCAG............--CAAT--AG-TAAT--G-C---T--A----G--A------C---T----C--GC-GCG-TGT----A--------A----GT--AT---T---G---C-A-	1333
MUS_1.CM.01.1085	-----------GTAT-ATC-----CA--A-----G----A------C--T-GGGCA............--CAAC-----AAAT---C------C--C--G-AG---G-G---A----GC-GC-G-G-TGT--C-A------C-AT--AGT--A---C--A--T--C-AG	1298
MUS_1.CM.01.CM1239	----------GGTA--ATCC----CCA-T----------A------A--T-GGCAA............--CGAT-----AAAT--G-C---C--C-----G-A------------A----G---C-G-GGTGT--C-A--------A----GTT--T--AT----T--C-AG	1300
MUS_2.CM.01.CM1246	-TG-G--G---GTGT-ATCA--G-CAGCA----------A---------T-GACAG............C-AGAC--AG-AAAT-----GA-T--A----G--A------C---T----CC--C-G-G-TGT--C-A------C-AT--AGTA-A---CT----T--T-A-	1381
MUS_2.CM.01.CM2500	-T--------GGTAT-CTCC--A-CAGCA---------AG-------C-GGCAA............C-AGAC--AG-GAAT--------C--A----GG-A-----T-------C--GC-A-G-TGT--C-A--------A----GTA-A---CT-A-----C-A-	1343
RCM.NG.x.NG411	-T--T--------A--A--C-ATA--GCT----------CT----CC--T-GAGCC............--CAAC--A--TG----G---A-C--------GG-AT-----TG-G--T--G---C----T--G---C-A---TGT--G-A--------A---AGTA-A---CT-A-----C-AG	1075
RCM.GA.x.GAB1	----------G--A-----CT-AA-AGCA-----G----CT----CC--C-GAGCA............--CAAC--C--TAAT--G--CA-TC-G--C-G--AT--------AG-A--G---C-A---TGG----A------C-AT--AGTG-A---C--A--T----AG	1075
DEN.CD.x.CD1	-T--------G--C--CTCCT-TA-AG-A--T--G---AC---------T-GACAA............C-AGGC--A--AAAT--------T--A-----G--G---G-T--TA--------C-A--GG-G--C-AG-----CCAA--AGT-TC----T---------AG	1420
MAC.US.x.239	-T--T--------A-----T-AG-AG-T--------CCAG------TAC-GACAA............C-GAAC--C--A--------CA-C-----C-GG--------CCAA-----G--GC-A---TGT--C----------A---A--A-A------A--TG---A-	1913
Gag	D I A G T T S S V D E Q I Q W M Y R Q . . . . . Q N P I P V G N I Y R R W I Q L G L Q K C V R M Y N P T N I L D V K	
H2A.GW.x.ALI	-------------A--A--C--AG-AG-A----G--CCAG------TTC-GGCCA............CGGAAT---G-G--------GA-C--C----G---------CCAGA-A--GC--C-G--GTGT--C--G-----C-A---A---A---C--A-------A-	1959
H2A.DE.x.BEN	-------------A--A--C--AG-AG-T----G--CCAG------TAT-GGCCA............C-AAAT--CG-A--G-----GA-C--C--C-G---------CCAAA-A--GC-GC-A--GTGT--C----A---C-A---A--T-A---CT-A-------A-	1966
H2A.SN.x.ST	----------G--A--A--C--AG-AG-A--G--G--CCAG------TTT-GGCCA............C-AAAT---G-A---------AGC--C----G---------CCAGA-A--GC--C-G--GTGC--C--G-----C-A---A---A---C--A-------A-	1411
H2B.GH.86.D205	-T------------C--C--C--AG-AG-------G---CAG------TAC-GGGCC............C-AAAT---G-------G---A-C-------G---------TCAAT-A-----GC-G---TGT--CC-------C-AT-------A---AT-A-------AG	1927
H2B.CI.x.EHO	-------G-----C--C--C--AG-AG-A-----G---CAG------TAC-GACCC............C-AAAT---G-------G---A-C--C----G---G-----TCAGT-A--GC--C-G---TGT--CC-G-----C-AT-----T-AT--A----------AG	1928
H2G.CI.92.ABT96	-T--------G--C--------AA-AG-A--------TCAG----CACAC-GGCAG............C-AAAC-----A---------A-T--Y----G---------CCART-A---C-TC-A---TGT--C--G-----Y-AT--A--A-A---C--A--TG---A-	1329
H2U.FR.96.12034	-T--T--------A-----C--AG-GG-A--G--G---CAG------TAC-GGCCG............C-AAAT--A--A--G--G--TA-C--C---CGG--------CCAAT----CC--C-A---TGT--T---------C-AT--A--T-AT--CT-A--T----AG	1445
MAC.US.x.EMBL_3	-T--T--------A------T-AG-AG-T--------CCAG------TAC-GACAA............C-GAAC--C--A--------CA-C-----C-GG--------CCAA---A-G--GC-A---TGT--C----------A---A-TA-A------A--TG---A-	1399
SMM.US.x.H9	----T---------------C--AS-AG-C--------TCA-------TAY-GGCAG............C-AAAC-----A--------CA-C--C--C-GR--G-----CCAAT-A--GC-GC-G---TGC-------------AT--A--A-A----T----TG-G-A-	1398
SMM.US.x.PGM53	-T--T------------------AG-GG-A--G-----TCAG------TAT-GGCAG............C-GAAC--C--A-----G---A-C-----C-GG--------CCAA--T--G---C-A--GTGT--C---------C-AT--A--A-A----------G---A-	1842
SMM.US.x.SIVsmH635F_L3	----T------------------AG-AG----------CCA-------TAC-GGCAA............C-AAAC--C--A--------CA-C-------GG--G-----CCAAT-A--GC-GC-G---TGT-------------A---A--A-A----T-A--TG-G-A-	1928
SMM.SL.92.SL92B	----------G--A-----C----CAAGT-----G----A-------TAT-GGGCA............C-AAAC--AG-------G-----C--C----G---G-----CCAG------C-CC-A---TGC--C--G-------A---A--A-A---C-----TG-T-A-	1356
STM.US.x.STM	-T--------G---------CT-A-CAG-----------TCA-------TACCGGCAG............C-AAAT-----A--G------A-C--C--C-GG--------TCAA--A--GC--C-A---TGT--C---------C-AT--AGTT-A----T-A---G---AG	1575
MNE.US.x.MNE027	----T--------A---------AG-AG-T--------CCAG------TAC-GACAA............C-GAAC--C--A--------CA-C--------GG--------CCAA-----G--GC-A---TGT--C-A--------A---A--A-A------A--T------	1393
DRL.x.x.FAO	----T--------C--C--CT-TA-AG-A--G--------AG----CC--C-GAGCC............C-GGATT-C---AAT--G----C---------G-CA----G-GG--T----G---C-A-G-TGT----C------C-AT--AGT--A---CT----TG-G-AG	1157
MND_2.x.x.5440	----T--------A--A--CT--A-AG-A--G--G----A-----CA--C-GGCAG............C-AGAC-AGG-AAAC-------GT--A--C---CA-------G-A--A--GC-GC-A---TGT-----C-------A----GTA-AT---T-A-----C-A-	1160
MND_2.CM.98.CM16	-T--T--T-----A--A--CT--A-AG-A--G--G----A-----CT----GGGAG............C-AGATG-AG-AAAT--G----G---A------CA-------G-TT-A--GC-TC-G--GTGT-----C-------AT--AGTA-A---C----------AG	1603
MND_2.GA.x.M14	----T--------A--A--C-G-A-AG-A-----G----A-----CA--T-GACAG............C-GGA--AGG-AAAT-----G-C---C-----GCAG------G----A--GC-GC-A---TGT--G--C-----C-A---AGTG-A---------------A-	1527
OLC.CI.97.97CI12	-T----T------C--------ATT-G---C-G---C-G-CT---G-A-ATGCAGGA.......................--AGAT-------C-C-C-----G------G-T---A-A--G--GC-G--GT-G--G-AG------TCG--AGTT-CAG-G--A-----C---	1509
GRV.ET.x.GRI_677	-T---------------C--CT-AA-----------------AG----CCTTC---GCC..................AGA--AGAC-----G-C-CAA--C-G--A----G-T--TT----C---C-A--GG----GCAG-----C-AT--CCAA-AGG-C--A-----TC--	1763
VER.KE.x.TYO1_patent	-T--------G--C--C--CT-AG-A--A-----GT---A------CTAT-C-GCT.............--C--C-GGG-AGAT-----T-CC--C--CCGG--------T--T--A---C-TC-A--GTGT--C-A------C-A---AGTATCAG-C--A-----T--G	1823
VER.KE.x.9063	-------------C---------AG-G--A-----T---A------TTAC-CAGCC.............--C--CAGAG-AGAT--T--G-CT--C----G---------T-----A--GC-GC-G--GTGT----A------C-A---AGT-TCAG-AT-A-----C---	1826
VER.KE.x.AGM155	-T--------G--C---------AG-G--A--G--GC---AG-----CTAT-C-GCT.............--C---AGGG-AGAT-----G-CC--C---CG---------C-----A--G---C-A---TGT----A------C-AT--AGTGTCTG--T-A--T--C---	1815
VER.DE.x.AGM3	-------------C--C--C--AG-G--A-----GC-G-A------ATAC-CAGCC.............-----CAGAG-AGAT--G--T-CC--C----G---G-----T--------G--GC----G--C-A---TGT----A------C-A---AGTGTCTG-CT-A----------	1327
TAN.UG.x.TAN1	-------G-----A--C--CT-TG--GCT--G--------AG----CTTTT---GCT............------AGAG-AGAT------CGT--C--C-G-G-----G-T---------C-CC-----C-CC-A---TGT----A------C-AT----T-TCAG-C--------C---	1786
SAB.SN.x.SAB1C	-------G--C--------C--AA-A--A--G--------A-----CC--C-GAGCA............C-GAATG--G--AAT------A-T---------G-------C--------TC-CC----TC-CC-G---TGT--C-A------C-A---AGTA-A----T-A-----T-A-	1916
MND_1.GA.x.MNDGB1	-T---------------TC---AG-GG-A------T-G-C---------AT-TGCAA............C-AAATG-A---AAT-----AC---C-----G--T-----T--A-----CA-G-----CA-G----G-T-G----A--GTC--T-T--A-TA--T--AACA--TG-----	1277
DEB.CM.99.CM40	-------------A-AC--C--AG-GG-A-----G----C---------C-GACCAGCAGGTCAAGGAC-AGG---C--AGAT------C----A----G-------G-C--A--A--TC-GC-A-G-TGT--G-A------C-AT--A----A---CT----TG-T-A-	1426
DEB.CM.99.CM5	-------------A-AC--C--AG-GG-A--G--G----CT--------C-GGCCAGCAGGACAAGGAGCAGGC--A--AGAT------C-G--A--C-GG------G-C-----A--C---C-G-G-TGT----A------C-AT--A--T-A----T----TG-T-AG	1417
TAL.CM.01.8023	-------------A--C--C----CAGCA------G-G-AG----CC--C-GGCAG............C-GAAT--AG-AGAT--G--GTCC--A--C--G--G--G-C---A-A--C---C-GCGCTGT----A------C-AT--AGTG-A---C-----TG-G-A-	1033
TAL.CM.00.266	----------G--A--A--C--T-CCGCT----GG-G-A-----CC--C-GAGCA............C-GAAC--CG-AGAT--G----C---A--C---------G-C---A----C---C-GCGCTGT----A------C-AT--AGTT-A---A-----TG-G-A-	1535
SUN.GA.98.L14	----------GGTA--ATCA--AG-AG-A--G--G---CAG----CTATGGCCTT......AATGGAGCAA--AGAG-TG-----CA---C--C------G-C---G-C--AGA---CA-GG-A-G-G----T---C-CC--CAGA-----TCAG-AG-A--A--T---	1890
WRC.CI.97.97CI14	-T---------GTA--AGCA-GTG-GG----TG--C---CC---T--G-TC-AAAT...............-----AG-AGAT----GTC---A-----------G-CC-TAGA---C--C-G---G----TCAG--C---GTT--AGTA---------------A-	1851
WRC.CI.98.98CI04	-T--------GGTA--AGC-T-AG-AG----TG--T-G-CC---T--G-TC-AAAT...............-----A--AGAT-------C-G-G--C-----G---G-GC-TAG---GC--C-A--GG-----CA---T---GTG--AGTT------T-A--T----AG	1869
WRC.GM.05.Pbt_05GM_X02	-T---------GTA--ATC---TA-GG-----G--T---C----C-AG-TC-GCAA...............-----A--AGATA------C---A----------------TAGA---C--C-G--GG----GCA---T---GTA--A--ATCT--CT-A--T--C-A-	813
LST.CD.88.524	-TG-G-----G-TC--ATCA--AG-AG-A-----GCA--CC----CA---GCAGAC..............A----C--AGAT----GA----C-----GA----G------A--CTA-GG-G---G-G--------AC--CAG--AGT---TG-AA---------A-	813
LST.CD.88.447	-TG-C------GTC--CTCA--AG-AG----G--GC---CT----CC--TGCAGAC..............A-C---G-AGAT------A-G---------GAG---G----T-AA-CTA-GG-A---G----T-----CC--CAA--AGTT--TG-AA-A-----C-AG	819
LST.KE.x.lho7	-TG-G-----GGTC--CTCA--AG-AG-A------C---CC----CT---GCAGAC..............A-T---G-AGAT------A-----------GA----G-C--T-A--CTA-GG-A---G----C-----CC--CAA--AGTA--TG-CA-------C-A-	1903
SYK.KE.x.SYK173	-------------A--A--CT-AA--GCT--G-------A------T--T-GGCAG............--CAAC--AG-A-A-----------A----G---------C--T--A--GC-GC-G--GTGT--GCAGG-C--C-A----GTTTCT--CT-A-------A-	1730
SYK.KE.x.KE51	-------------A--C------A-C--A---------AG-----A--T-GACAG............--CAAC--A--T-AG--T---C-G--C--C-GGCAG--------T--A------C-A--GTGT--GCA-G-C--C-AT---GT--AT---T-------C--G	1358
COL.CM.x.CGU1	-T--CA----C-AC--A--CT-AG-AGCA------G-G-CT---GGAGAG.........................--C--A......-CCA-C-------GGG-C------G-A-A-A-TC-GG-A--GG--A-CCAG--TGCA--A--AT--TCTG-A--A--T--T---	1437

```
H1B.FR.83.HXB2              CAAGGACCAAAGGAACCCTTTAGAGACTATGTAGACCGGTTCTATAAAACTCTAAGAGCCGAGCAAGCTTCACAGGAGGTAAAAAATTGGATGACAGAAACCTTGTTGGTCCAAAATGCGAACCCAGATTGTAAGACTATTTTAAAAGCATTGGGACCAGCGGCTACACT 1817
Gag                          Q  G  P  K  E  P  F  R  D  Y  V  D  R  F  Y  K  T  L  R  A  E  Q  A  S  Q  E  V  K  N  W  M  T  E  T  L  L  V  Q  N  A  N  P  D  C  K  T  I  L  K  A  L  G  P  A  A  T  L
H1A1.UG.85.U455             -----G-----A--------C--G--T--------TA-A---T---------C-----T--------A-------T--------C------------------C-----------------T-----C------T-C-------G------A--G----G-------T- 1272
H1B.US.90.WEAU160           -----------A--------------T----------------------------------------A-------------------------------------------------T--G-----------------------------------------GA--------- 1816
H1C.ET.86.ETH2220           -----------------T----------------------------T-----CT------T--A------A---A--T-------------------C--------------------------------------A--C------G------A--G----G----T--T- 1191
H1D.CD.84.84ZR085           ---------------------------------------C-------------------------A--A-------------------------------------------A--------------------C-----------------G-------T- 1334
H1F1.BE.93.VI850            -----G-----A----------------------A-A----T----GTC--------T-----------------C-----GGG-----------C--A------------------------------------------C-----------------A---G--------- 1160
H1G.SE.93.SE6165            -----G-----A-----------------------TA------T----TG-T-G-----T------------------T------GGC-----------C-----------------------------------------C--C----G------A----A--GA---T---- 1220
H1H.CF.90.056               -----G-----A-----------------------A----T-T-------T-------T--------CA-------T--G--G-----------------------------------A--T--------C--------A----G------A----A--G----T--A- 1164
H1J.SE.93.SE7887            -----------A-----T-----------------A-------T----G----------T---------A-------T-------------------T-----C-------------A--T--------C-----C----------------A---T---GA-------- 1140
H1K.CM.96.MP535             -----G-----A-----------------------TA------T------C--------T--A-----CA-------A-----G-----------------C---C--------------C--A-----------------C-----------G---------G----T--T- 1016
H101_AE.TH.90.CM240         -----G-----A-------C---------------TA-----------------C----G--A-----A-----------------C-----------------C-A--------C-----T-----C------T-C---------------A---A---GA------T- 1382
H102_AG.NG.x.IBNG           -----G-----A---------------T-------TA------T-------T-G-----T--A-----A-G-----------------C-----------------C-----------------T-----C------T-C-------G------A---A---G-------T- 1348
H1N.CM.95.YBF30             -----------A--------C--------------A-------C----------------A-----G--AA-------A-----G------------------AC-C--A--A--------A-------------ACAGC-CC----------A--G----GA-----CT- 1381
H1O.BE.87.ANT70             --G--------------A-----------------T--------C-----AT-------T---------A-T--A--A------------------------TC--C-T--T--------C--T--------C--ACAG-----G---T----A--G----GA-----CT- 1866
H1O.CM.91.MVP5180           --G---------A-----A--C-----T-------T-----T--C-----AT-------T---------A-T--A--A-----G--------------------C-T--T--G---T-A-------------ACAA---C-G---------A-------AA-----TT- 1841
CPZ.CD.90.ANT               -----------A--G-----C-----T-----G--TA-A------------AA-T-----A--A--G---------CCT--G---GCC---------------------AA----------C--T--------C--ACAC--CC-G--G--T------A--GA--CT-CT- 1238
CPZ.CM.01.SIVcpzCAM13       -----------A--G-------A---T---------A-----T---------CA-G-----T-----------T-----A--G----C-------------C---C-T-----G--------T-----------C---CAA---C----G--CC----T--G-G---CT-C-- 1394
CPZ.CM.05.SIVcpzEK505       --G---------A--GA-T---------T--C-----TA-----T--C-----------------A-----G--CA-T-----A-----G-----------------G--AC-C--A--A--------A--T--T--------A-A---CC-G-G-------A--------GA--CT-C-- 1365
CPZ.CM.05.SIVcpzMB66        -----G-----A--GT-A--C-----T-----G--A-A-----G---A-A--T---------T----G--A--A------A----A--A----------------------C--T--A-----G--------T--T-----T--T-------------T-C--A----G------A--G----GA-----CT- 1360
CPZ.CM.98.CAM3              --G-----T--A--G--T--C--G--T--------TA-A-----C-----C-----G--A--A-----AA----A--A-------------------T--C---C-C--A--A--G-----T--T-----T--T-----C-----A-AC--C----G---CC---------GA-----CT- 1207
CPZ.GA.88.GAB2              -----G--G-----G--T-----------------TA-A-----C---GT-T-------A--A--G----------A--T--G-----C-------------G---C-CC----A--G-----C------------------CAA--AC-T-GG---C-A-------GA-----C-- 1195
CPZ.CM.98.CAM5              --G-----T--A--G--A--C--G--T--------TA-----T----------T-------A--A--G--AA--------A--------------------T--C---C-T--A--T---------T--T-----T--T-----C--C--A-GC--C--G-G---C--A-------GA-----CT- 1510
CPZ.CM.05.SIVcpzMT145       -----------A--G--T--C--G-------------A-------C--G--CT----G--T--A-----A--------T--T-------------------C---C-----A-A--G-----T--------T--------C-----A-AC--C----G------A-------GA--C--T-- 1356
CPZ.TZ.00.TAN1              --G---------A-----A--C-----T---------A-A-----C-----AA-C-----A--A-----A--T--ACCA-------C----------------T--AC---A---------A-----------A-------------CA---C---------C-----G-A--GA--A---T- 1458
CPZ.CM.05.SIVcpzLB7         -----G-----A-----A--C-----------G--A----G---A-A---T--C-----AT-------A--A-----CA-------A--------------------C------C-T--A--G-----T-------------C--A-G---CC-G-G--------T----GA--CT-CT- 1363
CPZ.US.85.CPZUS             --G-----T--A--G-------------------A-----------C-----AT-------A--A----------T-CA--A-------CC-----------G---C-T----A-----C-T--T--C--T--T------------A----------G---C--------T-GA--C--C-- 1865
CPZ.GA.88.GAB1              -----G-------G-----------T-----G--A-A-----G---A-A-----C------T-----G--T---------------------A--G--------------T--C--TC-C--A--G--G-----A---------------------CAA---C-----------------T-G---C--C-- 1878
MON.CM.99.L1                --G-----T--A-----A--C-A---T--------A-A-----A-A-----C---TGC--C-----A--A---A-AGAT----CA--C---------------GC-G--GC--C--------------C---------C--ACTC--CC----G--CA--...----GT----GCT- 1481
MON.NG.x.NG1                ..............................................................................................................................------G--C--C---CTC--CC-C------A--...----GA--C-GCT-   47
GSN.CM.99.CN166             -----C--T--A--G------AG--A---------A-A-----A-A---T---GTGCT----G-----A-----AGAT----CC--G---------------C--T-GC--C----T--------T--------T---------A--C---CTG--C--G-----CA--...----GA--C-ACT- 1494
GSN.CM.99.CN71              --G--C--T-------A----AG--A---------A-A-----A-A---T----TGCT----G--T--A--G--AGAT----CA-----G------------C--TTGC--C----T--------T--T-----T--T-----A--C---CTG--C--G--G--CA--...----GA--C-GCT- 1500
MUS_1.CM.01.1085            --G---------A-----A----AG--T-----G--A-A-----G---A-A---T----TG------G--A-----GT-AGAC--A-C---C---------------CTC----C----A--A--G-----T---------G-----ACTC---C-G--GAGCA--...----GA--C--T-- 1465
MUS_1.CM.01.CM1239          --G---------A--G--A----AG--T-----G--A-----G---A--------C---TGCT----G--G------T-AGAT--A-CA--T--G-----------CTCC--T-----A--G--G-----T---------G------CTC--CC-G---AGCA--...----GA--C--CT- 1467
MUS_2.CM.01.CM1246          -----C--T--A-----A--C-A---T-----G--T-------G--T--------T----TGCT-------T--A--G---GAC--A-CA--C----------------C-GT--C--C-A--G---------C-----C-----C--G--C--ACTCG-C--G-----CA--...----G---C---T- 1548
MUS_2.CM.01.CM2500          --G--C--C-GA-----G--C-A---T-----G--A-------G--A------T----TG----------T--A-----CGAC----CT--C----------------C--T-T---------T--G-----C--T-----C--T--T--G--C--ACTCG--C-G-----CA--...----GA--A---T- 1510
RCM.NG.x.NG411              --G---------A--G--T----AG---------------TA-------C--GG-CT-G--G--A--A--G--AGAC-CA-CA--G--G-------------C--T-AC--C--A-T--------T--T--------T--T-----C----------AG-AC-T--G-GTC-----ATGAACC-----C-- 1245
RCM.GA.x.GAB1               -----------A--GT-A--C-AG--T-----G--TA-A--T------G-CT-G-----A--A--G--AGAC-C--CA------------------C--T-AC--C--A-A--------T--------T--------C-----A-TGG-AC-C--G-GTC-----ATGAACC--T-TT- 1245
DEN.CD.x.CD1                -----G--C--A---T------AG--T--------TA-A-----C---G-----------A-----G--AGAT--ATCA--G--GCT---------CACTT-AC-A-----A--------A--T--------A--T--------C---CAG---C-CG-------A-----TAACC-C---T- 1590

MAC.US.x.239                -----G-----A--G--A---CAGAG----------A-------C----G-T--------A--A--GA-AGATGCA-CA-----G-----------TC----AC--C--A-T--------T------------T-----------C---CTAG-GC-G--G-GGC----TGTGAATC-C--C-- 2083
Gag                          Q  G  P  K  E  P  F  Q  S  Y  V  D  R  F  Y  K  S  L  R  A  E  Q  T  D  A  A  V  K  N  W  M  T  Q  T  L  L  I  Q  N  A  N  P  D  C  K  L  V  L  K  G  L  G  V  N  P  T  L
H2A.GW.x.ALI                -------------G--A--CCA-AG---------TA-A-----C----GCT----G--A--A---A-AGAT-CA-CA-----G-----------TC----AC--C----A--G-----C-----C--------C--C--ATTAG-GC-G----G---A--GATGAATC----CT- 2129
H2A.DE.x.BEN                --G---------A-----G--CCA-AG----------A-------C----GCT-G--G--A--A---A-AGAC-CA-CA-----------------CC----GC--C-AA-A--G-----C-----C--------C--C---TTAG-AC------G-C----GATGAATC-C--C-- 2136
H2A.SN.x.ST                 --G------------G--A--CCAGAGT---------TA-A-----C--G-GCT-G--G--A--A---A-AGAT-CA-CA-----------------CC----AC--C-A--G--G-----C-----C--------C------TTAG-AC------G-C-A--GAT-AATC----CT- 1581
H2B.GH.86.D205              --G------------G-----CCA-AG----------TA-A-----C----GCT--C-G--A--A---A-AGAC-CA-CA--G-----------------C----AC--C--A-T--G-----T-----T-----------C---TTAG-GC-T--G-GC------ATGAATC-C--CT- 2097
H2B.CI.x.EHO                -----G-----A--------CCA-AG----------TA-A-----C----GCT--C-G--A------A-AGAC-CA-CA-----------------C----AC--C--A-T--G-----T-----T-----------C--ATTAG-GC-T--G-GC------ATGAACC-C---T- 2098
H2G.CI.92.ABT96             -----G-----A---------CAGTCT---------TA-A-----C--G-G---R-----A--A---A--GAT-CA-CA-------------------C-G--AC-AC--A-T---------C-----C----------------ATTAG-AC----G-GG------ATGAATC----TT- 1499
H2U.FR.96.12034             -----G---------G--A---CA-AGT--------TA-A-----C----G-T-G-----A--A---A-AGAT-CA-CA--G----------------TC-G--GC--C--A-T--G-----T-----T--------C--C--ACTCG-G--------GT-----TATGAATC----CT- 1615
MAC.US.x.EMBL_3             ---A-G-----A--G--A---CAGAG-----------A-------C----GCT-------A--A---A-AGATGCA-CA-----G-----------TC----AC--C--A-T--------T-----T------------C---CTAG-GC-G--G-GGC----TGTGAATC-C--C-- 1569
SMM.US.x.H9                 ------------A--GT-A---CA-AG---------TA-A-----C--G-G---------A--A---A-AGAC-C--CAS-G-------------TC----AC--C--A-T--------C-----C-----G--C-----ATTGG-GC-C--G-GTC----YAT-AATC-C--TT- 1568
SMM.US.x.PGM53              -----G--G--A--G--G--CCAGAG---C-----TA-------C----GCT-------A--A---A-AGAT-CA-CA--G--G-----------TC----AC--C--A-T--------T-----T---------------ATTGG-GC-C--G-GTC----CATGAATC-C--TT- 2012
SMM.US.x.SIVsmH635F_L3      ------------A--G--A---CA-AG---C-----TA-A-----C----G----------A------A-AGAC-C--CA-----G-----------CC----AC--C--A-T--------T-----T---------------ATTGG-GC-C--G-GTC----TATGAATC-C--TT- 2098
SMM.SL.92.SL92B             -----T--------G-------CA-AGT--------TA-------C----G-T-G-----A-----GA-AGAT-C--CA--G--G-----------C----AC--C--A-T--G--C--C--C--C-----T-----C--ACTAG-A-------GG-----CATGAATC-C--C-- 1526
STM.US.x.STM                -----G-----A---A-G--CCA-AGT----------A-------C--G-G-T-G-----A--A-----AGAT-CATCA--G--------------TCGG--AC--C-AA-T--------T--T-----T--T-----C--C--ATTGG-A--G----GTC----CATGAATC-C--C-- 1745
MNE.US.x.MNE027             -----G-----A--G--A---CAGAG-----------A-------C----GCT-------A--A---A-AGAT-CA-CA-----G-----------TC----AC--C--A-T--------T-----T------------C---CTGG-GC-G--G-GGC----TATGAATC-C--C-- 1563
DRL.x.x.FAO                 -----T--T--A-----A--C-AG-----------TA-A--T------G-C--C--G--A--A--CA-AGATGCA-CA-----------------C----AC--C--A-A--------A--T-----A--T-----C-----AGTG---C-C----GCC-A--GATGAACC-C-GT-- 1327
MND_2.x.x.5440              --G---------A--G--T----AG--------G--TA-A-----C----G----GC-G--G----GCA-AGAT-CAC----C-----C---------GC-G--A--AC-CA----G-----C-----C--------------AG---CCC-C----GG--A--CATGAATC-C---T- 1330
MND_2.CM.98.CM16            ---------------G--G--C-AG--------G---A-----T--C---G----GC-G--G----GGA--GAT-CAC-A--C----CC---------C----T---C-CA----G-----T-----T--------------AT-C--CC-T----GGC-A--CATGAATC-CT-TT- 1773
MND_2.GA.x.M14              -----G-----A--G--T--C-AG--------G---A-A--T------G----GC-G--G----G-A-AGAT-CAC-A--C----CC---------C----A---C-CA-----------C-----C------------------GC--C--G----G------CATGAACC-CT-TT- 1697
OLC.CI.97.97CI12            ------G-T--A--G--T--CTC---------------A-----T---GG-CT-G--G--A-------GG--C-------------CA-----------G-AGC----AA-A--G-----T--T-----T--T-----C--C--AGGAG-C--G-----G---...GT-AAAC--T-G-- 1676
GRV.ET.x.GRI_677            --G-----T--A--------CCAG------------A-A---------G-C--G-----A--A-----AC------T--T---------------C----T---C-TA----G-----C--T-----C--T--G---------ATTG---C-G----G-------ATGAATC-A--CT- 1933
VER.KE.x.TYO1_patent        --G-----T--A--G-----C-AG--T-----G--A-A--T--C--GG-AA-T-----A--A-----C---GG---A--G---C-A-------------T-A--AC-CA-T--------T--T-----T--T------------GTC--CC-G--G-GCC-A---ATGCACC-C--C-- 1993
VER.KE.x.9063               --G-----T--A-----T--C-AG--T--------TA-------C---G-AA-C-----A--A--G--C---GGA--A-----C-A-------------T-A--AC-CA-A---------T-----T-----------C--AGTC--CC-G---G---A---ATGCACC----CT- 1996
VER.KE.x.AGM155             -----G--C--A-----A--C-A---T---------A-A-----------G-AA-------A--A----------GGA--A--C---C-A-------------T-T---C-CA-T--G-----C-----C-----------C--AGTA-----G--G-GCC-A--GATGCACC-C--T-- 1985
VER.DE.x.AGM3               --G--G--C--A---G-A--C-A------C-----TA-------C---G-AA-------T-----G--C---GGA--A------C-G-------------T-A--AC-CA-T--G-----T--T-----T--T-----C-----AGTC--CC----G-GCC-----ATGCATC-C--T-- 1497
TAN.UG.x.TAN1               -----GG-C-----G--T----AG-----C------A-A--T---C--G-C--T-----A--A---A-AC-C--A--T-----------------------AC--C-CA----G-----C-----C------------C---TTGG---G----G-C-T---AT-CATC-A--T-- 1956
SAB.SN.x.SAB1C              -----G--T--A-----T--C-AG--T--------TA-A-----C--GG----G-----A-----GA-AGAC-CT-CT-----G--C----------C-GT-TC-----A----G-----C-----C------------------AG-GC-C----GG--A--CATGAATC-A--C-- 2086
MND_1.GA.x.MNDGB1           --G------------G-T----A------------TA-A-----C--TGTAA-G-----A--A----------GGA--A-----G-TG------CAGC-GCATC--C-TA-AG-------A-----A-------A--C---CAG-----G-G-AGC--A--GAA--GA--A--TT- 1447
DEB.CM.99.CM40              -----G--C--------A--C-A---T-----T---A----------G--CT-------T--A-----AGAT----CA------------------CACC---C-CA--A-T--------A-----A-----T-----C-G--TC--CC-G----G---A--T-AGAATC-C--CT- 1596
DEB.CM.99.CM5               --G--T--T-----------C-A------------TA-A-----C-----CT-G--G--T--A-----AGAT----CA--C---------------CACC---C-AA--A-T--G-----T-----T--------------GA-TA---C-----GG--A----AGAACC-C--TT- 1587
TAL.CM.01.8023              -----C--C-GA--G-------AG--G--C------A-------C-G-TGC--T-----A--A---A-AGAT----CA--C-----------------C-G--GC----A--G--G-----C-----C--------A--C---G-C---C-G--G--CA----T--CCA---G--C-- 1203
TAL.CM.00.266               -----C--C-G-----A--C-A---------------A-------C-G-TG---T--G-----A---A-AGAT----CA-------------------C-G--GC----------G-----T-----T--------G--C---G-C--C--------CA----C----GA--C--CT- 1705
SUN.GA.98.L14               --G---------A--G--T--C---------AC---TA------TC--GG--T-------A--ATC--GC---G----A------G-G-------A---G-AAA--C-A--G--------A-----A---------------ATTGG-AA----G---C-----GA-A-TC-AT--T- 2060
WRC.CI.97.97CI14            -----G--T-----GGAT--C-A--C---------TA-A---------GTAT-------A--A----GAG-C-CA--A------C-A-------------C-AGC-C--AA-A--G-----T--T-----T--T--G-----C--A--A--C-------G---A...-A-CAAC-A--CT- 2018
WRC.CI.98.98CI04            --G------------GAT----A-ACT-----G--TA-A---------GGTC--G------A--A----GAG-C-CA--A--G---C-G-------------C-AA--A---A-A--------T-----T--------------AG----C-----G-G---A...-A-CA-C-A---T- 2036
WRC.GM.05.Pbt_05GM_X02      -----G-----A---GAT----A-TCT--------TA-A--T-T---GGTC--T------A--A----GAGTT-CA--A----GC-A-------------C-AG--AC-AA-A--------A--T-----A--T--------C--ACAG-----G----GCC-A...-A-AGTC-A---T-  980
LST.CD.88.524               --G--T-----A-----A--C-AG-------CT--TA------T---GG-CT-------A--AGG--G-------T--A--G---G-G-------A---G-AAA--C-A--G--G-----A-----A-----T--C--C--A-TGG-AA----G--C--A---GA--GT--C--TT-  983
LST.CD.88.447               --G--G-----A-----A--C-A---A-----C----A-------T----G-AT-------A--AGG--GC-----T--A-------G-A-------AG--G-AGA--C-A--A--------A-----A---------------G--TGG-GA-------C--A---GA--G------TT-  989
LST.KE.x.lho7               --------C--A--G--T----A---A-----C---TA------T----G-A--G-----A--AGG--GC--T--T--A-----GG-G-------AG--G-AAA--C-T--G--G-----A-----A-----T--C--C-G-TTAG--A--------C-A---GA--G---AT-T-- 2073
SYK.KE.x.SYK173             --G--C-----A---ATT--C-A------------A-A-----CC-TTGCT-------A-----G--GGAC-CATCA--C---GGG---C-A--TC----AC-----A-T--------C-----C---------A--C-GACAG--CC-C--G---A---TCAA-...C----CT- 1897
SYK.KE.x.KE51               --------T--A---A-A----AG--T-----T--TA-A-------C-TTG---T-----A--A-----GGAC-C--CT-------------C-C---C-GT-A--AC-AA-T--G-----A-----A-----T--G--C-GA-GC---C-G-------A---TCAA-...C-A--CT- 1525
COL.CM.x.CGU1               --G---AGT--A--GGAT----A-AG-----------A-A--T---TC-G-C-----G--A--A-CT--AG--GGA--AA-C--GGCC------G-TA-T-AT---C-AA-A--GC----C-----T--C--C--A-GA---C-G----G-C--CAGAAG...C-AT-TT- 1604
```

```
                                                                                                        Gag-Pol fusion TF protein start
                                                  Gag p24 Capsid end Gag p2 start               Gag p2 end Gag p7 nucleocapsid start
H1B.FR.83.HXB2            AGAAGAAATGATGACAGCATGTCAGGGAGTAGGAGGACCCGGCCATAAGGCAAGAGTTTTGGCTGAAGCAATGAGCCAAGTAACAAATTCAGCTACC...............ATAATGATGCAGAGAGGCAATTTTAGGAACCAAAGA...................... 1950
Gag                       __E__E__M__M__T__A__C__Q__G__V__G__G__P__G__H__K__A__R__V__L__A__E__A__M__S__Q__V__T__N__S__A__T__.__.__.__.__.__I__M__M__Q__R__G__N__F__R__N__Q__R__.__.__.__.__.__.__.__
H1A1.UG.85.U455           ---------------------C--------G-----------------A-----G-----------G--------T------CA-C-GA--AGC..................---------------------------GG--CG......................... 1399
H1B.US.90.WEAU160         ------C-----------------------G-----G-----------A---------------------------------------A--------................G--------------------------G--C----...................... 1949
H1C.ET.86.ETH2220         ------------------------------G--------T-C---C--A--------G--------G----------------AC--A--A-C...................-------------A-A---------A-GG--CT-A-...................... 1321
H1D.CD.84.84ZR085         ------------------------------G--G--G--A--------A-----------------G--------------C----G-G-CT--G-TGCC.............----------A-A-A---------A-GG-AC----...................... 1470
H1F1.BE.93.VI850          ------------------------------G--------TA----------------------C--G--------T----C--ATTCAG-C.....................-------------A-A-T-------A-GG-------...................... 1287
H1G.SE.93.SE6165          ------------------------------G--------A--------A-----------A-----G-----------G-CCT--GG-G----AG-AGCC.............---------------A---------A-GG--C----...................... 1356
H1H.CF.90.056             ------------------------------G--------TA-T-----A-----------------G---------------------A--AA---AGCC.............-------------A------C----A-GG-------...................... 1300
H1J.SE.93.SE7887          ------------------------------G--------T--T--------G-----------------------------G--C---A-CAAC..................-----------A-----T--C------G----T-A-...................... 1270
H1K.CM.96.MP535           ------G-----------------------G-----G--TA-------A-----A-----------G-----------G---------C---T...................G------------A------C----A-GG---T---...................... 1146
H101_AE.TH.90.CM240       ---------------------C--------G--------TA-------A-----G-----------------------C-C-CA-C--G-A--...................-----------------------C-A-GG---G......................... 1509
H102_AG.NG.x.IBNG         ------------------------------G--------T--------A-----G-----------G--------T------CA-C-GG-CAAC..................G--------------------------GG---G......................... 1475
H1N.CM.95.YBF30           ------G--------G--C--C--------G--G-----A-CA--------------GC-A--A--G-T---TCA--G--GCAGC-GC--A-A--TAGT.............G-CT-TGCA--A--G--A--C----AAGG-AT---G...................... 1517
H1O.BE.87.ANT70           ------G------GT---C-------------------AACT--------C-----GC-A--A---------GCTAC--CCCAGC-AGATTTG-AAGGAGGATACACAGCAG--T-C-----A-----GC-AAA-CCA-TTAGG-A-...................... 2014
H1O.CM.91.MVP5180         -------------GT---C-----A-----------G--AACT--C-------A-A-AC-A--A---------GCTTCT-CCCAGC-AGATTTA-AAGGAGGATACACAGCAG--T-C-----A-----GC-GAA-CCA--TAG--A-...................... 1989
CPZ.CD.90.ANT             ----------T-A-----T-----A-----------C--A-C---------------G-----A-----T---GCTTCT-CT-AT---G--CAGGGAACC.........GCAG-CT-TC--------------GGA-AT-GAGG-G--AAA................... 1380
CPZ.CM.01.SIVcpzCAM13     ------------------C-----A--G-----------A-C---------------C--A--------T---CAATG--CCA--G-CA---A-GAACA..........GAC--TT-CT-T----A---GCCAGGGGCA-C--C----...................... 1533
CPZ.CM.05.SIVcpzEK505     G--G-----------------C--A--G--G--------A-CT--------------AC-A--A-----C---CA--G-C-CA--CAG--A---GT................G-TT-T-----A-----A-------AAGG-AT----...................... 1498
CPZ.CM.05.SIVcpzMB66      -----GGGATGA-C----C-TC--------G--------A--------A--T-----CC----A-----------T----C-CA-C-CAGTAA-GATGCA..............-A--G---ATTTAAAGGCCCA-AA................................ 1484
CPZ.CM.98.CAM3            ---------------------C-----------G-----A-C---------T-----AC----------T--------C--CA---CC-TA-A................GGTG-GT-CC----A--G--A---AA-G----A-C--CT...................... 1340
CPZ.GA.88.GAB2            ---G--------------C--C--A-----------C--G-CA--------T------C-A--A-----T---------A--CA-C-GA--AGC...............TCA---T-C-----A----CAGGGGGA--A-CT-CTCCT...................... 1328
CPZ.CM.98.CAM5            ------------------C--C--------G--------A-C---------T------C----------C------------CA------CT-A...............AATG-GT-T-----------A---AA-G---GAGT--AT...................... 1643
CPZ.CM.05.SIVcpzMT145     ------------------T-----A--G------------TCA--------T-----GC----A-----C--------G---AC--AA--...................GAA--CT-T-----A-----GGGACAA-AT-GA-CTCCTAAT................... 1489
CPZ.TZ.00.TAN1            ----------C-C-----C-----A-----G---------TCT----------AGA--C--------------GC-TC--C----GCAGGG-GA...............GTA-AT---C-----G---A--AAGACCACC-TT--A-....................... 1591
CPZ.CM.05.SIVcpzLB7       ---G-----------------C--------C--G--T--A-C------A--C--G--C--A--A-----T-----T----C--ATGCAG--...............ACAGTC--T-------A--G--A------AAGGA-C--A-........................ 1496
CPZ.US.85.CPZUS           ----------G-------------A-----G----------CT-----------G--C--A--------C---T-----A-G-A---CC-CT-A...............AGCG--T-TC-A----A---A--CGC-G----G-C-G-G...................... 1998
CPZ.GA.88.GAB1            ---G--------------T-----A--------------ATCT-----------G---C-A--------C---TCAATG--TCAG---CA--GG-GAGCA.........GATG-TT-CT-C----AG--AC-AGG-GC-GG--C--A-...................... 2017
MON.CM.99.L1              G--G--C---C-ACAG--------------G-----T--AATG--C--AT-T--GA--C----A-----T---GCAGGG-C--T-GCCAAC............ATGCCCATG-AT---G-A---.............................................. 1593
MON.NG.x.NG1              ----------C-CCA---------A--G--------C----CA------AGC--GT-A--A--A--G--T---GC-AC--C--TT---AGCAAC-TGCCA...ATG......-AC---G-A--AG-CA-AGGAGG-..................................  174
GSN.CM.99.CN166           G---------C-T--------C--A--G-----G-----AACT--C--AT-T--GC-C--A--A-----C---GCAAC--CCCTC--AGGCA---GTTCA.........TAC-AC---G------A----CCCCC-G-A-GG---......................... 1630
GSN.CM.99.CN71            G---------C-T--------C--A--G-----G-----AACT-----AT-T--C--C-----A--G--C---GCAAC--C-CTC--AGGC...-G-TCA.........TAC-AT---G----A------CCCCCAG-A-GG---......................... 1633
MUS_1.CM.01.1085          G--G------T--CAG-----C--------G---------ACT-----AT-C--GC--A----A--G--TC--GC-TCG-CCTTC...AA-CAGG-AGGC...ACACTG...-AC---G----AG---C---GGGAGCC-GGGG---G...................... 1604
MUS_1.CM.01.CM1239        ----------T--CAG-----C--A--G--------G--A-CA------T-C--GC-AA----A--G--T---GC-TCG-CCCTT...AG-CAA-A-AGT...CAGCTG...-AC---G----AG--.........GCC-GAGGG-A-...................... 1597
MUS_2.CM.01.CM1246        -A----G---T-ACAG--C--C-----GA-T-AG------ATG--C---T----GC-GA----A-----C----CTA-T-CTTTG...AGGCAA-A-ACCATTAACACCATC-AC---G-A--A---......CAG--TCC-AGG......................... 1684
MUS_2.CM.01.CM2500        G---------C-CCAG--T-----A-----T-----G--TATA--C--AT-C-A-T-AA----A-----C---GCATC--CTCT-...AG-CAG-A-ACCATTGAGACCATC-AC---G-T---G-G......CAAG-TCC-AG-......................... 1646
RCM.NG.x.NG411            G--G--G---C-A--------C--A--G-C------T--ACAG--C-----T--G--GC-A------------CAGATG--GCA-.....................AGTAAT--C---GCC--ACAG--AGGACC---A--AGG-CCC...................... 1372
RCM.GA.x.GAB1             ----------C-A-----C---------------------CAG-----A--TC-G--AC-A--A--G--C---CAAATGA-GCA--G-AAT......................-C---GCT---CA-TCAGCAAAC.................................. 1360
DEN.CD.x.CD1              G-----G---T--CA---C--C--------------G--A--A--A-----CC-TCCCAC---A--G---T-AGCTTC--C-TTTC-GAAGCAAGGGTTATGCCATGGTTCA-GG-AAGG-GCCCA-CC-CCAG-AGAA-GA............................ 1732
                                               Gag p24 Capsid end Gag p2 Spacer start        Gag p2 Spacer end Gag p8 Nucleocapsid start
MAC.US.x.239              ----------C----G--T-----A--------G--G--G--A--G-----T--T-AA-----A-----CC---AAG-G-CCCTCGCAC---TGC-AATCCCT......TTTGC-GCAGCC--ACAGA-GGGACCA.................................. 2213
Gag                       __E__E__M__L__T__A__C__Q__G__V__G__G__P__G__Q__K__A__R__L__M__A__E__A__L__K__E__A__L__A__P__V__P__I__P__.__.__F__A__A__A__Q__Q__R__G__P__.__.__.__.__.__.__.__.__.__.__.__
H2A.GW.x.ALI              ------G---C-A--C--C--------------G-----A-----G--A--C--T-AA-----A-----CT-A-AGG-G-CC-TG-CAC-----C-TATCCCA......TTTGCGGCAGCC--ACA-........................................... 2250
H2A.DE.x.BEN              ------G---C-A--C--C--------G-----C-----A-----G--A--C--GC-AA-----------CC-A-AAG-G-CT-TGGGAC--AGCC-TATCCCA......TTTGC-GCAGCC--ACA-........................................... 2257
H2A.SN.x.ST               ----------C-A--C--C--------G-----T-----A-----G--A--C--T-AA-----A-----CT-A-AGG-G-CC-TGGCAC----CC--ATCCCA......TTTGC-GCAGCC--ACAG........................................... 1702
H2B.GH.86.D205            ---G------C-A--G--C--C--A--GA-------C--A--G--G--------GC-AA-C--------CT-A-AAG-G-CCCT--CAC-T--AC--ATACCG......TTTGCTGCCG-T--ACA-AAAGCAGGG-A-............................... 2230
H2B.CI.x.EHO              ----------C-A-----C--C-----GA-------C--A--G--G--------GC-AA-------G---TT-A-AAG--CCTTG-CAC-TT-C---AATCCG......TTTGCCGCCGCT--ACC-A-AGCAGGG.................................. 2228
H2G.CI.92.ABT96           ---G------T-------C-----A--GR----------A--G--A--A--C--GC--A------------T-A-AAG--CCTT---CC-TA-AG--CTACCA......TTTGC-GCTGCT--ACA-AAA-CAGGAG-A............................... 1632
H2U.FR.96.12034           ------G---C-A-----C-----A---A----------A--A--A--A--T--GC-GA----A-----CT-A-AAG-G-CCTT-C-AC-TATGC-TATACCC......TTTGC-GCAG-C--ACAGA-AGGCGGCCCA............................... 1748
MAC.US.x.EMBL_3           ----------C----G--T-----A--------G-----A--A--G-----T--T-AA-----A-----CC---AAG-G-CCCTGCGAC---TGC-AACCCCT......TTTGC-GCAGCC--ACAGA-GGGACCA.................................. 1699
SMM.US.x.H9               ----------T-------C--------------R--G--A--A--A--AS-T--C-AA-----A--G---T---AAG-T-C-CTG-CGCA--GGC-ACTCCCA......TTTGCRGCAG-C--ACAGAAAGGACAA.................................. 1698
SMM.US.x.PGM53            ----------C-------C-----------------G--A--A--A--A--T--GT-AA----A--G---T---AAG-G-C-CTG-CGC---GGCAACTCCCA......TTTGC-GCAG-C--ACAGAAAGGACCA.................................. 2142
SMM.US.x.SIVsmH635F_L3    ----------C-------C---------A-------G--A--A--A--A--T---T-AA----A------T---AAG-G-C-CTG-GAC---ACCAACTCCCA......TTTGC-GCAG-C--ACAGAAAGGACAA.................................. 2228
SMM.SL.92.SL92B           ------G---C-A-----C-----A--G--------C--A--A--G------CC-CC-CA---A--G--C----AAG-C-CCCTC-CGGGGT--TTAGTG..............GCAGCA---TTTA-GGGAGCAGC---AGGCCA-....................... 1659
STM.US.x.STM              ------G---T-A-----------A--G--T--G-----A--A--G--A--T---C-GA----A-----CT---AAG-G--CTTCC-AC---ACC-ACTGCCC......TTCGC-GCAGCC--ACAGCAGGGA..................................... 1872
MNE.US.x.MNE027           ----------C----G--T--C--A--------------A--A--A------A-T-AA-----A-----CC---AAG-G-CCCTTGCAC---GGC-ACTCCCT......TTTGC-GCAGCC--ACAGAAGGGACCA.................................. 1693
DRL.x.x.FAO               G---------C--TT---T--C-----T-----------A--------------G--CA----A-----T----AG--CAGCAGGCAG-T...................--------------CAGAATCCCCCC--AGGT-C-CCG....................... 1454
MND_2.x.x.5440            G---------T--CTG--C--C--A--G-----------TAAAT-C-------CA-A-GA---A--G--C---CAG----C-CAGGCAG-T.....................G------------ACTCGGGGGGACC-CCG-GGG-T...................... 1457
MND_2.CM.98.CM16          ------G---CC-CT---C--C--A--G--------C--TAAAT----A----CA-A-GA---A--G--T---AGG---CCCAGTCA......................GCTG------------ATTCGGGAGGGCCACCG-GGG-C...................... 1900
MND_2.GA.x.M14            ----------T-ACT---------A-----G--G-----AAAAT---------CA-A-GA---A--G--C---CAGA-GA-GCAGTCA.....................GAAG-C----------ACTCGGGGGGACCACCG-GGG-T...................... 1824
OLC.CI.97.97CI12          G-----G---C-------C-----A--------G-----A---T----A--------C--A--A-----GC----A-C---GCAG-TGGTGCAACAGAGACCAGCACCA............................................................. 1785
GRV.ET.x.GRI_677          G--G------C-A-T---T--C--------------G--ACAA-------T-AGC-AA---TA---ATG-----TA-T-G-CAG...........................-AT---G-C--AGTG--ACC-CAG-AA--GGGCCCC....................... 2054
VER.KE.x.TYO1_patent      T---------T-A--G--T--------G--------C--AA--T-C--A----A---AA---A---ATG---CAGACCA-GCA----CA-.....................-AC---G-----CAG--AGG-CCA-AA-GA---.......................... 2117
VER.KE.x.9063             ----------C----T--C-----A--------G--C--AA--T-C--A---------A---A---ATG---CAAA-CA-GCAG-C-CA-.....................-AC---G-----CA---GGCAGCAG-AGGAGT---G....................... 2123
VER.KE.x.AGM155           T---------C-------C-----A--G--G-----C--AA-TT-C--A--C-A---CA---A---ATG---CAGA-CC-GCAG-G-CAG.....................-AC---G-A---CAG--AGG-GGA---GGAAG-CC-....................... 2112
VER.DE.x.AGM3             ----------T-A--T--C-----A-----G--------AA-TT-C--A----A---GA---A---ATG---CAAA-TA-GCA--GCCAG.....................-AC-------ACAG---GG-CAG--AGGAAG-CC-........................ 1624
TAN.UG.x.TAN1             ------G---C-------T--------------------A--A-----A---AGC-CA---TA---------CAA---A-GCA-GGA......................GTT-AC---G---AG---CACCGCGAG-AGGTAGGG-C....................... 2083
SAB.SN.x.SAB1C            ------G---C-T--T-----C--A--CA-T------G-ACAA---------C-TC-AA----G-----T----CAGCG-CCTTTC-GCA-CAA--AGTT.....GGGAAC--CT-TG-----CA---GGCAAGACCC-GGGGTCCC....................... 2228
MND_1.GA.x.MNDGB1         ---G------T--GA------------------T--G--ACAA-----A--C---T-AA---A---ATG-----AAC---GGT-GGACA-......TCA............CA--ATT-TGT-CAGCAG-GAGGGCCTC-AAG-.......................... 1574
DEB.CM.99.CM40            G-----G---C-ACAT-----C--A--G--G-----C--ACAG--A-----C---C---A---G--G--T---GC-TC--CCTT---AGAG--AGGGTCC.........CTAGGC---G-A--ACA-A-G-GAGGACCCCCTGGTTCT...................... 1735
DEB.CM.99.CM5             ----------T--CAT-----C--------------T---CAA--G--A--CC-TC-GA----G--G--T---GC-TCG-CCTT---AGA--AGG-TCC..........CTGGGG---G-----CA-A-AGGAGGACAAGGAGG-G-T...................... 1726
TAL.CM.01.8023            G---------T--CAG--------A--G--C--G-----TCAG-----AGC---C-CA-----A-----C---GCGG-G-CCTTG--A...CAG-A-ACCCAGCAAGTCATAGCC---G-A---...--AC-GGGAG-TCCG............................ 1339
TAL.CM.00.266             ----------C-CCAG--C-----A--G-----------CAG------AGC--GC--A-----A-----T---GCAG---CCTT---A...CAA-A-ACCCAACAAGTCATAGCC---G-A--ACAG--GC-AGGAG-CCCA............................ 1844
SUN.GA.98.L14             ---G--G--------G--C--C--A--G--T--G--G--GA-------A-G--AGA-CC----A---------GC-TCT-CTTTT-GGCAGCAAGTAGGA......AGACAAGCT---G-----CA-AA-TTACCACCC-GAA-C--CCAG................... 2205
WRC.CI.97.97CI14          ------G---T-A-----------A-----------G--A--AT----AAGC-A-A-AA----AC----TT-ACAAG---CTTTG...CAT...T-ACAG...........GCC---G-A--ACAG-CAGGAGCA-AACCT............................. 2142
WRC.CI.98.98CI04          ----------T----T--------A-----------G--A--AT----AAGC-AGA-AA----AC----CT-ACAGG-G-CTTTG...AGC...T-TCAG...........GC----G----ACA-CAGCCAG-G-AACCT............................. 2160
WRC.GM.05.Pbt_05GM_X02    ----------T-------------------------T--A---T----AAGT-A-A-AA----AC-----T--CAGG---CCCTG...--CCAAT-ACAG...........GC----G-A--ACAGCAGGCAG-C-A-................................ 1104
LST.CD.88.524             --------------A------------------------T-C------AAG--A-A-C--A--A--G------GCAAC--CC-TGC-GGGGCAG-TGAAA.........CATC-C---G-----CA-ACACCACC---A--TGC-CA-...................... 1122
LST.CD.88.447             ------G------GG---C-----A--------------A-C------A-G--GA--------A---------GCAAC--CC-TGC-AAGTCAA-TGAGA.........CAG-AT---G-A--AGTGACACCACCA-----TGC-CAG...................... 1128
LST.KE.x.lho7             ------G-------A---T--------------------T-CT-----A-G--A-A-A-----A---------GCTTC--CC-T-C-GCA-CAAGTAAGA.........CAA-AC---G-----GTCACTCCGC----A--TGCGCA-...................... 2212
SYK.KE.x.SYK173           G--G------C--CA---T-----A-----C---------TTA--C--------A-C-CA---A--G--C---GTAATG-CCCAGC-G----TA.................-AT---G-A--AG-GCCATCAAAAG--................................ 2018
SYK.KE.x.KE51             G---------C-TCA---T--C--A--G-----------ACAAT-C--A----A-C-A--A--A---------GTAATGAC-CA-C-CAGCTTG.................GGC-----A--AG-GCCA-GACAAG-A-G-A-TCCC....................... 1655
COL.CM.x.CGU1             G-----C---C-AG-------C--------G--------A-A---C-------AG--GC-A--A--G--T---CAG--GT-CCAGC-AGAGAGA...............ACC-AC-----TG-AGT-AAG-CAGCA-AATG-TTC-ATTGTCAGGGAATAGGGCACTTAG 1759
```

```
H1B.FR.83.HXB2             ...................................................................................AAGATTGTTAAGTGTTTCAATTGTGGCAAAGAAGGGCACACAGCCAGAAATTGCAGGGCCCCTAGGAAAAAG............... 2022
Gag                        .__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__K__I__V__K__C__F__N__C__G__K__E__G__H__T__A__R__N__C__R__A__P__R__K__K__.__.__.__.__._
H1A1.UG.85.U455            ...................................................................................-GA-GAA-------------C--------------A---CT-----A------T------------------............... 1471
H1B.US.90.WEAU160          ...................................................................................--A-C-A----------------------------A----T-------------------------------............... 2021
H1C.ET.86.ETH2220          ...................................................................................-GAGCAA----A--------C--------G---------CT------------------------------A............... 1393
H1D.CD.84.84ZR085          ...................................................................................--A-----------------C-------------------T---A--------T------------------............... 1542
H1F1.BE.93.VI850           ...................................................................................-GAG-------A-----T-----------------A----T------------------------A------............... 1359
H1G.SE.93.SE6165           ...................................................................................-GA-C-A-C--A-----------------G-----A--TCT-------------------------------............... 1428
H1H.CF.90.056              ...................................................................................--AT-------A--C-----C-----------G--A----T-------------------------------............... 1372
H1J.SE.93.SE7887           ...................................................................................-GA---------------------------C----A----T---A-A---C--------------A------............... 1342
H1K.CM.96.MP535            ...................................................................................--A-----------C-----C-------------------T------------------------A------............... 1218
H101_AE.TH.90.CM240        ...................................................................................--A-GAA-------C-----C-----T-G------A---CT------------------------A---C--............... 1581
H102_AG.NG.x.IBNG          ...................................................................................-GA-CAA-A-----------C--------------A---CT--------------A--------------G-............... 1547
H1N.CM.95.YBF30            ...................................................................................--ACCCA----A--------------------G--C--TTTG--A-----C--T-A---------A-G-GGA............... 1589
H1O.BE.87.ANT70            ...................................................................................GGA-C-A-A--A--------C-----A-----G--A--T-T---A--------TC-A--A------------............... 2086
H1O.CM.91.MVP5180          ...................................................................................GG-CCCA-A--A--C-----------A-----G--A--T-T---A-A---C--TC-A--A-----A---G--............... 2061
CPZ.CD.90.ANT              ...................................................................................-GACC-C-C--A-----T--C--C--T-----G--C--T--T--A----------A------A--A-GG--A............... 1452
CPZ.CM.01.SIVcpzCAM13      ...................................................................................-GA-AAA-A--A--C-AT--------A--G-----C--TCTG--------C--T-A---A--A----G----............... 1605
CPZ.CM.05.SIVcpzEK505      ...................................................................................----CCA-C--A--C--T--------T-----G--C--TCT---A-----C--T-A---A-----A------............... 1570
CPZ.CM.05.SIVcpzMB66       ...................................................................................-GA-----C--A--C--T-----------G-----T----T---T--G-------A------C--A-G---A............... 1556
CPZ.CM.98.CAM3             ...................................................................................-G--AAA----A--C--T--C--------G-----A--TTT---------C-----A-----C--A-GG---............... 1412
CPZ.GA.88.GAB2             ...................................................................................-G--GA--C-----C-----------A--G-----T--T-T---A--------T-AA--------A-G----............... 1400
CPZ.CM.98.CAM5             ...................................................................................----AAA----A--C-AT--------------------TGT---T-----C-----A-----A--A-G----............... 1715
CPZ.CM.05.SIVcpzMT145      ...................................................................................--A-AGA----A--C--T--------A--------A---CT---A--G--C--------A--CC-A-G----............... 1561
CPZ.TZ.00.TAN1             ................................................................................AAGGGTCAGC-GC-------T--C-----G----T---C--T-----A--------T-----T--A--A--G--A............... 1666
CPZ.CM.05.SIVcpzLB7        ...................................................................................-GAG--A----A----------------------------T---A--------T-AA-----C----G---A............... 1568
CPZ.US.85.CPZUS            ...................................................................................-GA-AAA-C--A-----T--------------G--A--TCT------------------T--A--A-G---A............... 2070
CPZ.GA.88.GAB1             ...................................................................................-GA-AAA-A--A--C--T--------G--------T---CT---T--------T-AA--A--A--A-G---A............... 2089
MON.CM.99.L1               ...........................................................GCCAGAGGACCACCCCAAAGGAAGGGACAACCC--A--C-----------A---TTT-------TG--G-A--------AA-----CCAA-GG-G-............... 1689
MON.NG.x.NG1               ..............................................................RGTCAGCCACGCCGACAAGGCRW-CAAA-C-G-----AT--------G---TTT--A--TGTG--A-AG--C--T-CT--T--C--A--G-CA...............  267
GSN.CM.99.CN166            ................................................................................GGA--A-CCCCA--A----AT--------AC--TTC--A---CT---A---G---TCCTAAA--C-AAG---GA............... 1705
GSN.CM.99.CN71             ................................................................................GGA--A-CCCCA--A----A---------A--GTTT--A--T-TC--A---G-C--TCCCAAA--C-A-G---GA............... 1708
MUS_1.CM.01.1085           ...........................................................GGCAACGGCAGCCAGGGAAACCGAGGA-A-CCC-GA--C-A---C-----GC--T-T--C---GT------GG-----CCCAAA--C-----C--A............... 1700
MUS_1.CM.01.CM1239         ..............................................................GGCAGCCAGGGAGGTCCCAGGGGA-ACCC--G-----A---C-----GC-GTTT--A----TG--A----GC---CCCAAG--A-AA-CT-G-............... 1690
MUS_2.CM.01.CM1246         ...........................................................GGAGTAATGGGAAAAAAGAGGGAA--C-GCACACGT--C-A---C-----AC--TTT--A--TTTG--A---G-----CCCAAG----A-TCT-CC............... 1780
MUS_2.CM.01.CM2500         ...........................................................GGAGGGGCCGGAGGAGGGCCTAGA-G--CCCCACG---C-----C-----GC--TTC--A--T-T-----A-G-----CCCAAA--A--AGTT-G-............... 1742
RCM.NG.x.NG411             ..............................................................CCTAAACTAGGGGGAGGACCT-G-T-CC-C--A--C-AT------------AC---T--T--T--T---T------AAA-AAG---A--G--A............... 1465
RCM.GA.x.GAB1              .....................................................AGGGGGCCTCCAAGAAGATCAGGAGGAAATCCA-A-T-A-GA----A---------T--GCC---A----TTT-T---T----T-AA--------A-GG---............... 1462
DEN.CD.x.CD1               ...................................................................................-G-GGAC---GA---------------C--ATT--A---CT-----AGG----T--AAAG--C-A--G-GT-ACT............... 1807

MAC.US.x.239               ................................................................................AGA---CCAA---------GG--------G-----G--A---T-T--A--GC-A-----A-----A--A-G-C--............... 2288
Gag                        .__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__._R__K__P__I__K__C__W__N__C__G__K__E__G__H__S__A__R__Q__C__R__A__P__R__R__Q__.__.__.__.__._
H2A.GW.x.ALI               ................................................................................AGA-G--CAA-------C-GG--------A--G---------T-G--A---C-A---C-A--A--C--A---C--............... 2325
H2A.DE.x.BEN               ................................................................................AGA---GCAA---G--A--GG--C-----A--G--G--A---T-G--A---C-G---C-A--A-----A-G-C--............... 2332
H2A.SN.x.ST                ................................................................................AGA-G--CAA-------C-GG-----C--A--G---------T-G--A---C-A---C-A--A-----A-G-C-A............... 1777
H2B.GH.86.D205             ................................................................................AGAGG--CA--G-CA--C-GG--C---------C-G--A-----------GC-A---------------A-G-C--............... 2305
H2B.CI.x.EHO               ................................................................................AAG-G--CA--G-CA--C-GG--C--C-----G-CG--A--T--------GC-G----A---------A-GGC--............... 2303
H2G.CI.92.ABT96            .............................................................................AAAAGG-GC-CAA-A--A----GG-----------G--G--A--T----TT---C-G----A-----C--A-G-C--............... 1710
H2U.FR.96.12034            ................................................................................AGG-G-CCA--C-GA----GG--C--------G-----A-----T--A--GC-G--T-AA---------A-G-C--............... 1823
MAC.US.x.EMBL_3            ................................................................................AGA---CCAA---------GG--------G--G--G--A---T-T--A--GC-A-----A-----A--A-G-C--............... 1774
SMM.US.x.H9                ................................................................................RGA-----AA--W------RG--C---R-A--G-----A---T-T--A---C-A-T---A---------A-G-C-R............... 1773
SMM.US.x.PGM53             ................................................................................AGA----CAA-------C-GG--------A--G--G--C---T-T--A-A-C-A-----A---------A-G-C--............... 2217
SMM.US.x.SIVsmH635F_L3     ................................................................................AGG-G--CAA-C-------GG--------A--G--G--A---T-T--A---C-A-----------------A-G-C--............... 2303
SMM.SL.92.SL92B            ..........................................................................GGGAATAAGCCT--CA-C-G------T--C-----A--GAC---A---T-T--G---C-G-----A--T--C--A-GG---............... 1740
STM.US.x.STM               ................................................................................CGA-G--CA--A--A--C-GG--------G----C----------T--A-A-C-G--T-AA-G---G--A-G-C-A............... 1947
MNE.US.x.MNE027            ................................................................................AGA---CCAA---------GG--------G-----G--A---T-T--A--GC-A-----A-----A--A-G-C--............... 1768
DRL.x.x.FAO                ...........................................................CGGGGACCCAGGGGACCCCCTCCA--A-ACCC--GA---CCA--C--C--GC-GTTT--A--T--CCTG---C-A--T-CAAAG-----A------............... 1550
MND_2.x.x.5440             .................................................................CCCCCGAGACAACCCCCT-GA-ACCC--GA--CCC------C--A--GTTT--A--TGT-CTG--GG-------A-----A--A--G-GA............... 1547
MND_2.CM.98.CM16           .................................................................CCCCCGAGACAACCTCCT-GA-ACA-C-GA--CCCT--C-----A---TTT--T--TGGGCTG--GG-C--T-TAAG----C-A------............... 1990
MND_2.GA.x.M14             .................................................................CCTCCGAGACAACCCCCT-GA-ACCCC-GA---CC---C-----A-GGT-T--A--TGTTCTG---G-C--T--ACT---C--A--G---............... 1914
OLC.CI.97.97CI12           .................................................................GGAGGAGGGCCTCAGCAAGG--GGC-G-G-----AT--------T-G-TTT-----T--TAAA-AT-----TTCACAAAAGG-TCC-.................. 1872
GRV.ET.x.GRI_677           ................................................................................CGAGG-CCGC-A--A--C--T-----------TTT--A--T-TGCAA--GG-A----A---A--A--AC-G-TC............... 2129
VER.KE.x.TYO1_patent       ................................................................................AGACCCCCAC-A-GA----AT--------A---TTT--C--T-TGCAA---C-A--TCC--AA--A-------CA............... 2192
VER.KE.x.9063              ..........................................................................AGACAGAGACCCCCAC-A--A----A------C--A---TTC--A----TGCAA--GC-A----CA-AA--A--A--G-TA............... 2204
VER.KE.x.AGM155            ................................................................................AGACCCCCGCCA-------A---C-----A---TTT--C----TGCAG--GC-G--TCCT-AG--A--A----TA............... 2187
VER.DE.x.AGM3              ................................................................................AGACCCCCA--A-------A---------A---TTT--C--T-TGCAA---C-A---CCT-AA--A--A--G-T-............... 1699
TAN.UG.x.TAN1              ..........................................................................AGAGGCAGAGGCCCACCCCGC--C-----A-----AC--ATT-----T-T-CAA-A-G---TCCAAGAG-AG-ACC---C............... 2164
SAB.SN.x.SAB1C             ........................................................TTAGGGGGAAGAGGTAGACCTCTGAACCCA-A-A----A----A---------A--GCCT--T---CTG--T---TT---T-A-----------G-C-A............... 2327
MND_1.GA.x.MNDGB1          ...........................................................GGACCAGTTAGACAACCTACTGGA-G--AACC--TC--C-----C---AAT-----------TGT---A--GTTC-T--A----------A-GG--A............... 1670
DEB.CM.99.CM40             ................................................................................AGA-GACAGA---GA--C-----------AC--ATT--A--TCTCCAA-AGG-C----AAAGA-----A--G-CC............... 1810
DEB.CM.99.CM5              .............................................................................CCTAGA-GACAGC---GA--C--T--------AC-GATT-----TGTGCAA--GG----T-AAAAG-----A---GTT............... 1804
TAL.CM.01.8023             ..................................................AAAGGGAGAGGGGGACCAAGAAGAACTCCCCCAGGACAAA-C-GA----AT------------GTTT-----T-TT--A-A---C--TCC---T---CA---GCCACGGGGG......... 1450
TAL.CM.00.266              ..................................................AGGGGGAGAGGGGGACCTAGAAGGAACCCCCCAGG-CAGA---G-----AT--C-----A---TTT--C--T-T---A-AG--C---CC------GCC--G-C--AAGGTG......... 1955
SUN.GA.98.L14              ...............................................GGGAGGTTTGTAAGAATAGGAGGAGGAGGACCTAGG---CCCA-G-CT-----T---------T---CC---A--TCT---A-ATC-G------A-AG--A-A----GGA............... 2313
WRC.CI.97.97CI14           ................................................................................AAAGG-CCAA---GA-----T--------AC--ATT-----T-TG--A-A-G-A--TCCCAAG--A-A----CCAGTCCCTATGAGAAGA 2232
WRC.CI.98.98CI04           ................................................................................AAAGGACCAA---GA--------------AC-GATT-----T-TG--A-A-G-C--TCCCAAG--A-A----CCAGTCCCCTTGAGAAGA 2250
WRC.GM.05.Pbt_05GM_X02     ................................................................................AGAGGACC-C-A-GA-----T--------AC--ATT-----T-TG--A-A-G----T-A-AAA--A-AA-GGTCTCCTCAGCAAAGG... 1191
LST.CD.88.524              ...............................................GGGAGATTTGTAAGAACGGGGGGAGGTGGCCCAAGA-GACC-C-C-CT-------------G--GCC---A--TT----A-A--TG------CAG-----AC-GG-A............... 1230
LST.CD.88.447              ...............................................GGAAGGTTTGTAAGAACAGGAGGAGGAGGCCCTAGA---CC-T-A-CT--C--T--------A--GCC---T--T-----A----TG--T--ACAG-----AC-GG--............... 1236
LST.KE.x.lho7              ...............................................GGCAGGTTTGTGAGAACAGGAGGAGGAGGTCCCAGG--ACCCC-C-CT--C-----------A--GCC---T--------A----TG-----ACAG--C--AC-GG-A............... 2320
SYK.KE.x.SYK173            ................................................................................AGATCA--GA----A--C-A---------GC--ATT-------TGCAA-A-G-C----AAAAG--ACTT---GCT............... 2093
SYK.KE.x.KE51              .............................................................................AGGAGAGG-CCAACA-GA--C--T--------GC-GTT---A--TCTTCAA-A-G-C---CCTAGG--A-AA--GCTT............... 1733
COL.CM.x.CGU1              CCAGAATGTGTCCCAAAAGACCAATAGGAGGAGCCGGTCGAGGAAGAGGCCGAGGACGGGGAGGCTTTAGAGGAGCTCCCAGA-G-CCA--G-G------T-CC--CAATC-G-----A----TGCAA---G----TCCC.............................. 1899
```

```
                                                                                              Gag-Pol -1 ribosomal slip site
                                                                                         Gag p7 nucleocapsid end  Gag p1 start                                    Gag p1 end  Gag p6 start
H1B.FR.83.HXB2          ............GGCTGTTGGAAATGTGGAAAGGAAGGACACCAAATGAAAGATTGTACT..........................GAGAGACAGGCTAATTT.TTTAGGGAAGATCTGGCCTTC......CTACAAGGGAAGGCC.........AGGGAATTTTCTTC 2137
Gag                     _.__.__.__.__G__C__W__K__C__G__K__E__G__H__Q__M__K__D__C__T__.__.__.__.__.__.__.__.__.__E__R__Q__A__N__F#__L__G__K__I__W__P__S__.__.__Y__K__G__R__P__.__.__.__.__G__N__F__L__
Pol                                                                                                                  F_#_F__R__E__D__L__A__F__.__.__L__Q__G__K__A__.__.__.__.__R__E__F__S__S
H1A1.UG.85.U455         ............-------------C--G--A-------------------C--C---..........................------------------.--------A--T--------......-A-------G----.........-----------C-- 1586
H1B.US.90.WEAU160       ............-----------------C--------------------------------..........................------------------.------------T---......-C-A-----------.........-----------C-- 2136
H1C.ET.86.ETH2220       ............------------------------------------C----C..........................------------------.-------GAC-T--------......-A--------------.........---------C---- 1508
H1D.CD.84.84ZR085       ............-----------------------------------------C---..........................--A---------------.--------A--T--------......-C--------------.........G-----C------- 1657
H1F1.BE.93.VI850        ............------------------GA--------------------C--C---..........................--A---------------.--------A--T--------......-A-------G----.........C--A-----C---- 1474
H1G.SE.93.SE6165        ............-----------------------------T-----------C--C--G..........................------------------.--------A--T--------......-A-------G----.........-------------- 1543
H1H.CF.90.056           ............------------------GA--------T--G--------C--C--A..........................------------------.--------A--T--------......-AG---A--G----.........---A--------C- 1487
H1J.SE.93.SE7887        ............------------------------------------C--C---..........................------------------.-----------T--------......-AG---A--G----.........-------C-----C- 1457
H1K.CM.96.MP535         ............----------------G-------------T--G---------C--C---..........................------------------.--------A-----------......-C-------G----.........-------------- 1333
H101_AE.TH.90.CM240     ............--T-----------C--G------------T----------C--C---..........................------------------.--------A--T--------......-A--------------.........G----------C-- 1696
H102_AG.NG.x.IBNG       ............----------------G--A-------------------C------..........................------------------.--------A--T--------......-AG------G----.........---A-------C-- 1662
H1N.CM.95.YBF30         ............-----------G----GC-A--------T-----------------AAAATGAA..........................GA-----------------.-----------G----T--C-........--T---A--G--A--.........---A--C--C-CC- 1710
H1O.BE.87.ANT70         ............--T--C------------C-------T-----------------C-GAAAT..........................GA-A-------A------.------C--ATA-------C-......GGGGGGCACG----.........---C----A-G-G- 2204
H1O.CM.91.MVP5180       ............--T--C------------C-------T-----------------C-AAAAT..........................GA---------A------.----------TA-------C-......GGGGGGCACG----.........---C----A-G-G- 2179
CPZ.CD.90.ANT           ............-----C----G-------C-----------------GC-T---A-C---C-AGCAACA..............AATACAG--A-A-TA-----.---------ACCGACC--CA-............-T-GT--GGGTGCAGACC------C---G-G- 1579
CPZ.CM.01.SIVcpzCAM13   ............-----C----------GC---G---G---------------C---T-A......................G-----A-TA------.---------GG--------------......-CG--GC--G----.........---C--C---G-A- 1720
CPZ.CM.05.SIVcpzEK505   ............-----C----------GC-A--G-----T-------G--C--C-GAAGT..............GGA--A--G-----G--------.---------G--------CT........--C---A--G----.........---A-----C-CC- 1691
CPZ.CM.05.SIVcpzMB66    ............-----------G----GC-A----G--T--G----G-A---C--AAAT......................--A---------------.---------AC--------------......-AG------------------.........-----------C---- 1674
CPZ.CM.98.CAM3          ............--A--C-----G-----GC-A--G--G--T---------A-C--CC-AGCT..............GGA--A-----------------.---------AGGT---T--C-GTG......G-GC--GG-AT-AAAGAGACC------C--C--G- 1542
CPZ.GA.88.GAB2          ............--A--------G--------A---------T--G-----G--C--CCAAAAT..............GAA-GA-------TG-----.--------GG-GG---------CCG......G-GC--G----.........------C----CC- 1521
CPZ.CM.98.CAM5          ............--A--C--------C---C-------G--TA-C---------A-----AACTGGA..................GC--------------.---------GAT---T--C-GTG......G-GC--G--AT-AAAAAGACC------C----CG- 1845
CPZ.CM.05.SIVcpzMT145   ............-----C-----G------C---G-----T--G---------------A-CAAAATAAT...............GA--G...AT-------.------A--A-AT------C-........-GGGG---G----.........------C-A-G-G- 1682
CPZ.TZ.00.TAN1          ............--T--C----GG------C-A--G-----T---------G--C--C--CACCAGAAAC..............AACAGC-CTGG--TA-----.---------ACG-ACC--C-T...............GT--GGGTGCAGACC------C---G-G- 1793
CPZ.CM.05.SIVcpzLB7     ............-----------------GC-----------T-GG---C---------ACATG..................G--AG...-T--------.---------AC-------------......--C-C---------.........---A--C----C-- 1683
CPZ.US.85.CPZUS         ............--G--C-----G----GC-----------T-----------C--C--AGCAGGA....................A-C--------------.---------ACAT---T-CC-ATCGTG...G-GC--GG-AT-AAAGAGACC------C--C--GG 2203
CPZ.GA.88.GAB1          ............----------CGG----GC-A--G----------------C--C--A.......................GA-------TG-----.---------AGGT--------CCG.......-GC--G----.........---C--C---G-G- 2204
MON.CM.99.L1            ............AAA---AC--T----CC-ACCG-------TCGCC-----C--CC-GCAGCCCCCC...........AAACAAA-C-A-GG--TG-----.---------TCCT-TCGG-C-............CAAG-A-GGGCC.......-A----C--C-C-- 1816
MON.NG.x.NG1            ............--T--C-TCCGC--C--C--------G--TRT-TCC---A-C--CC-A......AACGGAGGACAAAACCAACCCAGA-ACAG--TG-----.---------TCCT-TTGG...............CCCCG-CAAGAG......G---CCAAGGAACT  397
GSN.CM.99.CN166         ............AAG--C-TT--------C-GA-C---G---TTT-GC---CCA---GA...............ACAGGGACACCAATC-AGG--CT-T----.-----ACCTTGG-C--GGGCG............CC-CC-AG-CCC......CA----C----CAG 1829
GSN.CM.99.CN71          ............AAG----TT--------C----C-------TT-GCT-GGC-A--C-AAACA...............GGAACACCA-CC-AGG--CT-----.-----ACCTTGG-C--GGGCG................-C-AG-CCC......CA----C----CAG 1829
MUS_1.CM.01.1085        ............AAG----TC--------G-GA---------TTGGCA-G-C-A----GG...............TCTGATAATACC-A-GC--TG-----.--------G-TCC--TCGGCCT......-C-G--AA-G-----------.........GA----C--C-CC- 1824
MUS_1.CM.01.CM1239      ............AAA--C-TC----------G---G------TTGGCA---C-G--C-GG...............TCTGAAGGT-CC-A-TC-----------.--------G-TCCTCTTGGCCT......-C-G-GAA-G-----------.........-A----C--C-CC- 1814
MUS_2.CM.01.CM1246      ............A-A----TT--G-----T------------------TTGGCC-G-C-G--CCGA...............ACCGACACAGGCA-A--CGCT-TG-----.---------GATCTCGGC-T......-CGGTCTC-G-A-----------.........GA----C--C-CAG 1907
MUS_2.CM.01.CM2500      ............AAA--C-TT--G-----C-G------C----T-GCA---C-G--C-GA...............ACCAACTCAGACA-AG-CGCT-T------.---------AGGT-TCGG..........--T-C-CA-G-A-----------.........GA----C--C-C-G 1866
RCM.NG.x.NG411          ............--G--C----G-------G-------T--T-T------G----CC-A.......................A-G---T------.---------TGCTCCCTGGAA............-CAGG--AAACC...GAGGAACCT-CCCT-GA 1577
RCM.GA.x.GAB1           ............--A--C-------------TCCCC--AC--T-TCT-------C--C--A.......................AG--AATA-----.--------GAC--CCCTGGGG..........TCAGG--AAGCC...GAGGAACTT-CC-T-GA 1574
DEN.CD.x.CD1            ...CAACCAGGA-CG--C-TT------------CT---T--T---GCA--GA------GA...........AGTAACACTAATGATC--G------TA-----.---------TGGTCTTGGGAG............-AAG-AT--TCCCAGGAACTTTGTACAGTACT 1949
                                                                                               Gag-Pol -1 ribosomal slip site
                                                                                                        Pol start
                                                                                       Gag p8 Nucleocapsid end  Gag p1 Spacer start                               Gag p1 spacer end  Gag p6 start
MAC.US.x.239            ............--A--C-------------AATG-AC--TGTT---GCCA-A--CC-A..........................--C--------GGG---.------CCTTGGTCCATGGGG............AAAG-A---.........CC-C-----C-CCA 2397
Pol                                                                                                                   F__F__R__P__W__S__M__.__.__.__.__G__K__E__.__.__.__A__P__Q__F__P
Gag                     _.__.__.__.__G__C__W__K__C__G__K__M__D__H__V__M__A__K__C__P__.__.__.__.__.__.__.__.__.__D__R__Q__A__G__F#__L__G__L__G__P__W__G__.__.__.__.__K__K__P__.__.__.__R__N__F__P
H2A.GW.x.ALI            ............-----C-----------C---CC------T-TT---GC-A-C--CC-A..........................--A---------GG---.-------CTTGG-CCATGGGG............AAAG-A---.........TC-C--C--C-CCG 2434
H2A.DE.x.BEN            ............-----C-----G-----C---CC-------ATC---GC-A-C--CC-G..........................--A--------AGG---.-------TT-GG-CCA-GGGG............AAAG-A---.........TC-C--C--C-CCG 2441
H2A.SN.x.ST             ............-----C-----------C----C-------ATC---GC-A-A--CC-A..........................--A--------GGG---.-------TT-GG-CCATGGGG............AAAG-A---.........CC-C-----C-C-G 1886
H2B.GH.86.D205          ............--A--C-------------AAC-------ATC---TC-A-A--CC-A..........................--A--------GGG---.-------TTAGGACCCTGGGG............AAAG-A---.........TC-C--C--C-CCA 2414
H2B.CI.x.EHO            ............--A--C-----------C--C-------ATC---TC-A-A--CC-A..........................--A--------GGG---.-------TTCGG-CCCTGGGG............AAAG-A---.........TC-C-----C-CCG 2412
H2G.CI.92.ABT96         ............--G--C-----G-----G---CC---C---ATT---GCCA-A---C-A..........................-----------AGG---.-------TTTGGACCTTGGGG............AAAG-A---.........TC-C--C--C-CCA 1819
H2U.FR.96.12034         ............--A--C----------A----CC-------A-C---GCTAGC--CC-A..........................--A-CCA---TGGG---.------CTTGG-CCTTGGGA............AAA--A---.........TC-C--C--C-CCA 1932
MAC.US.x.EMBL_3         ............A-A--C-------------AATG-AC--TGTT---GCCA-A--CC-A..........................--C--------AGG---.------CCTTGGTCCATGGGG............AAAG-A---.........CC-C-----C-CCA 1883
SMM.US.x.H9             ............-----C-----G--------A-C---C--TGTT---GCTA-A--CC-C..........................--A--------GGG---.-------CTTGG-CCATGGGG............AAAG-A---.........CC-C-----C-CCA 1882
SMM.US.x.PGM53          ............-----C---------------AAC---C--TGTC---GCCA-A--CC-C..........................--A--------GGG---.-------CTTGG-CCATGGGG............AAAG-A---.........CC-C-----C-CCA 2326
SMM.US.x.SIVsmH635F_L3  ............-----C---GGG--C-----AACG--T--TGTT---GCCA-A--CC--..........................--A--------GGG---.-------TTTGG-CCATGGGG............AAAG-A---.........CC-C-----C-CCA 2412
SMM.SL.92.SL92B         ............--A--C-----------GG-A-------GTATTCA-GCCA-C--CC-A.......................A-CCAGA---GGG---.-----CCTTGG-CCATGGGG............AAAG-A---.........TC-C--C--C-CCA 1849
STM.US.x.STM            ............--T--------------C-AC-G--C--T--G---GCCA-A--CC-A..........................--A-------TGGG---.-----CTTTGG-CCATGGGG............AAAG-A---.........CCAC--C--C-CCA 2056
MNE.US.x.MNE027         ............-----C-------------C-AATG--C--TGTT---GCCA-A--CC-A..........................--C--------AGG---.-----CTTTGG-CCATGGGG............AAAG-A---.........CC-C-----C-CCA 1877
DRL.x.x.FAO             ............--A--C-TT-GG------GCTTT--AT--TATGT-A-G-A----C.......................CCA-A-----TG-----.---------T-CTCCCTGGGG............CA-CG--AAACC...CAGGAATTT-CC--AGG 1662
MND_2.x.x.5440          ............------TT--G------G-TCC------T-T-----GGA-C--C.......................CCA-AGAT--TG-----.---------T-C-CCCTGGGG............CA-TG-CAAGCC......CA----C--C-C.. 1654
MND_2.CM.98.CM16        ............--A--C-TC--G------G-TCT-------AT-----G-A----CC-A..........................-AGAT--TG-----.---------T-C-CCCTGGGG............CA-TG-CAAACC......CA----C--C-C.. 2097
MND_2.GA.x.M14          ............--A--C-TT--G-----GG-CAC------TATG----G-A-C--CC-A..........................-A-AT--TG-----.----------TGCTCCCTGGGG............CA-TG-CAAGCC......CA----C----C.. 2021
OLC.CI.97.97CI12        ............ACA--C-AC--G-----G--ACC---C---AT-GCA---A----C-GG.........GTGGGAGGACAACAACAAC----CA---TG-----.---------ATGGA--GGAGG.........G--A-C-GA---TCCCCTCAACAT-C-ACAGAGGA 2011
GRV.ET.x.GRI_677        ............AAA--C-TT--G--C--C--AATT--C--TATGGCA-----C--C-AG.......................A-TG-------A-----.------T-TGG-CATTGGGG............A---GC-AA...ACCAAG-AATTT-G-G-AAT 2244
VER.KE.x.TYO1_patent    ............AAA---CTA--G--------ATTG-------T-GCA-----C--C-GG.........................G------TG-----.-------T-TGGAC--TGGAT..............---GGCAAAACC...GAG-AATTT-CCCGCCG 2304
VER.KE.x.9063           ............AAA--CCTT-------G---CT--------TGGCA-----C--C-GG.........................G------TG-----.-------T-TGGAC--TGGA-..............---G-CAAAACC...GAG-AACTT-C-CGC-G 2316
VER.KE.x.AGM155         ............AAA---CTT------------CC---G---TT-GCA-----C--C-GG.........................G------TG-----.-------T-TGG-C--TGGAT..............---G-CAAAACC...AAG-AATTT-CCCGCAG 2299
VER.DE.x.AGM3           ............A-A--C-T--------G--ACC---G--TTT-GCA-------C-GA.........................G------TA-----.-------T-TGG-C--TGGAT..............----GC-AAACC...CAG-AATTT-CCCGCCG 1811
TAN.UG.x.TAN1           ............AAA--CCTA-------G---CC-------TT-GCA-------CCGC.......................AGTG-------C-----.---------GA--GCCAA-C-G..............-----CAAAGCC...GCGCAATTT---AGA.. 2277
SAB.SN.x.SAB1C          ............-----C-----------C-GCCC--AC--T--G-----------CAA..........................-A-----T------.-------TTTGG-CCCTGGGG............CA--G--AAACC...CAGGAATTTCCC---GA 2439
MND_1.GA.x.MNDGB1       ............--G--C-----T------GCAATG-AT--T--G-AAGCTC-A--CC--..............AAGCCAGCTCAGC--CAGAG--T------.------T-TGG-CCTTGGGG............TCCCTCCAAACC......G-------A--C.. 1792
DEB.CM.99.CM40          ............AAA----TT-------GC-------C---ATTGCA---A-----GGA.......................CAGACTCCGAG---A------.---------T-CTCCCTGGG-............CTCC---AGACC......CA----C-----AG 1925
DEB.CM.99.CM5           ............AAA--C-TC-------T-----G--C---AT-GCA--GA-C---GGA.......................CAA-TACC-AG---A------.---------TGCTCCCTGGG-............CTCCC--AGACC......CA--------T-GG 1919
TAL.CM.01.8023          ...CCTCCGGGGA------TC--G--C--GC--ATG-------GTGCCGC-C-G--C-GA.......................C--CC------A------.-----------GCT-TTTTAGG............-AAG-TC--CCC......G-TC--GGGA-CGA 1571
TAL.CM.00.266           ...CCGCCAGGCACT--C-AT--G-----C--ACC---C---AT-GCA---C-A--CT-C.......................TCAGC------C------.-----------TCT-TTTTAGG............-AA--TC--ACC......G-TC-GAGGA-CGA 2076
SUN.GA.98.L14           ...CCACCCGGCTC---C--------------ATG-----TA--CAAGCTC-A--CC-A.......................CAA-GA-A-----TG------.-------T-TGG-CCCTGGAA............CA-GG----GCC..........------C----CAG 2437
WRC.CI.97.97CI14        GGACAGGGACCCACA--C---GG-------G-A-T------TGT-CA--G-A-C---C-A.......................AAAATTAGA-------TA------.---------TGG-A-AGGGCA............-ATGCCCT-GAACAATTATCCAGT-CAGAAG- 2368
WRC.CI.98.98CI04        GGACAGGGACCCACT--C---GG-------G-AATT-----TGTGCAA-G-A-C--CC-A.......................AAAAGTAGA--G---AT-------.---------TGGGA-AGGGCA............-ATGCCCT-AAACAATTACCCAGT-CAAAAGG 2386
WRC.GM.05.Pbt_05GM_X02  ...GGACCTCCCAC---C---GG----------ACC--C---ATCCA----A-C--CC-A.......................GAAACAAGA--G---TA------.---------TGG-C-AGGGCA............-A-GCCCT-CAACAATTATCCAGT......... 1316
LST.CD.88.524           ............-----------C-----TGCTA---AC---AG----GCCC-G--CC-AAAACCA.......................A--G-GA---TG------.-------T-TGGTCCCTGGAG............-AATG-A--TTC......---C--C----CAA 1348
LST.CD.88.447           ............--A--C-----C-----T-GCA-G-A---TAGGT-TGCTC-A---C-AAAGCCT.......................A--G--A---TA------.------CT-TGGACCCTGGAG............-A-TG-T--TCC......---A--C----C-T 1354
LST.KE.x.lho7           ............--T-------C-----C-GTA---AG--TAG-T-TGC-C-G--CC-A.......................AAACCAA-AG--A---TG------.------AT-TGGTCCCTGGAG..........--GCAA-CCT--AGG......CAATT--CC---A- 2441
SYK.KE.x.SYK173         ............AAA--C-TT--T--------AAC---------T-GCA-GG-CC----GGCAACCCAAGAGAAATCAAGGCCCTCCA-TTGC-------------.---------AGGA---GGGGT..........-TCCA-G-----CCC.......--CC--C--C-CAG 2235
SYK.KE.x.KE51           ............AAG--C-TT--C------GGAACG------AT-GCA-GGC-G----GACAACCAAGAAAAGGACAAGGCAACCCTCCTCC------C------.---------AGGG---GGGAA...................CAGGAAACCCCC--CA-------C-G 1872
COL.CM.x.CGU1           ..........................................................................A-C-AG-----A------.-------GCA-C-A--GGGGTGGA................--TACAGACAGCCATCTTCCCC-C.. 1964
```

```
H1B.FR.83.HXB2          AGAGCAGACC......AGAGCCAACAGCCCCACCAGAAGAGAGCTTCAGGTCTGGGGTAGAGACAAC..................................................................................................... 2198
Gag                     Q  S  R  P  .  .  E  P  T  A  P  P  E  E  S  F  R  S  G  V  E  T  T  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Pol                      E  Q  T  .  .  R  A  N  S  P  T  R  R  E  L  Q  V  W  G  R  D  N  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
H1A1.UG.85.U455         ----------......--------------------C---A-T---TG--ATG----A-A---T........................................................................................................ 1644
H1B.US.90.WEAU160       ---------T......---A--------------------------------TCA---A---------......................................................................................................... 2197
H1C.ET.86.ETH2220       ----------......--------------------GAGTCT-AGACCAGAGCCAAC--CCC--C-ACCAGA................................................................................................. 1575
H1D.CD.84.84ZR085       ----------......--------------------C---AGCT-CGG-T-TG---AGGAGATA-C-...................................................................................................... 1718
H1F1.BE.93.VI850        ----------......--------------------C----------G---TCA-A-AG----T---...................................................................................................... 1535
H1G.SE.93.SE6165        ---A---GA-......---A-------------T-C---A---C--G---TC--A-AG----T-G-...................................................................................................... 1604
H1H.CF.90.056           -------G--......---A----------------C----------G---TC--A-AG----TG--...................................................................................................... 1548
H1J.SE.93.SE7887        ----------......--------------------C-------C--G--CTC--A-AG----T......................................................................................................... 1515
H1K.CM.96.MP535         ----------......--------------------C----------G---TC----AG----T---...................................................................................................... 1394
H101_AE.TH.90.CM240     ----------......--------------------C---A-A--GGG--ATG----A-----T---...................................................................................................... 1757
H102_AG.NG.x.IBNG       ----------......G--A--------T-------C--------TG--ATG----A-----T......................................................................................................... 1720
H1N.CM.95.YBF30         ---CA-C-A-AAGGAA------C--------G---CT------T-ATG---T-CA--AG--............................................................................................................ 1771
H1O.BE.87.ANT70         ----ACC-G-......-C-C---T---------................................................................................................................................... 2232
H1O.CM.91.MVP5180       ---AACA-GT......GTCC---T---------................................................................................................................................... 2207
CPZ.CD.90.ANT           ---AGGAGGAAGTAGTG-----------T-----CAT-----T--A......-CA--AG---CA........................................................................................................ 1637
CPZ.CM.01.SIVcpzCAM13   ---A---GA-......--------------------C------T-ATG---ACCAA-AGA-............................................................................................................ 1775
CPZ.CM.05.SIVcpzEK505   ---CA-C-A-AAGGAA---------------T---AT------T-ATG---A-CA................................................................................................................. 1746
CPZ.CM.05.SIVcpzMB66    ---AG-----......---A--------T------AT----------G---TCA---AG--............................................................................................................ 1729
CPZ.CM.98.CAM3          --G----GAA......G--A--C-----T--G--CCT------T-ATG---A-CA--A--CA---GG..................................................................................................... 1603
CPZ.GA.88.GAB2          -A--G-AGGT......G--A--------T-----TCC-----C----G---A--CA-AGC-----CA..................................................................................................... 1582
CPZ.CM.98.CAM5          --G--G-GAA......G--A--C-----T-----CATG-----T-ATG---T-CA--A-ACA----T..................................................................................................... 1906
CPZ.CM.05.SIVcpzMT145   -AG----GAA......G--A--------T-----TATG--ATC----G-A-T-CA................................................................................................................. 1731
CPZ.TZ.00.TAN1          ---A--CC--.........AGAG-A--GGAA......................................................................................................................................... 1815
CPZ.CM.05.SIVcpzLB7     ----------......------C--------G---AT--------TG---T-CA-CC-A-............................................................................................................. 1738
CPZ.US.85.CPZUS         -A-A---GAA......G----C-----T------AT----GAT---G---A-CAA-A-------GT...................................................................................................... 2264
CPZ.GA.88.GAB1          ---A----A-......---A-----G--------GAT------T-ATG---ACCA--AG--AGA........................................................................................................ 2262
MON.CM.99.L1            TC-C-TCTGTGCAGCCCAGCG--C-GA-T-TG----C..................................................................................................................................A 1854
MON.NG.x.NG1            TTCCAGT-A-CTCCTTGC--------------G-G---CCA-C..........................................................................................CCARGGGGCYTACAAGGTGAAC 462
GSN.CM.99.CN166         TA-CTTTC-TGCTGCC-ACAG-CC-G--GGA...G--G--AGAACC-CA-CA.................................................................................................................... 1878
GSN.CM.99.CN71          TA-CTTCC-TGCTGCC-TCAG-CC-G--GGA...G--G-T-GAACC-CA-CA.................................................................................................................... 1878
MUS_1.CM.01.1085        ---C-TCT-TCTTGCC-ACTG-TC-GC-AGA--AG--G-G-GAA--GGTCAA-CTCTCTCCAG--GT..................................................................................................... 1891
MUS_1.CM.01.CM1239      ---C-TCG-TTTTACC-AC-G-CC-GC-AGAG-AG--G-GAGA-C-GGTCAA-CTCTCCCC-C--GT..................................................................................................... 1881
MUS_2.CM.01.CM1246      TT-CATCG-TCATGCCTAC-G--C-CC-AGATA---C----CCGG-G-AC-TCTCCCC...ACT-GTAGGGCCCAAGGA......................................................................................... 1983
MUS_2.CM.01.CM2500      TT-CATCG-TCATGCC-TCAG-TC-CC-AGATA---C-----TAG---AC-TCTCCCC----GTGGGTCCACGGAAGGAGAA...................................................................................... 1948
RCM.NG.x.NG411          CTTC-TTGAGACCTAC..............TG-TCC--G...........................GAGAGAGGAAGGATGGAATCCCCAGTACGATCCAGCAGAAGAGATGCTCAGGAAATAC 1660
RCM.GA.x.GAB1           CTTC-TTGA-TCCCTCT-CT---GGGATGGAGAGCA-TT-CGA-CC..............................................................................TGCAGAAGAGATGCTAAAGAATTAT 1645
DEN.CD.x.CD1            -CCCTCAGATGAATCAG--TTTGC--TGGAGC----TGAG-GT................................................................CCAGCAGCAACCACAGTACCCACAAAATTTTCAAGAGGGGGGA 2035

MAC.US.x.239            T-GCTCA..................--TG-AT-AG-GGCT--TGCCA-CTG--CCCCC..................................................................................AGAGGACCCAGCTGTGGAT 2456
Pol                      H  G  .  .  .  .  .  .  S  S  A  S  G  A  D  A  N  C  S  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  P  R  G  P  S  C
Gag                     M  A  Q  .  .  .  .  .  .  V  H  Q  G  L  M  P  T  A  P  P  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  E  D  P  A  V  D
H2A.GW.x.ALI            TA-C-C-..................--TT--G-AG-G-TTA-CACCA-CAG-ACCCCC..................AGCAGAGCCAGCAGCGGACCTGCTGGAGCAGTATATGCAGCAGGGGAGAAAGCAAAGAGAGCAGAGAGAG 2544
H2A.DE.x.BEN            T--C-CA..................------T-AG-GGCT--TACCA-CAG-ACCTCC..................GGCAGATCCAGCAGCGGAACTGTTGGAGAGATATATGCAGCAAGGGAGAAAGCAGAGGGAGCAGAGGGAG 2551
H2A.SN.x.ST             T-GC-CA..................-AT---G-AG-GGCT--CACCA-CAG-ACCCCC..................GATAGACCCAGTAGAGGACCTACTAGAGAAGTACATGCAGCAAGGGAAAAGGCAGAGAGAGCAGAGAGAG 1996
H2B.GH.86.D205          T--C-CA..................--TG--T-AG-G--T--CACCATCTG-ACCCCC..............GATGAACCCAGCAGAGGGCATGACACCTCGGGGGGCGACACCATCTGCGCCCCCTGCAGATCCAGCAGTGGAG 2527
H2B.CI.x.EHO            TCCAGGC..................-C---A...G-GGATAGTGCCATCTG-GCCCCC..............GATGAACCCAGCATTCGGCATGACACCTCAGGGTGCGATTCCATCTGCACCCCCTGCAGATCCAGCAGAGGAG 2522
H2G.CI.92.ABT96         T--C-CA..................--T---T-AG-GTCT--CTCCATCAG-ACCACC..................................................................................GATGGACCCAGCAGTAGAC 1878
H2U.FR.96.12034         T-GC-CA..................--T---G-AG-G-CT--C-CCA-CAG-CCC.........................TCCAGCAGAACCAGCAGTGGATCTGCTGACCCCGACAGCCCCTCCAGCGGACCCAGCAGTGGAT 2033
MAC.US.x.EMBL_3         T-GCTCA..................--TG-AT-AG-GGCT--CGCCA-CTG--CCCCC..................................................................................AGAGGACCCAGCTGTGGAT 1942
SMM.US.x.H9             T-GC-CA..................GATG--T-AG-GGCT--CACCA-CTG--CCYCC..................................................................................AGAGGATCCAGCTGTGGAT 1941
SMM.US.x.PGM53          T-GC-CA..................GATG--T-AG-GGCT--CGCCA-CTG--CCCCC..................................................................................AGAGGATCCAGCTGTGGAT 2385
SMM.US.x.SIVsmH635F_L3  T-GC-CA..................GATG--T-AG-GGCT--CACCA-CTG--CCCCC..................................................................................AGAGGATCCAGCTGTGGAT 2471
SMM.SL.92.SL92B         T-CAG-C.......................-A--TC-TT--CACCGTCAG-ACCCCC..................................................................................AGATCCAGCAGCCCGGATA 1902
STM.US.x.STM            T-GC-CA..................-ATA--T-AG-GGCT--CGCCA-CTG-CCCTCC..................................................................................AGAAGACCCAGCTGCGGAT 2115
MNE.US.x.MNE027         T-GC-CA..................-ATG-AT-AG-GGCT--CACCA-CTG--CCCCC..................................................................................AGAGGACCCAGCTGTGGAT 1936
DRL.x.x.FAO             CCTC...........................................................................GCTGACCCCATCAGCTCCTCCATTGCCAGGATACCTCCAGGAGGATCCGGCCGAGAGGATGCTC 1730
MND_2.x.x.5440          ................T-CCATGC--TTGA-C---TC--CACC-CC..............................................................GATGCCAGGGTTGGAAGACCCGGCAGAAAAGATGCTGTTAGAC 1727
MND_2.CM.98.CM16        ................G-C-ATGC--TTGA-C---AC...-GC-CC...........................................................TCCAATGCCAGGGATGGAAGACCCAGCAGAGAGAATGCTGTTAGAT 2170
MND_2.GA.x.M14          ................T-CCATGC--CTGA-C---TC--CACC-CC..............................................................GATGCCAGGGATGGAGGATCCAGCAGAGAAGATGTTATTAGAC 2094
OLC.CI.97.97CI12        CAGAG-TGGACATCTTTCCCT-TG--C-TA-C-TG--................................................................................................................... 2048
GRV.ET.x.GRI_677        -C--AG--GACACAGTT-GT-TGGA-C-AA--G-CCCCCC..................................................AATGGAAACAGCTTACGATCCAGCAAAGAAGCTCCTC 2321
VER.KE.x.TYO1_patent    CT-CTCTTGGAGCGGA-CC-AGTG-GC-T--T---CCGAGCG--AC--CCC-..................................................................ATACGACCCAGCAAAGAAGCTCCTG 2381
VER.KE.x.9063           CC-CTCATGGAGTGGA-CC-AGCG-GC----T---CCGACACC...........................................................................ATACGACCCAGCAAAGAAGCTCCTG 2384
VER.KE.x.AGM155         CC-CTCTTGGGGCGGA-CCAAGTG-GC----T---CCGA-C-A.....................................................................CTCTACACCTTACGACCCAGCAAAGAAGCTCCTG 2376
VER.DE.x.AGM3           CT-CTCTTGGGGTGGAGCCAA-TG-GC----T---CCGAGTCC...........................................................................ATACGACCCTGCAAAGAAGCTCCTG 1879
TAN.UG.x.TAN1           .-CAAG-CGGAGCCGT......TC-GA-GG-T--CCCGAT-CCTGCGCACGGATTCCC.................................................................CACAGGTTCGCCAGCTGCGGGA 2350
SAB.SN.x.SAB1C          CCTC--TCAGACCAAC--CT--CC--ATGGA--GG--TT-C---AG........................GCCGGAGGAAAATTGGTATGCAGACAGACCTCCGACCAGAGGACCAGGGCCAGACGATCCAGCAACAGCCCTGTTAAAGCAGTAT 2570
MND_1.GA.x.MNDGB1       ..............................................................................GGCACAAGAGGTGACTCCAACAGCTCCACCATTAGAGGAGAAACCTCTGCAG 1844
DEB.CM.99.CM40          -AG-A-TTGGGAAGTACATC--GGAGAG-A--G----TC--GTA-A-TC-A-.................................................................................................... 1977
DEB.CM.99.CM5           -AG-G-TTGGGAAATACATC---GAGAG-A--G----TC--GTT-ATTCAA-.................................................................................................... 1971
TAL.CM.01.8023          G--A-TTC--AGTGGTGC-A--C--C--A--G--GAT.............................................................................................AGAGCCTCTCACAGCATCA 1627
TAL.CM.00.266           G--ATTTT--AGTGGTGC-A---T-G--A--G---AT.............................................................................................AGAGCCTCTCACAGCATCA 2132
SUN.GA.98.L14           TC-TGCCTT-TGCTCCGCC...........................................................TCTAGAGGATCTGACATTGGGGAACAGAATGACACCACCCCAGTCAAAGGCAGAG 2511
WRC.CI.97.97CI14        T---G--CTTGAACAATTTTGGGC--A-AA--A--ATGCCTTCTGCTCCTC-AAT-AG----GATC-AAACAT.....................................TCAGGAGCTGGAAAAATATTTATTGAAG 2469
WRC.CI.98.98CI04        T---G--CTTGAACAATTTTGGGC-G--...-A--AT-CC-TCTGCTCCTC-AAT-AGG---GATC-AAACAT.....................................ACAAGAATTAGAGAAATACCTCCTGAAA 2484
WRC.GM.05.Pbt_05GM_X02  ..............................T-A-A--AT--ATGCAG--GTGAATACCTTTGGGC-AGTGGT............GCCAACAGCACCAGTGGATCCTACCATGCAAGAACTAGAG 1398
LST.CD.88.524           TC-TGG--GGGGC...T-CAGG-GT...GATT--TTC--C-CCACC.......................................................CATGGAGAAAGTGCCTCCAACAAGAGCAGAGAGG 1422
LST.CD.88.447           T--TGG-GGGGAA...T-CAGG-GT...AGT----TC--CTCC-CC..........................................................AATGGAGAGGAGCCCAACACAAGCAGAGAGA 1425
LST.KE.x.lho7           T-G-AG--G-......T-CAGGG-G...AATT---TC--CTCC-CC...............................................................AATGGAGAGTGCTCCAACAAAAGCAGAG 2506
SYK.KE.x.SYK173         T---G--CGAGCCATCT-C.....................TCC-CCTCT-GA.....................GGACATCGAAGATGGGCCATGGCTCACATGGTCAGCACAGATGAGCCAACAAGCACAAGCGAAGGCACAGAAC 2339
SYK.KE.x.KE51           T--TG--GGAGACACC-ACAG-CC--C-AG-GGAG--.........................TTGGCCATGGCAGCAACAGACATGGGGAAATTATGCTTCACCCCAACAGAAGCACATAGCAATGGCCACCCCGCTACAGCCACAGTCC 1997
COL.CM.x.CGU1           ........................................................................................................................................................ 1964
```

```
                                                                                                                            Gag end
                                                                              Gag-Pol TF end Pol protease start        Gag p6 end
H1B.FR.83.HXB2           ..................................AACTCCCCCTCAGAAGCAGGA...GCCGATAGACAAGGA......ACTGTATCCTTTAACTTCCCTCAGGTCACTCTTTGGCAACGACCCCTCGTCACAATAAAGATAGGGGGGCAACTAAAGGAAGCTCTATTA 2324
Gag                       .  .  .  .  .  .  .  .  .  .  .  .  T  P  P  Q  K  Q  E  .  P  I  D  K  E  .  .  L  Y  P  L  T  S  L  R  S  L  F  G  N  D  P  S  S  Q  *
Pol                      .  .  .  .  .  .  .  .  .  .  .  .  N  S  P  S  E  A  G  .  A  D  R  Q  G  .  .  .  T  V  S  F  N  F  P  Q  V  T  L  W  Q  R  P  L  V  T  I  K  I  G  G  Q  L  K  E  A  L  L
H1A1.UG.85.U455          .................................G--CT-----GC---A-----.--T--A----G--.......--A-AC-------GT--------AA-------------------T------G---A-----A--A--G--G-TA----------- 1770
H1B.US.90.WEAU160        ..................................-------T-------------.....----A-------------.....G-------------------AA-----------------------------------A---------G--------------- 2323
H1C.ET.86.ETH2220        ...............GAGCTTCAGGTTCGAGGAAGC---A--TT---C---------.---T--A-----G--............AG-C---------------AA-----------------A--T--T--------A--------A--G--------G-----C--- 1716
H1D.CD.84.84ZR085        ...............................CCCT-C-AGAAA---G-A---A-...-GAC-A---------ACT............C---G---------AA------------------G-T--------------------A--G-----A------------ 1846
H1F1.BE.93.VI850         .......................................C---T---C---------.---A--A----GG---ACTGTA...CCC---C---G---------AA------------------T-A-----------A--------A--GA-------G--------- 1661
H1G.SE.93.SE6165         .......................................C---T-C-C----------...-AT--AG--A-----.............G-TA-ATC-C-------AA-------------G-------A------G---A--------A--G---TA-----C------ 1721
H1H.CF.90.056            .......................................C---T---C--------GCA--T--AG--------ACC.........---C---G----------A-------------G-------T--T---G---A----A---A--GT---G------------ 1671
H1J.SE.93.SE7887         .......................................C---T-C-C---A-------.-----AG--------ACT.............C-------------AA-------------G-------T-----------GA-----------G--G-G------------ 1638
H1K.CM.96.MP535          .......................................C---T---C--G--------...-A-C-A---------ACAGAG...CCC-----------------AA------------------A-T-----------AG--------A--GT---GA------------ 1520
H101_AE.TH.90.CM240      ........................GGGGGAAGAGAT---CT--TTA-C---------.---A--A---------ACATCC...TCC-------GT--------AA--------------------T------G---A-----A--A--G--G--A------------ 1898
H102_AG.NG.x.IBNG        ...................................-C-----T---CAC---------...A----GG-------GGCTATA...TCC-----------------AA------------------T-A--T---G---GA----A---A--G----TA-----C------ 1849
H1N.CM.95.YBF30          ....................GAAGAGCACACAGGGG-AGGAGATG---G--A-CC-GGA-AG--C---G--CTCTCTGTA...CCCA-----------------A------------------G--A-----G---A------AAAG--G---GA--------T--- 1918
H1O.BE.87.ANT70          ..................GATGGAGGAGGAAGTGAAGGGA-AGGAGA-TC--G-AC-GAAAGG-GGCCCG--C--GCT......C-----G--TG-C-------AA--C---------G-CA-----AA-A--T---GC--GAG-T------C--C---TGT----T-T-GC-G 2379
H1O.CM.91.MVP5180        ..................AATGGAGGAGGCAGTGAAGGAA-AAGAGA-TC--AGTC-GAA-GG-GATC-GG-A--GCT.........---C--A--TG-C-------AA--C---------G-CA-----AA-A-------GC-----G-T------T--T---TGT--G---T--C-G 2354
CPZ.CD.90.ANT            .......................................CAAGAGGACTC--A---GTCTCAA-GGG--GG----ACT......ACC---C-CGTA------G-AA--C------------A------AA-GA-GGA-G-TCTC---CA---A---AA-TGTC-G---------- 1763
CPZ.CM.01.SIVcpzCAM13    .......................................GGAGAAGA-CC-AG----ACA-GAAGAG-GG--CAGACTATA...CCC----C-----------AA------------G--------A--C--G---GA----AA------T--TGT-----C--G--- 1904
CPZ.CM.05.SIVcpzEK505    ..................GGAGGAGAAGACCACACAGGGGA-AGAGAG-G-AG-AA-GGA-AA--C---G-GCTCTCTGTA...CCCA--C-------------AA------------G------A--C-T---G---A-----A--AG--A----A--------TC-- 1896
CPZ.CM.05.SIVcpzMB66     ...........................GGAACCTGCCCAGAAGGAGG-AGT-A--AGCGA-GA--C---G--CA-GCTGTA...TCC--AGC------------AA--T------------------G-T------G---GA----A---A--GT----A------A--G--- 1870
CPZ.CM.98.CAM3           .......................................AAAGAGA---G--G----GAG-GA--AG--A-----ACCCCTGTATCC---G-----C-------AA-------------G---------C--GG-G----A----AA-----G----TA---------GC-- 1735
CPZ.GA.88.GAB2           ..............................GACAGG--AGAGT-AGA--GG-G----GAGTA-A-C---A-----GAG......CCC--TC-ATC----------AA-------------G---------C--C--G---GAG---CA--A------TGT------C---C-- 1717
CPZ.CM.98.CAM5           ................................................-G--A-AAA-AA-GAG--G-----ACCCCTGTATCC--AA-----C-------AA-------------G---------C-TGG-G----AG---AA-----GT---TA---------TC-G 2026
CPZ.CM.05.SIVcpzMT145    ................................GGAGGAGAAGAAAGTC---G----GGT-AG-GAG--G-GCA-GCCGCC...--C-----------------AA------------------------A--C--G----A--T-AA--A---T---TT-----CT-GC-- 1866
CPZ.TZ.00.TAN1           ..................GGCTCAGGAGCAGGAGAC-G-A-AGA-A-CAGT-GT-CCAACTG-CCC-CCACT---GATGACGA--A-AGGCGGGTTC-------A----A---------G------AA-GA-G---G-----AG-CCA---A---G-CTGTC--------T--- 1968
CPZ.CM.05.SIVcpzLB7      ...........................AACCACCCAG-AA-AGGAGG-C-G-G-A--GGACAA--C---G--CA-GCTGTA...CCC---C-------------AA--T----------------------A-T------------AG-C-----CT-CT-G--A-G---A--G--- 1879
CPZ.US.85.CPZUS          .............................................GA-A---G--A-AC-GGGAAA-GA-A-GG-...GCC......C-T--AGC------------AA-------------G---------C---G------AG---CA-----GA--GTA---------T--- 2384
CPZ.GA.88.GAB1           .......................................GAAGAGC---G--A-...GAAAGA-GG---G-GCAGTCT......A-----ACC---C--------AA-------------G---------A--C--G----AG---AA---------TGT-------T-GC-- 2385
MON.CM.99.L1             GGGACTGCGGAGCCAGTCAATCTCAACCAAGAGATTGG-T----..........-A-GAC-GG--CCCCA-----GACGAGG-G-G---T--ATC-C---T-AGCA--T--------GG-A-----AA-AAGAT-TG-GCT----------A---AAGGT-TCCATG-----G 2015
MON.NG.x.NG1             TTCTCACCTCCAGTGGGACCACCAGAGACCCC.....................A-CGCC-A-AGCCCCG-----GACGAGGAA-G---T--ATC-C---T-AGCA--C--------GG------AG--AGA---G--G-A--C-CT-----GTCCGT-TCCATGT----G  611
GSN.CM.99.CN166          ....................GCTCAATCTCTCGAC-C-AGTAGGGG-ACCC---AGCAA-GGACAGA-GG----GAGCAGGAACAC--TC-ATC-C-----GTCA--T--------GG-A-----AA-GAAGGA-G-CTCC----------A---G-GGT-TCCATGT------ 2028
GSN.CM.99.CN71           ....................GCTCAACCTGTCTCCGC-GGTGGGGG--CCC---A-CAA-GA-CAGA-GGCA--AAAGAGGAACAC--T--ATC-C-----GTCA--T--------GG-A-----AA-GAGAGA-G-CTCC--G------A---G-GGT-TCCATGT------ 2028
MUS_1.CM.01.1085         ..............................................--GTCCCGG-GAG-GG--AC-GGACAGGAA-AG--TC-ATC-------GTCA--C------CAGG------AG--AGAC--G--G-------AA-----GG-GGT-TC-ATG--GC-C 2008
MUS_1.CM.01.CM1239       ..............................................--GTCCCAA-GAT-GAGC-AGCGACAGGAA-AG--T--ATC-------GTCA--T------CAGG-------A-TAGAGA-G---T-----AA--A---G-GGT-C-CATG--GC-C 1998
MUS_2.CM.01.CM1246       ............................................A--GCCA--C-AG----G-A-AGACAG-AA-AC--TC-ATC-------GTCA--T--------GG-------AA--AAA---G--G-C--TAAT-----G-C-GT-TCCATG--CC-G 2100
MUS_2.CM.01.CM2500       ............................................CC-AACC--CCA--C-CCACCGGACAGGAA-A---T--ATC-------GTCA--C--------GG-------A--AGA---G---T-G-CCAT--T--G-CGGT-TCCATG--CC-G 2065
RCM.NG.x.NG411           CTAGCATTGGGGAGACAGCACAAACAGGAGCAGAGG-AAGA............-A-CAA-GA--A--TGGG-AGAGCCTACGA-G--G-G--GT-C-------AT--T--------A-G-----TG-AACATGGG--G-A-----A--------AAGGT---G--CT------ 1818
RCM.GA.x.GAB1            CTGAGGAGGGCAGGGGAACAAAAGAGACAACAGAGGCAGGA............A--GAGCAA--AGAGAG---GAGCATAT-A-G-AG-C-----C-------AT--G--------A-G-----AACTACA-T-GCTC-A----AA-----GAA-GT---G-TC------- 1803
DEN.CD.x.CD1             AACAGACAATCTTGTCAGGAAGACAAGAGGCAGAAC-AG-AGT-.........CC-TCA-GA-GAC--GG--ACAACGAGGGA-AG--T--ATC----GT-G-A---T--------AG------AA-A-GGCAG--CTG-G-TCA---A---AGGGTA--G-TC--TA-- 2196

                                                                                                                  Gag end
                                                                                  Pol Protease start        Gag p6 end
MAC.US.x.239             CTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAG-GAGAAAAG----GAG-AAGCAGAGA--AGCCTT-CA-GGAGGTGACAG-GGA---GCTGCA-----AT--T---------AGGA----AG-A-----TGCTC-T--T-AA--A--G-CTGTA----TAT--C-G 2626
Pol                      G  S  A  K  E  L  H  A  V  G  Q  A  A  E  R  K  A  E  R  K  Q  R  E  A  L  Q  G  G  D  R  G  F  A  A  P  Q  F  S  L  W  R  R  P  V  V  T  A  H  I  E  G  Q  P  V  E  V  L
Gag                       L  L  K  N  Y  M  Q  L  G  K  Q  Q  R  E  K  Q  R  E  S  R  E  K  P  Y  K  E  V  T  E  D  L  L  H  L  N  S  L  F  G  G  D  Q  *
H2A.GW.x.ALI             AGGCCATACAAAGAGGTGACGGAGGACTTACTGCACCTCGA............-C-GGGAGA--C-CCAC-CA-AGAGGTGACAG-GGAC--GCTGCA-----AT--T---------A--A----AG-A------GCCT-C--T-A---C--G-C-GT-----T-T--C-- 2702
H2A.DE.x.BEN             AGACCATACAAAGAGGTGACGGAGGACTTGCTGCACCTCGA............-C-GAGAGA--C-CCTC-CAGAGAGGAGACAG-GGAC--GCTGCA-----AT--T---------A--A----AG-A------GCGT-C--C-A--AT--G-CGGTA----TCT--C-- 2709
H2A.SN.x.ST              AGGCCATACAAAGAAGTGACAGAGGACTTCCTGCAGCTCGA............-A-ACAAGA--C-CCATGCAGAGAGACGACAG-GGAC--GCTGCA-----AT--T---------A--A----AG-A------GC-C-TG-T-A---C--G-C-GTA----T-T-GC-- 2154
H2B.GH.86.D205           ATGCTGAAAAGTTACATGCAGATGGGGAGACAACAG-GAGA............-AGCCGAGA--G-CC-T-CA-GGAGGTGACAG-GGA---GCTGCA-----AT--T---------AG-A----AG-A----A-GC-TGT--C-A---T--GTC-GTA----TAT--C-- 2685
H2B.CI.x.EHO             ATGCTAAAGAACTACATGCAACTAGGGAAGAAGCAG-AGGA............-A-CAGAGA--G-CC-T-CA-GGAGGTGACAG-GGA---GCTGCA-----AT--T---------AG-A----AG-A----A-GC--CT--T-A---T---TC-GTA----TAT--C-- 2680
H2G.CI.92.ABT96          CTACTGAAGAATTACATGCARCTGGGGAGGAAGCA--AGGA............-C-GAGAAAC-A-CC-T-CA-GGAGGTGACAG-RGAH--GCTGCA------C--T---------AG------AG-A-----TGC-T-C----AAA-T--G-C-GTA----TA------ 2036
H2U.FR.96.12034          CTGCTAAAGAGCTACATGCAACAGGGCAAGAAACA--AGGA............-A-CAGAGA--G-CCAT-CA-GGAGGTGAC-G-GGA---GCTGCA-----AT--T---------A-------AA-A-----TGC-T-C----A---------C-GTAA---TA---C-G 2191
MAC.US.x.EMBL_3          CTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAG-GAGA............AAGCAGAGA--AGCCTT-CA-GGAGGTGACAG-GGA---GCTGCA-----AT--T---------AGGA----AG-A-----TGCTC-T--T-AA--A--G-CTGTA----TAT----G 2100
SMM.US.x.H9              CTACTGAAGAATTACATGAAGGTGGGCAGAAGGCAG-GGGA............-A-CAGAGA--G-CCTT-CA-GGAGGT-AC-G-GGA---GCTGCA-----AT--T---------AG-A----AG-A-----TGCCT-C--T-AA-AA--G-CC-TA----TAT----- 2099
SMM.US.x.PGM53           CTACTGAGGAATTACATGAAGATGGGCAGAAAGCAG-GGGA............-A-CAGAGA--G-CCTT-CA-GGAGGTGAC-G-GGA---GCTGCA-----AT--T---------AGGA----AG-A---G-TGCCT-C--T-AA-AA--G-CCGTA----TAT----- 2543
SMM.US.x.SIVsmH635F_L3   CTACTGAAGAATTACATGAAGATGGGCAGAAAGCAG-GGGA............-A-CAGAGA--G-CCTT-CA-GGAGGTGACAG-GGA---GCTGCA-----AT--T---------AG-A----AG-A-----TGCCT-C--T-AA-AA--G-CCGTA----TAT----- 2629
SMM.SL.92.SL92B          GTCAA...GGAGTATCTAGAGAAAGCACAGAGGGA--AGA-............AAGGAG-AGC-GGCCTT-CA-GGAGGTGAC-G-GGA---GCTGCA-----AT--T-----C--AG-A----AG-A----G-GC-CTA---A---T--G-CCCTA--G-T----C-G 2057
STM.US.x.STM             CTGCTGAGAAGTTACATGCAGCTGGGCAAGAAGCAG-GAGA............-AGCAG-AA--C-CC-T-CA-GGAGGTGACAG-GGA---GCTGCA-----AT--T---------AG-A----AG-A------GCCC-T--T-A---T--G-CTGTA----TA-----G 2273
MNE.US.x.MNE027          CTGCTAAAAAACTACATGCAGTTGGGCAAACAGCAG-GAGA............AAGCAAAAG--AGCCTT-CA-GGAGGTGGCAG-GGA---GCTGCA-----AT--T---------AG-A----AG-A-----TGCTC-T--T-A---A--G-CTGCA----TA-----G 2094
DRL.x.x.FAO              CAAAAGTACATGGAGCAGGGAGCTCAGCAGAAGAG-CAA-AGAG.........AC-ACAA-A-CAGA-G-G--GCCCCTACGA-G-GG---AC-AC--------T--T-----C---TCA----AACTACA-T-G--G--G---AA------AAGGTA--G---T--C-- 1891
MND_2.x.x.5440           TATATGAAGAAAGGGCAACAACAGAGGGCAGCAGC-GGAG-............CC-GAAAGA-GA-A-G----GTCCCTACGA-GCAG---AC-AC--------T--T--------A-CA----AACTACAGT-G--G-A----A-------AAGGTA--G--A--TC-- 1885
MND_2.CM.98.CM16         TACATGAAGAAAGGGCAACAGCAGAAAGCGGAAAGC-A...............AC-GGAAAA--A---G-G--GTCCATACGA-GCAG---AC-AC--------T--T--------A-CA----AACTACAGT-G--G------AA------AA-GT---G--CT--C-G 2325
MND_2.GA.x.M14           TATATGAGAAAGGGGCAACAACAGAGGGCAGCAAAGGAGA-............CA-GCA-GA--AG-------GTCCCTACGA-GCAG---AC-AC--------T--T--------A-CA----AACTACAGT-G--G-A----A---A---AAGGTA-----C---C-G 2252
OLC.CI.97.97CI12         ...........ACAGATTTTGGAGAAGGGAACAAG..................---AAA-AGAGAGATG-----GGAGAAGAAACC--T--ATC-C-----AGCA--T--------AG-A----AA-GA-AGAGC-GT-T----AA--A--G-A-GTCACT--A--G--- 2189
GRV.ET.x.GRI_677         CAGCAGTATGCAGAGAAGGGACAGCGCCTGAGAGAGGAGAG............A--ACA-A-A-GGA-AC--A-GGAGAA-GAAGTGGAGGATGT----T-G--C--C---------AGGA----AA-GAAACG-G-C-TC----AA--AACG-C-GT-C----CT-G--- 2479
VER.KE.x.TYO1_patent     CAGCAATATGCAGAGAAAGGGAAACAACTGAGGGAGCAAAAGAG.........-A-TCCA---GC-ATG--TCCGGATTGGACCG-GGGA-ATT---T...G-AC--C---------AG-A----AA-AAAG---G-GT-T----AA---GTC-CC-TTA-G--A--GC-- 2539
VER.KE.x.9063            CAGCAGTATGCAGAGAAGGGAAAACAGTTGAGGGA-CAAGGGAA.........-AGAACA--ACC-ACA--TCCAGACTGGGC-G-AGGA-ATT---T...G-AC--C---------AG-A----AA-AAAA---G-C-TT--T-AA--AGTC-CTGTAA---TC--G--- 2542
VER.KE.x.AGM155          CAGCAGTATGCAGAGAAAGGGAAACAAATGAGAAATCAGAA-AG.........AA-CCCC--AGCGA----TCCAGATTGGAACG-GGGA-ATT---T...G-AC--C---------AG-A----AA-AAGG--CTGT-TA-----A--AACTGCCGTTA-G--AT----- 2534
VER.DE.x.AGM3            CAGCAATATGCAGACAAGGGGAAGCAGTTGAGGGA-CAAAGGAA.........AA-ACCA--AGC--TG--TCCCGATTGGACAG-GGGA-ATT---T...G-AC--C---------AG-A----AA-AAAA---G-TT-C----AA---GTC-CC-TCAG---AT----- 2037
TAN.UG.x.TAN1            GCATACGACCCAGCCAGGAAACTGCTGGAGCAGTATG-CAAAAAGGG......A--TCAATT--G-A-GC--A-GGAGAAGGAACTGGAGGACTA----T-G--T--C---------AGGA----AA-GAGT-AG--CT-C---CA---C--G-C-GTAC---T---G--- 2514
SAB.SN.x.SAB1C           GCTGTTCAGGGGAAACGGCAGAAACAGCAGTGGCA..................AA-CCACT--CC-C-AC--AGCCCCTACGA-G-GG---AC-GC---------T--T---------AG-A----AA-GAAA--TG-GT-C----A---A--GAA-GTCACT--------- 2722
MND_1.GA.x.MNDGB1        AAAACTCTGAGCACTTATCAGAAATTAGGGAGAGGGCT...............CAGGCA-AA---GA-GG----GAAGAG-GA-G-GGA---TCA------G--TA-T-------CAAG-A----AA-A-AAGAGG-CTCAG-G-AT--TGTCACT-TAAG--------C-- 1999
DEB.CM.99.CM40           ...........TACCAGGGTACCAGTACAGGGGCAGGAGAA-AAC-CCC--AG---GGAAA-A-CGAGG-GA-GGAAGTA...CCCGGA-----TGGGG--G-AA--T---------AG-A----AA-GAC----G--C-C-----AA--AAG-AA-GTCC-G-TG---A-- 2133
DEB.CM.99.CM5            ...........AACCAGGGTACCTGTGCAAGGGGA--GGAATG---CCC--G----CACAA---CCAGG-GA-GGAAGTA...CCCAGA-----TGGGG--G-AA--T---------AG-A----AA-GAC-G--G--T-TG-G-AA--AA--AGGGTCC-G-TG---A-- 2127
TAL.CM.01.8023           TGGCAGGAGGGGACTACCACATGGACCCCTCCGACC-AGGA............T-CCGCAG-ACCCA-GC-AC-GACGAGGGA-AG--T--ATC-C--GT-AGCCAGT---------AG------AA-A--A-AGG-GCTC----AA--AAGGGAGGT---G-TA--CC-- 1785
TAL.CM.00.266            TGGCAGGAGGGGACTACCACGTGGACCCCACCAGCC-AGAA...............GCCC-AAGA-ACA-CCACAACAAGGGA-AG--TC-ATC-C--TT-GGCCAGT---------AG------AA-A----A-G-GCT-----AA--A---GAGGT---G-TA--CC-- 2287
SUN.GA.98.L14            AGAGCTCTGGAGACCTACAGACTGTTAGGACAGGGCCTCAG............A-CACA--A--AGAGG-----AAGGGGGGAA-G--AGGAGC-A-G-T-G-ATATG------CCAG-A-----T-A-AG-AGG-CTTC--TAAT-------CTGT-ACT---T----G 2669
WRC.CI.97.97CI14         GGAGCACAATTGAGGGAACAACAGGAACACAAGAA..................-ACAGA-GGACAGA-G-G---AGTGGGGG--CCGGAA-GGC-C-------AA--T-----C-AACCA----AA-GA-ACA-TGC-TA----AA--A--TAGTGTCTC---AT--C-- 2621
WRC.CI.98.98CI04         GGAGCTCAATTGAAGGAGCAGGCACAGGCAAA.....................AA-GGAAGA--AGA-G-GA--AACGGGG---CCAGAG-GGC---------AA--T-----C-AA-CA----AA-GA-ACA-TGT-TA----AA--AA-TAGTGTCTC---CT-G--- 2633
WRC.GM.05.Pbt_05GM_X02   CAATATCTTCTGAGAGGAGCTCAGATGAAGAGGGA..................-A-AGA-AAAGA-ATGC--A-GGAGGGGAAAGTAGAA-ATC-C-------AA--T-----C-AACCA----AA-GA-ACA-TGT-TC--T--TC-AG-GAGTGTCTCT--GT----- 1550
LST.CD.88.524            GCAATAGAGACATACAGGAACCTAGGACAACAACTG-GGAA...............GCAA-AACA--TTCCAC-GAAGTGTG--G-CGAGCCTTGCCTGAATTTC-TCT-----CCAG------AA-A-AG-A-G--CT-G-CAAT------GA-GTAACT--C-----G 1577
LST.CD.88.447            GCCTTAGAAACATACAGGAACCTCGGACAACAGCTG-GGAG...............GCAA-A-CAGAT-CCAC-GAAATGTG--G-CGAGCC-TGCCTGAGTTTC-TCT-----CCAG------AA-A-AA-AC--TCT-G-CAAT-----G-C-GT-TCT-----TC-T 1580
LST.KE.x.lho7            AGAGCACTGGAGACTTACAGGACTTTAGGACAACAGCTGAAAAG.........-C-GCA--A-CA--T-CCAC-GAAATGTG--G-CGAGCCCTGCCTGAATTCACTCT-...-CCAG------AA-A-AA-GT---TT-G-CAAT--------C-GT-TC---A--G--G 2664
SYK.KE.x.SYK173          TCTCCCTCGAAGAAACCCCCCACCAACAGGGAAGTGCTCT-G--.........-A-GGA-AGC-GT-GG-----GGAGACCAA-AG--TC-ATC-C--T-----C--T---------AG-A----AA-GA-AGA-G--G-TG---AA--TG-C--GGT-C--ATGT--G-- 2500
SYK.KE.x.KE51            TCTCCAAAGACAGTCAAGGACCTCTTGGTCCCTGGG-AA--TT-.........A--GAA-AAAGA---G-T-A-GGAGAAGGGACC--T--ATC-C-----TCAA--T---------AG------AA-GA-GGA------TG---AA---ACCAAGGTAAC-ATGT-GC-G 2158
COL.CM.x.CGU1            ................................GAAGATG-GCAAAGA.........CCTGCCA--C-GGAGGG--A-GGGGGAGAG...--TC-ATC-C---T-G-A---T-----C--AG------AA-A--T-T-G--G-AG---AA--A---AAGTGCC----G--G--- 2093
```

```
H1B.FR.83.HXB2              GATACAGGAGCAGATGATACAGTATTAGAAGAAATGAGTTTG............CCAGGAAGATGGAAACCAAAAATGATAGGGGGAATTGGAGGTTTTATCAAAGTAAGACAGTATGATCAGATACTCATAGAAATCTGTGGACATAAAGCTATA............... 2468
Pol                          D  T  G  A  D  D  T  V  L  E  E  M  S  L  .  .  .  .  P  G  R  W  K  P  K  M  I  G  G  I  G  G  F  I  K  V  R  Q  Y  D  Q  I  L  I  E  I  C  G  H  K  A  I  .  .  .  .  .
H1A1.UG.85.U455             -----------------------C--------C--A-A----............-------A---------------A--------------------------------------------------T--------T------A-A--GA-----............... 1914
H1B.US.90.WEAU160           --------------------------------C----A----............-----G-A------------------------------------------------------------G---C-----------------------------............... 2467
H1C.ET.86.ETH2220           --C--------------------------------A-A----............-------A---------------------A-----------------T--------------------A---A----------T------A-A--G------............... 1860
H1D.CD.84.84ZR085           --------------------------------C--A-A----............-------AG-----------------------------------------------------------C--------------------------G------............... 1990
H1F1.BE.93.VI850            --------------------------------C--A-A----............-------A-----------------------------------------------A----------A-C--------------T--------C--G------............... 1805
H1G.SE.93.SE6165            --C--------------C-----------------A-A---A..............................................................A--------AG---CT--------TA-----A-A--G------............... 1865
H1H.CF.90.056               --------------------------------G--A-A----............--G----A--------------------------------------------------------G--AG--GC-----------------A-A--G------............... 1815
H1J.SE.93.SE7887            --------------------------------C--AGAC---............---C---A------------------------------------------------------A-CG--G---CG---------TGAG---A-A--G------............... 1782
H1K.CM.96.MP535             -----------------------------------A-A----............-------A---------------------------------------------------A--------AG----T--------T--------A--G------............... 1664
H101_AE.TH.90.CM240         --------------------------------T--A-A----............-------A------------------------------------------G----AG--A--------------T---------------A-A--G------............... 2042
H102_AG.NG.x.IBNG           --C-----------------------------C--A-A---A............-------A---------------------------------------------------A--------A-----T--------T------A-A--G--C---............... 1993
H1N.CM.95.YBF30             -----------T------------A-------GC-ACAA--A............GAG----A---------------------A-----------A-----------G-----A------A-T---ACAG----C--ACAG---AGA-----AG-T............... 2062
H1O.BE.87.ANT70             -----------------C-----C--A-CA-C---ACAA---............GA-----A------------------------T--A-----------A-------A-G-A------A-TG-GACAG-------AGAA---AGGG-G-TACAG............... 2523
H1O.CM.91.MVP5180           --------G------------------A-TA-C--ACAA--A............GA-----------C------------------T--A-----C-----A-------A-G-----A-CA-TG-GACAG-----G-ACAA---A-GG---TACAG............... 2498
CPZ.CD.90.ANT               -----------T-----C-----G-----G-G---TCA----............-A-----AC-----G-------CA--T--T--------G--------TTCC---CA-------A-CA-AG---C----C----TG-A-ACAGG-C--TAC--............... 1907
CPZ.CM.01.SIVcpzCAM13       --C-----G--T------------A--C----C--TCAA--A............-AG----TT--------------------------------C-----T-----G-A---A------A-TG--TCA-----G--AGAG--GA-A----TGCA-............... 2048
CPZ.CM.05.SIVcpzEK505       -----------T------------A----------ACAG---............GA-----A---------------------------------G-------------A----------A-TG-GA-A-----G--ACAA--GA-A-----AG-T............... 2040
CPZ.CM.05.SIVcpzMB66        -----------------C-----GC----------AGAA---............GG--T---------------------------------------C-A------------A------A-TG-TACA-----C--A------A-A-G---AG--............... 2014
CPZ.CM.98.CAM3              -----------T------------A----TA-TG-ACAAC-A............A------A-----G---------------------------------------T-A---A----CA-C---GCA---------TGAG---AGA---A-A-C-............... 1879
CPZ.GA.88.GAB2              --------G--T-----------------TA-T--ACAAC-A............GA--G-A------G---------T-A---A--C--------A-----T-------A---A------A-TG--GAA--------AGAG---A-A-G--TAGA-............... 1861
CPZ.CM.98.CAM5              --------G--T-----C------A----TA-TG-ACAA---............A-----AG-----G-----G---------------------------T-----T-A---A-----A-C---G-A---------AGAG--CAGA---A-A-C-............... 2170
CPZ.CM.05.SIVcpzMT145       --C-----G--T-----C------A-T-----CT-A-A----............-------A-----GT--------------------------------T-------A---A-T---AAGAG--AA---------AGAA---A-A--GATA-CC............... 2010
CPZ.TZ.00.TAN1              -----T-----------C-GT--T--TTGTA-C--C-AA--A............AAG---CAG-------C-----CC-----A-----A-----A---G-ACC---T--TG---CT--A-T--T-CAG--C----TG-CAATA-AG---TC-G---............... 2112
CPZ.CM.05.SIVcpzLB7         --------G--------------GC-------TT-AGAA--A............GGG----G-----------G---------A-------------------------AG--A-----A-T----CA---------AG-A---A-A--G--AC--............... 2023
CPZ.US.85.CPZUS             --C--------T-----------C-----CA-C--ACAAA-T............GA--G-C------G------------G--------------------A-------A-----------CG-CAAT---------AGAG---AGA-----ACAG............... 2528
CPZ.GA.88.GAB1              -----------T------------A----GAG---ACAA--A.............A---CTT---------------------A-----------------------C-A---A-T---A-TG---A-------G--AGAA--GAGA----TAG--............... 2529
MON.CM.99.L1                --------T--C--C--C---A-TA--C----TCA-CACG-A.....AATTTAGGG-AG-AT---C---T--G-AG-------------A--G--AA-G--AG----T-A---A--CA-AGCA----AAG-GAT-T-TGAA-ACAGAG-G-T-TGG............... 2165
MON.NG.x.NG1                --C--G--R--G--C-----CA-CG-CC----RCAAGACA-A.....AATCTAGG--A--AT---C---C-----G-------------A-----C--A--AC-G--T-AG--A-TCA-AGG-G--ACGG--TCCT--CAG--TA-A-C--TG-G-...............  761
GSN.CM.99.CN166             --C-----G--G--C--C--CA-CG-CC----CTCAGCAA-T.....GAATTAGATAC-CC------C----------G----------A--G-----A--AC-------GG--C--A-A--TG--GAAG--ATCT-TAA--AGA-A-G-ATC-A-............... 2178
GSN.CM.99.CN71              --C-----G--G--C-----CA-TG-CC-----TCAGCAA-T.....GAATTAGACAC-CC------C--------C-G-------G--A--G-----A--AC-----AGG-AC--AGA--TG--GAAG-GATCT-TAA--A-A-A-G-ATC-A--............... 2178
MUS_1.CM.01.1085            --C--C-----------C--CA--A-CC-G--TGCCCA-A-A.....CACCTAGA---CC-----------T--C--TCAG--------A--G--AAAC-ATC----C-A-G-A---AGGGGAG---AGG-TTGTT--AA----A-A-CTAT--GG............... 2158
MUS_1.CM.01.CM1239          --C-----G--------C--CA--G-CC-G--TTCC-ACA-T.....CAGATAGATAA-CC----------T--C--T-G-G-------C--G--AAAC-ATC----C---G-A---AGAGG-G---AAG--TCTT--AA--AGA-A-C-AT--AG............... 2148
MUS_2.CM.01.CM1246          --C--C-----G-----C--CA-CA-CC----CTCAGA-A-A.....AAAATTAGT--GGA--T--CT--C---G-AG-G--A--TT-A--G--GC-G-----T---A---A-A---A-GGG-G-GA-AG--A--T--AA--AGA-A-G-ATAG--............... 2250
MUS_2.CM.01.CM2500          --C-----G--------C--CA-TA-TC-G--GTCTGAAA-A.....GAAATAGATAA-GA------CT------G-G---------T-A--G--A--A-T--T---AG--A-A---A-AGCAG--AAAG---GT-TAA--ACA-A-G-ATC----............... 2215
RCM.NG.x.NG411              -----------------G-----TA--C--A-TT-AGA---A............GG--A--AT-----G--------T-----A-----A--G--A-A---A--T--T-A---A-T-A-C--TTGTAAA---ACT--AGCA--GA-G-CTA-ACAT............... 1962
RCM.GA.x.GAB1               --C--------------C------A-T----G---AGAA--A............GG-AATGAT----CC--------A-----A-----A--G--A-A---T--T---A---A---A-AA-CTGTGAA--T-----AGC-----A-A-GGA--CAT............... 1947
DEN.CD.x.CD1                --------G--------C--CA-TA---C----CAAGA-A-A...AATTTGGGAA---AGG------TC-------CAG----------A-----A-A---T--T--TCAGGCA--CCCAGG----GAA--TA-GC-TATG-ACA-AGT---GG--............... 2349

MAC.US.x.239                --------G--T------T-TA-TG--AC--G---AGAG--A............GGTCC-CAT-AT-CC--------AG----A-----A-----------T--TACT-A-G-A--CA-AA-TG--GAA------G-T-TA--CA-A-GGAT--A-............... 2770
Pol                          L  D  T  G  A  D  D  S  I  V  T  G  I  E  .  .  .  .  L  G  P  H  Y  T  P  K  I  V  G  G  I  G  G  F  I  N  T  K  E  Y  K  N  V  E  I  E  V  L  G  K  R  I  .  .  .  .  .
H2A.GW.x.ALI                --C-----G--T--C--CT--A--G---C--G---AGAG--A............GGGA-C-AC-AT-CC--------AG----------A--G--A--C--A--TACC-A-G-A-----AG-TG--GAA---A--G-ACTAAATA-A-G--TA-A-............... 2846
H2A.DE.x.BEN                --C-----G--T-----CT--A--G---C--G---AGAA--A............GGG-AC-AT-AC-CT--------AG----------A--G--A-----A--CACC-A-G-A--CA-AA-TG--GAA---A--G-ACTAAATA-A-G--TA-G-............... 2853
H2A.SN.x.ST                 --C-----G--T--C--CT--A--G---C--GCG-AGAG--A............GGGA-C-AT-AT-GT-----G--AG----------A--G--A--C--A--TACC-A-G-A---A-AA-TG--GAA---AG-G-A-TAAATA-A-G--TA-G-............... 2298
H2B.GH.86.D205              --C-------TT--C--CT--A--G---C--GG--AGAA--A............GGTA-C-AT-AC-CC--------AG----A--G--A-----G--C--A--TACC-A-G-A--CA-AG-TG--GAA------G-AGTG---A-A-G--TA-GG............... 2829
H2B.CI.x.EHO                --C--------T-----CT--A--G---C--GG--AGAA--A............GGCA-C-AT-AC-CC--------AG----T--G--A-----A-----A--TACC-ATG-A--CA-AA-TG--GAA------G-AGTA---A-A-G--TA-G-............... 2824
H2G.CI.92.ABT96             --C-----G--T-----CT-TA--G---C--G---AGAA--A............GG-A-T-AT-AT-GT--------AG----A-----A-----A-----T--TACC-A-G-A--CA-AA-TG--GAA--TA--G-GGTA---A-A-G-ATACAG............... 2180
H2U.FR.96.12034             A----C--G--T------T--A--G---C--GG--T-AA--A............GA-CC-TAT-AT-CT-------CTG------A---A-A---G--C--T--TAC--A-A-A-T-C-CA-TG--AA----A--G-A-TG--CA-A-C-AT--G-............... 2335
MAC.US.x.EMBL_3             --------G--T------T-TA-TG--AC--G---AGAG--A............GGTCC-CAT-AT-CC--------AG----A-----A-----------T--TACT-A-G-A--CA-AA-TG--GAA------G-T-TA--CA-A-GGAT--A-............... 2244
SMM.US.x.H9                 --------G--T------T--A-TG---C--GG--AGAA---............GGTCC--AT-AT-CC--T-----AG----A-----A-----C--C--T--CACC-A-G-A---A-AG-TG--AAA---A--G-A-TA--CA-AGT-AT--AG............... 2243
SMM.US.x.PGM53              --------G--T------T--A-TG-G-C--GG--AGAAC-A............GGTCC--AT-AT-CC--T-----AG----A-----A-----C--C--T--CACC-A-G-----A-AG-TG--AAA---A--G---TA--CA-GGT-AT--AG............... 2687
SMM.US.x.SIVsmH635F_L3      --------G--T------T--A-TG--AC--GG--AGAA---............GGTCC--AT-AT-CC--T-----AG----A-----A-----C--C--T--TACC-A-G-A---A-AG-TG-GAAA---A--G---TA--CA-GGT-AT--AG............... 2773
SMM.SL.92.SL92B             --C-----G-----C--CT-CA--G---C--GG--AGAG---............GGTCC-G-G-AC-C-----G---TG-------G--------A--C--A--TAC--A-G-A---AGAACAG--AAA------G-T-TA---A-GGT-ATA-G-............... 2201
STM.US.x.STM                --C-----G--T--C--CT--A--G-G-CG-GG--AGAGC-A............GGGCT-CA--AC-CC-----GG-AG----T-----A-----A-----A--TACT-AGG-A-T-A-GA-TG--AA------GG-A-TA--CA-G---ATA-A-............... 2417
MNE.US.x.MNE027             --------G--T------T-TA-TG---C--G---AGAG--A............GGTCC-CAT-AT-CC--------AG----A-----A-----------T--TACT-A-G-A--CA-AA-TG--GAA------G-T-TA--CA-A-GGAT--A-............... 2238
DRL.x.x.FAO                 --C---------------------A-TC---G-T--GAA--A............A------A-----G---C-G--T------A-----------AAC---T--T--G-----A---TT-A-TTGTGAA--TA-G--AGCA--GA-A-TTA--CAT............... 2035
MND_2.x.x.5440              -----T--------C---------A-TA----TT-AGAG--A............A------AT-----G---CG---T-----A-----------AGCA----G--------------TT-A-TTGTAAAG------AGCA---A-A-TTA--CAT............... 2029
MND_2.CM.98.CM16            --C--------T------------A-CA----TC-AGA---A............AA---T-AT--------C-G--T--T---A-----------A-CA----T----A------T-TTCA-TTGTAAAG--AC---AGCA--CA-A-CTA-ACAC............... 2469
MND_2.GA.x.M14              --C--T------------------A-CA----TG-AGA---A............A------AT--------CC----T-----A-----------AGCA--T-GG--G-A------TTCA-TTGTAAA---AC-G--AGCA---A-A-GTA--CAT............... 2396
OLC.CI.97.97CI12            --------G--------C------A-TA-G-GG-A-GA-A-C......TACATGGA---T-ATG---TT--T-G---T-----------A--G--AGC---T-GG----TG--A---AGATGT-GGA-A-G-ATTT-T...-AGA-AG--ATACA--............... 2336
GRV.ET.x.GRI_677            -----------------C--TA--A-TC------A-GAC---CACTTTCCCCCA-ATAA-CC----CGTT-C--GG-AG----A--T--A-----AGGG-TC-T---C-A-G-A---C-GGG-G---AAG--C--T-GGAG-ATA-A-TCATC-CC............... 2635
VER.KE.x.TYO1_patent        --C-----G--------C--CA--A-TA------ATGA---A.....CAATTAT-----TCC------G--C-----T--------C--A-----AGGCC-T--T---A-G-A---A-CG-C-GGGAAG--A-----AGAA-ATA-A-TTTTG-G-............... 2689
VER.KE.x.9063               --------G--------C--CA-CA--A------CAGACA-A.....CAACTAA-----CCT------G------G-A-----A-----------AGGAC-T--T---A-G-----AG-A-TG--GAAG--A-----AGAA-ATA-A-TTTTG-G-............... 2692
VER.KE.x.AGM155             --------G--------C--TA--A--A-G--T-CAGA---A.....CAATTAAGG---TC-------G--------AG----A-------G---AGGGT-A--C----A-G-A------A-TG--GAAG--C--T-GGAA-ACA-G-T-TTA-G-............... 2684
VER.DE.x.AGM3               -----G--G-----------CA-TA--A-----GCAGA---A.....CAATTAT-------C---------------------A-------C---GGGAC----T----A-G----AG-G-T-GGGAAG---AG-T-GGAA-ACA-A-TTTTG-G-............... 2187
TAN.UG.x.TAN1               -----------T--------CA--A--A-G---GAAGA-A-C.....CATTTAGA----C-C---CGG--T----G-AG----A--T--A-----GGGA---TC-----A---A-----CA-T--TT-AG-GA-T--AGA----A-AG--A---G-............... 2664
SAB.SN.x.SAB1C              --C--------------CT-----A--C-G-GT--AGAA--A............GG--AC-AT----------G---A--T--A-----A-----A-G---T--T----A-GCA--CC-CA-TCAGGAGG--A----AGAA-ATA-A-CTTG--A-............... 2866
MND_1.GA.x.MNDGB1           -----------T--------CA-C--TA-T----GA-A-A-A.....AAATTAAA-----AT---C-G-------T-------------A--T--AAACT-A-G-----A----------A-TG--TATG-------AA-A--GA-GGG-A-AT-T............... 2149
DEB.CM.99.CM40              --------G-----------CA-TA-TTC-----AAGA-A-A.....GATTTAGAG-C-CCT----CC-------CAG---------T-A--------C--A--T--C-A-TGT--CCCAGG----GAA--T-C---GGC-AATA-GGT---AG--............... 2283
DEB.CM.99.CM5               --C-----T------------A-TA-TTC---G-AAGA---A.....GATTTAGAT-C-CCT----CC--G----CAG-G-------T-A--G--A--------T----A-TGC--CCCAGG----GAG--TA----GGCA-ACA-GGT----G-T............... 2277
TAL.CM.01.8023              --------G--------C---A-CA--C-----GG-GACA-A.....GAATTAGA----GTGCCCC-C-A-----CAG--------T-G-C---A-AC---C-G---CA-AGC---AC-GGAG--GAA--C---TGGGAACACA-A-G------GG............... 1935
TAL.CM.00.266               --C--------------C---A-TA--C---T-GGAGAAA-A.....GAATTGGA----GT-CCCC-C-A-----CAG--------T-G-C---A-AC--AC----GCA-AGT---AG-GCC---GAA---C-GTGGGAAAACA-A-G----A-GG............... 2437
SUN.GA.98.L14               --C--G-----------C---A--G-GTCT---GATTC-G--......TCCATAGA---TCAC-----G--C-GGG-AG-G--T--T--A--G--ACAA--T-G-----A-G-A----CAG-TG-CT-TG-------TA-G-ATA-A-C-----C-............... 2819
WRC.CI.97.97CI14            ------------------------A----T-T--ATTTA...............AATAAG-AT---C-G--T-G---A--T--A--G--A-----CAC---G--T----A-TTAC--AGA--AG--AAA---AG---A-TA---A-A-C-ATAGA-............... 2762
WRC.CI.98.98CI04            ------------------------A----TATCCATTTAAAT............AATAAG-AT---C----T-G---A--T--A-----A-----CAC---G--T--T-A-TTA---AGGA-A----AA---AG---T-TA--GA-A-C--TAGA-............... 2777
WRC.GM.05.Pbt_05GM_X02      ------------------C--T---G----T-TTCAATTA..............GA----AT---C-------G---A-----A--T--A-----AAC-GC-----A--TA--A---A-AG-TG-CAGA---A----T-TA---A-GG---TA-G-............... 1691
LST.CD.88.524               --C--T---------------A-C--TTCT----ATTCAG-T.....AGAATAGA----CCT-AT-CT-----G-G-AG-------T--A--G--ACAA--T-------A-G-A-TCAGGG-TG-TT-T---C----AGCA--GA-G-CCA-AG--............... 1727
LST.CD.88.447               --C--T--G-----------TA-C--TTC-----ATTCAG-T.....AGAATAGAG---CCT-AT-C----T-G-G-AG----------A--G--ACAA--T-G-----A-G-A--CA-GG-TG-CT-T--------AGCA---A-A-C-A-AG--............... 1730
LST.KE.x.lho7               --C--T--G--T-----C---A-T--TTC-----ATTCAG-A.....AGAATAGA----CCC-AT-CT-----G-G-AG-------C--A--G--GCAA--T-GG----A-G-A---AGAG-TG-CT-T--------AGCA---A-A-CTA-AG--............... 2814
SYK.KE.x.SYK173             --------G-----------TA--A-TAG----GAAGA-A-T.....CAATTA-ACCA-CC-----TC---------AG----A---T-A-----AAAC--T-C---T--G--A---AGAA-T--TAGAT-CACTG-TGT-AA--CCTC--GA-A-AGAAAGCAAGTAGA 2664
SYK.KE.x.KE51               --C-----G--------C--CA--A--AG----CAAGACA-C.....CAGTTA-ATCA-CCT----CC---------AG--------T-A-----AAAC--A-GG--------A---A-C-----CAGAT-TC----AGAAA---CAG-T-GG-G-ATAAAGGAAGTAGA 2322
COL.CM.x.CGU1               -----------------------T--TA-----GAAGA-G-T.....CAATTAA----CGCT----C-G--GT-CAA--TCA-------A--G--AGCA----G-----A-------AGAG-T-A-A-TC--A-G--GG-A---A-AG--TAC-G-............... 2243
```

```
                                                      Pol Protease end_Pol p66 and p51 RT start
H1B.FR.83.HXB2          .GGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACTCAGATTGGTTGCACTTTAAATTTTCCCATTAGCCCTATTGAGACTGTA.........CCAGTAAAATTAAAGCCAGGAATGGATGGCCCAAAAGTTAAACAATGGCCATTGA 2628
Pol                      G  T  V  L  V  G  P  T  P  V  N  I  I  G  R  N  L  L  T  Q  I  G  C  T  L  N  F  P  I  S  P  I  E  T  V  .  .  .  P  V  K  L  K  P  G  M  D  G  P  K  V  K  Q  W  P  L
H1A1.UG.85.U455         .----------G----------------------------G--A------------------T-------------A-----T--------A-------.........-------------A---A-------------G----------------- 2074
H1B.US.90.WEAU160       .------------------------------------------------------C----------------------T--------A-------.........----------------------------------------------------- 2627
H1C.ET.86.ETH2220       .---------C-----------------------------C-----CA----------C----AC----A-----C-----A-----T--C-----A-------.........------------------------------C---------------- 2020
H1D.CD.84.84ZR085       .-------------------C-------------------------T-------C--------C-----------C-----A-----T--------A-------.........-----------------------------------------G---- 2150
H1F1.BE.93.VI850        .--------G-------------G--------------------------A-------------------T-------------AG----T--------A-------.........----------G--------------G----------------- 1965
H1G.SE.93.SE6165        .--G--GA------------------A-------------G-----A----------------------------------A-----T--------A-------.........----------------------------GG----------------- 2025
H1H.CF.90.056           .-------------------------------T----------G--A-A--------A----------C----------A-----T--------A-------.........-----------------------------G----------------- 1975
H1J.SE.93.SE7887        .--------G--A----------------------------G--CA-----------C-------T-------------A-----T--------A-------.........------------------------------A---------------- 1942
H1K.CM.96.MP535         .---------------------------------------C----C--------------------T-------------A-----T--------A-------.........-----------------------------G----------------A- 1824
H101_AE.TH.90.CM240     .---------------------------------------C-----A-------------------T----------C--A-----T--------C-------.........-------C-------------------A-----G--------G---------- 2202
H102_AG.NG.x.IBNG       .---------------------------------------C-----A-------------------T-------------A-----T--------A-----G.........-----------------------------G------------------ 2153
H1N.CM.95.YBF30         .-------------------A--------T--T--T--A-----------T--A--C--------C--T-------------A--A--T--------A-------.........-------------A--------------------G--A----------T---- 2222
H1O.BE.87.ANT70         .--A--------G--G--------T-----T--T--T------------A-A-----AGGAT-A-----T--AC-----C--C--T--A-----C--A-CCC-A--G.........----------C----A---------------A---------A-----------CC-AT 2683
H1O.CM.91.MVP5180       .--A--------G--G--------T-----T--T--TC----G-----CA-A-----AGGAT-A--A--T--AC--------C--T--A--T--C--A-CCC-A--G.........--------GC----A---------------A---------A-----------CC-AT 2658
CPZ.CD.90.ANT           .-CA------C-GT-G-----A-AC--A--A-----T--A--T------G-T--ATG--TTT-A--A--T--------------A-----TAAAG----A--A--G.........--C--T--------AGA------------A-----G---A----------GC-CT 2067
CPZ.CM.01.SIVcpzCAM13   .--G-----G--G--G-----A-----------T--------G------A-T--A--A-----A--G-----AC-GGTA--C--T--C--TT-A-----A-----C.........-----C-----G--A-----------C--G-----G-A-A--G----------T--AT 2208
CPZ.CM.05.SIVcpzEK505   .--C-----C--------G--A--------T--T--T-------------T-TC----A--A--------T---C----------A--A--T--------A--CA--.........-----G-----G--------T----------G-G---A-----G------T---- 2200
CPZ.CM.05.SIVcpzMB66    .--------T-----------C--------A--T--C--A-----G--A-T--------A------A---------------C-----C--T--------A-------.........--------GT---------------------G-----GG--C-----G--------A- 2174
CPZ.CM.98.CAM3          .--C--G-----G--G-----A-----A--G--T--------C------A-A-----A--A-----C--T--------------T--A--T--C-----A--A--G.........-------------------------------T-G---A-----------T--A- 2039
CPZ.GA.88.GAB2          .A-C------C----G-----A--T--A--T--T--T--A--------G---A--C--AT-A--A--T---C---------A--A--T--A---ACTC----G.........-----------------------CA-----A----GG--A--G-----------T 2021
CPZ.CM.98.CAM5          .--C-----------------A-----A--A--T--------C------A-A--A--A--A--------T---C----------T--A--T--CG-------A--G.........------------------------------G---A-----------T--A- 2330
CPZ.CM.05.SIVcpzMT145   .A----T--------------A--C--------T-----A--------G-CC-A--AA-AT-A--G--T---C--------C--A--A--T--A--A--C--A--.........AA---GTTTC----------G-----C--G--C--G--A-----G-----TC--T 2170
CPZ.TZ.00.TAN1          .-CC--T--CC----G---GAA--C--CA-T--T-----A--T------A-T--A-AG--AT-A--A--T--C------------T---------A--A---GTA--.........AA----C--------AGA-------------G--------A--G--G-----CC-CT 2272
CPZ.CM.05.SIVcpzLB7     .--------T-----------C-----A--A--T--C--A-----G--CA-T--A------A-----A--T---C------------T---------------A-------.........-------C-------------------------G--------A------------C--A- 2183
CPZ.US.85.CPZUS         .---T----T-----G-----A-----A--A--T--T-----C--G---A-T--A--A--A-----A--T--------------T--C---T-----A-A-------.........------G-----A-------------------T-G---A--G-------------- 2688
CPZ.GA.88.GAB1          .--G-----------------A--------A--T-----------------A-T-----A--AT-G-----T---C--GTG--C--A-----TT-A-----A-----C.........------C-----G--A---------------T-----G--A--G--------C--AT 2689
MON.CM.99.L1            .-CA-----------T-----A--T---A-A--T--TT-A--G------G-C--AG--A----G--GGTA--CC-C--CA-GG-TGC-G-GGGAGACCTCCAACCCCCATCTTTAGAG--C-CT-----AG------AA--G--A-----GT-G-----G--------AT 2334
MON.NG.x.NG1            .-CW-----T--GC-G-----R--C--AA-T--T--TT-G--G-----CG-CC-CT-AAG---G--GATA--CC-G--CA-GY--CAGG-AATCGCCC-AT-CCC-...TCCTTGA----G-CCC-T--A-------AA-----G------T-G-----------CC-CT  927
GSN.CM.99.CN166         .-C---T---C-G--T-----C-----A--A--T---T-A---------TGC--A-G-A--T-A--GGTA--------CA-GGTGCAGCAGAAG--A--AC-CA--.........GA---GC-CC-----G------AA--G--A-----GA-C-----G-----CC--T 2338
GSN.CM.99.CN71          .-CC--T---C-T-----G--C-----A--G--T---T-G---------TGT--A-GCA--T-A--AGTA--C-----CA-GGTGCAGCAGAAAC-A---C-C--.........GA---GC-CC---AT------AA--A--A-----GA--------------C---T 2338
MUS_1.CM.01.1085        .-CA---------------G--A--------A-A--T-----G--G------CTGT--A-GCA-AT----CATT---C----CA-GATACAGG-AAAAT-A--AC-AA--.........AA---G-CCC----AGA---GCAA-A--G----TG--C--G---------T---- 2318
MUS_1.CM.01.CM1239      .-CA----G--G--G-----------CA-----------G------TGT--A-G-A-AT----AAT---C--G--CA-GATACAGGAAAAA--A--AC-CA--.........AA---GGCC------GA---GGCAA-A--A----TG--C-----G------C--- 2308
MUS_2.CM.01.CM1246      .-CC-----C--G--GA-T---------CA----T--CT-A--G------TGC--AT-AA-AT-G-AAATA--AC----CA-GGTA--GGCAGGACAGC-ATTGAA-...CCCACTAA---T-GCC-T--AGA-A---AAA-A--G----TG--A-------------C-CT 2416
MUS_2.CM.01.CM2500      .-C------TC-----------------AA----T--TT-A-------CTGC--AT-AA-AT-G---ATT--C--G--CA-GGTAGGAG---AAC-G--AC--ACG.........AA---TGG-C-C--AG----G-AA--A--G----TG--A---------------T 2375
RCM.NG.x.NG411          .-C-CAT-----G---------A-----A--A--T--------C--G---G-------GCA-AT-A--AGC-T--C---------A--A--TAAAGCA--A--AA--.........AAG---G--------------CAA-----G--T-G---A-----------T---- 2122
RCM.GA.x.GAB1           .-C-CAT--GC----G-----A-----A--G--T--T--A--G------G-T--A-AGA-AT-A--AGC---AC---------A--A---AA--A--A---A--.........AAG---G--------T-T---CAA-----A---CG---A---------------T 2107
DEN.CD.x.CD1            .-AC----------------------CA-A-----CT-A--------CTATC-T-G-A--T-A--AGTA--AC----CA-GG-AG-ACAG-AA--GCG-C--CA-...GTCACC-------------A------T-T--------T-G-A-A------------T--A- 2515
                                                      Pol Protease end_Pol p66 and p51 RT start
MAC.US.x.239            .--G---A-CA-GAC---GGAC--C--GA-T-----TT----T------T--C-A--AGCTC-G--GATGT--C-------------AGCTAAAG-A---C-----------.........AA---CGCC-------------A------A------T-G--G--G--------AT 2930
Pol                      K  G  T  I  M  T  G  D  T  P  I  N  I  F  G  R  N  L  L  T  A  L  G  M  S  L  N  F  P  I  A  K  V  E  P  .  .  .  V  K  V  A  L  K  P  G  K  D  G  P  K  L  K  Q  W  P
H2A.GW.x.ALI            .-CC--CA--A-GAC---TGAC--C--AA---T--TT---C-----CA-T-----AGCCT-A--CATGT-A-----CC-A--AG--GC-AAG--A---C-AA--.........GAG----G-------------AA--C--G------T-A-G----------C--A- 3006
H2A.DE.x.BEN            .-CC--CA--A-GAC----GA---C--AA-------CT----C------A-TC----AGCCT-A--CATGT-A--------A--AG--GC-AAG--A---C-AA--.........AA-----C---G---------G-AA-----A----GGC-G-----------CC-A- 3013
H2A.SN.x.ST             .-CC--CA--A-GAC---TGA---C--AA-------TT----C-----CA-TC----AGCCT-A--CATGT-A------C-A--AG-CGC-AAG--A--AC-AA--.........AA-A---TGC-G-----------A------A------C-G-G----------C--A- 2458
H2B.GH.86.D205          .-CA--TA--A-GAC----GA---C--AA-A-----TT----C------A-T--A-A-ACCT-G--CATG-----------C--AG-GGCAAAGG-A--AC-A---.........AA---TG-G-----A--T----AA-----G-----GA-C-G---------TC-AT 2989
H2B.CI.x.EHO            .-CA------A-GAC---GGAC--C--AA-A-----TT----C------A-T--A-A-AGCT-A--CATG---C-------C--AG-AGCAAGG--A--AC-A---.........AA---CC-G---------T-A--AA-----G------A-C-G---------CC-AT 2984
H2G.CI.92.ABT96         .-C---T--TA-GAC---GGA---C---A-A--T--CT-------G--CA-T--AGT-A--C-A--CATGT--C-----------A---GCAAAGG-A---C-A---.........AA-----GGC----A--T--C---------G------A-C-GG---------CC-AT 2340
H2U.FR.96.12034         .-----CA--A-GAC----GAC--C--AA----T--TT----T--G-----T-----AG-AC-G--CATGT--C----C-----A--AG-AAAAGAA-CACTA--C.........ATT-GGGG-C----A-----C-A------A------T-A-G---G-----T--AT 2495
MAC.US.x.EMBL_3         .A-G---A-CA-GAC---GGAC--T--GA-T-----TT----T--G---T--C-A--AGCTC-G--GATGT--C-----C-------AGCTAAGG-A---C-----------.........AA---CGCC-------------A--T---A------T-G--G--G---------AT 2404
SMM.US.x.H9             .--A---A-TA-GAC----GA---C--AA-T-----TT----C---------C-A--AGCT--G--CATGT--C-----C-C-----AGCTAAGG-G--RC--A---.........AA-----C-C----A-------AA-----A------T-A-G---G-----GC-AT 2403
SMM.US.x.PGM53          .--A---A-TA-GAC----GA---C--AA-T-----TT----C------------A--AGCT--G--CATGT-C------C-C---G-AGCTAAGG-G---C--A---.........AA-----C-C----A-------AA--C--A------T-G-G---G-----GC-AT 2847
SMM.US.x.SIVsmH635F_L3  .--A---A-TA-GACG---GA---C--AA-T--T--TT----C------T--C-A--AGCT--G--CATGT-----------C-----AGCTAAGG-G---C--A---.........AA-----C-C----A-----G-AA--A--A---------G-G---G-----GC-AT 2933
SMM.SL.92.SL92B         .-AA---C--A-GAC----GA---C--GA-A--T--TT----C------A-TC-AG-AGCAT-A--GGT-T---------AC--AG-GGCAAAGG-A--ATA-AC---.........AA------G--G---GAG------------G----G-T-A-------------C--AT 2361
STM.US.x.STM            .--A--TA-TA-GACT---GAC--C--CA-T-----TT----C--G--C--A-----AGCTT-G--GATGT-A-----------AG-AGCTAAAG-A--AC-A---.........AAG--G-C-------A-------AA-----A-------A-A---------------AT 2577
MNE.US.x.MNE027         .--G---A-CA-GAC---GGAC--C--GA-T-----TT----T------T--C-A--AGCTC-G--AATGT--C--------C-----AGCTAAGG-A---C-----------.........AA---C-CC--------------AA-----A------T-G-GG--G---------AT 2398
DRL.x.x.FAO             .-C-T-G--G--G-----C-----C-----A--T--T--A--T------G-AC-CTG-A-AT-A--AGCT---C-T--------TG----T-AAG-A--AC-A---.........AA-----C-C----A---------------A-----GA-A------------CC-CT 2195
MND_2.x.x.5440          .-C-G---TC-G--G--C-----C-----A--T--T--A--T------G-AC-T-AGA--T-A--A--T---------------T-----TAAG--A-----A---.........AA-----C-C----A-------CA-----A--G-G-A-A-----G-------C--T 2189
MND_2.CM.98.CM16        .-C-T----CC----G--C--C--------A--T--TG-A--T------G--C---AGA-AT-A--A--T--AC------------AG----TAAAG-A--A--A---.........AAG----C-C----A-------CT--C--A--T---A-A---------------T 2629
MND_2.GA.x.M14          .-C-T----GC----G--C-----C-----A--T--T--A--T------G-TC-T-AGA--T-A--A--T---C-G---------T-----TAAG--A-----A---.........AA-----C-C----A---------------A-----G-A-------G-------C--T 2556
OLC.CI.97.97CI12        .-AGG----------G---GA------A--T--TG-GG-A-----G---G---AG-AGCAG-A--GATACG-C-GGTC-GCG-ACAGCTTT-AGA-A---TCCCT...ATTACTAAG-----------AGAG---CACA-GCCT---C-GA----G---------T--A- 2502
GRV.ET.x.GRI_677        .--CT--A-TC--A-----AG------AA---T--T--A----------A-T--AG-----GCA--CATG-AA---GT-A-GGGAG--CTAT---G-C---T--AG...GAAACAAA----C--C----AGA---G-AA-----A--T---T-G-------------C--AT 2801
VER.KE.x.TYO1_patent    .--A---A----GT-----G-A--T--CA-T--T-----A--T------T--C--G-C-C-GCA---GC-CGG---GTAA-GGGACAATTAT-AGAAA-A-T-CCT...GTCACA--T--C-----G---GA---GGCTCGG--A--CTGT--A-G----------TC-CT 2855
VER.KE.x.9063           .--C-----TC-TT-----G-A--C---A----T---------------T-AC-AG-A-CAGCA--AGCT-GAC--GTGA-GGGACAATTGT-AGAAC-A-T-CCC...ATCACC-----C---C-G--AGAT---GCAAGG--A---C-TA-A-----------TC-AT 2858
VER.KE.x.AGM155         .--A-----CC-CA-----G-A--T--CA----T--C--A---------CT-T--AG-C---GCA--AGC--AA---GTGA-GGGGCAATTGT-GCAGACA-TACC-...ATCACC--G---CGC-------GA---GGCCAGA--A---CG-T-G--G----------C-CT 2850
VER.DE.x.AGM3           .--G--CA----GA-----AGC--T--CA-A------------------A-A--AG-A-CAGCA--AGC--AA---GTAA-GGGTCAACTGT-AGAAC-A-T-CCC...ATTACC--T--G--------AGA---GGCTAGA--A--TTTCT-A-----------CC-CT 2353
TAN.UG.x.TAN1           .----AT---C--A-T-----A-----AA-A--T--TG-A---------A-TC-AG-A-CTT-A--AGCA-GA----TAA-GGGACAGCTAAA-GAGA-A-TACCC...ATAACCAAG---C----G--A--T--G---TCA--A--G---T-A-----------CC-AT 2830
SAB.SN.x.SAB1C          .-C---CA-T--G------GAA-----A--T--T--T--A---------G-C--AG-A--AT-A--AGTT---C-------GA-ACAA--AGAG--A--AC--A--.........AA---CC-C-------------CAA-----G----GGA-A-GG---------T---T 3026
MND_1.GA.x.MNDGB1       .--G--------GA----------T--AA-AG-T-----A--G-----CA-AA--GAAA-AT-A--AG-A-AA----TA--GG-ACAATTGT--GA-A-A-TACC-...ATAACAAA---G---------A------G-A-----A--C-G-A-A-----------T--A- 2315
DEB.CM.99.CM40          .TCAGATC-----------GA---C--AA-T--T--TC-A---------TAT--AG-AA-A--G--GGTAT--A-----------AG-ATTAGA-TA-ACCC-A---.................--------AGAG--GGCA-----A-----GA--------------C-TT 2437
DEB.CM.99.CM5           .TCAGAT--G---------TGA---C--AA--------T-G-----G---TAT---G-AA----G--GGTAT-CA-----------AG-ATTGGACT--ACAC-A---................--G--G---GAG---GCT----------GGA-------G-----CC-TT 2431
TAL.CM.01.8023          .-AC---------A----G---------AA--------T-A---------T-CC-AG-CA-AT----GGTA--AC-T---A-GGT--GC-CAGAA--ACCAGTAA--.........AA------G--G---GA----A---ACC------GA-C--G---------C---- 2095
TAL.CM.00.266           .-AC-----T--GA-T-----A--G--AA-T-----TC-A--G------T-T--AG-AA-AT--A--GTT--AC-T--CA-GGT--GCTCAGAC--ACC--TAACC.........AAG--T---------AGAG--T-A---ACC---C---A-C--G---------C--A- 2597
SUN.GA.98.L14           .--A-----G--G--G--G--A-----A--AG-T--C--A--G------A-TC-TT-AGCCT-A--GG-A-GA---GT---AG---C-CTTAG-GAAA---TCCC-...ATTACAAA---T--G-----AGA---GGCC-C---A---------G-----------CC--T 2985
WRC.CI.97.97CI14        .-CAGACA-TA-GAC---GAAA------A-A--T--T--A-----G---G-A--A-A-A-A--G--AATGCA----GCC---G-ACAATTAT-AGACA---T-CCT...ATAACAAA---G-GC--G--A------TGCTCA--A--T-G---A-----G-----C---T 2928
WRC.CI.98.98CI04        .-CAGATA-TA-GAC----AAA--T--AA-T--T--T--A--T------G-A--AGAAA-A--G---ATGCA---GGCC---G-ACAATTAT-AGA-A---TACCC...ATAACAAA-----GT--G---------TGCACA--A----G---G--G--G-----T--AT 2943
WRC.GM.05.Pbt_05GM_X02  .-CA-ATT--A-GAC----AAA------A-A--T-----A---------G--C-AGAAA-AT-G--AATGCA-C--GCA---G-ACAATTAT-AGA-A-A-T-CC-...ATAACAAA---CTC------A--T---TGTC-A--A----G-A-A--G---------TC-CT 1857
LST.CD.88.524           .--G-----T--GC-------A--C--A--AG-T---G-A--------CA-T--A-AC-CTC-A--GGCT-GA---GTG--AG-ACAATTAAG-GAGA-A-TACCT...ATAACAAAG--GC-T--G---------GTGT-----G----G---A-----G-------C-AT 1893
LST.CD.88.447           .--G----------T-G-----------A--AG----TG-A---------A-T--A---A-TG-G--AGCA-GA---GTAC-AT-ACAATTAAG-GACA-A-TACCC...ATCACAAAG--C-----G--------CTGT-----T--C-G---A-----G-----T---T 1896
LST.KE.x.lho7           .--G-----C---T----------C--A--AG-T--TG-A--G--G---A-T--A---GCA--G--GGC--AA--G-T---AG-TCAGTTAAG-GA-A-A-T-CCT...ATTACAAAG---TCT-----A--T--GTGT-----A----G---A--G--G-----TC-AT 2980
SYK.KE.x.SYK173         G--G--TC-T-----G--T-----G--A--A--T---T-A-----G---A-AC-T--CA-AT-G--AGT--AA---GTAA-GGTTCAA-CAG-AT-A---C-A--C.........AA-----GTC-G--A----AC-AA--ATTG--CCGGC-A---------------AT 2825
SYK.KE.x.KE51           G----GTC-TC----------------A--A--T--TT-A--------CA-A----GCA-AT-G--AGCA-AA---GTAA-GATAGCACAAAAAT-A--AC-AA-G.........AA---CC--C----------G-A---GCTA--T---T-A-----G-----TC--- 2483
COL.CM.x.CGU1           .--AGATA-C---A------AC--C--CA-A--T---G-A--T------A----A-AAG-AGGGAAA-TAGTAG--GC-CAATTATCAGATAGG--CCCAGTCACC.........AAG--G-CG-----AGAG--C--------G--------A--G--G-----C--A- 2403
```

```
H1B.FR.83.HXB2          CAGAAGAAAAAATAAAAGCATTAGTAGAAATTTGTACAGAGATGGAAAAGGAAGGGAAAATTTCAAAAATTGGGCCTGAAAATCCATACAATACTCCAGTATTTGCCATAAAGAAAAAAGACAGTACTAAATGGAGAAAATTAGTAGATTTCAGAGAACTTAATAAGAGA 2798
Pol                     T__E__E__K__I__K__A__L__V__E__I__C__T__E__M__E__K__E__G__K__I__S__K__I__G__P__E__N__P__Y__N__T__P__V__F__A__I__K__K__K__D__S__T__K__W__R__K__L__V__D__F__R__E__L__N__K__R_
H1A1.UG.85.U455         -------G---------------AC-----------AT----------------A------------------------------------------------T----------G-----C----------G--------------------C-----A--- 2244
H1B.US.90.WEAU160       -------G--------------------------------A--------------A---------------------------------T----------------------------------------------------------------------- 2797
H1C.ET.86.ETH2220       ----------G------------AC--C--------GA---A-----GC-------A---------G--------------C-----T---------------------A--G--G-----------G--------------------G-----C-----A--- 2190
H1D.CD.84.84ZR085       -----------------------AC---------------T--------------A---------G---------------------------A-----------------------------G------------------------------------- 2320
H1F1.BE.93.VI850        ----------G------------AC------A---CT---A--------A-----A----------------------------------------------------------G--------G---------A-----------A--------------A--- 2135
H1G.SE.93.SE6165        -----------------------AC-----------A---A------G-------A---------------------------C-----A--------------------G---------------------G-----------G--C-----A--- 2195
H1H.CF.90.056           -----------------------ACG-------------------A-----A----C----G---A---------G---------GC------A-------------A--G--G--T--------------------G---------------C-----A--- 2145
H1J.SE.93.SE7887        -----------------------AC-C--------G----A------G-------A---------G-G-------------------T--C------------------------------------------------T--------C-----A--- 2112
H1K.CM.96.MP535         -----------------------AC---------------A--------A-----A----------------------------G---------T--C--------G--------------------------------------T--------C-----A--- 1994
H101_AE.TH.90.CM240     -----------------------AC-----------A-----------G------A----C---------------------------------------------T----------G-----C--C--------G--------------------G--C-----A--- 2372
H102_AG.NG.x.IBNG       -----------------------AC---C-----------------------------A----------------A--------------------------------------------G--T-------------------------------------C--------- 2323
H1N.CM.95.YBF30         --AC----------G-G------AG---------------A--------------A--------T-G---A--------G--------T--C------A-T-----T-----A--G----T--C----------------------------G---T-A-----A--G 2392
H1O.BE.87.ANT70         -TA-----------G----C--GAC--C---A---CAG--A------C-A-----A----------G---A--A----------T--T-----A--TA-C-----T-----A--G-----TG------------------G--------T--G---T-A--------- 2853
H1O.CM.91.MVP5180       -TAG---G------G-----C--ACT-C---A---CA---A------C-------A-----C----G---A--A----------T--T-----A--TA-T-----T-----A--G-----T--C-----G----------G-----C----------T-A-----A--- 2828
CPZ.CD.90.ANT           --A----G------G-G--CC--AA----------GATA--T-A---GCA---AAT--G-----T-G---A-----A--T--C-----------A---A-T-----A-----A--G-------C-T-A-----------GC----T-------------T-A-----A--- 2237
CPZ.CM.01.SIVcpzCAM13   ---C----------T-----C--GAC-------A---CAG--A------------------G--A--T-----A--A--------C--------------TA-T-----A--T--A--G-----T---T-A--------------------------------T-A-----A--- 2378
CPZ.CM.05.SIVcpzEK505   ----------------G---C--AC-----------------A--------A-----A--------C-G---A------------------T------------A-T-----T-----A--G-----T--C------------------------------GT-A--C--A--- 2370
CPZ.CM.05.SIVcpzMB66    ---------------G-------AC-----------------A--------------A---------G-G-A-------------------------C---A-T-----T-----A--G-----T--C------------------------------GT-A-----A--- 2344
CPZ.CM.98.CAM3          ---C----------T--------AC----------CAG--A--------A-----A--G-------G---A-----C-----------T-----A---A-T-----T-----A--G-----T-----A------------C-G----------G---T-A-----A--- 2209
CPZ.GA.88.GAB2          --A-------G---G-G------AC-----------GG--A--------A-----A----------G---A--T--------------T--------CA-------A----G---G-----------A--G----------------------T------T-G-----A--- 2191
CPZ.CM.98.CAM5          ---C----------C--G--C---AC---------CCAG--A-----------G--A------------G--A--C--A--G--C--T--T--C--A---A-T-----T-----A--G----------A------------C-G-------T------T-A--C--A--- 2500
CPZ.CM.05.SIVcpzMT145   --A-----------G-G------ACT-----A----A---A--------------A-----------------A-----A-------T---------A--TA-T-----A--T---------G-------G-----C--C----------C-T--T---------------------------G 2340
CPZ.TZ.00.TAN1          -CA-G--G-----TG-G------AC------A----AGACAT-------------A--------TGC-G----A--A-----C-----T--C--A---A-T---------T-------G--T-CCT-----------------------------------G-----A--- 2442
CPZ.CM.05.SIVcpzLB7     ----G---------T-----T--AC------C---------A------C------A----------G---A-----------C-----T--------T--G-----T-----A--G-----T--C--A-----------------------------T------T-A-----A--- 2353
CPZ.US.85.CPZUS         ---C----------T--G-----AC---------CCA---A--------A-----A--G---A---G-G-A-----A-------T--------A---A-T-----T--C--A--G-----T--C-------------------------C--T------T-A-----A--G 2858
CPZ.GA.88.GAB1          ---C----------T-----C--AC----------CA---A--------------------A--G-----A--A--------------------------TA-T-----A--C--A--G---------------------------------C-----------A--------- 2859
MON.CM.99.L1            -CAG------G---G-G--TC-TAC-C-----AC-CA-------CTT--ATT---AC--T-GGA-CC--CC-A---CA-T--C-----T--CT-A--CTGT--------------------G----AG-GC-----------TGC-CA---------------------C---GC- 2504
MON.NG.x.NG1            -TA-------G---G-G--TC-CACCACC---RCCCGG------AT-G-ATT---CC--T-GGA-G--GCA-A---AACC---------T---T----TTGC--C--G----G-------G--T-AA--C---------G-TGC-CA----------G---T-A--C--AGC- 1097
GSN.CM.99.CN166         -TAGG-----G---G-G--C---ACCC----AACACA-------CTT-GATT---TC--T-GGA-------A--A--A-------------CT-A-----C-----------A--G--G---AAT-AC----------TGC--A-------T---A--T-A---G--GCT 2508
GSN.CM.99.CN71          -TA-G-----G---G-G--C---ACCC----AACACA---A---CTC---TT---CC--T-GGA---G--A-----A-------------CT-G-----C-----------A--G-------AAT-AC----------TG---A----------GA--T-G---G--GCT 2508
MUS_1.CM.01.1085        -CC-G-----G--C-GG--C--G-AG--C--AACCCAGC-----CT---A-T---CC----AGAGG---A--A--A--G-----T--T---T-----TGC-----A--T-G------G---AAT----G-------TGC--A----------GC--T-A---G--GCT 2488
MUS_1.CM.01.CM1239      -TC-------G--T--G--C----A--GC--AGT-CAGC-A---TT---ATT--ATC-G--AGA-G-G--A-----A--C-----T-----CT-----TGT-----A--C-G---G-----T-AAT------------TG---A-------T---C-G--A---G--GCC 2478
MUS_2.CM.01.CM1246      -TA-G--G--G---G-G---C--AA-------AC-CAGA-A---AT---ATT---TC--C-GGAGG--GCA------A-T--C-----T---T-A---------C------GA--G-------AG--A------C---TG---A-----------A-G-------G--GTG 2586
MUS_2.CM.01.CM2500      -CA-G---------G----TC---A-------AC-CAGAGA---TT----CTG--AC--T-AGA-G---CA------C------------------GC----C-----T----G-------G---AAT-C----------TGC--A----C-------A-G--C---G--GCC 2545
RCM.NG.x.NG411          GTA----G-----TG----T---AC---G-------AT-CC------------G--A---------T-G---A--A--A---------T------------A-----TGT--C--A--G-----TTCC--A------------GC-----------------T-G-----A--- 2292
RCM.GA.x.GAB1           --A-----------G----T---AC---------C-AT-CA-----G---------A-----------------A-----A----C--C-----C--A---A---CTGT--T--A--G------TCC--A-----------------G--------T--------A--------- 2277
DEN.CD.x.CD1            -TAC---G---T-------T---CA-C-G--AGTG---C--------------G--A-GGT-GGA---GGCGCCA--CACC--C--T--------A--T--G----TA--TCCA--G-------AA--AC---AT--G-TG---A----------GA--T-A---G--GCC 2685

MAC.US.x.239            --A-------G---GTT------AG------C---GA-A-------------T--TC-GT-GGAGG--GC-CCC--GACC-----------C--C--CAC------T-----------G--T-AG-AC----------TGC-GA-------T--G-----A---G-GTC 3100
Pol                     L__S__K__E__K__I__V__A__L__R__E__I__C__E__K__M__E__K__D__G__Q__L__E__E__A__P__P__T__N__P__Y__N__T__P__T__F__A__I__K__K__K__D__K__N__K__W__R__M__L__I__D__F__R__E__L__N__R_
H2A.GW.x.ALI            --A-----------G-G--C--AA-------C---GA-A-A-C-----GA-----CC--T-AGAGG-GGCACCT--AACT--C--T--T-----C--CAC------A--------G--G---AA-AC----------TGC--A-------T------T-A--C---GT- 3176
H2A.DE.x.BEN            --A----G------G-----C--AA---G--C---GA-A-A--------A--G--CC-GC-AGA-G-GGCACCT--AACT-----T--T-----C--CAC------A--T-------G---AG-AC---------G-TGC-GA-------T--------A-----GTG 3183
H2A.SN.x.ST             --A-----------G-G--C--AA---G--C---GAGA-A--------GA--G--CC-GC-AGAGG-GGCACCT--AACT-----T--T-----C--CAC------A--C-------G---AA-AC----------TGC--A-------T--------A--C---GT- 2628
H2B.GH.86.D205          -CAGG-----G---CT---CC-CAA------C---GA-A-A-----------G--AC--T-AGA-G--GCACCC---ACT-----------C--A--CACC---------------G-----T-AG-AC----------TGC--A----C--T-----GT-A--C---GT- 3159
H2B.CI.x.EHO            -CA----G------CT---CC-CAA------C---GA-A-A--------A--G--AC-GT-AGA-G-GGCGCCT---ACT------------T-G--CACC--C-------A--G-------AA-AC---------G-TGC--A----------------A--C---GT- 3154
H2G.CI.92.ABT96         -TA-G---------C-----C-CAGG---------GACA-A------C--------C--T-GGA-G--GCACCC--CACC-----G--T--C--A--CACC---A-T--C-----G--G---AA-AC----------TG---A----C----------------C-G-GT- 2510
H2U.FR.96.12034         -TA-G------T--G--------AA---G--A---GAGA-A-----------G--AC--T-GGA-G--GC-CCC--CACT-----------C-----CACC---------------G--G---AA-AC----------TG---A----C--T-----GT-A--C-GAGT- 2665
MAC.US.x.EMBL_3         --A-------G---GTT------AG------C---GA-A-------------T--TC-GT-GGAGG--GC-CCC--GACC-----------C--C--CAC----------------------T-AG-AC----------TGC-GA--C----T--G-----A----G-GTC 2574
SMM.US.x.H9             -GA-------G----TT------AG------C---GA-A-A-----G--A--T--CC-GT-AGAGGR-GC-CCT--AACC---------T--C--C--CAC------T------R-GR-----G-A-A---------G-TGC--A-------T------T-A--C---GTC 2573
SMM.US.x.PGM53          --A-----------GTT------AG------C---GA-A-A--------A--T--CC-GT-AGAGG--GCCCCT--AACC---------T--C--C--CAC------T-----A--G-------AA-A---------G-TGC--A-------T------T-G----G-GTT 3017
SMM.US.x.SIVsmH635F_L3  --A-------G----TT------AG------C---GA-A-A--------A--T--CC-GT-AGAGG--GCCCCT--AACC-----G--T--C--C--CACT-----T--------G-------AA-A-----------TGC--A-------T------T-A------GTC 3103
SMM.SL.92.SL92B         -TA-G-----G---C--------AC---G--C--CGAGA-A-----------G--AC--T-GGA-CG-GCCCCT---AC------T--T-----A--GACC-----T--T--------GA-T-AGGAC--G-------TGC-CA---------------T-A--C--A-TG 2531
STM.US.x.STM            --A-------G---G-G---C--AA---G--C---GA-A-A-----G-----T---C-GC-AGA-G--GC-CCT--AAC------T---------C--CACT-----T----------G-----AA-AC--G-----G-TGC-GA-----------------A-----AGTG 2747
MNE.US.x.MNE027         --A-------G---GTT------AG------C---GA-A-------------T--TC-GT-GGAGG--GC-CCC--GACC-----------C--C--CAC------T----------G-----AG-AC----------TGC-GA-------T--G-----A------GTC 2568
DRL.x.x.FAO             -CA-G--G-----TG-----C--GAA------A---GAGC-A-----G---------TC----AAGT-G---A--T--A--------------C--A--G--G---TG---T-----G-----TG-----C-----------------------------GC--T-A------GC- 2365
MND_2.x.x.5440          -TA-------G--CCT----C---CA------A--C-ATC-A--------A-----C--G--C--T-G-G-A--T--A--------------C--C--------CTGT-----A--G-----TG-A--C-GC--------------------T---C----G--C--AGTC 2359
MND_2.CM.98.CM16        -TA-G-----G--TTT----C---CA------A----GTC-A-----G--------CC-G--C--T--G--A--T--A--------T-----C--A--G--G---TG---C--A--G-----TG-A--C-GC-----------G--------T---C-GT-A--C--AGT- 2799
MND_2.GA.x.M14          -TA-------G--TTT----C---CAGAC---A---GAT--A-----------------C-----C--T-G---A--T--A----------T-----A---------TGT-----A--G-----TG-A--C-GC--------------------T--GC--T-A----GAGTG 2726
OLC.CI.97.97CI12        GTA-G-----G---G----T----AGA----A---GACAGAT-----G-A-----T--G--ACAG---GCA-A-G--TC-----A--GG----C---A-----TG-----A--G--G---AA-A-G-G-AT-----G--GA-------T--G-CT--A---GCAGCC 2672
GRV.ET.x.GRI_677        --AG---------TG----T---AC------A--C-A-C-A------G---G--A---T-A--T-GG--A--AGGA-------T--T-----A-----G----------------G--T-AA--AC---------TGC-T-----------G-----A--C--AGCT 2971
VER.KE.x.TYO1_patent    -TA----G--G--TG----T---CAG-----A---T-CC-AT-A--GC-------A-----CAGT-G-G-A--AGGA------G----------C---A-----TG---------G--G---AAT-CC-G-----G-TGC-------C--T-----GT-A------GC- 3025
VER.KE.x.9063           --A-G--G--------G--C---AC---G--A---GA-C--T-A-----A-----------CAGC--G--A--AGGA------G-T--T-----C---A-C--TGT--C-GA--G--G---AA---C-------G-TG--------C----------T-A-----AGT- 3028
VER.KE.x.AGM155         -TA-----------T---CC-GCA---------C-A-ACAT-A--GG-A-----A---T-AAGC-GGG-A---GGA--C---G----------A--------CTGT----G-----------AAT-AC-G-------TGC-G-----------G------C--C--AGCT 3020
VER.DE.x.AGM3           -CA-----------------C---CAG-----A---GACC-AT-A--G--A-----A------AGC--G--A--AGGA--G---G-------C--------G---TG------------------AGT-AC---------TG-----------T--G-----A--C--AGC- 2523
TAN.UG.x.TAN1           -------------------CC-GAC------AAC-CAG-CTC-A-----------A--------T--GG----AGGA--G-----T-----------C--G---TGT--CCG-----------AA--AC-G-----G-TG---A-------T----CT--C-----AGC- 3000
SAB.SN.x.SAB1C          --A-----------G-G--C---AAG-CC------GA---CT-A-----AC---AC-TT-GGA--G---A-----A-----------T--C--C-----C-----A----G--G-----T-AA---C-G-------T-C-CA-G--C-----GC-GT-A-----A--T 3196
MND_1.GA.x.MNDGB1       GTA----G------GTT-GTC-TCAGA----A---GATAGAT-A--GG-------A-----AGT-GGG-A-AT--A-G---AAT--------A--TA-C----------------G--G--T-AA-A-G-------------A----C--T-----T-A--C---TT- 2485
DEB.CM.99.CM40          --AGG---------G-G--C---AA---T--C-TA-AGCCC---AT-G---C---AC-G--AGTCCCGGCA-C---CACC--C-----T--CT-------G---TA--------G--G----AA-G---G---C---TGC--A----C--A----CCT-A--C---GCC 2607
DEB.CM.99.CM5           --AG----------G-----C---AA---G--CATA-GGCCT----T-G-A-C---AC-G--AGTCCCC-CA-A---CACT--C-----T-GCT----T--G---TT--------G--G---AG-A-----------TGC-TA----CC-G----CC----------GCT 2601
TAL.CM.01.8023          -CA-G---------G---GGC-GCAG-CC--CATAGAC--C----TG----C---AC-GT-GGAGG-G--------G-----C--------CT--------G-----T--C-GA--G--G---AG--AC---------TG--GA--------C---CC--G---C--GC- 2265
TAL.CM.00.266           -CAG----------G---GGC--CA-ATC--CATAGAC--C----TG----C---AC-GT-AGAGG-GG----A--G------------T---T-------G--C--T--C-GA--G--G---AA---C---------TGC-CA----C---C-C-CC--A---C-AGC- 2767
SUN.GA.98.L14           --AGG--G------G---GCC--CAGA--------GATAGAT-----GCA-----A--G---AGC-GGGCA-AC--A-GG--C--T--T--C--A---A-T-----T--------G-------AA-A-G----------G---A-------T----T---A---GGT-TG 3155
WRC.CI.97.97CI14        --A--A--------G-G-GGC--CAGA--------GATAGAC-A---GCA-----A------AGC-G-GCA-AT--A-G----------T--------TA-------T--T--A--G-------AA-A-G---------------GA-----------A--T-A---G-AGTT 3098
WRC.CI.98.98CI04        --A----G------G---GGC--CAGA----C---GATAGAT-A---GCA-----A------AGT---GC-TCC--A-G----------T--------TA-------A-----------G----G---AA-A-G-------G--GC-CA-------T---A--T-A---G-AGT- 3113
WRC.GM.05.Pbt_05GM_X02  --A----G-----TG-G-GGC--CA-A-G--A---GATAGAT-A---GCA-----A-----AAGT-GGGCA-AAG-AA-T-----T--------A---A-C-----A---GA--G-----T-AA-A-G----------------A-------T----G---A---GCAGT- 2027
LST.CD.88.524           -CA-----------G---G----CA--C-------GATAGAT-A--------------------A--CCC-G---AT--A-G-----T-----C--C---A-C-----T------------G-----AA-A-G---------G--GA------------A--T-A---G--TT- 2063
LST.CD.88.447           -TA-------G---G---GCC--CA--C-------GATAGAT-A-----A-----------A------GCA-ATG-A-GG--C-----T--C--A---A------------------G-------AA-A-G-G--------------A-T-----T---C----A---GCTTT- 2066
LST.KE.x.lho7           --A----G------G---GCC--CA--CT------GATAG-T-A-----A-----A-----A--TCC-G-G-AT--A-GG--C-----------------A-------T--C-----G-------AA-A-G---------G---A-T----------A-G--C---G-ATT- 3150
SYK.KE.x.SYK173         -G-T---G---T--G-G--CC-CAAG-C---AGTAGA---C---CTC-----C---AC--T-GGAG---GC-A-T---ACC-----T--T-----C----------TA--T-G------G---AA-AG--------G-TGC-CA-T-----T---A--T-A---G--GCC 2995
SYK.KE.x.KE51           --AG-------C-TG-G---C-GCG--C---AGT-T----T---CTTG-AA-G---C--T-GGA---GGCCTCC---AC------T--T-----C-----T----TA---C-A--G--G---AATTA--------G-TGCCTA-G-----T--GA----C---G-AGC- 2653
COL.CM.x.CGU1           GCA-----------G---GT---CA-A-G--A---GA-AGAT-----GCA--G--A-----AGA----GCA-AATTA-GG--------T-----A---A-----TG---T-G---------T-AA-A-G-------G------A----------G--G--A--C------ 2573
```


PLV Complete Genomes

```
H1B.FR.83.HXB2        ACTCAAGACTTCTGGGAAGTTCAATTAGGAATACCACATCCCGCAGGGTTAAAAAAGAAAAAATCA..................GTAACAGTACTGGATGTGGGTGATGCATATTTTTCAGTTCCCTTAGATGAAGACTTCAGGAAGTATACTGCATTTACCATACCTAG 2950
Pol                   _T__Q__D__F__W__E__V__Q__L__G__I__P__H__P__A__G__L__K__K__K__K__S__.__.__.__.__.__.__V__T__V__L__D__V__G__D__A__Y__F__S__V__P__L__D__E__D__F__R__K__Y__T__A__F__T__I__P__S_
H1A1.UG.85.U455       -----G-------------------------C--G---A-A--G--TC------------------..................----------A-------G--C------------------T---------AG---T--A-----------G--C----------- 2396
H1B.US.90.WEAU160     ----------------------------------------TT------------------------..................-----------------------------A-----------------------C--------------------- 2949
H1C.ET.86.ETH2220     -----------T-----------------G---------C--A-----------------------..................--G-------A-------G----------C--------T----------GT-----A--A-----------C----------- 2342
H1D.CD.84.84ZR085     -----------------------------G-------------T-----A-----G-------G--..................A-----------------C------------A---------TG--------T--------C--------------------- 2472
H1F1.BE.93.VI850      --------T-------G--------------------------T-------------------G--..................------------------G-----------------------------A-G--T---A-----C---------C----------- 2287
H1G.SE.93.SE6165      ----------------G--C-------------T------T-------------------------..................----------A-------G------------------------------------T--A--------------C--T--------- 2347
H1H.CF.90.056         ------------------------G-------------C--A------------------------..................--T---------------G-------------------C--T-----A---A-----A-------------C----------- 2297
H1J.SE.93.SE7887      ------------------------G-----------------A-----------------------..................-C----------------G-------------------C--T---T------T--------A----------C--T--------- 2264
H1K.CM.96.MP535       ------------------------------------------A-----A-----------------..................------------------G-------------------C--T-----A-----------------------C--T--------- 2146
H101_AE.TH.90.CM240   -----G-----T-------------------G-----A-----T----------------------..................----------A-------A------------------------T---------AG---T--A-------------C----------- 2524
H102_AG.NG.x.IBNG     --------T-------G--C--G----------T-------G--A---------------------..................--------G--A-------G------------------------------A-------T--A-------------------T--------- 2475
H1N.CM.95.YBF30       --C-----T--T--------G--GC-------T--------A-----A----GC------------..................-G-----TT------A--A-----T--------TG------G--CA----T--T--A--------A--T-------------- 2544
H1O.BE.87.ANT70       --A-----G--------G--A--GC----T--C--------G-GG--T--G--GC-A--GC----T..................-T-----CT-A-----A--A-----T-----C---TG---------CCC---T--T--A--A--------T--C--T--T----- 3005
H1O.CM.91.MVP5180     --A-----T--------G--G--------T--T--------A-GG--T-----GC-A-GGC----T..................-T-----CT-A-----A--A-----T-----C---TGC--T------CC------T--A--A--C-----C--C--T--T----- 2980
CPZ.CD.90.ANT         --A-----T--T-----GA-A----------T--T--T-----A--C--A-----GC-A--G------..................-G------T----------A-----C-----C--CA-A---------C-G-----T--A--A-----A--T--C--A--T--A-- 2389
CPZ.CM.01.SIVcpzCAM13 --C-----T-----------A-----G--C-----C--C--C--A-------------G----G------..................----------A-----------------------C--TTG---TC-----A-G-----------------A-----C--A--T----- 2530
CPZ.CM.05.SIVcpzEK505 --------T--T--------A-----------C--------A------C----G------------..................--------G--A-----A--A-----T--C------TG---------------A-T--T--A--------A--T--------------- 2522
CPZ.CM.05.SIVcpzMB66  -----------T--------C--GC----------C-----C-----T------------C-----..................--------TT-A-----A--G--C--T--C------TG----C-G-----A-----T--------C-----T-----------C-- 2496
CPZ.CM.98.CAM3        --A--G-----T--------C--G-----C--------C--G-----T--------------G---..................--------G--A-------A-----T--------CTGC--A---------A-T--T-----A-----A--------T-----C-- 2361
CPZ.GA.88.GAB2        --A-----T--T--------A--G-----C--T-----C--T-----T-----G--A--G------..................--------G-C--------A-----T--------CTGC-------------A-----T--A--------A--C--C-----C----- 2343
CPZ.CM.98.CAM5        --A--G--A--------G--C--------C--------C--A-----AC----G--A--G--G---..................----------A---------A-----T-----C--CTGC--A---------A-T--T--A--A--C--A--T-----T-----C-- 2652
CPZ.CM.05.SIVcpzMT145 --------A--T--------A---C-G--C--------C--A--T--------G------------..................--------CT-----------A-----C---------TG---TC-T--CCC------T--A--A-T----------C-----T----- 2492
CPZ.TZ.00.TAN1        --------T--T-----GT-A--GC----------C-----G----------G---A-G---TATG..................-G--------------A--G-----C--C-----CA-----C-G---CC----------A--------A--T-----------C-- 2594
CPZ.CM.05.SIVcpzLB7   -----------T--------------------T--C--T--G--A-----G-------GG------..................-C------T-------A--A-----C--C------TGC---C---------A-----T--A--A-------------C----------- 2505
CPZ.US.85.CPZUS       --A-----------------G-----C-----C--C--A-----A-----G-------G-------..................--------C--A-----A--G-----C--------TTG---AC-G---A-G--A--T--A--A--C--A-------------C----- 3010
CPZ.GA.88.GAB1        --A--------T--------G--G-----C-----T--C--A------------------------..................-G------T-A-----A--A-----C--C--C--TTG----C-G---A----T--------A------------------A--T--C-- 3011
MON.CM.99.L1          -----G-------TT--G--G--G--G-----C-----C--A--G--AC-TG-G----TGG-CCAT..................-G---A--G-A--CA--AAG--C--C-TC-ACAGCA-------GTGG---CCG-----A--A--C--G--G--CT-A-----CTC 2656
MON.NG.x.NG1          --A-----T----TT--G--A-----------T-----C--G--T--AC-CC-G--C-TGGC-AAT..................A----TA--G----CA--AAA-----C-T--ACAGCA-C--TC-G-------A-T--T-----A-------------T-A------TC 1249
GSN.CM.99.CN166       --G--------GGCT--G--C--G--G-----T--G--C--A-----AA-TCCGC-CCCTGC-GG-CTAGAGCAAAAGGAGCAT-----TA--A-A---A--AAA-----T----A-AGTA-A--A--GT----G--T-----------C--C--T---T-AG-----TC 2678
GSN.CM.99.CN71        --G--------GGCT--G--C--G--G--G--T--G--C--A-----AC--..................GAACAAAAAGAGCAC-----TA--A-A---A--AAA--C--T----ACAGT--A--A--GT-------A-----A-----C----------T-AG----ATC 2660
MUS_1.CM.01.1085      --C-----AC--GCA---------C-------C--------A-----A--..................ATAGAGCATGAGCAC--T---A--G-----A-CAAG--C--C-T--ACAGT--C---C-----CC------T--------C--A--C---T-GC-G---CA 2640
MUS_1.CM.01.CM1239    --C-----G---ACA--G--A--GC----G--C--------A-----A--..................GCGGAACATGAGCAT--C---A--G-A---A-CAAG--C--C-T--ACAGC--C--A------CC--CA-----A--A--C--A--C--CT-GC-G--CCA 2630
MUS_2.CM.01.CM1246    --A-----A--TGCA-----C--GAC---G--C-----------T--------GCCC-A--G.........GCA.........CAT--C--TA--G-A---A--CAG-----C-T--A-AGT--C--A-----CA----A---C-ACCC--C--G--C--CT-GG-C---GC 2738
MUS_2.CM.01.CM2500    --A-----------CA-----C--GAC---G--C--G--C--AT-----C-TTCT--A--G.........GCT.........CAC--C--CA--G-A--CA--CAA-----C-T--ACAGTA-C--AC-CT----G--G--T---CCT--C--A--C--CT-AG-T----- 2697
RCM.NG.x.NG411        --A-----T----TT--G--A--GC-G-----C--G-----A-G---A------C-ATGC-GGCA-..................A-T-----CT-A---A-T--G-----G----------TGC---C--T----GCCA--T--A-----C-----T--C--G----TAGC 2444
RCM.GA.x.GAB1         --A--G-----T-TT--G--G--GC-G-------------A-G---A-----GC-ATGTG-GAG-..................A----G---T-A--CA-C--G--------------TG--TTC-GT----GCCT--T-----A---------------A-----AGC 2429
DEN.CD.x.CD1          --C--G--------------A--------T-------------T-----A---C--C----GG--CAC..................A-C--T--GT-A--CT--AAA----------A-AGT--G--T--G---CC---T--TGCAC-A--C--A--------AG-G--GTC 2837

MAC.US.x.239          -----G-----TAC------C-----------------C--T-----AC--GC---A-GG---AG-..................A-T------------A-A-----------------C--CA-A--TC-----------A--T---C----C-----C-----TT----ATC 3252
Pol                   _V__T__Q__D__F__T__E__V__Q__L__G__I__P__H__P__A__G__L__A__K__R__K__.__.__.__.__.__.__R__I__T__V__L__D__I__G__D__A__Y__F__S__I__P__L__D__E__E__F__R__Q__Y__T__A__F__T__L__P_
H2A.GW.x.ALI          --------T---ACA--GA----G-----G--T--------A-----A--GCC------G-AG-..................A-C--T--G--------A--G-----T--C-----CA-A--AC-GC----GAG---T--AC---------------TC----ATC 3328
H2A.DE.x.BEN          --------T---ACA----A----GC-------T-----C--G-----AC--GCC--A--G---AGG..................A-CT-TA--T-A-----A--G-----C---------CA-A--AC--C-------T--T---C-----------------C----AGC 3335
H2A.SN.x.ST           ------------ACA----A-C--G--------T-----C--A-----AC--GCC-----G---CG-..................A-T--T--C--A-----A--G-----T--C-----CA-A--AC--C----G--T--T--AC---------------TC----ATC 2780
H2B.GH.86.D205        --C--G------ACA------CA-C-.G--T--T--C--C--G-----TG..GC-G-A----GGAG-..................A--------A-A--------A-----C--C--CAGTA-C--AC-----CC-A--------AC--------A-----C---T-G--ATC 3308
H2B.CI.x.EHO          --C-----A--TACA--G--C--GC-G--T--T--T--C--A-----AC-GGC-TCA--G---AG-..................A-----------A-----A--A-----C--C--CAGT--C--AC-----CC---------AC-A-----A--------TT-G--AGC 3306
H2G.CI.92.ABT96       --C--G-----TACT-----C--G--G--C-----T--C--T-----T--GGC-G-A----GGAG-..................A--------------A-A--G-----C--C--CAGCA----TC------T---T------C----C--A--T--C--TT----ATC 2662
H2U.FR.96.12034       --------T--TA-T-----C--G-----T-----C--C--A-----AC--GCTG-------GAGG..................A-------T--A-----A--G-----C--C--CAGCA-C--GC-GTG--------------C-A--------T-----TT----ATC 2817
MAC.US.x.EMBL_3       -----G--AC-T-ACAG-AG--C-A-TA----------C--T-----AC--GC---A-GG---AGG..................A-T------------A-A-----------------C--CA-A--TC-----------A--T---C----C-----C-----TT----ATC 2726
SMM.US.x.H9           --------T--TACA-----A--G-----G--------C--TR----AC--GC-----G--G-AGG..................A-C-----GT-------A-----------------C--CA-A--T-----------A------C-A--C-----C------T----ATC 2725
SMM.US.x.PGM53        --------T--TACA-----A--G-----G--------C--T-----AC--GC-----GG-G-AG-..................A-C------T-------A-----------------C--CA-A--TC-----------A------C-A--C-----T-----TT----ATC 3169
SMM.US.x.SIVsmH635F_L3 -----------TACA-----A--G--------------C--T-----AC--GC-----G--GGAGG..................A-C------T-------A-----------------C--CA----TC-----------A------C----C-----C-----TT----ATC 3255
SMM.SL.92.SL92B       --A-----A---ACA-----A--GC-G--T--C------G-----CC----GG-A-TGG--AGG..................A-------G--A---A-A--G-----C---------------A---C----CCCG--A--T--AC-A--C--A-----C--T-----ATC 2683
STM.US.x.STM          --A-----T--TACA---A----GC-----T--------C--T-----AC--GC-----GG-G-AG-..................A-C------T-------A--G-------------CA-A---C--------G-G---T---C-----------T--C---T-----ATC 2899
MNE.US.x.MNE027       -----G-----TACA-----C-----------------C--T-----AC--GC---A-GG--GAGG..................A-C--------------A-----C---------C--CA-A--TC-----------A--T---C----C-----T-----TT----ATC 2720
DRL.x.x.FAO           --A--G-------TT--G--G--G--G--G--------C--G-G---C------GCATGCG--CAG..................A-C--T-------------A--------C-----C---TGC--TC-G---CCG-----T--A-----C---------C-----T--ATC 2517
MND_2.x.x.5440        -----G--T----TT--G--G--G--------CA-------A-G---CC----GC-ATGTG-GCA-..................A-T------T----CA-A--A-----C----------TG------G----TG-----T--A-----------G--C-----T--ATC 2511
MND_2.CM.98.CM16      -----G-------TT--G--G--G--G-----C-----C--T-G---TC-C---C-ATGTG-GCA-..................A-T------------CA-A--G-----C-----C---TG---T--G-----G-----T--A-----C---------C-----T--ATC 2951
MND_2.GA.x.M14        -----G--T----TT--G--G--GC-------C--------G-G---C------C-ATGTG--CA-..................A-C---A--T-----A-A--A-----C---------TG---T--G-----GAG---T--A-----C-----------------T--ATC 2878
OLC.CI.97.97CI12      --A-----T--------G--A-------------T-----A-----------G-A-GGG--CAG..................A-T------T-A---A-A--A-----T-----C--TA-A--GC-----CC---A-AT--AG-A-----A---------A--T--ATC 2824
GRV.ET.x.GRI_677      -----------T-TT--G----GC-G-----T--T--C--A--G--CC-TC-G--A--G--GCA-..................A-C------A-A--CA-A--G-----C-----A---A-A--A--TGCA-G--A-----A--A-----A-----------C--CTC 3123
VER.KE.x.TYO1_patent  --C-----T----TT-----G--------G-----C--C--A-----A----G----TG-G-CAG..................A-------TT-A-----A--A--C--C-----A---CA-A--A--G---CC-A-T--T-----A---------T-----T--T--C-C 3177
VER.KE.x.9063         --A----------TT-----------G--------T--C--G---------GC-G--A-TG---CAG..................A--------T-A-----A--A--G-----T-----A---CA-A--A------CC---G--T--A--A--C--A---------A--C---TC 3180
VER.KE.x.AGM155       --A----------TT-----C--------T-----C-----A-----------G--A-TG--GCA-..................A----CA-TA-A---------G-----------A-AGCA-A--AC-G---CCT--G--T--A--A--C--A--T--C-----C----C 3172
VER.DE.x.AGM3         --A-----T--T-TC-----A--G-----C-----T-----AT------CG------TG-CGGA-..................A--------T-A--CA-A--G-----C-----A----A-A--A-----CCC---G--T--A---------C--T---------T--ATC 2675
TAN.UG.x.TAN1         --A--G-------TT--G--A--------T--T--C-----A-----A--G----GA--G---CAG..................A-C---------A---A-A--A-----C-----A-AGTA-A--A--G---CC----A------CCC-----A--T---------C--ACA 3152
SAB.SN.x.SAB1C        -----------TCA------C--G-----G-----C--C--A-----C---C-GC-A-GGG-GCAG..................A-T-----GT----CA-A--A-----C------AGCTGC--T------CC------TCA---A-----A-----C-----C--ATC 3348
MND_1.GA.x.MNDGB1     --A--G--T--TCAT---T-A--G-----T-----T--C--A-----AA--------TGT---AG-..................A-------C--A---A-A--G-----C------AGTA-A--TC-G---CC---T-AT--ACCC--------C-----GG----ATC 2637
DEB.CM.99.CM40        --A--G--GG-G-----G--G--GAC---T---------C--AT-T-CAC--CC-C-A-TGG--CAG..................A-C--T--CT-A---C-T-CA-----C-----A----A-AG-AC----CCC-ACA--TGCA--A--C--A-----C--AG----ATC 2759
DEB.CM.99.CM5         --C--G--TG-----------G---AC---G--C--G--C--T-G-----GCC-C-A-TGG--CAG..................A-------C--A--CT---CA-----C-----A----A-A-----G---CC-C-G--TGCACGT-----A-----C--A-------TC 2753
TAL.CM.01.8023        -----G-----TGCA--G--C---CAG--G-----C---T-T--AC--CC-C-------GCAG..................A-------CT-A---A--AAG-----C-----A-AGCA-C--A------CC--------GCCCGT--C-----T--C-----C--GTC 2417
TAL.CM.00.266         --C---------GC------A---CAG--G-----T--C---T-----CC-GCCTC---TG--GCA-..................A----G--C--A--CA--AAG-----C-----ACAGCA-C---------CCC-------TCAC----C--A--C--C-----C--GTC 2919
SUN.GA.98.L14         -----------TCAT--GT-A--G-----G--T--T-----A-----A-----G---TGT---CAG..................A-T------G-A---------G---------C--C---A-A--A--G---CC---T-AT--A---------A-----C--A-----GTC 3307
WRC.CI.97.97CI14      --A--G-----T-TT-----A--G-----------T--C--T-G---AC-G-G------T--GAAC..................A-TT-----G-A---A-A--G------------C--TA-A--A------CC---A-AT---C-A---------T-----AG----ATC 3250
WRC.CI.98.98CI04      --A--G-------TT-----A--G--G--C--T--T--C--A-G---AC---G------T---AAT..................A-TT-----G-T---A-A--G-----C--------CA-C--AC----CCC-----AT---C-------A--T-----TG----ATC 3265
WRC.GM.05.Pbt_05GM_X02 --A----------TT--G--A--------C--T--T-----A-G-------C-T------G-AAT..................A--------T-A---A-A--A---------------TTA-C--A------CC---A-AT--AC----C--A---------AG-----TC 2179
LST.CD.88.524         --A--------TCAT--GC-A--G---------------C--A-----AA----G--ATGC---CA-..................A-T------G-A---A-A--A-----C-----CAGCA-A--T------CCT--T-A---AC-A--------G--C--T-----GTC 2215
LST.CD.88.447         --A--------TCAT-----A--------C--T-----C--T-----AA----G--ATGC--GCAG..................---------GG-A---A-A--A-----C------AGCA-----C-G---CC-----A---AC----C--A-----C--AG----ATC 2218
LST.KE.x.lho7         -----------TCAT--GC-A--------T--T-----C--A-----TA-C--G--ATGC---CAG..................A-T------G-A--CA-A--A-----C-----CAGTA-C--T------CC-A-T-A---A--------A--------T------TC 3302
SYK.KE.x.SYK173       --A-----T--T-TC--G--A-----G--------C--C--A-G---AC-----CG-C-------..................C-----A-TA-A---T-AAAA----------AC-----C---C-G--CA----G--T---CCA--C--G--------AG-G--A-- 3144
SYK.KE.x.KE51         --A--G-----T--------A--G------------T-----A-G----C-C-GT-CA----G-...................A-G------T-A--CT--AAA--C--C----ACA----A--TC-----------G-----AC-------A--------AG-G--C-- 2802
COL.CM.x.CGU1         --A----------AT------C--------G--------C--A-G---TC-T-T-G-A--G-CCCAT..................A--------T-A---A-A--A-----C-----C---A-------G---CCG--T-ATCAA-----C--A--T-----A--C--ATC 2725
```

```
H1B.FR.83.HXB2          TATAAACAATGAGACACCAGGGATTAGATATCAGTACAATGTGCTTCCACAGGGATGGAAAGGATCACCAGCAATATTCCAAAGTAGCATGACAAAAATCTTAGAGCCTTTTAGAAAACAAAATCCAGACATAGTTATCTATCAATACATGGATGATTTGTATGTAGGAT 3120
Pol                      I  N  N  E  T  P  G  I  R  Y  Q  Y  N  V  L  P  Q  G  W  K  G  S  P  A  I  F  Q  S  S  M  T  K  I  L  E  P  F  R  K  Q  N  P  D  I  V  I  Y  Q  Y  M  D  D  L  Y  V  G
H1A1.UG.85.U455         --------------------AG-C--G---------------------G---------------------T----------G---------------------------C----TC----C------------------------------------C--------G--- 2566
H1B.US.90.WEAU160       ------------A---------------------------------------------------------------------------------------A--------------------------------------------------------------------- 3119
H1C.ET.86.ETH2220       --C---------A-------------------A--T------C--C------------------------C----------G---------C-CC--------------C---GGCC-CC--C------A-----------------T--------C------------- 2512
H1D.CD.84.84ZR085       ------------------------------------------A-------------------------------------------------T----------------C-------------------AG--------------------------------------- 2642
H1F1.BE.93.VI850        -G-C-------------T--A----G--C----------------------A-------------------------------T-------------------------C-----TGA----C------------C----C---------------------------G- 2457
H1G.SE.93.SE6165        ------T--------------G--------------------------------------------------------T--G--------------G------------C----GC-A-T--------A--G--G-----C------------------A---------- 2517
H1H.CF.90.056           ------------------------------------T---------------------------------------------G----------------------C---C-----G---------T--A--G------T-C----------------------------- 2467
H1J.SE.93.SE7887        ------------------------C--------------C-----A--------------------C-----------T--GT--------------------A-A--------G--AG---C-----A-----------C--G------------C--------G---- 2434
H1K.CM.96.MP535         ---T--------A---------G----------------------A-----A-------------------------------CA------------------------C-----T-A----------A--G-----A--C------------------A--------G- 2316
H101_AE.TH.90.CM240     --------------------A--C---------------------G-----------------------G-----------G---------------------------C-----T-A----------A--G-----------------A------C------------- 2694
H102_AG.NG.x.IBNG       -G----T-----------------------C--------------------------------G-----G--------T--GGCA------------------------C-----C-A----------G-----G-----C------------------A---------- 2645
H1N.CM.95.YBF30         ------------------T--T--------C-----T--------G-----A--C--------G-----------T--T--G----CA-----------TC---A--A--C---G-GA--C--------G---A-C--T--C--G-----------CC-C-----G---- 2714
H1O.BE.87.ANT70         -G-G--T--------C-----A--A-----C-----------C--C--G--A-----------T--------T-----------TCA-----------TC----T--A--C--G-G-G-C--C-----AT---AA--T-G---G-----------CC-A----------- 3175
H1O.CM.91.MVP5180       -G-G-----------C-----AG-A-----C-----------C--C--G--A--G---------T--------C-----T--G--TCA--------G--TC----T--A---------AGC--C-----AG---AA--T-----G-----A-----C--A---------- 3150
CPZ.CD.90.ANT           CG-G-----------C-----A--A-----TGT--T-----T--A-----A--C-----------------T--T--T---GCA---------C--G------CA--A-----GG-TA-GT------CAG---AA--T-----------------C-C------------ 2559
CPZ.CM.01.SIVcpzCAM13   -G----T----A--T--------C-----------T-----TT-A-----A-----------G--T-----T--C--T---------T----------TC-G--A--C--C---GC-A----C--T--T--CAC-G---C-------------C--A-----------G- 2700
CPZ.CM.05.SIVcpzEK505   -G----T----A-----T--T--------C---------------G-----A--C--------------------C--T--G--C-CT------------A----A--A----G--GA-C--------AC----C--A-----G-----------C-C---------C-- 2692
CPZ.CM.05.SIVcpzMB66    -G-------------------A--------C-----T--------A--G--A--G--------G--------------T--------------------T------------C------G--------A---A--------------T-----------A--------C- 2666
CPZ.CM.98.CAM3          -G----T----A---------A----------A--T-----T-A-----A-------------G-----------T--T-----C-------T----TC--ACC---A--C---C-G--GC--------T--A-A--T--C--G--T---------C-T-----G----- 2531
CPZ.GA.88.GAB2          -G----T----A--T-----T--A--G--C-----T-----T--A-----A------------G----------C---------C----------GG-T-----C-----C---C-GA---C-----AT-G--A--T--C-------------C--A------------- 2513
CPZ.CM.98.CAM5          -G----T--------------AG-------------T-----T-A--R--A------------G-----G-----C--T-----C-------------A-----T--A--C--------GC-------TC--A-A-----C-----T--------CC-T----------- 2822
CPZ.CM.05.SIVcpzMT145   C-----T----A--T-----A-----------A--T-----A-----T--A---------G-----T--TT-C--C--T--G--------------G------A--C--------C-A-GC-------A---A---T--C--G------------C--------G--C-- 2662
CPZ.TZ.00.TAN1          -C-C--T---A-C----------AA----T------T--C---T-A--T-A--T-----G-----T---------T--T--G--C--T-----------C----T-----C------G--C-C-----TG-G-AC--T--C-----T---------C-T--CA----T-- 2764
CPZ.CM.05.SIVcpzLB7     ------------A--------A--C--------A--------T--A--G-----------------------T-----T----------------G--T---------A--C---G---G--C-----T-----A--T-----G---------------A--------G- 2675
CPZ.US.85.CPZUS         -G-------------C-----A--------C-----T-----T-A-----A------------G-----------T-----G--C--T-----------TC-----T------------C-------TG-TA-A-------------T---------C-C--------G- 3180
CPZ.GA.88.GAB1          ------T--------C-----AG----------A--T----TT-G-----A------------G--T---T-T--C--------C--T---------TC-----A--C--C---G--A-G-----T-----TAC---T--C--G-----------CC-A-----G--G-- 3181
MON.CM.99.L1            -T---------CTGA---T-CT-AA--G-------T---G--C--C--G-----C--------G--C---A-TG-G--T---CA--CAGC-G--G-GC-ACC-C--GAGA-C--G---A--T-C-----GG--AC-T-AATA--G-----------CC-ACTGA------ 2826
MON.NG.x.NG1            AG----T----CAGA-----CCG-G----------T--GA---T-G--C-----R-----------C---A-------T--GTACTCAGCAT--G-GC-GC-RA-AGRAA-C--G-G-A--T-------GG-CACCC--ATC--G------------C-ACTCA----T- 1419
GSN.CM.99.CN166         AG----T---C--G-C----CTGAA--G-------T---A--CT-A--G-----T-------CG--G---A-C--C--T----AC-TTG-AG-TTCCT-GC--CGA-AGA-----G-T--GG-G-----TG-G--AT-AATA--G--------------ACTGA------ 2848
GSN.CM.99.CN71          AG--------C----C----CTGAG--G-------T---A--TT-A--T-----T-----G-CG--T---A-C--T--T--GGCC-CTG-AG-TTCCT-G---T-T-AGA------G-C--GG-G-----TG-G--A--AATA--G-----------C----TGA------ 2830
MUS_1.CM.01.1085        GG-------CTT-GG-----CA-GA--------A-T---A----A--------C--------G--T--TA-C--C-----GTTC-CAGCAG-C---C-------AGAGA---AG---GC---C-----TG-CAC-C--ATC---------------CC-CCTGA------ 2810
MUS_1.CM.01.CM1239      GG----T--CC--GG-----CA-GA--G-------T---A----G--------G-----------C--CA-C--C-----TAC-CGGCAG-----T-A---C--GAAA-C--GG--GC---------T--CACCC--ATC--G--------------C-C-TAA------ 2800
MUS_2.CM.01.CM1246      AG-G-----CAT-GG-----CT-AA----TC--A-TT--G---T-G-----G-----G------CA-C--T--T---GCC-CA-CAG-C--GC-TC------AAG----G--GG-C----AG---GCCT-A--AGTC---------------CC--CT-A----G- 2908
MUS_2.CM.01.CM2500      -G----T--CCTAGGG--C-C--AG--G----A-T--G--G--C--G--C--A--G----G--G--C--CA----T--T---GCA-CA-CTG-----T-A-------AAA-C--G---G------AG--TGC-T-A--AGTA----T--------CC-ACTCA----G- 2867
RCM.NG.x.NG411          ---T--T---C-AGG---T--C------------A--T-----A--T------T-G--C-----------G-----T-----T-----T---CA-CAGCT-AC-----T-----A--A-----G---G-G--------TG----C-------G--T---------------T-----G- 2614
RCM.GA.x.GAB1           AG---T---C-AGG------AG-G--G-----A--T-----A--T---------G--T-----------G-----T--C--C--C--T-GGCATCAGCC-AT--G------C----A-----GG--G-G--------TG-CA-C--T--C--G------------C-C-T------C- 2599
DEN.CD.x.CD1            C--T--------AGG------ACAA--G--CATT-T---A--CT-A-----A------GT------G--TA-C--T--T--GGCC-CAGCAGAT-----GC-TAGC-AA-----G--GGCCC----T--TG-GA--T-AATC--G--T---------C-C-TG----C-- 3007

MAC.US.x.239            AG----T----CAGAG-----A-AAC----CATT--T--G--T--G--T-----------G--G--------C--C------TAC-CT----G-C-TG-GC----A--C--C--G--GGC---------TG-GACCT-AGTC--G--T--------CA-C-TAA---CTA 3422
Pol                       S  V  N  N  A  E  P  G  K  R  Y  I  Y  K  V  L  P  Q  G  W  K  G  S  P  A  I  F  Q  Y  T  M  R  H  V  L  E  P  F  R  K  A  N  P  D  V  T  L  V  Q  Y  M  D  D  I  L  I
H2A.GW.x.ALI            AG---------CAGA------A-AA-----ATA--T--A--CT-A--G------------G---------------T--T---CAC-CA----G-C-G------------A--C-----GGC---C-AG--TG-CA--C--AT---------------A-C-TAA---CTA 3498
H2A.DE.x.BEN            AG--------AT-GA------A-AA-----ATA--T--A--CT-G----A---------G---------------T--T---TAC-CA----GGC--G---------A-----C------GC---C-----TG-CA--C--ATC--G-------------A-C-TAA---CTA 3505
H2A.SN.x.ST             A----------CTGA------A-AA-----CATA--T--A--CTCA-----------------G---------------T--T--GTAC-CA----GGC-GG--------A--A--C------GC---C--G--T--CA--C--AT---G-------------A-C-TGA---CCA 2950
H2B.GH.86.D205          AG---------CAGAG-----A-AG-----CATT--T--A--T--A-----A--G-----G--G--C-AGT----T-GT---TACTC-----G---GG-A-----T-----------GGCC--CAGC--TG-CA----AAT---G-----------CA-CCT-A---C-A 3478
H2B.CI.x.EHO            AG----T----CAGA------A-AG-------TT-----A--C--A-----------------G-----C---------T-----GTAC-C----G----GG-AC----C------C------GCC--CAAT--TG-CAC---AATC--G-----------CA-TCTC--G-C-A 3476
H2G.CI.92.ABT96         AG---------CAGA------A-AA-----CATC-----G--C--A--G--A---------G--C--------T--C------CA--CA----GG--GG-AC----C--C--C------GCC---GAG--TR--A----YATC-----T---------CA-TCT---G-CCA 2832
H2U.FR.96.12034         C--T--T----CAGAG-----A-AA-----CATC-----G---T-G--T--A--C-----------------A-----T--T--GTA--CA----GG--CC-AC-G-----A--------GGC----GA---TGCT--C--AAT---G--T---------CA-TCT-A---CCA 2987
MAC.US.x.EMBL_3         AG----T----CAGAG-----A-AAC----CATT--T--G--T--G--T-----------G--G--------C--C------TAC-CT----G-C-TG-GC----A--C--C--G--GGC---------TG-GACCT-AGTC--G--T---------CA-C-TAA---CTA 2896
SMM.US.x.H9             GG---------CAGA------A-AA-----CATC--T--G--AT-A--T--A--G------------------T--T--T--RCA--CTR---G---TG-A-----A--------------GC---------TG-GACCC-GATC----------------CA-C-TAA---CTA 2895
SMM.US.x.PGM53          AG---------CAGA------A-AA-----CATC--T--G--AT-A--T--A--G--------G--------T--T--T--GCA--CT----G---TG------GA-----C-------GC---------TG-GACCC-GATC----------------CA-CCTAA---CTA 3339
SMM.US.x.SIVsmH635F_L3  AG----T----CAGA------A-AA-----CATC--T--G--AT-A--T--A--G-----G--G--------T--T--T--GTA--CT----G---TG-A-----A-----C-------GC---------TG-GACCC-GATC----------------CA-C-TAA---CTA 3425
SMM.SL.92.SL92B         AG-G-----CC-AGA------A-AG--G--ATT-----G--C--G-----A---------G-----T--T-----T-----GGCA-C-----GGC-GG-A-----A--C--C------GCT--C-----TG-GC--C--GTC--G--T-----------ACTAA-T--CA 2853
STM.US.x.STM            AG---------CAGA------A-AA-----CATC-----G--T--G-----------------G---------------T--T---TA--CA----G---T--A---------A--C--G-G-GCC--C-----TG--AC-C-GATC-----------------A-C-TGA---CCA 3069
MNE.US.x.MNE027         AG----T----CAGA------A-AAC----CATT--T--G--C--G--T-----G-----G--G--------C--C------CAC-CT----G---TG-GC-G--A--C--C--G--GGC---------TG-GACCC-AGTC--G--T---------CA-C-TA----CTA 2890
DRL.x.x.FAO             GG-G--T---C--GGC-----A--C--G--------T-----T-----C--C------------------C-----T-----C------GC--CAGCTGAC--G-------C----A--C---G-GA-----------TG-GA----A-----G-------------C---T-------A 2687
MND_2.x.x.5440          GG-G-----CC--GGC-----A--C-----C-----T-----T-----C--A------------------C--T---------T-----GGCA-CAGCAGAC-----TC--A--G------AGG--A-GC-------AG-CT-A---------G--T---------C----T---G--GA 2681
MND_2.CM.98.CM16        GG-G--T---C--GGC-----A--C--G--C-----T-----T-----C--A-------------G--C--C-----C--T-----GGCA-CAGCAGAT---------C-----C--C---G-GAG-C-------TG----G------------T---------C-C-T---T--GA 3121
MND_2.GA.x.M14          GG----T---C-AGGC-----A--C--G--C-----T-----T-----T-------------------C--T--G---------T---GCA-CGGCAGAC-----TC--A-AA-A--C-A-G-GG--T-C-----AG--T-A--T-------------------C---T---G---A 3048
OLC.CI.97.97CI12        AG----T-----T-G------A-GA--G---ATA--T------T-A--T--A---------G--G-----TGT--C-----G----CAG--G-TC-TT-GA-G-------GG--G--GG-------T--G----A-T-A--C--G--T---------C---------G---- 2994
GRV.ET.x.GRI_677        AG----T---ACAGGG--------A--G-------T----TGT--G--T------------------T--TA----T-----G-A--CGGCAG----C--T------GAGA-C-A--GG--C-C---T-GGT---AA--TGTC-----------C-------GGT-G-CG- 3293
VER.KE.x.TYO1_patent    AG-G--T---C--GG---C--------G-----A-T---CTGT--C--G--A--G------------T--TA----C--------A--CAGCAG--TCC--T--G---GAGA-A-A--G-A-CTTG----CAC--ACC--TGTA-----------------A-GG-----T- 3347
VER.KE.x.9063           AG----T---C-AGG------TG----G-----A-T----TGC--------A--C---------G--T---A-T--T--T--G-A--CAGCAG-------T------AAAA-A-A-G-GG--TTAAA-CCAT-GACA--AGTG--G-----------------GG-----G- 3350
VER.KE.x.AGM155         GG-----------GG------C--A---------A-TT---TGC--A--G-----C-----G-----C--GA----T-------AC-CAGCAT-------TC----AGAAA-A-AG---G--TTAAA-C-GC-GACG--TGTC--G-----------CC-C-GG------- 3342
VER.DE.x.AGM3           AG----T---C-AGGG-----T-C------------T---CTGT---------A---------G-----C---A-T--T--T--G-AC-CAGCAG-TTCC--TC----AGAAA-A-A---GG-GTTAAA-CC-C--ACC--TGTG--------------CC-A-GG-----G- 2845
TAN.UG.x.TAN1           AG----T----TAGGG--T--A-----G-----A-T----TGT--A--T--A------------------CA-C-----T--GCAC-CAGCACAGCG-G--C--A-AGAAA-AG-----A-GGT--A---TT---AA--AAT---G-----------CA---GGA------ 3322
SAB.SN.x.SAB1C          AG-C--T---AG-GAG-----C--A-----------T--G-----T--G-----C--------G--------C-----CA----T--T--G-CA-CAGCC-AC-----TC-GC--GAA-----GC--A-G--C-----TG---A---A-----------------A---TAA---CCA 3518
MND_1.GA.x.MNDGB1       AG-T--T---C-AG-------A-AA-----CAT---T-----T-----T--A--G-----G--AGT---TGT--C--T---G-G-CAG-AG--TC-C-GC-G---GTA-------GA-CC-----ACAG--CAGT-A----------------------T------G- 2807
DEB.CM.99.CM40          -G-C--T---ATACA------AGAA------G---T--GA--A--G--T-----------GC-CT---------C--T--T---GCCTCAG-AGGG-G-CAG---C--ATC-----GG--A-G---------GC--A-A--AGTC--G--T-----------ACTA-----G- 2929
DEB.CM.99.CM5           GG-G------ACACA------AGAG------G---T--GA--A--A--T--A--C----GT-C---------C---------GGCA-CAG--GGC-GGCAGC-GA-A-T-----AGG--A--T-C------T--A----AGTC-----T--------CC-A-TG------- 2923
TAL.CM.01.8023          -C-G--T-----C-G--------GGC-G--C----T---G--A--C--C--A--G--------------CA-C-----T---TAC-CTTCAGATC-GC-G---AG-AAG-----GG--A--T-C---C-GG---CGC---TC--G-----------CC-ACTAA-----A 2587
TAL.CM.00.266           -C-G--------T-GG-----A-GA--------A-T---G-----A--C--A--------G--T--G--CA-C--C------CA--C-TC-GATC--T-A--GAG-AAA--C--GG-CA-GT-C--TC-GG---CAC---T---------------CC-A-TG------A 3089
SUN.GA.98.L14           AG----T---C--G-------A-AA-------GT---T-----C--A--------C----------AGT--TTGC--C--T--GG-A-CGG-AG-G-GCC---------AAA--------G---GT-------AG-TCAAT-G--C-----------------ACTCA------ 3477
WRC.CI.97.97CI14        AG----T------T-T-------AG------GTA--T-----A--G--T-------------------T--TTGT--T------TCA-CTG-AG--TCTC-A-----AATA------GG-A-G------A-T---CAAG-TCT------T---------C-A---A-T---- 3420
WRC.CI.98.98CI04        AG----T-----AT-T--T---AG--G---GTC--T-----A-----T--A--G-----G--T--C--TTGC--T-----GCA-CAG-AG---GTT-G-----TGTA--C-AG-----G-----TACT--GCAG--TCTC--G---------------A---A-----A 3435
WRC.GM.05.Pbt_05GM_X02  ------T-------A------T-CC------GTA--T-----A--C--T--A-----------G--T--TTGC-----T---GCA-CAG--G-TTCTT-GC-G---ATC--------GA--C----TA-AG-GCAG--AGTC-----T---------C------------ 2349
LST.CD.88.524           AC--------C--GA------A-AA-----------T------T-G-----A--------G--GAGT--TTGT--T--T---GAA-CAG-T-GTGG-C--C-T-CAGAG--C--G---TT-------------C-CC----C--G---------------TCA----T- 2385
LST.CD.88.447           ---T--T--CC--GA--------GA-----------T-----A--A--C--A--------G--GAGT---TGC--T--T---GC--CAG-AG--GG-C--C--TCAGAA----------TT---C--T------A-CT-G--------------------A-TCA------ 2388
LST.KE.x.lho7           -C-C--T---C-AGAG-----A-AA------------------G-----A--G-----G--GAGC--TTGT-----T--GG-A-CAG-AG--GG-C-TC-CTCAGAG-----G---TT-----------GA-C--T--C-----T-----------A-TCA------ 3472
SYK.KE.x.SYK173         C-----T----CCT-C--G--AGAA--------A-TT-CA---T-A---------------------T--CA-T--C---------C-CA--A-ATC-G--AC--C----A--------GA-GT--T----TC-CAC-C-AATA--G-----------C--A-TGA-T--CA 3314
SYK.KE.x.KE51           ----------TCTT-T-----AGAG-----C----T---G-----T---G--T--G-----A-----G--CA-T--C--T--GGCA-CAG-A-ATC-G--AC-CC-A---A-A--GG--A-G---T----T---A-A--AGTA--G-----------CC-T-TG-----G- 2972
COL.CM.x.CGU1           AG-G------TT-G---------C-------GT------A-----A--G-----T---------G--T---A----C------G-A-CAG-AG---G----C----A--G--C--G--GAG-GGA-A-TTG...CAA--AGC---G-------------A---C------ 2892
```


PLV Complete Genomes

```
H1B.FR.83.HXB2            CTGACTTAGAAATAGGGCAGCATAGAACAAAAATAGAGGAGCTGAGACAACATCTGTTGAGGTGGGGACTTACCACACCAGACAAAAAACATCAGAAAGAACCTCCATTCCTTTGGATGGGTTATGAACTCCATCCTGATAAATGGACAGTACAGCCTATAGTGCTGCCA 3290
Pol                       S  D  L  E  I  G  Q  H  R  T  K  I  E  E  L  R  Q  H  L  L  R  W  G  L  T  T  P  D  K  K  H  Q  K  E  P  P  F  L  W  M  G  Y  E  L  H  P  D  K  W  T  V  Q  P  I  V  L  P
H1A1.UG.85.U455           ----T--------------A------G-----------AT-A---GCT-----A-----C------T-C-TT--C-----------G--------------------T----------G--------T-------C-----------T---------CA------- 2736
H1B.US.90.WEAU160         ---------------------------------------------------------------------------T-----------------------A-----C---------------------------------------------------------------AA------- 3289
H1C.ET.86.ETH2220         -------------------A------G-CCC------A--T-A---G------A--A-A------T---------------G-------------------T----------G--------T-------C--------------------CA------- 2682
H1D.CD.84.84ZR085         ----T-----------A---------G----------A-AT-A---G-----------------G-----------------------------------T-------------------------G-----------T----ACA------ 2812
H1F1.BE.93.VI850          ------------------------------------T-A---G------AC---A-----T---T---------------------------------------GC----------------------G--------CAAT----- 2627
H1G.SE.93.SE6165          -----------------------G-------------T-A---G------A---AA----GT-----------T---------------------------------A-----G----------C--------G----A-----CA------- 2687
H1H.CF.90.056             -------------------A------G-------------T-A---GCT---T-----AA------T--------------------------C-----T---------A-----------------C---------------A--G--AAA------ 2637
H1J.SE.93.SE7887          -------G-------AA---------G-------A---------GG------A---A-------T------------T----------------------T----------A-----G--T-------C----------C--A-----CAA------ 2604
H1K.CM.96.MP535           -------------------A-C-----------------A--A---G------A---AA-----T-----------------G-------------C-----T----------A-----------------C--------------G--------CAA------ 2486
H101_AE.TH.90.CM240       ----T----------------C-------------------A---GCT-----A-----C------T----T-----------G--------G--------------------A-----------------C-G---------C----------AA------ 2864
H102_AG.NG.x.IBNG         ----T----------------------G-------------T----GG-----AC---AA-----T-----------------G--------------------T---------------------------------------C------G---AA--A--- 2815
H1N.CM.95.YBF30           -----------C---CA--A------GAGGC-G----A--C--T---G-T-----T----A------CT----G--C--T-----------------G-----C--G-----C--------A-------------A--C---------C------A---AA-T-A--- 2884
H1O.BE.87.ANT70           -A--T---CCCC-GACAG-A-------A--GG--T--ATT---T---G----C--A-ATCA-------T-C--T--C--T-------G-----A--G--------C--T--A-------G-----G-------A--C----------------T-C--CCAAT----T 3345
H1O.CM.91.MVP5180         -A--T---CC-T-G-CAG-A-------AG-GGG-C--ATT---T--GG-----T-A-ATCA-------T-----T--C--T--T-----G--------G--------C--TT-A-------A-----G----C--A--C--G--------------C--CCAAT----T 3320
CPZ.CD.90.ANT             ----TA-G-----TACTGCA------GA--TG-----AA----T---------T-ACA-GTC--------AGAG--T--T--------G-------------------T-AA--------A-----GT-A----A--C--------T------AAA---AA---A--- 2729
CPZ.CM.01.SIVcpzCAM13     ----TC-T--C--TACC--A--C---CAG--GG----A---T-A--G------T--C-TCAC-----GT----A--C-----T-----G-----A--------A--C---T-A--------G-----G--T--C--A---------------G-----A---AA---A--- 2870
CPZ.CM.05.SIVcpzEK505     ----TC---------ACA--A------GAGGC-G-G--AAGA--T--GAGT-----T----CT-------T----A--C--T---------G--------------A-----TT-G--------A-----------------A--------------C---A-A---CAAT-A--- 2862
CPZ.CM.05.SIVcpzMB66      --------A------A--TA------GA----G-G--A--AT----GGCT--C--A----AA-----TT-C-----T--T---------G-----A--G--G--A-----T--C--------A------------C--A--C-----------T----------CA-T-A--- 2836
CPZ.CM.98.CAM3            ----T--G---C--A-AG-A------GA---GG-----CTA--C-----G-----AC-T-CT-----GT----T--C--T--T-----G--------G--G--A--T--TT-G--------C-----G-------------G-----------G-----A---CA-T-A--- 2701
CPZ.GA.88.GAB2            -A--T---TCTT-G-AAG-A--C--G-A--G-G----AC-A--C---G-----T-AC-C--A-----TT--------C--T-----G--G--C--A--G--------T--T--C--------G-----G--T-----A--C-----------G-----A---CA-------T 2683
CPZ.CM.98.CAM5            -----C-G--G-----AA-A------GA---GG----ATTA--C--------------C-T-CT-----TT----T--C--T--------G-----A--G-----A--C--TT-G--------A-----G---------A--C-----------G-----A---CAAT-A--- 2992
CPZ.CM.05.SIVcpzMT145     -A--T---CCT----A-G-A-------AG-G-G----A-----C---GC----------A-AA-------T----A-----------G--G-----------------C-----T--C--------A-----GT-A-----A-----G--------C----AA---AAAT----- 2832
CPZ.TZ.00.TAN1            -A--TC-TA-TGA--A-G-A-----G-A-CTG---A--A----------G--------A-CA-------T-AGAG--C--T--------GT-----G--A-----------A-G---------C-----G--A-----AA------------T--AAA---CACAT-A--- 2934
CPZ.CM.05.SIVcpzLB7       -A--T--GA------A---A--C--G-AC--GG----A----------GCT--CT--C-A-A------GT--------C--T-------------------G-----A-----T--C--------A------------------A--C-------------------C---CA-T-A--- 2845
CPZ.US.85.CPZUS           -A--TC--A-CT---AAA-------GGA---GG----ACT---C------T--T--C-T-CT------T-C--T--C--T--T--G--G-----G-G-----A-----T-A--------C-------------A-----G--------G-----C---CA-T-A--- 3350
CPZ.GA.88.GAB1            ----TC-T-----T-AT--A-------A---GG-G--A--A--A---------T--C-T-AA-----GT-C--A--C----------------A--G--G--A--C--TT-A--------A-----G-----C--A--C-------------C--------TCAAT-A--- 3351
MON.CM.99.L1              -AA-TCACTCTC-GCCAG-A--C--G-A--T-G---G--TCA-A---GCGACC--AC-A---AAA--GA--CAA-----CCCAG-G---TT----G-T--CTAC--GG-G-AA---T-A--G-----G--A--C--ATCCC-T----A-A-TGCCAAA----A---A--- 2996
MON.NG.x.NG1              -A---CATTCCC--CCTG-A--C--G-AG-T-G--ACA-CCA-T---ACG-TR--A-CAGAAAA---GA--CAA------ATAG-G---T-----G-CA-GTAC-----G-AG---T----G-----------C--CACCC-T--------C--AAAG----AA------ 1589
GSN.CM.99.CN166           -A--TAGGA--C-G--AG-A--C--GCA-GT-G-GC-CA-AA-A---A-TTTA--AACATCC-ATAACA-CCAA--C--C--GGCC-----C--ACC---TTA---GG--AAG---C-C--G-----GT-G------A-GGG-----GGC-C-----GG-G-AA--C--- 3018
GSN.CM.99.CN71            -A--TAGGA--T-G-CAG-------GCA-GT-G-GC-CA-AA-A---A-TTTGT--AC-TCC-ATAACA-CCAG--C--C--GGCC-----C--AC----TTA----G--AGA---C-C--G-----GT-G--C---A-GGG-----GGC-C--A--AG-G-A---C--- 3000
MUS_1.CM.01.1085          -CA--AG----T--TCAAGC--C--G-A-GT-G---CAC--A-T---A-CATG--C--A-A-CTT--GA--CAA--C--TC-AG-----TT----G----CTAC---G---AA---T-A--G-----G--A--C--CCGGGGC----A-A-------CG---CAT----- 2980
MUS_1.CM.01.CM1239        -CA--AG---G---TC-GGA-----G-G-GT-G---CAC--A-T--GA-CATG--C----AACT---GA-ACAA--C--G--AG-T--GT-C---------CTAC---G---AG---C-A--G-----GT-A--C--CAGAGGG----A-A-------AG--ACA--C--- 2970
MUS_2.CM.01.CM1246        -A---AG---------AG-A-------GGCT-G-GA-AA--A-C---G-T-TC--C-CA--TAA---GA-ACAG--C-----AG-C-----C---CC---CTAC--GG-AGAA---C-A--G--C--GT-A-----GA-AGG-----G-A-CA-A--AG-G-A-------- 3078
MUS_2.CM.01.CM2500        -A---AG---------G-A-----G-GGCTGG-CA-AAGAG-T---G-C-TC--AACA--CAAA--GA-ACAA--C-----AG----G-----C----CTAC---G-AGAG---C-A--G-----GT-A--C--A-GGGGT----G-A-C-G---AG---AA--A--G 3037
RCM.NG.x.NG411            -A--TAGGAC-GC-CATG-A-----TCAG-TG--CA-AAG---C---G-G------C---CC-----TT--GAG--T-----T--G---TT----G-CA-G-----C--TGAA---------C----TG-------A--C------------AGAG--CAA-T-A-T- 2784
RCM.GA.x.GAB1             -A--TAG-AC--AGTT-G-A---GAC-A--TG--CA-AC-AT-A---G-T-----AC--TTC-----CT--GAG--C------------TT----G-TA---------ATT-G---------G-----G--G--C--AA-A-GT------------GAG--CAA-T-A--- 2769
DEN.CD.x.CD1              -A--TAG-TCCC---AA-----C--G-A--TGG-TA----A--A---G-CTTCT--GCAGTA-----CT--GAG--C--T--TGCC---TT---AG--AC---A--CA-A-AA---------G------T-A--C--CA-AC-C---CA-A-----AAA----AC------ 3177

MAC.US.x.239              G----AGGAC-GACCT-G-A---GAC-GGGT-G-TTTAC--TCA-AGG---TCT--AAT--CATA--GT--T-T--C-----AG-G---TTC--A-----T--C-----T-AA--------G--C---T-GTGG--AACA-------AGT-G--AAAG----A-T----- 3592
Pol                       A  S  D  R  T  D  L  E  H  D  R  V  V  L  Q  S  K  E  L  L  N  S  I  G  F  S  T  P  E  E  K  F  Q  K  D  P  P  F  Q  W  M  G  Y  E  L  W  P  T  K  W  K  L  Q  K  I  E  L
H2A.GW.x.ALI              G----AGGAC-GATTTAG-A---GAC-AGGTGG-CCT-C--T-A-AGG---TC--AAATG-CCTA---T--T----C-----TG-G--GTTC--A-----C-------A-AAA--------C----G---GTGG--AAC--------AGC-G--AAAA---CAAT----C 3668
H2A.DE.x.BEN              G----AGGAC-GGTTTAG-----GAC-A-GTGG-CCT-C----A-A-G---T---AAATG-CCTA--GT--T-T--T-----TG-G--GTTC--A-----C--------T-AA--------C-G------ATGG--AAC--------AGC-G---AAAC--CAA-----C 3675
H2A.SN.x.ST               GC---AGGAC-GATTTAG-A---GAC-G-GTGG-TCT-C----A-AGG---T---AAATG-CCT----T--T----C-----TG-G--GTTC--A-----C-------A--AA--------C---------GTGG--AAC--------AGC-G--AAGA---CAAT----C 3120
H2B.GH.86.D205            G----AG-AGTGATCT-G----CGAC-GGGT-G-GTCCC-A--A-A-G-GTTAT-AAATGACAT----T-CT-T--C-----AG----GTTC--A-----C-----G---AAA--------------G---TGG--AA-A--G----A-C-G--AAAA---CAA------ 3648
H2B.CI.x.EHO              G----AGGAGCGATCT-G-----GAC-GGGT-G-GTCTC-A--A-A-G-G-TAT-AAAT-ACAT----T-CT-T--T-----AG----GTTC--A-----C----------AAA--------G-----G---TGG--AA-G-------A-C-G--AAAA---CA---A--- 3646
H2G.CI.92.ABT96           G----AG-ACTGATTTAG-A---GATCA-GT-G-TTT-C----A-AGG---TCT-AAAT-AAACR--GT-CT-T--C--T--TG----GTTC--A-----C-------AT-AA--------A------T-RTGG--AA-A--G----A-C----AAAG--TAAC-----M 3002
H2U.FR.96.12034           G---TAG-A-TGATTTAG-----GATCAGTT-G-TTCAC----C-AGG-G-TA---AAT-A-ACA---T---G---C------G-----TT---------C--A---CA-AAA--------C--------ATGG---A-A--G----A-T-G---AAA--T-A---C--- 3157
MAC.US.x.EMBL_3           G----AGGAC-GACCT-G-A---GAC-GGGT-G-TTTAC----A-AGG---TCT-AAAT--CATA--GT-CT-T--C-----AG-G---TTC--A-----T--C-----T-AA--------G--C---T-GTGG--GACA-------AGT-G--AAAG----A-T----- 3066
SMM.US.x.H9               G---TAG-AC-GATTTAG-A---GAC-GGGT-G-TTTAC-AT-A-AGG---T----AAC--TATA---T--T----C-----AG-G--GTTC--A-----T--C-------AG--------A------T-GTGG--CACC-------A-C-G--AAAA----A-T----G 3065
SMM.US.x.PGM53            G----AG-AC-GATTTAG-----GAC-GGGT-G-TTTAC--T-A-AGG---T----AAC--CATA---T--T----C-----AG-G--GTTC--------T--C-------AG--------A------T-GTGG--AACC-------A-C-G---AAA----A-T----- 3509
SMM.US.x.SIVsmH635F_L3    G---TAG-AC-GATTTAG-----GAC-GGGT-G-TTTAC--T-A-AGG---T----AACG-CATA---T-CT----C-----AG-G--GTTC--------T--C-------AG--------A------T-GTGG--AACC-------A-C-G---AAA----A-T----- 3595
SMM.SL.92.SL92B           G-A--AG--GCC--ACAG-----GAC-A--TGG--ACAC--T-A---G-CATG--CAAC-ATCT---GT-C-GT--C-----AG-C--GTT---A--GA-T--A-----A-AG--------A---TTGT-AT-C--CA-G--G----A-T-----AAA----AA--C--T 3023
STM.US.x.STM              G----AG-AC-GATCTAG-----GAC-GGGT-G-TTTAC--T---AGG---TC--AAAT-AT-TA---T--T----C--T--AG-G---TT------G--T--------T-AA--------A-----GT-ATGG--AACA--G----A-C-G--AAAA----AAT-A--T 3239
MNE.US.x.MNE027           G----AGGAC-GACCT-G-A---GAC-GGGT-G-TTTAC--T-A-AGG---TCT-AAAT--CATA--GT--T-T--C-----AG-G--GTTC--A-----T--C-----T-AA--------G------T-ATGG--AACA-------AGT-G--AAAG----A-T----- 3060
DRL.x.x.FAO               G---TAG-CC--A-CATTTA---GAGCAG-TG--TA-A--AT-A---AC----T-AAA-TTC-----G--CGAG--C--G---------TT----G-CA-T-----T-ATGAA--------G--CC-G--G-----AA-G--G--------G--AGAA--TCGCT----T 2857
MND_2.x.x.5440            G---TC--AGTGCCTCTG-A--C--T-A--TG----TA--AT-A--GG-----T--CA-TTC-----G--CGAG--C----------GTTC--A--G--------C--TGAA--------A----TG--G--C--AA-G-----------T---AAA---CA---A--- 2851
MND_2.CM.98.CM16          G---TAG--TTGCCCCAG-A--C--C-G--TG--TC-A---T-A---G-C--C--C---TTT-----G--CGAG--C--------G--GTT---A--G-----A--C--TGAG--------A--CAT---G--C---A-G------------G---AAAG--CA---T--- 3291
MND_2.GA.x.M14            G---TC-GAGTGCGTCAG----C-AC-A--TGG--A-TAG-T----GG-----T-AAG-TTC-----G--CGAG--C-----------GTTC--A-----------T--TGAG--------G----TG--G--C--AC-G------------C--AAAG---CA------- 3218
OLC.CI.97.97CI12          -A-----GCC--G--CAG-----TT-GAG---G--A-AC-------GG-G-TCT--AG---T-ATAAC--AGAA--T--TCCA-----GT-----G-G--T--A--T-ATGAG--------G---AC-T-AT-C---A-G-G-----AGA--GCAAGA---AA-T----- 3164
GRV.ET.x.GRI_677          -A---CAT--TGAGACTAGA----ATCA-C-GG----CATAG-A---A-GATG---C-AGAAAAA--T--AGAA--C--------G---GTC--A-G------G---GGGAA--------G---A--T-G-----GA----------CA-TA-CAAA----AAT-A--C 3463
VER.KE.x.TYO1_patent      --C-AGA-A-TGA-CACACC---GAC-A-TT-G----AC--T-A---AC-A-AT-ACAAGCC-----CT-AGAA--C-----A--G--GGTG--A---------A--T-ATGAG--------A--CA----TTGG---C-C------GA-C--AGCAGA---CAA---GAG 3517
VER.KE.x.9063             --C-AGA---TGA-TACACC---GAC-AGTT-G----AC----T---GTCA-G--AAAA-CA-------T-AGAA--C------------GGTG--A--GA-G--A---ATGAA--------A--CA--T-ATGG--AC-A-------CA-C--A-TAG----AACT-AGA- 3520
VER.KE.x.AGM155           -AC-AGA---GGGTCCAA-----GATCAGCT-G--C-AACA--T--GA-TAGAT--CAAGAA-------T-AGAA--------G-----GGTG--A-G---------C--TGAG--------A---A--T-ATGG---C----------AGT----AAG-----AAT-AGAG 3512
VER.DE.x.AGM3             --C-GGA---TGA-TACAC----GATCGGTTGG----AC-A--A---ATGA-AT-AAGTGCC-------T-AGAA-----------G---GTA------A----A--T-ATGAG--------A--CA--T-GTGG--AC-C--G---CAGA--AGCAGC----AAT-AGA- 3015
TAN.UG.x.TAN1             -A--T--GCC-GAGAAAG-A---GATGA---GG----A---G-A---A-G-T----C-ATTAA-A--CT-CGAG--G-----T--G--GGTG------AG---------ATGA---------A--CA-GT-GT-C---TC-C-C----GTC-CA-CAAA----A-A-A--- 3492
SAB.SN.x.SAB1C            G---TAGGCC--AG-CAG-A---CT-GT--TGG--C--C--T-A---G-CT----AGAA-CT-----GT---AA--C--T--A--G--GTT---A--G--T--A----A---C--------G-----T-GT----AA-G------CA-C-G---GAG--CACC--A--- 3688
MND_1.GA.x.MNDGB1         -A----ATAC-GC--AAG-----GAG-A-GCT----TA--AT-A--GGCTTTAT-AA---CA---AACT-AGAA-----T--A--G---T----------------C--T-A---------G-----GT-A--C--A-----G----AGA--G-AAAGG-TCAA--A--- 2977
DEB.CM.99.CM40            -A--T---AGG-A--AT--A--CCT--AG--GG---CAC-AT-A-----GTTC--CC-AGA-A-----T-G-GG------CCTG-G--GT-----G----C--A--C--T-AC--------G-----GT-A--C---A-A------CGGT-GAT---AG-TACC--A--- 3099
DEB.CM.99.CM5             -AA-T-----T-AG-TCTCA---TT-GA-C-GG--A-AC----T--GA--TT---CC-AGAAAA--GT-A-GG-----CCCGG-G---T-C----G-G--T--C----A--A---------G-----GT-A-----AA-G------CACT-GTT---A----A---A--- 3093
TAL.CM.01.8023            G---TGA-AGCT---AA---------G-G--T-G-G-GA-----T---GCC-TC--TAAC-A-GTA--GA-CCAG-----G--GGCC---TTC---G-CA-G-----TA-TAAG---C-T--G--C--G--GT-C--A--G-------A------AAA-----A---A--- 2757
TAL.CM.00.266             GCA-TGAGAGCT-G-A-G-A-------GG-T-G---GA--A--A---GCT-TCT-AAAC-A-GTA--GA-CCAG--C-----GGCT---T-C--AG-CA---G--CA-TAAA---C-A--G--C-A--G-----T-GT---A----G----A------------T-AAT----- 3259
SUN.GA.98.L14             ----T-AT--G-A-AA-A-----GA-GA--T-G-GA-AC-AT----G--G-TA--CA--GAA---AATT-AGAA--------G------T-----GG---G--A--C-ATAAG--------G--CAT-T-A--C--A--C-G---GAGA--G--AAG---AAAT-A--C 3647
WRC.CI.97.97CI14          -A-----G--C-AGACAG-A---CT--AC--GG----A--AT-A---A-G-TA-------CA---AATT-GGAG--T-----G-----G-TA---TT---------C-GTAAC--------A---AT-T-G-----A-----G---CA-A-----GAG---AA-T----- 3590
WRC.CI.98.98CI04          G------T----A-CCAAG----TT-GAG---G----A---T-A-------TG-------CT---AATT-GGAA--T-----A--G---TTG--ATT---G--G--C-G-AAG------------ATT--A----------G----CA-A-----GAA---AAA--A--- 3605
WRC.GM.05.Pbt_05GM_X02    -A--T-AT----AGAAAG-A---TTGCA--TG--TC-----T-----G-T-T---T--A-CT---AAC--GGAG--T-----A--G--G-TA--ATT-----------GTAAA--------G---ATCT-G-----A---------GA-A-----GAG---AAAT-A--- 2519
LST.CD.88.524             -A---AGG--G-A--AAGGA-----TCAGGC-G--A-A--AT-A---G---T-T-AA-A-CT---AACT-AGAA--T-----G--G---TT---AG-CA---A--C-A--A---------G----TC--G--C--A---G---GA-A--G--AGAG--AAA------ 2555
LST.CD.88.447             -A--TAGG--C-AGAAAGG----GATCAGGC-G--A-A--AT-----G-T-TCT-AG---CA---AATT-A-AA--------G--G---TT---AGC---------C-AT-A---------G----T---T-----A---GG---GA-A--G--AAA-CTAAA------ 2558
LST.KE.x.lho7             -A--TAG---G-G-AAAGGA---GATCAGGC-G--A-A--A--C---G---T---TA---CA---AACT-AGAG--------A--G--GTTC--AGC---G--A--C-AT-A----------------T---G----------GG---GA-A--G-AAAG--TAAAT-A--- 3642
SYK.KE.x.SYK173           ----TAG-TC-GAGAA-GCA---CAGGAG-T-G--C-AC-AA-AGTGAC-GCA--A--A-AAGT----T---AGGTC---A-AG-----TGG---G-CC--TA----A-G-AG---T-A-----CACT--T-----A--C--G---CAGT-G---AAA----AA--C--T 3484
SYK.KE.x.KE51             --A--AG-TC-GA-AA-G-A---G-TCA--TTG--AGT--AA-TGT-A--AG---A--AGCAGTA---T-CT-T-TT---CCAG-----TGG---G-CA--TTC--C--G-AA---T-A--G--CACG--A--C--------G----GTC----AAAGG-TCAA--A--- 3142
COL.CM.x.CGU1             -A--TC-GCCTT---A-G---------GGGC-G--C-------C---G-TGCA--C---TAC-AT--GT-AGAG--------A--G---T-C---GC---T-------A-AAG--------C------T-A------A-GC------A-T----AAAGG-CAAAA-T--- 3062
```

```
H1B.FR.83.HXB2          GAAAAAGACAGC......TGGACTGTCAATGACATACAGAAGTTAGTGGGGAAATTGAATTGGGCAAGTCAGATTTACCCAGGGATTAAAGTAAGGCAATTATGTAAACTCCTTAGAGGAACCAAAGCACTAACAGAAGTAATACCACTAACAGAAGAAGCAGAGCTAGA 3454
Pol                     _E__K__D__S__.__.__W__T__V__N__D__I__Q__K__L__V__G__K__L__N__W__A__S__Q__I__Y__P__G__I__K__V__R__Q__L__C__K__L__L__R__G__T__K__A__L__T__E__V__I__P__L__T__E__E__A__E__L__E
H1A1.UG.85.U455         ------------......--------------T--------A-----A--A---C-A-------------------TG---------------A----C-G---------T--C--G---G----------------TA--G--A-C--G--T-----------AT---- 2900
H1B.US.90.WEAU160       --------A--T......---------------------------------A--------------------------G--------------A----C-----------------G--G-----------------A-------A------------------------ 3453
H1C.ET.86.ETH2220       -----G--T---......--------------T-----A-----------A-----A--C-------------------------------C-G-----------------G---G----------------CA--G--A-------T-----------AT---- 2846
H1D.CD.84.84ZR085       --G-----A---......--------------T-----------------A-----A-----------C-------T-----A----------A-------------------G--------G----------G-------------------------AT---- 2976
H1F1.BE.93.VI850        A-C--G------......-----------------------------A--A--C-A--------------------T--------------------C-------------------G---G----G-----------CA--G-G----G--T-C---G-----T---- 2791
H1G.SE.93.SE6165        --C--G--A---......--------------T-----A-----------A--C-A--------------------T----------------C---C------------------G--GG----------------CA--G--T----G----C---------AA-G-- 2851
H1H.CF.90.056           ------------......--------------T-----------------A--C-A--------------------T--AAT----------A----C------------------G--GG--------T------CA----------G---A----G-----AT-G-- 2801
H1J.SE.93.SE7887        --------AGA-......--------------T-----A-----------A--C-A--------------------T----A----------A----C-----------------A---GG-T---------------CA--G-----T-G--TAG---------AT-G-- 2768
H1K.CM.96.MP535         --C--G--T---......--------------T-----------------A-----A--------------A--A--T-----------------AA-----G-----------------G---GT---------------CA--G-----T----T-C-----------T---- 2650
H101_AE.TH.90.CM240     ------------......--------------T--------A--------A--C-A---------------A-----TG----------G---A----C-G-------------C--G---G-T---------------CA--G------G--T-----------AT---- 3028
H102_AG.NG.x.IBNG       ------------......--------------T--------A--------A--C-A---------------------TG---A------A--A---C-G----G------C--G---G-----------------TA--G--G----G--T--G---------AT---- 2979
H1N.CM.95.YBF30         -----G--TGTA......--------------T--------A-----A--A--G--A------------------C--T-----A--C-G-----AA--GC-C------T-AA-C---------------TT-G----------G-CAACT-T----------------AT---- 3048
H1O.BE.87.ANT70         A-C--G--TGTG......-----A--A-----T-----A--AC--A-A--A--GC-A---------------A--C--T-A---A----G---G--AG----G-----GT-AA-------C-------GT--T------------G----TT---GTAG---G---------G-- 3509
H1O.CM.91.MVP5180       --C-----AGTG......-----A--A-----T-----A--A-----A--A-----A---------------A--C--T-A---A----G-----AAG----G--C--GT-AA-C-----------------T--T-G-----G---G----TT---GTA----G-----A----- 3484
CPZ.CD.90.ANT           --GCC---TGAT......-----A--T---------C-----AC----A--A-----A--------------------C----------A--C---ACT-A---G--G----G----A-C------GT---GAGT--------TAG-G---A-A-G--TAGG---------AT---- 2893
CPZ.CM.01.SIVcpzCAM13   -----G--GGTA......-----A---------TC-T-----AC-CA-A--A-----A--C--------------------T-----A--A---A---AA--G--G--C---T-AA-A---------A---AGGT-------T---G-T--T--C---CC---------AT---- 3034
CPZ.CM.05.SIVcpzEK505   --G-------CA......---------------T-----AC-------A--A-----A--C-----T--C-----C----------------------AA---C-T--C---T-AA---------G-A-----T-------------G-CA----T---AG---------AT---- 3026
CPZ.CM.05.SIVcpzMB66    --------A---......---------------T--C-----A---A-A--A-----A----------T-------A--T-----A--C-G-----AA---C-G--C--G---A---------A-------TT----------TG--A-CT-T---AC---------AT---- 3000
CPZ.CM.98.CAM3          C-G-----G-C-......---------------T--T-----AC-------A--C-A---------------A--A--T-----A--A-------AA--G-----C--G--GA-A-A---G--A-----T--------CA--G-GA-CA-G---C-G---------AA-G-- 2865
CPZ.GA.88.GAB2          C-G-----G-AT......---------A--------T-----AC-G--A--T-----A--------------------C--T-----A-----GA---AA-----G------T-AA-C------G-T--G--TT---------A--G-CGAGA-G--TAG---------AT---- 2847
CPZ.CM.98.CAM5          C-G-----G-T-......---------------T--T-----A-----A--A-----A---------------A-----TT--------A-------AA--G-----C-------A-A-A---G--A-----CT-G--T------G--G-CA-G---C-G---------AA-G-- 3156
CPZ.CM.05.SIVcpzMT145   ---C-G--AGTG......---------T-----T--T-----------A--C-----A---------------A--A--T-----A------A-T-A---GC-T------T-AA----G---G-G--G-GT--G---------G--G-T--C--CCC------T--A----- 2996
CPZ.TZ.00.TAN1          ---CC---GCAG......-----A--G---C-T--C--------G--A--C---C-T--------C-----A-----T-AT--A--A---ACT-AAG--C-----C---T-GA--------GTA----G-T----T--GCC-G--GA-A-G--CAGG---------AT-G-- 3098
CPZ.CM.05.SIVcpzLB7     -----G--A--T......-----------------------A--G-----A-----A-------------------A--A---------C--C--GA---AA--GA----C-GG--T------G---GTA--GAATTC-----------G-CA-CT-T--TAG---------AT---- 3009
CPZ.US.85.CPZUS         C-------A-TT......-----A---------T--T-----A-----A-----C-----C---------------A--T-----A--A---A---A---G-----C---T-AA-A-A---G-T-----T-----T-----TG--AATT-T---C-T--G-----AA-G-- 3514
CPZ.GA.88.GAB1          --------GGTA......---------------T-----A--AC-GA-A--A--G--A--------------------T-----A--A---A---A-------------G--GA-A---------A--GAA---G-----T---G-T--T--C---CC---------AT---- 3515
MON.CM.99.L1            --CC-G--ACAA......----------------G--C--A-----GA-A--A--GC-T---------GCA--A--C--TT----A--A---ACC-AAA-CC-G--C---AGTA-C-----GG-ACGGAGC--G-----TAC-G-GGTC--CT----TTT---TC-AGCT-- 3160
MON.NG.x.NG1            ---C-----CAG......-----A-------GG--C--A--A-----A--------A--C------GCG-----C---------A--C--GAC--A---TC----C-R-TGTA-CC-R--GRTA-----C-----G--CA--G-GGAG---T-----CTT---CC-AGC--- 1753
GSN.CM.99.CN166         --CC-G--T-C-......-----A--G---------C-----A--G--A---------A-------T-TCA--A--AC---T--T--A-GGAC--AA--G--G--C---TGTA----G---G-T----GG--------CCAGG--GAG--C--T--G--G---C--A---- 3182
GSN.CM.99.CN71          --TC------T-......-----A--A---------C--A--A--G--A-----GC-C-------T-TCG--A--AC-T--T--C--A-GGAC--AA--G-----C-G-TGTA---------G---------GG--------CAGG-GGAG--C--T--G-------C-AA---- 3164
MUS_1.CM.01.1085        C--C----AGTT......-----A--T-----TG-G--A---C-G--A---------A---------GCA--AGCA--------T--C--GAC--A---------C---TG-A-C--G---GTA---T-T--G--T----AGG-C-AGT-G------G-------C--GC--- 3144
MUS_1.CM.01.CM1239      C-GC----A-A-......-----A--A-----T-----------G--A--C---------------GCG--AGCC--------AG-----ACC-AA--GC-----C---TGTA-A--G---GTT---T-C--G---------A-G-T-A--------G-----TC-AT---- 3134
MUS_2.CM.01.CM1246      --CC-G----C-......-----A--G-----A--T------C-G--A-----G--A--C------GCC--AG-C--TT--------C---ACT-AA--T-----C-G-TGTA-C--G----TA--G-GG--------GC--G-GGAG--C-GT--GC----CC-AT---- 3242
MUS_2.CM.01.CM2500      --CC----A-C-......-----A--A-----G--T-----AC----A--A---------------GCA------A---T----A--C--GACC-AA--TC-G--C---TGTA-C-------GTA-----C--T-----AC-G--GA---T--C--GC-G--TC--T---- 3201
RCM.NG.x.NG411          --------AGAA......-----A--A-----T-----------A--G--A---------A-----------------A--TT----A--A--GAC--A-G-G-----C-----AA-A--G---GTT--G---T-GGAT--CAAGG--GA-T-T---AG---------AT---- 2948
RCM.GA.x.GAB1           --G-----AGAA......-----A--T--C--T--T-----A-----A--A--G--A--C-----------A------AGT---C-A-GGACT-A-G----G--C--GT-GA-A--G---G-A------T--GAT---AA-G--GA-A-G---A----------AA---- 2933
DEN.CD.x.CD1            TTGCC---GGATGGGCAG-----G--G---CA--T------C-----A--AGT-------------------T-G---T----A--A--GAC--AAG-----------GAG-A---------GTA----ACT--GG----ACC--TT-----C--G--G-G-------GC--- 3347

MAC.US.x.239            C---G---G-C-......-----A--G-----T---------------A--AGT---A---------GC---A-----T-----T--A---ACC-AA--TC-C----GGT-AA---------AA-TGA-T--------G-A-G-T-AGTGG--T--GATG-----AGC--- 3756
Pol                     _P__Q__R__E__.__.__T__W__T__V__N__D__I__Q__K__L__V__G__V__L__N__W__A__A__Q__I__Y__P__G__I__K__T__K__H__L__C__R__L__I__R__G__K__M__T__L__T__E__E__V__Q__W__T__E__M__A__E__A
H2A.GW.x.ALI            C-G-----AGTA......-----A----------C--A--ACAT----TGTCC-A---------GCA--A--C---------A--A--GACC-AA--C-------G---AA---------AA-TGA---C-------G-G-G-AGTGG------CT------AGC--- 3832
H2A.DE.x.BEN            C-G-------TA......-----A----------C--A---C-------AGTC--A--------GGCA--A--C--TT----A--A---ACC-AA--C-------G---AA---------AA-TGA---C-------A-G-G-AGTGG------CT------AGC--- 3839
H2A.SN.x.ST             C----G--AGTA......-----A----------C--A--AC-G-----TGTCC-A---------GCA--A--C-------------A--GACC--AA-C-------GGT-AA-C-------AA-TGA---C-------AGG---AGTGG------TT------AGCG-- 3284
H2B.GH.86.D205          --------AGTT......-----A--G----CA--T--A--AC-G--A--AGT---A--C------GC---AC-C-TT--T--A-----GAC------CA----C-----AA----G----AG-TGA-C-----------A-G---AGTGG------CT------AGC--- 3812
H2B.CI.x.EHO            --------GGTT......-----A--A--------T---------------AGT---A---------GC---AC---T---G---------GACC-----TA-----------AA-A--G----AG-TGA-C----------AGG---A-TGG--T---TTG------GC--- 3810
H2G.CI.92.ABT96         --------G-CT......-----A--A-----T--T--A--AC----A---GT-C----Y------GC----C-A--T-----A--C---AC--AAA-CC-G--Y---T-AA-------T-AR-TGA-CT-G--G----A-G---AGTGG------TTG-----AGC--- 3166
H2U.FR.96.12034         ----G---TGTT......-----A--A--------T-------------A--AGT-C-A---------GCA--AG----T-----A--A--GAC--AA--TC-T------T-AA---AG--G-AA-TGA----C-------AGG---AGTGG------ATG-----AGC--- 3321
MAC.US.x.EMBL_3         C---G---G-C-......-----A--G-----T---------------A--AGT---A---------GC---A-----T-----T--A---ACC-AA--TC-C----GGT-AA---------AA-TGA-T--------G-C-G-T-AGTGG--T--GATG-----AGC--- 3230
SMM.US.x.H9             C---G---A-C-......-----A--A---A-T-----A--A-----A--AGTGC-A---------GCA--A-----T-----A--A--GACT-AA--TC-T--C-G-T-AA-C--------AA-TGA-TT---------AGG-T-AGTGG--T--GATG------GC--- 3229
SMM.US.x.PGM53          C---G---A-C-......-----A--A-----------A--A-----A--AGT-C-A---------GCA--A-----T-----A--A--GACT-AA--TC-T--C---T-AA-C--------AA-TGA-TT---------AGG-T-AGTGG--T--GATG-----AGC--- 3673
SMM.US.x.SIVsmH635F_L3  C---G---G-C-......-----A--A-----------A--A-----A--AGTGC-A---------GCA--A-----T-----A--A--GACT-AA--TC-T--C-G---AA-C--------AA-TGA-TT---------AGG-T-AGTGG--T--GATG------GC--- 3759
SMM.SL.92.SL92B         -----G--GC-A......-----A--A-----------A--AC-------AGTTC-A---------GCA----A--T-------------AC--AAA-CC-G--C---A-GA-C-----G-AA-TGA-T--G-------A-G-T-AGTGG------TT------AGC--- 3187
STM.US.x.STM            C-G-G---GTT-......-----A--G-----T--C--A--AC-G--A---GTC--A--C------GCA--A-----T-----A------ACT-AA--TC-G--C---T-AA-C-----T-AG-TG-----G--T--G-AGG---AGTGG------ATG------GC--- 3403
MNE.US.x.MNE027         C-------G-C-......-----A--G-----T---------------A--AGT---A---------GC---A-----T-----T--A---ACC-AA--TC-C----GGT-AA---------AA-TGA-T--------G-A-G-T-AGTGG--T--GATG------GC--- 3224
DRL.x.x.FAO             --T-----A-TT......---------T-----T--T-----A-----A--AGTGC-T--------T-----A-----TT-C--A------AC--A-G-GC-C------T-AA---G---G-----C-C---GAT----A-G--GA-TGG--TAGG---------A----- 3021
MND_2.x.x.5440          --------A---......-----A--A-----T--T-----A-----A------C-T-----------C--A--C--TT-C--A------AC--AAG-GC-C------T-GA-A--G--GG-A-G-C-C---GATA--A--G-GGA-TGG--TAG----------A----- 3015
MND_2.CM.98.CM16        --G-----AGAA......-----A--A--------C--A-----G--A------C-T--------------------TT-C--A------AC--AAG-GC-C----G---AA-------GG-A---C-C---GAT---AA-G--GA-TGG--TAG----------AT---- 3455
MND_2.GA.x.M14          --------G-CA......-----A--A-----T--T-----AC-G--A--A---C-C--------------A--C--TT-C--A------AC--AAG-GC-C--C---T-GA-C-----GG-A---C-T---GAT------G--GA-TGG--TAGG---------AT---- 3382
OLC.CI.97.97CI12        --------AGAA......-----A--------T--------A-----A---C----A--------TTCA--A--C--TT----A--A---ACG-AAGCT-----------GA---AG---C-T----GGT-GCT---TAA-G-GGAGTGG---------G---A--GA--- 3328
GRV.ET.x.GRI_677        CCCTT---AG-A...GAA-----A--A--CA-A---------G----A--AGTTC-A---------------A-----T-----A------ACC-AA--TACC--GCCA-GT-G-----G-AA--GAAC--CCT-----A----GT-TGG---------G------GC--- 3630
VER.KE.x.TYO1_patent    --------TGAA......-------------------C--------------T------C-A---------GCA--AT-G--T-----TC---GGACC-A-A-TA----C--GT-AA---------AG---AAT--GTT---GC--G-GA-TTGG---CCT--G-----AGCT-- 3681
VER.KE.x.9063           -----G--ACAA......---------T-----T--C--A--AC-----A-----------C------GC----C-------C---T-A-G-AC--AA--GA----C-GGT-AA-C--G----AG---AAT---TT---G-A-G-G-A-TGG--T------G--G--AGC--- 3684
VER.KE.x.AGM155         A-G-----ACAA......-----A--G-----TC-T-----A--G--A---------A---------GCA--AT-A--T-----AT-G-G-AC--AAA-TA-C------G--A-----------AG---AATT--TT---C--GG--GA-TGG--CCC---G-----AGC--- 3676
VER.DE.x.AGM3           --C-----AGAA......---------A-----T-----A-GAC-------------C-A---------GCA---C-----------AC-C-G-ACT-AAA-C--G-------T-AA-C--------AA--GAACT--CT----AC-G--A-CTGG------G---------AGC--- 3179
TAN.UG.x.TAN1           --GC----ACA-......-----A--G-----T--T--A--A-----A--AGTG--A--------T-----AT-A--T-----A-----GAC--AAGCCC----C---T-A--AC-----G-TC---ATT------CT---G-G-A-TGG--G--GC-G------G---- 3656
SAB.SN.x.SAB1C          ----GG--AGAA......-----A--G-----T--T--A--A-----A--A--G--A-----------------A-------A-------A--GACT-AA--T--G----G---AA---------G-T-G-C-TT-G-----GA--G-C-A-TGG---------G-----A----- 3852
MND_1.GA.x.MNDGB1       ---TT--CAGAACAG...CCA--A--A-----A--------A--G--A--T-----A--------TGCA---T-A--T--T-----C---AC--AA---C-G--C--G--AA-A------GGACT-AACA-------GAA-G-CA---A-G------------AGA--G-- 3144
DEB.CM.99.CM40          --TG----AGAA......-----A--GC--A-A--T------------A--AC-G--A------------C--A-----TT-------A---AC--AA--C--------GGGCAA-C-----GGT-CCG-G-T--------CCCGG-TA-C--CT-T---------TC-AGC--- 3263
DEB.CM.99.CM5           ---G-G--GGAG......-----A---C-CA-A--C-----AC-G--A--AC-G--A-------------------A---T----A--A--GACC-AA--C--G--C--GGCAA-C-----GGTTCC--GG--------TAA-G--AGC--TT-----G-------C-AGC--- 3257
TAL.CM.01.8023          ------C-G-CA......-----A--A-----TT---------C-G--A--A---C-A------------------A--A--TT-------C---AC--AAGCCC----C---TG------G--TGTA----GCT-GTT------AGG-TA--T-G----------------AGC--- 2921
TAL.CM.00.266           --T---CT--CA......-----A--------GT-----------A-A--A---C-T----------T-C--A--A--TT-T--T------AC--A-GCCC-C--C-GGTA---C--G---GTA----G---TCT-----A-G---AGT--T----G----------GC--- 3423
SUN.GA.98.L14           CCTCT---AGAAGAG...CCA--A--G-------------AA-T--A--AGTTA-C--------TTCA--AT-A--TGA----C-C-G-AC--AAG-------------AA-------G-AT---C--T--G----GAA-G--AAGA-G------------TAG-GA--- 3814
WRC.CI.97.97CI14        ---G---TA-AG.........--A--A-----TG-------A--G--A--A-----A--C-----TGC----C-A--TGA---A--AC-GAC--AA--------C-G-T-GA-------G-AT---C--T-G--------A-G--GA-TGG------------TTTAGA--- 3751
WRC.CI.98.98CI04        --GG---TA--A.........--A--A-----TG---A--A--G-----A-----------------GCA--------TGAG--A---C--AC---A--------C---T-AA---------A----C-C--C--------A-G--A--TGG-----G-------TTAGA--- 3766
WRC.GM.05.Pbt_05GM_X02  ---G--ATA--G.........--A--A-----TG-------AC-GA-A--A-----A---------GCA--A--A--TGAG--T--A---AC--AAG----G------T-AA---G---AT---C-T--------C--A-G--GA-TGG--CATG---------AGC--- 2680
LST.CD.88.524           --G-T-----CAAGGAAAACT--A--A-----A--T-----AC-G--A--AGTT--A---------GCA--AC-G--TGAT--A--C-G-AC--AAG-GC-C------GT-GA---------GTA--GC-CT-GGA-C--C--G-GA-TTGG-----G------CTAGA--- 2725
LST.CD.88.447           --C-T----CTTCAGAAGACA--A--T--CA------------C-----A--AGTC-------------GCA--AC----TGAG--A----G-ACT-AAG--C-T------T-AA---G---GTA-----T---GAT-----TG--A-CTGG-----T-C----TTAGA--- 2728
LST.KE.x.lho7           --G-TG--TCTAACAAAAACT--A--A---C-A---------AC-G-----AGT-C-T----------GC---AT-G--TGAT--T----GGAC--AAG--C-C--C---T-AA-A--G---GTA---C-CT-GGA----A-C---AACTGG-----G-----CTT-GA--- 3812
SYK.KE.x.SYK173         A-T-T---TGA-GAA...ATC--A--G---C-GT-------A--GA-A--AGT-C-T------------C---------TT---A-----GAC--AAG-GC-C-----GTGTA-C---------T---CA-T-----------C-CA----T-------C---T---GC--- 3651
SYK.KE.x.KE51           ----TTACAGAATCA...CCA--A--T-----GT-------AA--A-A--AGTCC-A----------------A--A--T--T--T--C---TGT-AAGC---GA----GTG-A-A-----GG-A--------T--C--CA-TG--GA---G--------G-----AGC--- 3309
COL.CM.x.CGU1           ---C-G--T-AG......-----A--A-----T-----A--A--G--A--AGCC--A------TT---C--AT-A--------A--C-GGAC--A-G-GA--A-----TG-AC--A---T-AG--GCATT-GTTG--G-A-G-G-AG--T-GT-C---------AGAG-- 3226
```

```
H1B.FR.83.HXB2          ACTGGCAGAAAACAGAGAGATTCTAAAAGAACCAGTACATGGAGTGTATTATGACCCATCAAAAGACTTAATAGCAGAAATACAGAAGCAGGGGCAAGGCCAATGGACATATCAAATTTATCAAGAGCCA...TTTAAAAATCTGAAAACAGGAAAATATGCAAGAATGA 3621
Pol                     _L__A__E__N__R__E__I__L__K__E__P__V__H__G__V__Y__Y__D__P__S__K__D__L__I__A__E__I__Q__K__Q__G__Q__G__Q__W__T__Y__Q__I__Y__Q__E__P__.__F__K__N__L__K__T__G__K__Y__A__R__M__
H1A1.UG.85.U455         -T-------G----G-------------C--T--G--------A--------------------C--G----------------A--A------A---------------------------------...-----------------------------------AA- 3067
H1B.US.90.WEAU160       G-------------G--A-------------G--------------------------------------GC-----------------------------G-----------------...----------------------G---------G--- 3620
H1C.ET.86.ETH2220       -T-------G----G--A-------G---------------A-T-------------------T----------A-----A-T-A----------T----T----C-----A---...--C---------------G--G-T-----A--GAG 3013
H1D.CD.84.84ZR085       --------------G---------G-----A-G--------------------------------T------A--A-------T--------------------------A---...----------------------G----------- 3143
H1F1.BE.93.VI850        -T-----A---T--G---------G---------------G--A-------------------------A--A--AG-C--G---------------------GA-C---...-----------A----------G-------A-G--- 2958
H1G.SE.93.SE6165        GT-------G----G--A-------G------T----------C-------------------A-----------G----A--A--A----T--A--------------------------...-AC--------------------G-------A--G-G 3018
H1H.CF.90.056           -T-------------G--------G-G-------A----------A--------T-------------------------G-----A---------------------------G-------...-------------G-----------------A----- 2968
H1J.SE.93.SE7887        --------------A-------------------------G--A---------T--G-------A-----------G-G-----A--A----TG-A--------------------------G-------...-------C-----------G----------A--G-- 2935
H1K.CM.96.MP535         -T-A-----G-------------------G-----G--A------------------------------A--A---A-T-A------------------A-----------------...CA------------------G--G-------------- 2817
H101_AE.TH.90.CM240     GT-------G----G-------------ACC--T--G-------A--------------------G-------G------A--A-----G-A--------------------------...------------A----------------C----GAG 3195
H102_AG.NG.x.IBNG       -T-------G----G--A---------------T-------G--A-----------A----------G-------T------A--A------A--------------------------...----------T-A-------------------A--A-- 3146
H1N.CM.95.YBF30         ---A------------G-----AT-----------CC-G--------C-----------GG------A---G---------T--A-----A--A-----T--G----------G---------G---TT-...CA-------T-A-----------G--------A----- 3215
H1O.BE.87.ANT70         -T-A-AG------------A-GGT-----C--------G-----G--A--C---C-A--TGAT--G--TC--TGG-TTA-T--T--------A--AGGG-AG--------T--C--G--A-----G--AGA-...CA---G--C--C--------G------A-T--GCAA- 3676
H1O.CM.91.MVP5180       -T-A-A------------A-AG--------G------------------A-----CC-G--TGAC--------GTGG-TTAGT--T--------T--AG----G--------T--C--GG-A-----G--TGA-...CA---G--C--T-----------------T--GCAA- 3651
CPZ.CD.90.ANT           -T-A-A------T-A-C-A------C-GC-GAA-A--G-G---TAC--C---C-G--TGGTCT-CCAC---A----ACC---------A--A---TC---A---------C---------A------A-TGA-...GGA---CTGT-A---G-------------T--GCCT- 3060
CPZ.CM.01.SIVcpzCAM13   -T-A-------------A--AG-T-GTAC-------------G--T------A-T---GAT--G--GC-C---------G----A--A------AG----G--G-----------G--C-T---G--A-A-...CA--G----T-A--G--------------C---CA-- 3201
CPZ.CM.05.SIVcpzEK505   GT-A--T--G--T--G-----AT-----------T----------G-CC------A-----GAC------A-----------G--C-----------A-----T--------------G-----------CTT-...CA-------T-A--G-------------------A----- 3193
CPZ.CM.05.SIVcpzMB66    -T-----------T-----A--AT------------------------G-CC----------C---------A------------------------A--T-----A--G----T-----------TC--G--A-AG...-A---G--CT----------G---------T--G---- 3167
CPZ.CM.98.CAM3          -T-A-A---G-----R------T---RG--T-----G-----G-----C----------GA---G--A-----------N-------A-----AA-TA-T--G-----T-----------TC--G--A-AG...CA---G--C--A--------C--------T--GCAA- 3032
CPZ.GA.88.GAB2          GT-A-A------T--G--A---T--------------G-----CACC--C----------GA------AC-----------G-T--A------AATA-T--G--------T-----G----T--G--AA--...GGA-GG-----A-------------------C---CAA- 3014
CPZ.CM.98.CAM5          -T-A-A------T------A---T-----G--T--T--G--------T--C-----T---GA---G--G-----------GG----A--A-----AA-TA-T--G-----T--C--G----TC--G--A-AG...CA---G--CT-A--------G--------T--GCA-- 3323
CPZ.CM.05.SIVcpzMT145   ------------TCA-------A-----------------------CC--C---------TGA---G--A-------------------------A--A--G--T--G-------------A-T---G-------...CA-------C--A--G--------------GC-A-CAA- 3163
CPZ.TZ.00.TAN1          GT-A-A------T-AGC---------------AAG--T--A----CA--C-----T--TAA-TT-CCTC-GCA-----C---------------------A--G-------------G--A-----G--AGA-...GGG-------T-A--------------------A-TCAC 3265
CPZ.CM.05.SIVcpzLB7     -T-A---------T--G--A--AT-----A-C------------G--C-----------GG-T------C-G---------G----A--A--A--A--GA-T--------C---------C----G--A-A-...-----------------------------------C---CAAG 3176
CPZ.US.85.CPZUS         GT-A-A-------------A-------G-------------G--C-----------GA------A---G---------G----A--A--A--AAGGA-T--G---------C-------T-----A-GG...CA---G-----A--G---------------C---CAA- 3681
CPZ.GA.88.GAB1          -T-A---------T--G------AG---GCAC-------------G--A--C-----T---GAC------GC-T------------------A-----CA-CT----G-----T-----G--A-T--G--A----...CA---G---T-----G-----G--------C---CAA- 3682
MON.CM.99.L1            G--A--G-----T--G--------C--G--GGA-TCCACC--G-GA-----C-----CAACG-GCCA---AG-TG--GT--ACCTCA-TA-CAG--AA------GGC--CAG-T-C-TC-----AAA-......TC-GTAT----G-----------T---C-A---T- 3324
MON.NG.x.NG1            -T-A--------T--G--A-------GTC---AGTCGAC---G-GA-----C-----CCAGCTC-----G-AG-T---G--TACCTCC-T--YAG---A--G---GGG--CRGGTA--TC--G--ARAT......-GGGT---A--G-----C----TC----AG--T- 1917
GSN.CM.99.CN166         -T-A--------T--G-----CT-----C--T-TTC-G-G--T-GC---------G--GA---GCCGC-CG-G-TG-----TACCTCC-TA--AG--CAG------GGG--CATGT-CAGC-----AGGC.......--G-TGT-A---T---CC----TC----AG-CC- 3346
GSN.CM.99.CN71          GT-A-----------------------C--T-CTC-G-G--CAGC--C--C---G--GA---GCCAC-TG-G-T------TACCTCC-TA--AG--CAG------GG---CATGT-CAG---G--CAGC.......--G-TG--A-G-T----C----T-----A--CC- 3328
MUS_1.CM.01.1085        -T----------TCAG-----C--TC--C-GA----C--G--G-CA--C---------GA-C-GCCA---G-G--------TGTCTCA-TA--AG-CTCA--G---GG----A-CT-CACC--GAGCAAG.....GGGGTAT-A-GG-----------T---C-A-G-A- 3308
MUS_1.CM.01.CM1239      -T-----------CAG--A-----T--GC-GG----C--G--G-CA--C-T---T---GA-C--CC-C-GG-G--T-----TGTCTC-TTA--AG-CTCA------GGG---A-CT-CCTC---A-TAAG.....GG--TAT-A---T-T-------T---C-A-G-C- 3298
MUS_2.CM.01.CM1246      ---A--------T--G--A--C--G--GC--GAG-CCGG---G-CC-------------GA----CC-C--G-CTT---G---GT-TCCTTA---G--CAA------GGG--CACCT--ACC-----CAG-.......--C-TG--A-G----------G-T---C-AG--C- 3406
MUS_2.CM.01.CM2500      GT-------------G-----CA-G---C--GAG--CA-A----CA--C-----T---GA----CCA-----TCTG------GT-TCC-TC---G--CAA------GGG--CAC-T-CACC------AAG.......--C-TGT-A--G-----------T---C-A-G-T- 3365
RCM.NG.x.NG411          -TAT-A--------A--CTG-C--G--G--GAA-A-------G--T------A-----GA----CCA--------TA--G-T-----ATTA-AAGG---G-------GC--C--G--AG-G--G-GAGAC...-CA---CC---A--G-----G--G-------A-CAA- 3115
RCM.GA.x.GAB1           -TAT-A--------AGAT-------------AAGT-G-----G------------TGA-AAG---CC----G-G---A-C--T------TTA-AAGGT--A--G---T-C--------AG-A--G---T--...GGA---CCA-----------C--------T-AGCA-- 3100
DEN.CD.x.CD1            G--TAA----GCACA---A--AT--------AGTA-TTCA---AGT------AGG---GAG---CCT--G---------G--CTCT--ATTA--AG--AAT--G---GG----AC---CA-A--G--AAA-......C-TCTA--C---------C-----CCAT-AGGAC- 3511

MAC.US.x.239            -TAT-AG-----T-A-ATA-----C-GTC-GGA-CA-G-A---TGT-----CC-AGA-GGC--GCCA---GA---CACGG--ATA---AGTCA-G-CAAT--G---T-T---A-----C-C-----AGAC.........----TA------GT--------T-----AG--A- 3920
Pol                     _E__Y__E__E__N__K__I__I__L__S__Q__E__Q__E__G__C__Y__Y__Q__E__G__K__P__L__E__A__T__V__I__K__S__Q__D__N__Q__W__S__Y__K__I__H__Q__E__.__.__D__K__I__L__K__V__G__K__F__A__K__
H2A.GW.x.ALI            ---A-AG--G------ATT--CT---GTC-GGA-CA-G-G--GCAC-----CC-AGA-GA---G--G---GA---AC-G-C--A--AG-TCAAG-CAAT-----------A----AC-C--G-GAGA-.........----T---A---GTG-A---G-------A----- 3996
H2A.DE.x.BEN            G--A-A---------A-ATT--CT-G-GCC-GGA-CA-G-A---TAT-----CC-AGA-GA------A---GAG---AC---C--A--AAGCCAAGG-CAT-----------CA----AC-C--G--AGAG.........----TC--A---GT------G-------AG--A- 4003
H2A.SN.x.ST             ---A-A---------A-ATC--CT---GCC-GGA-CA-G-A---TGC-----CC-AGAGGA---G--GC--GA---AC-G-C--A--AG-TCAAG-CAAT--G---------A-G--AC-C--G-GAGG-.........----T---A---GT---------------AGG-A- 3448
H2B.GH.86.D205          G--ACAG--G--T-A-ATC--CT--G--C-GGA-CA-G-A---TCC--C--CA-GGA-AGGGT-CCGC--GA----AC-G-------AA-CCTAGC-AAT--G---------CA-----C----G-GAAAT.........---GTC--A---GT---------------AGG-T- 3976
H2B.CI.x.EHO            -T-CCAG--------A-ATC--C---G--C--GAGCAGG-A---TCC-----CA-AGA-GGGGT-CCT---GA----AC-G-G-----AA-TCTAGC-AAT--G---------CA-G---C----G-GAGAT.........----TC--A---GT---------------AGG-T- 3974
H2G.CI.92.ABT96         -T--CA---------AGATT--C---R--C-GGARCAGG-A--C-CC-----CA-AGA-GR-G-GCCT--RGA---CACTG-------AA-TTTAG-CAAT--G---------CA--G--C------GAAAC......-G--TC--A---GT---T----------AG--A- 3330
H2U.FR.96.12034         -T-T-----G--T-AGATT---T-G-GTC-GGA-CA-G-G--G-CT-----CAGGGA-GA-G--A-TC--GA----AC-G-G-TT--AA-TCAAG-TAAT--------C--CA-G--CC----G-GTGAC......-GG-CAT-A---GT--------T---C-AGG-A- 3485
MAC.US.x.EMBL_3         -TAT-AG-----T-ATATA-----C-GTC-GGA-CA-G-A---TGT-----CC-AGA-GGC--GCCA---GA---CACGG--ATA---AGTCA-G-CAAT--G-----T---A-----C-C-----AGAC.........----TA------GT-A------T-----AG--A- 3394
SMM.US.x.H9             -TAT-A---------AGATA----CC-GTC--GA-CC-G-G---TGT--C---C-GGAGGG---GCCAC--GAG---CC-G--ATA---AGTCA-G-TAAT------T-----A-----C-C------GAC.........----TA--A---GT---C----T-----AG--T- 3393
SMM.US.x.PGM53          -TAT-A---------AGATA-----C-GTC--GA-CAGG-G---TGT--C--CC-AGAGGG---GCCA---GAG---AC-G--ATA---AGTCA-G-TAAT------T-----A----CC-C-----AGAC.......--G-TA------GT---C----T-----AG--T- 3837
SMM.US.x.SIVsmH635F_L3  -TAT-A---------AGATA-----C-GTC--GA-CA-G-A---TGT--C--CC-AGAGGG----CCAA--GAG---AC-G--ATA---AGTCA-G-TAAT--G---T-----A-----C-C-----AGAC.........---GTA--A---GTG--C----T-----AGG-T- 3923
SMM.SL.92.SL92B         -T-------G--T---ATT--------TC--GAGCAGG-A---AGA-----CAGAGA-GATG--CCGC--GA----ACGG-T-TA---A-TCAAG-CAAT--G----GT---A-G---C------GTGAC......-GG-T---C--GGT---G--G-T-----A---A- 3351
STM.US.x.STM            -TAT---------T-A-ATT-----C-GTC-GGAGCAGG-A--GCAA--C--CCGAGA-GAC--GCCTC--GAG---AC-G--GT---AG-TCA-G-CAAT------T-C---A-G--AC-C-----AGAC.........----TAT-----GT---C--G-T-----A---A- 3567
MNE.US.x.MNE027         -TAT-AG-----T-A-ATA-----C-GTC-GGA-CA-G-A---TGT-----CC-AGA-GGC--GCCA---GAG--CACGG--ATA---A-TCA-G-CAAT--G---T-T---A-G---C-C-----AGAC.........----TA------GT--------T-----AG--A- 3388
DRL.x.x.FAO             -TAT-A------T-A-CTA--AA-------G-A-A-G-----G--A------C-G---GA----CC-C---A---CA-G-----A--A-T---TA-T--A------T-------G--AG-A-----TGAT...AAC--GCCA--A----------------T-AGG-A- 3188
MND_2.x.x.5440          -TAT-A---------AGCTA--AG-GC-G--GGA--------------C--CC-G---GA----CCT-----G---A-GG-G--A---TTAACA-----A--G----GT--C-----AG-A--G--AGAC...AAC---CC---C--GGT-------G------C--G-CA- 3182
MND_2.CM.98.CM16        -TAT-A---------A-CTC--AG-GC-G--GGA---G---------A--C---C-A---GAG--GCCT-----G---A-GG-T--A--ATT-ACA-----G--------GT--C-----AG-G--G---GAT...AA---GCCG-----GGT---G--------C----CA- 3622
MND_2.GA.x.M14          -TAT-A------T-AGTTA--AG-GC-G--GGA------------T-----C-G---GG----CCTC-G--G--CA-GG-------A-T-ACA--G--A-------GC-----G--AG-A--G--AGAG...AA---GCC---T--GGT---G--G-----T--G-CA- 3549
OLC.CI.97.97CI12        -T-A-A---G-------------T----G---GA-CA-G-A------ATG---A-AGAGGAGG-GCCTC-CCAGTGTACT...TTATGCT-TACAA-G--AACGGTAG-----AG---A-GG--GA---AT......GG--CCT-A--G-------GG-T-CA---G-A-- 3489
GRV.ET.x.GRI_677        -TATAAGA-C--TCA--G---AG-GC-G---A--CA-G-A---ACA--C---------TCTC-----A---------AC-G-T--A-----A--AG----G-----------C---T-CACC-----AGGG......GC-GTAT-A--GGTG----G------C-AGCAA- 3794
VER.KE.x.TYO1_patent    -TAT--------TGC---------T---AC-GA-CAGG-A---ACC-----CA-A---GG--T-CCTA-T-GG----C-G------ATT--AAGG---A--G----GT--C---T-CA-A-----AGG-......C--GTCT-----GT---------CA-C-AGCAA- 3845
VER.KE.x.9063           GTAT-----G----A---A---T-G---AC-GA-CA-G-A---ACT-------CA---GG---GCCTC-T-G-----C-G-C-----ATTA-AAG-T--G-------GT-----GT--A-A--G--AGG-.........----TCT-----GT------G-------A-CAA- 3848
VER.KE.x.AGM155         GTAC-A---------AG-----C------AC-GAGCA-G-A--TACT-------CA---GA----CC-C-T-GG----C-G-------ATTA--AG-T--G------T-----C---T-CA-G--G--AGG-.........----TCT-A--GGT---G--G-TC--C-A-CA-- 3840
VER.DE.x.AGM3           -TAT-----------A------CT-----ACGGA-CAGG-A--GACC--C--CA-A---GG--G-CC-A-C-G-----C-G-G--A--A-TA-AAGG---T-------GT--C---T-CA-G-------GG-......C--GTAT-A---GT---T-----CA---AGCA-- 3343
TAN.UG.x.TAN1           GT-CAG-C----TCA------CT----------GAGCA---G----CT--C----------AGT---TCT--G-G---CACT--TACC----T--ATG-T--G------GG-------T-CA-A--G---GG-........---GTA--A-----------------A---AG-GC- 3820
SAB.SN.x.SAB1C          -T-A-A------T---C----CT---G-C--AAGCAG--G---CAA-----------TG-TCTTCCT----GG--TA-GG-T-TT--A-T---AG-T---------GG------G--A--C--GCCAGA-...AA-----TA-----GGTT--C----------A---C- 4019
MND_1.GA.x.MNDGB1       -TAT-A-C----T-A------CT-GGCT---GA-CA-G-A--GTCT--------T--TAAT--G--A---TAT-T-AG-T-T-----AACAACAGG---AG-TATAT---T----TGGA-G----GA......AA---GGT-T-A-G-G----G-------GG-A-CA-- 3308
DEB.CM.99.CM40          ---TAG---G----AG--A--C--G--GC--GA---GTCA-----T--C---A-AGAGCAGG--CCA------------T--TCC--ATTA---A----A------GG----GTG--CAGA---CCTAAG......GG--TAT-A--G-----------TCT-T---GAT- 3427
DEB.CM.99.CM5           ----AGG--G--T-A---A-C--CC---C-GGA---CTCA--G--A--C--CA-AGA-GATG-GCC-C-----T----G--TTCC--AATA--AAG---A--G---GG---GT----AGA---ACAAAG......GG-TTG--A--------G----TCT---AGGAC- 3421
TAL.CM.01.8023          -T---------------CA------GC---GGAGACC-----TTCA-----CA-GGA-GG-G--CCA---GA------GC--ACC--ATT--CAG-----------GGC--CATGG-AA-A--G-GAAA-......CGGCTG--A--G-----G--G-T-----AGCA-- 3085
TAL.CM.00.266           -T-A-----------G--A--CT--GC----GA-ACT--C--TACA-----CC-GGA-GG-G-GCC-C-GGAG------C-GACC--ATT--CAG----T------GGC---ATG--AA-A----GCAA-......CGGCTCT-A--------C--G-T---C-A-CA-- 3587
SUN.GA.98.L14           -TACCAGAGT--TCAG---G-AT--C-G--GT-----TCA--CAGT--C-----A---GAT-----A-----TTGCAG-G-G--A---GTAAA------AAT--TA--C-T----TGG-TG--G-GAAA-......C--GTAT-A-GGGT-----G----CA--AG-GAG 3978
WRC.CI.97.97CI14        GT-A-A-ATC--T-AGC----AT---------GA-CA-TCA----CC--C---A-AGG-AA-G-CTTAA--GC--A-TT----A-A-CAGG-CCAGG-CAATGGGCCTATACAAT--GAG-AAT--G-GA-...AAG---CCA------GT---C--------GG---CAA- 3918
WRC.CI.98.98CI04        GT-A-A-AT---T-A-AGC-----------GGAGCA-ATG--G-CT------ACTGGGAA-G--TTAG-GGC--A-TT----A---CAGGATCAGG-CAATGGGCCTATAG-ATT-GAG-AGC--GAGA-...AAG---CCAT-A--GGT---G--------GG--GCAA- 3933
WRC.GM.05.Pbt_05GM_X02  ---A-A-ATC--T----CT--------G---T--CA---A---AGT-------GAGG-GG-G-CTTAA--GCT-AG-T----A-A-CAGGAACAGGGCAATGGGCCTACACAAT--AAG-AGC--GAGAG...ACA---CCAT-A---GT--------T-A--------C- 2847
LST.CD.88.524           -TAT-AGACC--T-AGACA-----------AA------G--G-CT--C-----T---AGT-----AC-TG---T-AGGG----AC--A-CAA-A----GGTTATA--T-TC---TGGAGG----GAAAT......GCC-T---A-G-G------G----CA----CAA- 2889
LST.CD.88.447           -TAT-A-C-----CA-C----A---------AA-C-G--G----CT---------T--TA-T-----A--GG---T-AG-G-----C--AGTAA-A-G--AATTATT--C-TC--GTGGAGA----GAAAT......TCT-T-----G-G------GG---CA----CAA- 2892
LST.KE.x.lho7           -TAT-GGC-G----A----G-A--T------AAGA-G--G----CC-------------GA---G--AC-T--T-TCAGGG-----C-AA-CAAAA-G--GAT-ATT--T-TC--GTGGAGA----GAAAT.......--C-TCT-A-G-G-T--G-GG---CA----CA-- 3976
SYK.KE.x.SYK173         -T-A-A---G--T---C-A--CT----G---GA-CAGGCA--GTCT-----------TAA----CCT---GA----C-C--CACT--A-TA---TCCCAG------GG---CATG---A-G--------A-AAAGGCCC-CCC--A-T----------GACA----A--CCT 3821
SYK.KE.x.KE51           -A----T-----T---------A--C-----T-GA-CA-ACA--GTCT------C-G---GA----CCA---GA----C-C--TAGC---TTA--A---CAA--G---GGG--CAT----A-A----G----CAAGAA---CC-----TC---------ACA-G--A-GCAT 3479
COL.CM.x.CGU1           G--ACA---T----A-C----AA-C------GA---T--G--------C---ATA---CATG-G---A-TTGG-T---C--TTTC-G-TTA---A----A------GGC---GC-G-C-GC---A-A-AT......GG--CG--A-G--G-------C--AAC-C-GGC- 3390
```

```
H1B.FR.83.HXB2          GGGGTGCCCACACTAATGATGTAAAACAATTAACAGAGGCAGTGCAAAAAATAACCACAGAAAGCATAGTAATATGGGGAAAGACTCCT...AAATTTAAACTGCCCATACAAAAGGAAACATGGGAAACATGGTGGACAGAGTATTGGCAAGCCACCTGGATTCCTGAG 3788
Pol                     R  G  A  H  T  N  D  V  K  Q  L  T  E  A  V  Q  K  I  T  T  E  S  I  V  I  W  G  K  T  P  .  K  F  K  L  P  I  Q  K  E  T  W  E  T  W  W  T  E  Y  W  Q  A  T  W  I  P  E
H1A1.UG.85.U455         --TC---T---------------------------A-TG---------G-GT--------------------------------TC---...--------G---A------------------------------------G---------TG----------G--T-------------A 3234
H1B.US.90.WEAU160       --------------------------------------------------G-----------------------------------------...-------------A------------A------------------------------------------------------------- 3787
H1C.ET.86.ETH2220       --AC--------------------G--G-------C--T---------G---G--CTG-------------------------------...--------G-T-A-----C--G--A-------------G-----------C-------------------------A 3180
H1D.CD.84.84ZR085       ------------------------G---------------------G---T-----------G---------------------...--------G---A-------------------------------T--C-------------------------- 3310
H1F1.BE.93.VI850        --TCG---------------------------A-----A-----G---G-TTT-------------------G----GAT--CCT--G-------A-------TG--A--G--------T-----------T--------------------------A 3128
H1G.SE.93.SE6165        --TC--------------------G-------A-T------------G-------G-----------------------C...------------A-------G--A------------T-----------C-------G------------------ 3185
H1H.CF.90.056           -AAC----------------A----------------A----------G---T-T--------------------------A-T----...--------G---A--T-------A----------G--C--------------------------A-----------A 3135
H1J.SE.93.SE7887        --A-----------------------------G----A-TG-----------G--TTG---GC------------------A------...--------G---A----------GA-----------G-------------C--C-----G------------------ 3102
H1K.CM.96.MP535         -ATC--------------------G------------A---------------G--------G-----------------A------...--------G---A----------A-----------G--------------A-------------------C-----A 2984
H101_AE.TH.90.CM240     --TC---T---------------G-----------A-TG---------G--G---------------------------C---...--------G---A----------GA--------------------TG----------G--T-------------A 3362
H102_AG.NG.x.IBNG       --TC--------------------G-------------T----------G--G-T-TG----------------------C---...--------G---A----------GA--G-----------------TG----------G--T-------------A 3313
H1N.CM.95.YBF30         -ATC------T---------A-------G---GTT--A-TG--AAGG---G-GG-A--------T--------T---------------...--------G-T-A--AG-----------GTG-----GG----------C--TC-----------A--T----------- 3382
H1O.BE.87.ANT70         A--CCT-------A------A---G-------G----A-T-A-C--G--GG-GT-TCA----TCT---A-T--C--------ATTG---...---------G-----AG-CACT-GA-----T---------------G-G--C-------------------------A--A 3843
H1O.CM.91.MVP5180       A--CCT-------A------A---G------GG----A-T---C--G--GG-GT-TCA----GCT-----T---------G--ATTA---...------C-GG-----AG-TACT-GA-----T--------T------G-----A-----------G-----------------A 3818
CPZ.CD.90.ANT           CA--AA-T-----------G--T-GG------G-T-GA-T---------------GGATT------T-----T--C------G--GTA--A...------C--T-A------ACT-G------T--------TG-C------T-----C-------G-----------A--A--A 3227
CPZ.CM.01.SIVcpzCAM13   -ATCAA-A--T--A------A-T-G------GG----A-TCA-A---------TG-TGTG-----T-----C--T---------A--A---...-G------G-T-A--AG-------A----GT-----G-----------G-----------C---------G-----------C--C--- 3368
CPZ.CM.05.SIVcpzEK505   --TC-A-T--T--------CA---G---------T--A-TG--A-----GG-GG-ACT------T-----T--C---------A-------...------C-G-T-A--AG-G-----A----GT-----G--C--------------------------G--A--T-----------T 3360
CPZ.CM.05.SIVcpzMB66    -ATC---A--T------------------G------T----T--A------G--G--CT--------------C--C---------AGTA---...-G---C-GG--A--T-----------------------GG-G-----------C----------------A--A-----C-----A 3334
CPZ.CM.98.CAM3          -ATCA--A--T--------CA---G-------G------T---A---------G-T--T-----T--------C--------A--A--A...------C-G-T-A--TG-T---------GGTC---------C------T--------------------------C--A--C 3199
CPZ.GA.88.GAB2          -A-CA--T-CT--------CA---G----C--G-T--A-TG----C-----G-TT--C-T-----T---A-CC-T--------CTGT----...-------GT-A--AG-G----G----GT------G-C----------T----------------------C-----A 3181
CPZ.CM.98.CAM5          -ATCA--A--T--------CA---G-------G-------TG-------------G----T-----T--------C--------A--A--A...------C-G-T-A--TG-T----------GTC---------C-----T--A-----------------A--A--C 3490
CPZ.CM.05.SIVcpzMT145   -A-C---A--T----------A-T-GG--GC-TG-T--AA----A----------G-T-----------T---------------A--A--A...--------G---C--TG-T-----A----TG-----G-C------G----T--C-----G--T----------- 3330
CPZ.TZ.00.TAN1          CA---A-------C-----GA---G-------G---GACTGA-A--G------GG--AT--G------A----T-----T-TTGTG---...------TT-T-A--TG--TCC--A--G------AGCCAG-------T--T--C-------TT------G-A------ 3432
CPZ.CM.05.SIVcpzLB7     --TC---T--T--C---------------G---G------T-A-------G---G---T------------C------------A-TA--A...--G----G-T-------T-----A---G-------G---------------T------------------T-----C-----A 3343
CPZ.US.85.CPZUS         -ATCA--A--T---------A-C-G---GC--GTT--A-TG-----------G-T--T--G-----T--C--T---------GTA--A...---------T---TAG-G--G------GTC----------------T-----A----------G-----------A--T 3848
CPZ.GA.88.GAB1          --TCA--A-----A-----CA-C-GG---C-GG-G--A-----A---------TG-T--T-----------------T----------A--A---...-----GGT-A--AG----G--A----GT------G-----------G-----A--C-----G--A-------C-----A 3849
MON.CM.99.L1            -AA--A-----T-C---ACCTATC-G--G---G----T--GC-AGCC---T-GGGA-A---GGC------C-C--------CGACT---C...-TT---CGC--A--AG-GGTG--A--GCA------CG---------G----TA--------AT--AT---G-A--A--C 3491
MON.NG.x.NG1            AAAA-------T-C--CACATACC-G-----GG----T--T--AACC---C--GGG-AG--GGC-T-G--C--T-----------CTA--A...-T-----G-T-A--AG--A-C--A---CAG------C----------G-C--CA-----------T---T-----A--A--- 2084
GSN.CM.99.CN166         -AAA---A----TA--CTCCTATC-G-----TG----T--TT--ACC---G--GGT-GG--G--TCAG--T-CC----------GTA--AGAC-------GGA-C--AG--GTC------CA------TG---------TGA-T--------G--A--A-----A--CACC 3516
GSN.CM.99.CN71          -AAC---A--T-TA--CTC-TATC-G-----GG----T--TA--ACC---G--GGT-GG-----TCAG--T-CT----------AGTA---GAC-----C-GGA-C--AG--GTC-----GCA-------T--G---------TGA-C--------G--A--A----------CACC 3498
MUS_1.CM.01.1085        --TCA-TA----------AGCTATC-G-----GG----C---A--GCC-------GGA-G-------C-----C--T------G---C-C---GAC--GG-C-G-A-C--AG-GATC--A---CAG------CCAG---------T--CAC-C------G-T-T-G-----A---ACC 3478
MUS_1.CM.01.CM1239      -ATCA-T---T--C--CAG-TATC-G--G---T-T--T--CA-AGCC-G-----GGA-GG---TCAT-----C--T------G---C-C--CGAA--GG-C-GGA-T--AG-GATC-----GCAG-----TCA--------------AC--------G-TAT-A------A-----C 3468
MUS_2.CM.01.CM1246      -AACA-----T-G---CCCCTACC-----GC-GG----A--CC-ATCT-G-GCT-GT-AG--GGC-C-----CTGT------G------C--AGAC----GC-GGA-A--GG--GTC------CAG------C-AT------G----C-CA------A-A--T----------C--C 3576
MUS_2.CM.01.CM2500      -AACA-----T-TC---AC-TACC-G---C--G-----------T-AGC---GGCTG-A-G----TC-T-----CTGT-------------C---GAC----G--G-A-A--GG-GGTC-----GCAG-----CCAC------G----A-CA-----GA-T---A---G-----A--- 3535
RCM.NG.x.NG411          A-ACA--A--T--A-----AA-T-G-ATGC-GG---GCTT---A--G-------TG--A----GCT---------T----------CT---A...--------G-G-----A---G---GA---GT-------CAG------T-G--T--C------G-TTT-------------A 3282
RCM.GA.x.GAB1           AAACA--A-----C-----AA-T-G-ATGC-GG---G-TT---A---------TG-A-A----GC------T--T-------G-CTG--A...-C---C-G------A---G-G-GA--GGT-------CTA----C--T-CC-A--C------G-TT--------A--A--T 3267
DEN.CD.x.CD1            AA--AAG---TTA---CCCATATC--------G-CA-A-TGA--GC-------GGA-G----GC-T-------C-----C-GACT----...-CT--C-G---C--AG--AGC----T-GAT-----TG-------T----A---------TT-AT------A-----C 3678

MAC.US.x.239            A-AA-A-A--T--C----GA--G-G--T----G--C-T-T-A-A--G------GGA-AG---GCA-----G--C------C--GTC--A...-----CC-CT-A--AG-TG-G-----TGT-------CAG-----------C---------G-TA---------A--G--A 4087
Pol                     I  K  N  T  H  T  N  G  V  R  L  L  A  H  V  I  Q  K  I  G  K  E  A  I  V  I  W  G  Q  V  .  P  K  F  H  L  P  V  E  K  D  V  W  E  Q  W  W  T  D  Y  W  Q  V  T  W  I  P
H2A.GW.x.ALI            AAAA-A----T--C--C-GG--C-G-TTG---G--C---T---T---------GGA-A----GCAC-G--C--T------CGA-TA--A...-G----C-C--A--AG--G---GA-----C------CAG------GAT--C--C-------TG--A------C--A--C 4163
H2A.DE.x.BEN            AAAA-A----T--C----GG--C-G-TT-C--G--C---T---T-G-------GGA-A---GGCAC----C--T------CG--TA--A...-------C-C-----AG-GG-G-GA--G--C------GCAG------GATA-C--C-------TG--A------C--A--- 4170
H2A.SN.x.ST             AAAA-A-------C--C-GA--C-G--TCC--G--C-A-T---T---------GGA-A----GCAC----C--T------CGA-TA--A...-------C-C--A--AG--G---GA--T--C------CAG------GATA-C--C-------TG--A------C--A--C 3615
H2B.GH.86.D205          AAAACA-G-----C--C-GG----G--T-C-GG--C-T-T---T-G------GG--A----GC-C-----C--C------G---TA--A...GTG--CC-T-----AG--G---GA--G--------CCAG------------T--C-------TA---------C--A--- 4143
H2B.CI.x.EHO            AAAACA-T-----C----GA----G--T---GG-TC-T-T---C----------GGA-AG---GCAT-G--C--C------G---TA--A...-TG--CC-T--A--AG--G---GA--G--------TCAG------------T--C-------TA---------C--A--A 4141
H2G.CI.92.ABT96         AAAA-A----T--A----GA----G-TT---GG-TC-T-TG--C-----R---GGA-AG---GCAC-----------------ATTA--C...TTC---C-T-----AG--G---G---C--C------CA-------------C----------TA--T------C--A--T 3497
H2U.FR.96.12034         A-AA-A-T--T--A----GA--TCGC-TGC--G--C-T-TG--A--G--G---GGA-A----GC-------------------A-T---A...-TG--CC-T-----AG--G---GA--GGTC------CA-----------T--C---------G-T-------G-C-TA--C 3652
MAC.US.x.EMBL_3         A-AA-A-A--T--C----GA--T-G--T----G--C-T-T-A-A--G------GGA-AG---GCA-----G--CGTA---C--GTC--A...-----CC-CT-A--AG-TG-G-G---TGT-------CAG-----------C---------G-TA---------A--G--- 3561
SMM.US.x.H9             AAAA-A-A--T--A----GA--C-GCTT----G--C-T-T-------G------GGA-A----GCA---A----T-------C--GTG--A...-G----C-TT----AG--G-G-RA-----TT------CA-------------T-----------TA---------A--A--A 3560
SMM.US.x.PGM53          AAAA-A-A--T--A----GA--C-G-TT----G--C-T-T---CG-G------GGG-A----GCA---------T-------C--GTG--A...------C-TT----AG--G-G-G------TT------CA-------------T-----------TA---------A--A--A 4004
SMM.US.x.SIVsmH635F_L3  AAAA-A-A--T--A----GA--C-G-TT-C--G--C-T-T-------G------GGA-AG--GGCAC--------T------G--GTG--A...-----CC-TT----AG--G---GA-----TT------CA-------------T-----------TA---------A--A--T 4090
SMM.SL.92.SL92B         AAAA-A-A--T--A----GAA---GGTT-C-GG-TA-T-T---T-----G---GGA-A------TC----G------------------A--A...TTT---C-C--T--AG--G---G---GGTT-----TCA-------------T------------T--T---------A--- 3518
STM.US.x.STM            AAAACA-A--T--A----GA--T-G-TT---GG--C-T-TGA-T--G--G---GGG-A----GCA-----G--C------C---TC--A...-----CC-CT----AG--G-G-GA--GGTC------CA-------G----T---------G-TA--A------A--A--A 3734
MNE.US.x.MNE027         AAAA-A-A--T--C----GA--T-G--T----G--C-T-T-A-A--G------GGA-AG---GCA-----G--C------C--GTC--A...-----CC-CT-A--AG-CG-G------TGT-------CAG-----------C---------G-TA---------A--A--A 3555
DRL.x.x.FAO             A-AA-AGT-----C------A-G-G-ATG---G---GACT---A-----G---G-A-A---GGCAC-G-----C-------G-TTG--C...-TT--CT-CT-A--AG--G---G----GTT------CA-------CAG-----C-----G-TA--A------C------T 3355
MND_2.x.x.5440          A-AA---------A-----AT-G-GGGT-C-TG---G-TTG--A-----GG--G-T-AG--GT-TC----------C----------CTA--G...-----CT-T--C--GT--G-G-G---GGT------CCA-------CAT--C-----------TA--A------C--A--- 3349
MND_2.CM.98.CM16        A-AA---T--T--A-----GC-T-G-GTGC-TG---GACT---A--G--G--TG-A-A----GCTT----G--C---------ACTC----...--G---T-T------A---G---GA---GTC------CCAG------C----A-----------------A---------C--A 3789
MND_2.GA.x.M14          AACA---------A-----GT---GGGTGC-TG---GATT---A------------G---A---GT-------G--T------G-ATTG--A...-----CT-T--C---C-GG-G-G----GT-------TCAG------CAT--T--------G--A--A------C--A--A 3716
OLC.CI.97.97CI12        T-AC---T--T--A---A-GT-T-G---GA-G----GA--CA-AG-----G---GGA-A------T-----G--T---------TTA--G...-TT---C-GG-A--T-CTAC--GAA--GAT-------G-----GCATTT--TG----------TACAT------A--A--C 3656
GRV.ET.x.GRI_677        -A-AAA-T--T----------C---GGACTC--G--C-CCTT--C--------CTGT-AG---GCAC-TACC--T--------GACT---A...CG-G-AC----C--AG--G-C---A-----------T-TG------CAG--C-----------TAT-------A--A--A 3961
VER.KE.x.TYO1_patent    A-AACA----T--A-----AC-TCGCAC----G-T-GTTT------G--G--TTG--A----GCTC----T--T-----G-TATTA--A...GTTC-AG----C--G---G---GA--GGT-------CA-------G-G--T--C------G-TA-G----------C--A 4012
VER.KE.x.9063           A-AACA----T--A-----AC-TCGCAC---GG-C-GCCT---A---------CTGT-A---GGCAC-G--C--T------G-ACTA--A...-TT---G----C--AG--G---G----GT-------CA-------G----C--C-------TG-G-------------A 4015
VER.KE.x.AGM155         AA-C-A-T-----C-----GT-GCGTGT-C--G---GA-T---A--G-------GGG-A---GGC-C--------T------C-ATTA--C...-CT---G----C--AG-GG-G-G---C--------CA-------G----C-----------T--GT------A--C--A 4007
VER.DE.x.AGM3           AAAACA-T--T--C-----GT-CCGTGT---GG---GATT---A-------C-TTGT-A----TCTT----T----------G--TTG--A...GTCC--G----C--A---G-G-G----GT-------CA-------G-T--T--C-----G-TA-GT---------A--C 3510
TAN.UG.x.TAN1           AAAA-A-T--T--A-----GT-T-G-GT---GG---GAATG--C----GGG-GTG--AG---GCG--CACCT-T---------ATTA--C...-TCA-GG--G-A--AG--G---GA---GT-------CA-------GT--C-----------T---A---C-A--A--- 3987
SAB.SN.x.SAB1C          AAACA--A-----C-----AT---G-ATG---G---GCCT---A--G--G---GGA-AG-----------C--C------C---TA---...-TTA-GG-------AG--G---GA--GCTC------CA-------T-T--C---------G-T----------C--G--A 4186
MND_1.GA.x.MNDGB1       AAACA--A--T-G------CC-C-TGA----GG---GT--TAC---G--GG--GGA-G------T--------C-----TTTTGTA--A...---A-GC-GA-A----CTAC--G---G-T-------GAT------CAT------------GTGT--A------A--A--A 3475
DEB.CM.99.CM40          AA--A--T---TA-------T-CC-T------G-TA----TA--T-T------GGA---------------TT-T-------GA-TA--A...C-G--C-GG--C--AG-GGT---A---GA-------C-AT------CATA-TC-C-----------G--------C--A--C 3594
DEB.CM.99.CM5           A---AA-T--TTA------CT-TC-T---C--G-TA-----A--T-T------GG------G--T-----CTAT---------AGT---C...--------GGT---AG--GT-------GA-------T-AC------CATA-CC-C-----G-----T-------C-----C 3588
TAL.CM.01.8023          -AACA--A----G---CCCCTATC-----C--GTG-GA--CA-----------GGA-A---GGC------CT-T------C--GTG--C...-T-----GGG-A--AG-GGTC------GGAG-----GCA--------T---------------TG-T-A---C-G--CACC 3252
TAL.CM.00.266           -AACA--A----G---CCCCTATC----G----T--GA--TA----G------GG--AG--GTC------CT-T-----GC--GTA--A...GT---C-GGA-C--AG-GGTC------GGA-------CAG------------AC-C------TG-T-A-----A--CACC 3754
SUN.GA.98.L14           ---CA--A--TGAG---CC-TGTC----G---G---CT--CC-A-----G---GGA-G-------------T---------TTTGTG--A...--GA-AC-GG-A-----------GA--G-T----AGTCAG-------G----C--C-----GTGT--A------C------ 4145
WRC.CI.97.97CI14        -ACA-A-T--T-G------GT---G-TC-A-GG-TC-T--CA----G--G---GGG--------T--------C--------G-GTA--C...------C-G-A-A------G-------T-T------TT-T------T----C-----------TGT-A------C--A--C 4085
WRC.CI.98.98CI04        -ACA-A-A-----C-----GC-T-G-AGC--GG-TC-T--CA----G------GGG--------T--------C----------GTA--C...---------GA-A------G-------T-TC------CT--------T----TC-C-----G-T-T-A---------A--T 4100
WRC.GM.05.Pbt_05GM_X02  --CA-A-T--T--A-----GC---G-AC----G-TC-TTGTA-A--G------GGA-----GGCT--------T-----G--ACT---A...--------G-A-C--AG--G-------T-T------C-TG------T----A--C------G-----A------C--A--- 3014
LST.CD.88.524           A-A----T----TG---CCCC-TC-G------G----A--T--A--G------GGA-A-------------C--T-----TTTTGTA--A...--GA-AC--G-A--TG-CACT-GA-------------AT-------T-A-C-T--------TA--T------A--A--C 3056
LST.CD.88.447           AATCA--T-----A---CCCC--C--A----GG-----------A---------GGA-A---G--T------C--T-----TC-TGTG--C...--GA--C-GG-A--GG--AC--GA---GTG-----TCA---------GT--TC-------G-T----------A-----A 3059
LST.KE.x.lho7           A--CA--A-----A---CCCC--C-GA-----GT---A--TA-T--G--G---GGA-A-------------C--C-----CTTTGTG--A...---A--C--G-C--AG-GACT-GA---GTC------GCAC-------GC--CC-C-----G-TT--A---------A--C 4143
SYK.KE.x.SYK173         TT-CA--A--T-G------CTATC--TCTC-TG-TC-ATTGT-AA-T------GGG-TCC--TCAC--TGGTAT---------GTA--A...-C---CC-T--C--AG--A---GA--GGA-------A------------------------A---G-A--A--A 3988
SYK.KE.x.KE51           AT-CGA----TCAC------TACC-GGC---GG-TC--CTGA--A-T------GGA-T-C--GC-C--TGG--T---------A-TC---...G----CC-C--C--AG--A---G----GA-------A-------------TC-C-----------A--A---T-G--A--A 3646
COL.CM.x.CGU1           AAAA-CAG---TT----AG-A-GC-GG-G---G-TTCA-TTA-A-----G---GGT-GG---GC-------C-CC----------GTG--C...---A-G---G-C--AGC-A-G-GA--GGAT-------CAG--------GT--C---------AGTG-T------C--A--- 3557
```

p51 end and p66 RT continue Pol p15 RNAse H start

```
H1B.FR.83.HXB2        TGGGAGTTTGTTAATACCCCTCCCTTAGTGAAATTATGGTACCAGTTAGAGAAAGAACCCATAGTAGGAGCAGAAACCTTCTATGTAGATGGGGCAGCTAACAGGGAGACTAAATTAGGAAAAGCAGGATATGTTACTAATAGAGGAAGACAAAAAGTTGTCACCCTAAC 3958
Pol                   _W__E__F__V__N__T__P__P__L__V__K__L__W__Y__Q__L__E__K__E__P__I__V__G__A__E__T__F__Y__V__D__G__A__A__N__R__E__T__K__L__G__K__A__G__Y__V__T__N__R__G__R__Q__K__V__V__T__L__T
H1A1.UG.85.U455       -----------C-----------C---A-------------------------C-------C--------G--A------------------------T---A----GC------------G----C--G-C-------------G-----TT------- 3404
H1B.US.90.WEAU160     -----------C-------------------------------T-------------------------------T------------------------A--------------------------------C------------G------T------- 3957
H1C.ET.86.ETH2220     ------------------------C----A-----------------C-----------------C-----T-----T-------------A-------T----A-----A-----------G---------G---------G--G---A----TT-T----- 3350
H1D.CD.84.84ZR085     -----A-----C----------T-----A----------------------G---------A--------------T----------------------T-AA------------------------G-C----------G---------C--T-T-- 3480
H1F1.BE.93.VI850      -----------C--------C--TC--------C-----------------A-C---G-------C---------C---------------------T----T--A-----C---AA-----------------G---A-----A------------T------- 3298
H1G.SE.93.SE6165      -----------C----------TC----A--G--------T-G------A-C---------CC-----------T-A----------------------T--------A----------G----------------G-C-A------AG------A--A-T-------- 3355
H1H.CF.90.056         ---------------C-------ATC----A----------T--------A-C---G-------C-----------T-A----A---------------T------A-------------------------------C----G-----------AG-------------T------- 3305
H1J.SE.93.SE7887      -----------C--------C--TC----A--G-----------A-----A--G---------A-G-----------T--------------------T-----------------AC------G-----G---------G-C-A---------G-----A--T-------- 3272
H1K.CM.96.MP535       -----------C-----T------C----A---C-------------------A-C-----------------G-------T----C---------------C-T-----A-----AAG----G-----------------G-C---------------G------T--A---- 3154
H101_AE.TH.90.CM240   ----------------------TC----A-----------------A-----A----C-------------------G--T-------------------GT-----------GC--------------G-----C---G-C-------------G---A---TT------- 3532
H102_AG.NG.x.IBNG     -----------C----------TC----A---------------------A----C---------------------T--------------A-------T--A--------GA-------------G-----C---G-C-------------G-----TT------- 3483
H1N.CM.95.YBF30       -----A-----C--C--T------C-T--A-----------T--------A-C---G--A--CAGT--G---------T--------------A--------T----A--A-----G----------T-T---G--AG----G---------G------G---T-TA-TG- 3552
H1O.BE.87.ANT70       -----A-----C-GC--A--C--A--GA-C-----------AG-C----A-GT-----T--TA-G--G-----------AT--------------A--------T--A-----A---C------G---------------AG-ACA---G-A---G--GA-AA-A-AAT--GA 4013
H1O.CM.91.MVP5180     -----A-----C-GC--A--C--A--GA-C-------------------A-C-----T--T-----G-----------T--------------A--------T---A-T--A---C------G--G-----------AG-ACA-----A---G--CA-AA-A-AGT--GA 3988
CPZ.CD.90.ANT         -----------------A--C--A---A-T-G-C-C-----TA-TC--TT-GC---C--T--TCC--AG--T--------T-----T---------------T--AA-CT--C----G--G--G-----C-----G--AG-C---------AGT-GG--AAAACA----CA 3397
CPZ.CM.01.SIVcpzCAM13 ---------CA----C-----C--A-----A----------TAGT-------C---C--T---CC-ACG-----C--A-A---------------A-----A--T-----A--A--GAA---------------CA-A--AG-C-AG----A-------A-AA---GT---GA 3538
CPZ.CM.05.SIVcpzEK505 -----------C--C---------AC----A----------T----------C------A--TAGC--G-----------AT--------------A--------T-----A------------G----------T---G--AG----G--T--G------G--GAC-T-TA-CT- 3530
CPZ.CM.05.SIVcpzMB66  -------A------C--------------A---C-T--------A-----C---C------CC-------------A--T---------------------A--A--C---------G---------------A--AG-C-A------------GA-CA---T-TT---- 3504
CPZ.CM.98.CAM3        -----N--C--G-----------C-T--A-------------A--G--A-C--------CC-----------G--------G-----------------A-----A--GAA------------C--C--C---G--------C-----G---A-A---A-A-T--GA 3369
CPZ.GA.88.GAB2        -----A--C--C-GA-----A---C-T--A--G--------T--A--G--A-------T---CA-----------AT-----T----------C--T---------A-G-AG-------------------T-----AT----A---G--G--CA-TT---T-GGA 3351
CPZ.CM.98.CAM5        -----A-----A-----------C-T--A----------------A--G--ACC-----------CC---G--G-----A------------------------A--A--A--GAA------------T--CA----AG----G--C-----G---A-AA-A-ATT--GA 3660
CPZ.CM.05.SIVcpzMT145 -----A--CA-A--C-----A---C----A--G--------------------C---G--T---A----------A-A------------------C-----A--A--A---AG------------C-----G--A--C--G--G--------GA--CAT-GT--TGA 3500
CPZ.TZ.00.TAN1        -----A---A----C-----A--AC--A-C-GGC--------A-TC-GTT-TCT--C-----CCC--A-----------T----------------A-----A--C-G---AAG----G-------------A--A--C-----C---T-C-GGTCAAAGGA-T--GA 3602
CPZ.CM.05.SIVcpzLB7   -----A-A-------------T--CAT--C--G--T--------A-----C------T---CC---G-------A--T--------------------A--A----G-----------------------A--AG----G--G-AG--G---A-AA--C--T---- 3513
CPZ.US.85.CPZUS       -----A-----C-----------C-C--A---C-T-----TA---------C---GG-----AG--G-------A--T-----G-----A-----CC-AC-A--A--C--GAA-----G----------------G----G--T----------A--A-A--TT--GA 4018
CPZ.GA.88.GAB1        -----A---A-C-------A-------C-------------AGT-------C------T---CC-ACCA---T--T-AT--------------A-----A--T-----A--A---ACT--G--------------A--AG-C-A---G-A-------A-CA-T-G-T--GA 4019
MON.CM.99.L1          ATA---GCG--GT--T----A-A-C-GC-C-G-CAG-----TACTC-G-TAC----C-------A---G--TCC---A-A---C--G--------C--C--T---ACAT-A--GGA---G--------T-----A--C--CTTT--GGC-AG---G-CC-----AT--GA 3661
MON.NG.x.NG1          ATA---CCA--AC-CT-T--A-A-C-GC-T-G-CA-------ACC----T-C--R-T-----T-A-CAC--TCCC----A-----------------G--------A-CAT-A---GA---G--G-------------C--C-G--C---CCA-AG---G--AA-TC-AT--GA 2254
GSN.CM.99.CN166       ATA--AGCAA-AC-C--A--A-A-C--C---G-CA-------ACCC---TCTCG------C---A--AT----TT--T-AT-------------A-----A--T---ACCT-C---C-----------T--T-----G---C----C-----A---T-GG-CCA-AGAAT--GA 3686
GSN.CM.99.CN71        ATA--AGCA--AC-C--A--A-AT--GT---G-CAG------ACCC---TCTCG------C---A-AAT----TC----AT---------C--------A--T---ACCT---------------------T--C-----G---C----C-----A-T-T-G--CCA-AGAA---GA 3668
MUS_1.CM.01.1085      ATA--AGCA--GC-----A-C-A-C-GT-A---CAG----T-ACA--G-TACC------AT--AA--A---TCCC--A-AT-----G-----A-----A--T----TAT-C-----G-----------------C--C--A--C-C---G-A-G----G--AA-T-G----GA 3648
MUS_1.CM.01.CM1239    ATA---GCA--AC-----A-C-A-C-GC--CG-CAG----T-ACA--G-T-CC---G--AT-GTC--A---TCCC--A-AT--------------A-----AC-T-AA-TAT----G-----G--G----------------C-C-----A-G-----G--G--T-G-T--GA 3638
MUS_2.CM.01.CM1246    ATA--AGCA--AC-C-----ATA-C-TC---G-CAG----TTACC--G-TCCC---G-------A-AC----CCC----A------G-----A-----A--T---A-CT-----C------G--T---------A--CG----G---CAGG-G-G--C-A-A-ATT-GGA 3746
MUS_2.CM.01.CM2500    ATA---GCG--GC-C--A----A-C-GT---G-CA-----T-ACTC-G-T-CC---G--TC---AG-CG---CCC---AT-----T---------C----T--AA-CT-C--------T-GG-----G-----C--AG---------CAGG----G-CCA-AGA----GA 3705
RCM.NG.x.NG411        -----A-------G---T--A--AC--A-A-GGC--------A-TC---TA--G--T--T---CC-----A---GT--AT--------------------A--T---A-T-G---G--G-----G-----------A--AG----G---A-G----G--AAAAGAAT--GA 3452
RCM.GA.x.GAB1         -----A-------GC--A--A--TC-TA-T-G-C-CG-----A-CC---TT--G--T------CC-----AG---GT--A-------G------------------AA-T-G----A------G--G------------A----------A-G----G--AAAAGAAT--GA 3437
DEN.CD.x.CD1          -----AA-G--ATCA--T--A---C-TC-A-G-AAT-TT--TA-T--G-TATC---G-------ACAA---GCCC----AT---AC------A-----A--T--AC---------CA-----G-----------GG-C-CC-AT--C-AG-----G--A--A---T-GGA 3848
```

Pol RT end Pol p15 RNAse H start

```
MAC.US.x.239          -----T---A-CTCA--A--A--GC----A-G----GTC-T-A-TC---T---G--C--T----AG----A--------AT---AC------AT--TG---T-AAC--T-A---GA---G------------A-C--AG----G--C-A-G-C-----AAAAGTGT--GA 4257
Pol                   _E__W__D__F__I__S__T__P__P__L__V__R__L__V__F__N__L__V__K__D__P__I__E__G__E__E__T__Y__Y__T__D__G__S__C__N__K__Q__S__K__E__G__K__A__G__Y__I__T__D__R__G__K__D__K__V__K__V__L
H2A.GW.x.ALI          -----C-----ATC------A--AC-G--C-GGC--GCA-TTA-CC---TA-----T--T---C----C----G--------CAC---C---C-CTG---T---C-AT-A---GA---------------A-A--AG-------G---G-C--G--GAAGGTA---GA 4333
H2A.DE.x.BEN          -----C-----ATC------A--AC-G--C-GG---ACA-TTA-CC---TAGG---T--T---CC---C-----G---------CAC------AT--TGC--T--AC---T-A---GAG-----------------A--AG---------A-G-C-----AAAAGTAT--GA 4340
H2A.SN.x.ST           -----C--CA-ATC------G--AC-G--C-G----GTA-TTA-CC-G-T------T------C----C----------------CAC------AT-CTGC--T-A-C-AT-A-G-GA---------------CA-A--AG-----------G-C--G--GAGGCTAT--GA 3785
H2B.GH.86.D205        -----C-----CTCG-----A--A---A-A-G-C--GCC---A-CC---TC-----C---C---A---GAG--------A---CAC-------T-CTGC--T--AACCT-A--GGA------------------C---G-C--G---A-G-T--G---AAAGTGT--GA 4313
H2B.CI.x.EHO          -----T-----CTCA-----A--A---A-A-GG---GCC--TA-CC-G-TC-----C---C---A-----T------T-A---CAC------AT-CTG-----AA-CCT-A---GA---G---------------C--AG-C--G---AGG-T------AAAC-AT--GA 4311
H2G.CI.92.ABT96       -----C-----GTC---A--G-ATC----A-GG---GTC---A-CC---TC--------TC---AGCAGA----G----AT---AC---C---T-CTGC--T-AAACCT-C--GGAR-----------C--C--C--AG-C--------A-G-C--G--AAARGT-T--GA 3667
H2U.FR.96.12034       -----T-----ATCC--A--A---C-G----G-C-GGCA--TA-TC---TA--------AT--C--...-A---G--A-AT---AC------CT-TTG--------CTT-A---CAG--G---------C-----G--AG----------A-G-C-G---AAAAC-TT--GA 3819
MAC.US.x.EMBL_3       -----T---A-CTCA--G--A--AC----A-G----GTC-T-A-TC---T---G--C--T----AG-T--A--------AT---AC------AT--TG---T-AAC--T-A---GA---G------------A-C--AG----G--C-A-G-C-T---AAAAGTGT-G.. 3729
SMM.US.x.H9           -----C-----GTCA--A----------C-G-C--GTT-T-A-CC---TA-----G--T---CAG-----------A--T---------A-AT-CTG---T---C--T-A-G-GA---------------C-----G--AG-------C---GGC--GACAAAACTG--GGA 3730
SMM.US.x.PGM53        -----C-----GTCA--A----------C-G----GTC-T-A-CC---TA--------T---CAG---------G--A--T-----------AT-CTG---T---C-AT-A-G-GA------G-----G-----C-----A--AG-------C---G-C--G-CAAAAC-TT-GGA 4174
SMM.US.x.SIVsmH635F_L3 -----C-----GTCA--A----------C-G----GTC-T-A-CC---TA-----G--T---CAG-----------A--T--C----------AT-CTG---T---C--T-A-G-GA---------------C-----G--GG----G--C---G-C----CAAAACTTT--GA 4260
SMM.SL.92.SL92B       -----C--CA-CTCC--A--A--G---A-C-GG---GT--TTA-T----TA-----T--A----A-AAG-AG--GTA-AT--CA-T-----CT-CTGC--T--AA-C-G---GA---G---------------C--AG-C--G---A-G----G--CC-GC-AT--GA 3688
STM.US.x.STM          -----T-----ATCC--A-----T-----A-G----GTC-TTA-T----TA---------T-G-A---A-------A--T-----G-----CT--TG-------C-AT-----GA---G-------C--CA-A---G---------A-A-C-----AAAAG-TT--GA 3904
MNE.US.x.MNE027       -----T---A-CTCA--G--G--AC----A-G----GTC-T-A-TC---T---G--C--T----AG----A--------AT---------------T--TG---T-AAC--T-A---GA---G------------A-C--AG----G--C-A-G-C-----GAGAGT-T--GA 3725
DRL.x.x.FAO           -----A---A-ATCA--A--C-ATC-CA-C--G---------TA-C--GCT--GT--C--A---CC---G-A---GTT-AT-----G-----A-----------A-TT-G----GA---------T--C--------CTCC--G----A-G-------AA-AG--A--GA 3525
MND_2.x.x.5440        -----C--CA--TCA--A--A--A---A-A-GG---------A-CC-GCTA----C--A---CC---G-A---TGTA-AT------------------A----AA-CT-C--GGA---C--G----C---TA---CTC------T-A-AGC--G--A--AG--T-GGA 3519
MND_2.CM.98.CM16      -----A-----ATCC--T--A-A-C-TA-CGGG--G-----TA-CC-GTT--G------TG-GCC-----A---TGTG-AT-----G-----A-----------AA-CT-C---GA---G---------------A--AGC----AAT-A-TCC-G---AA-AG-----GA 3959
MND_2.GA.x.M14        -----A--CA-CTCA--A--A--AC--A-A-GG--G-----TA-CC-GTTA--------A---CC---G-AG--TGT--AT-----G-----------A--T--AA-CT---GGA---G------------CTA---AGC---G----A-AGT-----AA-AG--T--GA 3886
OLC.CI.97.97CI12      ATA--------ATC--ATT--GAAAATCAA-GGAC--TC-GGG-------T-CC--C------G...-AG-T--------------G----A---TA---T---A-A-GC--GGA------G-----------C-G-TCAG-G----AG--T---A-A--ACA-A--CA 3823
GRV.ET.x.GRI_677      -------------GC--A--A-T-C----A---C-G-----TTCC----TA--------A--CAA-----A---TGTT-AT-----G----------ATC--AA-T---C--------T--G--------------C-GT-AG-G-------A-AGT-G-A--AGGGAAT--GA 4131
VER.KE.x.TYO1_patent  -----T-----C-GC-----A--T--GC-C---C---------ACA---ACA------------CCCAAG-AG--CGTT-A-----------A--TGC-----AA-TT-A---GA---------------CA-CT-AC-ATAC----A---G-G---A-AA--AT--GA 4182
VER.KE.x.9063         -----A-----G-G-GTA--A--TC-----------G-----ACTC-GACC--G-----T---CC-----AG--TGTA-AT--------------CTG---T--AA-CT-A---GA---------------CA---AC-ATAT----A---G-G---A-AG-AAT--GA 4185
VER.KE.x.AGM155       -----C-----C-G-GTT--G---------A-CT--G-----TACAC-GACT--G----------CCCG---AG--TGT--A------------A--CTG---T--AC--T-G---GAG-----------C--CA-A--CC-ACA---C-A-----G---ACAACAG---GA 4177
VER.DE.x.AGM3         -----A-----C-G-----A---C---A-------------TACCC-GACA------------CC-AAG-A---TGT--A-----G-----A--TTG---T--AA-TT-A-GGGA----------------A-C--AC-ATAT--G-A-----GG--G-AA-AAT--GA 3680
TAN.UG.x.TAN1         ATA--A--CA-CTCC--A--G------A-C-----------TA-C--G-TA-----T------CC-AA----A-GGTG-AT---TGC---------TGC--T--AA-TT-C-----------G-----C-TCC-A-----C-C----GA-TC-C-G--CA-AGAAT--GA 4157
SAB.SN.x.SAB1C        -----AA-G--C-G---T----AA--GA-T-G---G-----A-A----TA-----T-----CCC-----A--C-GT--AT--------------------T--AA-TT-C---GA------------------CT-A--AG-------GGAT-----G--G--AG-AT--GA 4356
MND_1.GA.x.MNDGB1     GTA--A---A-C-GC--A---ATG----AA-GGGA------TAGC--GTCCCC------TC---AG--G-T------A-AT-----T-----A----------------C-G----A-G----------------A---AG-------TTTT----GG--A-AAGAATATCT 3645
DEB.CM.99.CM40        ------GCAA-AC-------A-ATC-G--C--GC-C-----TG--C-G-T-TC-----------CC--AT-----C---A-------------------------------T-C-------G-----G----T--C-----C--AG-GT-G----A-----GT--AAAATG-T--GA 3764
DEB.CM.99.CM5         ------GCG--GC-------A-AAC----C-GGC-------TG--C-T-TATC---G--A---CC--AT--T--C---AT-----------A--------T-----AT-A---A-G-----------C--CA-AT-AG-AT-G----A---GGG---AAAA-GT--GGA 3758
TAL.CM.01.8023        ATT-TCCCCA-AC---------------C-GGC-------TA-CC---TATC---G-------A-----------A-A------G-----T--T--CC----A--AT-A--GGA---C-----G---C-----G--C-CGTC-----A-G--C-T--A--AGGA--T-G 3422
TAL.CM.00.266         ATC-T-CCCA--C-C--A--A--A-----T-GGC-------TA-CC---T-C-G--G-------A---------C--G-AT--C-----C--T-----CC--C-A---T-A---GA---C--G---------C--G---GCA-C-----A-G--C-T--CA---AGT--GA 3924
SUN.GA.98.L14         -TA--A-----C-GC--T---AAG--G-A-C--GAG-----TACTC---CC-C------AG-GCC----...--C----A------G-----A------G-A-A-CTAGAA--GAG---G------------A-A--CC-GG-------TCC-G----AAG-AAT-GGA 4312
WRC.CI.97.97CI14      ACA--ACA---GTC---G---TT-C-T-CT-GGGA------TA------T-CT------TC--CC--A--T------T-AT--A-------AT-T--A----A--C-G---GA---G------------...TGG-CC-A-TC--A-G---------A-GTG-GCA 4252
WRC.CI.98.98CI04      ACA--ACA---ATCA------TA---G-CA-GGGA-------A-AA---TACTG--G--T------AG-TG----A-AT--CA-------AT-T--A--T--A--C-G---GGA---G--------C-TC...TGG-CA-ACTCT-A-G-------AA-T-ATT--TA 4267
WRC.GM.05.Pbt_05GM_X02 GTA--ACA---GTC---T---TTTC--AAA-GGGA------ACA-----TATC-----T----A---G-T----GAAA-G---A--------T-T-----T-AA----GC---GA---G------------...TGG-C--AGACT--GA-C--GA-ACAG-AAA-C-- 3181
LST.CD.88.524         -TA--A--CA-CTCC--A----AA----AAC-GGA------TAGC--G--AGC------------A---TA----T----A------G-----A--T--AG-A--A-TAGG----AC-----------G---A-A---C-ATC-----A-G-G---A-AAAAGAAT--CA 3226
LST.CD.88.447         -T---A---A-CTCA-----A-AG----AAC-GGA------TAGC-----AGC-----T----A---G-T---T----AT-----------A--T--AG-A-AA-TAGGG---AC------G-----T-----C--AC-ATC-----AGG-G------AAAGAAT---- 3229
LST.KE.x.lho7         -TA--A--CA--TC------G-AA----AAC--GAG-----ATT-GG---GC----------A----G-T---C----A-----------A-----AG-A-A--TAGGA---AC-----------------A-A--AC-ATC---G-A-G-G-----AAAGGAGT---- 4313
SYK.KE.x.SYK173       GTCA-A--CA-A-G------G--A--G--T-GG-GG-AC--TA-C--G-TACC----------CC--A-----T----------C--------A-------T--A--C-GC--GAC-----T-----------AG-A-G-GAT--G-C---G-G---ACAGTAT--GGA 4158
SYK.KE.x.KE51         GTAA-A-G---GC-C-----CATA-----T-G--GG-AT---A-T--G-TTTC---G--AG--...-A---------A--T-----C--------C-----T---A-T-GC---GA----C-------T--CA-A--AG----G--T-A---G-C-T-AAAG-GGT-GGA 3813
COL.CM.x.CGU1         ATA--------CTCAT--T--TATG----A------GTA-GGA-C--G-TTTT---G--TC-GAAG-AG---CC---T-AT-GGAC------A-G-TGCTC----A-ATTAGG-GCG-----G-------GGA-A-A-TCA-A----GA-G--G----A----AT--TCA 3727
```

```
H1B.FR.83.HXB2               TGACACAACAAATCAGAAGACTGAGTTACAAGCAATTTATCTAGCTTTGCAGGATTCGGGATTAGAAGTAAACATAGTAACAGACTCACAATATGCATTAGGAATCATTCAAGCACAACCAGATCAAAGTGAATCAGAGTTAGTCAATCAAATAATAGAGCAGTTAATAA 4128
Pol                           D  T  T  N  Q  K  T  E  L  Q  A  I  Y  L  A  L  Q  D  S  G  L  E  V  N  I  V  T  D  S  Q  Y  A  L  G  I  I  Q  A  Q  P  D  Q  S  E  S  E  L  V  N  Q  I  I  E  Q  L  I
H1A1.UG.85.U455              ---G-----------A--------A-----T-----CC-------C-----------A----C---------T-----------------G----------G--------G-----------CAGG-----------AA-------------------A--C-----G 3574
H1B.US.90.WEAU160            ----------------------------------T---C--------------------------------------G--------C--------------------------------------------G------------------G-----------A---------- 4127
H1C.ET.86.ETH2220            ---A-------------------A--------G--CC-G-----G--A--------A----C-------------------------G----------G----C-TG----------A-G-----------A-----------------A---------- 3520
H1D.CD.84.84ZR085            ---------------------------------------A--------------------------------T-----------C------------------------------A-G----------A------G----------A---------- 3650
H1F1.BE.93.VI850             G--G--C-----------G------------------T--------------A----C-----G-----------------------------------G--G----------A-G----------AA-------------------A-----C 3468
H1G.SE.93.SE6165             ---A--------C--A---G----A-----G------C-G-----------A--C--AA---C----------------G--------G-----C--------------------------AGG------G----A-------------------A---C------ 3525
H1H.CF.90.056                G--A-------------------A------------------------A-----A--G--------G--------G-----T-----G-----C---------------------C----A-G---------------T--------------G-A------- 3475
H1J.SE.93.SE7887             ----------------------AC----C--C------T-------A-G------A--GC---------------------T----------------G--G--T--A---------------A-G----------A-------------------G--------- 3442
H1K.CM.96.MP535              ---A---------------AG----A----------C-G-T-------A---------A----C-------------------------G--------------------------G------A-------------T-----T-----------------A------ 3324
H101_AE.TH.90.CM240          ---G-----------A--------A-----T--G--CC--T----C-----------A----C---------T----------------------------------G----------CAGG-----------AG-------C------------G--C------ 3702
H102_AG.NG.x.IBNG            ---G-----------A--------A-----T------C--T----C--A--------A----C---------T-----------------G-----------------------G----------CAGG---------------------------------A--C------ 3653
H1N.CM.95.YBF30              A-----C--C-----A---G-------------T--CCT-A-G--C--A--A--G--A---CGG--T------------C--T-----T--G-----TA-G-----A-----TT----G------A-------------A--G--G-GC-----------AG--C-C---- 3722
H1O.BE.87.ANT70              ---G--C--C-----A--AG----A---ATG--G--A-TA-----C--A----------CAA-GA-AC------T------------T----------------G--CG----CTCCT-C-----TACA--G--------CCCTA----TC-G------------G-AC----C-- 4183
H1O.CM.91.MVP5180            A--G-----C-----A---G----A---ATG--TG-A-TAA----C-----------CAAGGAGC---------------------------T---G--C-----ATCCT-C------ACA--G-----C--CCCTA----TC-G--G---------G-AC----C-- 4158
CPZ.CD.90.ANT                AA-G--C--C-----AC-AG-A--A-----G--G---CT-A-G---C-AG-------AAC-GGCCC---C--T-----C-----T--------------G---G--T-G----GTACC-------------------CCCC----GG-AG-------CC--A-A-------- 3567
CPZ.CM.01.SIVcpzCAM13        GA----C--T-----AC--G-A------GGG--CT-G-----------A--A--C--T-A-CA-C----G-----------T-----T-------TG-----G--T-----GT------------------C-----------A-----A---------------AG-A-----T- 3708
CPZ.CM.05.SIVcpzEK505        A--A--T--C--C--AC-AG----A------G--TG-CCTCA-G--C--A--A---G-A---CA---------T-----C--T-----T--G---TCC-------A-----CT-----------------A-------------A-----A-----G---------AG-AC-T---- 3700
CPZ.CM.05.SIVcpzMB66         A--A--C--T--------AG-A--A---------T--AC-AT-G----A--A-----A-A-G-----------T-----C--T--T------------C-G-----A-------GT-----T---AC--------------AA----A-----G---------G-AC-G--C- 3674
CPZ.CM.98.CAM3               AA-T--C--T--------AG-A--AC-GAC---TG----------A--A--A---G-A-A-CATAT---T--T-----C--T--T--C-------TG-----G--A-----CT-C--G--------------G-----A--G--T----T---------AG--------- 3539
CPZ.GA.88.GAB2               A--G--C--T--C--A---G-A--A---G-G---C-CAGAA----CC-A-----C--A--GCA----------TG-G-----------C-------TG--G---------AA-TT-C---------C---------G-----AA----G-----------------AG--C------- 3521
CPZ.CM.98.CAM5               AA-T--T--T--------AG-A--AC--AC---TG-------G--A----------G-A-A-CATAT---C--T-----T--T--T--T-------TGC-------A-----CT-T--G------------------------A-----AGGGTCT--------AG--------- 3830
CPZ.CM.05.SIVcpzMT145        A--T--C--T-----A--AG-A--A----------G---TAA-G--C--A-----C--A---CAG-----C--TG-G--C--T-----C-------TGC-G-------T-A---T-----C-----C--C------------TA----A-----G---------AG--------- 3670
CPZ.TZ.00.TAN1               GA----C--T-----AC-AG-A--A---TGG---G-AG-------C--AA-A--C--A---GC-C-G-----T-----C-----T--C-------TTA-G---G-TT-A--G-G-TT-----------------C--CCCCA----AG-G-----T--TC-AA-------C-C 3772
CPZ.CM.05.SIVcpzLB7          A--A-----C---------G-------------C---C-AT----C--A--A--C--A---CCT--G-----T-----T--T--T-----------------------C-------C-----G-----G-----T-----A--G--G--------C-----AA-TC------- 3683
CPZ.US.85.CPZUS              AA-T--T--T-----A---G-A---C-CAC---GG-A--CT----A--AA-A-----A-AGAATAC---T--TG----C--T-G---C-------TC------------C--CT-T--G----C-------------------A--G---------G---------AG-A------- 4188
CPZ.GA.88.GAB1               AA-T--C--T------C-AG-A--A---A-G--TT-GCT---G--C-------------A-ATCA-C-G--T---------G--T-------------TG-----G--T-----GT----G--------C---------------A------------G---------AG-------T- 4189
MON.CM.99.L1                 AC-G--C--C--C--A---G-A---C--G-G--CG-G-TA-----CC-CAGA---GGACC-CCTC-GA-G------A-T-----T--T--G---T-C-G--C--TC-AGCCT-CTGC--G--GATCTCA---AGTCCCA----GG-GGCC--C---C--G-AC-CT--- 3831
MON.NG.x.NG1                 GC-G--T--C--C------G----AC--G-----G-GCT-T----A-----AA--GG-CCCGCTAGTA-G--T-------------------G----TCA-G------C-AGC-T--A-G-----GGTCTCG--G--CCCAA----AG-GA----C--CC------C-CC--G 2424
GSN.CM.99.CN166              A--A--T---------C-AG-A--AC-C--T---G--CT-T-G--A--AA-A--AGG-CC-CC-A--A-G--TC-G--C-----T--C--G---T-A-G--G---T-G---AGT--G-----GGT-TCCACC--T-C-C----GG-A-----C---C-AG-AC-GC--- 3856
GSN.CM.99.CN71               A-GA--T--------AC-AG-A--AC-C--G---G--CT-T-G--A--AA-A--AGGACC-CC-AG-A-G---T-G--C-----T--C-----C-T-A-G--G---T-A---AGC--------AGT-TCCACC--TCCC-----GG-A-----C---C-AG-AC-GT--G 3838
MUS_1.CM.01.1085             AAGT--T--C--C-----AG-A--A--GG-G---G-C-TAT-G--C--AA-A--AGG-CCTCCTAGGA-G--TT----C-----T--T--G---TC-----G--AG-AGCTT-T--G---C-AG--TCCACT---CCA-----AG-G----------C--G--C--C-GT 3818
MUS_1.CM.01.CM1239           AA-T--T--C-----A---G-A--AC--G----TG-C-TGT-G--C---A-A--AGGCCCTCCCAGCA-G--T-----C-----T--C------TGC-G--G--AG-GGCTT-T-----CC-AG--TCCACC--GCCC--G--AG-AG-------TC-A--A---C-G- 3808
MUS_2.CM.01.CM1246           AA-------C--C---C-AG-A---C-C-----C---CTCT-G---C-AAT----GG-CC-CC----A-G---C----------T--T------------G-----A--CAC----GC------GGT-TCA--G--CCCTC-G--GG-G--G------C-----AA-GC-G- 3916
MUS_2.CM.01.CM2500           AA-T-----C--C--A---G-A--A--G--G--T--A-TAA-G--CC--ACA---GGACC-CC-A--A-G--------G-----T--C--G-----CC-G--T-----CAC----GC------GAC-TCA--G--TCCCC-G--AG-GA-G--C---C-A-----GC-C- 3875
RCM.NG.x.NG411               AA-T--T--C-----A--AG-A-------GG----G-A-TA--G--C--AA-A--C--A----AA-AGG-------------------T--T-G-C---TG--T--C--AT-AGC--G-AGT-----CAC------GAGCCCCC----AC-A--G--T-----A--AA-G---G 3622
RCM.GA.x.GAB1                A--A--T--T-----A--AG-A--A---G-----G-A-TA--G--A--AA-A-----A--GCCTA-------------C-----T-----G---TT-AT-----TC-AG------------------ACT-----C-----G------G-CAG----T---A-T---A-G---G 3607
DEN.CD.x.CD1                 A--A-----C-----A---G-A--A---AT--G-G-C-TAA-G--C--A--A---AGTCCCAACA-G--C---------C--T--T------------CC----C-----AGC--GTAC---CAC-AC-TC----CATC--A----GG-A--G--C---C-AGCTA-GCA-G 4018

MAC.US.x.239                 AC-G--T--T-----AC-AG-A--A--GG-----T--CTCA-G--A---ACA--C--A--GCC-A-G-C---T--TA--GT---T----------TTA-G-----A--AAC--G-TGC--TACAG--TCA--GAGCAG-C----T-------------AG-AA-G--T- 4427
Pol                           E  Q  T  T  N  Q  Q  A  E  L  E  A  F  L  M  A  L  T  D  S  G  P  K  A  N  I  I  V  D  S  Q  Y  V  M  G  I  I  T  G  C  P  T  E  S  E  S  R  L  V  N  Q  I  I  E  E  M
H2A.GW.x.ALI                 AC-A--T--C------C-AG-A--A---G---CT-CGCG--G--AG-AACA--C--A--TCC-A---CC--T--TA--GT---T-----G----T-A-G-----AG-AGC--GC--G---ACAG--TCA---AATAGAA---A--------C-----AG-AA-G---- 4503
H2A.DE.x.BEN                 AC-A--T--C------C--G-A--A---G---TCT--CGGA-G--AC--GCA--C--A--CCC-A-G--T--T--CA--GT---T-----G----T-A-G--G-AG-AGC--GC--G---ACAG-GTCA---AATAGAA----G--C--G--C-----AG-AA-G---- 4510
H2A.SN.x.ST                  GC-A--C--C------C-AG-A--A---G----CT--GCGA-G--AG-AACA--C--A--TCC-A-G-CC-----TA--GT---------------T-A-G-----AG-AGC--GC------ACAG-GTCA--GAGTA-AA----A---------C-----AG-AA-G---- 3955
H2B.GH.86.D205               AC-G---------C--AC-AG-A--AC-TG-----T--GCAT----A--AACA--C--A-A-CC-C-----T-----CA--GT---T----------TCA-G-----A--AGCT-----G---ACAG---CA------CCAA----AGCAA-------T--AG-AA-G--C- 4483
H2B.CI.x.EHO                 AC-A-------------C-AG-A---C-TG-----T--GCA-----AC-A-----C--A---CC-C-G--C--T--CA--GT---T----------TCA-G-----AG-AGCT---------ACAG---CA------CC-A----A-GAG-------T--AG-AA-G--C- 4481
H2G.CI.92.ABT96              GC-G---------C---C--G-R--AC-TG-G---T--GCAA-G--C---------C--A--TCC-A----C--T--CA--GTR--T--C--G---TCA-G--R-----AGC--GR------ACGG---CA-----CCC------A---A-------T--AG-AA-R--T- 3837
H2U.FR.96.12034              AC-G--C--C--C--AC-AG-A--A---G----CT--GC-A-G--C--AG-A-----A---CC-A-----------T--G-----T---------AT-A-G-----A--AGC--G---G---ACAG---CA--G--CCCAC----A-G-A----C-----AG-AC-T---- 3989
MAC.US.x.EMBL_3              .ACA--T--T-----AC-AG-A--A--GG-----------AT-GCA-TG-A--C--A--GCC-A-GAG---T--TA--GT---A-T----GTATGT-AT--G-AT-A-A-CAGGTTC--TACAG--TCA--GAGCAG-C----T--C-----------AG-AA-G--T- 3898
SMM.US.x.H9                  AC-G--T--C-----AC-AG-A-----GG----CT-----------C--AGCA-----A---CC-A---C---T--TA--GT---T--C-------TTA-G--C--AG-AGC--GT-----CAC-G--TCA---AGTAGA-----A--C--G---------A-AA-G--T- 3900
SMM.US.x.PGM53               AC-G--T--C-----AC-AG-A-----GG----CT-C---------C--AGCA-----A---CC-A-G-CC--T--CA--GT---T--C-------TTA-G--C--A--AGC--GT-----CAC-G--TCA---AGTAGA-----A--C--G---------G-AA-G--T- 4344
SMM.US.x.SIVsmH635F_L3       AC-G--T--C--C--AC-AG-A-----GG----CT-C---------C--AGCA-----------CC-A---C---T--TA--GT---T--C-------TTA-G--C--AG-AGC--GT--G--CAC-G--TCA---AGTAG------A--C--G--------AG--A-G--T- 4430
SMM.SL.92.SL92B              AC-GG-T--C------C-AG-A---C-G--G--CC-G-TAT-G--C---A-A-----TCC-AGTA-G--C--TG-------G--------G---TC--GAAC--A--AAC--G---G---TCAG--TCG--T-----TA----GGCA-----T--T--A--AC--G-TC 3858
STM.US.x.STM                 AC-G--T--T-----AC-AG-A--A---G-----T--GCCA-G--C---GC---C--A--TCC-A---C---T--T--GGT---------------TGA-G---------AAC--G---G--CACAG--TCG--GAGCA-------A-----G---------AG-AA-G--T- 4074
MNE.US.x.MNE027              AC-G--T--T-----AC-AG-A--A--GG-----T--CTCA-G--A---ACA--C--A--GCC-A-G-C---T--TA--GT---T----------TTA-G-----A--AAC--G-TGC--TACAG--TCA--GAGCAG-C----T--C-----------AG-AA-G--T- 3895
DRL.x.x.FAO                  A--A--C--C----------G----A-----G--C---CTGT----C---A----C--T--TCC-A----C--T-----G--G--T-----G--C--CA-G--T--A--ATCCT-CTCC-----AATC-----CAATCCTA----G--------T-----A--AA-G--C- 3695
MND_2.x.x.5440               A--A--G--C-----A---G-A--A--GAC---T--AA-GA-G---------A------A---CCCAGG--G--T-----C--T--T--C--G------C----C--AC--TC-----GCC---------GTCA--CAACCCCA----T-GGG----------A-TCA-G---G 3689
MND_2.CM.98.CM16             GA-T--C-----C--A--AG-A--A--GG-G-----AA-AA-G--A--A--A--C--A---CC-A--------T-----G-----T-----G-----TA-G--T---C-ATCG---GC---C--C---TCA--CAACCCCA----A-GAG----T-----A-TCA-G--CC 4129
MND_2.GA.x.M14               AA-------T--------AG-A--A---G-G-----ACTGT----C---A-A-----A--TCCTAGG-C---T-----C-------------------------C--AC--TC---CACT-------G-TCA--CAACCCTA----T-GAG-------CA-C--AA-G---G 4056
OLC.CI.97.97CI12             G--G-GT-GT--------AG-A--AA--AG---T---ATAGA---CC-TAG---C--AAA-AGT--G--C--T--T--------T---TTG-----CA-G-----AG-GTT--AGA-------AGTGTC----GA-CCCC-CA-AC-AG--T-----A--G----GGAG- 3993
GRV.ET.x.GRI_677             AA----C--T--C--AC-AG-A--A---AC----G--A-GA-G--A---G----CAGT---GA-A-T-----T-----C-----T--T-------T-A-GAAC---T-GAC----TGT---C-GG-----A-C---CCC-----GG-A--G------C-AGCCC----G- 4301
VER.KE.x.TYO1_patent         AA----T--C------C-AG-A--A---AC---T--AA-AA-G------G-A--CAGT--GCCTA-T--G-----------------T-----------A-G-----TT-GAC---------CACA--------T---CCA-----AG-G-----T----CCTTAA-G---C 4352
VER.KE.x.9063                GC-G--T-----C---C-AG-A--A--GGT---C--AA-AA-G--C---G-----AGT---GA-A------------------------T----------CA-G-----AT-GAC-------G--CACA--------C--CCCA-----AG-A--G--C----CATTAA-GG--- 4355
VER.KE.x.AGM155              AA------------AC-AG----AC-GAC---C--AA-AA-G--C---G-----AGC--CCCTA----C--T------------T----------GA-G--C--AT-GAC-------G--CACA--G------C--CCCAC----AG-A------------CA---A-GG--C 4347
VER.DE.x.AGM3                AA-T--------C---C-AG-A--A---ATG--C--AA-AA-G--AC-AG-A---AGT--GCCTA-T-------------------T-----------A-G-----AT-AACT--C-----CACA--G------C---CCC---A-AG-A-----T----CA-TAA-GG--C 3850
TAN.UG.x.TAN1                G--T--C--G--C--AC-AG-A--A----------G-CCT------A--AA-AC--AGTAAGAGCA----T--T--------G--T-----G----TCA-G------C-GAC---------TACA---------C--C-T-C----A------------T--AG-AA-G--T- 4327
SAB.SN.x.SAB1C               GA-T--T--C--C-----AG-A------G-G--C---CTGT----C--AAGA--C--T---AGCA-------T---A--------------------TA-G--T--T--AGC--GGG----TACAG--TCA--TAATA-CA----AC-A-----T------G-AC------ 4526
MND_1.GA.x.MNDGB1            AA-T--C--C------C---A--A-----T---G-AA-A------C-AG-A---AGT---AGTT-T--T---------------T----------T-G----T--AC-AGC-AGCAG---TAC-G---CA--TCACCCCA----A--GG----------ATTAA-G-A-G 3815
DEB.CM.99.CM40               GA-T--C--T--------AG-A--A---G----C--A-TGT-G--C---G-A--AGGCCC--C-A--A-G--T---------------C--G------C-G--G--A---TT--A---C---TC-G-G-CA---CATA--A----AG-AA-GG-----C--GCC--GCA-G 3934
DEB.CM.99.CM5                AA-T--T--C-----A---G-A--A---G----C--CCTA--G--C---G---GGGACCC-CCA-GA-G----------------C--G-----T--------AT-GTT--A---C--CACGG---CA---CAT---A----AG-AA-G--C---C-AGCC---CA-G 3928
TAL.CM.01.8023               CA-T--T--T---------G-A--A---G-----G-GCTAT-G--A--AA-AC-C--T--TCC-A-G-----------C-------T-------TC-AT--C---T-AGT--G-A-T--CACAG-G-CA--CAACA--A----GG-AG-C--C---C-----C--T--G 3592
TAL.CM.00.266                AA-A--T--C--C--A---G-A--A--GG-----G-ACTAT-G--AC-CA-A-----A--GCCCA-G-----T-----C--------T--G---TC-AT--C---AT-GGC--G-AGC--CACAG---CA--TAACA--A--A-AG-AG-C--T--CC---TAC--C--- 4094
SUN.GA.98.L14                AA----C--C------C-AG-A--AC--G-G--T--AA-AA-G--C---G-----AGTA-GAGCTC---T--T-----------------G-----C---A--C-GC-CAGTAAGAG---CACAG---CA--T------C----A--GG-G--CG------TCA---GGC 4482
WRC.CI.97.97CI14             AA-T-----T------C-A--A--A--GAT----T-AA-A-----AC-AG-A---GG---TCA-A--A-G--T---A-C------AGT-------TCA-G--G--AG-AA---AT--T--CACAATG-----G--TCCT-----AC-GG----T--T--ATT-----A-G 4422
WRC.CI.98.98CI04             -A----C--T------C-AG-A--A---GCG--TT-GA-GT----CC-AG-A---GG---T-C-A-GA-C--T--TA-T-----TAGT-------TTA-G-----AG-AA---AC--C--TAC-ATT-----G----CTC----AC-AG----T-----T-TTC---A-C 4437
WRC.GM.05.Pbt_05GM_X02       AA-T--T----------C--------C-GATT---T-AAGAA-G--A--AG-A--CAGTTC-GATA-G-----T---A-C------AGT-------TT-----------AA---AT--C---AC-------------TCCC---A-GC-AG----------C-TTC---A-G 3351
LST.CD.88.524                A--T--T--T------C--G-A--A---G-G---G-A-TGT-------A--A---AGCAA--C-AC---T-----------T-----C-------T-A-GAAGG--T-AAGTCAGAG---TACAG---C---GCATCCTA----G--GA-C--C---A--G--C-GCAG- 3396
LST.CD.88.447                A-----C--C-----AC-AG-A--A---G----TG-C-TAT-G--A--AAGA--CAGTAAGAGCAGG--C-------------------------TGA--AA-G-TT-ATCTCA--GG--TACAG---CA---CATCCTG----A--A--G--T------G-A-GT-A-- 3399
LST.KE.x.lho7                ------C---------C-AG-A--AC--G-G---G--CTAA-G--A--A--A---AGCAATAGTA-------T-----------T----------T-A-GAA---AT-GTC-CA-AG----ACAG---CA---CACCCCA----A--AG-C--C-----A--A-GC-AGC 4483
SYK.KE.x.SYK173              GC-A--------C--AC--G-A--A---G-G-GCT-GCTCA-G--A--------C--AAA-GACA-------T------GT---T--------CT-T-AT--C--TC--ATGA--TGC--CAC-A-C-CA---CACCCAA----GG-G--G-------C-AG--GCC---- 4328
SYK.KE.x.KE51                AA-------C-----A---G--------G-----G-A-TAA-G--A--A--A--C--A---CC------T---------------T--T--G-----------------AT-AATGAACTGT--TACATCCTCG---CACCCAG-----G-A--G--T--GC-AA--GCT--GG 3983
COL.CM.x.CGU1                ---AGGGT--------C-AG-A---C-GAC--G-G---TG------C-AA-AC--GG-CC-AA-AG-A-G--TT-G---------AGT-TT-----------G--A---ACT-GG--G---T--G-TCAGCTTC--TCTAGTAGAGGGG-G--T--TC--GCAGGT--GG 3897
```

Pol p66 RT, Pol p15 Rnase H end Pol p31 Integrase start

```
H1B.FR.83.HXB2        AAAAGGAAAAGGTCTATCTGGCATGGGTACCAGCACACAAAGGAATTGGAGGAAATGAACAAGTAGATAAATTAGTCAGTGCTGGA...ATCAGG......AAAGTACTATTTTTAGATGGAATAGAT...AAGGCCCAAGATGAACATGAGAAA...TATCACAGTAAT 4283
Pol                   K  K  E  K  V  Y  L  A  W  V  P  A  H  K  G  I  G  G  N  E  Q  V  D  K  L  V  S  A  G  .  I  R  .  .  K  V  L  F  L  D  G  I  D  .  K  A  Q  D  E  H  E  K  .  Y  H  S  N
H1A1.UG.85.U455       ----------A-----C---T-------------G--------G--------------G---------------------T-----...------......--G--G--------------G------...-----T-----G--C-----A---...------T-C--C 3729
H1B.US.90.WEAU160     ----------------C--A------------------------------------------------------------T-A---...------......--------------------G------...-----------A------------...------------ 4282
H1C.ET.86.ETH2220     GC-------G------C---T----------------T--------------------------------------A---AG----...------......-----G--G--C---------------...-----T-----A--G-----A---...--------C--- 3675
H1D.CD.84.84ZR085     ----------------C-----------------T---------------------------------------------AG----...------......--------------G------------...-----T-----G------------...--------C--- 3805
H1F1.BE.93.VI850      ---------G------C---T----------------------G------------------------------------------...G-----......--A-----G-----------G------...-----A-----A--------A---...-------AC--- 3623
H1G.SE.93.SE6165      ----------------C---T----------------------G-----------------------C------------AG----...------......-----------C--------C------...--A--------A--G--------G...-------AC--C 3680
H1H.CF.90.056         -G--------------C---T-------------------------------------------------------T---T-----...G----A......-----G-----C--------G------...--A--T-----A--------A-GG...-----T-AC--- 3630
H1J.SE.93.SE7887      -G--------------C---T-G--------------------------------------G------------------T-----...------......-----G--G--C--------G------...--A--T-----A--T-----A---...-----T--C--- 3597
H1K.CM.96.MP535       ---------G-A----C--AT-----------------------------------------------------------------...------......-----G-----C---------------...-----T-----A--------A---...-----T-AC--C 3479
H101_AE.TH.90.CM240   -------G--A-----C---T-------------------G--G------------------------------------T-A---...------......--G--G--------------G------...-----T-----A--------A-G-...--------C--- 3857
H102_AG.NG.x.IBNG     -------C--A-----C---T----------------------G------------------------------------AA----...------......--G-----------------C------...--A--------A--G-----A-G-...--------C--- 3808
H1N.CM.95.YBF30       ---------GA--T-----CT-T--------T-----T-----T--------------G--G-----C--------T--CT-A---...--T--A......---A--T---CC--------T-----A...--A--------A--T---C-G---...--------C--- 3877
H1O.BE.87.ANT70       --------C----G-----TA-------T--T--C--T-----C--A-----------A--A--------------A--CAAG-AT...--T--A......--G---C--G-CC---A---------C...C----A-----A--T-----A---...-----T------ 4338
H1O.CM.91.MVP5180     --------CGA--G-----TA-------T--T--T--------C--A-----------A--A--------------A--CAAA-AC...--T--A......--G---C--G-CC-G--A---------...C----A-----A--T-----A---...-----T------ 4313
CPZ.CD.90.ANT         ---GA---C--A-T--C--CT-C--------------T-----C--A--------------G-----C------------CAA--G...-----A......C---C--C---C-G--A--C-------...--A--T-----G--C-----T---...-----T--C--- 3722
CPZ.CM.01.SIVcpzCAM13 -------C-----T--C--TT-C--------T-----T--G--C--A-----------------------G---------T-----...------......-----------CC-G------------...--A--A--G--A--------A-G-...-----T--C--- 3863
CPZ.CM.05.SIVcpzEK505 ---------GAA-A-----TT-C--------T-----T-----T-----------------GA----C-----------CA-A---...------......------CT--CC--------T------...-----T-----A--------A-G-...-----T------ 3855
CPZ.CM.05.SIVcpzMB66  ----A--G--A--G-----CT----------T-----------------------------A-----------------CT-A---...--T---......------C--C-CC-G-----G------...--------G--A--G-----A---...--C--T-A---- 3829
CPZ.CM.98.CAM3        -G--------AAGT--C--TT-------G--------T-----G---------------------------------T--T----G..A-----A......------T-C-----------T------..T---CA-A-G--A--------G--A...-----T-AC--- 3697
CPZ.GA.88.GAB2        -G--A--G--------T--A--------------------------A--------------G--------G--------AG--C--...-----A......--G--G--------------T------...-----A-----A--C-------G-...-----T--C--C 3676
CPZ.CM.98.CAM5        ----------AAGT--C--CT-------G-----------------------------------------C------T--T-A-C-...--T--A......------C-C-C---------T--C---...-----A-G-A----------A--G...-----T-AC--C 3985
CPZ.CM.05.SIVcpzMT145 ----------AA-T-----AT-------------------G--T--A--------------G--------C-----T--T-A-G--...-----A......------T-T--C--------------A...--A--T-----A--------A-G-...--C---TC---- 3825
CPZ.TZ.00.TAN1        -----AC-GCAA-T-----A--------------C--T-----T--A--G--T--------G-----C------G--T--AAAAAT...--T--A......---A--T---CC-G---------TA--...G-A--A--G--A--C-----T---...------------ 3927
CPZ.CM.05.SIVcpzLB7   ----------A--T--CT-AA----------T-----T--G-----A--------------------C------------T-A---...--T---......------C-T--C--G------------...CGA-----G--A--G-----A---...--C--T-A---- 3838
CPZ.US.85.CPZUS       -G--A-----AAGT---A-CT----------------T-----------------------A--------C---------T-A-G-...-----A......------T-T-CC--------C------...--A--A--G--A--C-----C---...-----T------ 4343
CPZ.GA.88.GAB1        ----A-----AA----C--CT-C--------------------T--A-----------G--G-------------------A----...------......------G---CC--------------C...-G---------A--------A-GG...-----T------ 4344
MON.CM.99.L1          -------CC----A-TCT--T-------------T-----------A-----------G--------CGCC-----GG-CCAG---...--T---......--C---C--G--A-G--AAATT-G--A...CCC--AAGG--A--T-----A--G...-----T------ 3986
MON.NG.x.NG1          G---AA-GC----A---T--T-------C--G--C--------C--A--G--------G---R--C----------A--YCAA--G...-----A......C-G--C--C--CA-G--GAAC--T--A...CCR----G---G--T-----A---...--------C--- 2579
GSN.CM.99.CN166       --------GCCA----CT-AAGT-----T--G--------G-----A-------CAC-G-G--G--C------------CAAG---...---C-C......C----G-T--CC-G--AAAC------G...CCT--AGT---A--C-----A--G...-----T------ 4011
GSN.CM.99.CN71        GG--A--GGCC-----CT--AGT-----T--------------G--A-------CAC-G-G-----C------------CAGA--G...---C-C......C-------C-CA-G--GAAC------G...CCC--AGT---G--C-----A---...-----T------ 3993
MUS_1.CM.01.1085      C---A---GCCA-A--CT-AT-------G-----C--T--G--T--A-----------GG----G--C--------A-A-AG----...------......C----C----CA-G--ACAG------A...CCT--AA-G--A--T-----A---...------------ 3973
MUS_1.CM.01.CM1239    CC----GGC----A--C--AT-G-----G-----C--T-----T--A-----------GG-T--G--C---C-G--A--CCA----...--T---......C-G-G-----CA-G--GCA-------G...CCT--AA----G--C-----A--G...--C--T--C--- 3963
MUS_2.CM.01.CM1246    GC--AA-TGCCA---T-A-AT-CC-A--------C-----G--C--A-----G-----GG-------CC-CC----T--CAGG---...---C-C......C-G--------A-G--AAAT------C...CC---AGTT--A--C--------G...--C--TTC---- 4071
MUS_2.CM.01.CM2500    CC--A---GCAA-A--C---T-C-----------C--T--------A--G-----CC-GG-------CC-CC----T---AAA--G...---C-C......C-G--T--G---A-G--AAAC--T--G...CCA---AT---A--C-----T---...------TCC--- 4030
RCM.NG.x.NG411        GG------G-A--G--C--CT-C--------G-----T-----C-----G--------G--G--------------G--CAG----...--T--A......C-G-------CC-----C--G----GA...--A--A--G--A--------A---...-----T--C--- 3777
RCM.GA.x.GAB1         GG--A---GCA--G--C--TT-C-----G--C-----T--G--C--C-----------GG----------------T---AAA---...-----A......C-G-----G-CC--------G-----A...--A--A-----A--------A--G...-----T-A---- 3762
DEN.CD.x.CD1          GC------G-AA-A--CA-CA-C-----T-----C-----------A-----C---AC-G--A---C--GC-----T---AAA--G...-----A......C--A-C-----C-----ACA---TCCA...C-A--T-----A--------T---...--C--T--C--- 4173
```

Pol RNAse H end Pol p31 Integrase start

```
MAC.US.x.239          -----TC-G-AA-T---G-A-----------------------T-A--------CC--G--A----CC-CC-----T---CAA--G...--T--A......C---T--C--C--G--AAAG------G...CCA--A-----A--------T---...--C--T------ 4582
Pol                   I  K  K  S  E  I  Y  V  A  W  V  P  A  H  K  G  I  G  G  N  Q  E  I  D  H  L  V  S  Q  .  G  I  .  .  R  Q  V  L  F  L  E  K  I  .  E  P  A  Q  E  E  H  D  .  K  Y  H  S
H2A.GW.x.ALI          --------GCCA-----G-T--G-----C-----C--------C--A--G------C-GG-------CC-T------A--CAG--C...-----A......C-----T-G--C----AAA-------G...CCA--T--G--A--------A--G...-----T--C--- 4658
H2A.DE.x.BEN          -G------GCA------G-T--------C-----C--T-----C--A--------CC-GG-------CC-T------A--CAA--C...-----A......C-----T----CC-G--AAAG-----G...CCC--T-----G--------A---...-----T--C-T- 4665
H2A.SN.x.ST           --------GCAA-----G-T--------C-----C--T-----C--A---------C-GG-G-----C-C-------A--CAG--C...-----A......C-----T----CC---GAA-------A...CCC--T--G--G--------A---...-----T--C--- 4110
H2B.GH.86.D205        ----A--GGCA--A---G-A-G------------T-----G---C-G--T--T---C-GG-------CC-CC----A---CAA---...-----A......C-G--CT-G--CC---AAA-------A...CCA-----G--A--G-----A---...------TG-C-- 4638
H2B.CI.x.EHO          ----------AA-A---G-A-G------------T-----G---C-G--T--T---C-GG-------CC-CC----G--CCAA---...--T--A......C--A-C------C----AAA------A...CCA--T-----A--------A---...-----T-A---- 4636
H2G.CI.92.ABT96       ----A---RCTC-A---G-A-G------------T--T--------A-----T---C--G-G-----CCRTC----A--CCAG---...--T--A......C----------C--GR-AAA------A...CCT--T-----A--------A--R...-T---T------ 3992
H2U.FR.96.12034       GG--A---GCA--G---G-T-GC--------T--T--T--G--T-----T--C---C--G--------C-T--G--G---CAG---...--T---......C-------G---C---AAA-------A...CCA--T--G--A------------...--C--T--C--- 4144
MAC.US.x.EMBL_3       --...-TC-G---T---G-A--------------TTAG--T--A----------CC--G--A---G-CC-C-----T---CAG--G...T-T--A......C----T--C--C--G--AAAG-----G...CCA--A-----A--------T---...--C--T------ 4050
SMM.US.x.H9           --------GCAA-T---G-A-----------T-----T---RR---A----A----C--G--------C-TC----T---CAA-A-...--T--A......C---C-----C-----AAA-------A...CCA--G--G--A--G-----A---...--C--T--C--- 4055
SMM.US.x.PGM53        --------GCAA-T---G-------------T--T--T--------A---------C--G--------C-CC-G--T---CAA---...--T--A......C---C-----C-----AAA-------A...CCA--A-----A--G-----A--G...--C--T--C--- 4499
SMM.US.x.SIVsmH635F_L3 --------GCAA-T---G-A-----------T-----T--------A---------C--G--------C-CC-G--T--CCAG---...--T--A......C---C-----C-----AAA-------A...CCA--A-----A--G-----A--G...--C--T--C--- 4585
SMM.SL.92.SL92B       -G--A---GCA--A--CA-A-G------T-----T--T-----C--A-----G--CA-TG-G-----CCGT-----T---CAA---...--A--A......C-G--C--G--CC---AA-T------G...CCA--A-----A--C-----C---...-----T--C--- 4013
STM.US.x.STM          --------GCAA-T---G-A-----------------T--------A---------C-GG-G------C-T--G--A---CAG---...--T--A......C-G--C-----CC---AAA-------A...CCA--T-----A--G-----A---...-----T--C--- 4229
MNE.US.x.MNE027       -----AC-G-AA-T---G-A--------G--------------T-A--------CC--G--A----CC-CC-----T---CAA--G...--T--A......C---C--C--C--G--AAAG------G...CCA--A-----A--------T---...--C--T------ 4050
DRL.x.x.FAO           GT--A---GCA--T--CT-AAAT--------------T--------A-----T-----GG-G-AG--C---C----T---AGA---...--A--A......C----GT----CC-----AAC--G--A...--A--A--G--A--------TTT-...--C--T-A---- 3850
MND_2.x.x.5440        GC------G--A-T-----CT-------C-----C-----G--T-A------T--------GA-------------A---AAG---...--A--A......C-G--------------G---------...--A-----G--A--------C---...-------A---- 3844
MND_2.CM.98.CM16      -C------GCA--A--C--A--T--------------------C--A-----G-----G------------C----G--CAGA---...G-A---......C-------G--------A--C------...--A--A-----A--G-----T---...--------A--- 4284
MND_2.GA.x.M14        CC--A---GCT--A--CT-A--------------T--T-----C--A-----C--------A---C----------A--CAAA---...-----A......C----------CC-G--A---------...--A--T--G--G--------C---...--C--T-A---- 4211
OLC.CI.97.97CI12      -G--A----GAA-A---A-T-G------T--------T--G-CT--A--G--C--C--G--------C------------AAA---...-A---ACTAAGC----GA-G----G-CA-AC-------G...CCA--A-----GAGC--CA-T--G...------TCAA-- 4154
GRV.ET.x.GRI_677      -----AGGC-------CT-ACA------------T--T--G--G--A-----C---AC-G--A-------------A--CAAA---...--A--A......C-GA-C--C--C------A-------A...G-A--A--------C----CA--G...--C--T-AC--- 4456
VER.KE.x.TYO1_patent  -----C--C-AA-A---T--CAG--------------T--------A-----------GG-GA-------------G---AAA--C...--T--A......-G---TT----C-----AAA------A...G-A--T-----A--G-----A-G-...-----T-A---- 4507
VER.KE.x.9063         -G--AAGGGCAA-A--C--ACA------G--T-----------T-A--G---------G--A---C----------A---CAA---...--A--A......-G---T-----A--GAA---------A...G-A--T-----A--------C-G-...-----T------ 4510
VER.KE.x.AGM155       -G--A---GCCA--------CA---------T--T--T-----T--A--G--C-----G--A---C----------A--CAAG---...G-T--A......-G-A--T-G--CA-T-GCA-G-----A...G-A--A-----A--------T-GG...-----------C 4502
VER.DE.x.AGM3         ----AC-TC--A-A--CT--CA------------G--------G--A-----C-----G-GA---------C----A---CAA--G...-G----......----A-TT---------AAA------A...G-A-----G--G--------A-GG...--C--T-A---- 4005
TAN.UG.x.TAN1         -T--A--TGCTA-----T-AA-C--------------T-----G-A--G-----C-----G--A------GC----A--CCAG--G...--T---......C----G-----C-----CA----G--A...G-A----G-ATC--------T---...------T-C--- 4482
SAB.SN.x.SAB1C        -G-----GGCA--G---A-A--------------T--C--T--G-A--T--C------GG--A-T------C----T--CAA----...--A--A......C----------CC-G--CA----T--A...G-A--A-----A--G-----T---...--C--TGCC--- 4681
MND_1.GA.x.MNDGB1     G---A-----AA-T---T-AAGT---C----------------G--A-----G-----G--A-------GC-----A---T-A---...-----A......------CT---CC--C-AAAT-----A...CCA--A--G--A------------...-----T--C--- 3970
DEB.CM.99.CM40        GG--A---C-AA-T-----A--------------C-----G-----A---------C--G--------------------AAA---...--A---......-GGA-C--C--C--G--AA-GG----G...G-A--------G--C-----A---...--C--TG-G--- 4089
DEB.CM.99.CM5         --------C----A----TAAT------------C--------T-A-----G-----T--G------C-----G--A--AGA----...--A---......-G-A-C----C--G--AA--G-----A...G---T----G--T-----------...------TG---- 4083
TAL.CM.01.8023        GC-----GCT--G--C--TA----C------C--T-----------A--------------------C--GC-G--G---CAA---...------......-G-A-T--C---G----ACA---CCCA...G-A--A--G--A--G-----A-GG...--C--T-AC--- 3747
TAL.CM.00.266         GC-----GGCA-----CT----T---C-G--G--C--------C--A--------------G--G-----G-----A--CCAA---...------......--GA-C--T---G----ACAG--CCCA...G-A--A-----A----------G-...--C--T-A---- 4249
SUN.GA.98.L14         -------CC----A--C--A-GT--------G--------------A--------C--G--A-----C--C----T--CCAG----......--AAAGAGAC-G---A-G---A----AAAG-----A...CCA--TGT---A-------G---G...-TC--T-AC--- 4643
WRC.CI.97.97CI14      GG--A---G-AA-A--CA-A-AT-----------T--T-----G--A--G--C--AC-G--A------G-------G------A--G...-T--ATCA....---A-TA-G---C----AAAT---ACA...CCA--AAT---AAGT---ACT---...--C--TCAAGG- 4580
WRC.CI.98.98CI04      -G--A---G-AA-T---A-T-AG-----G--------T-----G--A-----T--AC-G--A-T-----------A------A---...-T--ATCA....-------A-G--C----AAAT---AC-...CCA--GAT---AAGC---ACT---...-----TCAGGGC 4595
WRC.GM.05.Pbt_05GM_X02 GG--A---G-AA-A---A-CA-------G--------T-----G--A--------AC-G--A-----GC------A------A--C...-T--ATCT....--G---A-G--C-----AAAT---ACC...CCA--AAT---GAGT---ACA---...-----TCAAGGG 3509
LST.CD.88.524         -G--A--TGCA--T--C--A-G------------T--T-----T--A--G-----CC-GG-GA-----C-------T-T-AAA---...--T---ACAAAAC-G--CA-G---C-G--GAAG-----A...CCA--TGT---A--G---GA--G-...--T-----A--- 3557
LST.CD.88.447         -G--A--CC-A--G-----A-G------------T--------T--A--G------C--G-------CC-CC--------AAA---...--T--ACAAAAAC-------G---C---GAAAG-----G...CCC--A--G--A--GT--AGC---...-T---T-A---- 3560
LST.KE.x.lho7         -G--A--TC-A--T-----A-G------G--T--T--T--------A-----T---C--G-G-----CC-CC-----A--AAA--T...--A--ACAGAAAC-G--CA-G---C----AAAG-----A...CCT--AGT---A------AGC---...-TC--T-AC--- 4644
SYK.KE.x.SYK173       --------GCCA-A--G--A--------C-----C--T--------A--------------GCT--T------C------AAA---...--A--A......---A-C-----CC----AA----TCCC...C-A--A--G--A--T-----A-GG...--C---TCA--- 4483
SYK.KE.x.KE51         -G--A------A---CA-AA-C------T-----C--T--------A---------C-G--------C--------A---AAG---...--T---......----A---G---C-G---AAC--CCCA...--A--A--G--A------------...--------C--- 4138
COL.CM.x.CGU1         CT--A---GCCA-TC--G-A--C---TGT--------T-----G-A--------------TGA-----C--AAG--A......------G-A--A......C-G---A-G-GGA-----AA------A...GCA--AG---A--C---C-A--G-...-T---T------ 4046
```

```
H1B.FR.83.HXB2          TGGAGAGCAATGGCTAGTGATTTTAACCTGCCACCTGTAGTAGCAAAAGAAATAGTAGCCAGCTGTGATAAATGTCAGCTAAAA...GGAGAAGCCATGCATGGACAAGTAGACTGTAGTCCAGGAATATGGCAACTAGATTGTACACATTTAGAAGGAAAAGTTATCCT 4450
Pol                     _W__R__A__M__A__S__D__F__N__L__P__P__V__V__A__K__E__I__V__A__S__C__D__K__C__Q__L__K__._G__E__A__M__H__G__Q__V__D__C__S__P__G__I__W__Q__L__D__C__T__H__L__E__G__K__V__I__L
H1A1.UG.85.U455         --------------------------T-----------G-----G--G-----------------A---------A------...--G---------------------------------------------G---------T-------C-----C-------------A----- 3896
H1B.US.90.WEAU160       ----------------------------------------------------------------------------------...-----------------------------------------------------------------------G-----A------- 4449
H1C.ET.86.ETH2220       ----------------A---A-----TA-C-----C------C-C-------------TT----------------------...--G--------A----------A-T-----------G---------T-----------C-------G---A-C----- 3842
H1D.CD.84.84ZR085       ------------------------------------G-------------------T-------------------------...---------A------------------------------------------T---C-------------------------- 3972
H1F1.BE.93.VI850        --------G-----------------T--------A------------------T---------G-----------------...--G---------C-------------------------------G---------T-------C-------------C--G------- 3790
H1G.SE.93.SE6165        --------------C-----------TT-------A-----------------G----------------------------...--G----------------------------------------------------T-----------------------A-----A- 3847
H1H.CF.90.056           ---------G----------------T--A------A------------------T--------------------------...--G---------------------------C---------------------T-------C--------G-----C-------T-- 3797
H1J.SE.93.SE7887        --------------------------T--------------------G-------T-------------------------G...--G---------------------------------------------G---------T--------------C------------- 3764
H1K.CM.96.MP535         --------------------------T--------A-------C-----------T-------C------------------...--G-------A-----------------------------------------------T-----------------------A------ 3646
H101_AE.TH.90.CM240     ------A-------------------TT-------A---------G--------A--A---------------A--------...--G----T--------------------------------------G---------T-------C-----C-------------C----- 4024
H102_AG.NG.x.IBNG       ----AG--------------------T--------A--------------------G-------------------A-G---...--G--------------------------G------------------------T---------------------G---A-----T- 3975
H1N.CM.95.YBF30         ----A---------C------------T-A--C--CA---G---------------------C-----C-------------...--G---------------G--CA-T-----------G-G-------T-------------C-----G------A-C----- 4044
H1O.BE.87.ANT70         ----A----C-A--------A---GGA--A-----A--G--G--C--G-----CA-T--T-----CC--------TA-----...--G----A--T----T--G-------C-------A-G-------A-------C-----A-G-----C---A-C--AA- 4505
H1O.CM.91.MVP5180       ---------T-A--------C---GGAT-A-----AA-------C--G-----CA-T--T--T---CC------C--TA-----...--G----A-C------T----------AC--C----AG---------A-G----------------------C---A-C--AA- 4480
CPZ.CD.90.ANT           ------T--T-A--AGA---A-AC--T--T--C---A-T--G--T--------TA----ACAG--------------CG----G...---------G-------------------C------------------G-------C--C---------T-----A---A- 3889
CPZ.CM.01.SIVcpzCAM13   ----A---TT----GG----A------T-A-----CA------------G-----G--T-A----------------G------...------G--TC------G--G--T--T-----C-------C-------GA-G------C-------------C---A-C---A- 4030
CPZ.CM.05.SIVcpzEK505   ----AG--------C------------T--A--C--CA------T----------------T-------G----AT-G---...------------A-----G---A-CA--------------GG---------T-------C-------G--G-----A------- 4022
CPZ.CM.05.SIVcpzMB66    -----G--C-----C------------T--------CA--------------G--------A-AT-------G----AT------...------G----A-------G----------------------------T----C--C-----C--G-------GA-C----- 3996
CPZ.CM.98.CAM3          --------T-----A--------C--TA-A---A-A----G--C------C--------C-T-----------C--A--------...------G----------G----------C--C------C--------T----C--C-----CC------------------- 3864
CPZ.GA.88.GAB2          --------C------GA---A--CC-A--C-----G-------C-----------------CA-----------------T---...--------A--A----------G--T--C--------G--C-----G--G--C--C--C--C-------G---A-A----- 3843
CPZ.CM.98.CAM5          ----A---T-----C--------C--T--A-----------G--C--------------T-----C--G--C---T-----...--G--------------------------------------------C--T------T-----C-----------------A-C----- 4152
CPZ.CM.05.SIVcpzMT145   ----AG---------CAG--C------T-A-----AA---G--T---------------CAA----------------------...--------A-------------T--------C--------------GA-T------C--C----------------A-A--TA- 3992
CPZ.TZ.00.TAN1          ----A---TT-A---GA---A-A---T-----C--A--T--G--T--------TA-T--TCAG---CCA-----C--TA-----...------G--T--A---------G--G----AC-------A---C-----A-----C-----C--CC-----------G---C---A- 4094
CPZ.CM.05.SIVcpzLB7     --------T-----C--------C--T-------CA-T-------------G-------T--T-------------------G...------G--A-----------G--------------C---------------GT--------------CC-----------------C---T- 4005
CPZ.US.85.CPZUS         ----C---T-----C--------------------A--G--C--T--G--G-----G--------------------C------...------G----A----G--G----T--C------T--C---G--------------------------------G---A-C--TA- 4510
CPZ.GA.88.GAB1          ----A---T------------------TT-A-----CA-------------------G---CAT--------G--C--G------...-----------------G--G--------C-----G--T-----G--------C--CC-------C-----G---A- 4511
MON.CM.99.L1            ----A-TACT-AAGAGA-ACC-ACC--A-T--TA--C-GT-G--C-----------CAA-CAT---C-C-------AAACC-T...------C---AATCA-----G---A--GCAGAAGT----G-------GA-G--C--C--C---C----G----GCA-A--TT- 4153
MON.NG.x.NG1            ----A-TATC-CAGGGA-AC--A-GG-A-T---A-AC-CC----C-----G-----CAA-CAT--CCCC--G-----CAC-C--...------C---AA---------G-----Y-CAGA-AT---G---------A-G-----C--T----AT-------GTA-C----- 2746
GSN.CM.99.CN166         ----AGTACC-AAGAGA-C-A-AC--AA-C--CG-AC--T-G--T-----------GAA--AA---TCC--G------G--C-T...--G---C-A-AA-----G--G---A-TGCAGAATT---C--G-----GA-G--C--C--G-----G---------G--C---T- 4178
GSN.CM.99.CN71          ----AGTATT-AAGGGA-C-A-AC--AA-C--GG-AC--T-G--T------------AA--AG--CTCC----------A--C-T...--G---C-A-AA-----G-------A-TGCAGAATT---C---------A-G--C--C------C-G-----C---------T- 4160
MUS_1.CM.01.1085        ----AGTACC-CAGGGA-A-A-ACC-AA-C---G--T-GT---C-------------AA-CAT---CCA-----C---AC-C-T...------C-A-AAACA--G-----CA-TGCAGAGAT---GG---------A-G-----C--C--CC-G--G------A-C--T-- 4140
MUS_1.CM.01.CM1239      ----AGTATC-CAGGGA-A-G-AC---A-C---G--T--C-G--T--G---------AA-CTT---CCC----------AC-C-T...--G--GC-A-AAACA-----G--CA-TGCAGAATTG---G-G------A-G----------T--C-------------A-C---T- 4130
MUS_2.CM.01.CM1246      ----AGTATC-CAGGGA-A-A-AC---A-A--CA-CA-CT----C--G---------AA--AA--CTCAGC------AAC-C-T...--G---C-A-AA-----G-----CA-TGCAGACTTG--GG---------A-G-----C-------C-G-----C--G--C---T- 4238
MUS_2.CM.01.CM2500      ----A-TATT--AGGGACA-A-A---TA-A--CA-AC-TC-T--C--G---------AA---A--CTCAGC---C--AAC-C-T...------C-A-AA-----T-----CA-TGCAGACAT---GG---------A-G--C--C--C--------------A-C--T-- 4197
RCM.NG.x.NG411          --------------AGAA--A---C-AA-A--T-AAA------------------------CAA--CCCC--G------G------...----------AG-A--------G-----TGC----------T-CT------A-G-----C-------C-----------A-C---A- 3944
RCM.GA.x.GAB1           ---------T----AGAA------C-AA-T----AAA----G-------------------ACAG---CCA---------AG------...--G-----A--C-----G-----G--TGCA---------G-CT------A-G--C--C--------G-----------A-A---A- 3929
DEN.CD.x.CD1            ----A--AG--CAGAGA-A-G-A-GG---A---A--A-G-----T--G-----CA-TAATGT----CCA-----------AC-C-T...--G---C-A--CAGA-----G--G--TGCCTC--T-----C-------GA-G-----C--C--C-----GAAT-----A--AA- 4340

MAC.US.x.239            GTA-A--A-T---TATTCA-A---GGAT-A--CAGAA----G--C-G-C-G-------A--C---------------T-AG---...------G--T--A-----G--G-C-A-T-CAGA--T---G-CT------A-G--------C---C----G------A-A---A- 4749
Pol                     _N__V__K__E__L__V__F__K__F__G__L__P__R__I__V__A__R__Q__I__V__D__T__C__D__K__C__H__Q__._K__G__E__A__I__H__G__Q__A__N__S__D__L__G__T__W__Q__M__D__C__T__H__L__E__G__K__I__I
H2A.GW.x.ALI            GT--A--A-C-AT-CCA-A-A---GGAT-A--CAA-C-G--G----G-C-G-----AA--CA---CCC---------AG--G...------G--T--A-----G-----GA-TGCAGAA-T---C-CT------A-G--C--C-----C-------------C---A- 4825
H2A.DE.x.BEN            ATA-A--A-C-AA-CCA-A-A---GGAA-A--C-T-C----------G-C-G-----AA-TCA---CCC----C--A-AG---...------------A-----G------A-TGCAGAAAT---CG-T------A-G--C-AC-----C-------------A-C--TA- 4832
H2A.SN.x.ST             GTA-A--A-C-AT-CCA-A-A---GGA-----CAAAT----G----G-C--------AA--CA--CACCC---------AG---...--G--G--T--A-----G------A-TGCAGAATT---C-CT------A-G--C--C-----C-------------A-C--TA- 4277
H2B.GH.86.D205          GTA-A--A-C---TCCA-A-A--CGGAA-T----AAT----G------C-G-----AATTC------------C--A-A----...--G-----T--T--------G---A-TGCAGAC-T---G-C-------GA-G--C----------------------A----AA- 4805
H2B.CI.x.EHO            GTA-A--A-C-A-TCCA-A-A---GGGA-T----AAT----G----G-C-------AATTC------------C--A-A----...--G-----T--T--------G---A-T-CAGAA-T---G-C-------A-G--C---------------G-----G-----AA- 4803
H2G.CI.92.ABT96         GTA-A--AGT-AA--CA-A-G---GG-A-T--T-AAT-----------C-G-----AATTCA-ACC-C-GT--C--A-AG---...------------T--------G---A-TGCAGAATTG--T-CT------A-G--C--C--C------------------A---A- 4159
H2U.FR.96.12034         GTT-A--AGC-ACAGCA-A-A---GGTA-A----AAT-------C---C-------AAT-CT---C--GTT--C----AG---...----------------------G---A-TACAGAAGT---G-C-------A-G--C-----C--C-----G---------G--A- 4311
MAC.US.x.EMBL_3         GTA-A--A-T---TATTCA-A---GGAT-A--CAGAA----G--C-G-C-G-------A--C---------------T-AG---...---------T--A-----G--G---A-T-CAGA--T---G-CT------A-G--C-----C---C-----------A-AG--A- 4217
SMM.US.x.H9             GTA-A-AA-T---TATTCA-A---RGTT-A--TAGGC-----------C---------A--CA-----------C--C--G---...---------Y--A-----G------A-TGCRGRA-T---G-CT------A-G--C---------CC----G--C---A-A--TA- 4222
SMM.US.x.PGM53          GTA-A--A-T---TATTCA-A---GGTA-A--TAGGC-----------C-G-------A--CA---C-------C--C-AG---...-------------A-----G------A-TGCAGAA-T---G-CT------A-G--C---------CC----G--C---A-A---A- 4666
SMM.US.x.SIVsmH635F_L3  GTA-A--A-T---TATTCA-A---GGTA-A--TAGGC-----------C-G-------A--CA-------G---C--C-AG---...-------------A-----G------A-TGCAGAA-T---G-CT------A-G--C-----G--CC-------C---A-A--TA- 4752
SMM.SL.92.SL92B         GTA-AG-AGT-A---CAGA-A-A----A-C----AAT-------T---C-G-----AATGCT---A-C-----C--A-AG--G...---------G--A-----T--GAC-A-TGCAGAGGT---C-C-------GA-G--C--C--C--C-----------G------A- 4180
STM.US.x.STM            GT--A--AGC-A-TCTTCA-G---GG-A-A---AGGC----------GC-------G-AT-CA-----C--G--C--C-AG---...---------T--A-----G------A-TGCAGAGTT---G-CT-----GA-G--C----------C-------C--GA-A--TA- 4396
MNE.US.x.MNE027         GTA-A--A-T---TATTCA-A---GGAT-A--CAGAC----G--C---C-G-------A--CA---------------T-AG---...---------T--A-----G--G---A-T-CAGA--T---G-CT------A-G--------C---C----G------A-A---A- 4217
DRL.x.x.FAO             ------T-CT-A--ACAA--A---GGTT-A---GGGA----G---------------G--ACAG---CC---------AA--C-T...------GC-A--A-----C--------TGCA--C-------CT------A-G--C--C-----------------GA-A---A- 4017
MND_2.x.x.5440          --------GT--T--CAG--C--------A--C--CA----G--------G-----T--ACAA---CCA---------AC---G...------C-TG-T-----G--G---A-TGCAGA------G-C------GA-G--C--C--C-----------------A-C---A- 4011
MND_2.CM.98.CM16        -----G---T-A--ACAA--C--CTGTA-A---AACA----G-------------------CAG---CCA--G-----AAC----...------GC-A--A-----C--G-----TACATCC-------CC------A-G--C--C-----A-G-----------A---A- 4451
MND_2.GA.x.M14          ------T--T-AAG-CAG--A----G-A-A--C---A----G---------------T--ACAG--CCCA--G------G-----...------C-T--C--------G--G--TGCGGCC-------C-------A-G-----C--C---C----------G------A- 4378
OLC.CI.97.97CI12        GCACA-AGTT--CAGCAA--G--CC-AA-A----TAA----G------AGG------ATGAA-----AG-----A-AGGG---...--G-TTC-T-G-------AT---G-CAGTGG-GGTTTAT-GG-----GA----C--C--T-----------G--G--G---TG 4321
GRV.ET.x.GRI_677        ------AGT----TACAG--A---GGAT-A--TAA-A-------------G--------GGCA---CCC-----C--AA---G-...------C-T-A---------G------GCCTCCATT-A--CT-----GA-G--C--C--C---------------------AA- 4623
VER.KE.x.TYO1_patent    ----A-AACC-A--AGA-ACA-A-GGG--T----AAA-------------G-----G---TG---CCA---------A----G...------C-AG-------------G--TGCCTCA--T----C------GA-G--------T---C-------------AG--A- 4674
VER.KE.x.9063           ----A-AATT-A--GGACACA---GGAT-A----AGA----T--C-----G-------A-TG---CCA--TGCCA-AA-----...--G---C-A--A-----G---------GCCTCC-----GG--------A-G--C--C--G---G-----------A-AG--A- 4677
VER.KE.x.AGM155         ------AATC-A--AGACACA---GGAT------AAA-------T--------T-----A-TG--CCCA--------AG-----...--G---C-A--A---------------TGC-TCA------G-G-----GA-G--C--C------A-----------A-AG-GA- 4669
VER.DE.x.AGM3           -----GAACT-A--AGACAC--A-GGG--A----AAA-T--G-------------------TG---CCA---------A-----...--G---C-AG-C-----G--------TGCCTCG-----GG-------A-G--C------------------C--G--A---A- 4172
TAN.UG.x.TAN1           ---CA-TTT--AAGAGA--CA--CGGTA-A---G-AC-G-------------------TGCT--CCCA--------AA-TCG-...------C-G--A-----G--------TGCC---GT---GG-C-----A-G--C-----------G---------A-C--TA- 4649
SAB.SN.x.SAB1C          -----GAGT---CAACAA--A---GG-T-A--TG--A--------------------GGCA---CC------C---A------...--G--GT-TG-A-----G--------TGC----------G-G-----A-G----------CC-----------A-C---A- 4848
MND_1.GA.x.MNDGB1       GAAGC-CA-T-AAGAGAGA-A--CC--T-A---G--C------C---C-G--T--GCAA--T--CAG---G--CTGT-ATC-T...------GC---AA-G----GAC---TGC-TCA-TT---G-C-----GA-------C------C-G---AATC--A----TA- 4137
DEB.CM.99.CM40          ----A-CA-T-AAGGGA---A-ACC-A--A--TA-AC-CA----T--GC----TA--A-ACAA--CCCC----------AC--G-...--G--GC---CA-----T--G--G--TGCCTCAATG--C--C-----GA-G--C--C--C---C-G---AAC---A-A--AA- 4256
DEB.CM.99.CM5           ----A-CA-T--AGAGA---A---C-GT-A--TA-CT--A----C--GC-G--TA-T--GCAG---CCA--G--C---A----G...------GC-T--------T--------TGCCTCCATG--GG--------A-------C--C--CC-----AAC-----C---A- 4250
TAL.CM.01.8023          ----AG-ACC-AAAGGCCAGG-----A--A--CA-CA-------C----CC--CA---AGGCA--CCCC----------G-CC--...------C-A-A-ACA-----GACCA--GCAGCAGT---G-C------GA-G-----C--C--CC-G------C-G--A--ATG 3914
TAL.CM.00.266           -----G-ATC-TAAGGCAAGA-----GT-A--CA-CA-------C--G-CT--TA---AAGCA--CCCC---------AG--C--...------C-T-A-ACA-----GAACA-TGCAGC-GT---G-C-------A-G-----C--C-----G--G---C--A-C---TG 4416
SUN.GA.98.L14           GCAGCCT-TT--CAGGAGATG---G--A-A--C-TA--G--------GC-G-----CAATGAA----CAC-------A-AG---...--G--------AACA-----------TGCCTCAGTG--C--C-----GA-------C--C--CA-G----A------C--AA- 4810
WRC.CI.97.97CI14        ----AG--CT-ACAAGAA--A---C-T--T--T-AAT-GA--------AG----A---AGGAA---AG--CC--C--AGGG---GGA--G-C-C-A--------TATG-----T-A-GAGGTT--G-C-------A-G--C-----T-----------G---A----TT- 4750
WRC.CI.98.98CI04        ----A----T-ACAAGAA--A---C-TA-T----AAT--A-T------AG--------AAGAA---AG--CC-----AGG---GGA---CTC-AC-A-----TATG-------A-GAAGTG----C-------A-G-------T---------------A-C--T-- 4765
WRC.GM.05.Pbt_05GM_X02  ---G-G---T-A-AGGAA--A---C-T-----CA-CA-TC-T------AG--------AAGAA----C--CC------GG---GGG---C-CC-T--T-----CATG-------A-GAGAT---G-C------GA-G--C--C--C--C--G---------A-C--AT- 3679
LST.CD.88.524           GCA-A--ACT-A-AAGAAA-G-A-GGGA-A--C--CA-G-----T--GC--------AATGAG----C---G--------AG--G...--------------AACA--T---------TGTATCAGTG--T--T--------G--------T--CC-------GC--A-C---A- 3724
LST.CD.88.447           GCT-A--ATT-A-AAGAAA-A---C--A-A--C---A-G-----T--GC-------GAAT--T---A-------------AC---G...--------------AACA------------TAC-TCA-----T--T------A---C--C--T--CA-G--G---C-G--A---A- 3727
LST.KE.x.lho7           GCA-AG-ATC-A-AAGAAA-A---------A--C--CA-G-----C---C----T--CAATGA-----CA--C-----AAAG---...--------------AACA--------G--TGTCTCAGTG--T--T-----G-----C-----T--C---------C-G--C--TA- 4811
SYK.KE.x.SYK173         AT-GAGTATC-TAGGCAG--G---C-T--A--C-GACAG----------C---TA--CAACAA---CCA--------AAAT-G-...------C-A-AA-----G---------TGTAGA-AT-TAT-AT-----GA-G--C-----T---GAG-----C-----A---TG 4650
SYK.KE.x.KE51           GTAGA-TAC--AAGGCAA--A---C-T--A--CAGACAG-----C----CT--------AA-TG---CCC-----C--AA-C-G-...------C-A-AA-----G----------ACAGA--T---T-C-------A-G--------C--C---------------A--TTG 4305
COL.CM.x.CGU1           GT-CAGTATC-AAAAGAAC-A---GGGT-A--CA-------------------TTGG-AA--A---TCAG-G------AAC---...----C----TG-T-----C--GC----T-A---CTAT--CT--------T---------T---GAG--------G------T- 4213
```

```
H1B.FR.83.HXB2              GGTAGCAGTTCATGTAGCCAGTGGATATATAGAAGCAGAAGTTATTCCAGCAGAAACAGGGCAGGAAACAGCATATTTTCTTTTAAAATTAGCAGGAAGATGGCCAGTAAAAACAATA...CATACTGACAATGGCAGCAATTTCACCGGTGCTACGGTTAGGGCCGCCT 4617
Pol                          V  A  V  H  V  A  S  G  Y  I  E  A  E  V  I  P  A  E  T  G  Q  E  T  A  Y  F  L  L  K  L  A  G  R  W  P  V  K  T  I  .  H  T  D  N  G  S  N  F  T  G  A  T  V  R  A  A
H1A1.UG.85.U455             A--------C-------------C--C-----------------C-------------A-----G--------C---A-AC------------------------------GT----...--C--A-------------------A-C---G-A----AA--A-T-- 4063
H1B.US.90.WEAU160           ----------------------------------------------------------G---------A-----------C---A-C------C-------------------------------------...-----A--------------------A--A---------A--------- 4616
H1C.ET.86.ETH2220           ---------C-------------C--C-----G-----G-----------------------A--A------------C----AC-----------G---------C-GGGT----...-----A--T---------T--C------A--AA-G-A----AA--A---- 4009
H1D.CD.84.84ZR085           -----------------------C-----------------------------------------G--------C----C---------------------------------GT-G-G...-----A--------------------A-------A----A--------- 4139
H1F1.BE.93.VI850            ---------C--------T----AG--C-----------------------------A-----------C--C--CA-A-----G----------------------------T----...-----A--------------------A----CG------A---AT--- 3957
H1G.SE.93.SE6165            A----------------------C----------------------------------A---------------C---A-A------------------G----------C-GT----...-----A--T------------------A-----G-A--A-A----A--A- 4014
H1H.CF.90.056               ---------C-------------C-------------------C--C-------------AA--------------C--C--G--G---C-----A-C----------------GT----...-----A-------------------GA-----G------A---A---- 3964
H1J.SE.93.SE7887            -----------------------C-----------------------C-------------A------G-------T----A-A---------------C-----------------GT----...-----A----------------C-------A---G-G-T--G-A---A---- 3931
H1K.CM.96.MP535             ---------C--------T--C--C-----------------------C-----------G--A--A--------C--C----A-AC------------------------------GT----...-----A----------C-----------A-CA--GT-----A---A---- 3813
H101_AE.TH.90.CM240         ---------C--C--G------------------------------C-------------A-----G--------C-----GC-----C--------------------------GT----...--C--A-----C--T--------------A-C---G-A----AA--A---- 4191
H102_AG.NG.x.IBNG           A--------C-------------C-----------------------C-------------A-----G--------C---A-A-------------------------G--GT----...--C--A-----------------------A-----G-A--A-A---G--A- 4142
H1N.CM.95.YBF30             T--G--G--C-----G--------C--CT------------------T-----G-----A-----------------A------G-----T---------------------GTT---...--C-----T-----ATC----------TA----C--T--A-AA--A---- 4211
H1O.BE.87.ANT70             A--T--T--C-----G--A-----G-TC---------------G--A----------------A--A-----T--C--C--C--G------C-G--T-C----------T--T--GT----...-----A-----C--GCCT-----T--AA--A-A--TA-G-A---T--A- 4672
H1O.CM.91.MVP5180           A--T--T--C--------A----AC-T-----------G--G--A----------------A-----------T--C-----C--G--------------C----------T--C--GT----...-----A----------ACCT-----T--AA----AG-CA-G-AA--T--A- 4647
CPZ.CD.90.ANT               A--------G------T-T-----C-TC--------T------A--GG---AT--G-----A-AAGT---------C--C--G------------CA-C------------------------...--C-----------AGCT---------AA----AG-A--A-AA--G--A- 4056
CPZ.CM.01.SIVcpzCAM13       A-----------A------------C--------------------C--------------A--A-----T-----C--CA-A-------------A-C-------------G---GTCT--...--C--A--T-----G-CA-----T--AA--AA-G-A----AA--A--A- 4197
CPZ.CM.05.SIVcpzEK505       T--------C-----G--------C--CT-----------------C-----T-----------A--------T---------A-CC----G----T---------------------GTC---...---------T-----G-----C-------A----C--T--A-AA--A---- 4189
CPZ.CM.05.SIVcpzMB66        A--G-----A-------------G-----G-----------A--A--T--T----------A--A------------C---A-C--------G----C--------------------GTC---...---------T-----A--T--C--T---A--AG---T--C-A----A--T- 4163
CPZ.CM.98.CAM3              ---G-----A-------------G--CC----G-----G--C-----T----G--T--A-----------T------A--C-------------------------T-----GT---C...--C-----T-----GCCT--C--T---A--AA---A-----A---A---- 4031
CPZ.GA.88.GAB2              ---G-----C-----G--A--------CT-----------------A--------T--G--T--A---------T--------A-C-----------------G-------------G-AG---...--C--A--T-----A-CA-----------A----AG-A--A-AA--A---- 4010
CPZ.CM.98.CAM5              A--G-----A-----G-----------CC----G-----G--C-----T-----------A--A--------T------A-----------------------------T-----GT-----...---------T-----GCCT--C--T--TA--AAC--A--C-AA--A---- 4319
CPZ.CM.05.SIVcpzMT145       A--G-----C-----G-----------C--------------------C--C--T--T--------A-------------------G-CC-T--G-------------------G--T---...--C--A--------GCCT---------TA--AA-G-A--A-A----A--A- 4159
CPZ.TZ.00.TAN1              A--------G--------T-----T-TC-----------------C--A----A--------AAGA-----C--T--C--CA-CC-------G---------------T-----G-A-----...-----A--T-----ACCA-----T--TA--A-AG-A--G-A----A---- 4261
CPZ.CM.05.SIVcpzLB7         A--------G--------T-----C-----G-------------C--A--T------------A-----------T-------A-C-----------TCG-------------------TT----...---------T-----T---------T--TA-CA--G-A--C-A----T--A- 4172
CPZ.US.85.CPZUS             A--G-----A-----G-----------CC---------G--G--C-----T-----------A--A--G-----C-----A-C----------------------T-----GTG--T...---------T-----ATCT--C--T--AA--AG----A-------A--A--T- 4677
CPZ.GA.88.GAB1              A--G--------C-----------C--------------------C-----T--G-----A--A---------T-----C--G-----------------T-------------------T--T...--C--A--T-----GCCA-----T--AA-------G-A--C-A----T---- 4678
MON.CM.99.L1                ---------A--------------T---G--TGG--CA-GA-CC-A---AGG--------AA-ATGT----GGATAGCC--G--GG-GC-G--C-CT-TG------G--C-CTCAG---...-----A--T-----GCCA----------A-CCAAGAAT-CGAA--A--TG 4320
MON.NG.x.NG1                A-----C-----------T-----C--CG-TTGG--CA--A--C----CAG----------A-A-------GCAT-GCC--G--GG--C-T----CG-C-------A--C-TCATG---...--C-----------ACCA----------A--CAGGAAT--GCA--Y--TG 2913
GSN.CM.99.CN166             A--G--G--C--C-----A-----G---G-CTGG--TAGGA-A--A--CCAG---------CAGAC----T----TGAAG--C---G----------C--C------T----CCCACC--...-----A--T------CC-------GT-TCCAAGGAAT-GGAA--A---- 4345
GSN.CM.99.CN71              A--G-----C--C-------------G---G-TTGG--TAGGA-A--A--CCA----------CAGAC----T----TAAAG--A---G----------C--C---------------CTCACC--...-----A--T------CC--------T-TC-AAGGAAC-GGA---A--T- 4327
MUS_1.CM.01.1085            A--------C-----G--T-----G--C--GTGG---A-GA-AT-G---AG----------AG-C-G-----T-TAGGC--AC-GG-G------TC-CT------T-CG-C-CA---C...-----A--T-----GGCT--C-----T-A----AGAAT-CGAA--A--A- 4307
MUS_1.CM.01.CM1239          A-----C--C--C--G-------------C-CGTGG---A--A-AT-A--CAGG--------AAG-C-------CCTAGGC--G---G----------CTCTT----T-----C-CA-----...--C--A--T-----AGC----C-----TTA----AAGAAT-CG-A--T--A- 4297
MUS_2.CM.01.CM1246          ---G-----A-------------A-T-TC---TGG--CAGGA-A---T-CCA------T--AAG-CT-------CCTAGAG--G-C---C--G----CC-C--------A-T-GTCAG---...--CC-A---------T-C---C--T--TA-CAAGGAAT-CCAA--T-TTG 4405
MUS_2.CM.01.CM2500          A--------A--------------T-T----TGG--TAG-A-----T-CCA-----AT--CAG-CT--------TAGAG--G------C-------C--C------CA---C-CAG--C...--C--A---------T-C---C-------A--AAGGAAT-CCAA--A-TTG 4364
RCM.NG.x.NG411              A--------C-----------------C-----------------A--A-------G-----AA-A--G-----TC----C--G---------------CC-------------G---CATT--...--C--A--------A-CA--C--T--AA----AG-A--GCAA--A-A-- 4111
RCM.GA.x.GAB1               A--------C-----G--A--------------------G--------A--A------------------AA-A--G------C----C--G--G---C-------C---------------GG-AGC--...--C--A--T-----AGCA---------AA----AG-A--ACA---G-T-- 4096
DEN.CD.x.CD1                A--------AA---C----------G--CC----GA--A--A--T-G-----G--G-----CA-A----------CTA---T-AC-GC-GA----T-CT----------A-C-CC-A-T--...--C--A---------ACCA-----T-T-A--CAAG-CA-GCA---A--A- 4507

MAC.US.x.239                A--T-----A--------T-------TC-----------G--A------CA---G----AAGAC-G------CTA-----G---------G-----C---------TA-T-C-CATC--...--C--A--T-----TGCT--C--TG-TTCGCAAGAA--A-A-ATG-TTG 4916
Pol                          I  V  A  V  H  V  A  S  G  F  I  E  A  E  V  I  P  Q  E  T  G  R  Q  T  A  L  F  L  L  K  L  A  G  R  W  P  I  T  H  .  L  H  T  D  N  G  A  N  F  A  S  Q  E  V  K  M
H2A.GW.x.ALI                AA-------G-----T----------TC------------G--C--C--GCAG---T----AAGAC---------CTC--C--A------C-G--TA-T----------A---C-CACT-G...--C--A--T-G---TGT---C------TCACAGGAA--A-A-ATG-TAG 4992
H2A.DE.x.BEN                A--------A-----T--A-------TC---------------C--C---CAG---T----AAG-C-G------CTC--C--A------C-G--CA-T--G------A---CGCACT-G...--C--A----------CC---C-----TTCACAGGAA--G-A-ATG-TGG 4999
H2A.SN.x.ST                 A--------A-----T--A-------T----------------C--C---CAG---T----AAG-C----G---CTC--C--AC------C-G--CA-T--G------A---C-CATT-G...--C--A---------TGC---C-----TTCACAGGAA--A-A-ATG-TGG 4444
H2B.GH.86.D205              A--G-----C---------------G-T-----------G--A--A--CCA---G-----AAGAC-G-----TCTC--C--AC----G--G--CA-C---------TA-C-C-CACC--...--C--A-----C--TGC---C------TCAC-A-GT--A-A-ATG-TAG 4972
H2B.CI.x.EHO                A--G---------------------TC---------------A--A--CCA----------AAGAC-G-----TCTC--C--G-----GC-G--CA-C---------TA-C-C-CACC-G...--C--A-----C--TGC---C-----TTCACAAGAT--G-AAATG--AG 4970
H2G.CI.92.ABT96             A-------------G--------C-T-G--------------A--C---CA----------AAG-C----------TA--C--G-----------------G---------TA-C-C-CACC--...-----A---------TGC---C------TCGCAAGA---G-A-ATG-TAG 4326
H2U.FR.96.12034             C--G-----A-----G--------C-T---------------G--C--GCAG-----T--TAG-C-------CCTG--C--AC----GC----------G-----TA-C-CTCATT-G...--C--A--T-----TGC---C-----TTCACAGGAA--A-A-ATG-TAG 4478
MAC.US.x.EMBL_3             A--T-----A--------T-------TC---------------A------CA---G----AAGAC-GCACTAT...--C--G---------G-----C---------TTATTT-CAC--CTACACACACAGT-----TGCT--C--TG-ATCGCAAGAA--A-A-ATG-TTA 4384
SMM.US.x.H9                 A--T-----A-----G-GT-------TC-----G---------A--C--GCRG---------AAGAC---------CTG-----GC-----------CA-C---------CA-C-C-CATC-G...---------T-----TGC----------ATCACAAGAA--G-AAATG-TTG 4389
SMM.US.x.PGM53              A--------A-----G--T-------TC-----G---------A----GCAG---------AAG-C---------CTG-----GC-----------TA-C---------CA-C-C-CATC-G...---------T-----TGC----------ATCACAAGAA--G-AAATG-TTG 4833
SMM.US.x.SIVsmH635F_L3      A--T-----A-----G--T-------T------G---------A--C--GCAG---------AAGAC---------CTG-----G-----GC----T--C---------CA-C-C-CATC-G...---------T-----TGC----------ATCACAAGAA--G-AAATG-TTG 4919
SMM.SL.92.SL92B             A--------A-----G--T-----C-T-------------G--A--A--TAGG--------AAGAC-G-------TG--C--C------A-----TC----------TA---GCATC-G...-----A-----C--TGC-----------TCCCAGGA---G-A-ATG-TAG 4347
STM.US.x.STM                A--------A--------T-----G-T------G---------A--C---CA-----------AGAC-G-------TG-----G--G--G------A-C---------T--C-C-CACC-G...--C--G--T------GCT--C-------TCGCAGGA---A-A-ATG-TAG 4563
MNE.US.x.MNE027             A--T-----A--------T-------TC---------------A------CA---G----AAGAC-G------CTA-----G---------G--A-C---------TA-T-CGCATC--...--C--A--T-----TGC---C--T--TTCGCAAGAA--A-A-ATG-TTG 4384
DRL.x.x.FAO                 A--G-----C--C-----A------C-C---------------A--A--------------AG--A-------T---------GC-----------T--C--------------TC-CACT-G...--C--A--T-----ACCA----------A---AA-A---AGCAA-A-TA- 4184
MND_2.x.x.5440              A--------C-----------------T----------G--A--A--------G-----AA-A--------GC-C--C--G-----GC-------T--G-------------CATC--...--C-------------TCCT--C--TGT-A---AA-A---GGCCA-A-T-- 4178
MND_2.CM.98.CM16            A-C------C--------A---A-G---C-------------A--A--CA-----------AA-A---------C-C--C--G-----------------T----------------CATC--...--C--A--T------CCT----------A---AA-AA--AGCCA-A-T-- 4618
MND_2.GA.x.M14              A--------C-----G------------C-------------G--A--------------AA-A---------C----C--GC-----C---------------------G---CATT--...-----------------TCCA--C--TGT-A---AG-A---AGCCA-A-TG- 4545
OLC.CI.97.97CI12            TT-G----CA---AC--GGTCA--G-TCC--T-T-G----A-CC--AG-------GAT--AG-AA-T----TGA-A---G-GC-T-G-C----TA-TCA-------T-------T---T...--C-----------G-CA-----TGTGAA-CAG-A-A-GCA---A-TA- 4488
GRV.ET.x.GRI_677            A--------A-----------------TC-----------G--G--C---AG------T---A----G------C-C-----GC-G---C-GTT--C------------G---CATC--...--C-----T------CCA--C--T---TC-CAG-AT--GGCA--G-TG- 4790
VER.KE.x.TYO1_patent        A--T--G--C----------------TC---------------C--A--TAGG--------AA-A-----G---A-G-----A-----A--CTGA-T---------TA---C-CAGT--...--C--A--------GCCT--C--T---TCCCAAGAA--GGCA--AATA- 4841
VER.KE.x.9063               A-------------G-----C----TC---------------C--A--TAGG--------AAGA--G-----CA-A--C---C-----A--TTGA-T----------A--GTCCAGT--...--C--A--T-----ACCA-----T--TTCCCAAGA---AGC---AAT-- 4844
VER.KE.x.AGM155             A-----G--C---------------G-T-----------G------C--TAGG--------AA-A--G------A-G--CT-G------A--AT-------------CA-C-CTCACC-C...-----A--T-----ACCA----------TTC-CAGGAA--AGCT--TATG- 4836
VER.DE.x.AGM3               A--------C---------------TC--------------------A--TAG-----------A-A----------A-A---T-A-----GA--CT-A-T---------CA---CCCA-C-G...-----A---------ACC------T--GTC-CAAGAA--AGCA--AATG- 4339
TAN.UG.x.TAN1               A--------C-----G--------C-TCT---------------C--AG-TAGG--------AA-A-----------A-A--C--G------A-CATTA-T--G-----TA-T-C--A-T--...-----------------GCCA-----TGTTA--CAGGA---ACAAA-TATT- 4816
SAB.SN.x.SAB1C              A--G--T--------T--T-------T----------------A--C---CAG--------AA-A-CT------C-C--C--GC-G--G-----CA-C---------CA-C-C-CAGC-G...-----A--T-----T-C---C--T--TA-CCAGCAA--AGCA--TATT- 5015
MND_1.GA.x.MNDGB1           A--------G--------TTCA--C-TC--GA-G------------AA-----------T--AA-AA-G------G-G-----G-----G-------C-CA------TA-T-GT-A-C--...--C--A--T-----GCCT--C--T--TA--CAGGAA--AGAAA--ATG- 4304
DEB.CM.99.CM40              A--G-----C-----G--T---------------GA-CA-GA-CC-A--------------AA-A--G-----TCTG--C--G---C-G---G--C--------------CTCAG---...--C--A---------ACC---C--T-T-A--CAAC-A--C-A---A--T- 4423
DEB.CM.99.CM5               A--------G--------T-------T---G--GA-TA-G--CT-AT-----------T--AA-----------C-TA--CT-A---C-G--G-GG-C-----------CA-T---CAG---...--C--A---------ACCA-------TAA--CAGC-A--C-A---A--A- 4417
TAL.CM.01.8023              T--------A-----G--A-----T--C--T--GA--A-GA-CT-G--CAG-------C--A-G-------C--CCTG------C--GC-GG-G--CA-C----------A--G--CATC--...--C--A---------GCCG--C--TGT-TCG--AGA-A-GCA----A-AG 4081
TAL.CM.00.266               T--------A--C--G--------G--C-----GA-TA-GA--C-G---AGG--G--G----G---G--C--CCTC-----C---C--G-G--CA-C---------TA---GCCACC-G...-----A---------GCCA--C--TGT-TCA--AGAAA-GCA----ATGG 4583
SUN.GA.98.L14               A-----G--G-----G--TTCA--G-----G-T---------AC-A--CAAT--GCA----A--AC------CAC--GG--G-----GC-CTGT-CC-TG------------CAG---...-----A--T-----GCCA-------T-A--AAAGAT--GGA---A-TA- 4977
WRC.CI.97.97CI14            AAAT-----C---------TCA-----CC--------A--A-C--AT--------GAT-CCT-TC-G------CTA---ACC--GC-GC-G---A-T---------CA-C--G-A-----...-----------------G-CA-------AGTC-CAG-AAA-GCA---TAT-- 4917
WRC.CI.98.98CI04            AAAT-----C--------ATCA--G---C--------A--A-C------------GAT-CTT-TC-------TCTA--C--CC--C--A-C-----T---------TA-G----AG--T...-----G--T-----A-CA-----T-AGTCACAG-AAA-GCA-A-TAT-- 4932
WRC.GM.05.Pbt_05GM_X02      AAAT-----C-----G--ATCA--C--CT-------GA--A-A-----A----GAT-CCT-TC--------CT---CACCA--C----------------------------AG--T...---------T-----A-CA-------AATCACAG-AAA-GGA----ATA- 3846
LST.CD.88.524               -AAT-----C---A-G--ATCA--C-TC--G-TT--------G-------AT---------A-AACT------A-----T-G-----G---TGTA-C---------T------CA---T...--C--A--T-----TCCA-----TGTGA--AAAGAT--ACA---A-TGA 3891
LST.CD.88.447               TAAT-----------G---TC-----T---G-T---------A-------AT---------A-AAC-------C-C--C--G------C-GTGCA-C-------------C-GCA-G-C...-----A---------ACCA--C--TGT-A-CAAGGA---ACAA----TAA 3894
LST.KE.x.lho7               AAAT-----C-----G---TCA--C-TC--G-TT-----------A----AT--G-----AA-AAC---GT--C----C--G---------TGCA-T---------T------CA----...-----A---------TCCA-----TGTAA--AAGGA---ACA---A-TAA 4978
SYK.KE.x.SYK173             T--------AA-CAC---T----------------A-CA-GA-CT-AAA-AGG-----G---G-T-----------TG-----CA-GC--A----CA-T---------A-----CAG---...--C-----T-----ACCA--C--TGT-A---AC-A-T-C-AA--A--A- 4817
SYK.KE.x.KE51               T--------AA-CAC---------G-----C-----A-TA--A-AT-GAA-AG-------------A-------G-CTA--C-----GC--A------C-----------A-T---CAT---...-----A---------TCCA--C-----T-A---A-G-CT--GC---A---- 4472
COL.CM.x.CGU1               A--------A--C--CTGT-C-CTG-TCTGCTGG---ACCA-AT-AAAGAG----------G-A-------GTAGGGCC--CA-------G--TA-TCAG---GAG--T-G-CA-G-T...--C--A--T-----GCCA-----TGTAA--CAGCACT---A---A--AG 4380
```

```
H1B.FR.83.HXB2             GTTGGTGGGCGGGAATCAAGCAGGAATTTGGAATTCCCTACAATCCCCAAAGTCAAGGAGTAGTAGAATCTATGAATAAAGAATTAAAGAAAATTATAGGACAGGTAAGAGATCAGGCTGAACATCTTAAGACAGCAGTACAAATGGCAGTATTCATCCACAATTTTAAA 4787
Pol                        C  W  W  A  G  I  K  Q  E  F  G  I  P  Y  N  P  Q  S  Q  G  V  V  E  S  M  N  K  E  L  K  K  I  I  G  Q  V  R  D  Q  A  E  H  L  K  T  A  V  Q  M  A  V  F  I  H  N  F  K
H1A1.UG.85.U455            ----------AAAT---C-A----------G--C--------------------------G-----C-----C--G--------------C-----G----------G--A--------C-----------------------------T----------- 4233
H1B.US.90.WEAU160          -------------G----------------C-------------------------GA-------------------------------------------------------A----------------------------------------------- 4786
H1C.ET.86.ETH2220          ----------A--T--TC-A------------------------------------------------------------------------C-----G----------A--A----G--C-----------------------------T----------- 4179
H1D.CD.84.84ZR085          -C--------A--T----------------------------------------------------------------------------------------------------A----------------------------------------T----------G 4309
H1F1.BE.93.VI850           ----------A--T---C-A-----------C----------------------------------A-------G----A--G--C-------------------A----------------------------------------------------- 4127
H1G.SE.93.SE6165           ----------AAAT----CA----------------------C----------------------G-------------C--C--G-----C--G-----A-----------------------G-------------T----------- 4184
H1H.CF.90.056              ----------A-AT---C-A----------G------------------G-------------------------------G--C-----G----------C--A--A-----C-----------------------------T----------- 4134
H1J.SE.93.SE7887           ----------A-AT-----------------------------------------------------------------------C------------------A--A--------C-----------------------------A----------- 4101
H1K.CM.96.MP535            ----------A--TG---------------G------------------C---------------------------------------------C-------------------G------------C--A------------------------------- 3983
H101_AE.TH.90.CM240        ----------CAATG--C-A----------G--C-------------------------------G--------------C-----G----------G--A--------C----A-------------------------T----------- 4361
H102_AG.NG.x.IBNG          ----------AAATG---CA--A---------------------C---------G--------------------------------------C--------A--------C-----------------G-------------T----------- 4312
H1N.CM.95.YBF30            ----------AAAT-----A----------G--A----------T--------G---C-----G--C------------------------AA-C--------A--A---------A-----------G--------G--T-----T----------- 4381
H1O.BE.87.ANT70            ----------CAAC--AC-A--T--G-------A--A---A--T----A-------------------G-C--------G-------ATC------CAG-----G--G--C--A--A-----CT-A-GA----------------------TG-T----------- 4842
H1O.CM.91.MVP5180          --------A-A--C--AC-A--T--G----G--A--A--T----A-------------------G-C-----------------ATCT------CAG-----G--G--C--A--A--G---T-A--A--------------------C--TG-T----------- 4817
CPZ.CD.90.ANT              ----------TAAT---C-A-------------A--A------A---------------G-----C--------C----G--C-------------AA-T-----C--A--A-------AT-A-----------GT---------GCA---T----------- 4226
CPZ.CM.01.SIVcpzCAM13      ----------CAAC-----A--AA-G------------C--A--------------G-------C-----C-----------------C--T-A---A--T----------A---------A--A---------------------G----GT----------G 4367
CPZ.CM.05.SIVcpzEK505      ----------AAAT---C-A-----G--------A-----------T-------G--T-C--------C----------------------AA-T---------A--A------T-A--A---------------------C-----T----------G 4359
CPZ.CM.05.SIVcpzMB66       -------------T---C-A--A--------------------T---------G------------------------------A-----T---A---------A--A------------A--------C-----------------T----------- 4333
CPZ.CM.98.CAM3             -C--------A--C---C-A-------------------T----A------------G-------A-------G-------A--G--C-------AA-T-----A--A--A-----CT-G-G--------C--G--------G-----T----------- 4201
CPZ.GA.88.GAB2             ----------A--C--A--A--A--------A--T--T----A-----------G---------C-----C----------A--G--------G--A--T-----C--A--A-----CT-A-----------------T--------T----------- 4180
CPZ.CM.98.CAM5             ----------T--C---C-A-----G--------C----------A---------------G-----C--------------------C---------A-T-----C--A--A--G--G--A-------T--C-----------C-----T----------- 4489
CPZ.CM.05.SIVcpzMT145      ----------A--T--T-GA--------------A-----C--A------------G----------------------A--G--C----AG--A--C-------------------T-A--A--------------------C-----T----------- 4329
CPZ.TZ.00.TAN1             -C--------ACA---TC-A--T--------G-----A--T----T--------------------------C--------C---------AG--A--C--G--C--A--A--G--A--G-G---------ATC--------G-AT-------------- 4431
CPZ.CM.05.SIVcpzLB7        -----------CAG---C-----------------------------G---------G-------A-----C------C-T-------A-----G--------------------A--G--C--A--A---------------------------T----------- 4342
CPZ.US.85.CPZUS            -C--------A--C--AC-A--A--G----------A----------A-------G---------C--------------------C-------AA-C--G-----A--A--G---T-A-------T-----G-----------------T----------- 4847
CPZ.GA.88.GAB1             ----------A-AC-------------------A-----T----A--G-----------G-------CT-A-------GC---------A----------T--G-----A--A------T-A--A--------------------G-----T----------- 4848
MON.CM.99.L1               CA--------AAAT--TC-T--CACCACA--GG-A--------C--A-----C-----G--T----GAA-------G-C--C---AG-G-CC--TAAGA-AA-C--G---G---T----T-C--CCCA------T--GC-CAA---C-G------TTA--------- 4490
MON.NG.x.NG1               CA--------AAAC--AG-A--CAGTAC----G-G--------C--A----------T--------AAC-------GC--A-C--A-CC-CAT----GA-AA-C--G---G-A-T-C-GTTC--CTCT------C-GGC-CAAA-GC-C-A---T-TA--------- 3083
GSN.CM.99.CN166            -C--------A-AC--AC-A--CACCACA--GG-G--A-----C--G----------G--------AAC-----C---C--C-C---G---CA--CACGA-AA-C-----AG---TCAC-T-CT-AG-A---------GC-CA----TGC-A---T--T--------- 4515
GSN.CM.99.CN71             ----------AAAC--AC-A--CAGTACA--GG-G--A-----C--G----------G--------AAC-----C---C--C-C---G---CA--CCA-A--A-T------G---TCAC-T-CT-AG-A---------GC-CAA---TGC-A---T--T--------- 4497
MUS_1.CM.01.1085           -C--------CA-C--AG-A--TACCACA--GG-A--A--T--C--T--G-----------G-----AAC-AA--C---C-G-----AG--GCC---CACA-AA-C--G--AG---TCACTT--T-GG-A---------GC-CAA-----C-----TTTA--------- 4477
MUS_1.CM.01.CM1239         ----------TA-C--AG-T--CACCACA--GG-A--A-----C--T--------G------------AAC-AA--C---C-G-----AG---C----CAGA-AA-C----AG-A-TCACCT--T-AG-A---------GC-CA------C------ATG--------- 4467
MUS_2.CM.01.CM1246         CA--------TAAC--AC----CACCACA--TG-C-----------A--G------------------AA-GCA--C---C-G------G---CC--TCACA-A-----G--AG-A-TAAC-T--T-AG-A---------GC-CAA-----C-----T-TA---AA---- 4575
MUS_2.CM.01.CM2500         CA--------A--C--AG-A--CAGTACA--TG----T--T--C--A--G--C-----G---------AA-GCT-----GC---------G-G-CC--TAACA-A--C-----AG-A-TCC-GT-C--AG-----------GC-CAA-----C-----TTTA---AA---- 4534
RCM.NG.x.NG411             ----------TCA---AG----CACC------G-A-----------T-----C--------G-------A------C-TC----G----T--A-----G--AA-T-----------A---A-GT-AG----------T-----------GC----T----------- 4281
RCM.GA.x.GAB1              -C--------TCAG--AG----C-CC------G-A--T--------T-----------------G---AGC----------C-------A-T---C----A---A--------A--A--A---A-AT-AG-A--------C-----------T--GG-T----------- 4266
DEN.CD.x.CD1               ----------AAAT-----A--CACCACA---G----A--T---G----A-------G--TA-T-----AA--A---C--GC--------G--G------C--TA----G--AG-T--A---AGAT-AG-A-----C--GC-------ACCC----T-TT--------- 4677

MAC.US.x.239               CA--------A--G--AG----CACC-----GG-A--A--------A--G----G--------G---G-A-----C-CC-CC-G--A--TCAA---ATAGAA-C--G--A--A--AA-TTCAG-AG-A--CAT----TT---------TCATTG-ATG--------- 5086
Pol                        V  A  W  W  A  G  I  E  H  T  F  G  V  P  Y  N  P  Q  S  Q  G  V  V  E  A  M  N  H  H  L  K  N  Q  I  D  R  I  R  E  Q  A  N  S  V  E  T  I  V  L  M  A  V  H  C  M  N  F
H2A.GW.x.ALI               CA-------TT--T--AG---ATCC------G-A--T-------A----C---------------G-A-----C-CC-CC---A--TCAG--A-CAGAA-T----G----AA-TACAG-GG-A---AT-----T---------TCATTG-ATG--------- 5162
H2A.DE.x.BEN               CA-------TA--T--AG-A--ATCC------G-A--T----C--A----C--G-----------G-A-----C-CC-CC-------TCAG--A-TAGAA-T-----A----AA-TACAA-AG-A---AT-----TG--------TCATTG-ATG--------- 5169
H2A.SN.x.ST                CA------ATA--T--AG-A--ATCC--C---G-A--T-------A----C----------G---G-A-----C-CC-CC---A--TCAG--A-CAGAA-T-----G----AA-CACAG-AG-A---AT-----T---------TCATTG-ATG--------- 4614
H2B.GH.86.D205             CC-------TA-----AG-A--AACT------G-A-----T--C--A------------------G---G-A-----CC-TC-CC-G--A--TCAA---ACAGAC-C-----C--A--A-T-TCAA-AG------TT----T-------ACTCA-TG-ATG--------- 5142
H2B.CI.x.EHO               CC------ATA--G--AG-A--AAC---C---G-G-----T-----AG------G-----------G-A-----CC-TC-TC-G--A--TCAG---ACAGAA-T-----------A-T-TCAA-AG------TT--GTT-------ACTCA-TG-ATG--------- 5140
H2G.CI.92.ABT96            CC-------TA--C---G----A-C-------G-A--A-----C-----G-------G-----------A-----C-TC-TC--------CAA---ATA-AA-T--------A--AA-TTCAA-AG-A--CAT---GTT----------CATTG-ATG--------- 4496
H2U.FR.96.12034            CA-------TA--T--AG----TACC------G-A-----T--C--G-----C--------G-----G-C------C-CC-TC-------CCAA---AGAGAA-C-----A-----AA-TACAA-GG----CAT---C-T--------CTCA-TG-ATG--------- 4648
MAC.US.x.EMBL_3            CA--------A--G--AG--GCAC-CC--T-GG-A--A--------A--G-----G--------G---G-A-----C-CC-CC-G--A--TCAA---ATAGAA-C--G--A--A--AA-TTCAG-AG-A--CAT----TT---------TCATTG-ATG--------- 4554
SMM.US.x.H9                C---------A--R--TG-A---ACC---R-GG-A--T--T-----A--G--C---R-G-----G---G-G-----CC-TC-TC----A-CCCAA---ATAGAA-T-----A-----AA-TTCAA-AG-A--TAT-----T-------A-TCATTG-AT---------- 4559
SMM.US.x.PGM53             CC--------A-----TG-----ACC--C--GG-A--T--T-----A--G--C-----------G---G-A-----CC-TC-TC----A-CCCAG---ATAGAA-T-----A-----AA-TTCAA-AG---TAT-----T---------TCATTG-ATG--------- 5003
SMM.US.x.SIVsmH635F_L3     CC--------A--G--TG-A---ACC------G-G--T--T-----A--G--C-----------G---G-A-----CC-TC-T-----A-CCCAG---ATAGAA-T-----A--A--AA-CTCAG-AG---TAT-----T---------TCATTG-ATG--------- 5089
SMM.SL.92.SL92B            CA------TTA---G-GG-A---TCC------G-A--G-----C--A--------G-----------G-C---G-CTT-C-TC-T-------AC----ATA-AA----G--A--A--A--GTCAG-AG---CCT-----T--------C-CATTG-ATG--------- 4517
STM.US.x.STM               CC---------------AG-A--AACC------G-A--T--------A--G--C-----------G---G-C------C-TC-T-------CCCAG---ACAGAA-T---------A--AA-CACAG-AG-A--T-T-----TC---------TCA-TG-ATG--------G 4733
MNE.US.x.MNE027            CA--------A--G--AG----CACC-----GG-A--A--------A--G----G--------G---G-A-----C-CC-TC----A--TCAA---ATAGAA-C--G--A--A--AA-TTCAA-GG-A--CAT----TT---------TCATTG-ATG--------- 4554
DRL.x.x.FAO                -C--------TAAG--AG----TACCACA---G-------T--C-----G---------G-------A-----C--TC-GC--------------------AA-T-----------A---A-A--AG----------T------------C-G--T----------- 4354
MND_2.x.x.5440             -C--------TCA---AG----CACCACA---G-A--T-----C--A--G-----------------G-A-AA---C-TC-TC-T-----G--C-----A---A--T--G------A------AGG--AG-A---------G-----------GC-A--T----------- 4348
MND_2.CM.98.CM16           ----------TCA---AG----CACTACA--------------------T--G-----------A---GG-A-AA--CC-TC-TC-C--AC----C-----G--A--T-----C--A--A--GA-A--AG-A--------T-------------A--T----------- 4788
MND_2.GA.x.M14             ----------TCAG--AG----CACCACA-----A-----T--C------------------T--G---G-A-AA---C-TC-TC-T---GT--------AG--A--T--G-----A--A--GAGA--GG----------------------C----T----------- 4715
OLC.CI.97.97CI12           -C------TTA--G--TG----T---ACG--C-----G-GG-----T------------A-CA-----G-A-AA---G--TG------G-G--AT---GAGAA-C-----A-----AAC----T-GG-A------T-GGCT-----TA-C-----T--T--------- 4658
GRV.ET.x.GRI_677           -C-------GTAAT--AG----CACCAC------A--T--T--C--A--GTCA--G--TAGT-----AGC-----C-G-C-GC-C---G---C--CTCT--AA--------G-TTG---GAGAT-GG----------G--------TACGCAT-------------- 4960
VER.KE.x.TYO1_patent       ---------GAAA---TG-A--TAC-ACA--T--A--A--T--C-------TC------TC-A-----AGC-----C---C-------AG-G--A--T--GA-AA--------G-TTGCC--T--ACAG-----------TG-----TTGCCAT--T----------- 5011
VER.KE.x.9063              ---------GAAAG--AG----CACCACA---G-G--A---------A--GTCA--G---TC-------AGC-----C---C--C-C--AG---A--T---A-AA--------G-TTG-C-GT--ACAG-A----------TT------TGCCA--------------- 5014
VER.KE.x.AGM155            -C-------GAAAGG-AG-A--CAC-ACG--GG-A--A--T-----A--GTCC--G---TCTA-----AG------C---C----G--AG-G--A--T---A-AA--------G-CTG-C--T--ACAG-A----------TT-----CTGCCA---T----------- 5006
VER.DE.x.AGM3              ---------GAAA---AG-A--CACCACA--TG-A-----T--C--T---TCA----CTCTA---GAG-----------C-G----AG----A--T--GA-AA--------G-CTG-C--T-CACAG-A----------TT------TGCCA--------------- 4509
TAN.UG.x.TAN1              ---------GACA-G-AG----CAC-ACA--G--C-----T--C--A---TCA-----AGT------AG-------G-C--------AG---C-----A-A-AA-T-----AG-CTG--C-TT-T-GG-A--G-----C-TC-----CTGCCA---T----------- 4986
SAB.SN.x.SAB1C             -C-------GAAA---AG----CAC------GG-A--------C-----------G-----------------GC-G-----AG----C--------AA-T------G-T--A---AGAT-GG-A---------CAT---------GCA---T----------- 5185
MND_1.GA.x.MNDGB1          --------TTA--G--AG-A--CAC----------A-----T--C--A-------G--G----G---AA--AA-----GT-TC---AG--T-G--T-AGA-AA-------AG-TTGCA--G-AT-A--A---------GCC-----CAC------T--T--------- 4474
DEB.CM.99.CM40             -C------CTA--T--TG----CACCACA--GG-A--A--T--C--A-----C--G----G------AG--A---C-GGATTC---AG-T-CC--CAA---AA-C-----G-T--A--GAGAT-GG-A---------GC------TACC-A---AATT--------- 4593
DEB.CM.99.CM5              --------TT------AG-T--TACTACA--GG-A--A--T--C--A-----------G---------AGC-AA---GG-TCC---AG---C---CAA---AA-T----AG-T--A---AGA--AG-A---------GC------CACC-----TATT--------- 4587
TAL.CM.01.8023             CG------CT-AA---AG----CACCACA--GG-A--A--T-----A---TCC-----GTC------GAA--A---C---C-G--G-----G-CA--TCA---AA-T--G---G-A-TGC-GT-C--GTCC-----G--GGC-CA---TACC-----T-TT--------- 4251
TAL.CM.00.266              TA------CTCAA---AG----CTCCGCA---G----A--T-----T---TCC--G----C-------AAC-AA------C-GC-C-----G-CA--CAC---AA-T--G---G-A-TAC--T-C--GTCC--------GGC-CA---GACT-----TTTG----A---- 4753
SUN.GA.98.L14              --------TTA--C--AC----TAC-ACA--G--A--T-----C--T--G------------------G---G-A-A---C--G-T------AC-------CA-CAGAA-T-----AG-T--AC-GG-AT-G--A--T--------TC-----TC-GCA---T--T--------- 5147
WRC.CI.97.97CI14           -C------TTA-----AG-A--A--------G--C-----T-----T----------------A-----AG--A---C-GGAGG-----AG-C------CACAGTA-T-----AG-T--A---A-AT-AG-A------T--GCG------T-------T----------- 5087
WRC.CI.98.98CI04           -C------TTA-----AG-A--A--------G--C--T--T-----T----------------A-----AGC-AA---G-AGGC---AG---A---CACAGAA-TC----AG-C--A---A-GT-AG-A------T--GC-------C-------T----------- 5102
WRC.GM.05.Pbt_05GM_X02     -C------TTA--C--AG-A-----------G-------------T--G-----------A-T--GAG--AA---G--T------AC---A---CA-AGAA-------AG-T--A---A-AT-GG-A------T--GCT------A-C-A---T----------- 4016
LST.CD.88.524              CC------TTAA-T--AG----CAC-ACA--G--A--A--T--C--T-------------------GG-A-AA-------TT-----C----C----A-AGA--------AG-T--AC-G--G--A-----------GTT---------CCA---T--T--------- 4061
LST.CD.88.447              T-------CTA--T--TG-A--TAC-ACA-------T--T--C--T--G-----G-------T---G-C-AA-------C-C---AC----------A-AG---C-----AG-T---C----AT-A--------T--C-T-------AC-CA---T--T--------- 4064
LST.KE.x.lho7              CC------ATA--G--AG-A--TAC-AC---G--A--T--T--C--T--G--------G--T------G-A-AA--C----TT-----A-GT--------A-AGA--------AG-T--AC-G--A--A--A----------T---------CCA---T--T--------- 5148
SYK.KE.x.SYK173            --------TGT--C--AG-A--CACTACA--T-----A-----C--A-----C--G---A-------G--C-AA----GGT-T-----AG--GC---CTC---AA-T-----CG-C-TCAC----T-GC-A--------GGC-------ACC------TTA--------- 4987
SYK.KE.x.KE51              ----------A---G--G-A--CACCACA-----A--A--------A-----C--G---A--------AAC-AA----G-C-------AG---CC--CAAT--AA-T--G--AG-A-TA---AGAT-AG-------G--GGC------CACG-----TTTA--------- 4642
COL.CM.x.CGU1              TG------CTA-----AGCA--CACCACA---CAC-----T-----T---TCA--------T-----GCAG-GA--C--G--TG-T--A-G--AG--TAAGA-AA-G-A------A--A---ACAT-AG-A-GCAA----GC---------C-ATGCG-T---------- 4550
```

```
H1B.FR.83.HXB2          AGAAAAGGGGGGATTGGGGGGTACAGTGCAGGGGAAAGAATAGTAGACATAATAGCAACAGACATACAAACTAAAGAATTACAAAAACAAATT...ACAAAAATTCAAAATTTTCGGGTTTATTACAGGGACAGCAGAAATCCACTTTGGAAAGGACCAGCAAAGCTCCT 4954
Pol                     _R__K__G__G__I__G__G__Y__S__A__G__E__R__I__V__D__I__I__A__T__D__I__Q__T__K__E__L__Q__K__Q__I__._T__K__I__Q__N__F__R__V__Y__Y__R__D__S__R__N__P__L__W__K__G__P__A__K__L__L
H1A1.UG.85.U455         ------------------A--------------------A------------------------------------------...T---------------------------------------------G-C--CA-----------------A--A-- 4400
H1B.US.90.WEAU160       ---------------------------------------A------------------------G--C--C----------------...----------------------C-------------------G-------------------------T-- 4953
H1C.ET.86.ETH2220       ----G----------------------------------A---T--------T--------G--------C-C-----C-------...TT--------------------------------A--------G-C--TA---------------C--A--A-- 4346
H1D.CD.84.84ZR085       ---------------------------------------A-------------T---T---------G-----------------C...-----------------------------------------G----A-------------------T-- 4476
H1F1.BE.93.VI850        ---------------------------------------A-----------T----------------G-------------------...--G-------------C-----------------------C---G-------------------A-- 4294
H1G.SE.93.SE6165        ----------------------------A---------A-------------T---T-------------C------------------...---------------------------------------G-C---G---------------C--A--A-- 4351
H1H.CF.90.056           ---------------------------------------A------------------------------------------------...T---C-------A-------------------------G-C---A-------------------A---- 4301
H1J.SE.93.SE7887        ------------------A--------------------A-------------------------G----------------------...-------------------------------------------G-C---A-------------------A--A-C 4268
H1K.CM.96.MP535         ------------------------------------------T----------------------------------------------...TT---C-------A-------------------------G-A---A-------------------A-- 4150
H101_AE.TH.90.CM240     ---------------------------------------A----------------------------------------------...----------------------------------------G-C---A-----------------A--A-- 4528
H102_AG.NG.x.IBNG       ---------------------------------------A---T--------T---T-------------C--------G---...-T--------------------------------------------G-C--CA-----------------A--A-- 4479
H1N.CM.95.YBF30         -------------------------C--------------A-----------------------G--A-C-A-T------C-------...TT---G------------------A--------G----CA---------------C--A--T-- 4548
H1O.BE.87.ANT70         -------------------------C------A-----G--A--------T-----T--C-A--------A-C--------------...TT---N---C--A------C-------A-------G-C--TA-C-----------G---C------- 5009
H1O.CM.91.MVP5180       -------------------------C--------G---C--A-------C-----T--C-A--------A-C----C------------...TT------CA-C-----------C-------A--T-----G-C--TA------------G---C-A----- 4984
CPZ.CD.90.ANT           -------------------------CAC-T--AC-G------T-------C----------------G--A-CTC---------T-----...TT----------C-A-------C----------T------G-C--TG-G---------------C-A--T-- 4393
CPZ.CM.01.SIVcpzCAM13   -------------------------C---------------A-------------------------A-----C---------------...TT------------A------------------A--------G-C--GA------------G--C-CC--G-- 4534
CPZ.CM.05.SIVcpzEK505   ------------------A----C---------------A-------------------T-------A-C-A-GC-----C--------...TT---G--------------------------A--------G-G--CAC----------------C--A--A-- 4526
CPZ.CM.05.SIVcpzMB66    ----------------------------A--------A--------A-------C---------T--------C-A-------C-------...T---T---------A------------------------G-C--TA-----------------T--A--A-- 4500
CPZ.CM.98.CAM3          -------------------------C-----A--------A-------TA------C--------------------A-CTA-TC---------...TT---G------------------C----------------G----CA--------------T--C-GA--T-- 4368
CPZ.GA.88.GAB2          -------------------------C-----A--------A---------------GC-----------T------AC----C----------...CT-----------A-----------------A--------G-C--CA---------------CTCT----- 4347
CPZ.CM.98.CAM5          -----------------------T-C-----A----G--A-----G--TA------C--------------------A-CTA-T---------...TT---G----------------------------------G----A--------------T--C-GA--T-- 4656
CPZ.CM.05.SIVcpzMT145   ----------------------------A--G-------A-A--G------A-------C--------------------A-----C--------...TT-----------C-A--------C-------A--------A-G----A---------------G--A---- 4496
CPZ.TZ.00.TAN1          ------------------A-----C----------C--T--------C--A-T---A-T----G--A---C---------------...TT---G---------------------T------GC---G----A-------G-------GCGA--A-- 4598
CPZ.CM.05.SIVcpzLB7     ------------------A-----------A--------A-------A-----------------T---------C-------------...T----G-------A----------------C----G----TA---------------C--A--A-- 4509
CPZ.US.85.CPZUS         -----------------------T-C-----A--------A-------CA------C-----T---AC------AG-CTT--------------...TT---G-----------------C---------G----A----------------C--A--T-- 5014
CPZ.GA.88.GAB1          -------------------------C---------------A---------------------------A-GC---------------------...TT---G------A--------------A--------G-C---A-------------G---CCT-A-- 5015
MON.CM.99.L1            -----G--A--A--A---ATATGGCCC-T-TA--C---T-TA--A---G--CAC----AT--G--CTAC--ACC-C-A-C--C--------...-----T--TC--A-----------T----ACAG-TGCCG-C--T-AG--------------GCGC--A-- 4657
MON.NG.x.NG1            ----RG--A--A--A---ATAT----C--ATA--G--GT----CA---G--TCAT----TT--G--CTAC--C-C-C-A-C--C--------...-----T--TCG--A-----------T---AACAG--GCTG---TTCG-----G---------C-T--T-- 3250
GSN.CM.99.CN166         -----G--A--A--A---ATATGT-CC-TACA-----------A-T--G--TCAC----AT--G---AC--C-C--A-C-C-----A-...T-C---T------A----------------C-AG-AGCT--C--T------C-----------GTA----- 4682
GSN.CM.99.CN71          -----G--A--A--A---ATATGT-CC-TACA--G--GT----GA---G--CCAC----AT--G---AC--C-C--A-C-C-----G-...T-C---T------A----------------C-AG-AGCT-----T-A----C-----------GTA--T-- 4664
MUS_1.CM.01.1085        -AG--G--A--AG-G---ACCTATCAAAT-CT--G--G---A-CA---GC--CAC----AC--G--CTAC--C-C------C------AA...TAC---T--------------------T---ACAG-TTCTG---GTC------G-----C--CGTCT-A-- 4644
MUS_1.CM.01.CM1239      -AG--G--A--A--A---ACCTAGCA--T-CT--G------A-TA-T--GT--CAT----AC--G--CTAC--C-C------TC----AA...T-C---T--------------------T---ACAG--TCTG-C---TC------G-----T--TGCCT-GT- 4634
MUS_2.CM.01.CM1246      -----G--A-----A---ATATG-CCC-TACA-----GT----TA-T--GC--TAC----AC--G---TAC--C--C-------T--C-A-...-----T--TC--A----------------AAGAG-AGC---C--TTA----CTGA-------TCGA--T-- 4742
MUS_2.CM.01.CM2500      -----G--A--A--A---ATATG-C-C-CACA-----GT----CA---GC--TAT----A---G---AC--C--A------CAC--AA.......T--TC--A----------------AAGAG-AGCG------AG----C-------T--TC-A----- 4695
RCM.NG.x.NG411          ------------------------------------A----G---A----------------T-----T-------A-CCA-----------C-------...T---------------------------T---A--AG-A--GG---AG-AG-----G--------T--A---A- 4448
RCM.GA.x.GAB1           ------------------------------------A---------A----T---------------T--GC---C--TA-----------T-------...T--------------------------------A--AG-A--GG---A----G-----G---T-----T-----GA- 4433
DEN.CD.x.CD1            ----GG--A--AC-A---ATATG-CA----CA------TACA-TA----GC--AAT---C-TT--G---TAC--C-T...ACC------......C-----T--TCG------AA---C--------A--GG-AGCT-----G-AA---C-----------TG----G-- 4838

MAC.US.x.239            ----GG--A--A--A---ATATG-C-C---CA------T--A-TA----G--CA-T-----ACA-G-G-TAC--TTTCA----TC-A---AC...T-----T--A--------------C--------A--AG-----G---A---G-----G-----C-GTG----AT- 5253
Pol                     _K__R__R__G__G__I__G__D__M__T__P__A__E__R__L__I__N__M__I__T__T__E__Q__E__I__Q__F__Q__Q__S__K__.__N__S__K__F__K__N__F__R__V__Y__Y__R__E__G__R__D__Q__L__W__K__G__P__G__E__L
H2A.GW.x.ALI            ----GG--A--A--A---ATATG-CCC---CA------C-CA-CA-T--G--CT-C-----ACA-G---TAC--TTCC-C----C-A---A-...TTG---T--A-------C-C---C-----------AG-----G---AG--G-----------T-GGG----A-- 5329
H2A.DE.x.BEN            ----GG--A--A--A---ATATG-CCC---CA------C--A-CA---G--TA-C-----ACA-G---TAC--TTCC-C----G-A---A-...T----TT--A-------C-A---C--------A--AG-----G---AG--G-----------T-GTG-A--A-- 5336
H2A.SN.x.ST             --G-GG--A--A--A---ATATG-CCC---CA------C--A-CA-T--GG-CA-TG----ACAGG---TAC--TTCC-C---GC-A---A-...T-----T-A----------------C----T---A--AG-----G---AG--G-----------T-GGG-A--A-- 4781
H2B.GH.86.D205          ----GG--A--A--A---ATATG-CCC-T-CA------C----TA----G---A-C-----GCA-G---TAC-GTTC--C---GC-A---A-...TT----T-----------C-A---C--------A--AG-----G---A---C-----G-----T-GTG-A--AT- 5309
H2B.CI.x.EHO            ----GG--A--A--A---ATATG-CCC-T-CA-----------CA----G---A-T-----ACA-G---TAC--TTCC-C----C-A---A-...TT----T-C---------C-----C--------A--AG-----G---A---C-----G-----T-GTG-T--AT- 5307
H2G.CI.92.ABT96         ----GG--A--A--A---ATATG-CCC-T-CA---------A-TA----G--TA-T-----ACA-G---TAC--TTTCA-----C-A---A-...T-----T--A--------------C--------A--AG-----G-C-AG--C-----------T-GTG----TT- 4663
H2U.FR.96.12034         ----GG--A--A--A---ATATG-C-C-T-CA--------CA-TA----G---T-C-----ACA-G---TAC--TTCC-C---TC-A---A-...T-----T--A--------------C--------A--AG-----G-C-A-T-G--------C------G-CT-A-- 4815
MAC.US.x.EMBL_3         ----GG--A--A--A---ATATG-C-C---CA------T--A-TA----G--CA-T-----ACA-G---TAC--TTTCA----TC-A---AC...T-----T--A--------------C--------A--AG-----G---A---G-----G-----C-GTG----AT- 4721
SMM.US.x.H9             ----GG--A--A--A---ATATG-C-C---CA------T----CA-T--G--TA-C-----ACA-G---TAC--TTCCA----TC-A---A-...T-----T--A--------------C--------A--AG-----G-C-AG--G-----------C-GTG----AT- 4726
SMM.US.x.PGM53          ----GG--A--A--A---ATATG-C-C---CA------T----CA-T--G--CA-C-----ACA-G---TAC--TTCCA----TC-A---A-...T-----T--A--------------C--------A--AG-----G-C-AG--G-----------C-GTG----AT- 5170
SMM.US.x.SIVsmH635F_L3  ----GG--A--A--A---ATATG-C-C---CA------T----CA-T--G--CA-C-----ACA-G---TAC--TTCCA----TC-A---A-...T-----T--A--------------C--------A--AG-----G-C-AG--G----G------C-GTG----AT- 5256
SMM.SL.92.SL92B         ----GG--A--A--A---ATATG-C-C-T-CA-------------A-T--G--CA-C-----AC--G----CC--T-C---A-TTC-----A-...T-----T------------------------------A--AG-A---G-C-A---G-----G---------G------- 4684
STM.US.x.STM            -A--GG--A--AC-A---ATATG-CCC---CA------T----TA-T--G--CA-C-----ACA-G---TAC--TTTCA----TC-A---A-...T-----T-CA--------------C-----------AG-----G---A---A-----G-----C-GTG----AT- 4900
MNE.US.x.MNE027         ----GG--A--A--A---ATATG-C-C---CA-----GT--C-TA----G--CA-T-----ACA-G---TAC--TTCCA----TC-A---AC...T-----T--A--------------C--------A--AG-----G---A---G-----G-----C-GTG----AT- 4721
DRL.x.x.FAO             --------------A----A-----------CA-------------------T-----CT-T---C-CTT---A-GCA--C------C--A-T----...CTC-----------------------------------A--GG-A---G---AGT-G---------------G-A--TG- 4521
MND_2.x.x.5440          --------------A----A-------C-T--A--------------------A-C---------C-T---A-CCA------G-C--A-T----...T-------------------A--------------A--AG-A--GG-C-A--AG---------------G-----A- 4515
MND_2.CM.98.CM16        --------------A----AA------C----A------------------T-------------C-C-T---A-CTA--------C-CA-T----...CA----------------------------------A--AG-A--GG-C-A--AG---------------G-A---A- 4955
MND_2.GA.x.M14          --------------A----A-------C----A--------------G--T------A-------T--T-T-----CCA--------C--A-T----...T------------------------------------A--AG-A--GG-G-AG--C---------------G-A---A- 4882
OLC.CI.97.97CI12        -AC--G--A--ACAA-----AGGT-CA-TG-CA--GGA-T-----A--CA----TAC-----T-C-...CAA-T-----ACAC-C--A-T-AA...TTT----AATTTTCAGGATATCGAGTG--TTATA-A-A--A-G-AGGC-AG---TGT--C------G-A----- 4822
GRV.ET.x.GRI_677        -----G--A--A--A-----TAT-TC-AGT-CA------T-G--TA-T--GC--A-----C-AC--G--CTA--TACTC-------C-----C...CA-------TTG------AA---C--C-----A--AG-T---G-----G-G--------------GCGA---A- 5127
VER.KE.x.TYO1_patent    -----G--A--A--A-----AC-G-C-T---CA--G---C--A-TA-T-----A-----C-AT--G---TAC--C-T-------CCA-----...CA-------TT-------A-A--C--C-----A--AG-G---G-C--TG-G---------------C-AT-AA- 5178
VER.KE.x.9063           -----G--A--A--A--------TA-CAC---CA------C--A-TA-T--G---A-----C-AT--G--CTAC--AC--------CCA-----...CA-------TT-------A-A--C--C-----A--AG-G---G-----G-G-----G---------C-AT-AA- 5181
VER.KE.x.AGM155         -----G--A--A--A-------CTA-CA--T-CA--G---C--A--A-T--G---A-----C-AT--G---TC--CACTC------CCA----C...CA-------TTG------A-A--C--C-----A--AG-----G-----G-G-----G-----T--TCGC--GA- 5173
VER.DE.x.AGM3           -----G--A--A--A--------TA-CAC-G-CA--G---T--A-CA-T--G--TA-T---C-AT--G--TTAC--C-CC------CCA-----...CA-------TT-------A-A--C--C-----A--AG-G---G---TG-C--------------G-C--T-AA- 4676
TAN.UG.x.TAN1           -----G--A--A--A----TATG-CCC---CA--G--GC--A--GC--A---G---A-C---C-AC--G---TAC--C-CA------C----CAA...CA-------TC-------AA---C--C-----A--AG-A---G-G---G-G--------------C-CT----- 5153
SAB.SN.x.SAB1C          ------------------------TC-----CA-----C--A-CA-T-----CAT-----AC--G----C---AC-C-T---C--A-----...T--------------------------C-------------AG-----G-C--TG-G-------T--G--T------A- 5352
MND_1.GA.x.MNDGB1       CA--GG--A--AC-A------ATG-CA-----A--G----------A-T--G--CAAT-----AT--G--TA-C--T-TCA------T-------...T------A-TT------AA------C-T---A--AG-A---G---AG--G-----------T-GT-TC--TT- 4641
DEB.CM.99.CM40          ----GG--A--A--A---ATATGTC-C-GATA------C-----A---G--A-------A---G---A---ACT------C--A-------...T-----C--TT-GGC--CAA---C--------A--AG-TGCTG-----AC---------G-------CC--G-- 4760
DEB.CM.99.CM5           ----GG--A--A--A---ATATGGCA---ATA--G---C-----A---G--A-T------T-GG---AC--ACC------C--A-------...T-----C--TT-GG--ACAA---A--------A--AG-AGCTG-C---AC------G-----C--T-CCT-A-- 4754
TAL.CM.01.8023          ----G------AT-A---ATATGT--C---CA---GC-C-CA--A----G--CTAC----AC--------A-C-ACCC--------C--------...CAC--TT--TC-G-----AA---G--C--T--AA-GG-TGCT-----T------C-G--------TC-C---G- 4418
TAL.CM.00.266           ----G------AT-A---ATATGT-CC---CA---GC-----A-TA----G--CTAC----AC---------C-C--C--------C--------...CA---TT--TCGG-----AA---C--------AA-GG-TGCT-----T------C-G--------TC-C---A- 4920
SUN.GA.98.L14           CA--GG--A--AC-A-----AATG-CA----CA--G--GT-TA-CA----G--TAATG----TC--G----AC--T-C-----G---AT--A-...T--------TT---A---AA---C--C-----A--AG-A---G-----T-AG-----------T--T--A--T-- 5314
WRC.CI.97.97CI14        GA---G--A--AT-A------...TCCC-T-CA------T-----A-T--G---A-CT----TC--G----AC--C--ACT---C------AA...TT----T--A---------A---G--------AACAG-AGCC----AG-AG---C---------GT-------C 5251
WRC.CI.98.98CI04        GA---G--A--AT-G-----A...TCCC---CA------T-----A-T--G--CA--G----AC--G----C--TC--AC---C------AA...TT----T--A---------A-------------AACAG-AGCC----AG-AG---C----T----GT-------- 5266
WRC.GM.05.Pbt_05GM_X02  GA---G--A--AT-G-----A...TCCC---CA-------------A-T--G---CA-T----T---G------TC--AC---C------AA...TTC---T--A---------A------C------ACAG-AGCCTCG-AG-AG---C---------GT--AT-GG- 4180
LST.CD.88.524           CA--GG--A--AC-G-------TATCAC---CA--G---TACA-CA-T--G--CAATG----TT--G-----C--T-CA--------TT--A-...T--------TT---A---AA------C-----AC-AG-A---G-----AG----G------T--TC-A--T-- 4228
LST.CD.88.447           CA--GG--A--AC-A------CT-TCAC---CA----GT-TA-TA-T--G--TAATG-----T--G---AC--T-T----------T--A-...T--------TT---A---AA------C-----AC-AG-AC--G----C-AG----G-----T---C----T-- 4231
LST.KE.x.lho7           CA--GG--A--AC-A------CTT-CAC---CA--G--GT-TA-TA-T--G--TAATG----AC--G----AC--T-TC--------TT--A-...T--------TT---A---AA------------AC-AG-A---G----T-AG--------------GC-A--T-- 5315
SYK.KE.x.SYK173         -----G--A--A--A------ATATCAC----A-----GTACA-TA-T--GC--TAT-----AC-----TTAC--C--AATAC--C-TC-......C-----T--TCG------A-------C---C-CC-GG-G-AG...AACGAG-----G---------TCGT--A-- 5148
SYK.KE.x.KE51           -----G--A--A--A------ACT-CAC---CA--T--GTATA-TA----GC--TAC-----AC-----CTAC--A-TAC---C-C-.........CA----T--TCG------AAA--A-----T--AC-TG-T-CT...T-TGA----C-G---------TTCT--A-- 4800
COL.CM.x.CGU1           -----G--A--AT-A------A-G---C--T--------CA------A-G-GC-AT--T---AT--G-T--AC--A-CC--AC----TT-CAAAATCA----T--A---------AA--CC--C-GG-AA--GCA-....-CAGG-GAG---C---------G-G-A--GG- 4717
```

```
                                                                                                            Vif start                             Pol p31 integrase end
H1B.FR.83.HXB2           CTGGAAAGGTGAAGGGGCAGTAGTAATACAAGATAATAGT...GACATAAAAGTAGTGCCAAGAAGAAAAGCAAAGATCATTAGGGATTATGGAAAACAGATGGCAGGTGATGATTGTGTGGCAAGTAGACAGGATGAGGATTAG......................... 5096
Vif                                                                                                          M  E  N  R  W  Q  V  M  I  V  W  Q  V  D  R  M  R  I  R##
Pol                        W  K  G  E  G  A  V  V  I  Q  D  N  S  .  D  I  K  V  V  P  R  R  K  A  K  I  I  R  D  Y  G  K  Q  M  A  G  D  D  C  V  A  S  R  Q  D  E  D  *
H1A1.UG.85.U455          ------------------------------C------...--T----G----A-------------------------------------------------A-----G------------A------......................... 4542
H1B.US.90.WEAU160        -------------------------------------...------------------------------------------------------------------------------------......................... 5095
H1C.ET.86.ETH2220        -------------------------------------...--------G----A----G--G-------A--------------------------------C----------G------------A------......................... 4488
H1D.CD.84.84ZR085        ------------------------------C------...--------G----A---------------------------------------------------------------------------A......................... 4618
H1F1.BE.93.VI850         ----------------------C--------C------...--A----GA---A----G--G-------------------------------T--------------G-----------------A......................... 4436
H1G.SE.93.SE6165         ------------------------------C----AC...--A----G----A--------------------------------------------------G------------------------......................... 4493
H1H.CF.90.056            ------------------------------C------...--A---------A--------G-G----A-----------------------------------------------------------A......................... 4443
H1J.SE.93.SE7887         ---------------------------G--C------...--A----G----A-------------------------A----------------------G---------------A......................... 4410
H1K.CM.96.MP535          ------------------------------C--C---...--A----G----A----------------T-------------------------------G---------------A......................... 4292
H101_AE.TH.90.CM240      ------------------------------C------...--T----------A-----------------------------------------------G-----------------......................... 4670
H102_AG.NG.x.IBNG        ---------------------------G--C------...--T----G----A-------------A---------------------------------G------------------......................... 4621
H1N.CM.95.YBF30          G--------A----------G-----T--------CG-G...--T----------C---C-T--G--------A--A---------------------G-----------G------------G----------AA--C--GAAATGGAATAG............. 4702
H1O.BE.87.ANT70          G-----------G--A--------C-----------GG-A...-----T--G-----A--------G--G----A--A--C--A--G--------------------AC---A--A--------G-----ACA--AAG-G-AAGCGTGGAACAGCCTAGTGAAATAC 5176
H1O.CM.91.MVP5180        G-----------G-----------C-----------AG-A...-----T----G--A----------G----A--A--C--A--------------------AC---A--A------A-------ACA--AAG-G-AAGCATGGAACAGCCTGGTGAAATAC 5151
CPZ.CD.90.ANT            G---------------------C--CA----CC-AGAG...--A--T--G-----C--TC----G-----------T--A-A---G---A-----A---A-A--A-AGA---GA-T-----G--------A----------ATATCTGGAAATC............. 4547
CPZ.CM.01.SIVcpzCAM13    G-------------------------T--G---C-AG-A...--A-----------A--T----G--------A--A-----A--------------A-----------------------------------AAGCATGGAATAG............. 4688
CPZ.CM.05.SIVcpzEK505    G--------A----------G-----T--------CG-G...--T-----------C---C-T----G--C--A--A-----A---------------------------G------------G----------AG-C--GACATGGAATAG............. 4680
CPZ.CM.05.SIVcpzMB66     ------------G--A--------GC-C-----CC-AGAA...--A-----G-----C--TC----G-------A--------------------------------------------------A--C-AAACATGGAATAG............. 4654
CPZ.CM.98.CAM3           G-------------------G-----TA----G-GAGAG...--AG-T------A-A--C-----G-------A--A--------------------T---------A--A------G--G----------AG-C-AGGCCTGGAATAG............. 4522
CPZ.GA.88.GAB2           G-----------G--T--------T--T--G--C--CG-A...--T----------A--C-----------T--A--T----AAC----------G---G-----------------C----TG--G-----------A--C--AAC......................... 4492
CPZ.CM.98.CAM5           G-----------------------G--TA----G--CGAG...--AG-T--------A--C--------G--T-----A--C--A------------------------G-A--A-----G-------A--------C-AAACCTGGAATAG............. 4810
CPZ.CM.05.SIVcpzMT145    G-------------------G-----T------C-GGAC...--A--------G-----T-----G-----T-----A-------------------------CA---A-------------------------AG---AAACCTGGAATAG............. 4650
CPZ.TZ.00.TAN1           G-------------------------A----AGGAGAA...-----T---------A--C--G-----------A-----A-AA--G-----------------------CA-G-G--A---ATGA-------A-----AC----AACATGGACAAG............. 4752
CPZ.CM.05.SIVcpzLB7      G-----G--------A--------GC-G------C-GGAA...--A-----G-----C---C-T--------T--A--T----------------------A-----C-A-----C-----GA-------------A----AAGCATGGAACAG............. 4663
CPZ.US.85.CPZUS          G-----------------------CA-G--A--CGAG...--GG-T---------A--C--------------A--T--A-AA--C---------------------C----A--A-----G-------------AG---AGGCTTGGAATAG............. 5168
CPZ.GA.88.GAB1           G-----------------------G--C--G---C-AG-G...--AC----G-----A--------G-----------------A------------------------------------A----------AGACTTGGAATAG............. 5169
MON.CM.99.L1             ---------------T--C-----C---A--AC-G-GGAGGGC--T--CCTGAC---C--T--G-----G-------A--C-AACCC-----C-CCAC-GAAAAT-TG-G--G-GACACCAACCAATATA-TCTCAGAA-GC-AGATGGATTGGCTAATTAG....... 4820
MON.NG.x.NG1             ------G--G-----A--TT-G--G--TA-GACAG--CAAGGT--GG---T-AC---A---C-C--G--G-----------C-CCC------CTCCA--GAA-AT-TG-G-AG-AAGCCCAGTGCACAT---ATCAGA--GGT-GATGGAATGGCTGATTAG....... 3413
GSN.CM.99.CN166          ---------------A---------G-C---AC-C-AGCAGGT--G--T-TTAC---C--T-----G---------------C-A-CCA------G-GGCA-A-A---AA----TGG---GT----CCCATACAA--A-T--C-GAGAGGAAGGTAGACTGGTTGCTTAG. 4851
GSN.CM.99.CN71           ------G--------A--------GG-C---ACCC-AGCAGGT--G--C-TCAC---C--T-----G---------------C-A-CCA------G-GGCA-A-----AA----TGG--AGT---GCCCATACAAG-A-T--CAGAGAGGAAGGTAGAATGGCTGATTAG. 4833
MUS_1.CM.01.1085         G----G---------A--------C--CA--ACAG-CTCAGGA--GG---TCAC---T--T---------G-----A--A--C-AACCC-----CGCCA--GAA-AT-TG-G-AG-AAGTCC-AT-CAG--G-CCTCAGAA-GG-AGATAGATTGGATAATTAG....... 4807
MUS_1.CM.01.CM1239       G-----G--------A--------CG-CA--ACAG-ACAAGGAC-GG---TCACT--A--T---------G-----A--A--C-A-CCA-----CGCCA--GAAAAT-TG-G-AG-AAGTCCAAT-CAG--G-CC-CAGAA-GG-AGATGGATTGGATAATTAG....... 4797
MUS_2.CM.01.CM1246       T--------A-----A--TC----G--CA--ACA--GGAAGGT--G--CGT-AC---C-------------------A--A--C-A----------C-CCAG-CAAAAT-TG---AG-GAGCCC-ATTCAGT---T-TCAGAA-GG-AGATGGATTGGCTGATTAG....... 4905
MUS_2.CM.01.CM2500       ------G--------A---A--------CA--ACAG-CCAAGGT--G--T-T-AC---A-----G--G--G----------T--C-AA--C-----C-C-A--GAAAAT-TG---AG-GACCCCAATCCAGTTT-T--CAGAA-GG--GATGGACTGGCTGATTAG....... 4858
RCM.NG.x.NG411           ----G----A---------T--C--G--C--G---C-AGAG...--TC----G-----C--T--------GTGT--------A-AA---------C-G-A-AGA--TG-A-AGA--GACCAATATGGAAG-TAGAC-G-----AAGTTGA.................... 4593
RCM.GA.x.GAB1            ---------A-----A--------C--C--G--GG-G-C-GGA---T-G--G-----C--C--G-----------A-----A-AA--A-----C-G-A-AGAT-TG-A-AG--GGCCAATTTGGCGG-TAGAC-G-----AA-TTGA.................... 4581
DEN.CD.x.CD1             T--------------A-----T--C--CA-GACAC-GGCAGGA-----CTTCCCG--A--T---------G-----A-----C-A-CCA---AATGC-G-A...T-TA-GA-G-GAGAACA-AAC-AAG--AT----AGT--GACTGACATATAA............... 4989
                                                                                                            Vif start                         Pol, Gag-Pol, and p31 Integrase end
MAC.US.x.239             G--------G-----A-----CA-CT--A-G-TAGGG-CA...-----T--G-----A--C--------G--T--A--T--C-AA---------GG-GGA-AA-AG-TG---AGCA--TCCCAC-TGGAGG-TACC-GA--GGCTAGAGAGGTGGCATAG.......... 5410
Pol                        L  W  K  G  E  G  A  V  I  L  K  V  G  .  T  D  I  K  V  V  P  R  R  K  A  K  I  I  K  D  Y  G  G  G  K  E  V  D  S  S  S  H  M  E  D  T  G  E  A  R  E  V  A  *
Vif                                                                                                          M  E  E  E  K  R  W  I  A  V  P  T  W  R  I  P  E  R  L  E  R  W  H  S##
H2A.GW.x.ALI             G--------G--C--A-----CA--G-TA-G-TAGGG-CA...---------------A--------G--G--C--------C--A--C------GG-AG-CAA-AGCTG---AG-G--CCCCACTTGGAGGGT-CCAG---GG-TGGAGAAGTGGCATAG.......... 5486
H2A.DE.x.BEN             G-----G--A-----A-----CA--G-CA-G-TAGGG-CA...---------------A--------G--G--C-----T--C-----C------GG-AGACA--A-CTG---AG-A--CCCCACCTGGAGGGT-CCAG---GG-TGGAGAAATGGCATGC.......... 5493
H2A.SN.x.ST              G-----G--G--C--A-----CA--G-CA-G-TAGGGGC-...-----------A--A-A--------G-----T--------C-AA--C------GG-AG-CAA-AGATG---AGCG--TCCAACTTGGAGGGT-CCAG---GG-TGGAGAGGTGGCATAG.......... 4938
H2B.GH.86.D205           G--------G-----A-----CA-C---A-G-TAGGG-CA...--A--C--------A--C--G-----------A--T--A---C-C------GG-GGA-AA-G-TTG---TG-A---CC-AC-TGGAGG-TACCAG-C-GGCTAGAGAGATGGCACAG.......... 5466
H2B.CI.x.EHO             G--------G-----A-----CA-C---A-G-TAGGG-CA...--A--C------A-A--C----------------------A---A-C------GG-GGA-AA-A-TTG---TGCA---CC-ACGTGGAGG-TACCAT-C-GGCTAGAGAGGTGGCACAG.......... 5464
H2G.CI.92.ABT96          G--------G-----A--T--CA-C---A-G-TAGGG-CA...--A--C-----T--A--C--------------A-----A-AA---------GG-GGA-AA-AG-TG---AGCA--ACCAAC-TGGAGG-TACCAG-C-GACTGGAGAAGTGGCATAG.......... 4820
H2U.FR.96.12034          G-----G--A-----A--G--CC-T--TA-G-TAGGG-CA...-----T--G---A-A--C--G-----------------C--A---------GG-GGA-AA-A-TTG---AG-A--TCCAACTTGGAGGGT-CC---A-GGTTAGAGAGATGGCACTC.......... 4972
MAC.US.x.EMBL_3          G--------G-----A-----CA-CT--A-G-TAGGG-CA...-----T--G-----A--C--------G--T--A--T--C-AA---------GG-GGA-AA-AG-TG---AGCA--TCCCAC-TGGAGG-TACC-GA--GGCTAGAGAGGTGGCATAG.......... 4878
SMM.US.x.H9              G--------G-----A-----CA-CC--A-G-TAGGG-CA...--G--C--G-----A-----G--G-----T--A--T--C-AA--C------GG-GGA-AA-A-TTG---AG-G--TCCCACTTGGAGG-TACC-G---GGCTAGAGAAGTGGCATAG.......... 4883
SMM.US.x.PGM53           G--------G-----A-----CA-CT--A-G-TAGGG-CA...--G--C--G-----A--------G-----T--A--T--C-AA---------GG-GGA-AA-A-TTG---AG-G--CCCCACTTGGA-G-TACC-G---GGCTAGAGAAGTGGCATAG.......... 5327
SMM.US.x.SIVsmH635F_L3   G--------G-----A-----CA-CC--A-G-TAGGG-CA...--G--T--G-----A-----G--G-----T--A--T--C-AA--C------GG-GGA-AA-A-TTG---AG-G--TCCCACTTGGAGG-TACC-GA--GGCTAGAGAGGTGGCATAG.......... 5413
SMM.SL.92.SL92B          ---------A--------T--T-----TA-G-TAGGG-CA...--A--------G--A----------------A------C-AA---------GG-AG-CAA-AGATG-G-AGCA---CCAGT-TGGAGG-TC-GT--A-GA-TAGTGGAGCGCTGGCATAG....... 4844
STM.US.x.STM             G--------G-----A-----CA-CG--A-G-TAGGG-CA...-----T---------A-----G-----G--T--A-----C-AA---------GG-GGA-AA-AG-TG---AG-G--TCCCACCTGGAGG-TACCAG---GGCTGGAGAGGTGGCATAG.......... 5057
MNE.US.x.MNE027          G--------G-----A-----CA-CT--A-G-TAGGG-CA...-----T--G-----A--C--------G--T-----T--C-AA---------GG-GGA-AA-AG-TG---AGCA--TCCCAC-TGGAGG-TACC-GA--GGCTAGAGAGGTGGCATAG.......... 4878
DRL.x.x.FAO              ---------G-----A--G--G--C--CA-G--AGG--C-...---T-G------A-A--T-----------------------AA------------A-TG---ATA--C--ACCAACA---AG---T--........................................ 4648
MND_2.x.x.5440           T--------A-----C--T--G-----CA----AGGAGT-...--TT----G--G--T--T---------------A--A----A---C---------ACAG---ATA-CA--CCCCACA---G---ATAG........................................ 4642
MND_2.CM.98.CM16         T--------A-----A--G--G-----CA----GGGG-CA...---T-G--G-----A-----G-----------A-----C--A-----------ACAG---ATA-----CCCAAC----A-GC-T--T--.................................. 5085
MND_2.GA.x.M14           T--------A-----A-----G-----CA----AGGA-CA...---T-------T-----C-----------T--A--A--C---------------ACAG---ATA--A--CCCCACA---AGGA---------GACA-C--GA....................... 5024
OLC.CI.97.97CI12         G--------A-----A-----CC-GC-TA----AGGAGA-...A-GTATTTCTCT-----TCTG--G--------A--AG-G-A-CC-------GTGG-T-GCAGT-CCA---TA...............T--.................................. 4937
GRV.ET.x.GRI_677         ---------A-----C--G--G-----TA----GGGGGAA...-----C--G-----C--C--G-----G--T-----T--C-AA---------G-GAGA-AAA--ATG---AG-GAG-GTAGT-TGGAGGGT-TCAGA--GGCAAATAAGCAGATGGAGGGGGATAGTG 5294
VER.KE.x.TYO1_patent     ---------G-----A-----G--CC-CA-G...G-CG-AAGT---C----G--T--A--------G-----T--A--T----A--------A-CCCA-ACAAAG--TG-G-A--GAG-GT-ACGTGGA-GGTACCAG--GA-CTGATAACTAAATGGCAGGG...AATA 5342
VER.KE.x.9063            ---------G-----T-----G--C--CA--...G-AG-AGAA--GC-G---A-T--C------------------A-----C-AA-------A-CC-AGA-AAAG-ATG-G----GAG-GTAAC-TGGA-GGTCTCAG--GACCTGATAACCAAATGGCAAGG...AATA 5345
VER.KE.x.AGM155          ---------A-----C--G-----TC-CA-G...G-AG--GAA--AC-G--G-----T--G-----G--------A-----A-AA--C----AGCC-AGA-AAA--TTG-G----GAGACTCACCTGGA-GGT-CA-GA-GAAGTGATCACCAAATGGCAGGG...GATA 5337
VER.DE.x.AGM3            T--------G-----T-----G--C--CA--...GGAG--GTG--AT------A-TAC---------G--------A--T--A-A--------A-CC-AGA-AAAG-ATG-G----GAGAGTAACTTGGA-GGT-CC-GA-GAGCTGATAACTAAATGGCAAGG...GATA 4840
TAN.UG.x.TAN1            ------G--A-----A--G-----C---A----AGG-GAG...--GT-G--------A-----G-----G-----A--T--C-AA--A------G-GAGA-AAA-T-TG-G---CAAGACTAAC-TGGAGGGTT-G............................... 5286
SAB.SN.x.SAB1C           ---------A-----A--------T--------AC-GG--...--AT-G---AC-A-C---------------C--A--T--A-AA-----------------.......---TTG---AG-CAG-CCC-TCTGGA-GGTA-C-G-CGGACAGCAGGAGAGGTGGACTAG....... 5506
MND_1.GA.x.MNDGB1        G--------A--------------TT--A--T--C-AGAA...--G-----GA----A--T--------GTGT--A--A--A-AA---------G-G...-GT-G-AAGA--AG-CAG--TAACTTGGA-AGT-TCT--............................ 4774
DEB.CM.99.CM40           G--------A-----T-----G--GTGCA-GACAG-GGTAGGA--T--C-----G--C--T-----G--------AG-A--C-AA--A---------G-C----ATA--A-G-T-AACACACA--A-TCCA-T--AT-A.............................. 4899
DEB.CM.99.CM5            G--------A-----A--G--G--GTGTA--ACAG-GGTAGGA--------G-----A--------------C---G-G--A-AA-----------GGAG---ATA--A-GAC-GACACACA--A-CCCAGT--AT-A.............................. 4893
TAL.CM.01.8023           G--------A-----A-----C--CC--AGGACAG-CGAAGGA--GG---TTAC---T--C-----G--G--T--A--A--C-AACCC------C--GCT----GTAACA-GAC-GACT---A-G--TCCA----AC-----GCAGAGATGGGTAGGGACAATTAG.... 4584
TAL.CM.00.266            T--------A-----A--G--T--CT-G-G-ACAG-CGAAGGA--GG---TCAC---C--T-----G--G-----A--A----A-CCC------C--GCTT---GTAACA-GAT-GACC---A----TCCA----AC-----GCAGAGATGGGTAGGGACAATTAG.... 5086
SUN.GA.98.L14            G--------G-----A--------C--CA--C-G...G--GAGA----CCTG-----A--T--G-----------AC-AG-A-A---C------GG-G-ATCCT-CTCA-TG--GATG----G-T-A.......................AAGAGAATGGGAAGATAGAA 5458
WRC.CI.97.97CI14         T--------A-----A--T--------G--ACACCGGAAGGT...----T-AC---T--G---------------C-T--A-AA-T--GGAATGG-G-AGGAATG-A-AGGAG--CCTCT-AT-AG-ATT--............................ 5381
WRC.CI.98.98CI04         T--------A-----A--CT-----G-GG--ACCCCAGAAGGT...----T-AC---T--T-A---G--------A--A--A-A--TC-GGAATGG-G-AGGAATG-A-AGGAG--CCTCTAAT-AAGATT--............................ 5396
WRC.GM.05.Pbt_05GM_X02   ------------G--A--TT----TG-G--GAC-CC-GAAGGA--TT--CTTAC---T--C---------------A-----A-A-AC--GG-ATGG-A-AGGAATG-A-AGCAG-ACCACTAAT......T-A................................ 4307
LST.CD.88.524            G--------A-----T--T-----GG--A-G--GGGAGAG...A----TTTCTCTA-C--C--G--G--------AT-AG-G-AA---------G-GGGACCAAAG-A-AG-C-GA--AGCATGGA--AC--AT-G.............................. 4361
LST.CD.88.447            G--------A-----T--T-----GG--A----AGGAGAC...AC---CTTCTC---C--C-----------T---C-AG-A-AA---------G-GGGACCAAAG-A-AG---GG---GCCT-CA--ACA-CT-G.............................. 4364
LST.KE.x.lho7            G--------A-----T--T-----GG--A----AGGGGAG...A----CTTCTC---C--C---------------AC-AG-A-AA---------G-G...GGAC--AAG---AG-GAGAGTAGCCTGGATA-CA-CT--............................ 5448
SYK.KE.x.SYK173          T--------------T--------GG-G---ACAG-GGAGGGA-----TTTT-C---A-----------G-----A--A----CA--CC-----G--AGA----ATA---G-TCCCAC----AG-A-GAC-CCA-GACA----GATAG..................... 5296
SYK.KE.x.KE51            G-----G--A-----A--------G--C--GACACCAGA-CAAC-GG---T--C---A--T-----G--------A-----C--CAG-G-----G--AGAG---ATA---G-AC-CACC---A--T--C-TCCA-ATCAA---GATAA..................... 4948
COL.CM.x.CGU1            G--------------A--G-----C--CAG-A--TC-CAAGGTAC-T-GTTT--GAAA--T----------T--------CC--ACTCC-ATATGG-G-AGAT-TG--CAG---GAA-T--TT-TCC-ATGG-C-AA----AGCCGAAACCGTTAAGGGAATGGATTAA. 4886
```

```
H1B.FR.83.HXB2            ..............................................AACATGGAAAAGTTTAGTAAAACACCATATGTATGTTTCAGGGAAAGCTAGGGGATGGTTTTATAGACATCACTATGAAAGCCCTCATCCAAGAATA...AGTTCAGAAGTACACATCCCACTA 5217
Vif                       .__.__.__.__.__.__.__.__.__.__.__.__.__.__.__._T__W__K__S__L__V__K__H__H__M__Y__V__S__G__K__A__R__G__W__F__Y__R__H__H__Y__E__S__P__H__P__R__I__.__S__S__E__V__H__I__P__L_
Pol
H1A1.UG.85.U455           ..............................................---------C-------------T--C--------C--CAA-------CAA--T---------------------------TAGA---T-----G--...------------------------ 4663
H1B.US.90.WEAU160         ..............................................-----------------------------------A-----AA--------A--------C---------------C-------A--------------...--------------T--------- 5216
H1C.ET.86.ETH2220         ..............................................---------T----------G--------C--A----A--G----AT-----G------------T-----C---AGA-------AGG--...---------------------T-- 4609
H1D.CD.84.84ZR085         ..............................................C--------------------T--------C--A----AA-------AA-----------------C-------C--------C----A----...---------------------T-- 4739
H1F1.BE.93.VI850          ..............................................C--------------------T-T-------------AA-----C-AA-----C------------T-C-----AGG---------G--...------------------------ 4557
H1G.SE.93.SE6165          ..............................................------C-C---G-------T---------C--AA-------A--C-----------C-------C----AGG---------G--...------------------------ 4614
H1H.CF.90.056             ..............................................C----------C-------GT--------C--A----A--------A------------------T-T-------A----------G---...---------------------T-- 4564
H1J.SE.93.SE7887          ..............................................C------------------GT--------A-------AA--------AAA---C-G----G-----------C---AA--------A----...------------------------ 4531
H1K.CM.96.MP535           ..............................................T--T-----------------T--------C--A----AA-------ACC-----A---------------------AGG-------A----...------------------------ 4413
H101_AE.TH.90.CM240       ..............................................---------C-------------T---------A-C--AA-------AAAAG----------------T---------AG-------AGG--...--C-----------T--------- 4791
H102_AG.NG.x.IBNG         ..............................................---------T-------------T-----------C--AA-------A---T-------------------T------TAGG-------A-G--...T----------------------- 4742
H1N.CM.95.YBF30           ..........................................................C-----------T----------G--AAA--G--A-AA------A-----------T------CA-A---C----A----...--------------T------G-- 4811
H1O.BE.87.ANT70           CATAA.......................................................................---CAGG--TA-A--GA-CGA-AAC---GG--C--------T----TC-AGAA---------G-T...------AGT---T-T--T---G-- 5272
H1O.CM.91.MVP5180         CATAA.......................................................................A--CA-G--TAA--G--CGC-AAC---CG------G-----T----TC-AGGA-------A-G-C...-----G-CG--GT-T--T---G-- 5247
CPZ.CD.90.ANT             ..........................................................C-----G--------C--A-GG-AGA-TAAAGT-CT--AACC----AAG-----------T-------ATGA---C---AG-A--...G-AGA---------T--A------ 4656
CPZ.CM.01.SIVcpzCAM13     ..........................................................CC----T---T-T---------AAAA--AA------A-AA-AT-----C--C------C--------CAT---A---------G-T...TC------G--------T--TT-T 4797
CPZ.CM.05.SIVcpzEK505     ..........................................................---G-----G--T--C---------G---AA---G--AC-A--T-----C------------T-------CAGA-A------A----...---------A-----------T-- 4789
CPZ.CM.05.SIVcpzMB66      ..........................................................CC-----------T--C---------G--TA-------C--AAAT-----C---------C----------AA-A------A----...-----------------------G 4763
CPZ.CM.98.CAM3            ..........................................................CC--------TGT----------AAAA--AA------A-A--AC------------G--C--T-------TT--------T-A-G--...---------A----------TA-- 4631
CPZ.GA.88.GAB2            .................................................------T-----------GTTT--------CAAG---AAA--G--C--A--------------G--C--T-----T--T--CA----T-A-G--...-----------C-----A--CA-- 4610
CPZ.CM.98.CAM5            ..........................................................CC--------AC----------AGAA--AA---G--A-AAAAG-----------G-----T-----G--TGA-A----T-A-G-G...--------------T-----TT-- 4919
CPZ.CM.05.SIVcpzMT145     ..........................................................------G---T-T--------CAGA---AA------A-AAAAT-----C------------------TG--A-CA-C--T--G--G...GC----C-GA--T-T--T---A-- 4759
CPZ.TZ.00.TAN1            ..........................................................CC----T------T------C-T-ACAA-CAAATGCTG--AA-AT---AAG-----------T-------CTGA-ACA----A--G-...GCAGGG---A-------A--T--- 4861
CPZ.CM.05.SIVcpzLB7       ..........................................................---G------T-T-----C-----A--CA-------A--A--C------------------T-----G--TGAAA------A-G--...---------A----T--------- 4772
CPZ.US.85.CPZUS           ..........................................................-C-G------T-T--C------AGA---AA--------AA--------C--C--G-----T-----G---A--A----C--G---...-A------------T------TA- 5277
CPZ.GA.88.GAB1            ..........................................................------T---T-T-----T---AGG---AAA--G--A--A--T--------C---------T-----TCA----A------A-G-T...GC---T---A-------A--TT-C 5278
MON.CM.99.L1              ..........................................................A-GCACT---TGG--C--ACTCACAGG-A-T...--ACCCTTTGTC...---GT--------CC-GCTG-ACA---AG---T-CTCACAGAACA--A-TA-AC----CT-- 4926
MON.NG.x.NG1              ..........................................................AGC-ACC---T-TA-C--A---ACAGGGAAA...-TGCCATTTGT-...---GT----------C-GCTTA----C-AGC-CT-TTCGCAA-GCA-GA-TA-GC-T--TT-- 3519
GSN.CM.99.CN166           ..........................................................AGCTACT--------C--A-GGTCAGGCAAA...A-ACCATTTGTA...---GTC--------CC--TTG-AA--C-A----T-TACACAAAACA-GA-T-GAT-AG--A-G 4957
GSN.CM.99.CN71            ..........................................................AGC-ACT-----T-----A-GGTCAGG-AAA...A--CCATTTGT-...----TC--C------C--CTA-AA---A---GT-TACACAAAACA--A-T-GAT-AG--T-- 4939
MUS_1.CM.01.1085          ..........................................................GGC-AC----T--T-C--CCT-TCAGGGA-A...--AGCCTTTACA...---GTG---------C--TTGTAC----AG--GT-CTCCCAGAGCA--A-CAGA--A---T-G 4913
MUS_1.CM.01.CM1239        ..........................................................GGCTAC------T-----C-T-TCAGG-AAA...A-ACCATTTGT-...---GT---C-----CC-GGTA-AA--C-A----T-TTCCCAAAACA-GA-CA-GC-AG--T-G 4903
MUS_2.CM.01.CM1246        ..........................................................A-GCACT-----TT----C-TCTCAGGGAAA...---CCATTTACA...---GTC--C------C--TTG-AG----AG---T-CACCCAGAACA--A--A-A--A--TT-G 5011
MUS_2.CM.01.CM2500        ..........................................................A-GTACC--------C--C-TATCAGGGAAA...--ACCATTC-CA...--C-TC-----T--CC--GTG-AG----AG---T-CACTCAAAATC-GA-TA-G--A--CT-G 4964
RCM.NG.x.NG411            ........................................GCAGTGGCAT-CCTT-GTAAAGTATC-T--GT---A-GGAAAGAAG-AAGCTAAAGATT-GGA--A-GTACCC---TTTA-A-T-CCTTGGGGGTGGT-GTCT...C----------C--------CT-G 4720
RCM.GA.x.GAB1             ......................................ACA..ATGGCAC-CTTTGGTAAA-TATC-T--GT-C-AAGGAAAGAA--CAGC-AAAGA-T-GGA--A-GTGCCT--CTTTA-A-T-CCATGGGGATGGT-GTCG...CA---------T-----A--TT-- 4708
DEN.CD.x.CD1              ......................GATTGGGGAAACCAAATTAGAACATTGGAT--GC-C---GAA-TCTATAATATACA-AACAAAT-A-TT-AAAG--AC-GT-A--ATACC---TTTAGAATC-CATGGG-ATGGTAT-CC...-TGA--A--A------------G 5134

MAC.US.x.239              .......................................................CCTCA--AA-T-T-TGA-ATATA-AAC-AA--ATCT-CAA-A--TT--C-A-GTGCCC---TTTA-G-TCG-ATGGGCATGGT-G-CC...T-CAGCAG----AT-T-------- 5522
Pol
Vif                       .__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__._L__I__K__Y__L__K__Y__K__T__K__D__L__Q__K__V__C__Y__V__P__H__F__K__V__G__W__A__W__W__T__.__C__S__R__V__I__F__P__L_
H2A.GW.x.ALI              .......................................................CCT-G-CAAGT-C-TAA-GTACAGAACAAA--ATCT--AA-A--TG--C-A-GT-CCC--C--TA-G-TGG-ATGGGCATGGT-G-CT...T-CAGCAGG---ATAT-----T-- 5598
H2A.DE.x.BEN              .......................................................CCT-G-CAAGT-C-TGA-ATACAGAACAAA--ACCT--AGGA--TGC-C-A-GT-CCC--C---A-G-TGG-ATGGGCATGGT-G-CT...T-CAGCAGG---ATAT-------- 5605
H2A.SN.x.ST               .......................................................CCT-A-CAAGT-T-TAA-ATACAGAACAGG--ATCT--AG-A--TG--C-A-GT-CCC--C--TA-G-TGG-ATGGGCGTGGT-G-CT...T-CAGCAGG---ATAT-----T-- 5050
H2B.GH.86.D205            .......................................................TCTGA-TAAGT-T-TTA-GTATAGAACAGG--A-TTGCAACA--TC-CT-A-GTCCCT--C---A-G-T-G-ATGGGC-TGGT-G-CT...T-CAGTAG-A--ATAT-T--C--- 5578
H2B.CI.x.EHO              .......................................................TCTAA-TAA-T-C-TGA-ATATAGGACAAA--ACTTGCAACA--TC-CT-A-GT-CCT--C--TA-G-T-G-ATGGGC-TGGT-G-CT...T-CAGTAG---ATAT-T--C--G 5576
H2G.CI.92.ABT96           .......................................................TCTGA--AA-TTCTTAA-GT-TA-AACAAA--ACCT-CAG-A--CGGTT-A-GT-CCT------A-A-TTG-ATGGGC-TGGT-G-CT...T-CAGCAGG---ATAT-----T-- 4932
H2U.FR.96.12034           .......................................................CCTGA-CAA-C-C-TGA.GTATA-CACAAA--A-CT-CAG-AA-TG--T-A-GTGCCC------A-G-T-G-ATGGGC-TGGT-G-CT...T-CAGCAGG---ATAT-T--T--- 5083
MAC.US.x.EMBL_3           .......................................................CCTCA--AA-T-T-TGA-ATATA-AAC-AA--ATCT-CAA-A--TT--C-A-GTGCCC---TTTA-G-TCG-ATGGGCATGGT-G-CC...T-CAGCAG----AT-T----C--- 4990
SMM.US.x.H9               .......................................................CCTCA-CAA-C-C-TGA-ATATA-CAC-AA--ACCT-CAG-A--CT--T-A-GTACCC----TA-A-TTG-ATGGGCATGGT-G-CT...T-CAGCAG----AT-T----GT-- 4995
SMM.US.x.PGM53            .......................................................CCTCA-CAA-T-CTTGA-AT-TAGAAC-AAG-ACCT-CAG-A--CT--T-A-GTACCC----TA-A-TTG-ATGGGCATGGT-G-CT...T-CAGCAG----AT-T-------- 5439
SMM.US.x.SIVsmH635F_L3    .......................................................CCTCA-CAA-C-C-TGA-ATATA-CAC-AAG-ACCT-CAG-T--CT--C-A-GTGCCC----TA-A-TTG-ATGGGCATGGT-G-CT...T-CAGCAG----ATTT-----T-- 5525
SMM.SL.92.SL92B           .......................................................CT-CA-TAA-TTC--TA-GTATA-AACAAG--A-CT--AG-AA-CC--C-A-GTGCC-------A-A-T-G-ATGGGCATGGTAC-C-...GCC--TAGG--GATTT----CT-G 4956
STM.US.x.STM              .......................................................CCTCA--AA-C-TTTAA-GTATA-CAC-AA--A-CT-AGC-A--C---C-A-GTGCCC--C--TA-G-TTG-GTGGGCATGGT-G-CT...T-CAGTAG---GATTT-T--CT-G 5169
MNE.US.x.MNE027           .......................................................CCTCA--AA-T-T-TGA-ATATA-AAC-AA--ATCT-CAA-A--TT--C-A-GTGCCC-----TA-G-TCG-ATGGGCATGGT-G-CC...T-CAGCAG----AT-T-------- 4990
DRL.x.x.FAO               .........................GAGGTACATGGTAGAAAGGTGG-AC-CCCTGT-CAA-TATC-TA-G----A-GGA-AAAAGTATTT--AA--AT-GCAT-A-GCACCT--CTT-C-ATGCTCA...GG-TGGT-G-CT...CA-AGTC-GTGGAC------CT-C 4787
MND_2.x.x.5440            ......................ACAGGACGTCAAGTTGAGAGATTGG-AT-CA-T-GTAAA-TATC-TA-GT---AAGGA-AGAA-CATTT--AA-ATT-GCAA--A--CCCC---TTTC-ATGCTCA...GGATGGT-G-C-...CAC--TC--AAGA-G-----CT-T 4784
MND_2.CM.98.CM16          ............................ACATGCAGTAGAAAGATGG-AT-CACTGGTAAA-TACC-CA-AT---AAGGA-AAAA-CATTTG-AC-AAT-GGAT--AA-GGC---TTTC-ATGCTCA...GGATGGT-G-C-...CAC---C--AA-T-T--T--C--- 5221
MND_2.GA.x.M14            ........................................AAAATGG-AT-CATT-GTCAA-TACC-CA-AT---A-G-A-AAAA-CACTT--A--A-T-GGAAC-A-TCCAC--CTTTC-ATGCTCA...GGATGGT-G-CG...CAC--TC--AAGATT-----CT-C 5148
OLC.CI.97.97CI12          ............................GCTAAATAATGGGCAGGT-GA-CATGTG-AA--CA-GGCTA-GTTAGCC--CAAA-TTT--GG-TACCATT-GATA-GGC-A-ATTG-T--A-GT-CCCAAGGG-ATGGT-G-C-...CA-GG-A--A-TT-T--T---G-- 5076
GRV.ET.x.GRI_677          ACTTACAAGATCAGGAATAA....................................................................AC--TTGCCTT-GGAA-ACAGACAT---TGGC-G-TGCAATGG--GTTTT-G-CC...TACAGCC-GT-CATT-----CT-- 5393
VER.KE.x.TYO1_patent      GTCAGATATTGGATGAGACA.................................................................-A-A--TTTG-AAT-GAAC-A-----TG---T--C-AATT-CATGGGC-TGGTAC-CC...-TGAGTAG-TATGTA--A---A-- 5444
VER.KE.x.9063             GTCAAATATTGGATGACTAA.................................................................-A----CCT--AAT-GCAA-ACATG-TG--CT--C-AATT-ATTGGGC-TGGTAT-CT...-TG-GCAG-TACAT-------A-- 5447
VER.KE.x.AGM155           GTTAGATATTGGATGAATAA.................................................................-A----TCTG-AAT-GGAA-ACA-A-TG---T-TC-AATC-CTTGGGCATGGTAC-CT...-TGAGCAG-TATGTA--A--C--C 5439
VER.DE.x.AGM3             GTGAGGTACTGGATGAGGAC.................................................................TA-A---TTAGACT-GAAA-A-CGA-TG--CT--C-AATT-CATGGGCATGGTAC-C-...-TGAGTAG-TATG-G--A--C--- 4942
TAN.UG.x.TAN1             ..................................TGGAGAGCATAT-GACAAATGG-AGGGCA--GTGA-AT-CCACATGAGGAACC--TT-CAGGATT-GACA-A-CT--TG--CT--C-ATGTG-ATGGGC-TGGTAC-C-...T--AGTAG-T-C-T------CT-- 5419
SAB.SN.x.SAB1C            .......................................................CCTAG-GAA-T-C---ATGCATGTGTCCAA-CA-TGT-T-CATT-GC---A-ACCCCT---ACTA-AAT---ATGGA--TGGTACTCC...TA-CA----TGGGT---A--C--G 5618
MND_1.GA.x.MNDGB1         .............TCAGAGAATAGAAAAGTGGCACTGGTTAGTAAG--G-CA--TGGCA-GG-CC-CTGCAA---ATG-G-AAGG-T-TTGGTGGCT-TATCCTCA--T--TGGC-T-TA-----T-GTA-AC-TGC--T-A-...GTAGTGATTA--ATA-ATAGGGAC 4928
DEB.CM.99.CM40            .....................................ATTAGATAGGTGG--C-GTGC-A--AA--G-G-AA-ATACA-GACCAA--AATT--A--CT-T-CAA-GGATACAC---TTTG-ATT---GAACTCATATTAT-C-...CAAA--A--A--GTAT----CA-C 5029
DEB.CM.99.CM5             .....................................ATTAGATAAGTGG--T-GTGCCA--AA--G-GGAA-GTATA-GACCAA--AATT-CAAGCAAT-AA--GGATACAC---T-TG-ATT---AAGGGCATTCTAC-CT...CAGA--A--A--GT-T----CG-C 5023
TAL.CM.01.8023            .............................................................GTATC-CATGT-C-AAACCAAG...TACTT-CAA-A-TT-ATA-GGGTCCAT---TT-C-ATT---AATGGCCT-TTTC-CG...CAGG-CA--TGGAT-C----C--- 4687
TAL.CM.00.266             .............................................................GT-TC-CATA----AAACCAAG...TACTT-CAA-A-TTGATT-GGGT-CAT--CTTTC-ATTG--A-ACGCCT-CTAT-C-...CAAG--A-GTGGATAC----GT-- 5189
SUN.GA.98.L14             GACAGTATAAGATAGTTAGAATAGTATGGCTTATAGATAGAATTGCTGT-GA----TT-----ATTT-AGAAGA---C--AGGGA-ACA----A-GATT-GGTA-C-ATGTATGGAACAGG-ACTG-ATGGG-GTGGTAC-C-...TACAATA--A-TATT--T---G-- 5625
WRC.CI.97.97CI14          .....................................TAAAGGTAA-TTCA---G-T-GACCTATT-TATAGCA-ATCTAA--AA--CAGGGCAGGCAAAT-TTC-A-GGG--TC-TGGC---C-GCAG-AG-ATGCTTT-CT...--GCATAT----A-A--A---A-- 5511
WRC.CI.98.98CI04          .....................................TAAAGGAAGGTTTAAA-G-T-GGC-TATTGGATAGC-CATCTAA-GCA-AAT---AAGGCAA-T-TT--A-GGG--AC-TGGC--ACTCAGG-AG-ATGCTTT-CT...--A-ACCT-T-GAGA--A--CT-- 5526
WRC.GM.05.Pbt_05GM_X02    ...............................AATCTCAAAAGGAAG-TTTA-A-G-T-GAGCTATTGGATAGCAGA-CTAA-ACA--AA---CAAGCAA---TCC-G-T-G-CTCATGGC---T-CATG-AG-ATGTTTC-CT...--ACATAT----AGA--T--T--T 4443
LST.CD.88.524             ........................................GAGCTT-GT-GA----TT-C---AT-TCA-AAGG-ATC--TCAGA-A-T---AA-GATT-GGTTAG--TACATGG-ACAGGA-T-G-ATGGG-GTTTTAT-CT...TACAATA--T---TA--T---T-- 4488
LST.CD.88.447             ........................................GAGCCT-GT-GA---GTTCC---AT-TCAGGAGAGCAC--AAAGAGACTC--AA-GATT-GGTAAGC-TACATGGAACAGGG-T-G--TGGG-ATTTTAT-CT...TA-AATA-TA--GTG--A--TT-- 4491
LST.KE.x.lho7             ........................................ATCCTT--T-GAA--GTT-----AC-T-GT-AGGT-AA-CACAGAG-CA----AA-ATT-GGTAGGC-TACATGGAACAGG--T-G-TTGGG-ATTCTAT-CC...TA-CATA--T--GTG--A--T--- 5575
SYK.KE.x.SYK173           ..............................................-TTGCA-C-T-TAA--AA--C---TA-GTATA-AAGCAA--AATTG-AG-AA-CCACA-A-A-ACAT--CT--C-AAT-GAATGG--ATGGTAC-CT...TA--GCC--TGGACA--A---G-- 5417
SYK.KE.x.KE51             ..............................................--TGCA-C-T-TA---AAGGT----A-GTATA-AAGCAA-CAATT--AG-AA-CCACA-A-A-ACAT---T-TC-GATT-CATGGG-GTGGT-G-CT...C-AG--C--TGGG-A--A---G-- 5069
COL.CM.x.CGU1             ............................................................................-A-GCAG-TTCCAG--TGCCAATAT-AT-A--TAG--A-AACAG-GTC-CC-GAATGGTGG-C-GA-...G--AA-AT--AGTTTTA-TT-GGG 4977
```

```
H1B.FR.83.HXB2          GGG......GATGCTAGATTGGTAATAACAACATATTGGGGTCTGCATACAGGAGAAAGAGACTGGCAT................................................. 5280
Vif                      G  .  .  D  A  R  L  V  I  T  T  Y  W  G  L  H  T  G  E  R  D  W  H  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
H1A1.UG.85.U455         ---.......--A--------A---G---G---------------------------A---------C.................................................. 4726
H1B.US.90.WEAU160       ---.......--A-G--A---A---------------------------------------------................................................... 5279
H1C.ET.86.ETH2220       ---.......--G-------AA------A-------------T----A-------------T------................................................. 4672
H1D.CD.84.84ZR085       --A.........-------C-----G-------------------------------------A------............................................... 4802
H1F1.BE.93.VI850        -A-.......--A-T--A---A-------------------G-----CC------------A------................................................. 4620
H1G.SE.93.SE6165        --A.......-------C-C-A---G-----------------G--------------A---------A................................................ 4677
H1H.CF.90.056           --A.......--A-----G--A--C-----C-----C---------A---------------A------................................................ 4627
H1J.SE.93.SE7887        --A.......--G----T---A-------------------------G--------------T-----C................................................ 4594
H1K.CM.96.MP535         ---...........---AGAG--A---G----------------------T-----------G---------............................................. 4476
H101_AE.TH.90.CM240     --A.......--G--------A-------G-----------------A----------AG--------C................................................ 4854
H102_AG.NG.x.IBNG       ---..........---------G--A---G---G----------------------------C---------C............................................ 4805
H1N.CM.95.YBF30         --T.....C-G--A-----A--G-C-GTC--T--------G--AACA---------CAGTCT------................................................. 4874
H1O.BE.87.ANT70         ---.......-TG--CCATG-A---G----C-----------AT--ATGC----G--------TGAA---............................................... 5335
H1O.CM.91.MVP5180       -CA.......--A---GATA-A--GG-C--C-----------AT-AATGC----G--------GGAA--C............................................... 5310
CPZ.CD.90.ANT           CCAACATTG---AAA-A-------G-G---GTG-TC-----AT-A--ATGT------C-CCCA-----C................................................ 4725
CPZ.CM.01.SIVcpzCAM13   AAA...GATA--T-A-A---AA----C-----C-------C-T-AACC-------------CG-------............................................... 4863
CPZ.CM.05.SIVcpzEK505   --A.......--A--A-A---A------GTT--C--------AT-AATGC----------GCCA------............................................... 4852
CPZ.CM.05.SIVcpzMB66    --A.......--A---A---A--------------------A--CACA-------------C-------................................................ 4826
CPZ.CM.98.CAM3          -AA.....A-G--A-A---A--GG-------C--------AT-AATGC-----C-GCAT-C-------................................................. 4694
CPZ.GA.88.GAB2          A-A.........-----A-A-A-A--GG-CTGT-----C---AA-T-AATGC--------CA-CCA------............................................. 4673
CPZ.CM.98.CAM5          -AT.....ACA--A-A---A--GG-C--------------A---ACA-----GC--CAT-CT-----C................................................. 4982
CPZ.CM.05.SIVcpzMT145   AAT.....C-G----AGC-AA-TG-C--------------A--AATGC-------------T------................................................. 4822
CPZ.TZ.00.TAN1          ACA...GAAAGAT-A-A---A--GG-TTT-CAT-----------AGCCTGT---------CCA------CTAGGTCATGGC..................................... 4939
CPZ.CM.05.SIVcpzLB7     ---.......--G--C-A---A------GT---T--------A--TACAC--------C-G--------C............................................... 4835
CPZ.US.85.CPZUS         -AT.....A-A--A-A-A-A--G--T-T---G--------GT--A-A---------CA--C------C................................................. 5340
CPZ.GA.88.GAB1          A-A...GATT--T---A---AA--G-G-----C-------CCT-AAGCC----------G-C-------................................................ 5344
MON.CM.99.L1            -ACCTAGGAATCA--CA-GAA-G-GAGT--TGGGCAACATA-T--G-G...AT-AC--TTT-T---G-CTTAACAAACGTGGGACCTAGTGCACTCTCACCAACAGTGTACAACAAACAGGCCTACACAATAGAATGGGTTTATTGGATGCAAAAGGTGGAACGG..... 5088
MON.NG.x.NG1            -A-...ATG-G-ATC-C-CAARAGGRGGA-CA--GGGCAACATACATRGAGATCACG-TCCTT---G-CAYTACCAATGTAGGGCCCGCGTCACTATCTCCATCAGTCAAATGGAAGCAAGCATACATCATAGAATGGGTATATTGGAGGCAGCCCAGGGTGCCACAGCA 3686
GSN.CM.99.CN166         -ATACTAGGA-G-TAG-GGAA-A-...GT-GA-GCAACATA-A-AG-G-T-ACTATCCT-TGGGACACAACCAGTCATGGACCAGCTAGCTTATCCCGGTCCACCTATTACCAACAAGCAGTCATAATAGAATGGGTATACAAC.......................... 5098
GSN.CM.99.CN71          -ATACTAGGC-G-TAG--GA--A-...GT-GA-GCCACATA-A-TG-G-T-ACCATCCT-TGGGACACAACTAGTCATGGACCAGCCAGTCTATCCCGATCCACCTACTACCAACAGGCAGTCACAATAGAGTGGGTTTACAAC.......................... 5080
MUS_1.CM.01.1085        ACACAAAATACCCGGGA--CA-GTGC-GT-GA-AG-ACTTACA-AG-ACTGAC--TCCTGTGGGATGTGACCAATGTGGGACCAGCAAGTTGCTCTAAAAGCACATGGTGGAAGCAATCCTTTATCCTAGAGTGGGTATATCTA.......................... 5057
MUS_1.CM.01.CM1239      ACACAAAACACCTTAGA--CA-G-GAG---GA-GTCACCTTCA-AG-G-T-AC--TCCT-TGGGATGTGACTAATGTAGGACCAGCAAGCCTATCCAAAAGTACCTATTGGAAGCAGTCATACATCCTGGAATGGGTATACATG.......................... 5047
MUS_2.CM.01.CM1246      A-CATAAACAGAATAGA-GGA-GCCC----GAGAC-ACATACA-AG-G-T-ACCAT-TTGTTAGATGTCACAAACGTAGGGCCAGCCAGTTTATCAAGGTCCACCTATTGGCAGCAATCCTATATCCTGAAGTGGAGATACATC.......................... 5155
MUS_2.CM.01.CM2500      ACAATAAAGA-GAGAGAGGAA-AC-A-GT-GA--C-ACATACA-A--A-T-AC-ATTCT-TGGGATGT-ACCAATGTAGGGCCTGCCAGCCTCTCAAAGTCTACTTACTGGCAGCAGTCATATATATTGGAGTGGGTTTATGTT.......................... 5108
RCM.NG.x.NG411          --A...........-ATAACACT-AG-T--AGGT-ACAACATATTGG-GTCTCAC--T---AAA-GGATGGTTAGGAACCTATGGA................................ 4795
RCM.GA.x.GAB1           -A-...........GA-GGATC--AGTT--A-ATAACCACCTATTGG-ATTTGAC-GT---AAA-GGATGGCTAGGGACATATGGG................................ 4783
DEN.CD.x.CD1            A-A..............GC--CTTAC-TGGA-GTAAA-CTCTATTGGCATTTGACCCC---AAAAGGATATCTGTCACAAGTAGCA................................ 5206

MAC.US.x.239            CA-...........GA-GGAAGCCATTT-GA-GTACAA--GTATTGGCATTTGAC-CC---AAAAGGGTGGCTCAGTACTTATGCA................................ 5597
Vif                      Q  .  .  .  .  E  G  S  H  L  E  V  Q  G  Y  W  H  L  T  P  E  K  G  W  L  S  T  Y  A  .  .  .  .  .  .  .  .  .  .  .
H2A.GW.x.ALI            CAA...........G--AGAAGTCATCT-GAGATACA--CATATTGG-ACCT-AC-CC---AAAAGGATGGCTCTCCTCTTATGCA................................ 5673
H2A.DE.x.BEN            CAA...........G--AAAAGTCATCT-GA-ATACA--CATATTGG-ACCT-AC-CC---AAAAGGATGGCTCTCCTCTCATGCA................................ 5680
H2A.SN.x.ST             AAA...........G--GAAAGTCATCTGGAGATACA--CATACTGG-ACCT-AC-CC---AAAAGGATGGCTCTCCTCCTATTCA................................ 5125
H2B.GH.86.D205          AAC............-A-GGA-C-TGGCT-GA-GTCCAA--ATATTGG-ACCT-ACCCC---AA--GGATTCTTGAGCTCCTATGCT................................ 5653
H2B.CI.x.EHO            AAA...........GA-GGA-C-CATCT-GA-GTCCAA--ATACTGG-ACCTGACCCC---AA--GGATTCTTGAGTTCCTATGCT................................ 5651
H2G.CI.92.ABT96         ACA............-A-GAA-C-CATTTGGA-ATACAA--GTATTGG-ACTTGACCCCT--AAA-GGATGGTTAAGCCAGTATGCA................................ 5007
H2U.FR.96.12034         AAA............-A-GGATC-CATTTGGAGGTGCAA-CCTATTGG-ATTTGAC-CC---GAA-GGATGGCTATCCTCTTATGCA................................ 5158
MAC.US.x.EMBL_3         CA-...........GA-GGAAGCCATTT-GA-GTACAA--GTATTGGCATTTGAC-CC---AA-AGGGTGGCCCAGTACTTATGCA................................ 5065
SMM.US.x.H9             A--...........GATGA--CTCATTTGGA-GTACAA--ATATTGG-ATTTGAC-CC---GRAAGGGTGGCTCAGTACTTATGCA................................ 5070
SMM.US.x.PGM53          A-A...........GATGAATCTCATTTGGA-GTACAA--ATATTGG-ATTTG-C-CC---AAAAGGATGGCTCAGTACTTATGCA................................ 5514
SMM.US.x.SIVsmH635F_L3  A--...........GATAA-ACTCATTTGGA-GTACAA--ATATTGG-ATTTGAC-CC---AAAAGGATGGCTCAGTACTCATGCA................................ 5600
SMM.SL.92.SL92B         -A-...........GAGGGAAGTCACTTGGAGGTACAA-TGTACTGG-ATCTGACCCC---AAAAGGATGGTTATCAAGCTATGCA................................ 5031
STM.US.x.STM            CAA...........G--GAA-C-CACCT-GA-GT-CAA--GTATTGG-ATTT-AC-CC---AAAAGGATGGCTCAGTGAATATGCT................................ 5244
MNE.US.x.MNE027         CAA...........GA-GAAAGCCAGTT-GA-GTACAA--GTATTGG-ATTT-AC-CC---AA-AGGGTGGCTCAGTACTTATGCA................................ 5065
DRL.x.x.FAO             AAA...........GACAAAAGT-A--T-GT-GTAACA-TC---TGG-ACCTCACCCC----AAAGGGTGGTTAAGTAAATATGCCATA.............................. 4865
MND_2.x.x.5440          AA-...........GATGGTTC--AG-TT-TCATCACC-TCT--TGG-ACTTGACCCC---AAAAGGGTGGCTGTCTCAATACGCC................................ 4859
MND_2.CM.98.CM16        -AA...........GA-GAT-A-CAG-TC-T-ATCACCATC---TGG-ACTT-ACCCC---AAAAGGATGGTTATCAACTTATGCC................................ 5296
MND_2.GA.x.M14          AAA...........GATGGCTC--AG-TT-TTAT-ACT-CC--ATGG-ACTTGACCCC---AAAAGGATGGTTATCTCAATATGCA................................ 5223
OLC.CI.97.97CI12        CCA..............GG--GTGAG-T-GA-AT-ACACTCTACTGGCATATTAG-CAT-GGAAAGGATTACAAAATTCAGGTGGGGTG.............................. 5151
GRV.ET.x.GRI_677        TCA............-A-GAT-ATTAC-T-GA-GTGAATAT-TAT--C-ACCTCACCCC---AA-AGGATGGCTCTCAAGTCATGGA................................ 5468
VER.KE.x.TYO1_patent    --A............-A-CAT-GGGA--T-TGTGTAGACCTATATTGGCATTT-AC-CC---GCAAGGATGGCTATCCACATATGCA................................ 5519
VER.KE.x.9063           --A...........TT-GAA-G-GA--T-CAGATAGACT-CTATTGGCATTT-AC-CC---AA-AGGCTGGCTCTCTACTTACGCC................................ 5522
VER.KE.x.AGM155         CCA...........G--AGT-G-GA--TCCATGTGGATATCTATTGGCATTT--CTCC-A-ACAAGGATGGCTCTCAACTTATGCA................................ 5514
VER.DE.x.AGM3           ---...........CA-CAT-G--GT-T-CATGTAGATCTATATTGGCATCTGAC-CC---AAA-GGATGGCTATCAACATATGCT................................ 5017
TAN.UG.x.TAN1           --A...........G--GAA-G--A--T-GT-GTAGATT--TACTGGCACCT-ACGCC---ACA-GGATGGCTCTCTACCTATGCA................................ 5494
SAB.SN.x.SAB1C          AA-...........GATGGA-CTC---T--A-GT-ACTAA-TATTGGCATTT-ACCCC---GAAAGGATGGTTAGAAACCTATGCA................................ 5693
MND_1.GA.x.MNDGB1       .......................A---G-TT--T-GT-A-AA-CTATTGGCATTTGC---T---GGTAGGATGCTTAAGTACTTATGCAGTA............................ 4997
DEB.CM.99.CM40          ACA............-A-GAA-GGGG-GT-GT-GTAGAATTA--ATGGTGCTTGACCCC---GAAAGGATGGCTACCCACTATGGCA................................ 5104
DEB.CM.99.CM5           ACA...........GA-GAA-G--A-GT-GT-GTGGAAATAT-CTGG-GCCTG-CTCC---AAAAGGGTGGATGCCCTCTATGGCG................................ 5098
TAL.CM.01.8023          -AAGAAAAA--GAA----G-A-AGTAC-TT--CATCACTCAGTATTGGTGCTT--C-CC----AA-GGATACACATCAGACTGGGAT................................ 4774
TAL.CM.00.266           -AAGAAAAG--CA-C---G-A-A-TATTT---CATAACACAGTATTGGTGTTT--C-CC----CATGGATATACGTCCGATTGGGAC................................ 5276
SUN.GA.98.L14           ACT...............-AT-GG-C-GT-GT-GTAA--ATCTATGGACACCTCAC-CC--CAAAAGGGTTGATTAACCAGTGGGGA................................ 5697
WRC.CI.97.97CI14        CA-...........G--CACTGGGAG-TGCA-AT-A-AT-CTATTGG-ATTTGTC-CC---AAA-GGATGGATGGACACCATGGGATTT.............................. 5589
WRC.CI.98.98CI04        CAA...........G-GCATTGGGA-TT-CA-ATCA--T-CTATTGG-ATCT--C-CCT--AAA-GGATGGTTAGACACCATGGGGTTT.............................. 5604
WRC.GM.05.Pbt_05GM_X02  CAA...........G-GCATTGGGA--TGCAGATAA--T-CTATTGG-ATTT--CTCC---AATAGGATGGATAGATACTTTGGGATTC.............................. 4521
LST.CD.88.524           -AT...............ACA-G--CTTT-GT-GTAA-AAT-TATGC-CATCTG-C-GC--GGA-AGGATGGATTTCACAGTGGGCA................................ 4560
LST.CD.88.447           -A-...............ACA-G--CCTT-GT-GTGA-AT-CTATGCACACTTG-C-GC--GAA-AGGATGGATCTCTCAGTGGGCA................................ 4563
LST.KE.x.lho7           -A-................-CA-G--CTTT-GT-GTCA-AATATATGCACATTTG-C-GC--GAA--GGATGGATCTCACAATGGGCA................................ 5647
SYK.KE.x.SYK173         --A...............GAT-G--CT-T-TGGATCAC-TTCTAT--C-ATTTG-CCCC---AA-AGGATGGTTACATATGCAGGGA................................ 5489
SYK.KE.x.KE51           --A...............CAA-GTG--TT--TTATAAAATTCTAT----ATTTGAC-CC---GAAAGGCCTGTTAAGAACAGAGGGA................................ 5141
COL.CM.x.CGU1           TCA...............GCTA-T-AGTT--GCATAATATACT--GGC--TTT-ACCCC---TA-AT--CAT............................................... 5034
```

```
H1B.FR.83.HXB2             ...............................TTGGGTCAGGGAGTCTCCATAGAATGGAGGAAAAAGAGATATAGCACACAAGTAGACCCTGAACTAGCAGACCAACTAATTCATCTGTATTACTTTGACTGTTTTTCAGACTCTGCTATAAGAAAGGCCTTATTAGGAC 5419
Vif                        __.__.__.__.__.__.__.__.__.__._ L  G  Q  G  V  S  I  E  W  R  K  K  R  Y  S  T  Q  V  D  P  E  L  A  D  Q  L  I  H  L  Y  Y  F  D  C  F  S  D  S  A  I  R  K  A  L  L  G
H1A1.UG.85.U455            ...............................--------T--G------------------CTG--A-------------------T-----C-----------T--------C---C----T-----------------A-----C-----G-GA---A--------- 4865
H1B.US.90.WEAU160          ...............................--------A---------------------C---------------------------C--------------------------C-T-----T-----------G--C-----------T---A--------- 5418
H1C.ET.86.ETH2220          ...............................--------T------------------ATTG-GA--C----A----------------GC--G--------C---------A--C----T-----T-----G---A----C-----G--A---A--------T 4811
H1D.CD.84.84ZR085          ...............................C-------------------------G------------------------GC--G------------------A-------T-----T-----G------------C-----A---A--------- 4941
H1F1.BE.93.VI850           ...............................C------------------------C-GGG--AG----G-----A---------GC--G--G-------G--C---A-A-----T-----T----------A-----C-----G--A---A--------- 4759
H1G.SE.93.SE6165           ...............................--------T--G----------G-----C-G-GA--------A-----------G-----C-----------T--G---------C----T-----T-----------------C-----G--A---A--------- 4816
H1H.CF.90.056              ...............................--A--C----------------------ACTG--A--G------------------G---GC--G------------------A--C----T-----T----------A-----C-----G--A---A--------- 4766
H1J.SE.93.SE7887           ...............................-------------------------C-G-GA--G--C--A-----------------GC--G-------------------A---G---------T-----------------C-----G-----A--------- 4733
H1K.CM.96.MP535            ...............................-------------------------CTG--A--G--C--A------------A-----C----------------------A-A-----T-----T----------A-----AG----G-----A--------- 4615
H101_AE.TH.90.CM240        ...............................--------T------------------C-G-GA-A--------------A---T-----C---------A---G---------AC-A--T----G-----------------C-----G--A---A--------- 4993
H102_AG.NG.x.IBNG          ...............................--------T--G--------------A-C-G--A---------------A----------CT--------------G-----C--------T---A-T-----------------C--------A---A--------G 4944
H1N.CM.95.YBF30            ...............................C-A--A--T-----A--------------ACT--GA-A---C-AG---------T--T------A-G------A-G-----A-----TC----T-----T------A---C------C-----GC-A--GG-C-----GA 5013
H1O.BE.87.ANT70            ...............................-----A--T--G--TAGT---------C-TT-T----AG----AA-----GA-T---------AC-------AGGA-G--A------C----T---AC-------A---CA--A--AG-C--G--A---A-CC------ 5474
H1O.CM.91.MVP5180          ...............................-----A--T--G--TAGT---------CAAT-C----GAG----AA-----GA-T--T------AC-------AGGA-G--A------C----T--CACA------A----A--A--A--C--G------A-TC----G- 5449
CPZ.CD.90.ANT              ...............................--A--A--T--T--GAG-------------ATGTGGA-AG----T-----------------A---AC---T--T---A-G--A--C-AA-----T--CCC------------TCAA--AG-C---C-A--AA-CC-----G 4864
CPZ.CM.01.SIVcpzCAM13      ...............................--A-----T------------C-G--------TGGGATC----GTT-----G-----T---T-TG-------TA--T-G--C--CTCTC-A--T-----T------G-----------C--C----GA--TA-------G- 5002
CPZ.CM.05.SIVcpzEK505      ...............................--A--A--T-----A--------------AC--GGA-T---C--G-------A----T------T-G-------A-G-----A--C--T-----T-----T------A---C------C--C--GC-A--GG-C-----GA 4991
CPZ.CM.05.SIVcpzMB66       ...............................C-A--C--T--G--AAGT-----------AC-G-GA--T-----G------A----T------T--G----------T----C---T----C--T-----T------G----A-----C-----G------A--------- 4965
CPZ.CM.98.CAM3             ...............................C----G--T-----G--------------ATGTGG-TC------T----------------GA-G-----TA-----------C--TC-A--T-----T------G---CT-----C--C--G--A--TG-T-----GA 4833
CPZ.GA.88.GAB2             ...............................C-A--C--T-----G--------G------C-G-GA-C------A--T--GA----T---C--AC------T-GG--G--A--CTGT-----T-----T------G---C--------G-C----GA---A-CC----GG 4812
CPZ.CM.98.CAM5             ...............................C----A--T-----G--------------ATGTGG-TC---C-AA------------T------ACG-----TAG-----------C--TC----T-----T------G---CT-----C--C--G--A--TG-T-----GA 5121
CPZ.CM.05.SIVcpzMT145      ...............................C----A--T-----AG-A-----------AG---GAG-T--C--T--C--G-------------AC------TAGG-----A--C--TC----T-----T------A---CAA----C--C--G--C--TA-C----A-A 4961
CPZ.TZ.00.TAN1             ...............................A-A--ATTA............---------AC--GGA-A---C--T-------A------------AC-----------T-G-----CACTAGG--T---AC--------G-T-CAGGA--AG-TC-GC-A--AA--------G 5066
CPZ.CM.05.SIVcpzLB7        ...............................--A-----T--G--A--T--C--G---AAC-G----A-------T-------A----------GT-G-----T--TA-G------T-A-----T---A-T------G----A-----C-----G-----A--------- 4974
CPZ.US.85.CPZUS            ...............................--A--G--T-----A--A-------------AC--GGA-CT--C-AG------A---T-----GA-G------A-G------------------T-----T------A---C------C--C-----A--TG-T-----GA 5479
CPZ.GA.88.GAB1             ...............................--A--C--T-----------TC--------CT-GGATCT---GT-------G--G--T--CTTCAC------TAG-T-G--C--CTCTC-A--T-----T--C---G---AA-----C--C----GA--TA-------G- 5483
MON.CM.99.L1               ....GTCCCTTTAGATCCCGTGTGG......AATCCAA-CCC-AGAGAT-G---GGTC...-G--CAT-G---GCT--C--TC-GACG--G--C--G--GA-T--GA--G-A---ACAC----T----CA--C--CCA-C-G-TG-ACG-GC-C-GA--AA-CAG----G 5245
MON.NG.x.NG1               AGAGGGAGTAGTAGATCGAGTGTGGTTTCCTAGTCCCAGAA--CCGGA-CGG---ATTGAT-G-GTCT-G---GCT--CT-CC--ACA--A--TG----T------A----C--CACCCGC------TCA-----CCAG-C-----ACG--C-T-GA--AA-TAG---GG 3856
GSN.CM.99.CN166            ...............................AGAA-CA-ACT--GAG-A-G---GG-AGAC-T--TTT-G--C-----TA-TC-GAC---A-G-G-----ATG---A-T--A---GGAA-A------CCA--C---CAGA--GAA-AC---C-G-G---AA--AGG--TG 5237
GSN.CM.99.CN71             ...............................AGAA--A--CT--GAG-A-GG---ACAGAC-T--CAT-G--C--T--TA-TT-GAC---A--CG-----ATG---A-C--A--CACAAGG------CCA------CAGA--GAA-ACG--C-G-G---AA--AGG--TG 5219
MUS_1.CM.01.1085           ...............................AG-A-A--TCC-CAGGA--GG---CA-GATT-C-TCT-G---TCT---T-CT-GAG---CAC---G--CATGAG-A-C--A---ACCC-G--T---TCT--C--C-ATAG-CAG-AC--CCC--G--T-A-CAG----- 5196
MUS_1.CM.01.CM1239         ...............................AG-AAAGGTCC-AGAGAT-G----C--GACT-TGTTT-G---CA----T-CC-CACA-----GA----CATGA-GA-C--A--CACCC-CC----CTCT-----C-ATAG-CAG-A---CCCG-G--T-A-CAG----- 5186
MUS_2.CM.01.CM1246         ...............................AG--CAGCCCC-CAAAAT-G---GATAGAC-TGGT-CAT--CGAG--CTTCC-------A---G-GA-CATG--GA-T--A--CACAC-C-----CTCT-----CCA-C-GAGA-AC--CCA--G---AA-CAG----G 5294
MUS_2.CM.01.CM2500         ...............................AG--C-GCCCC-CAAGAT-GG--GGT-GAC-TGGTAT-G--C--T--CT-CC-T-----A---T-G--CATG---A-T------ACT---C----CTC-------CAG--ACAG-AC--C-T--GA---C-CAG---GG 5247
RCM.NG.x.NG411             ...............................G-A--AATA-C-TA-ATAGATA-CA-AT-TG-TCCCT-T--CTT----G-TA-T-----AAC-G-------TA--T--G-A---AATC-G--T---CCT------A----TAAA--A--T---C-A---A-C------G 4934
RCM.GA.x.GAB1              ...............................G-A--AATCTT-................----TA---GGAGAT---GTA---G-T---TTT---TGGACT-----TAGTT----A---AAAAT---T---CCA------A----TAGA--A--C---C-AAGTC--C----GG 4910
DEN.CD.x.CD1               ...............................A-CA-CATA-CTTAT...G--C-CCTC--C-C-GG-T-G...GCT---G------------T-AC------TG-GA----A---AAAG-CC-T--CTCT--C---A-CAGACAA--AG-C---C-A------AGG---G 5339

MAC.US.x.239               ...............................G--A-GATAACCTGG-A-............TC-----AC-T-T-G---G-T---ACA--AA-CTAT------ATTT--C-G---AGCAC---T--CCCT--C---A---CGGGA-AAG-G----G----A-CAGG---G 5724
Vif                        __.__.__.__.__.__.__.__.__.__._ V  R  I  T  W  Y  .  .  .  .  S  K  N  F  W  T  D  V  T  P  N  Y  A  D  I  L  L  H  S  T  Y  F  P  C  F  T  A  G  E  V  R  R  A  I  R  G
H2A.GW.x.ALI               ...............................G-AA-AATAACTTGG-AT............-C-G-A-AG-TCT-G---G-T--TAC---A--CTGT------TCC-----A---GGCAC---T--CTCT--C---A-G-CAGG--AAG------GA---A-CAG---GG 5800
H2A.DE.x.BEN               ...............................G-AA-ATTAACCTGG-AT............-C-G-A-AG-TCT-G---G-T--TAC---A--CTGT------ATC-----A---AGCAC---T--CTCT--C---A-G-CAGG--AAG------GA---A-CAG---GG 5807
H2A.SN.x.ST                ...............................G-AA-A-TAACTTGG-AT............-C-G-A-A--TCT-G---G-T--TAC---A--CTGT--G---TCC-----A---AGCAC---T--CTCT--C---A-G-CAGGC-AAG------GA---A-CAG---GG 5252
H2B.GH.86.D205             ...............................G-AA-A-TAAC-TGG-AT............G-G-G--AC-T-TAT---G-T---ACA-----TG-G---------G---C-G---GG--C---T--CTCT--C--------C-AA--AAG----G-GA---A-CAGG---G 5780
H2B.CI.x.EHO               ...............................G-AA-A-TAAC-TGG-AT............G-G-G---C-T-TAT---G-T---ACT-----TG---------G-T--C-G---GG--C---T--CTCT-CG---A---CTAA--AAG----G-GA---A-CAGG---G 5778
H2G.CI.92.ABT96            ...............................G-AA-ATTAACCTGG-A-............-C--GA-A--T-TAT---G-T--GACA--A--GAC------T---T--T-G---GGA-C---T-----T--C---A---CTGG--AAG------GA---A-CAGG---G 5134
H2U.FR.96.12034            ...............................G-AA-GTTAACTTGG-A-............-C--GAG-T-T-T-G---G-T---ACA-----TTGC---------T--C-G---GG-AG---T--CTCT--C---A---CGGG--AAG-G--G-GA---A-CAG----G 5285
MAC.US.x.EMBL_3            ...............................G--A-GATAACCTGG-A-............TC--G-GACCT-TTGGACAG-TGTA--A-CAG----TGCAGG--TT-TTCCTGCA-AGCACTTA--TC-CTC---A---CGGGA-AAG-G----G----A-CAGG---G 5192
SMM.US.x.H9                ...............................G--A-AATAACCTGG-A-............TCC-GA-AT-TCT-G---G-T-C-ACA--A--TTAT------ACTT--C-G---GGCAC---T--CCCT--C--------GGGA-AAG--C---G----A-CAGG---G 5197
SMM.US.x.PGM53             ...............................G-AA-AATAACCTGG-A-............TCC-GA-A--TCT-G---G-T---ACA--A--TTAT------ACTT--C-G---GGCAC---T--CCCT--C--------GGGA-AAG--C---G----A-CAGG---G 5641
SMM.US.x.SIVsmH635F_L3     ...............................G-AA-AATAACCTGG-A-............TCC-GA-AT-TCT-G---G-T---ACA--A--TTGT------ACTT--C-G---AGCAC---T--CCCT--C--------GGGA-AAG--CA--GA---A-CAGG---G 5727
SMM.SL.92.SL92B            ...............................G-TA-GATTACTTGG-A-............TC-G-A-A--T-T-G---G-T---AC---A--TG-------T--GT--GCA---AGTAC---T--CCCT--C---G---CGCAC--GG-G---C-A---A-CAG----G 5158
STM.US.x.STM               ...............................G--A-GATAACTTGG-AT............-CC-GA-AT-TCT-GT--G-T---ACA--A--CTGT-----T--G---C-G---GGCAC---T--CCCT--C---A---CGGGA-AAG-G----GA---A-CAGG---G 5371
MNE.US.x.MNE027            ...............................G--A-AATAACCTGG-A-............TC--G--AC-T-T-G---G-T---ACA--A--CTAT------ATTT--C-G---AGCAC---T--CCCT--C---A---CGGGA-AAG-G----G----A-CAGG---G 5192
DRL.x.x.FAO                ...............................ACAATAG-ACATA-T..............C-GG---AG-T-T-T--CTT-A-------AAC-AC-------AG-A----C--CT--G-A--T--CCCT------A----TCAA--A------GGT-T----C-G---T 4989
MND_2.x.x.5440             ...............................--TACAGTA-A-TAT............CAA--GG-A-AT---TT---TT-CA--------A-TGACG-----TAG-A----C--CGG-G-A--T--CCCA------A----TCAA--A---------A-----G--T---G 4986
MND_2.CM.98.CM16           ...............................A--ACAATA-A-TAT............-A-TTC-CA-AC--CTA---T--CA-------CC--AC-------AGGA-G--A---TG-A-A--TC-CCC-------A---ACAA--A------C-A---C-GC-----A 5423
MND_2.GA.x.M14             ...............................A-CACAGCA-A-TA-............-AC---GG-GAC---TAT--T--CA-------A-T-AC-------AG-A-G------TG-G-A--T--CCCA------A---CTG-C---G-G-------TGC-G-AC---G 5350
OLC.CI.97.97CI12           ...............................AATTT-A-ATATA-AGTA............GGG-TTTC--GG-TG...T-TT-GACT--A--GTAT-T---TTTCA-GG-A--C--AGTG-------C-----GGA-TC-A-----------G--A--TA-T--G---A 5275
GRV.ET.x.GRI_677           ...............................G-A--GTTATCCTAT-A-......CATCAA--GGGA...----AG---G-------T--A-G-AC-------AG-A-G--A--C--A-----T---A--------A----TAGA--C--CCA-C----TA-CAG---GG 5598
VER.KE.x.TYO1_patent       ...............................G-A---ATACA-TATGTA...AGCAATTTAG--TCT-A------A---G--T---------CTAC------TAGTA----A---GGTC-C--T---A-T------AA---AAGA--C--CCA-C-A--TC-GAGG--C- 5655
VER.KE.x.9063              ...............................A-A--AATACA-TA--TG...-G-GCAGTAGG-...-AT-TC--A---G--T-------A-C-AC------TGG-A-G--A--C--AC-C--T---TCA------A----AAGG--C--TCA---A--TA--AGG---G 5655
VER.KE.x.AGM155            ...............................G-A--AATACA-TATGTT...AGCCTAGTA--TG-T-A-------A---G--T----T--CA-TAC-------TCCA-G--A---TGTC----T---AC-------A----TAGA--C--CCA-C----AC--AGG---A 5650
VER.DE.x.AGM3              ...............................GA---GATACAGTATCTA...AGCAAT---G-TCCTT-G-----G---G--T-G--T----C-AC------TAGC--G--A---ACCC----T---ACT------A----AAGG--C--C--G--A---C----G---- 5153
TAN.UG.x.TAN1              ...............................G-A-CAATAA-TT-T......--GAATT--C----C-CG--C-AA---G-----ACT-----TG-G--G-----TA-G------TGCC-C-----CCCT--C---A---TAGA--C--CCA-C-----A-CAG---GG 5627
SAB.SN.x.SAB1C             ...............................ACA--AATA---TA-.............--T---GG-GA--GGTTT---G-GC-T--T---TGGAC---G-----TA-T------TG--CC--T--CCC-------A---TAGA---G-CCA-C-A---A-CAGG---G 5820
MND_1.GA.x.MNDGB1          ...........................AGCATAGAA-CAGTA-TTAGA...CCGCC-CCCTTTG-G--AGAG-GGT-T---G-GA--ACT--A--GG-------T--T-----A---T-AC---TT-A------C--CATG---AG---AG-T-TG--A---A-CAGG---G 5139
DEB.CM.99.CM40             ...............................G-A-CAATA-C-TGG............GAA-C---A-C--GG-AG---G--C--ACA--A--TT-------T--C--T--A--C--AAGA------CCA-----C--CC-ACAG--AG-G---C-A---A--AGG---G 5231
DEB.CM.99.CM5              ...............................G-A--AATCTCTTGG............GA-C-G-CAG-G-GGCAG---G--T--ACA-----C-----T--T--TT-G--A--CT-AAGA-----CCCT-----C--CC-ACAG--AG-C-C-C-A---C--AAG---G 5225
TAL.CM.01.8023             ..................ACAATATCCATTGA-T-GTCA-TCTCGGATG-GACC-T-GATGC----GTC---TC----G-GT--ACA--A--GG----------C--G-----CACCA-G--T--CCCA------A---C---A--AG-C---C-A---A-CAG---GG 4925
TAL.CM.00.266              ..................ACAATATCCATTGA-T-GTCA-TCTCGGAT----TCC-AGACG-----GTG--CGC---TG------AC---AA-TG-G-----T--CT-G------ACCA-A--T--CCCA------A---CG--A--AG-C---C-A---A-CAG---GG 5427
SUN.GA.98.L14              ...............................-GTA--AT--A-TGG............-TTT-T--C--T---CAA---G--A----T---CTTG-G-----T--GA-G--C--CT-TA-G--T-----T----GGA-CTCTAGGTG---C----GA--TA-G-----GG 5824
WRC.CI.97.97CI14           ...............................AACAT-AGAATG--G-AT............G----T----GGGAA---TTTA--ATG--ATC-GAG---A-TA--A-C--A--CAGTC-AC-T-----A----GGA--C-GAA--TGG-G-----TA-AA-TAGGAC-G 5716
WRC.CI.98.98CI04           ...............................AACATAAGAATG--G-AT............G----T----GGGAA---TTTA-TATG--CTCTGA------TGT-A-G-----CAGTC-AC-TG---------GGA--C-AAA--TAG-T---G-TA--A-TAG-TC-G 5731
WRC.GM.05.Pbt_05GM_X02     ...............................AACATAAGAATG--A-GG............G----T-A--GGGAA--TTTTA--ACTGTAA--T-G-----TGTCT-G--A---TCTC-AC-------G----GGA-CC-AAG---AG---A---CA-AA--AG-CAGG 4648
LST.CD.88.524              ...............................G-ATCAATA-A-TGG............-TCT-TGGA-AC---CAG---G----G--T--CATT-----T--T---A-GC-----G--A-G--T----------GGGTGC-TAGG-AC---GTT-G----A--AGG---G 4687
LST.CD.88.447              ...............................G-ATCAATA-A-TGG............-TTT-TGG--AC---CAA---G--A----T--CATT------A-T---A-G--A--CTGTA-A--T---A----C-----TACAAGG-AC--T--G-----T----GG---G 4690
LST.KE.x.lho7              ...............................G-TTCCATA-A-TGG............GTCT-T-GAGAC---CAA---G--A-------CAT-------------A-G--C--CTATA-A--T---A----C--CAT-C-AAGA-AC--T----G---AC-TC-G---G 5774
SYK.KE.x.SYK173            ...............................A-TA-GATTCA-TA-CAA...TGGAACCA-TGG--C.........--TG-CT--ACA--A-C-G---------GC--T--A---AAT-T---T--CCCT--C---A---CAAGA--GG-C-ACC----AG-CAG---GG 5616
SYK.KE.x.KE51              ...............................G-A--ATTAATTTGGG--...C-TCC-TCTGG-TG-.......--T---G-----ACA--AC--AC------TGC-T--G-A---AGTC-A-----CCCT-----------TAGG--AG-CCA-C-A--AC-TAG---GG 5271
COL.CM.x.CGU1              ...............................AGA--A-CACC------TTAT----TACA--T-GGT-AG-GGGAA---G-TT--AG-TGGC---A-----TAGGGTAC--G--CA-CATAAGGGGGA-G---G-ACGCATAGAG-AAGA-GC--GAAAAGCTA--A-GG 5173
```

```
H1B.FR.83.HXB2            ACATA......GTTAGCCCTAGGTGTGAATATCAAGCAGGACATAACAAG.................GTAGGATCTCTACAATACTTGGCACTAGCAGCATTAATA...............ACACCAAAAAAGATA...AAGCCACCTTTGCCTAGTGTTACGAAACTG 5547
Vif                       H  I  .  .  V  S  P  R  C  E  Y  Q  A  G  H  N  K  .  .  .  .  .  .  V  G  S  L  Q  Y  L  A  L  A  A  L  I  .  .  .  .  .  T  P  K  K  I  .  K  P  P  L  P  S  V  T  K  L
H1A1.UG.85.U455           -A--......----A-------------------------------.................-----------------T-----T--GAA-------G--...............--C----C--G-GC-...----------------------A----T-A 4993
H1B.US.90.WEAU160         -TT-......----TT----------------------G---------.................--------------------------A------------...............---------------...----------------------A------- 5546
H1C.ET.86.ETH2220         -T-G-......-----------------C-----------------T---.................-------------------------GA-------G---...............-AG---------GC-...----------C-----------GT---T-A 4939
H1D.CD.84.84ZR085         -T--......----T-----------------------------------.................-------CT----G--T--A------A------------...............G-----------G-...---------------C-------A---G--A 5069
H1F1.BE.93.VI850          -T-G-......A-------------A-C-----------------T---.................-------C--------T--------A------------...............G-T---G-G----C-...--------GC----------CCA------A 4887
H1G.SE.93.SE6165          -A--......-------------------------------------TC--.................-----------------T------T-GAA--T----G--...............---T------G--GT...-G-----------C---------G--T-A 4944
H1H.CF.90.056             GTG--......-----A---------A-C----C-----------AC--.................------A----------------T--A--------G-G...............G-----------------...----------------------GA--G--A 4894
H1J.SE.93.SE7887          -A--......-------------------C--C-----------------.................-----------------TC---------A------------...............-G--------G--G-...----------------------CA------A 4861
H1K.CM.96.MP535           -T-G-......-----T-------------G--------------------.................-------CT-------T--------CA--------G--...............G-T----G--G-CC-...---------G-----------A------A 4743
H101_AE.TH.90.CM240       -AG--......-----A-G-------------C-T---------------.................-------C--------T--------GAA--------C-...............-----------G------...G--------C----------A...G--A-A 5118
H102_AG.NG.x.IBNG         -A--......-----A-------------------------------T---.................-------G--------T--------GAAT--T---G--...............G------C-----C-...----------------------G---GT-A 5072
H1N.CM.95.YBF30           GACC-......--ATTA-----------------C------G--C--AC--.................-----CA-C--------TC-A------A----C-GGG-G...............GG-G----G----G-...---------C--A--------G--T--G--A 5141
H1O.BE.87.ANT70           -G-G-......--ACTGA-C-A---------C-CTA---------GTC--.................-----GA-A------CTGC-A--T---AG----G--G--...............-A-G---G--GC-GG...------T--CC-A--C-----CCA----T-A 5602
H1O.CM.91.MVP5180         -G-G-......--GCTGA-C-A---------C-TG----------GTC--.................-----GA-A-------T---A--CT-GAA----G--G-G...............-A-GT-----GA-AT...------T--CC-A--C-----CCA--G-T-A 5577
CPZ.CD.90.ANT             -A-G-......A-ATTGA-ATAC--CC-C--CA-GAAG-------GTC--.................-----GA-A--G--G--T--A--TT-TTGTAAGA-CC--......GAATTCAGAGG-TACCCC---GGTCCA-GAAG--AG--C------C-GT-T-TCT-A 5004
CPZ.CM.01.SIVcpzCAM13     -T--T......--AGAA--CC------CT--CAG--A------C-GGC--.................-----CA-C-------T-C-A--TT--AA---T---G--...............-A-GA-----G-C-T...-GA-----C--A--C-----AG-T---T-A 5130
CPZ.CM.05.SIVcpzEK505     GACC-......--A-TA----A---------C-CT-----G-----AC--.................-----C--C-------TC-----T--AT---C-GGG-G...............GG-GT-C-G----G-...-GA-----C--A--------G--C--GT-A 5119
CPZ.CM.05.SIVcpzMB66      -A--......-------------------C-A-------------------.................-----CAG-T-------T------T-GT--A---------...............-A-AGC---CCA-G-...-GA-----C-----C-------GA---T-A 5093
CPZ.CM.98.CAM3            -ACA-......--ATAT----A---------C-C------G---C-AC--.................-----G--AT-------T--A--C---AG---C-GGG--...............-G-GT-GG-----A-...--A-----C-----------C--T--GT-A 4961
CPZ.GA.88.GAB2            -G--T......--G-CAAAGGT--------TC-CTAGG--G---CCTC-A.................--TCAG--A--------TC----CA---T-CA-G--G--...............-AGA----G--A-AG...---------CC-A--------C--T-TCT-A 4940
CPZ.CM.98.CAM5            -ACA-......--ATAT----A---------C-C-A----G--CC-AC--.................-K---G--A--------T--A--C---AG---T-GGG-G...............GG-GT-C------A-...-CA-----C-----C---A--GSC--GT-A 5249
CPZ.CM.05.SIVcpzMT145     -GGC-......--CCT-----A--------TC-C--A---G--C-G----.................-----G---T---------G---T---AAGCAG--GG-T...............-A-A-C--G--A-AT...-GA------C----CTCA--AG-A--GT-A 5089
CPZ.TZ.00.TAN1            -A-G-......A-ATTGA-ATTC--CC-C-T----T-------C-GAC--.................-----GA----G----T---A--TT-CAG-AAGG--G-TGAGAGCCAA...GAT-A--AGCC----GG-CCA-G-AGG--C-----ATC------A-----A 5209
CPZ.CM.05.SIVcpzLB7       GAG--......----------A--------T--C---------C--G---.................------A-C--------T------A-CT-T--C---T--...............CA-AG-C-GCCA-C-...-G---G-------------C-GC---T-A 5102
CPZ.US.85.CPZUS          GACC-......--A-TT--C-AA--------C-CTA----G--C--T---.................-----C---T-G---T------TT--AAG--C---G--...............GG--AG-GC--A-G-...-GA-----C--A--------G--T---T-- 5607
CPZ.GA.88.GAB1            -GC-T......--AGCG---C-T--------CA---A------C-GGC--.................-------CT----G-T------C---AA---T--------...............T--GAG-G--G-CAT...-GA--------A--------CG-C---T-A 5611
MON.CM.99.L1              -G-AG......C-CTTGGGACA----C--C---T----AC---CT--...............CCGAAG-----GA-C-CGTTGAC---A-AGAA-TT---C--CC-TGCCCTAGTGCGATATGG-G---CCCCC-C-AAGTCAT-CG--CCATTGGACC-CC--G..... 5389
MON.NG.x.NG1              -AGCG......M-CCTAGGG-AC---C-CC-CAT-CAGAC---CT--...............AATAAG------A---CT-C-AC-C-A-AGAA-TT---C--CC-TGCTTGCTTCAAGTATGG-AA-TGC-CC-CCGAGCCAT-CT-CACATTGGACC-CC--....... 3999
GSN.CM.99.CN166           -GCAG......A-CCTAGGA-AT------C-C-CG-A--CT--CCTGC--.................------A-A-CT--GACGC-T-AG-AGTT---C--CT-TGCATATGTCAAACATT............................... 5336
GSN.CM.99.CN71            -ACAG......C-CTTAGGATAT-----CC-C-CG-A--CT--CCCAC--.................------A-A-CT--CACGC-T-A-AGGCT---C--CT-TGCGTACGTTAGATACT............................... 5318
MUS_1.CM.01.1085          -GCAC......AC-CTAGGGGAT------C---C--AC-C---C-CA---.................------C---CT-CT-CT--ACA-G-CCT---T---T--GCATTAAAGAAGC................................... 5292
MUS_1.CM.01.CM1239         -ACC-......CCCTTAGGGGAT------C-C-C--G--C-----CA...AAG.................------C-C-CT-CCACTC-ACAGG--TT----C--T--ACATTAAAGAAACATGG-AAG-GT-CC-CCCAGCCA-AGA-C-CC................. 5314
MUS_2.CM.01.CM1246        -ACAG......C-CCTAAGA-A------GC-CATCAAGAC---CT--CC-AAG.................------A---CT-T--CTC-A-A-AAGCT------CT-TGCATACATCAAACAT---AATGG-GCA-GCGCCTCCAAGT-ACCCT--CTC............ 5431
MUS_2.CM.01.CM2500        -A-GG......A-CCTAGGG-TT---C--C-C--GA---C---CCT-CCCAAG.................------A-C-CT-CT-C-C-A-AGAAGTTG--C--CT-CGCATATATCAAACAGGG-GATGG-GCA-C-CCCTCCAAGT-ACCCT--CCC............ 5384
RCM.NG.x.NG411            -GCA-......--GCTG-T-........-----GG-T-G-GAA--------GA...............GATCAG-----GAGC---------TC-----A--AA---T-GGGCTAGGCAGCAGCTAAAGG............................... 5033
RCM.GA.x.GAB1             -A-A-......---CTAGTG........-----CC-TC----GG-------GA...............GATCAG-----GA----G----TT------A--CA---G-GGGCTAAA...AGTCAGTTAGATAGGT........................... 5012
DEN.CD.x.CD1              -AG-G......--CCTGAGC-CT--C-CCAGGA-GAAG-C----T--C-TCCA......AGGACAGGTCC---TAGGGA--CGCCA-CATT--AGTATAT-G--T--......CAAGCAGCC-TTAG-T-TGGA-AGGTT-CCA-CATCACA--C-CCAGC-TT-GTT-C 5491

MAC.US.x.239              -ACA-......C-GCTGT--T-C--CAGG-TC-CGAG--CT-----GT-CCAG.................---CC-AGC-----G-----A-----GAA--T-G---GCG.................................................... 5811
Vif                       E  Q  .  .  L  L  S  C  C  R  F  P  R  A  H  K  Y  Q  .  .  .  .  .  .  V  P  S  L  Q  Y  L  A  L  K  V  V  S  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
H2A.GW.x.ALI              -A-AG......T-ATTGT-CT-C--CA-T----CCCAG-CC-----GTCACAG.................---CCG--A--C----TTC----CT---TG-T-G-GCA-CAAA............................................. 5890
H2A.DE.x.BEN              -A-AG......T-ATTGT-CT-C--CA-C----CCCA--CT-----AGCACAG.................---CC---A--T------C-A--C----T--T-G--CA-CAAA............................................. 5897
H2A.SN.x.ST               -A-AG......C-ATTAT-CT-C--CA-C--C-CCCA--CC-----GT-CCAG.................---CCG--A--C--G-TTC----CT---TG-T-G-GCA-CAAA............................................. 5342
H2B.GH.86.D205            -A-AG......A-ATTGT-CTAC--CA-C----C-T---CT--CG-AGG-CAG.................---CC-AGCT----G-TTC-A--C---AGG-TCG--CAGGAAGGAAAAA........................................ 5876
H2B.CI.x.EHO              -A-AG......A-ATTGT-CCAC--CA-C--C-C-T---CT----CAGGCCAG.................---CC-AG-T----G-TTC-A--C---AG--TTG--CA-GAAGGAAAAG........................................ 5874
H2G.CI.92.ABT96           -GCAG......A-ATTGT-CT-C--CA-C----C-A---CT-----A-G-CAG.................--GCCTAGCT-G----TTC-A--CT--CA--T-G-GCA-AAAGGAC........................................... 5227
H2U.FR.96.12034           -ACAG......T-ATTGT-CT-C--CA-C-T--C-T-G-CT----GAC--CAG.................---CC----T------TTC-A-----CAG------GCA-G................................................. 5372
MAC.US.x.EMBL_3           -ACA-......C-GCTGT--T-C--CA-G-TC-CGAG--CT----GGT-CCAG.................---CC-AGC-----G-----A------AA--T-G---GCG................................................. 5279
SMM.US.x.H9               -A-A-......T-GCTGT--T-C--CA-G-TC-CGAA--CT-----G--TCAG.................---CC-AGC-----G--TC-A------A---T-G---GTC................................................. 5284
SMM.US.x.PGM53            -G-A-......T-GCTGT--T-C--CA-G-TC-CGAA--CT-----G--CCAG.................---CC-AGC-----G--TC-A------A---T-G-G-GTC................................................. 5728
SMM.US.x.SIVsmH635F_L3    -G-A-......T-GCTGT--T-C--CA-G-TC-CGAA--CT-----A--TCAG.................---CC-AGC-----G--TC-A------A---T-G---GTC................................................. 5814
SMM.SL.92.SL92B           -GCAG......--GTTAT-CTAC----GC--CGCG-T--CC---C--TCTAGT.................--CCA-AGC-----GCTG-----C---AAG-T-G--T--CAAA............................................. 5248
STM.US.x.STM              -A-AG......T-ACTGT--T-C--CAGG-TCACGAA--CT-----G--CCAG.................---CC-AGCT----G-----A--TT--AA--T-G-GGA-C............................................... 5458
MNE.US.x.MNE027           -ACA-......C-GCTGT--T-C--CAGG-TC-CGAG--CT-----G--CCAG.................---CC-AG------G-----A-----GAG--T-G---GTT................................................. 5279
DRL.x.x.FAO               -T-G-......A--CAAG--....--CT-T-TC-CTAG---G--C--G...............GGTCAG-----GAG---T--G--TC-T--CT-GCA---CCAC---AAGTACCTTAAGC...................................... 5085
MND_2.x.x.5440            -A-G-......C-GGTAG--.......--T-C-TC-CCTGG-------GA...............GGTCAG-----GA-------G-TT-----CT--CA---C-ATC-CAGAGGTAGAAAGC.................................... 5082
MND_2.CM.98.CM16          -G-G-......C-A-CAGTG....--CT-C-TC--CTGG--------A...............GGGAAG-----GAGC--G--G--------TT-GCT-AGT-AT-C-GCCTACTGTAACA..................................... 5519
MND_2.GA.x.M14            -A-G-......A-AGTAG-G....---T-CAGC-CCTGG--------A...............GGTCAG-----GA--------CT------TT--AG---C-AC--CAAGTTTTGCAGAA.................................... 5446
OLC.CI.97.97CI12          TGCC-......-A-CAT-GATAC--CC-GCTGG-GA----T---CTT...............GGG....--CCTG--GT----G-TGC----TT-GAG----G-C-G-A................................................ 5359
GRV.ET.x.GRI_677          -G-AG......TA--CGTGG....--CAC--TCA-G-A---------AGGTCAG.................---CA---A--G---CTT----------TT----AT-C-A................................................ 5682
VER.KE.x.TYO1_patent      ---G-......T--GT-TTC......--C-G-T--C--A---G-----A-GCACA.........GGACAG---CC----T-G--G---C-A--T--GCTC---CATCA-A............................................... 5745
VER.KE.x.9063             -G-G-......T-CGT-TTC......--C---TC-CC-AG--------A-CCACA.........GGAAAG---CCC--A--G--G----A--C--TCTG--CCATCAGA................................................ 5745
VER.KE.x.AGM155           ---GG......T-C-T-TTC......--C---T--C--G---------ACTAACA.........GGTCAG---CCC--CT-G-----T--A---T--CT---CCATCA-A............................................... 5740
VER.DE.x.AGM3             -G-GG......T-C-C-TTC......--C-G-TC-CC-AG-----C--G--AACA.........GGACAG---CCC---T-G--------A--T--CCTT---CACCA-A............................................... 5243
TAN.UG.x.TAN1             -A-GC......T--TTGTGG....--CAC---CA-G-A------CGTAGCAGAAAACCATTGGGGTCAG---C-G--C--G--G-TT--A--C--GA---T--AT-C-GACTTCCTAAGAA....................................... 5738
SAB.SN.x.SAB1C            -A-AG......TA-CTTTGG....----A--C-C--G-TG------C-GCCAACA.........GGGCAA---CC---AT----G--T--A-----TAGG-T--AT-C-A................................................ 5910
MND_1.GA.x.MNDGB1         -AGA-.......--GTTAAAAGTT---AG--T--C---T--C-----AGCA............CAAGGT--TCTC---T-G--G-TTC-CTGCT-GAG--TCA-CT.................................................. 5226
DEB.CM.99.CM40            -G--C......C-ACTG-AACA----C-GC-GT-CCATA-GGGACCAG-AGGT.........AAAGGACC-CCT--A-----G---A-A-----TAG--T-GC-G--AAAC.............................................. 5327
DEB.CM.99.CM5             -G-A-......C-ACTA-AACA----C-GC-AT-CCATA-GGGACCAG-AGGT.........AAAGGACC-CCC--AT----G-----A------AGGA--GC-G--AAAT.............................................. 5321
TAL.CM.01.8023            -GG--......C--CTTG-GGT-------GTCAG-CATAA-GG--GG..............GACTCGCC-CC---CT----G-T-C-A--CT--AAG-T-GC-G-TAGGC.............................................. 5015
TAL.CM.00.266             -GG-G......C-ACTAG-AGTT----G-GTCAG-CACAA-GG--GG..............GACACGCC-CCG--------G-T-C-A--CT-GA-----GC-G-TAGAT.............................................. 5517
SUN.GA.98.L14             -A-A-......A-ACTA-A-GA---CAG-A------T--CT--C--A..............GGCCTG--TCTC---T-G--G-TTC-CTGCT-GAG--TTC-CC.................................................. 5908
WRC.CI.97.97CI14          GT-A-ATAAAC--G-TAGA-GCA---C-G-----CAA-------GCA..............GGA...--GCTTA-CT-G--GAC----TGCT--CAG-TC--GCA-CATGAGCAACTCCAG-G-G--................................. 5826
WRC.CI.98.98CI04          GA-GGTTAGATA-A--AAACGCA---C-------CAA-------GCA..............GGA...--GCTTA-CT-G--GAC----TGCT--CAG-TC--GCA-CATGAGCAACTCCAGGG-A--................................. 5841
WRC.GM.05.Pbt_05GM_X02    GTTATTTACATCC--TAGA---C---A-------TCA-------CTA..............GGA...--GCTTA-CT-G--GGC---ATGTT-GCAG--C--GCA-CATGAGCAACTCAGG-AGAG-................................. 4758
LST.CD.88.524             -A-AG......A--CTG-AC--C---A--ATG-CTA-T--------A..............GGTTGT--TCTC--AT-G--G---C-CTGCT-GAG-CA-C--CA-C................................................. 4774
LST.CD.88.447             -A-AG......A-A-TGG-ATTC----CTATG-CCA-T--------A..............GGTTGT--TCTC--AT-G--G-T-C-CTGCT-GAG-CA-C--CA-C................................................. 4777
LST.KE.x.lho7             -A-G-......A-A-CTATGT-T--CC-G...-CTTT---------A..............GGTTGT--TCTC---T-G--G-TTA-CTGCT-GAG-CA-C--CA-C................................................. 5858
SYK.KE.x.SYK173           -AC--......T-G-CAT-ACAC--CTGGACC-C-CATACTG-CC-G.......................---CCC--A--T--G--T--A---T--CA--TG-ACC--AAAG.............................................. 5700
SYK.KE.x.KE51             -G-AG......T-G-CTT-GCAC--CTGGA--TTCCATAA-G-AC-G.......................---CTT--AT-G-----T-----TT--CA-AAG-ATT--AGCAAG............................................ 5357
COL.CM.x.CGU1             G-T--......CCCT-GAA-CCT-----T-TC--G-T------CCTA..............GGA...--GTTGAAGT-G---G-AC-ATGT--TCTG--CC---G-GAGAAA...GAGAGG--C-ACC-G---.............................. 5280
```

```
H1B.FR.83.HXB2              ACAGAGGATAG................................................................................................................................................ 5558
Vif                         _T__E__D__R#__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__
H1A1.UG.85.U455             -----------................................................................................................................................................ 5004
H1B.US.90.WEAU160           -----------................................................................................................................................................ 5557
H1C.ET.86.ETH2220           GT--------A................................................................................................................................................ 4950
H1D.CD.84.84ZR085           -----A-----................................................................................................................................................ 5080
H1F1.BE.93.VI850            GT------C--................................................................................................................................................ 4898
H1G.SE.93.SE6165            G----A-----................................................................................................................................................ 4955
H1H.CF.90.056               GT---------................................................................................................................................................ 4905
H1J.SE.93.SE7887            GT---------................................................................................................................................................ 4872
H1K.CM.96.MP535             GT---------................................................................................................................................................ 4754
H101_AE.TH.90.CM240         -----A-----................................................................................................................................................ 5129
H102_AG.NG.x.IBNG           G----A--C--................................................................................................................................................ 5083
H1N.CM.95.YBF30             -----A-----................................................................................................................................................ 5152
H1O.BE.87.ANT70             -----A-----................................................................................................................................................ 5613
H1O.CM.91.MVP5180           -----A-----................................................................................................................................................ 5588
CPZ.CD.90.ANT               -----A-----................................................................................................................................................ 5015
CPZ.CM.01.SIVcpzCAM13       -----------................................................................................................................................................ 5141
CPZ.CM.05.SIVcpzEK505       -----------................................................................................................................................................ 5130
CPZ.CM.05.SIVcpzMB66        G----------................................................................................................................................................ 5104
CPZ.CM.98.CAM3              -----A-----................................................................................................................................................ 4972
CPZ.GA.88.GAB2              G----A-----................................................................................................................................................ 4951
CPZ.CM.98.CAM5              G----A-----................................................................................................................................................ 5260
CPZ.CM.05.SIVcpzMT145       G----A-----................................................................................................................................................ 5100
CPZ.TZ.00.TAN1              --------C--................................................................................................................................................ 5220
CPZ.CM.05.SIVcpzLB7         G----A-----................................................................................................................................................ 5113
CPZ.US.85.CPZUS             -----A-----................................................................................................................................................ 5618
CPZ.GA.88.GAB1              -----------................................................................................................................................................ 5622
MON.CM.99.L1                ........................................................................................................................................................... 5389
MON.NG.x.NG1                ........................................................................................................................................................... 3999
GSN.CM.99.CN166             ........................................................................................................................................................... 5336
GSN.CM.99.CN71              ........................................................................................................................................................... 5318
MUS_1.CM.01.1085            ........................................................................................................................................................... 5292
MUS_1.CM.01.CM1239          ........................................................................................................................................................... 5314
MUS_2.CM.01.CM1246          ........................................................................................................................................................... 5431
MUS_2.CM.01.CM2500          ........................................................................................................................................................... 5384
RCM.NG.x.NG411              ...AT--CAGAG......GGCAGAGAAAGAGTGCCA......GAGGCCCCCACTGGGGCTGGAGATGTAGAGTTTGCCCCCTGGCTTCACAGAATGCTAACAGAAGTCAACTTAGAAGCC...AGGTTGCACTTTCATCCAGAGTTCATTTTCCGTCTGTGGCGTACTTG 5185
RCM.GA.x.GAB1               ...AT--CAGAA..............AGAGCCCCC......GAGGTCCCCACTGGGGCTGGAGAAGCAGAGTTCCAGCCTTGGCTACGGGACATGCTAGAAAAGGTCAACTTGGAAGCC...AGGTTACACTTTCATCCAGAGTTCATTTTCCGTCTGTGGCGCACTTG 5155
DEN.CD.x.CD1                -AGA....................................................................................................................................................... 5495
                              Vpx start
MAC.US.x.239                ...AT-TCAGAT......CCCAGGGAGAGAATCCCA......CCTGGAAACAGTGGAGAAGAGACAATAGGAGAGGCCTTCGAATGGCTAAACAGAACAGTAGAGGAGATAAACAGAGAGGCGGTAAACCACCTACCAAGGGAGCTAATTTTCCAGGTTTGGCAAAGGTC 5966
Vpx                            M  S  D  . . .  P  R  E  R  I  P  . . .  P  G  N  S  G  E  E  T  I  G  E  A  F  E  W  L  N  R  T  V  E  E  I  N  R  E  A  V  N  H  L  P  R  E  L  I  F  Q  V  W  Q  R  S
Vif                          _##D  V  R  . . .  S  Q  G  E  N  P  . . .  T  W  K  Q  W  R  R  D  N  R  R  G  L  R  M  A  K  Q  N  S  R  G  D  K  Q  R  G  G  K  P  P  T  K  G  A  N  F  P  G  L  A  K  V
H2A.GW.x.ALI                ...AT--CA-AC......CCCAGAGAAACAGTACCA......CCAGGAAACAGTGGCGAAGAGACTATCGGAGAGGCCTTCGAGTGGCTAGACAGGACAGTAGAGGCCTTAAACAGAGAGGCGGTGAATCACCTGCCCCGGGAGCTCATTTTCCAGGTGTGGCAAAGGTC 6045
H2A.DE.x.BEN                ...AT-ACAGAC......CCCAGAGAAAGGGTACCG......CCAGGAAACAGTGGAGAAGAGACCATTGGAGAGGCCTTCGAGTGGCTAGAGAGGACCATAGAAGCCTTAAACAGGGAGGCAGTGAACCATCTGCCCCGAGAGCTCATTTTCCAGGTGTGGCAAAGGTC 6052
H2A.SN.x.ST                 ...AT--CAG-C......CCCAGAGAGACAATACCA......CCAGGAAACAGTGGCGAAGAAACTATCGGAGAGGCCTTCGAGTGGCTAGACAGGACGGTAGAAGCCATAAACAGAGAGGCAGTGAACCACCTGCCCCGAGAGCTTATTTTCCAGGTGTGGCAAAGGTC 5497
H2B.GH.86.D205              ...AT-...GAT......CCCAGGGAGAGAGTGCCA......CCAGGAAACAGCGACGAAGAAACAGTAGGAGAAGCATTCGCTTGGCTAGAAAGAACAATAACAGAGCTCAACAGGGTAGCGGTCAACCATTTGCCCCGAGAACTTATTTTCCAGGTCTGGCAGAGGTC 6028
H2B.CI.x.EHO                ...AT-...GAT......CCCAGGGAGAGAGTACCA......CCAGGAAACAGCGGCGAAGAAACAGTAGGAGAGGCATTCGAATGGCTAGAGACAACATTAGAACATCTCAACAGAGTAGCAGTCAATCACTTGCCCAGGGAACTTATTTTCCAGGTCTGGCAGAAGTC 6026
H2G.CI.92.ABT96             ...AT-...GAT......CCAAGGGAGAGAGTMCCA......CCAGAAAACAGCGACGAGGAGACAATCGGAGAGGCATTCGAATGGCTAGAAAGAACAGTMACAGAGATCAACAGGATAGCAGTCAATCATTTACCCAGGGAGCTTATTTTCCAGGTTTGGCGAAGGTC 5379
H2U.FR.96.12034             ...AT--GAGAT......CCAAGAGAGAGAATCCCA......CCAGGAAATTCAGAAGAAGAAACTGTAGGAGAGGCTTTCAGTTGGCTAGAAAGAACTGTGAGAGATATCAACAGGGCAGCAGTGCAACATCTTCCCAGGGAACTTATTTTCCAGGTCTGGCAAAGGTC 5527
MAC.US.x.EMBL_3             ...AT-TCAGAT......CCCAGGGAGAGAATCCCA......CCTGGAAACAGTGGAGAAGAGACAATAGGAGAGGCCTTCGAATGGCTAAACAGAACAGTAGAGGAGATAAACAGAGAGGCGGTAAACCACCTACCAAGGGAGCTGATTTTCCAGGTTTGGCAAAGGTC 5434
SMM.US.x.H9                 ...AT-TCRSRT......CCCAGGGAGAGRATCCCA......CCTGGAAACAGTGGAGAAGAAACRRTAGGCGAGGCYTTCGACTGGCTACACAGAACASTAGAAGAAATAAACAGGGCAGCAGTGAATCATTTGCCRAGGGAGCTAATTTTCCAGGTCTGGCGAAGGKC 5439
SMM.US.x.PGM53              ...AT-TCAGAT......CCCAGAGAGAGGATCCCA......CCTGGAAACAGTGGAGAAGAAACAATAGGAGAGGCATTCGAATGGCTAAACAGAACAGTAGAAGAAATAAACAGGGCAGCAGTGAATCACTTGCCGAGGGAGCTAATTTTCCAGGTTTGGCGAAGGTC 5883
SMM.US.x.SIVsmH635F_L3      ...AT-TCAGAT......CCCAGAGAGAGGATCCCA......CCTGGAAACAGTGGAGAGGAAACAATAGAAGAGGCTTTCGAATGGCTAAACAGAACAGTAGAAGGAATAAACAGGGCAGCAGTAAATCACCTGCCGAGGGAGCTAATTTTCCAGGTTTGGCAAAGGTC 5969
SMM.SL.92.SL92B             ...AT-ACAGAC......CCAAGGGAAAGAATCCCA......CCAGGAAACAGTGGAGAAGAAACAATAGGAGAGGCCTTCGAATGGTTACACAATACAGTGGAGGCCCTGAATCAAACAGCAGTACAACACCTACCAAGAGAGTTAATTTTCCAGGTCTGGAGAAGGTG 5403
STM.US.x.STM                ...AT-TCAGAT......CCCAGAGAGAGAATACCG......CCCGGAAACAGTGGAGAAGAGGCAATAGAGGAAGCATTCGAGTGGCTACACAGAACGGTAGAGGACATAAACCGAGAGGCAGTAAACCATCTACCGAGGGAACTGATTTTCCAGGTCTGGCGAAGGTC 5613
MNE.US.x.MNE027             ...AT-TCAGAT......CCCAGAGAGAGAATCCCA......CCTGGAAACAGTGGAGAAGAGACAATAGGAGAAGCCTTCGAATGGCTAAACAGAACAGTAGAGGAGATAAACAGAGAGGCAGTAAACCACCTACCAAGGGAGCTGATTTTCCAGGTTTGGCAAAGGTC 5434
DRL.x.x.FAO                 ...AT--CAGAAAGACAGTCAGTGGAGAGAGCTCCAGCGGAGCCAATGGGAGCAGGAGAG.........GTAGAGTTAGAAGAATGGCTACAGAGGAGTCTCTTAAGAATCAACCAGGAGGCTCGATTACACTTCCACCCAGAGTTCCTCTTCCGTCTTTGGAACACCTG 5243
MND_2.x.x.5440              ...AT--CAGAG..............AGGGCACCAGAGGCGCCAGAAGGAGCAGGAGAG.........GTAGGACTGGAGCAATGGCTGGAAACGTCACTGGAGAGAATCAACCGGGAGGCCCGGTTACACTTCCACCCAGAGTTCCTTTTCCGTCTCTGGAACACATG 5225
MND_2.CM.98.CM16            ...AT--CAGAA..............GGGGCCCCAGAGATCCCAGAAGGAGCAGGAGAG.........GTGGATCTAAATACCTGGTTGGAAAGATCTCTGGAGAAAATCAATCAGGAGGCAAGATTACACTTCCACCCAGAGTTCCTTTTCCGTCTCTGGAACGCATG 5662
MND_2.GA.x.M14              ...AT--CAGAA..............AGAGCACCAGAGGCACCACAGGGGGCAGGAGAA.........GTGGGACTAGAACAGTGGCTGGAAAGATCATTGGAACAAGTGAACAGAGAGGCTCAATTACACTTCCACCCAGAGTTCCTTTTCCGTCTTTGGAACACCTG 5589
OLC.CI.97.97CI12            ........................................................................................................................................................... 5359
GRV.ET.x.GRI_677            ........................................................................................................................................................... 5682
VER.KE.x.TYO1_patent        ........................................................................................................................................................... 5745
VER.KE.x.9063               ........................................................................................................................................................... 5745
VER.KE.x.AGM155             ........................................................................................................................................................... 5740
VER.DE.x.AGM3               ........................................................................................................................................................... 5243
TAN.UG.x.TAN1               ........................................................................................................................................................... 5738
SAB.SN.x.SAB1C              ........................................................................................................................................................... 5910
MND_1.GA.x.MNDGB1           ........................................................................................................................................................... 5226
DEB.CM.99.CM40              ........................................................................................................................................................... 5327
DEB.CM.99.CM5               ........................................................................................................................................................... 5321
TAL.CM.01.8023              ........................................................................................................................................................... 5015
TAL.CM.00.266               ........................................................................................................................................................... 5517
SUN.GA.98.L14               ........................................................................................................................................................... 5908
WRC.CI.97.97CI14            ........................................................................................................................................................... 5826
WRC.CI.98.98CI04            ........................................................................................................................................................... 5841
WRC.GM.05.Pbt_05GM_X02      ........................................................................................................................................................... 4758
LST.CD.88.524               ........................................................................................................................................................... 4774
LST.CD.88.447               ........................................................................................................................................................... 4777
LST.KE.x.lho7               ........................................................................................................................................................... 5858
SYK.KE.x.SYK173             ........................................................................................................................................................... 5700
SYK.KE.x.KE51               ........................................................................................................................................................... 5357
COL.CM.x.CGU1               ........................................................................................................................................................... 5280
```

```
H1B.FR.83.HXB2            .......................................................................................................................................................................... 5558
Vif                       .__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__._
H1A1.UG.85.U455           .......................................................................................................................................................................... 5004
H1B.US.90.WEAU160         .......................................................................................................................................................................... 5557
H1C.ET.86.ETH2220         .......................................................................................................................................................................... 4950
H1D.CD.84.84ZR085         .......................................................................................................................................................................... 5080
H1F1.BE.93.VI850          .......................................................................................................................................................................... 4898
H1G.SE.93.SE6165          .......................................................................................................................................................................... 4955
H1H.CF.90.056             .......................................................................................................................................................................... 4905
H1J.SE.93.SE7887          .......................................................................................................................................................................... 4872
H1K.CM.96.MP535           .......................................................................................................................................................................... 4754
H101_AE.TH.90.CM240       .......................................................................................................................................................................... 5129
H102_AG.NG.x.IBNG         .......................................................................................................................................................................... 5083
H1N.CM.95.YBF30           .......................................................................................................................................................................... 5152
H1O.BE.87.ANT70           .......................................................................................................................................................................... 5613
H1O.CM.91.MVP5180         .......................................................................................................................................................................... 5588
CPZ.CD.90.ANT             .......................................................................................................................................................................... 5015
CPZ.CM.01.SIVcpzCAM13     .......................................................................................................................................................................... 5141
CPZ.CM.05.SIVcpzEK505     .......................................................................................................................................................................... 5130
CPZ.CM.05.SIVcpzMB66      .......................................................................................................................................................................... 5104
CPZ.CM.98.CAM3            .......................................................................................................................................................................... 4972
CPZ.GA.88.GAB2            .......................................................................................................................................................................... 4951
CPZ.CM.98.CAM5            .......................................................................................................................................................................... 5260
CPZ.CM.05.SIVcpzMT145     .......................................................................................................................................................................... 5100
CPZ.TZ.00.TAN1            .......................................................................................................................................................................... 5220
CPZ.CM.05.SIVcpzLB7       .......................................................................................................................................................................... 5113
CPZ.US.85.CPZUS           .......................................................................................................................................................................... 5618
CPZ.GA.88.GAB1            .......................................................................................................................................................................... 5622
MON.CM.99.L1              .......................................................................................................................................................................... 5389
MON.NG.x.NG1              .......................................................................................................................................................................... 3999
GSN.CM.99.CN166           .......................................................................................................................................................................... 5336
GSN.CM.99.CN71            .......................................................................................................................................................................... 5318
MUS_1.CM.01.1085          .......................................................................................................................................................................... 5292
MUS_1.CM.01.CM1239        .......................................................................................................................................................................... 5314
MUS_2.CM.01.CM1246        .......................................................................................................................................................................... 5431
MUS_2.CM.01.CM2500        .......................................................................................................................................................................... 5384
RCM.NG.x.NG411            TGTGGAACACTGGCATGATAGGCTTGGAAGAAGCCTTGAGTATGCAGGCTATAGATATCTGCTTCTGATGCAAAAAGCTCTGTTTATTCATTGCCAATCAGGGTGTTCTCAGAGA.........CATGGACAGGGACAAGCAAGGGAAGCAGGAGAAAGAATCCAGATTC 5346
RCM.GA.x.GAB1             TGTGGAACACTGGCATGACGTACATCAGAGAAGTTTAGAATATGCTGCTTATAGGTACTTACTGCTGATGCAAAAGGCTTTGTTTATCCACTGTCAGACAGGGTGTAGTCAGAGA.........CATGGACCCAATCCTAGG...GCAGTAGGAGAAAGGATAACCATCC 5313
DEN.CD.x.CD1              .......................................................................................................................................................................... 5495

                                                Vif end
MAC.US.x.239              TTGGGAATACTGGCATGATGAACAAGGGATGTCACCAAGCTATGTAAAATACAGATACTTGTGTTTAATACAAAAGGCTTTATTTATGCATTGCAAGAAAGGCTGTAGATGTCTAGGGGAAGGACATGGGGCAGGGGGATGG...AGACCAGGACCTCCTCCTCCTCCCC 6133
Vpx                       _W__E__Y__W__H__D__E__Q__G__M__S__P__S__Y__V__K__Y__R__Y__L__C__L__I__Q__K__A__L__F__M__H__C__K__K__G__C__R__C__L__G__E__G__H__G__A__G__G__W__.__R__P__G__P__P__P__P__P__
Vif                       _L__G__I__L__A__*
H2A.GW.x.ALI              CTGGAGATACTGGCATGATGAACAAGGAATGTCACAAAGTTACACAAAATATAGATACTTGTGCTTAATGCAGAAAGCTATGTACACACATTTCATGAAAGGGTGCACCTGCCTGGGGGGAGGGCATGGGCCAGGAGGATGG...AGATCAGGACCTCCCCCTCCTCCCC 6212
H2A.DE.x.BEN              CTGGAGATATTGGCATGATGAACAAGGGATGTCAGCAAGCTACACAAAGTATAGATATTTGTGCCTAATGCAAAAAGCTATATTTACACATTTCAAGAGAGGGTGCACTTGCTGGGGGGAGGACATGGGCCGGGAAGGATTGGAAGACCAAGGACCTCCCCCTCCTCCCC 6222
H2A.SN.x.ST               CTGGAGATACTGGCATGATGAACAAGGAATGTCAATAAGTTACACAAAGTATAGATATTTGTGCCTAATGCAGAAAGCTATGTTCATACATTCTAAGAGAGGGTGCACTTGCCTGGGGGGAGGACATGGGCCGGGAGGATGG...AGATCAGGACCTCCCCCTCCTCCCC 5664
H2B.GH.86.D205            TTGGGCATACTGGCGTGAGGAACAGGGCATGTCAATTAGCTATACCAAATATAGATACTTGTTGCTAATGCAGAAAGCAATGTTTGTGCACTATACAAAGGGCTGTAGGTGCCTGCAGGAGGGCCATGGGCCAGGGGGATGG...AGATCAGGACCTCCTCCTCCTCCTC 6195
H2B.CI.x.EHO              TTGGGCATATTGGCGTGAGGAACAGGGCATGTCAATTAGTTATACCAAATATAGGTATTTGTGTCTAATGCAAAAAGCAATGTTTATACACTTTGCAAAGGGTTGTGGATGTCTGCGGGAAGGCCATGGGCCAGGGGGATGG...AGATCAGGACCTCCTCCTCCTCCTC 6193
H2G.CI.92.ABT96           CTGGGAATATTGGCATGATGAACAGGGCATGTCAGCTAGTTACACTAAGTATAGATATTTGTTGCTGATGCAAAAGGCTATGTTTGTACATTGTAAGAAAGGATGCACATGCTTGCAGAAGGGCCATGGGCCTGGAGGATGG...AGACAAGGACCTCCTCCTCCTCCTC 5546
H2U.FR.96.12034           CTGGCAATACTGGCGTGATGAACAGGGCATGTCAGAAAGTTATACTAAGTATAGATATCTCTGTCTGATACAGAAAGCTTTATTTATGCATTGCAAAAAGGGATGTCAATGCTTGGGGGAAGGCCATGGACCAGGAGGATGG...AGATCAGGGCCTCCTCCTCCTCCCC 5694
MAC.US.x.EMBL_3           TTGGGAATACTGGCATGATGAACAAGGGATGTC......................................................................................................................................... 5467
SMM.US.x.H9               CTGGGAATACTGGCATGATGAAATGGGGATGTCAGTCAGCTACACTAAATRCAGATACTTGTGCTTGATACAGAAAGCTCTGTTTATGCATTGCAAGAAAGGGTGTAGGTGCTTAGGAGGAGAGCATGGGGCAGGGGGATGG...AGACCAGGGCCTCCTCCTCCTCCCC 5606
SMM.US.x.PGM53            TTGGGAATACTGGCGTGATGAAATGGGGATGTCAGAGAGCTACACAAAATACAGATACTTGTGCTTAATACAGAAAGCTCTGTTTGTGCATTGCAAGAGAGGGTGTAGGTGCTTAGGAGAAGAGCATGGGGCAGGGGGATGG...AGATCAGGGCCTCCTCCTCCTCCCC 6050
SMM.US.x.SIVsmH635F_L3    CTGGGAATACTGGCATGATGAAATGGGAATGTCAGAAAGCTACACAAAATATAGATACTTGTGCTTGATACAGAAAGCTCTGTTTATGCATTGCAAGAAAGGGTGTAGGTGCTTAGGAGAAGGGCATGGGGCAGGGGGATGG...AGAACAGGGCCTCCTCCTCCTCCCC 6136
SMM.SL.92.SL92B           TTGGGAATACTGGGTAGATGAGCAAGGATACTCACCAAGCTATGCAAAATATAGATATGTACAGTTAATGCAAAAAGCAATGTTTCAGCACTTCAGAAAGGGATGCACATGCAGAGGAGAAGGGCATTCACAGGGGGGCTGG...AGAACTGGACCACCACCTCCTCCCC 5570
STM.US.x.STM              TTGGGAATACTGGCATGATGAGCAAGGGATGCCCGGGAGCTATGTTAAATATAGATACTTGTGCTTGATACAGAAAGCTTTGTTTATGCATAGCAAGAGGGGATGCAGATGCTTGGGGGAAGGACATGGGGCAGGAGGATGG...AGACCAGGCCCTCCTCCTCCTCCCC 5780
MNE.US.x.MNE027           TTGGGAATACTGGCATGATGAACAAGGGATGTCACCAAGCTATGTAAAGTACAGATACTTGTGTTTAATACAAAAGGCTTTATTTATGCATTGCAAAAAAGGCTGTAGATGTCTAGGGGAAGGACATGGGGCAGGGGGATGG...AGACCAGGACCTCCTCCTCCTCCCC 5601
DRL.x.x.FAO               CATGGAGCACTACCATGATGCTCTTCAGTTATCTTTTACTTACAGCAAGTATAGATACCTACTTTTGTTACAGAAGGCCATGTTCATGCACTTTCAGCAAGGATGCTCATGTCTGCAGGGAAGGCATCCACCTCCCCTCAGACCAGCAGGAGATAGACTTCCTCCTCCTC 5413
MND_2.x.x.5440            TGTAGAACACTGGCATGATAGACATCAGAGATCTCTTGATTATGCCAAGTATAGATACCTGCTGTTGATGCATAAGGCCATGTATACTCACATGCAACAGGGATGCCCATGTAGAAATGGG...CGCCCAAGG..................GGA.........CCTCCTC 5365
MND_2.CM.98.CM16          CATAGAACACTGGCATGACAGACATCAGAGAAGCTTAAGCTATGCAAAGTATAGATACCTGCTGTTGATGAATAAAGCAATGTTTACTCACATGCAACAAGAATGTCCTTGTAGAAGTGGG...CATCCG...............AGA...GGA.........CCTCCTC 5802
MND_2.GA.x.M14            TGTAGAACACTGGCATGATAGACTTAGAAGAACCTTAAATTATGCCAAATATAGGTATCTTCTGCTGATGCAGAAAGCCATGTTTGTTCATATGCAACAAGGATGCCCATGTAGAAGTGGA...CATTCA...............AGA...GGA.........CCTCCTC 5729
OLC.CI.97.97CI12          .......................................................................................................................................................................... 5359
GRV.ET.x.GRI_677          .......................................................................................................................................................................... 5682
VER.KE.x.TYO1_patent      .......................................................................................................................................................................... 5745
VER.KE.x.9063             .......................................................................................................................................................................... 5745
VER.KE.x.AGM155           .......................................................................................................................................................................... 5740
VER.DE.x.AGM3             .......................................................................................................................................................................... 5243
TAN.UG.x.TAN1             .......................................................................................................................................................................... 5738
SAB.SN.x.SAB1C            .......................................................................................................................................................................... 5910
MND_1.GA.x.MNDGB1         .......................................................................................................................................................................... 5226
DEB.CM.99.CM40            .......................................................................................................................................................................... 5327
DEB.CM.99.CM5             .......................................................................................................................................................................... 5321
TAL.CM.01.8023            .......................................................................................................................................................................... 5015
TAL.CM.00.266             .......................................................................................................................................................................... 5517
SUN.GA.98.L14             .......................................................................................................................................................................... 5908
WRC.CI.97.97CI14          .......................................................................................................................................................................... 5826
WRC.CI.98.98CI04          .......................................................................................................................................................................... 5841
WRC.GM.05.Pbt_05GM_X02    .......................................................................................................................................................................... 4758
LST.CD.88.524             .......................................................................................................................................................................... 4774
LST.CD.88.447             .......................................................................................................................................................................... 4777
LST.KE.x.lho7             .......................................................................................................................................................................... 5858
SYK.KE.x.SYK173           .......................................................................................................................................................................... 5700
SYK.KE.x.KE51             .......................................................................................................................................................................... 5357
COL.CM.x.CGU1             .......................................................................................................................................................................... 5280
```

```
                                        Vpr start                                                                                  Vif end
H1B.FR.83.HXB2             ....................ATGGAACAAGCC.......................................................CCAGAAGACCAAGGGCCACAGAGGGAGCCACACAATGAATGGACACTAGAGCTTTTAGAGGAG 5633
Vif                                              W  N  K  P                                                             Q  K  T  K  G  H  R  G  S  H  T  M  N  G  H  *
Vpr                                             M  E  Q  A                                                            P  E  D  Q  G  P  Q  R  E  P  H  N  E  W  T  L  E  L  L  E  E
H1A1.UG.85.U455            ....................------------.......................................................-----------G-----------------GT--GC-----A-G--T-----T-G-----A--- 5079
H1B.US.90.WEAU160          ....................------------.......................................................------------------------------T-------------------------------A 5632
H1C.ET.86.ETH2220          ....................------------.......................................................-----------G-----G-------A---T-----------G--------------------- 5025
H1D.CD.84.84ZR085          ....................------------.......................................................--G-C-------------G-----------T--------------T----------------- 5155
H1F1.BE.93.VI850           ....................------------.......................................................----G------G-----------------T-----------G-------AA----------- 4973
H1G.SE.93.SE6165           ....................------------.......................................................-----------G-----------A--A--CT-----------G--T----A--G-----A--- 5030
H1H.CF.90.056              ....................------------.......................................................-----------G--------------------------------------------------- 4980
H1J.SE.93.SE7887           ....................------------.......................................................-----------C-----------A------T--C----------------------G------ 4947
H1K.CM.96.MP535            ....................------------.......................................................-----------G------------------A--------------T-----A----G------ 4829
H101_AE.TH.90.CM240        ....................------------.......................................................--------...-G-----------A--A--CT--------------T----A--G--------- 5201
H102_AG.NG.x.IBNG          ....................------AG----.......................................................-----------G------------------GTC------------------A--G-----A--- 5158
H1N.CM.95.YBF30            ....................-------G---A.......................................................--------TGC--------------A--CT-T----------G--------AT-A-----A--A 5227
H1O.BE.87.ANT70            ....................-----------A.......................................................--T--GA-T--G--A---GCT-AA-----CTT----------G----------C-----A--- 5688
H1O.CM.91.MVP5180          ....................------------.......................................................-TG---A-T--G--A---GCT---------TT----------------------CC-G--A--- 5663
CPZ.CD.90.ANT              ....................------------.......................................................--GC-G--TG-G-----------A--A---ATG-----G---TTGT----AAC------A--A 5090
CPZ.CM.01.SIVcpzCAM13      ....................--------C---.......................................................-----G-------------CC--A--A---T-------G---G--T----AA-A-----A--C 5216
CPZ.CM.05.SIVcpzEK505      ....................--------G--A.......................................................--------T-----A--------------T----C------G-TT----AT-G--G--A--C 5205
CPZ.CM.05.SIVcpzMB66       ....................---------CT-.......................................................-----------G--A-----A-----A---T-T-------------------A-A-----A--A 5179
CPZ.CM.98.CAM3             ....................--------G--A.......................................................--------T--G--T----CA--A------T-----------G--T----A--GC-G-----A 5047
CPZ.GA.88.GAB2             ....................------------.......................................................-----------G--A---GCAG-A--A--TT-------G---G-------AA-A--G-----C 5026
CPZ.CM.98.CAM5             ....................---------ATA.......................................................-----------G-------CA--A--A---T-----------R-------AA-A-----A--- 5335
CPZ.CM.05.SIVcpzMT145      ....................------------.......................................................-----G-----G--A---TCA--A-----CT-TG-GC-G---CTGT----CACA--G--A--A 5175
CPZ.TZ.00.TAN1             ....................-----------A.......................................................--GA-C---A-C-----G-----A--A---T---C---G---CTGT----CA-CC----A--A 5295
CPZ.CM.05.SIVcpzLB7        ....................----------T-.......................................................-------------------------------T-T-----G---G--T-----G-A-----A--A 5188
CPZ.US.85.CPZUS            ....................----------TG.......................................................-------------------GCC--A--A---T-------G---G--T----TG-C-----A--A 5693
CPZ.GA.88.GAB1             ....................-----------A.......................................................-----G--------T----CA--A--A--TT--C-G------G--T-----ACC--G--A--A 5697
MON.CM.99.L1               ....................----TT......................................................................CCCAT-G-A-GAC-GGCCCT--AG-CCGCCA-TTG---- 5434
MON.NG.x.NG1               .....................................................................................................CGT-GTTCCCAT-G-GCGC-CCGCTCTG-AG-CCGCCA--TG---C 4045
GSN.CM.99.CN166            ....................------AG-AT-CCACCCAGTCATCCT........................................ATGCCTTGG-TCA-C-G-AG-GT-CCCA--AC--TG-CTATCG-T-A-A-CGC-C-GTG---A 5426
GSN.CM.99.CN71             ....................------AG-A--CCACCAAGTCACCCT........................................-T-CCTTGGATCA-C-G-AG-GT-CCCA--AC--TG-C-GTTG-T-A-AGCGCCA-GTG---A 5408
MUS_1.CM.01.1085           ....................------AGCTTGCCCCCCAGCCACCCA........................................G-GCCCTGG--CTCCAG-GT-GTCCCAA-CAC--TGCGCCA-G---A---AGCCA-GTG---A 5382
MUS_1.CM.01.CM1239          ...........................................................................................-TGG--CTCCAG-GT-GTCCCAA-TAC--TGC-GCA-G-C-A-C--GCCA-GTG---C 5372
MUS_2.CM.01.CM1246          ...........................................................................................-TGG-T-A-T-GT--A-C-CCAA--ACT-TG-TTCA-G-G-A--CCGCCC-GTG---A 5489
MUS_2.CM.01.CM2500          ...........................................................................................-TGG-T-A-C-GC---GTACCAA--AC--TG-CGCA-G---A-A-CGCCA-GTG---T 5442
RCM.NG.x.NG411             TTCCGGGAATG...TAA..............................TCAAGCACTCCTCTCATTGCAGAGATGGAGATGCTC--G-----TG---AA-------A--A--CT-TG--------CT-A-G--TACCC-TATA--- 5459
RCM.GA.x.GAB1              TACCGGGGATG...TAA...............................TGATGGCCCTCTCTTTACAGAGATGGAGCTGCCC--T--G--TG----T--G--AC-A--A--TT-TG--------TT-A-G--TACCC--ATA--- 5424
DEN.CD.x.CD1               ....................----C-..............................................................................--AA-CACTC-AAGGCA-CT--AG--AGCAG--TG---T 5540
                                  Vpx end     Vpr start
MAC.US.x.239               CTCCAGGACTAGCATAA..---...........................................GAAGAAAGACCT------A-TG----A-----A-----A---TGGG--------GT-G-G---G--C-G--A--A 6228
Vpx                        P  P  G  L  A  *
Vpr                                             M                                            E  E  R  P  P  E  N  E  G  P  Q  R  E  P  W  D  E  W  V  V  E  V  L  E  E
H2A.GW.x.ALI               CTCCAGGTCTAGTCTA...---...................................ACTGAAGCACCAACAGAGTTTCCC--G-CC-GAATG--A--C--CCA--GAG-CAGGG---G---GT-A----AG-CC-GAGA--A 6318
H2A.DE.x.BEN               CTCCAGGTCTAGTCTA...---...................................ACTGAAGCACCAACAGAGTTTCCC--------TGGGACC----G------CTT-GGG-G---C---GT-A----AAC-C-GAG---A 6328
H2A.SN.x.ST                CTCCAGGTCTAGTCTA...---...................................ACTGAAGCACCAACAGAGTCTCCC--G--G--TAGGACC----C----------GGGG---G---GT-A----AACCC-GAGA--- 5770
H2B.GH.86.D205             CCCCAGGCCTGGCCTA...---...................................GCAGAAGCAGCCCCAGAGATCCCT-----GA--G-GAAC-----A--A--A--GTGGG-A--G----T-GGG---A-CC-G-----A 6301
H2B.CI.x.EHO               CCCCAGGTCTAGCCTA...---...................................GCAGAAGCAGTCCCAGAGATTCCT-----G--TA--AAC-----A--A--A--GTGGG-AC-G---GT-G-G--CG-CC-G-----A 6299
H2G.CI.92.ABT96            CTCCAGGTCTAGCATAG..---..............................................GCAGAAGAAATCCCC--------TG-G-CC--C-----A------TGGG--------GT-G----AG--C-G--A--A 5644
H2U.FR.96.12034            CACCAGGATTGATCTA...---..................................................GCAGAAATACCT-----G---GGG-CC-----A--A-----TGGGG-C--G---GTGAG---AA-AC-G-----A 5788
MAC.US.x.EMBL_3            ...................................................................................................................................................... 5467
SMM.US.x.H9                CTCCAGGACTAGCATAA..A---..................................................ACAGAAAGACCT--------TG---CC-----------A--GTGGG--------GTRG----AG--C-G-----A 5702
SMM.US.x.PGM53             CTCCAGGACTAGCATAAA.A---..................................................ACAGAAAGACCT--------TG---CC-----------A--GTGGG--------GT-G--A-AG--C-G-----A 6147
SMM.US.x.SIVsmH635F_L3     CTCCAGGACTAGCATAAA..---..................................................GCAGAAAGACCT--------TG---CC-----------A---TGGG--------GT-G-G--AG--C-G-----A 6232
SMM.SL.92.SL92B            CACCAGGTTTGGCCTA...---.....................................................GAACATGCT-----G--TG-GACCAAC-CA--A--A---TGGG---G----T-AG---TG-AC-G--A--A 5661
STM.US.x.STM               CTCCAGGCTTAGCATAG..A---..................................................ACACACAGACCT-----G--TG----C--------A--A---TGGG---G---GT-G-G--AG--C-G-----A 5876
MNE.US.x.MNE027            CTCCAGGACTAGCATAA..---...................................................GAAGAAAGACCT--------TG----C-----A-----A---TGGG--------GT-G-G---G--C-G-----A 5696
DRL.x.x.FAO                CTCCTCCTCCA...TGA.......................TTGGTCTAATAATTGTGTCTTACAGATGGAAAGGCCA--G--G--TG----A--G-CA-----A--CT-TG----G---CT-G----TAC-----TA--A 5527
MND_2.x.x.5440             CTCCAGGGATGGCTTAA..............................TTTCTCTGTTTTACAGATGGAACAGCCA--T--G--TG-G-CT--C-----A-GA---T-T---A-----CTGG---GTAC---G-CA--A 5473
MND_2.CM.98.CM16           CTCCAGGGATGGTCTAA................................CCTCTGTGTTACAGATGGAGCAACCA--C--G---G-G-CC--T--A--A--A---T-----------CT-A-T--CACC----CA--A 5908
MND_2.GA.x.M14             CTCCAGGATTGGCCTAA................................CTTCTGTCTTGCAGATGGAACAGCCG----G--TG-G-CT-----A--A--A---T-T-----G---CTGA----TACA----CA--A 5835
OLC.CI.97.97CI12           ....................----C-...............................................GAAAGAGTT-----G--T--G-------CC----------TGG-G--T---GTGAG---CACAA-G--A--A 5437
GRV.ET.x.GRI_677           ....................----C-TC--GAAGAGATCCAAGAGAA....................CCTTTACCAGGATGG-TG---AT-TGG-AT-T-G-C---------TGGG-C------CT--A---CA-GC-CAG---T 5787
VER.KE.x.TYO1_patent       ....................----CCTC--GGAGA.......................GATCCAAGAGAGGCAAGACCAGGAGAAGT----ATTTGG-AT-T-AGC---------GTGGG-C------CT-AG---TA-G---C-A--T 5850
VER.KE.x.9063              ....................----C-TC--GGAGA.......................GGTCCCAGAGAAAACAGGCCAGGAGAAGT---GAT-TGG-AT-T-A-C--A--A---TGGG-C------CT-AGG--CA-G-----A--T 5850
VER.KE.x.AGM155            ....................----CCTC--GAAGA.......................GATCCCAGAGAGGAGAGACCAGGAGGA-T----ATTTGG-AT-TCAGC---------GTGGG-C------CTCAG---TA-GG----A--A 5845
VER.DE.x.AGM3              ....................----CCTC--G-AGA.......................GATCCCAGAGAAGCAAGACCGGGGGAA-T----ATATGG-ATTTGAGC---------GTGGG-C------CT-AG---CA-GC----A--T 5348
TAN.UG.x.TAN1              ....................----C-G---GAAGA.......................GATTCCAGGGAAAGAAGACCCGGATGGTT----ATTTGG-AT-TCAGC---------GTGGG-C-----ATT-AG---CA-GG---CA--A 5843
SAB.SN.x.SAB1C             ....................----CCTC--GAGGGTGGCTCCCACCAGTAGGAGGGGATCCTCCCAAGGATCCCCCCAAGAATCCCAGAGAAGAGATACCAGGATGG-T----ACATGG-ATTTG-CC--A-----GTT-G-C------CTC-G---CA-G---C----C 6060
MND_1.GA.x.MNDGB1          ....................---...................................GGCCAGAAGAGAGATGAGCAAGTAT-------T-----A--T-CC--A------T-----C-G---CT-GC---TAC-A-G-----A 5316
DEB.CM.99.CM40             ....................------AGGTA-CCACCCTCACACCCA.........................................---C-TTT-ACCTCAAG-ACA-TACCCATGACA-GGCT-G--TTG-AGC--GCCA-GC-A--T 5417
DEB.CM.99.CM5              ....................------AGGTATCCACCCTCACACCCA.........................................---C-TTT-AC-TCAAG-ACAGTACCAATGACA-GGCT-GCACT--A-C-AGCCA-GC----T 5411
TAL.CM.01.8023             ....................------AGGTTG...............................CCACCCAGTCATCCT-TGCC-ATAAC-TCAAGG-TC-TACCAA---CA-GA-CTCAATT-GA-AGA--CA-GAGA--A 5105
TAL.CM.00.266              ....................------AGGTTG...............................CCACCCAGTCATCCT-T-CCGCTGACCTCCAG--TC-TTCCAA---CA-GGCCTCA-CT-GA-AG---CA-GAGA--A 5607
SUN.GA.98.L14              ....................----CCAGC.........................AGGAGAGAGCCAGTAGAACAGCCA---------G-G-------C---A------TTTGCAC-----CTGGC---TACCA-G--A--A 6004
WRC.CI.97.97CI14           ...................................................................................................GGCGCCTT--GGAG-ACCT----CA-T-CC-GA-GAC--TCTAGG-A-GC-A--- 5881
WRC.CI.98.98CI04           ...................................................................................................AGC-CCTA--GGAG-ACCT----CA-C-CC-GA-GAC--TCTAGG-A-GC-A--A 5896
WRC.GM.05.Pbt_05GM_X02     ................................................................................................-AGG-CCCCAA--GGAG-ACCC-AGGCA-TACC-GA-GAT--TCTAGG-A-GC-A--- 4816
LST.CD.88.524              ...................---TCCAG-CAAAGA.........................CAGCAGCGCAGCCCAGAAAGGCCC-----G---TTT--T--T-CA--A------TGGGC-C-----CT-GC---CACAA-G-----A 4876
LST.CD.88.447              ...................---TCCAG-A-AAGG.........................CAGCAAAGGAGTCCAGAAAAACCT-----G---TTT-------CA---------TGGGG-C-----CT-GC---CACCA-G-----A 4879
LST.KE.x.lho7              ...................---TCCAGGCAGAGG.........................CAGCAAAGAAGTCCAGAGAGACCA-----G---TTT--A----CA---------TGGGG-C-G---TT-GC---CACCA-G-----A 5960
SYK.KE.x.SYK173            ....................----CGG-G--TTTC..................TTCAATCCCTCCCAGCATGTGCAAGGAACACCATGGTTCTTCATTCCAAG-AATGT-G--TT-ACC---A--GTGATCAATGTCACT-TAAAGGC---ACT- 5817
SYK.KE.x.KE51              ...................G----C---G--TTTC..............TTCAGAGTTTACCAGCGGCAGCCAGGGGAACCACTATTTTACATTCC-AGAA-T-TCAATTG-GAC---G--GGAGTAAT-A-TG-CACT-CCAAAGC---AC-- 5478
COL.CM.x.CGU1              ...................---AG-GTCCTG...........................................................................A-AGGGACTTCAC-.---TTG-AG--C-AGG-----A-C 5330
```

```
                                                                                                                                           frameshift insert in HXB2
H1B.FR.83.HXB2             CTTAAGAATGAAGCTGTTAGACATTTTCCTAGGATTTGGCTCCATGGCTTAGGGCAACATATCTATGAAACTTATGGGGAT.......................ACTTGGGCAGGAGTGGAAGCCATAATAAGAATTCTGCAACAACTGCTGTTTATCCAT..TTTCAG 5777
Vpr (frameshifted)                                                                                                                                                                          F  R
Vpr                        _L__K__N__E__A__V__R__H__F__P__R__I__W__L__H__G__L__G__Q__H__I__Y__E__T__Y__G__D__.__.__.__.__.__.__.__.__T__W__A__G__V__E__A__I__I__R__I__L__Q__Q__L__L__F__I__H_##_F__Q_
H1A1.UG.85.U455            ------C-----A------------------CAG----------A-----A-----------A-C--C---------.......................--------AG----T-----T-----------T--------T-A--------T---..------ 5222
H1B.US.90.WEAU160          -------G-----------------------G-A---------A-------A-------------------------.......................-------A-----------------------C------------------T---..------ 5775
H1C.ET.86.ETH2220          --C---C-G--------C-----C--------ACCA---------AA------A---T-----------C---------.......................-------T-G-----A-----TT---------C-----------AA-------------..------ 5168
H1D.CD.84.84ZR085          -------G---------C------------------A-----T---A-------A---T--------------------.......................---C-----------T-----T-----------------------A--------T---..------ 5298
H1F1.BE.93.VI850           --------------------G------------CCA-----T--C--------C-----------A-C--C--------C.......................--------AG-----T-----T--------G--AT----------A-------------..------ 5116
H1G.SE.93.SE6165           -----A---------------------------C--------T-----------A--G---------A-C-----------.......................--------A------T-----------------A--A--------A-------------..------ 5173
H1H.CF.90.056              A----------G--------G-----------AG-A---------CAA-----A--G---------A-C--C-----A---.......................--------T------T-----TT---------CG-----------A--------T---..------ 5123
H1J.SE.93.SE7887           -----A-----------------C---------CCA-----T---A-T-----A--GT--------AGC--------A---.......................--A----A------C------------------A-----------AT-------T---..------ 5090
H1K.CM.96.MP535            -------GA------------------------CCA----------AATC-G--A--G---------AC------------.......................--------A----C-T-----T-----------AT----------A--------T---..------ 4972
H101_AE.TH.90.CM240        -----A---------------------------CCC---------A---------A--GT-C------A-C-A--------.......................--------A---G--T-----T------------GT-----------A-------G-T---..------ 5344
H102_AG.NG.x.IBNG          -----AC----------------C---------CCG------------A------A------------------A------.......................--------A---G--TA----T------------T-----------AT-----G-T---..------ 5301
H1N.CM.95.YBF30            T-A--A------------GC-C---------A-----------A-----G-----A------C-------A-C--A-----A--C.......................--C----AG--G--A--G--A--T--C--G--A--A------T-A---------------...-AT-- 5370
H1O.BE.87.ANT70            --A--AGCA-----A--A--------C------CC------A--C-C---G--A---T-C--T-----G-----------C.......................--------T------TATG--A--T--------CT-A------T---------C------...-AT-- 5831
H1O.CM.91.MVP5180          --G--AG-A-----A--A--------C------CC----T-A--A-C--GT-----GT-C--T-----G--------A--C.......................--------A------TATG--A--T--------CT-A---------A-------C------...-AT-- 5806
CPZ.CD.90.ANT              A-A--A-----G--A--A--------C--ACA-CC-ACAT-A--ACA-C-T--AA-TTGGG-A----C--A---C--A---.......................T-A---AG-------C-G-TTT-G----TTC-C--T---A--GCT--------CA------...-AT-- 5233
CPZ.CM.01.SIVcpzCAM13      T-A---------------A-A---C-A---C--A-G------T--CCAA-----C--G-----T--ACC--C---------C.......................--G---A---G-----G--A--T---------C--A---------A-------G-T------...--T-- 5359
CPZ.CM.05.SIVcpzEK505      --A--------G---C-GC-C--C------C--CC------A------AC--------T-CT-----A-T--A-------A---.......................--C---AG----A--G-----C--T--G-CA--A---------T--------A-------...-AT-- 5348
CPZ.CM.05.SIVcpzMB66       -----A-GG-----A--A-----C--C---C-AGA-----------CAGC----A--G---------ACC-----------A---.......................--------------G-----G--T-------------C-----------T--------T------...-A--- 5322
CPZ.CM.98.CAM3             --A---------------AC-C---------C-AGCA-----A------A-----A--GTT---T-----C----------A---.......................--A---------T--A-----A--C--T-----A---------G--T---------------...--T-- 5190
CPZ.GA.88.GAB2             T-A--A--C--G-----C---------------ACCA-----A---CAG-----A---TTA-----CA-T-A--------A---.......................--A-----A-------A--G--------C-----A--T-----CT-----------G-G----...--T-- 5169
CPZ.CM.98.CAM5             --A--A-C---G---C-A--------------C-CCC------A---A-AC----A--GT-C-----------------A---.......................--A---AT---G--A------A--T--T------AT------T--T---------G------...--T-- 5478
CPZ.CM.05.SIVcpzMT145      A----AC-A-----A---------C------C--ACCC---T-A----CAC----AG-------T---------A-----A--C.......................T-C------G------AC----T-----------------T-------T-A-------G------.------ 5318
CPZ.TZ.00.TAN1             A-A--AC-A-----A--G-A---C------A--ACCAATAT-A--G--GG----AA-TTGGG---TCACC-T-------A--C.......................T-C----AG-----AC-G-AAT----C-AG--CT----GAG-GCTT-------C---C...-ATC- 5438
CPZ.CM.05.SIVcpzLB7        --C--AC-G-----A--A-----C---------ACCA---------CCAA--G--A---T----A------C--C------A--C.......................--C----AG-----TC----T--C----------C------------C---------T------...-A--- 5331
CPZ.US.85.CPZUS            --A--AG-A--G---T-GC-C------------ACCC-----A--A---------A--GT-C------CAGT---------A--C.......................--------AG-----A-----A--T-----G--CT-A--------TT-------------...-AT-- 5836
CPZ.GA.88.GAB1             T-A--A--------A--A--------C------ACCA---T-G---CAA------A---TTC--T------C---------A---.......................--A----T------A--G-----C---------CT-A------CT------------------...--T-- 5840
MON.CM.99.L1               --A--CG-A-----CC-A-A---C--CT-A---GAAGA-----G-----A-TT--G-----GG-AACG--GCTGCC--CT--CCCGGCCTGGAATGCAGACCAGGCATGGGCA---T-T-CCATT-ATTAT-CT---TGGG-C---ACCA-C--T-A--GG--C...-A--- 5601
MON.NG.x.NG1               --A--CG-G--G--CC--TYG--C--CT-A-A-GAAGA-----G-----A-CT--G-T--GG--AC----TTGCC--CC--CCCAGGGTGGAACGCAGACCAGGCATGGGCA---T-Y-CTATT-ATTAC-CC--GCGGG-----AC-A-RA-A----GG--C...A-G-- 4212
GSN.CM.99.CN166            A-A--CG-G--G--A-AG-A---C--CT-A---GACGA-----G---AA-TT----TG--G-GACG--GCTCCC--CC---CCAGATTGGACAGTGGATCAGGCAG-AATA---T-T-CAAT--ATTAC--C...-GGG-----AC---T--C----GA--C...--T-A 5590
GSN.CM.99.CN71             A-A--CG-A--G--A-AA-A---C--CT-A---GAAGA-----G---AA-TT----TG--G-AACG--GCTCCC--CC---CCGGATTGGACAGTGGATCAGGCAG-AATA---T-T-CAAT--ATTATG-C--G-GAG----GAC-T-A--C----GA--C...--T-- 5575
MUS_1.CM.01.1085           G-A--CG-A--G--A-AA-A---C---T-C--AGAAGA-----G---AA-TT--A-TG-GG-AACG--GCTGCCA-CC---CCTGATTGGACAGTGGATCAGGCGG-AATA---T-T-CCAT--A-TATG-T--G-GAG-----AC---TG-C----GG-----...------ 5549
MUS_1.CM.01.CM1239         --A--CG-A--G--A-AA-A---C---T-C--AGAAGA-----G---AA-CT--A-TG--G-AACG--GCTACCA-CC---CCTAATTGGACAGTGGATCAGGCTG-AATA---T-T-CCAT--ATTAC--T--G-GAACA---AC---C--T--C-GG-----...-ATC- 5539
MUS_2.CM.01.CM1246         --A--CG-A--G--A-AA-A---C--CT-C--AGACAA-----G---AA-TT--G-TAGAG--ACA--GCTCCCA-CA---CCACAATGGTCTGTAGATCAGGCAG-CATAA--T-T-CCATT-ATTAC--T-A-TCAG-C---ACC--CT----C-GA--C...-AT-- 5656
MUS_2.CM.01.CM2500         --A--CG-A--G--A-AA-AG--C--CT-C---GACGA-----G---GA-TT--G-C-GAG--ACG--GCTCCCA-CA--CCCTAATTGGACTGTTGATCAGGCAG--ATA---T-T-CCAT--ATTAC--T---TGGG-A--GGGT--C--C--C-GG-----...---C- 5609
RCM.NG.x.NG411             A-AC-AG-A------AAG-A-------A-ACCTGAA-T---TACACAAGA---AA-TT----A-----GCAGC----A---.......................-G--TA-AG-----AA----T--G----CCT-G--TA-TA--GCA--T--CT-G--C...------ 5602
RCM.GA.x.GAB1              T-GC-AG-A------AAG-A-------A-ATATGC--T---AACGCAAA----AG-TT--G-A-----GCAAC--------.......................-GCATA-AG------C-G--A--G--TC-GC-G--T---AG-GCCT-----C-T-------...--T-- 5567
DEN.CD.x.CD1               A-C-CAG-G--G--AAGA-A---C--CAGC-AAGAAGA-A--AC---AA--T--G-T--CTG--GGTC-TTACCA-C-CTCCCACATTGGACAGAAGGACAAGTC-TGGCA--C-CC--CAT--ATT-C---------CA----GA--GAAA-A-GG-AG------...-AT-- 5707
                                                                                     Tat exon 1 start
MAC.US.x.239               --G--AG-A-----T-A-A-------GA-CCTCGC-T---AAC--CAC-T--TA-T---------A-T-GAC----A--C.......................--CCTT-AG----CA-G--AAC-C--T-----C--C----G-GC---C--C--G-----------..------ 6371
Vpr                        _L__K__E__E__A__L__K__H__F__D__P__R__L__L__T__A__L__G__N__H__I__Y__N__R__H__G__D__.__.__.__.__.__.__.__.__T__L__E__G__A__G__E__L__I__R__I__L__Q__R__A__L__F__M__H__.__F__R
Tat exon 1                                                                                    M__E__T__.__.__.__.__.__.__.__.__.__P__L__R__E__Q__E__N__S__L__E__S__S__N__E__R__S__S__C__I__.__S_
H2A.GW.x.ALI               A-A--AG-A------T-A--G------GACCCTCGCAT---AAT--CTC-T--CGGCT---------ACT-GAC----A--C.......................--CCTT-A-A---CCAGG-AGC-C--T-ATGCC--A----G-GCC--C--C--G--C..------ 6461
H2A.DE.x.BEN               A-A---G-A-----CT-A---------GA-CCCCGC-T---AAT--CTC-T--CT-CT-----C--A-T-GAC----A--C.......................--CCTT-A---C-CCAG--AGC-C--T-A--CC--A----G-GCC--C--CG-G--C..------ 6471
H2A.SN.x.ST                A-A--AT-A------T-A-AG--C---GACCCTCGC-T---AAT-ACTC-T--CA-CT---------CT-GAC----A--C.......................--CCTT-A---C-CCAG--GGC-C--T--G--C--A----G-GCC--CC-CT-G--C..------ 5913
H2B.GH.86.D205             A-A---C-A-----CT-A-AG------GA-CCTCGC-T---AAC--CGC-T--TA-CTT------CAGT-GGC----A---.......................--CCTT-------CA-G--AGC-C--T-A---C--C----G-GCC--C--CC----C..------ 6444
H2B.CI.x.EHO               A-A--AC-A-----CT-A-AG------GACCCCCGC-TAT-AAC--CAC-T--CA-TTT------CA-T-GGC----AA--.......................--CCTT-A-----CA-G--AGC-C--T-A-C-C--C--G-GGGCA--C--CC----C..------ 6442
H2G.CI.92.ABT96            A-C--AG-A-----CT-A-A-------GACCCTCGC-T---AAC--CGC-T--TA-CT----------T-GGC----A--C.......................--CCTC-AG----CA-G--AAC-C--T-A---C-----GAG-GCT--C--CC-G--C..--TC- 5787
H2U.FR.96.12034            --A--AG-A--------A--G------GA-CCTCGC-T---AAC--CGC-T--TA-CT---------A-T-GGC----A---.......................--CCTT-A-----CA-G--AGC-C--T-A---C------AG-GCAA-C--CC----C..------ 5931
MAC.US.x.EMBL_3            .....................................................................................................................................--C----G-GC---C--C--G------..--T-- 5496
SMM.US.x.H9                -KA--AG-A-----CC-G-AC--K---GA-CCTCGC-T---AATY-CGC-T--TA-CT----------T-GAC----A--C.......................--CCTT-A-----CA-G--AGC-C--T-----C--C--GGG-GCA--C--C-----------..------ 5845
SMM.US.x.PGM53             G-A--AG-A------C-G-A---C---GA-CCTCGC-T---AAC--CGC-T--TA-CT----------T-GAC----A--C.......................--CCTT-A-----CA-G--AAC-C--T--G--C--C----G-GC---C--C--A--C..------ 6290
SMM.US.x.SIVsmH635F_L3     A-A--AG-A-----CC-G-A-------GA-CCTCGC-T---AAC--CGC-T--TA-CT----T-----T-GAC----A--C.......................--CCTT-A-----CA-G--AAC-C--T-----C--A--GAG-GCA--C--C-----------..------ 6375
SMM.SL.92.SL92B            T-G--AG-A--G--AC-G-A-------GACCCTCGC-T---TAC--CT-----AA-TT-CG-A-----T-----C-------.......................--AATA-A---G-CA-GG-AA-----T-AG--AT-A---A-GGCAT----CT-A--C..--T-- 5804
STM.US.x.STM               A-A--AC-A-----CC-GC-G------GA-CCTCGC-T---AAGC-CGC-T--TA-CT----------T-GAC----A--C.......................--CCTT-A-----CA-G--AGC-C--T-AG--C--C--G-G-GCC--C--C--G--C..------ 6019
MNE.US.x.MNE027            --G--AG-A------T-A-A-------GA-CCTCGC-T---AAC--CGC-T--TA-TT----------T-GAC----A--C.......................--CCTT-AG----CA-G--AAC-C--T-A---C--C----GGGC---C--C--G--C..------ 5839
DRL.x.x.FAO                A-AC-AG-G------T-G-AG------GA----CGC-TAT-G---AATG-----TC-TGGG---------GA-----A---.......................T-A-TA-A-------C-CCAGC-T----T-C-CT-----AG-GCA--T--CT-G-----..--T-- 5670
MND_2.x.x.5440             A-AC-AG-G------T-A-AG-----CGA----CGC-T---A----CAG----CTC-TGGG-G-----GCAAC-A--A--C.......................-----TA-A---T--CC--CAGC-------C---T-----AG-GCA-----CT-A--C..--T-- 5616
MND_2.CM.98.CM16           G--C-AG-A------T-A-AG------GA---ACGC-TAT-G----CAG----AAGCTGG-----CA----C---------.......................--A-TA-A---G---CG-CAGC-T--CTC-------T---A--GCT----C---G--C...-AT-- 6051
MND_2.GA.x.M14             A--C-AG-A------T-G-AG------GA----CGC-TAT-G----CAG----AAGCTGG--T---------A--------C.......................--C-T--AG--G---C-GAAGC-T--C-CCC-CT-A--GAG-GCT--A---C-G-----..--T-- 5978
OLC.CI.97.97CI12           A---TTGTA--G--AAAA--G--C-----C---GAG-TAT-GGTCC-AG-TTTT-T-TTCTGTGC-C---GA.....................CATATCAATGATGAGAT-AGTACCTCTTTGCTGT-C--T--GC-A--A---AG-GCCT-----CAA------...-AT-- 5583
GRV.ET.x.GRI_677           --A--CG-A-----CAGA--G--C---GGA-T--ACAT---AATCC-AG--T--A-TT-CTGTGTA--GGAGGGAA--AGA............CATAATACCCCATGGAAT-AGAT--GCT-CAAGTACTAT------G-T---A-GTCTA-----G-A-----..------ 5942
VER.KE.x.TYO1_patent       --C--CC-G--G--GAGGCTG------GGG--AGAACTC--GTTCCAAG--T--A-CT-CTGTC-G--GGAAGGA-AAAGA............CATGGTACTCCC-TGAT--A-A-G-CCT-CAAATATTAT--GC-AG-A---A-GGCT--C---G-G-----..--TC- 6005
VER.KE.x.9063              --C--TC-G-----CAAATT-------GGC--AGAGCTC---TTCCAAG-TT--A-TT--TGTC-G--GGAAGGA-A-AGA............CGGGGAGCCCCA-TGAC--AGA-G-CCT-TAGATACTAT--GT-GG-T--GA--GCA--C---G-T-----..---C- 6005
VER.KE.x.AGM155            A-C--C--C-----CAAACTG------GGCC-AGAACTC--AT-CCAAG--T--A-TT--TGTC-G--GGAAGGG-A-AGA............CAGGGAAGACCC-TAGC--A-A-G-CAT-TAAGTATTATC-CT-AG-T--GA--GCT--C---G-G-----..---C- 6000
VER.DE.x.AGM3              A-C--TC-G-----CAAG-TG------GGGC-CGAGCTC--GTTCCAAG--T--A-CT--TGTC-G--GGAGGGA-AAAGG............AATCGCACTCCC-TGCTA-A-A-G-CTT-TAAATATTAT-A-T-GG-----A--GCT--C---G-G-----..---C- 5503
TAN.UG.x.TAN1              --C--TC-G-----ACAAC-C------GGG--AGAACTC---TTCCAAG--T--A-CTTCTGTC-G---GAAGGG-AAAGG............AATGGGGCACCA-TGATA-A-A-G-CTT-TAGATATTAT--GT-GG-A--GA-GGCTT-A---G-G-----..--TC- 5998
SAB.SN.x.SAB1C             --C--TTCA--G--CCAGT-C--C--C--A----A-CTC--TTTCC-TC-TT--TGGA----TGTG---GAACCA-CCAT-.........GATCATGGGCAAACC-GGCT--A---CTG-T-TAAATATTGT------------A-GGCTT-A---G-G----...A-G-A 6218
MND_1.GA.x.MNDGB1           A-A---G-A-----AAGA-AG--C--C---CTC---ATC--AA---CAG--TCAG--T--TGTGTGC---ACACA---AG-GAG.....................GAAGA---CT-T-A-A--TTT--T-CC...T-AA--A-TAG-GCCA-T-GGG----C...C-AGC 5459
DEB.CM.99.CM40             --A--CG-G-----CC-A-A------CA-C--AGAAGAA--ATT---AG-TT--A-T---TGTGTA--CTTACCC-CCC-ACCAGATTGGACGGGAGAACAAGCATGGGCT---TCG--CATT-A-TAC----A---AG-T---A--A--T-A-GGT-G--C...C-G-- 5584
DEB.CM.99.CM5              --A--CG-G------T-A-A---C--CA-A--AGAAGA-T-ATGG--AG-CT--A-T--CTGTGTA--CCT-CC--CCC-ACCAGATTGGACAGGAGAACAAGCATGGGC---GTC---AATT-A-TAT----A---AG-A---AG-A----T-GGC-T-----...C-G-- 5578
TAL.CM.01.8023             G-A-G-G-C--GATGC-C-T---C--CA-A--AGAAGA-G--AT---AG-TT--A-T--CTGTGTA--GTTACCA-CA--GCCTGATTGGACAGGCGAACAAGCATGGGCT--TTC---CAT--ATT-TGCTC-G-AGGGTA-CG-GA--T-A--CC-A-----..----CA 5272
TAL.CM.00.266              G--CGAG-C---ATGC---T---C--CA-A--AGACGA-G--AT---AG-CT--A-T---TGTGTA---TTGCCA-CA--CCCTGATTGGACAGGAGAACAGGCATGGGCT--CTC---CAT--ATT-TGCT--G-AGGGTA-T---A--T-AC-CC-A-----..----CA 5774
SUN.GA.98.L14              A-C-G-G-C-----G-AA--G-----C--AGCAGCAATAG-TGCCCAAG-CTT----T-CTGTGCG.......................GACTCTACAGGGTCAGAAAA--A---GTGTATGAGAGC---C-CTT-G--CA-TAG-GCCT----C--T-----...C-GCC 6147
WRC.CI.97.97CI14           --CTTTGCAATG--AAAACAG-----C--A---GAAGTC--TGTAA-AG--TTT-TTT--TG--C-C--......................CAGAGAGAGTATGAGGAAGTCC-TACTA--C-TTATT-CCAC--GG-A------AG-GG---C--C--G-----...C---C 6027
WRC.CI.98.98CI04           --CTTTGCAATG--AAAACTG-----C--A---GAAGTC--TGTAA-AG--TTT-TCT-CTG--CGC--......................CAGAGAAATTATGAAGAAG-TC-CACTA--C-TTATT-TCAC--GG-A--A---AG-GG---C--C--G--C...C-T-C 6042
WRC.GM.05.Pbt_05GM_X02     --GTTTGCCATG--AAAAC-G-----C--CC--GAAGTC--TGTAA-AG--TTT-TCT--TG--CGC--......................CAGAGAAATTATGAAGAACA---CACTA--C-CTATT-TCAC--GG-C--T---AG-GGAT-A-A-C-A--C...C---T 4962
LST.CD.88.524              A-C---G-A--G--AAGAC-C-----C--AGAAGCAATTT-GGCCCAAG-CTC-G--T-CTGTGTC.......................AGCAGTACCGGGTCAGAAGA--AG-CTTGCATGAGAT-C--C-CTT-AG-TA-TAG-GCT--C-A-T-A--......CT 5016
LST.CD.88.447              A-A--AG-G--G--CAGAC-C-----C--AGAA-CAATTT-GGCCCAAG-CTCAG--T-CTGTGTC.......................AGCAGTACCGGGTCAGAGGA-AA--CCTGCATGAAAT-C--T-CC--AG-AA-TAG-GCA--C-----T--......TT 5019
LST.KE.x.lho7              A-A--AG-G--G--AAGAC-C-----C--AGAAG-AAT-T-GGCCCAAG-CTC----T--TGTGTA--C-G-ACA--.........................TCAGAAGA--AG-CCTGCATGAAAT-C--C-CCC-AG-CA-TAG-GCTT-----C-T--C...C--CC 6103
SYK.KE.x.SYK173            G-AGT--CA-----ATCC-AG------A--CCTCAAGAAA-AT----AG-TT--A-T--ATC--TAA-CGAAG-G-CA-GGACTGAT...TCCCCGACAATGGCTTGGGAAAG-ACCA--TTG-A---GG-----GCA--CA-T-TTA----C---GAA-----..-TGC 5981
SYK.KE.x.KE51              G-AGCCC-A-----ATGC-AG------A-AGC-GAAGAAA-ATT---AG-TT--A-T--ATG--TA--GGTGG-A-CA-G-CCAGAAGAAACACCCACCATGGCTTGGATTAG-T-CA--CT--AT--GCAT---GCA--CA-CTTTA----T---GAA-----..-TGC 5645
COL.CM.x.CGU1              A-A-CCCGG--G--CAGA--G--C--C---TT--C-CACA--TCA-TGA--AT-TGGGG-TGTAC-AG--GGC-CT---GGAGA.....................C-G---A-G-CT-GTCTG--TTATG-G-----CA--G-GTGTGCC--C-ACT-G--C...G-TCA 5476
```

```
                          Vpr premature end (HXB2 only)                              Tat exon 1 start   Vpr end
H1B.FR.83.HXB2            AATTGGGTGTCGACATAGCAGAATAGGCGTTACTCGA..............CAGAGGAGAGCAAGAAATGGAGCCAGTAGATCCTAG.................................... 5850
Vpr (frameshifted)        __I__G__C__R__H__S__R__I__G__V__T__R__.__.__.__.__.__Q__R__R__A__R__N__G__A__S__R__S__*
Tat exon 1                                                                                   _M__E__P__V__D__P__R##__.__.__.__.__.__.__.__.__.__.__
Vpr                       _N__W__V__S__T__*_
H1A1.UG.85.U455           ---------C-A-----------------A---T--C-...............GG---A----GC--G-------------------A.................................... 5295
H1B.US.90.WEAU160         -----------------------------A------C-...............-----A---------------------------A---.................................. 5848
H1C.ET.86.ETH2220         ------A--C-AG----------------A--TTA---...............-----A---------------------------A.................................... 5241
H1D.CD.84.84ZR085         -----------A-----------------A------C-...............-----A-------------T-----------A.................................... 5371
H1F1.BE.93.VI850          ---------C--T----------------A--GT--C-...............-----A---T-------------------------.................................... 5189
H1G.SE.93.SE6165          ---------C-A-----------------A----A-C-...............-G---A--G-T---GG-----C--G---------A.................................... 5246
H1H.CF.90.056             ---------C-A----------------AA--------...............-----A---T---------C-------------A.................................... 5196
H1J.SE.93.SE7887          ---C-----C-AT----------------A---TA-C-...............--A--A----GG--G-------------------A.................................... 5163
H1K.CM.96.MP535           ---------C-AC----------------A---TA-C-...............-----A----GG--------T-------------A.................................... 5045
H101_AE.TH.90.CM240       -----------A-----------------A---TG-C-...............GG---A----GC--G------A-TG---------A.................................... 5417
H102_AG.NG.x.IBNG         -----------A-----------------A---T--A-...............GG---A----GC--G--------TG----------.................................... 5374
H1N.CM.95.YBF30           G-----C--C-AG--C-----------GA-C----CT...............--A...--GAG---G-----A-------------.................................... 5440
H1O.BE.87.ANT70           ------A--C-A------T--------AA---AC-C-............AGAGGA--AG--AG---------T-----------GA.................................... 5907
H1O.CM.91.MVP5180         ------A--C-A------T--------AA--CTC-C-...TCTAACACAAGAGGA--AG--AG---------T-----------GA.................................... 5891
CPZ.CD.90.ANT             -CAC--A---GCT--------G-----ACA-GGA-CC..................................C-----CG--GA.................................... 5282
CPZ.CM.01.SIVcpzCAM13     -G-------C-AC--------------TA-A-TA-C-...............--A------AG-GC-----------A----C---GA.................................... 5432
CPZ.CM.05.SIVcpzEK505     G-----C----A---------G-----AA-C----CT...............--A...--GAG---G-------------------GA.................................... 5418
CPZ.CM.05.SIVcpzMB66      -----------AT----------------A-C-A-ATT.........AATCAG-----A---AG-------------A---------A.................................... 5401
CPZ.CM.98.CAM3            GC----C----A---------------GA-C-TC-CT...............GG-...--GAGG--------------------GA.................................... 5260
CPZ.GA.88.GAB2            G-----C----A---------G-----GA--GAA---...............---------AG---G-GA-..............................ATGGATCCTATAGATC 5241
CPZ.CM.98.CAM5            GC----C----A---------------GA--CT--YT...............G--...---AGG---------A---------Y--A.................................... 5548
CPZ.CM.05.SIVcpzMT145     ------C--C-AT----------------A-A-AC-CT...............-----A---AGG---G-----A---A----G---GA.................................... 5391
CPZ.TZ.00.TAN1            CCA---T---ATC--C-----------ATCA..........................................................TGAATCCCATAGATC 5484
CPZ.CM.05.SIVcpzLB7       ------C----A-----------------A----A-C-...............--A--A---AGG---G---------A----------.................................... 5404
CPZ.US.85.CPZUS           G-----C---AAT--------G-----GA--T-CTTGGCAAGGAGAACTCCT--A...G-GAG---G------T-----------A.................................... 5921
CPZ.GA.88.GAB1            -C-------C-AG---------------A-CCTC-C-...............--A--A---AG-TCC------T---A--------GA.................................... 5913
MON.CM.99.L1              GGAA--C---TAC-----ATATGC--AGCAA-TAA--AGATACCCTGTTCTTAG-CCA-TGAG-G-G-C--CTC--G-CCC-A--AGTTCTGTTCCACAGGCCGATCCAGACAATCCCAGAAGACCTTCCAGATATAGAATGGATGAATGA....TGGAACCTGTTGACC 5767
MON.NG.x.NG1              GGA-------TAT--C--GCATGCCC-A--GRT---CAGGTACCCTYGRCTG-GACCACT-AG-G-C-C--CCC-AG-ACCCA--AGTTCTGTCCCTCATGCAGACCCAGAGAGGCCTTTGCACCCGTCCAGATACCACCACGATGAATAG..TGCAASGATAGTGAACC 4380
GSN.CM.99.CN166           -GA---A--CTTT--C--GGTC-ATTTATCA...G-GAGATACCCAGCAATTAG-CCCTC-AGGG-C-CC-CCC-ACCAGACAGTA-CTCTGTTTCTCATGCCGATCCTGAGCAGCCTAGGAGACCCTCAAGATATCGCATGGATGAATAG.....AGACCCGGAAGAGT 5752
GSN.CM.99.CN71            -GA-------TTC--C--ATAC-ATA-AA--GTCA--AGGTACCCAGTAATCAG-CCCTT-AGGG-T-CC-CCC-ACCAGATAGTA-CTCTGTTCCTCATGCTGATCCTGAGCAGCCCAGAAGACCTTCAAGATACCGCATGGATGAATAG.....AGACCCGGCAGAGT 5740
MUS_1.CM.01.1085          -GAC--A--CTAT-----GTAT-ACA-A--A-TGA--AGATATCCAACCATAAG-CCTTTGAG-G-C-CACA-C--CC-CCT--TA-TTCAACCCCAAATGCTGACCCTGTACCTCCTTTGAGTCCTTCTAGGTACAGGATGGATGAATAA....TGGATCCATCAGTAG 5715
MUS_1.CM.01.CM1239        -GAG--A---TAT-----GTACTCTAA-ACA-TCA-GAGATACCCCAACATAAGACCCTTGAG-G-G-CACA---TCC-CCT--TA-TTCAATGCCTAATGCTGATCCTACTCCTCCTTTGAGACCCTCTAGATATAGGATGGATGAGTAA....TGGATCCTTCAGTAG 5705
MUS_2.CM.01.CM1246        GGAC--A---TAC------TATGCTCAAACA-TCA-GAGATACCCACCGTTG-GCCCATTGAG-G-G-CGCAGC--CCAGAC--TA-TTCCATGCCAAATGCTGACCCCACACCATCTTTGAGACCCTCCAGATATAGAATGGATGAATAA....TGGACCCTTCAGTTG 5822
MUS_2.CM.01.CM2500        -GAG--A--CTAC-----GTAT-GTCAG----T-A--AGATACCCACAGGTG-GCCCCTT-AG-G-G-CGCA-C-TCCAGAC--AA-TTCAACACCAAATGTTGACCCCTTTCCTCCACGAAGACCCTCCAGGTACCGCATGGATGAGTAA....TGGATCCAGAAGTAG 5775
RCM.NG.x.NG411            GCA---A---GA-GGG-----G--T--A-GAG-GA--GGAGGA.........A-TCCTCTTAG-TCC-T-CTTCA-TCG---AA-ATCTTGTAA.................................... 5687
RCM.GA.x.GAB1             --A---A---GC-GGG--T--G--T--AACAT-CA--GGAAGT.........A-TCCTCT-CG-TCC-T-CCGCAA-CG---AA-ATCATGTAA.................................... 5652
DEN.CD.x.CD1              -G-A--A---TTC-----GGA-GC--AAAGAGT-A--CACTAT.........-CA-AC-T-AG-CCCTTAC-CC-TC-C---GAAG--CCATGA.................................... 5792
                                                                                                         Vpr end
MAC.US.x.239              -GGC--A--CATC--CTC------C---CAAC--G-GGGAGGA.........A-TCCTCTCT--GCT-TACCGC--TC----AG-AT-CTATAA.................................... 6456
Vpr                       _G__G__C__I__H__S__R__I__G__Q__P__G__G__G__.__.__.__N__P__L__S__A__I__P__P__S__R__S__M__L__*
Tat exon 1                E__A__D__A__S__T__P__E__S__A__N__L__G__E__E__.__.__.__I__L__S__Q__L__Y__R__P__L__E__A__C__Y__N##__.__.__.__.__.__.__.__.__.__
H2A.GW.x.ALI              -GCA--A---G-C-G-TCA---G-T---CAA--AA-GGGAAGA.........A-TCCTCTCT--GCT-TACCGA--CC----AA-AT-CAATAA.................................... 6546
H2A.DE.x.BEN              -GCG--A---AAC-GCTCA-----T---TAA--AA-GAGAAGA.........ACTCCTT-CC--GCTGCACCGA--CC----G--AT-CACTAA.................................... 6556
H2A.SN.x.ST               -GCA--A--CG-C-GCTCA--G--T--TCAGC-CA-GGGACGA.........A-TCCTTT-T--GCT-TACC-A--CC----GG-AT-CGATAA.................................... 5998
H2B.GH.86.D205            -GCC--T----A---CTCA--G--T--ACAAT-AG-GGGAGGA.........A-TCCTCTCT---CT-TACCGC--CC-TA-GG-AT-CGATAA.................................... 6529
H2B.CI.x.EHO              -GGG--T--C-A---CTCA--G--T--ACAAC-GG-GGGAGGA.........A-TCCTCTCT--GCT-TACCGC--TC-TA-GG-AT-CAGTAA.................................... 6527
H2G.CI.92.ABT96           -GGA-AC---A--TGCTC---G--T--A-GG-ACGC-GGCGGA.........A-TCCTCT-T---CT-TACCGC--TC----GG-GT-TTCTAA.................................... 5872
H2U.FR.96.12034           -GGG--A-----G--CTCA--G--C---AGC-GCA-GGGAGGA.........A-TCCTCTCA---CTGTACCGC--TC-T--GG-GT-CTCTAA.................................... 6016
MAC.US.x.EMBL_3           -GGC--A--CAAC--CTC------C---CAAC--G-GGGAGGA.........A-TCCTCTCT---CT-TACCGC--TC-T--GG-GT-CTATAA.................................... 5581
SMM.US.x.H9               -GG---A--CA----CTC------C---CAGT--G--GGARGA.........A-TCCTCTCT---CT-TACCGC--K-----GG-GT-TTATAA.................................... 5930
SMM.US.x.PGM53            -GGA--A---GCT--CTC------C---CAA-G-A--GGAGGA.........A-TCCTCTGT---CT-TACCGC--TC----GG-GT-CTCTAA.................................... 6375
SMM.US.x.SIVsmH635F_L3    --G---A---GCC--CTC------C---CAAT--A--GGAGGA.........A-TCCTCTCT---CT-TACCGC--TC----G--AT-CTATAA.................................... 6460
SMM.SL.92.SL92B           GCA---A--CACC---TCA--G--T--ACAGC-AA-GGGACCA.........A-TCCCCTG-G-TCT-TACCGT--GC----GA-GT-TTATAA.................................... 5889
STM.US.x.STM              -GGC--T--C--C--CTC---G--T---CAAC-AG--GGAGGA.........A-TCCTTTG----CT-TACCGC---C----GGAGT-TTGTAA.................................... 6104
MNE.US.x.MNE027           -GGC--C--CAAC--CTCT-----C---CAAT--G--GGAGGA.........A-TCCTCTCT---CT-TACCGC--TC----AG-AT-CTATAA.................................... 5924
DRL.x.x.FAO               -TA---A--C-AGG-A--T---C-T--A...C-AGC-GGACGCTGG......A-CCCATTTCG--CCTT-CC-AGGGCG-ACCAAC-ATTGTAA.................................... 5755
MND_2.x.x.5440            -TA---A---A-GGGA--TC-C--C--ACAAGAAG--GGAAGATAT......A-CCCACTTCG-TCCTT-CCGAG-CCG-ACAA-CCCTTGTAA.................................... 5704
MND_2.CM.98.CM16          -TA---A--CA-GG-A---C-T--C--ACAAGAAG--GGGGGATGT......T-CCCCTT-CG-TCCTT-CCGAGATCGGACAA-CCCCTGTGA.................................... 6139
MND_2.GA.x.M14            -CA---A--CA-GG-A---C----T--ACAAGAAG--GGCAAA......TATA-CCCCCTTAG-TCCTT-CC-AGGCCG-ACAA-CCCTTGTAA.................................... 6066
OLC.CI.97.97CI12          --A---A--C-C-GGAGTGGAGC-TA-AAGA-AC-CTCACAGA.............................AGATC-GA-GGAAT-CAAGAA.................................... 5647
GRV.ET.x.GRI_677          -TG---T---A--AGG--AG--CCTTTTTCCC--TACGAAGAGAGGAGAAATGGACAAG---G-G--...-CCC-ACCCCCT.................CCTCCAGGACTTGCATAG.......... 6039
VER.KE.x.TYO1_patent      -TG------CA-GAGA--GCAGCCCTTT-AAC-ATACGAGGAGAGGAGAGATGGACAAG-G-G-G-C-GA-C-AATC--GTC.................CCACCAGGACTTGAGTGA.......... 6105
VER.KE.x.9063             TTG---T-----TAGA--ACAGCCCTT--AAC--TACGAGGAAAGAAGAGATGGACAAG-G-G-G---GA-CGAA-C--GCT.................CCACCAGGACTTGATTAG.......... 6105
VER.KE.x.AGM155           GTG---A-----CAGG--ACA-CCCTTT-AGC-ATACGAGGAGAGGAGAAATGGACAAG-G-G-G---GACC--GAC--GTC.................CCACCAGGACTTGATTAA.......... 6100
VER.DE.x.AGM3             GTG---A--C--CAGA--ACA-CCCTTT-AAC-ATACGAAGAAAGGAGGGATGGACAAG-G-G-G--CGA-C--GGC-CGT-.................CCACCAGGACTTGATTGA.......... 5603
TAN.UG.x.TAN1             -TG---C--CA-GAGA--A-CTCCTTTT-AGC-CTACGAAGAAAGGAGAAATGGAGTCG---GGG--CGG-AT-G---AGAG..............CCTCCTCCAGGACTTGCATAG.......... 6101
SAB.SN.x.SAB1C            -GG-A-A---T-TA-GCCA-AG-CTCA-CC-G-GTATGGACCAGGAGCAGGAGGCCC-CCCC--G-TCTG----GAGC-TC-GGAGGAGCTGCATCGGCCGCTCCAGGCCTG...TGA.......... 6333
MND_1.GA.x.MNDGB1         TCAA------GATGGA-C-TTC............A-GGAAAGAAGACCACAA-T-CCCCCCT--G--TTCA-GC-A--AG--GATAGATTATAA.................................... 5541
DEB.CM.99.CM40            -GAA-CT--CTTT--C--GGAG-G--AG-C---CA--GGG...TATCCTAACATT---CC-TTG-C-GGAA--AAT--AGA-GTTAG-GACGGAGAATAA..........CCTTTCACCTTTGCTT 5697
DEB.CM.99.CM5             -GAA-CA--CTTC-----AGAG-GG-AA-C----A-GAGA...TACCCCAATATT--ACCCCT--C-GGAA--AA---GGAGGT-AG-GATGGAGAATAA..........CCTCTCTTTTACATCA 5691
TAL.CM.01.8023            GCAA-----CTTT---GAGCAGCA-A-AAGA---GCCCACTACCCCAACATA-GACCCCTTT-CG-C-GGA-G...-AC-CCATGC-ATAA.................................... 5360
TAL.CM.00.266             GCAA-----CTAT---GAGCAGCA-A-AAGA----CGCACTACCCCAATATC-GACCCCTTT--G---GA--G-AG--G-CCATGC-ATAA.................................... 5865
SUN.GA.98.L14             -...AT-----CT.............................TCTT-C-TC---ATC----G----T-AG-A-ATGT-A-CCCAAGGCCACCAACAAGACCAAGACCAGGGCAAGGGGACATTAGAAGAGGCCTACAAGACTAA.......... 6259
WRC.CI.97.97CI14          CCAA-------CTA-ACAACA-GAGAAAAGGGAA-A-CAATTTCAAAGAGCAG-ACAATTGTAT-----CTTG-AG--A-C-G-AGGAGATCAGGGAGAAGTTGAGCAAAGA.......... 6139
WRC.CI.98.98CI04          CCAA-------C-A-ACAACA-GA-AAGAGAGAA-A-CAATTTCAAAGAGCAG-ACAATTCTTT-----CTTG-AA--AGC-G-AGGAGATCAGGGAGAGGTTGAGCAAAGA.......... 6154
WRC.GM.05.Pbt_05GM_X02    -CAA--A----CCA-AGAA-A-GAGAA-AGAGAAA--GATTTTCAAAGAGCAG-ACAATTTTAT------TTG-AA--GGCT--AGGAGATCAGGGAGGAGTTGAGCAGAGA.......... 5074
LST.CD.88.524             GCC-CAA----CCAG-TATGC-GC--C-AGA.........GCAGGAACAGCA--CCCA-C--A-GC-GCACCTAGACCA-CTGGAAGAGATATACAAAGAGGCCATTACTGA.......... 5119
LST.CD.88.447             GCC-ACC----CCAGCTATGC-GC--CAAGA.........GCAGGAACAGCT--CCC--C--A-GC-GCACC-AGACCC-CTAAAAGAGACATACAAAGAGGCCGTTACTAA.......... 5122
LST.KE.x.lho7             -...ACT----CCAGCTATGC-...-CA-CC-GA......GCAGGAACAGCC--TCCA-C--A-GC-GCACC-AGACC--ATGAAGGAGACATACAACGAGGCAGTGACTAA.......... 6203
SYK.KE.x.SYK173           TGCA--A----CT--AC-G-CC-G-TAT-CC-GA-ACAGAGGGTACCCCCAT-CTTCTTAG.................................... 6042
SYK.KE.x.KE51             TGCA--A----C---G--G-C--AGTA--CG-GA-ACAGGGGATACCCACAT-CTTCTTAG.................................... 5706
COL.CM.x.CGU1             TGCG---CAAT-T-GCTTTCCCTCTACGCAA-AC-CTTAG.................................... 5516
```

```
                                                
```

```
                                                                Tat Rev exon 1 end   intron start
H1B.FR.83.HXB2           AAC...............AGTCAGACTCATCAAGCTTCTCTATCAAAGCA...GTAA.........................................................................GTA....................GTACATGTA........ 6061
Tat exon 1               _N__.__.__.__.__.__S__Q__T__H__Q__A__S__L__S__K__Q
Rev exon 1               __T__.__.__.__.__.__V__R__L__I__K__L__L__Y__Q__S__
H1A1.UG.85.U455          ...............GGC---A--GA-------A--CT-A--C-------...----.........................................................................---A................TTA-CTT-----........ 5510
H1B.US.90.WEAU160        ...............AAC----------------A---------------...----.........................................................................---....................---------........ 6059
H1C.ET.86.ETH2220        ...............AGC---A--GA-------AA-CT-A----------...----.........................................................................---CCAAATAATAG.............------..ATGGT 5463
H1D.CD.84.84ZR085        ...............AGC----------------A-C--A--C-------...----.........................................................................---....................---------........ 5582
H1F1.BE.93.VI850         ...............AGC-----AGTG------AA-------C-------...----.........................................................................---GTGTTA..........AATA---T-----........ 5410
H1G.SE.93.SE6165         ...............AGC---A-AGGC-------A-C--G--C-------...--G-.........................................................................---GAAAAA..........ATTA---T-----........ 5467
H1H.CF.90.056            ...............AGTTTG--AGA-------AA----A----------...----.........................................................................---....................T--TCATA-........ 5407
H1J.SE.93.SE7887         ...............GGC---A------------A-CT-A--C-------...----.........................................................................---AATCAA...........GTA-C-T-----........ 5383
H1K.CM.96.MP535          ...............AAC---G---AC-------A-C-----CG------...--G-.........................................................................---GTGCTA..........AGTA---------........ 5266
H101_AE.TH.90.CM240      ...............AGC---A--GA-------AA-C--A--C-------...----.........................................................................---................ATAA---T-----........ 5632
H102_AG.NG.x.IBNG        ...............AGCC-----GA-------AA-C--G--C-------...--G-.........................................................................---GTAATA..........ATTA-C-T----G........ 5594
H1N.CM.95.YBF30          ...............AGC---A---G--------A-CT-A--C--G----...----.........................................................................---AAACCTGTATAT......................... 5654
H1O.BE.87.ANT70          ..................CA--CAGA----A---A-C--G--C-------...----.........................................................................---ACGCTA............................... 6109
H1O.CM.91.MVP5180        ..................TA--CAGA-A--A---A-C--G--C--G----...----.........................................................................---ACGCTG............................... 6096
CPZ.CD.90.ANT            GCTGAA.........AGC---G-A-A-A------A-C--G--------T-...----.........................................................................---CTCTTA............................... 5496
CPZ.CM.01.SIVcpzCAM13    -G-...............---A-A-A-A------A-C--G--CG------...----.........................................................................---TATTATCTGACCCT....................... 5648
CPZ.CM.05.SIVcpzEK505    -G-...............-A-A---G--------A-C-----C-------...---C.........................................................................---AAGCTATATAT.......................... 5631
CPZ.CM.05.SIVcpzMB66     -G-...............-------A--------A-C--G--C-------...--G-.........................................................................---ATTCATGCAGATATATGCTTA-GCTT-A......... 5631
CPZ.CM.98.CAM3           GCA............AGC---AC-CA-A------G-------CAGC----...----.........................................................................---ACTTCACATGCT......................... 5477
CPZ.GA.88.GAB2           GG-...............---A-A-GCA------A-C-----CG-C----...--G-.........................................................................---TCCCCTCTTT........................... 5457
CPZ.CM.98.CAM5           TCA............AGC---AC-CA-A------A----A--CG-C----...----.........................................................................---ATTTTGCATGCT......................... 5765
CPZ.CM.05.SIVcpzMT145    --T...............CAGG-T...A--A-G-A-C-----CG------...----.........................................................................---AAAGGCTCAC........................... 5600
CPZ.TZ.00.TAN1           -GT..................AC--A-A------A-C--G--CG-C----...----.........................................................................---CCC.................................. 5687
CPZ.CM.05.SIVcpzLB7      -G-...............----C-CA-T------A-C--A----------...----.........................................................................---CTCCTATAAGTTGTAGTCAATA-.............. 5629
CPZ.US.85.CPZUS          --GAGC............---GC-GA-A------A-C-----CGC-----...----.........................................................................--GCCCTTTTGTGCT......................... 6138
CPZ.GA.88.GAB1           GCAGGC............---A-A-A-A------A----A--C-------...----.........................................................................---T.................................... 6119
MON.CM.99.L1             GT-CCT.........GGCCT-AGC-GCAG-A-GAA-C--GC-AG------...--G-.........................................................................---..................................... 5976
MON.NG.x.NG1             -CG.........ATTGGCGA--G--GCT----G-A-AA--CGGG-GTCA-...----.........................................................................--G..................................... 4592
GSN.CM.99.CN166          G-ACCT...GATCCTAGCGTGTC-T--A------AC---GC--AGGTC--...----.........................................................................---AGGC................................. 5974
GSN.CM.99.CN71           G-ACCT...GCTCCTAGCGTG-C----A------ACC--A--CA--TC--...----.........................................................................---AGGC................................. 5962
MUS_1.CM.01.1085         ..................CTG-C--GC---TC--A-C--G--CA--TCA-...----.........................................................................---GC................................... 5920
MUS_1.CM.01.CM1239       ..................GTG----A-----CGAA-C--G--C---TCA-...----.........................................................................--GAT................................... 5910
MUS_2.CM.01.CM1246       ..................CTG-C-T-AG----GAA-C--G--CG-TTC--...----.........................................................................---AT................................... 6027
MUS_2.CM.01.CM2500       ..................CTG-CA-ACA------A-C--A--CA-CTCA-...----.........................................................................--GAT................................... 5980
RCM.NG.x.NG411           --G............AAG-A-----T-A---C-TT-GA-AC---TG----...----......................................................................................................... 5841
RCM.GA.x.GAB1            --G.....................G--G--A--TT-C--G--C-T--T--...----......................................................................................................... 5797
DEN.CD.x.CD1             ............CAAAGCG--AC-C--G------AAG--G----------..-----.GTTTGATGGATCTCCTCTCATCACTGATTTTAAGCTTTTCAATAATCATTATTTGCTTAGTTTGTTATATT--TATTATTACAGTGATAGTTCTT-AC-CATAGG....... 6091
                                                                Tat Rev exon 1 end   Tat Rev intron start

MAC.US.x.239             ..................GC-A--G--A--AC-T-----GC------CA-...----.........................................................................--...................................... 6603
Tat exon 1               _.__.__.__.__.__.__A__K__A__N__T__S__S__A__S__N__K
Rev exon 1               __.__.__.__.__.__.__L__R__L__I__H__L__L__H__Q__T__
H2A.GW.x.ALI             ..................-T-A--G-----TCGT-----GC----G-CA-...--G-.........................................................................--...................................... 6693
H2A.DE.x.BEN             ..................-C-A------C-TCGC-----GC-C--G-CA-...--G-.........................................................................--...................................... 6703
H2A.SN.x.ST              ..................-C-A--G-----TCGT-----GC----G-CA-...--G-.........................................................................--...................................... 6145
H2B.GH.86.D205           ........................---AC-GC-C-----GC-C--G-CA-...--G-.........................................................................--...................................... 6670
H2B.CI.x.EHO             ........................---AC-AC-T-----GC-C----CG-...--G-.........................................................................--...................................... 6668
H2G.CI.92.ABT96          ........................G--A--TC-TT----GC------CG-...----.........................................................................--...................................... 6016
H2U.FR.96.12034          ........................G--T--TC-T--G---C------CA-...----.........................................................................--...................................... 6157
MAC.US.x.EMBL_3          ..................GC-A--G--A--AC-T-----GC------CA-...----.........................................................................--...................................... 5728
SMM.US.x.H9              ........................G--A--AC-TT----GC----G-CA-...----.........................................................................--...................................... 6071
SMM.US.x.PGM53           ..................GC-A--G--A--AC-TT----GC-C----CG-...----.........................................................................--...................................... 6522
SMM.US.x.SIVsmH635F_L3   ........................---A---C-CT-C--GC------CA-...----.........................................................................--...................................... 6601
SMM.SL.92.SL92B          ........................G--A---C-TT-GA-AC------C--...----.........................................................................--...................................... 6030
STM.US.x.STM             ........................---T---C-AT----GC------CA-...----.........................................................................--...................................... 6245
MNE.US.x.MNE027          ..................GT-A--G--A--AC-T-----GC------CA-...----.........................................................................--...................................... 6071
DRL.x.x.FAO              ............TCAAGAGTATCTGAGAC-GGT-ACA-G-CTGTGGCAAG...----.........................................................................--GAAAAAG............................... 5918
MND_2.x.x.5440           ............GAGAAAGTA-CA-GCA--AGTCAAG-AGCT-TGGGAAG...----......................................................................................................... 5858
MND_2.CM.98.CM16         ............GAAAAA-TATCAGAAG--AGT--AG-G-CTGTGGGAAG...----......................................................................................................... 6293
MND_2.GA.x.M14           ............GGAAAA-TATCACAGG--AGT-AAG-AGCTGTGGGAAG...----......................................................................................................... 6220
OLC.CI.97.97CI12         ........................T-G------CAAGA-TCTAGGG-CAC...----.........................................................................--...................................... 5791
GRV.ET.x.GRI_677         ........................GA-A------ACC--A--CGGC----...----.........................................................................--...................................... 6201
VER.KE.x.TYO1_patent     ........................CT-G--TTG---C--T--CAGC-C--...----......................................................................................................... 6277
VER.KE.x.9063            ........................C--A---GCCTA----A-CAGG-C--...----.........................................................................--...................................... 6279
VER.KE.x.AGM155          ........................G--G---GCAT-C--G--C-GC----...----.........................................................................--...................................... 6274
VER.DE.x.AGM3            ........................CT-G---GCTT-C--GA-CA---A--...--G-.........................................................................--...................................... 5777
TAN.UG.x.TAN1            .....................TCA-A------GAA-CT-G-----C----...----.........................................................................--...................................... 6266
SAB.SN.x.SAB1C           .....................ATTT-G---A-TCAAGT-TCTCT-C-CA-...C--G.........................................................................TA...................................... 6477
MND_1.GA.x.MNDGB1        ..................-TA-CTGG-AG-GGT-AAGAAGCTAT-CGAAG...----.........................................................................--...................................... 5736
DEB.CM.99.CM40           -CGCCAGCTGCTGAAAGT-A----C--G------A--T-G--C--G----...----.........................................................................-C-AATCTGCATGTAAATGCTAAA-GGA-CA-T....... 5946
DEB.CM.99.CM5            GCT.........GCGAATGA-A--GA--------AG-T-A--C--G----...----.........................................................................---AAGCTGCATATGTATATAGTATC---........... 5927
TAL.CM.01.8023           -CTTCT......AGCAGC-A-----AGG--A-G-A-C--G----------...----......................................................................................................... 5514
TAL.CM.00.266            -CTCCT......AGCAGC-A----GAGG--A-G-A-C--G----------...----......................................................................................................... 6019
SUN.GA.98.L14            ........................-AGG--A-GCAAGAGGCTGT-CGAAG...----......................................................................................................... 6416
WRC.CI.97.97CI14         .......................................GCTG-TTCAAG...----.........................................................................--...................................... 6289
WRC.CI.98.98CI04         .......................................GCTA---CAAG...----.........................................................................-A...................................... 6304
WRC.GM.05.Pbt_05GM_X02   .......................................GCT----CAAG...----.........................................................................-C...................................... 5224
LST.CD.88.524            ........................GAAAC-G-GCAGGA--CTGTGG-CAG...----.........................................................................AA...................................... 5284
LST.CD.88.447            ........................GAAG--A-GCAAGA--CTGTGG-CGG...----.........................................................................-A...................................... 5287
LST.KE.x.lho7            ........................GAGAC---GCAGGA--CTGTGG-CAG...----.........................................................................-G...................................... 6368
SYK.KE.x.SYK173          ........................T--G-GG-T-A-AGCGC-C-T-CAGG...----......................................................................................................... 6218
SYK.KE.x.KE51            ..................GAGG-CTT-GC-GCGCA-AGCTC-A---CT-G...----......................................................................................................... 5894
COL.CM.x.CGU1            .............................................-CA--...----.........................................................................--...................................... 5637
```

```
                            Vpu start (ACG in HXB2)
H1B.FR.83.HXB2             .ACGCAACCT.........ATACCAATAGTAGCAATAGTAGCATTAGTAGTAGCAATAATAATAGCAATAGTTGTGTGGTCCATAGTAATCATAGAATAT.............................AGGAAAATATTAAGACAAAGAAAAATAGACAGGTTAATTG 6191
Vpu                         T  Q  P  .  .  .  I  P  I  V  A  I  V  A  L  V  V  A  I  I  I  A  I  V  V  W  S  I  V  I  I  E  Y  .  .  .  .  .  .  .  .  .  .  R  K  I  L  R  Q  R  K  I  D  R  L  I
H1A1.UG.85.U455            .-T-AC----.........T-GGA---CTGG------AC--GGC-GA--------C----CT----------A------A-T------GGT--------AAA.........................-AA---T-GC---AG----A---------------T-AA 5642
H1B.US.90.WEAU160          .-T----T--.........T-G-A----T------------C-----------GG---------------------------C------------.................-------------------------------------- 6189
H1C.ET.86.ETH2220          TGATTT--TAGCAAAAGTAGATTAT-G-A---T------------CA--------C----C-----------------A------C-TAT---------.................------T-G-------------G------T---------A 5603
H1D.CD.84.84ZR085          .-T----T--.........T---A----T-------------------T----GC----------------------A-------T-----------.................C-T-G----AA---G---------------T--------- 5712
H1F1.BE.93.VI850           .-T-TC-TA-.........T-GTT-GC-A---GG----C-------A--------C---------C-----------A-T------TAT---------.................-A----C-GG----G------------A-T-AA---TA-A 5540
H1G.SE.93.SE6165           .-T---GT-A.........T--GT----T-----------G----A--------T-C---GC---C-----------A--------T-T---------.................---G-----AG--A---G-A--G-----GA-A----C--- 5597
H1H.CF.90.056              .-T-T-TATA.........T--GG-T---G-AT-GG--CGCT-G-----AC-TTT--C---GCC...G--A--------A--------TAT---------.................-AA---T-GG----G----AG------------C------ 5534
H1J.SE.93.SE7887           .-T-AT----.........T-G-A-----C--------------TA----------TT-CC-T-------G-A-----A-T-----GTAT---------.................-A----C-G-----------G--------T-A-------- 5513
H1K.CM.96.MP535            .-T-GTGT-C.........T--G----TTC-AT-G---C-TT-G-------C-CT----T--GC-...---A----A---A--------TAT---------.................------T-GG---A-----A--G----A--T-------A- 5393
H101_AE.TH.90.CM240        .-T-AC----.........T-GGA---TAGT----------G-C-GA-------GC----CT-----------A-----AA-T------GCT------GT-.................-A-------C----G----------------------G--A 5762
H102_AG.NG.x.IBNG          .-T-------------.........T--A-----ACT----------G-C------------T-C---GC---C------------AG-------TAT---------.................---------AGG-A---G-AG---------------C--- 5724
H1N.CM.95.YBF30            .-T--TG..................TC-T-G-G-T-CA----G----G--C----G-T-GC------G--A-A--C---G-AT--C--TAT-G-------.................-A-------AA-TTG--GGA-------A-ACACA---GAC 5775
H1O.BE.87.ANT70            .-T---T-A-.........-GGGACC-GC--------A--ATTA-TAGT-CTTTGC-GT-T---AATG--A--T-A---GGGT-TA-T......CTTAGAAAATAT............TTAGAACAAAAGGAAC--GACAG--AGG-G---G-----CTTGA-AGGT-AA 6251
H1O.CM.91.MVP5180          .-T---T-AA.........GAGAACC-GC----CT--A----T---AGT-CTTTGTGTC-T---AATG--C--A-A----TGT-TAAC......CTTAGAATTTAT............TTAGTGCAAAGA-AAC--GATAG---GG-GCAGG-----CTTGAAAG-T-AA 6238
CPZ.CD.90.ANT              .-T-ACTAA-ATATTTGAGTATG-TT-TC-T--CT-TAGTAT-G--C--TGGATT--CTGT---C-T......A-ACTC-AT-AGC-T......T-TA-AATATAT.........AAGCAGCAGCAAATAGAT--T-AGAG--AT-------TC-----AGT-----G-A 5644
CPZ.CM.01.SIVcpzCAM13      .-T-AT-TTA..................----CTTT-GGTTGCTT-GCCA-T-CCTT---TT--C-TAAC--TT--A-A---AGG-ATT--......TGG............AGATTATGTAAA............C---ATAAGCTTG--GT-G----T--G-ACC-TGC-T 5772
CPZ.CM.05.SIVcpzEK505      .-T-TTGTTG.........C-TAT--AGT---G-T-CA---G-----C-A---AG-CTC------T------A--T---G-A----------......T-TAGAATATAC............AGA............G--G--AA-GT-G--GA-------TCTCA-C---GAC 5761
CPZ.CM.05.SIVcpzMB66       .-T-G-TATA......GTGCAG-A-G---GCTT-T------TTC--A-TA---A-T--G-T----TC-----AA-A---GTA-A---T......T-T............AAACTCTGTAAA............G-GGATAGG--G----AG--------T--A------- 5764
CPZ.CM.98.CAM3             .-T--T-A--......TGG...GA-CA-A---GCT--A-T-GTA---G-A---A---T--T-----T--T--A-CC---GGG----CT......TTTA-AGAGTGG............AGAAAAGGAAAAGAAG--GA-AAC---GC-CA------TAGG-CT-----A- 5619
CPZ.GA.88.GAB2             .-T--TTT-C......ATGTGGGT-GCTA---G-C--A-T-G-A---GTACTCTTT--G-T---AAT-----G------GGT--T---......-G-ATCTCTGTCTAT......AAAAGGTGGAAAAGACAC---GA-GAGCA-AG--TC...--T--TTT-A-T...A 5602
CPZ.CM.98.CAM5             .-T--T-AT-......TGG...GAGCA-A---GTT--A----T----G-A---A-C-T--T--T-TT--T--A--T---GGA----CT......T-TA-AAAGTGG............AGAGAAATAAAAGAAG---GGAAC---ACCCA--G---TT---AT---GCA- 5907
CPZ.CM.05.SIVcpzMT145      .-T---G-TA............GA---T--GCT----A--TT---CA---CTTT---GC-TG----T---T---CT---ATAGC--CT......T-T........................AAAGAGTAT-AA--GT--CA-CA-GT---GCG------A--AC-TCAG- 5727
CPZ.TZ.00.TAN1             .-T-AT-AAA......ATAG--GTGGG-AGT-TGTC-ACTAATG-CA---GCATTC-TTGT---TT-C-GA--T-AATAGGGGG-..........-GC-TGCTAATAGGT......ATAGGTATAAGAAGAGA-TT-GA-AGGGA-AGGCA-C-TCA-AGAGTA---GAAA 5835
CPZ.CM.05.SIVcpzLB7        .-T-G-C-TA............AT-GA-T---G-T--A---GC-----TA---A-C----C----T------A--A----TA-A--C-......T-TC-GCTTTAT.................-AAG-G-ACA---------AGGC----A-T-AA------- 5756
CPZ.US.85.CPZUS            .-T--TTAA-......TGGT-TGA-----GTTT-----C-TT-GG-A--...-A-GGT--TT---T-G-GA-AA-C---GGAC----T......-CTAGATTGTGG............AGACAAATAAAA-TA---GA-AAT-CT---CA-G-G--TC-A-ATC-TT-A- 6280
CPZ.GA.88.GAB1             .-T-ACT-TG........TTAGT-.GGCT---TTC-CA--TT-G---GCT--...--TGC-TGGAAC--TTG-A-A---GGATACA-T......ATTA-ATGGGGATAT......AGAAGGTATAAAAGACAT-GGC-TGAG-C-G--........--T--G--AC-T-ACC 6258
MON.CM.99.L1               .-T-A-TTA-...TGGTGGTC--TGG---C-AT--CTTACAGCC-GA-CC-CAT-GCCC-CCCC-T-GC--C-TG-GCT-GG...TGG.....................CGATATTATAAAATCACTAAG---TTT-A-AG--T-G-CCA-G-----C-G---C-G---C 6118
MON.NG.x.NG1               .-T-A-TA-C................TATTGGAT-GC--CCAT--CT--TTGGTGTG-T--TG-GCTTT---CGCCT--C.................ACCCTATACCTTCTCTATCAGTCTTATAAGGAGCAC--GCGGG-GTCC-GCTTCG--CAG---TTTCA--GGC 4728
GSN.CM.99.CN166            .-T---T---............G--GC---TTGGTGGTGG-GTGCT-CTA-TAT--CCT-C---TATC-TTG-...GTAG-GC-TC-T......-CTCTCTAC....................TTAGCTT--G-T-AG-GGGT-A--G-----CC-A-ACCAAC-CAA- 6101
GSN.CM.99.CN71             .-T-AGTG--............G--GCTC-CTGGTGGTGG-GGGCT-CT---AT--CCT-T---TATT-CTGC...-TAG-A--CT-T......-CTCTCTAC....................TTGGCGT--G-C-AG-GG-T-A-GG----GCC-A-G...A--CCA- 6086
MUS_1.CM.01.1085           .-T-A-CTA-...TGGTATT--G--GCT--CATTG--AC--GCA-CTACT-T-TG--TGCT---TTTGCCT------TAG-TTACCA-....................AGGTGGTGTAAGCCCAAGAAG.....................GT-GAG-T---TG-C--CC 6044
MUS_1.CM.01.CM1239          .-T-A-TTA-...TGGTATC--G--GC--CTTT-G--AC--GCA-CTACT-TCTG---GCCT--TTTGC-T-C---CTAG-ATATCAG....................AGGTGGTGCAGACCAAAG.......................GT-GA--TG--CG-C--CA 6031
MUS_2.CM.01.CM1246          .-T---TTAC...TGGTATC--GG-GCT-CCAT-G--AC---TA-TTATC-T-TT--TGCTT---T-GCCT--A-CCTAG--TATCA-....................AGGTGGTGTCAACCTAAG..................-C-CC--T-GA--TG-ACG-CC-CC 6154
MUS_2.CM.01.CM2500          .-T-A-TTGG...TGGTGGT-TG--GC--C--T-G--ACT--CA--TATT-T-TG---GCTT---TTGC-T-----CTTG-ATATCAG....................CGCTGGTGCCAGCCTCAG................-A-GGGCAGGT-GAG-T--ACG-T--CC 6110
RCM.NG.x.NG411             ....................................GTAT-G-A-G...TCCTGGTCTG--CCT--TT-TAGATGC-TAG-TTATAAG............GTGATAGGAGATATAATTAGAGAAGTTTTT-TTTGT-G-G-TTATAG--A-T-G-AG-TA-ATC-TTAGA 5960
RCM.GA.x.GAB1              ....................................GTAT-G-G-G...CCCTGG-CT-GT-CT--TT-TAGA-CC-CAG-T-A-AAG............GTTTTTATAGAT..................CTTTT-G-TAC--T-ATT-A-GG--A-T-G-ATAAG--AA 5898
DEN.CD.x.CD1               .-T--TTAAG.........................ATC-AGT---G-CC-A---A-TACG-TTGCT-G-T-T-T-CA--TGTTATT-CCTGG--AGC--CCATAGGT............GTAGGGTATCTTGCTTAT-G-GCC-AT-A-AGTTAT-G-GA---GCTTAGGTA-A 6227

MAC.US.x.239               ...................................................................................................................................................................... 6603
H2A.GW.x.ALI               ...................................................................................................................................................................... 6693
H2A.DE.x.BEN               ...................................................................................................................................................................... 6703
H2A.SN.x.ST                ...................................................................................................................................................................... 6145
H2B.GH.86.D205             ...................................................................................................................................................................... 6670
H2B.CI.x.EHO               ...................................................................................................................................................................... 6668
H2G.CI.92.ABT96            ...................................................................................................................................................................... 6016
H2U.FR.96.12034            ...................................................................................................................................................................... 6157
MAC.US.x.EMBL_3            ...................................................................................................................................................................... 5728
SMM.US.x.H9                ...................................................................................................................................................................... 6071
SMM.US.x.PGM53             ...................................................................................................................................................................... 6522
SMM.US.x.SIVsmH635F_L3     ...................................................................................................................................................................... 6601
SMM.SL.92.SL92B            ...................................................................................................................................................................... 6030
STM.US.x.STM               ...................................................................................................................................................................... 6245
MNE.US.x.MNE027            ...................................................................................................................................................................... 6071
DRL.x.x.FAO                ...................................................................................................................................................................... 5918
MND_2.x.x.5440             ...................................................................................................................................................................... 5858
MND_2.CM.98.CM16           ...................................................................................................................................................................... 6293
MND_2.GA.x.M14             ...................................................................................................................................................................... 6220
OLC.CI.97.97CI12           ...................................................................................................................................................................... 5791
GRV.ET.x.GRI_677           ...................................................................................................................................................................... 6201
VER.KE.x.TYO1_patent       ...................................................................................................................................................................... 6277
VER.KE.x.9063              ...................................................................................................................................................................... 6279
VER.KE.x.AGM155           ...................................................................................................................................................................... 6274
VER.DE.x.AGM3              ...................................................................................................................................................................... 5777
TAN.UG.x.TAN1              ...................................................................................................................................................................... 6266
SAB.SN.x.SAB1C             ...................................................................................................................................................................... 6477
MND_1.GA.x.MNDGB1          ...................................................................................................................................................................... 5736
DEB.CM.99.CM40             ...................................................................................................................................................................... 5946
DEB.CM.99.CM5              ...................................................................................................................................................................... 5927
TAL.CM.01.8023             ...................................................................................................................................................................... 5514
TAL.CM.00.266              ...................................................................................................................................................................... 6019
SUN.GA.98.L14              ...................................................................................................................................................................... 6416
WRC.CI.97.97CI14           ...................................................................................................................................................................... 6289
WRC.CI.98.98CI04           ...................................................................................................................................................................... 6304
WRC.GM.05.Pbt_05GM_X02     ...................................................................................................................................................................... 5224
LST.CD.88.524              ...................................................................................................................................................................... 5284
LST.CD.88.447              ...................................................................................................................................................................... 5287
LST.KE.x.lho7              ...................................................................................................................................................................... 6368
SYK.KE.x.SYK173            ...................................................................................................................................................................... 6218
SYK.KE.x.KE51              ...................................................................................................................................................................... 5894
COL.CM.x.CGU1              ...................................................................................................................................................................... 5637
```

```
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                
```

```
                          Env signal peptide end Env gp120 start
H1B.FR.83.HXB2            ......GCTACAGAAAAATTG...............TGGGTCACAGTCTATTATGGGGTACCTGTGTGGAAGGAAGCAACC.........ACCACTCTATTTTGTGCATCAGATGCTAAAGCATATGATACAGAGGTACATAATGTTTGGGCCACACATGCCTGTGTACC 6451
Env                       _.__.__A__T__E__K__L__.__.__.__.__.__.__W__V__T__V__Y__Y__G__V__P__V__W__K__E__A__T__.__.__.__T__T__L__F__C__A__S__D__A__K__A__Y__D__T__E__V__H__N__V__W__A__T__H__A__C__V__P
H1A1.UG.85.U455           ......AA-G--C--C-----...............--------G----C-------------------A--T--GTT.........-----C------------------------------------G---A--G-------C-----T----------------- 5896
H1B.US.90.WEAU160         ......---G-------C---...............------------------------------------A---------.........---------------------------------------------A---------------------------------- 6446
H1C.ET.86.ETH2220         ......-GA-TG-GG--C---...............----------------------------A-----A--T--T-G-.........C-T-----------------------------------------A----------C----GT---TT---------C-- 5860
H1D.CD.84.84ZR085         ......---G----T--C---...............------------T---------------A-----A---------.........-----G------------------------T-----A-A-------C-------A-C-----T------------------ 5972
H1F1.BE.93.VI850          ......---G----C--C---...............------------------------------------A---------.........--T-------C--------------------------------A-G-----C-----C-------------------------- 5797
H1G.SE.93.SE6165          ......--CT--A-T--T---...............----T-------------------------G-A--T--GAT.........--------------------C--------T-----AG-G-T--AAGC--------C-----T------------------ 5854
H1H.CF.90.056             ......---G--C----C---...............----T----A--------------------A--G---AA.........-------------C--------------------G--------G-----AAAG--------C-----T--------A-------- 5791
H1J.SE.93.SE7887          ......--A-A----G-C---...............-----------------------------------A--T---AG.........-----T-----C----T-----------------AG---T--AAA---------C-----T--------T-------- 5770
H1K.CM.96.MP535           ......AA-G----C------...............------------------------------------A-------A.........C----------C------------------------------G-AG--------------C-----T--C-------------- 5650
H101_AE.TH.90.CM240       ......--CT----C--C---...............-----T----T-----------G---------GA--T---GAT.........-----C---------------------C--G---C----G-----A--G--C-----C------------------------ 6019
H102_AG.NG.x.IBNG         ......AA-G-T---C-----...............--------G----C---------------------ACG---GAG.........-----C----------------------------------------A----------C-----T------------------ 5978
H1N.CM.95.YBF30           .........T-T---C--CAT...............-----A-----G--C-----------A--A----GA------GAG.........--A-----T--C-----T---------------CC--AG--------CT--C--CA-C-----------A--A-----T-- 6038
H1O.BE.87.ANT70           AGC......CTTAG-C-GC-A...............-AT-CA---------GC-----G-----A--G-A--T-----A.........C-AGTA----C-----T------------CCT-ACAAGC--T--AAAG------A-------AT----A--------T-- 6499
H1O.CM.91.MVP5180         AGT...TA--GTA--C-----...............-AT-C------T----C-----------A--G-A--G--G-A.........C-AGTA----C-----T------------CCT-ACAAGC--T--ACAG------A-------AT----A-----C--T-- 6489
CPZ.CD.90.ANT             AAG...-GG--GA-TG--GAC...............-AT--A-----A-TC-----A--C-----C----GAA-T--G--A.........C-T-------------CA--A---CTCCATGACAAG----------G--C-----A-----A--TACCAGT-----G-- 5877
CPZ.CM.01.SIVcpzCAM13     ACA.........TCTG-TGAA...............-----A-----G--------A--------------GA----TC-A-.........--AGTA--C--C-----T---------------C--AGC-------CT-----CA-A-----------C--A-------- 6029
CPZ.CM.05.SIVcpzEK505     .............TCTG-----...............-----A-----G--C---------------C----GA--T---GAG.........--AGT------C-----T----------G--CC--AG--------CT--C---A----------G--A--A--C--C-- 6000
CPZ.CM.05.SIVcpzMB66      ............--T--CA--...............-----G-----T--C-----A--------C----GA--G---GA-.........--A-----G-----------------A------C-GA-CC----A--G--C-----A-----TT-T--A--T------T- 6033
CPZ.CM.98.CAM3            ACT...-G--GT--TC-G---...............----CA-----G--C--------------------A--G---GA-.........--A--A--C--------C-----G--GTC---TCT-A-C-A-----CC------A---------T----A--------T-- 5870
CPZ.GA.88.GAB2            AAGACA-AAT-T--TC---G-...............-----A-----A--C--------T--A--C---------G--T--T.........--AC-AT-G--C------G--A---A-C---T-TAA-AGAG--ACCT------A--------------A--------G-- 5856
CPZ.CM.98.CAM5            ACT...-GCCAT--G...-G-...............----CA-----G--C-----A-------A-----A--G---GAA.........--A--A--T--C-----C-----G--CTC----CT-A-C-A---A-CC--C---A--------T----A--------C-- 6155
CPZ.CM.05.SIVcpzMT145     GGT......GA------T-G-...............--------T--T-----------C--A--A---G---G----AA.........------C--T--C-----C--C---------GAGT---AG------A-CT-----CA-C---------C--A--T--------- 5996
CPZ.TZ.00.TAN1            ATT...-GATTTAGC-CT-AT...............-ACAC------G-T------A--------T-----A--G--CCAA.........C-A--CT-G--------C--T------G-TATTAC-AG--G---TAA---C--CA-A-----A------AA------G-- 6071
CPZ.CM.05.SIVcpzLB7       TTG...ATA-TTAGT-GCACTAAAAATGAAAATTTA-----G-----G--C-----A--G-----C------A------GAA.........G-A-----C--------G--------A------C-AA-GG------C---C---A-A------T-T--A--T--------- 6016
CPZ.US.85.CPZUS           CCA...-GATTGTCT-GC--A...............----CT-----A-----------------A----GA--T-T-GAG.........--A--CT----C------C--T------A--G-----CA-GCAG----CC--C--CA--------------G--A-----C-- 6522
CPZ.GA.88.GAB1            ACC......T-T--G....-A...............-----A-----A--------A--------T---C-T--T--TGA-.........C-GGTA--C--------C-----C-----G---C-----G---C--AG---------CT------A--------------G--A--------- 6515
MON.CM.99.L1              TTG...-TG-AT-GGTCTAAC...............---AC---C-----C--C--T--T--A--A-----ACCG-----T.........C-TC----C--------C--T---C-A--TTACGGGTC--A-----C-GGG--CAA----CTGG-GTC-T----CC-C-- 6366
MON.NG.x.NG1              RAT...T-AGG-A-C----CA...............-----------A-TC-----A--G--G--C-AT--ACCT--G--G.........C-TC-AT-G--C-----T-----CC--G-YTATGGCTCA-A-----C-GGG---AA----CTAG--TC-T----CC-C-- 4979
GSN.CM.99.CN166           GGT...TTG-AGTGTG--AAT...............---ACG-----G--C----------G--A----G---C-----T.........C-AC-AT-G-----C--C-------C-AG-TAT-GC-AG--AT--ACC-GGA--CA-----ATTT-TACA--A---T-G-- 6340
GSN.CM.99.CN71            TCG...AGGT-TA--GGGAAC...............---ACG-----G-TC----------G--A----GA--T----AA.........C-AC-AT-G--C-----C---------AG-CAT-AC-AG--AT--ACC-GGA--CA-----AT-T-TACG--A--CT-G-- 6325
MUS_1.CM.01.1085          ATA...AT-TGG-G--G-CAA...............--------------T------------G---------A--T--T---.........C-AC-A--C--C---------------C--CAT-GCAA---G---ACC-GGG-----A---ATA--CACA--A--CC-C-- 6295
MUS_1.CM.01.CM1239        ATA...ATATGG-GC--TCA-...............------------G-TC-----------G--A------A--T-----T.........C-AC-A--C--C------------------CAT-GCAA-C-GG--ACC-GGG-----G---AT---CACA--A---C-T-- 6282
MUS_2.CM.01.CM1246        AGA...AAACAT-CC--GCAA...............--------G--G-T----A-------G--A-CA---A-T--C--A.........C-AC----C--C---------A------TAT-GCAA-C-G----CC-GGG--C--G-T-ATA--GACA--A--CC---- 6414
MUS_2.CM.01.CM2500        CTG...ATATGG-G----CA-...............------------G-T------A--C--G--C-------T-----A.........C-AC-AT----C-----C------------CAT-GCGA-C-GG---CC-GGG-----G---ATA---ACA--A--CT---- 6358
RCM.NG.x.NG411            ATA...-TAGG---G---CAA...............-AT--G-----A-T------T--C--A--A------A-CAGT--A.........GTTC--A-G-------T......................--T--TAACGCAC-ATCA----GA--GTTAAAT---A---- 6134
RCM.GA.x.GAB1             GGA...T-ACA-A--CCGCAA...............-ATA-A-------TC-----T--C--A--A---GAA-CAGC--A.........GTGC--A-G------TG......................--T--TAATAC-C-ATCA----GA--T-TAAAT---A---- 6078
DEN.CD.x.CD1              TGT...A-A--------GCA-...............-AC---------A-TC-----CA------CA----G-A-CAGTG-A.........G-TC-G--G--C--CA--.....................G-GAT--ACAGAGGA--A-----AG--GCAT-A-----T-- 6445
                          Env signal peptide end Env gp120 start                                                 V1 loop start                                     V1 loop end
MAC.US.x.239              GGG...ATCTATTGT-CTC-A...............-AT---------T------T-----A-CT---G-A-T--G--A.........-TTC-C--C---------.....................--CA--AATAGGG--AC-----GA---AC-CAG--CC---- 6779
Env                       _G__.__I__Y__C__T__L__.__.__.__.__.__.__Y__V__T__V__F__Y__G__V__P__A__W__R__N__A__T__.__.__.__I__P__L__F__C__A__.__.__.__.__.__.__.__T__K__N__R__D__T__W__G__T__T__Q__C__L__P
H2A.GW.x.ALI              TTA...-TATATTGT---CAA...............-AT--G--T--T-T------C--G--A-CA-----AA-T---T--.........-TTC-C--C---------.....................--CA-AAATAGAG--AC-----GA--CATACAG--CT---- 6869
H2A.DE.x.BEN              TTA...-TATATTGT-GCCA-...............-AT--G--T--T-TC-----CA----C-C------AA-T---T-T.........-TTC-CT----------.....................--TA-AAATAGAG-CAC-----GG--CATACAG--CT-G-- 6879
H2A.SN.x.ST               TTA...ATATATTGCGTCCAA...............-AT--G--T--T-TC-----C--G--C-------GAA-T---T--.........-TTC-C--C---------.....................--TA-AAATAGAG--AC-----GA--CATACAG--CT-G-- 6321
H2B.GH.86.D205            GGG...TT-GT-TGT---CAA...............-AT--T--T----TC-----CA----C-CA----G-A-C-----A.........GTTC-C--CA--------.....................--CACAAACAGAG-CACC----GA--TGTACAG---C-C-- 6846
H2B.CI.x.EHO              GGG...TA--TG-GC--GAAC...............-TT-----T----TC-----TA----C-CA-----AA-T---T-A.........-TTC-C--C--------T.....................--CAGAAACAGAG--ACC----GA--TGTACAA--CC-C-- 6844
H2G.CI.92.ABT96           GGG...TTCTATTGT-CCAAA...............-AT-----T--T-T------CA----C-CA----G-A-C--G-G-.........GTAC----C--C-------.....................--CA-AAATAGGG-CACA----G---TACACAG--CC-T-- 6192
H2U.FR.96.12034           GTG...ATATACTGT-CTCA-...............-AT-----T----T------CA------CA----G-A-C--G--A.........-TTC-C--C---------.....................--CA-CAATAGAG--ACA----GA--CACACAG--CC-T-- 6333
MAC.US.x.EMBL_3           GGG...ATCTATTGT-CTCAA...............-AT-----------T------T-----A-CT----G-A-T--G--A.........-TTC-C--C--C-------.....................--CA--AATAGGG--AC-----GA---AC-CAG--CC---- 5904
SMM.US.x.H9               GGG...AT-TATTGTGTTCAA...............-AT--A-----A-TC-----T-----A-CA----G-A-T--G--A.........GTTC-C--C--C-------.....................--CAGAAATAGGG--AC-----GA--CACACAA--CT-G-- 6247
SMM.US.x.PGM53            GGG...ATCTATTGC-CTCAA...............-AT---------T-T------T-----A-CA------A-T--G--A.........-TTC-C--C--C-------.....................--CA--AATAGGG-CAC-----GA---AC-CAA--CT-G-- 6698
SMM.US.x.SIVsmH635F_L3    GAG...AT-TGTTGTGTTCAA...............-AT--A-----A-TC-----T-----A-CA------A-T--G--A.........-TTC-C--C--C-------.....................--CA--AATAGGG-CAC-----GA---ACACAA--CT-G-- 6777
SMM.SL.92.SL92B           GGG...A-CTGGTGTGC-CA-...............-AT--A---A---T------TA-T----CA----G-A-C--T--G.........-T-C-C--C--C-----G.....................--CC--AATAGAG-CACA----GG--CGT-CAG--CT-G-- 6206
STM.US.x.STM              TTG...A-CTATTGC-CTCA-...............-AT-----G--A-T------T-----A-CA----G-A-T--G--G.........-TTC-C--C--C-------.....................--CA--AATAGGG--AC-----GA---AC-CAA---C---- 6421
MNE.US.x.MNE027           GGG...ATCTATTGC-TTCAA...............-AT-----------T------T-----A-CT----G-A-T--G--A.........-TTC-C--C--C----T-.....................--CAG-AATAGGG--AC-----GA---AC-CAG--CC---- 6247
DRL.x.x.FAO               AGT...ATAGG-.....-CAT...............------A--G--G-TC-----AACC--CAAA----GACC-----T--.........--ACA---GA-C------T.....................--G--TAAT---TCAT-C----TG--TACAAGT--CA---- 6088
MND_2.x.x.5440            GTA...ATAGT--G-G--CAA...............------G-----G--C-----CAC---CAAA---C-TA-G----GG.........--ACAC--G----------.....................-----CAATA--TCAT-C----TG--CACAAG------C-- 6039
MND_2.CM.98.CM16          CTA...-GGTT--G---CCAA...............------G-----A--------CAC----AAA---C-TA-G---GAG.........--ACAC--G--C-------.....................-----AAATG--TCCT-A----TA---AC-AGT-----T-- 6466
MND_2.GA.x.M14            ATA...-G--T--G----CAA...............------A-----A--------CAC----AAA---C-CCC-----GA.........--GCA---T----------.....................--T--CAACA-CTCCC------TA--CACAAGT-------- 6395
OLC.CI.97.97CI12          GGAATCAG--T--GG...CAA...............-AT--G-----G-TC-----AAC---AAA-----GA--TAGT-AT.........GAT--GT--A-G----TG.....................----CTAACA-A-G-C-G----T-----ACA--T--CC---- 5958
GRV.ET.x.GRI_677          GGA......-T--GT--CC--...............-AT--G-----G-T------AA-C--A--A-----AA-TT----A.........GTTCAGGC------CATG.....................--GCCCAATACC---A-G-----A--CACCAA---CA---- 6371
VER.KE.x.TYO1_patent      .......-TGTT-AGT--GCAA...............---A-A-------T------AA----A--A-----AA-CAGCT-G.........GTGCAGGCT--C--CATG.....................--TCCCACCACA-GCT-A-----T--TAC-AAT--CA---- 6449
VER.KE.x.9063             ......TTA-GT----G-CAA...............------------G-TC------A--------A------A-CAGCT-A.........GTACAAGCC--C---ATG.....................--TCCTACTACC-GGC-A-----T---AC-AAT--CA---- 6452
VER.KE.x.AGM155           ATA...AG----A--C-GCA-...............---A-A------G-TC-----A-----A--A-----AA-CAGCT-A.........GT-CAAGCT-----CATG.....................---CCTACTAC--GGT-G-----A--TAC-AAT--CA---- 6450
VER.DE.x.AGM3             CTT...AA----AGGC--CAA...............------------A-T------A-----A--A-----AA-CAGCT-A.........GTACAGGCT--C--CATG.....................---CCCACCACC-GAC-A-----A--TAC-AA--CGA---- 5950
TAN.UG.x.TAN1             CTG...TTATAT-GG-C-CA-...............-ATA-T-----G-T------TA-C--A--A-----AA-CAGCT-A.........GTGCAAGCT--C---ATG.....................---CCTAATACC--CC--------A---AC-AAT---A---- 6448
SAB.SN.x.SAB1C            CAG..................................-AT--T-----G-T------TA-C--A--A-----AA-TAGTT-A.........GT-CAAGCC------AAG.....................--CCCCAATACG--CC-A------T-TAC-AAT--CA-C-- 6647
MND_1.GA.x.MNDGB1         GGT...AT--T--G----CAA...............-AT--G-------TC-----A-----A--A-----------T-AA.........--ACA-T-GA-------T.....................-----TAATTCA-G-C-C----TA--CAC-AAT--CA---- 5905
DEB.CM.99.CM40            AAA...CAG-A-A-GC--CAA...............-ACT-G-----A--------A-----A--C---GTA--T--C-AA.........GTTGA-T-G--C---A--G-CA-CT-ATC--AG..................TCA-GA-----AGTGACA--A--CC-G-- 6131
DEB.CM.99.CM5             GCA...C-A---A-GC--CAA...............-----G-----A--------A--G--C--A---C-CA-T--C--T.........GTTGGCT----C---A--G-TA--AG-TC--A-..................TCA-GA-----TGTGACA--A--CT-G-- 6100
TAL.CM.01.8023            ACC...T-CG-----C-GCA-...............-TCC-------G--C--C--A--G--A--A-----AA-T-----G.........C-TC-AT-G--C--C--C.....................--TAT--C-AGGG-CCAG---T-T---TCCCAA-------- 5723
TAL.CM.00.266             TTT...ATA-T--GG--CCA-...............-AT-----CA----------A-----A--------AA-T-----A.........C--C-AT-G--------C.....................--CAT--C-AGGG-CCAA---T-A---TCCCAA--C--C-- 6203
SUN.GA.98.L14             GGT...AG--T--C--T-CA-...............-AT--A-----A-TC------ACT---AA----G-ACC----GTA.........GTGC-AT--A-A-----T.....................T---CTAATA---GCT-G----TG--TAC-T-A--CT-G-- 6585
WRC.CI.97.97CI14          ATC...AG-TT--CTGT-CAA...............-AT--A-----G--C-----A-----ACA----G----CAGT--A.........GA-TTC---C-G-------.....................-----AAATAGATCAT-A----TA-GCACACG----A-T-- 6450
WRC.CI.98.98CI04          TTA...-GG-T--GG---CAA...............-ACA-A-----G--C-------------CAA---G----GAGT-AA.........GATTTCT-GC-C-------.....................-----AAACAGATC-C-G----T--GCAC-CA---CA---- 6465
WRC.GM.05.Pbt_05GM_X02    TTA...-GATT--G---GCAA...............-ACA-A-----G-----------G--ACA----G---------AG.........GA-TT-A-G--G-------.....................------AATAGATC-C------TA---ACACAT---C---- 5388
LST.CD.88.524             GGA...-AA-AG--G---CAA...............-AT--A-----A-TC-----A-----AAAA---G-A--T--TGTA.........GTAC--T--A-C-----T.....................T---CTAATA---GCT-G----T---TACAT-T---C-C-- 5460
LST.CD.88.447             GAG...AAA---A-G--GCA-...............-AT--A-----G-TC--------G---AAA---G-A------GTA.........GTGC-AT--A-------T.....................T---CTAATA---G-T-A----T----ACAT-A--CA---- 5460
LST.KE.x.lho7             ATA......TGG-G----CAA...............-AT--A-----A-T---------G--AAAT---G-T--TAATGTTTCA.....GTGC-CT-GA-------T.....................T---CCAATAC--GCC-C----T---GACAT-T--CT-G-- 6538
SYK.KE.x.SYK173           TTT...ATGGA-A--C--CAA...............-AT--A-----A-T-------A---ACAT---G-A--T---TAT.........G-AC----C------A-C.....................--CTCCCATA-AGGA-GA-----T--TA-GAAT-----G-- 6400
SYK.KE.x.KE51             GTA...----GGTT-G--CAA...............-AT---------A-T------A--T--CAAT---G-A--T--TCAA.........GTAC--A-G----------.....................--CCCCCATACGGGA-G------A---A-GAA--------- 6078
COL.CM.x.CGU1             GTG...TTA...AGT--TAAA...............-----A-----A---C-A--A--------CA---G-A------GATGTAAATGATCAGCAGT-C--C--CTTT--TTCCT-CCC--AG.....................A--CAACAGGTGTTG-GG---T---- 5807
```

```
H1B.FR.83.HXB2           C...ACAGACCCCAACCCACAAGAA.........GTAGTATTGGTAAATGTGACAGAAAATTTTAACATGTGG......AAAAATGACATGGTAGAACAGATGCATGAGGATATAATCAGTTTATGGGATCAAAGCCTAAAGCCATGTGTAAAATTAACCCCACTCTGTG 6603
Env                      __.__T__D__P__N__P__Q__E__.__.__.__V__V__L__V__N__V__T__E__N__F__N__M__W__.__.__K__N__D__M__V__E__Q__M__H__E__D__I__I__S__L__W__D__Q__S__L__K__P__C__V__K__L__T__P__L__C__
H1A1.UG.85.U455          -...---------------------.........A---AT-----C------------G-A-------------.....-----A---------T----------A-----------C------C------------------------G--------T-----C- 6048
H1B.US.90.WEAU160        -...---------------------.........----------A---------------------------.....-----A-------------------A-------T-----------------------A---------------------------- 6598
H1C.ET.86.ETH2220        -...----------G----------.........T-G-GT---A---------------------------.....-------------G--G---------C--------------------G--------------------G--G------------- 6012
H1D.CD.84.84ZR085        -...---------------------.........A--A-C--AA-----C-------------------------.....-----CA-------G--C----------------------T---------T-----A----------------------G----- 6124
H1F1.BE.93.VI850         -...---------------------.........---T-TC--AA-----A-----------G-T-------.....-----A------------A------ACA---------------------------------------------G---------------- 5949
H1G.SE.93.SE6165         -...---------------------.........A--A-TA---A------A------T--------------.....-----A-------G-----A-------------------------------G-----------------------GC-------T------- 6006
H1H.CF.90.056            -...--------------------G.........A-G--CA---AG-----A-----G-GC-----T-----.....G-----A-------G--G---------ACA--C----------------------------T-G--A---------------------------- 5943
H1J.SE.93.SE7887         -...-----T--A-G---------G.........A-GAATC--CC------A--------C------------.....-----------------C---------G--A-------------G-------G------T----------------GA-------T--A---- 5922
H1K.CM.96.MP535          -...--G------------------.........----AGA---A------A--------C------------.....-----A---------G---------ACA---------------------G-----------A--T------G-G--------T------- 5802
H101_AE.TH.90.CM240      -...---------------------.........A--CACC---A------A---------------------.....-----A---------G--------G-----G-----------------------T-----------------G-----T--T-----C- 6171
H102_AG.NG.x.IBNG        -...--------T------------.........A--CAT----A------A--------G------------.....-----A-----------------------A-------T---C------C------------------------G--------T-----C- 6130
H1N.CM.95.YBF30          T...--T--T-----T---------.........--GC----ACCC-----A--T-----A-----T------.....G-----A-A----C---C--A-----A-------T------C-G----A--G---T------C-----T-------------T-A----- 6190
H1O.BE.87.ANT70          T...----------CT---T-T---.........TATCC----CAC-----------TG-C-----T--A---.....-----T-------------A-----G--A--C--T--T----------C--G--T--T--A--T-----TC--A-G--TTTC--G----- 6651
H1O.CM.91.MVP5180        T...-----T-----T-----T---.........T-TCC-C-A-GC-----------T--C---G-T--A---.....-----T-------G--C--A--------A--C--C--T-----G----A--G--TT-------T----AG---A-G--TTTCT-A----- 6641
CPZ.CD.90.ANT            A...-T---T--AG-T--TATT...............-----TAG-C-C---...--CTC-GTC-GGTTT-ATGCTTAT...-------T-T---------AGT---ACA--A-----GN-ACAA---TCC-A-------AT---------------C----A--TA-G---A 6026
CPZ.CM.01.SIVcpzCAM13    T...--------T--T--T------.........-----TC-TCAT-----A-----GG-C------T------AAT.......-C-G-------------------A--A--C--C-GT---C-C----C--G--T--T--A-----------GC----T--TT-A---- 6181
CPZ.CM.05.SIVcpzEK505    -...--T------------------.........---C--A-ACC---------------CG---G-T------AAA......CA-T---------T--A-----A--A--C--T--T--C--G----A--G--TT----A--C-----T--G---------T-A---- 6152
CPZ.CM.05.SIVcpzMB66     -...---A----T------G-G--G.........A--A-C-AAC-----A--T---T-C--C--TGCA---GAA......CA-T-------G--A-----------------G---A-C-T----C--G--TG---A---------------TT--TT-G---- 6185
CPZ.CM.98.CAM3           T...--------T------------.........--GCAGA-CCC------A--T-----C-----T---------......-C---AC-----C------A---A--ACT--C--T--T---C-G----------TT-G--A-------T--GC-G--G---A-A---- 6022
CPZ.GA.88.GAB2           T...---------TCT---G----G.........--CA-CC-AA-T-----------G-A-----TG-A---.....G-C----CT---------G--A-----G---------T-CT--CC------C--G-----T-GA--T-----------G-A---TT-A---- 6008
CPZ.CM.98.CAM5           T...GT---T--T-----C-----G.........--GTATA--CC------AT-T----GC--------------.....-----A-T----C------A-----G--A-----T-GT---C-G-----------T-----A-------T---C-G--A---A-A---- 6307
CPZ.CM.05.SIVcpzMT145    T...--T--T----CT--T-----G.........---C-C--ACCT-----C--T---G-G-------------GAA......----T----------T--A-----G--A--C-----A---C-C----A-----T--T--A-----------------A--CT-A---- 6148
CPZ.TZ.00.TAN1           T...TT---T-----T--TT-T---.........---ACCC-A-CC------T--AT--GG-----T---GAA.....G-----T-------GC--G-----A-A--A------T-ATCAC-TTTC-A--G--TT-T-----T-------------A---T-T--CA 6223
CPZ.CM.05.SIVcpzLB7      -...-----T--TG-T--TA---C-.........A--A-C-AAGC-----A------------------T-----ACA.........CA------A-T---------AC-----G----T--C------------------A------A----GC-------T--T---- 6168
CPZ.US.85.CPZUS          T...--T-------T----------.........---CAT---CC---------T-----G---G-------------.....G-----A-T----C-----A------G--A-----T--T--CC-----C--G---T-----A--T---A----G---------T-G---- 6674
CPZ.GA.88.GAB1           T...-----T---GT--T--G----.........---T-TC-TCC------A-T----TCA-------------.....------A-T-----G--C--A---------A--C--T------C-C-------G--T--C------T--------G--G----TT-A----- 6667
MON.CM.99.L1             G...--------T-C---G---TCT...........C---ATT-G--CA-C--T--GG-G--C--TGCT-ATCAA.......--CT---------G---G-A---GTA-------G-AGTC-C-G-TCTCG--GGCTT-------T-----G--------A--CA-G---- 6515
MON.NG.x.NG1             T...T----T--TGTT----GC-CT...........C-CCCTC-C---A-A-------GC-----TGCA-ATAAC.......--CT---------G---G-----GTG--A--C----AATCAC-C-TCTCA--GGC-T---------------G--------TA-G---- 5128
GSN.CM.99.CN166          -...T-G------TCA--GGC---G.............--GCC-T----CA-A-----G--A--C-----C-AT...........-------T-----------TG--G---GA--T--C--GG-ATCAC---TTA---GGC---G--A-------C----G--T---A-G---- 6489
GSN.CM.99.CN71           -...T--------TCA--GTCT--G.............--GCCAC----CA-A-----GG-A-----T--C-AT...........-------T-T---------TG--G--AGA--T--C--GG-ATC-C---TTA----GGC----GA--------C---C-G--T--GA-G---- 6474
MUS_1.CM.01.1085         -...--G--T--TGCA--C--G---.............--GCA-T-G---A-CT-T-GGG-AAAGTTTGACATCTATAAG..---T-T---------C-------AC---T---G-G-GGTC-A-C-TTA---GGC-T----A--------C-----------TA-G---- 6447
MUS_1.CM.01.CM1239       -...-----T---GCA--C--G---..............---CAAT----A-CT-T-GGG-A-A-TTT-ATGTCTATAAA..---T-------GACC-------GT---T---G---GGTCCA---TCA-C--GGC------A--C-----C-----G-----TA-G---- 6434
MUS_2.CM.01.CM1246       -...--G------GCG-----G--G..............---CAAC----CA-CT--A-TG-GCACTTT-ATATATATAAG..--CT-T---------T------AC---T---G-G-AGTCCA-C-TTAC---GGC---T-----G-----G---C----A--TA-G---- 6566
MUS_2.CM.01.CM2500       -...---------AGC---T----G..............---CAAT----CA-C----GGG-A-A-TTTGATATCTATAAA..--CT-T-----G--T--A---AC---T---G-GCGATCCA---TCA---GGCT--T--A--C---------C----T--CA-G---- 6510
RCM.NG.x.NG411           A...GA--GAGG--TAT---C----..............---CCA----------T-----G--A---G-TGCT---GAA.........---AG-T--TAT-----AGCAA-GTCTA--G-GTGGCA-C-T-AT---TCC-CTT-G--A---------GG----A---A-G---A 6283
RCM.GA.x.GAB1            A...GAG-GAGGA-TAT-T-C---G.............--T-CAA---------T------CGA---G-TGCT---AAT.........---AG-T-ATAT-----AGCAA-A--TA--G-GTGG-A-C-T-AT---TCC-CT-----A-------C-G-----G----T-A---A 6227
DEN.CD.x.CD1             T...GA-------T-T-A-AGG--GGTCTGGATG-A-AATGCCAC-G--TACTTT--C-TC-A-C---ATCCA.........-TA-TAGACCAG-T-GT-CA-A--AT-A-AGAT......C-G-ATA-A--G-C-T-C--A--T-------C--------A--TG-A---- 6597
                                                                                                                                                       ⌐V2 loop start
MAC.US.x.239             A...GATA-TGGTG-TTATTC----............--GGCCC-T----T------GC---G-TGCC---AAT.........---ACAG-CAC------GCAATA-----G--TGGCAAC-C-TT--GACCTCAA------T-----------T-----T-A--CA 6928
Env                      __.__D__N__G__D__Y__S__E__.__.__.__.__V__A__L__N__V__T__E__S__F__D__A__W__N__.__.__N__T__V__T__E__Q__A__I__E__D__V__W__Q__L__F__E__T__S__I__K__P__C__V__K__L__S__P__L__C__
H2A.GW.x.ALI             A...GACA-TGATG-TTAT--G---...........A--GCTT-G-----------GGC---CG-TGCA---GAT.........---ACAG-AAC------AGCAGTA--A---G-CTGG--AC---TT--GAC-TCAA---A-------C--G----A--TT-A--A 7018
H2A.DE.x.BEN             A...GACA-TGATG-TTAT--G---...........A--A-TT-------------GGC----G-TGCA---AAT.........---ACAG--AC------AGCAGTA--A---G-CTGGCA-C---TT--GAC-TCAA---A-------C--GC----A--T--A---- 7028
H2A.SN.x.ST              A...GACA-TGATG--TAT--G---...........A--GCTT-------------GGCC--CG--GCA---AAT.........---ACAG-AAC------AGCAGTA--A---G-CTGG---C---TT--GAC-TCAA---A----C--C---C----A--CT-A---- 6470
H2B.GH.86.D205           A...GACA-TGGTG--TACACT--G...........A-CAG-C----CA-A-----GGC----G-TGCA---GAT.........---ACAG--AC-C-----GCAGTA--T---G-GTGG--AC-C-TT--AACCTC-A---A-------C---C---------G---- 6995
H2B.CI.x.EHO             A...GATA-TGATG--TATACC---...........A-CCAAT-----A-A-----GGC----G-TGCA---GAT.........---ACAG--AC---T---GCAACAA-----G-GTGG---C-C-TT--GACCTCAA---A-------T-------------G---- 6993
H2G.CI.92.ABT96          T...GATA-TGATG-TTACAGT---...........C--GCAT-G-----A------GC----G-TGCC---GAT.........--CACAG-AAC-------GCAGTA--A---G-TTGG-A-C-G-TT--GAC-TCTAC---A--T-----T--------A--T--G--CA 6341
H2U.FR.96.12034          T...GATA-TGGTG-TTATTC---G...........T--GCAT-------C-----GGCC---G-TGCC---AAT.........--CACAG-AAC---G--AGCAGTG--A---G-CTGG---C-T-TT--GAC-TCTA---A----------G--G--A--C--T---A 6482
MAC.US.x.EMBL_3          A...GATA-TGGTG-TTATTC----...........T-GGCCC-T-----T-------GC---G-TGCT---GAG.........---ACAG-CAC-------GCAATA-----CG--TGGCAAC-C-TT--GACCTCAA------T-----------T-----T-A--CA 6053
SMM.US.x.H9              A...GATA-TGGTG-TTACTC----...........T-GGCAA-T-----C-----GGC----G-TGCT---GAT.........---ACAG-CAC------AGCAATA------G--TGG-ACC-C-TT--AAC-TC-A-T----C---------C-C--------A---A 6396
SMM.US.x.PGM53           A...GATA-TGATG-TTATTC----...........T-GGCCA-------T-------GA---G-TGCC---AAC.........---ACAG-CAC-------GCAATA------G--TGG-A-C-C-TT--GACCTCTA---A--C-----------------CT-A--CA 6847
SMM.US.x.SIVsmH635F_L3   T...GATA-TGATG-TTACTC----...........T-GGCA--C---A-C-----GGC----G-TGCT---AAT.........---ACAG-CAC------AGCAATA------G-GTGG-ACC-C-TT--AAC-TC-A-T----C---------C-T--------A---A 6926
SMM.SL.92.SL92B          A...GATA-TGGAG--TATTC----...........T-GGCCC-T-----T-----GGCC--CG-TGCT---GAT.........---ACAG-AACT-----AGCAATA--A---G-TTGG-A-C-C-TT--AAC-TCTA-T--A--T-------G-----T---T-A---A 6355
STM.US.x.STM             A...GACA-TGGTG-TTACTC----...........T-GGCAA-------C-----GGC----G-TGCC---GAC.........---ACAG-CAC------AGCAATA------G-CTGG-ACC-T-TT--AAC-TC-A-T--A--T-----------G--T-----A---A 6570
MNE.US.x.MNE027          A...GATA-TGATG-TTATTC----...........T-GGCCC-T---A-T-------GC---G-TGCT---GAG.........---ACAG-CAC-------GCAATA------G--TGGCA-C-C-TT--GACCTCAA------T-----------------T-A--CA 6396
DRL.x.x.FAO              A...-GTTTATTGC-TTATG-----.........AC--A-A-TAAC---A-AGAG----------CTG-CCCTATGACAG-------GG--A--A-----GCATGG-GT-CGC-GTCATCCA-GATA---GC-GTG--T--A--------T---A-C-AT--TTAT---- 6246
MND_2.x.x.5440           -...-GTCTGTTGC--TATG-----.........CAGCATA-CCCC---A-A-----G-------CAGGCCCTATAGAAG-G-----A--A--GAC---AGCTTGG-GA-CA--TTCTTCCA-GATA---GC-GTAT----A----------GG-----A--TTAT---- 6197
MND_2.CM.98.CM16         -AGCTTGTTG-A-T-TGA-G-GC--.........CC-A-TCC-AAC-TAACATGGA-TTT...C-CAGGTCCCATGGAAG-------GG-A---ATG--AGCATGG-GA-C---CTC-TCCA-GATA--CGC-GTAT----A--T----G---C-G------TA----- 6624
MND_2.GA.x.M14           A...-GCTTATTGC-TTATG-G--G.........CA-C-CA-CCCC---A-A-------------CAG-CCCTATACAAG-G-----AG-----AGG--AGCATGG-GA-C---TTC-TC-A-GATA---GC-GTT-----A--G-----------G--A---TA----- 6553
OLC.CI.97.97CI12         -...T-TTTTGATG-AGATG-GAT-.........-A-CA-CCCA-CCC--GAGT----G-GG-ATTTTATAATGTGTCAG-C-----AG-A---AG---AGCATGGTTT-GC--TGCTTCC--G-TA---GC-GTT--G--A--------G--G-----T--TT-T---- 6116
GRV.ET.x.GRI_677         A...GAT--T-ATG-TAATAC---G...........--GCCTC-A...--CA-T------GC---CG-GGCT---GAT.........---CCGC-----A----AGCAG-GAGTA-----CATCTAC-C-TT--A----CGA-G-G---T-----T--GC-CT----CA-A---A 6520
VER.KE.x.TYO1_patent     A...GAT--T-ATG--TATAC----...........---CCTC-A...---A-C--T---CCA---G-GGCA---GGTGAT-G---CCCAT-AA---C---AGCAGCCAGTA-C--CCATTTAC-C-TT--G----CTA-G--A--T-----G--G---T-A-----A--CA 6604
VER.KE.x.9063            A...GAT--T-ATG-TTATAC---G...........--GCC---A...--CA-T--T--GCCC---G-AGCC---GCAGAT-GG---CCAT-A--G-C----GCAGGAAGTA----TCATTTG--G-TT--A---CAT-G--A-----C---------T-T--T--A---A 6607
VER.KE.x.AGM155          A...GAT--T-ATG--TATAC----...........---CC-C--...---A-A----GCCA---G-AGCA---GCAGAC-G----CC-T-A----C---AGCAGGAAGTA-C--TCA-CTGC-G-TT--A--G-CAT-------C--------GC--T-A--T--A---A 6605
VER.DE.x.AGM3            A...GAT--T-ATG--TACAC---G..............---CCAT----CA-C--T---CCA---G-AGCA---GCTGAC-G---CCC-T-A----C---AGCAGGAAGTA-----CA-CTGC---TT--G--G-CT--G-------------------T-A--TT-G--CA 6105
TAN.UG.x.TAN1            A...GAT--T-ATG-TTATACT---...........---CAG--A...-----CT-T--G--A---G-AGCA---AAGGAC-GG---CCAT-A--G-C----GC-G-GAGTA-C---CATTTGC-C-TT--AAGC-CT--G--------C------C-G--A--TA-G--CA 6603
SAB.SN.x.SAB1C           -...GAT--TGAACCAGA-GG-AC-ATAGCAGAA---CCCA-ACCT---A-C--------A---G-TGCT---AAAAATCGG--CCCTC-A----GC--AGCAG-GAGTA-C---CA-CT---G-TT--AAGT-CTT-------------T--GC-CT-A---A-G--CA 6814
MND_1.GA.x.MNDGB1        T......TCATTGCCAGATT-T--TGAG.......----A-A-TCCTG--A-A-AG----------CAGGACTTATAAGGG-----C-G--A--TT-T--AGCATGGC-T-C---GGGA---A-G-TA---ACC-TA--T-------------GA-T-A----TAT---- 6063
DEB.CM.99.CM40           ACATG-CTTAGTG-GAGA-G---TCCCAATGCCAAAT--GACCCA---C...TTTA-TGC----G-T-ACCCC...............--A-AG--G---C-ATGGC-------G-CT-----G-ACA-A--GTCAT-C--A-----------GC----A---TAT---- 6283
DEB.CM.99.CM5            ACATG--ATGGTT-GAGA-----TCCCTATGCCTAAT--G-CCCAG...TATTTT--CTC----TCA-ATGTT...............----A-C----AT-ATGG--C--C--G-CT--C----ACAGA---TCTT-T--A--C--------G---------TAT---- 6252
TAL.CM.01.8023           A...T-------TC-GAAT-T---G.............--GCCA--C---T----T--GT-C---G-TG---AT.........-----CT-------G--C--C--AGCA--T-----GGA-GCA--G-TCTTAAC-GCAACC-----T--CACC---C-----T--TA-G---- 5872
TAL.CM.00.266            A...T----T--TC-GAAT-TC--G..............---CCA--G---C-C--T--GT-C--CG-TG---AT.........-----T-------G--C--C--AGCA--A--C--GGAT--CC-C-TTTTGAC-GC-AC--GA------ACT--G-----A---A-A---- 6352
SUN.GA.98.L14            A...GATTTA-AA-CTTATGC----...........--GCCCA-AAC-GGGT-AGA---G-------CAGA-G-AATAAGC--T---C-A--A--GC-G--AGCATGGC-A-C---G-C-TC-A-GGTA---GCC-TTA-G-----T--------GA-C-AT--CTAT---- 6743
WRC.CI.97.97CI14         A...CATTTGGAGC-TTTGG----G.........CAGTACA--CCTG-GA-A-A---------C-CA-GAAAAGTAGCAG-----C-A--A--GC-G---GCATGGACA-GAG---CTTCAA-GATT--CACCTTAT----------------A-C--A--CTAT---- 6608
WRC.CI.98.98CI04         A...CATATAGAGC-TTGGG-----.........CA-TATC-TCCT-----A-A-----------CT-G-AA-GTAGCAG-C---C-G--AA-CC----AGCATGGACA-GGG-T-CATC-A-GATT---ACT-TAT----A--T-----T--GA----A---TA----- 6623
WRC.GM.05.Pbt_05GM_X02   T...CATATAGAAG-AAT-AC----.........CA-TATA---A-GG-A-A-A-----------CA-AAAA-GTGGAAG-------GG-A---C-G--AGCATGGACA-GAG-G-CT---A-GATA---ACTCTGT-G--A--C---------A---AT---TAT---- 5546
LST.CD.88.524            A...GATCTG-AATCTTATGCTC-G.........A-CCCTA-TCAT---A-T--TCT--------CTCAAGAAATTAAGG-C---C-A--AA--C-G---GCATGG-GA-C---G-CG--CA-GGT----GCT-TTA-G---------AC----A---AT--TTAT---- 5618
LST.CD.88.447            A...GACTTG-AGTCTTATGC-C-G.........A-TCCTA-TCAC--CA-A-GTTT------C-CT-AAGA-ATAAAAG-T---C-A--TA--C-G--AGCATGGAGT-C---G-C----A-GGT----ACT-T-A-G-----T---------G---AT---TA----- 5618
LST.KE.x.lho7            A...GATTTG-AATCTTATGC---G...........---CCTA-CTAT---A-TT-------C----CA--ACCAGTAAAAG-C---C-GG-TA--C-G--AGCCTGGTCT-C---G-ATGC-A-GGTA--CAGT-TTA-G--A-----------A-C-AT---TAT---- 6696
SYK.KE.x.SYK173          G...T---CAGATC-AAT-G---T-............AGGG-CAAC-T-ACAGG---T-C---CCAGCA---AATAGC-GTC-CATG--AAGGC-G--A--ATTA-----C--GTCAGC-----TCTTG---GC-AAT-GA--C-----G--GC--G-A--CA-G---A 6555
SYK.KE.x.KE51            A...T-TTCTGAAC-GAT-G---TC............AG-G-TAAT-T-AGTGGG---T----C-CTGC----AATTCTTCCC---G---AAG-C-----C-TTTGC----C--GTCTGA-C-T-TCCTG--GGCTGAT-G---T-------------T-T--TA-G---A 6233
COL.CM.x.CGU1            ACCTC-GCCTGGA--G---GT----.........CA-AACA--CC------A------GC----G-TC-T-TT......--G---AG-T-CAGT-GGG--G-ATGGATAATA-C-CAG-C-ACC-TA--A-----G----GA-------CT--G--G--AG--TA----- 5962
```

```
                          V1 loop start
H1B.FR.83.HXB2            TTAGTTTAAAGTGCACTGATTTGAAGAATGATACTAATACCAATAGTAGTAGCGGGAGA.................................................................................................... 6662
Env                       V  S  L  K  C  T  D  L  K  N  D  T  N  T  N  S  S  S  G  R  .__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.
H1A1.UG.85.U455           ---CG--G-T---CA-A-CA-C-CC-T-A-C-AC-CC-AT--C-AC-CC-AT........................................................................................................... 6101
H1B.US.90.WEAU160         ---C------T------A--G----TGTGAC--A-TTG-AG---GAG-C--ATACC-ATAGTAGTAGTGGAGGG..................................................................................... 6672
H1C.ET.86.ETH2220         ---C------C--T-A--C-A-C--A--CA-------GGTT-CA-A--A---T.......................................................................................................... 6065
H1D.CD.84.84ZR085         -C-C------T---------GCATC--GAA-C-GC-C-GAT----AC--C-CA.......................................................................................................... 6177
H1F1.BE.93.VI850          -G-C------C--T---A--GCC-CC---A-C-G-C-GGAA--GCCAG-GGCA.......................................................................................................... 6002
H1G.SE.93.SE6165          ---CC-----C--T------G-A-CC--CA-AGGC--C-AA-GA-A--A----ACTGATAATAGCACTGAAACCAACAAT............................................................................... 6086
H1H.CF.90.056             ---C-C----C--T---A--G-C-GA--CA----CTC--A--GC-C---C--T.......................................................................................................... 5996
H1J.SE.93.SE7887          -C-C------T--T-G-A-CA-C-CT-G-A-C-G------A-C-----AC--TA-T....................................................................................................... 5978
H1K.CM.96.MP535           -C-C------T--T-------AC--AGGAACC-A--GC-------A-GCA-CTA-C-CT.................................................................................................... 5861
H101_AE.TH.90.CM240       ---C------T--T--CA--GCT--TTTGACC-A-GGC-GT-GC-AA-CC-AT-TCTCTAAC................................................................................................. 6233
H102_AG.NG.x.IBNG         ---C----G-T--TCA-A----C--C--CAGCTAC-GC-A--G---C-ACCTAACT--T.................................................................................................... 6189
H1N.CM.95.YBF30           -A-C-A-GCTT--T-AC---AGCT-TGGG--GGAA-GG-A-----CA-A-.................................................................................................ATGACAACAAGA 6252
H1O.BE.87.ANT70           -ACAAA-GG----T--AA-CA-AGCTGGAACA--A---GAA...................................................................................................................... 6692
H1O.CM.91.MVP5180         -ACAAA-G--C--TGTA---C--C-A-CAA---AA-CAGG-...................................................................................................................... 6682
CPZ.CD.90.ANT             -A-AAA-G--T--T--A-GA-AC--TGGAACAC---CA--ACCA----CA-CAACA--TACA..................................................................................GTAACACCAAAGACAACA 6106
CPZ.CM.01.SIVcpzCAM13     -A-CAA-G--A--T-G-A-CG-T-CT-GAC-A-A--GC--T-GC-ACTC--A-............................................................AATAACCCTACCCTCAAGAACAATACTTCAGGT 6267
CPZ.CM.05.SIVcpzEK505     -A-CCC--TCA--T-GCTCA-G--G--G--TG-AC----GT...................................................................................................................... 6193
CPZ.CM.05.SIVcpzMB66      ---C------C--TT-ACTG--T--ATGCAT--AAG-A-ATGGC-AC-CC-C-AACT-C...................................................ACTGTTCAAATTAGCACTGGTAATGATAGCACTGCTAAC 6283
CPZ.CM.98.CAM3            -C-CCA-GG-A--T-GAA-GG-T-CTTTCA---GC-C-T-T-----G-A--AAACTTCTACTATGACTACT.................................................................................AACAGT 6099
CPZ.GA.88.GAB2            -GCAA----C---T---TC-A-A--TGC-AC-GGA--CGAA-C--A-GC--CT--A.....................................................................................AACGGAATT 6073
CPZ.CM.98.CAM5            -C-CCA-GC-A--T--GACCC-T-----CAT--GC-C---T----C--AC--T........................................................................................ACA 6363
CPZ.CM.05.SIVcpzMT145     -G-CCC-T-CT--T-A-A--CCC-CT---AC-T-CTGC--T.................................................................................................. AAC 6192
CPZ.TZ.00.TAN1            -A-AGA-G-CA--T...ACAA---CT---ACC--A----AA-CC......................................................CTGAATTCGGCAACAACAACCTTAACACCAACAGTAAAT 6303
CPZ.CM.05.SIVcpzLB7       ---C----C-T----G-ATCCCT--ATT-----ACTC--G-...................................................................................................................... 6209
CPZ.US.85.CPZUS           ---C-A-G-CT--TCT-A-CCCCG-T-G-A--............................................................................................................AGT 6709
CPZ.GA.88.GAB1            -A-CC--GC------G-A-GGCT--CTT-AGCCAGGCA-AA--CCTA-CA..........................................................................................AAC 6720
MON.CM.99.L1              ----GA-GCTC--TGT---AG-T--C-CC-T-T-G---G---G--CC-C-CCT-CCCCTAGTACCCCCACACCC...................................................................TGGGGAAATTGG 6601
MON.NG.x.NG1              -A--GA-GCG---TCAAARGA-C--TG-AAGC--A-CAGTG-CRCCA-C--C-ACC-CYWCC...GCCACCACACCA.................................................................TGGGGAAGTTGG 5214
GSN.CM.99.CN166           -GG-GA-GT-T--T-ACCTCACA--C-CATCC--A-GCGAACCA-CA-CA-CTCCA--GCCACCAAATGTCTCCACAACCACG.................................................CAGTGGGGAAGTTGGGGG 6590
GSN.CM.99.CN71            -AC-AA-G--A--T-AGCTCCCA--T-CGAGC--A-CA--A-CACCGTCC-CAACA-C-GCCACAACATCACCCTCATCCCCAACGCCAACACCA.........................................TGGGGAAATTGG 6581
MUS_1.CM.01.1085          -A-AAA-G--A--TG-GA-AG-A-C----T-C--C-GAT-T..........................ACGGGAGGGAATAGCACCGGAACCACAGCACCCCCAACAACACCATGGGGATCATGGGGCAAT 6551
MUS_1.CM.01.CM1239        -C-AAA-G--A----G-AGAA-A--TGTAACA-GGCCAT-A............................ACAACAACAACAGCATCAACACCAACAAAAACAACACCAGTGACCCCTTGGGGGAAATGG 6535
MUS_2.CM.01.CM1246        -A-AAA-G-----TGACCT-C-C--T-T-AG--TGTC--GT..........................ACCACCGCAGCAAGCGCCACAACAAGCGGGCATACAACCGTCACCCCTTGGGGCAGATGGGCA 6670
MUS_2.CM.01.CM2500        -G--AA-G--A--T--CA-GC-A--TTCAACA-GGCCG---.....................................ACAAAGCCGACCACGCCCAGTACCACTCCATGGGGAAGGTGGTCT 6596
RCM.NG.x.NG411            -A-CAA-GT-T--T----CA-----TGG-AG-TGGG--GGT-C-GCC-CCCAATCACCTACAACAACAACAAGAGCAACAACAAGAACAACA...AAAGAAATAGGGAAAGATTGTTATGCGAAT..................CAAACAGAGTGCAGT............ 6420
RCM.GA.x.GAB1            ---CCA-G--T--TT-A-CAA-A--TGG-AGCTGGG--GGA-TCCC--CCTCA-CACC-CCAACAACAACAAAAACAACAACACAAAGAACT...ATAGGTGTAGAAAAGGAATGTACTGCTGGC..................AACGAAACATGTGAG............ 6364
DEN.CD.x.CD1              -C--AA-G--T--T--AAGAAAAG--G--A-C-GC-CA-A-GCCGTC-CCGTGTCCTCTACCACAACCACAACCACAACCATAAAGACCCCAACCACA.......................................GAGTGG 6701

MAC.US.x.239              ---C-A-G-GA----A-A-AAGTG---CA----GATGGGGATTG-CA-AATCAATA-C-ACAACAGCATCAACAACATCAACGACAGCA...........................TCAGCAAAAGTAGACATGGTCAATGAGACTAGTTCTTGTATAGCCCAGGATAAT 7071
Env                       I  T  M  R  C  N  K  S  E  T  D  R  W  G  L  T  K  S  I  T  T  T  A  S  T  T  S  T  T  A  .__.__.__.__.__.__.__.__.__.__.__.__.__.__.  S  A  K  V  D  M  V  N  E  T  S  S  C  I  A  Q  D  N
H2A.GW.x.ALI              -AGCAA-G-----T-GCA-CA-A-GC-CA--G-GC-CA---CATCCCCG---CCAG-GAGCACACTC..................................AAACCCCTGATAAATGAGAGCGATCCATGCATAAAGGCAGACAAC 7131
H2A.DE.x.BEN              -GGCAA-G--T--T-GCAGGG-TC-AGGGA----C-CG---CCG-A-CCC--GACCTCGAGTTCCACAACCTCGAGACCACCCACATCC.......................GCAGCCTCCATAATAAATGAAACTTCTAACTGCATAGAAAACAACACA 7165
H2A.SN.x.ST               -AGCAA-GCGT--T-ACAGCACA-CTGCAA-A-AC-CA---TCC-CACCA-CAACC-CCACAACAGCA.................................AACACAACAATAGGAGAGAATTCTTCATGCATACGCACAGACAAC 6583
H2B.GH.86.D205            -GGCAA-G--C--T-G-A-AACCG-A-CAA-CC-AGGG-ATGCC-------CTACC-CCACTAAGCCTACTACCACCTCT.....................CGTGGGCTGAAAACGATTAACGAAACAGACCCATGCATAAAAAATGACAGC 7126
H2B.CI.x.EHO              -G-CAA-G-----T-A-A-GAC-TG--GCTCAG---GC-AAG-G-CC-C--CGTCCTCCGCGTCCCTCAGATCTTCTACTCAG..............................ACCCTACTCAACGAAGATAGCAAATGTATTCAAAATGACAGC 7118
H2G.CI.92.ABT96           --GCCA-G-GA----A-A-AACTG---CAA-C-AATGGGGTCTA-CAG-G-C-ACC-CTGTAACCCCTGCAACAGTTCCCACAACTAAAATGGTGACAGCT.......GAACTAGTGAATAGCACTAGCCAGTGCTTAATGTACGATAAT 6484
H2U.FR.96.12034           -AGCAA-G-GA--T-G-A-AACTG-A-CAA---GATGGGGATTA-CAG-A-CA-CAGC-ACAACAACTAGCCCAGCAATGCAAGAATCAAGT.................AAAGTAATAAATGAAACAGAGCCTTGTATACGAAATAATAGC 6616
MAC.US.x.EMBL_3           ---C-A-G-GA----A-A-AAGTG---CA----GATGGGGATTG-CA-AATCATCA-C-ACAATAACAACAGCAGCACCAACATCAGCA..........CCAGTATCAGAAAAAATAGACATGGTCAATGAGACTAGTTCTTGTATAGCTCAGAATAAT 6202
SMM.US.x.H9               -A-CAA-G-G---T-A-A-AAGTG--MCA-MC-GATGGGGTTTA-CAG-AGSACCAGC-CCAACAACAACACAAACATCAACAACACCACCATCACCAATAATAGCA.......AAGGTTGTAAATGACAGTGATCCTTGTATAAGAAGTAATAAT 6545
SMM.US.x.PGM53            ---CCA-G-GA----A-A-GACAG-A-CA--C-GATGGGGGTTA-CA-AG-AT-AA-T-CCAACAACAACAACAACAACAACAACAAAAAGCCCAAAAGCA...........GAAGCAATAACAGCAAAAGTTATAAATGAAAGTGATCCTTGTATAAGCAACAATAAT 7005
SMM.US.x.SIVsmH635F_L3    -AGCAA-G-GA--T-A-A-AACTG---CA----GGTGGGGTTTG-CAG-A--A-CAGAGACAACAACAACAGCGAAATCAACAACATCAACAACAACAACAACA..............GTAACACCAAAGGTTATAAATGAAGGTGATCCTTGCATAAAAAATAATAGT 7081
SMM.SL.92.SL92B           --GC-A-G--A--T-A-A-AAATG---CA--C-GATGGGGGTTG-CA--AGCA-CT-CTACTACTAGCTCACCAACTACTACTAGCCCCTTAACTGCTGCTAGCCCATCAGGAGAG...........GAAATCGTTAACGACACTATGTCTTGTACAAAGAACAACAAT 6513
STM.US.x.STM              --GCAA-G-G---T-A-A-GAATG-A-CA--C-AATGGGGATTG-CAG-G-AAACAGT-ACAACAGTGACACCAACAGCAGCAGCAGCAGCAACA.......................AAGCCAGAGTTAGTAAATGAAACAAGCTCCTGTGTAAGTAACAACAAC 6713
MNE.US.x.MNE027           ---C-A-G--A----ACA-AAGTG-A-CA---AATGGGGATTG-CA-AATCATCA-C-ACAACAGCAGCACCAACAACAAAAACAACAACAACA.....................AAGGAAATAGAAGTGGTCAATGAAAATAGTACTTGTGTAAATCGTGATAAT 6542
DRL.x.x.FAO               -G-AAA-GG-A--T-G----C-C--A-TA-TA-A-GCAG---GC-CA-CACCAACACC-ACTACA...........................GTAGCTACTACTGTTAATACTAGTGTTGCCTCAACAACAACAGCAAATGACACTTTGGAAATAGCTGTAAATAATACA 6389
MND_2.x.x.5440            -G-AGA-G--A--T--A--AGGTC-A-----A--AG-GCAAGCA-CAGC--AGACC-C-ACACCTGTACCCACAACAACAACCCCCTCCACTACCACT....................AGCTCCAGCACAAATAAGACAACAACTCCTGTGTTGGTTGTTGAGAAACAA 6346
MND_2.CM.98.CM16          -C-AGA-GC----T--AA-AGGA--AG-CCT-G---CCC-T-TACCC-C--CATCC-CCACAACTACCACAACTACCACC...............................AAGACGGTGGCAAATACTACTGAGTTGGACATAGACACCAAT 6746
MND_2.GA.x.M14            -C-AGA-GC-A--T--AA-GGGAG-A--CAG--G--CC-G--C--CC-C-CC-AT-CCCACTACCACCACTCCCACCACTCCA.....................................GCAGTGACGGGGCTGGAGGTGACGGTACAG 6666
OLC.CI.97.97CI12          ---AAA-GC-A--TGAA---CCATCAGCAA--TT--CAGAACC--T--CACCTATC--TACAGCTGCTCCA..........................ACAACTGTCACTAGCAGTACAGGAGGTCAGACCACTACAATAGGGTTAGGAGTTGATAATAGT 6250
GRV.ET.x.GRI_677          ---AAA-GTCC--TGTA--GC----TGG-ACAG-C-CG--A--GGCC-CC-CTACTGC-ACTACAACAATGACT.......................ACCCCCTGTCAGAAT.............................TGCAGTACAGAGCAGATA 6627
VER.KE.x.TYO1_patent      -C-AGA-G--T--TGTA--G--A--TTCCACA-GAG-A-GGGCG-CA-CACCTACA-CGACGCCGAAATCTACCGGC...................CTACCCTGTGTAGGCCCGACGTCAGGT...GAAAATCTACAGTCCTGTAATGCAAGCATTATA 6741
VER.KE.x.9063             -A-AAA-G-GC--TGTA--G-----CTCAAG--GA-GA-A-CCA-C-CCA-CTTCAGCTTCAACAACAAAAGCTCCAAAAACAGGT...................GATCCCTGCATTAAATCTACCAATAATAATGTTAATTTACAGCCCTGTAATGCCTCCCTTATA 6756
VER.KE.x.AGM155           -C-AAA-G--T--TGTA--G--A--AGGCTCCG-A-CCT-T-CCCCAGCA-C-TCT-CTACGGCAGGAACCAAA...................CTACCCTGTGTTAGAAATAAAACAGAC...TCCAACCTACAGTCATGCAACGACACCATCATA 6739
VER.DE.x.AGM3             ---AAA-GTCC--TGTA--A-----CTCCTC-GAGCC-----CC-C-CC--AAA-T-CCACGGCCTCAACAACCAATATCACAGCCTCAACAACC.........ACTTTGCCGTGTGTCCAGAACAAG...ACAAGTACTGTGTTAGAATCATGTAATGAAACAATCATA 6263
TAN.UG.x.TAN1             -A-AGA-G--T--T---A-A--A-CA-G-ACCG-CCC---ATCA----CCCCTACATC-AGCAGCACTACG...................GATCCCTGTCCAAATACC...............GACGAAAGTAGCTGTAACGCCACCTTAGTT 6722
SAB.SN.x.SAB1C            -A-AAA-G--C---TA-CGC--AG-AGGA-GAG-AGCA--A-CA-CATCACCATCA-C-TCAACAGCACGCCCCGAGGTAGTGAGT...................................GTGGGCTTTAATGACTCAGTGATA 6924
MND_1.GA.x.MNDGB1         ---AGA-GC-A--TCAG--AACAG-A----TAT-AGCA--A-CAGC--AGCCTATA-CTACACCTACTACTACATCTACAGTTGCAAGTAGTACA.........................GAGATTTACTTAGATGTAGATAAAAAT 6185
DEB.CM.99.CM40            ----CA-GC------TAA-AACATCA-CAA--C-A-CAC-A-CA-AC-CATCAACA-C-ACAACAACAATAGCAACAACAACA.......................................AAAACAACTACAGATTGGTCA 6387
DEB.CM.99.CM5             -GC-GA-G--C--T-G-AG-ACA-CA-CAACA--A-CAG-AGCA-CA-CA-CAATA..................................................................AAAACAACAACTGACTGGTCA 6329
TAL.CM.01.8023            -G--GA-GGC---T--A--CA-C-CA--AA--CTC-C-GG--CAGTAGCACC-CCA-C-ACAAATTCAACAACAAACTCAGTAACATCC...............................AAGAGAACCAATGACACCAATGTGTGG 5988
TAL.CM.00.266             -A--GA-G-----TGAG--CA-A-CA--GA-CCT--CA-T--G-GCC-A--C-ACA-C-CCAGCCCCGACG...................................................GCCATGAGCAATAGCAATATGCCGTGG 6450
SUN.GA.98.L14             -C-AAA-G-----TC-ACCGAAACC--CAAC-C-CTCCT----CTCA-CAGT-AAA--TAGCTGTGATTATTGGACGACCACCACGGCTAAGACTACAACTCAGACTACAAGTAGTACCAGCTCCACTGCTAGTACAACTACTCCAATGCCTCTAGATTGGAATTGTACT 6913
WRC.CI.97.97CI14          -A-AAA-GC-A--T---CCAGGAG-AGGGAG-GAG-CATTT.............ATCC-TGTGAATCTAATACTACAAGAGCT..................ACTACAAGTGCTACTACAACAGTTGGTACAGTTAGTACTACTCCAGTTATGCAGGTAGAAAACAATTGC 6748
WRC.CI.98.98CI04          -G-AAA-G-CA--T--ACCAGGTGCACCA---T----C-AT-CG-AGGACTTTAACCCCTGTGAAATTAGCACAGGTGCACCTACTATTAGTACCATAAGTGCTGCACCTACTACTAGTACCATAAGTGCTGGTACCAGGACACCAGCTATGCCAGTAGATAATAATTGT 6793
WRC.GM.05.Pbt_05GM_X02    -G-AAA-GTCT--T--ACTAGG-G-TG-GA-A-GC---TAT-CAGAAGAC-TA-AACTGACAACAGTAAAA.................................ACAACAACACCAGTAACTCCGCCACAGAATACTACCCCTAGTATGGATGTAGAGAATAACTCT 5680
LST.CD.88.524             --C-CA-G--A--T--A-GGGAAGGACTA-G--AG--G-GA-CA-CA-CA-CAACA-C-CCAACAACAACACCC...ATGCCTTGCTTTATAACAGAATCTACTACTACTACTACAAAGAGTTCAGGGACTGAACAACCATCTTTAATAGAAGATACAAATTGTACACGA 5785
LST.CD.88.447             -G--GA-G--A--TGAA-GGGAAGGA-TA-GA--A-CA-AAGCA-CA-AA-CAAAA-CG...TCAACAACACCC...ATGCCTTGCTTTATAACAGAAAAAACCACAACCACCACCAAAAGATCTACAACTGGGGGACCAGAATTAATAGAGGACATAAATTGTATTAAA 5782
LST.KE.x.lho7             -G--AA-GC----TGGG--AG-T...............--A--A-CACCA-CAACA-C-CCAAAAACTACTACACAGATGCCTTGTTTTATCAACGAACAGGTA.........ACAGTTAAGAACCCGGGGAATGAAACAAGATTAGAGGAAGATCTTAATTGCACAAGA 6842
SYK.KE.x.SYK173           ----GA-GTT---T--ATTGGAC--C-GCCC-G---CA-GT.................................................ACCCCTACTACAAGTCCCCCTACTACACCTCCAAATGAAACGTGG 6641
SYK.KE.x.KE51             -C--AACGTTA--TGTA--GGAC---G-CAG--G--G-GGT---CT--CC-CAACA-TC...............................................AGTCCAACAACAACAACAAACAAAGCAGCAACACCACCTGCATGG 6337
COL.CM.x.CGU1             C-CCAA-G-TT-----CA-AG-A--T-GGAC-GAA---GG--C-----CAGTA-CCCC-ACAACA.............................................ACAAATAATAGTGCTTCAGATTGGGAT 6054
```

```
                                     V1 loop end   V2 loop start
H1B.FR.83.HXB2          ...ATG...ATAATGGAGAAAGGAGAGATAAAAAACTGCTCTTTCAATATCAGC...ACAAGCATAAGAGGTAAGGTGCAGAAAGAATATGCATTTTTTTATAAACTTGATATAATACCAATAGATAATGATACT.................................AC 6790
Env                      .  M  .  I  M  E  K  G  E  I  K  N  C  S  F  N  I  S  .  T  S  I  R  G  K  V  Q  K  E  Y  A  F  F  Y  K  L  D  I  I  P  I  D  N  D  T  .  .  .  .  .  .  .  .  .  .  .  T
H1A1.UG.85.U455         ATC-CA...GATGGA-T--GG-A---A--G-----T---------------G-C-...---GAAC------A----AAA-------T----T--C---------G----------G---A---CA----AAC-GA-AAC...............................A 6235
H1B.US.90.WEAU160       ...GAA...-AG------G-G-----A--G-----------------------G---C-...---CT-------AA----AGAA---CT---------C-----------------G----G---------C------A-..................................-- 6800
H1C.ET.86.ETH2220       ...--C...-AT-GT-CC--T-AT--A--G-----T--T-----------A-C-...---GAAC------A----AAAAG---G-C-------C--------------------G-----C-TA----C-G-T--...........................-- 6193
H1D.CD.84.84ZR085       CTGCCA...-CGG--A--CC------A--G--------------------A-C-...---GTAG---CG-A----AG-A--C---T-C-----C---------G-------G--G---A-----------GGG-AAAAATGAAATCAAT.........GATACTTATGG 6332
H1F1.BE.93.VI850        ..........................--GC-------T-----------G-C-...---GAAG------A----AA--T----TT--C---GC---------GG---------G---------GC---A-C-A-AGC...........................-G 6112
H1G.SE.93.SE6165        AGC-CT...G-GGATA-CCC------A------------------G-A-C-...---GAA-------A----AAAA----------C--G-----C----G-------G--G-------TA------CA.....................................-- 6214
H1H.CF.90.056           ...........--GGA--CAGG---G--AC---C---T-----------G-A-CT...---GTAC------A----CA--------T-C-----C-C------G-------G--G-------T-----CA---G-.............................-- 6121
H1J.SE.93.SE7887        .........G---GTAGTCCT-ATATC--G-C------------------A-CT...---GAA-------AA---AAG-A--C----G--C--GC----C----G--AA---G--G-------T----G-A-C-A-............................-A 6103
H1K.CM.96.MP535         .........G-GG-AAGCCC--C---A--------T--------------A-C-...---GAA----A--A----AA-A-A--G----C----C----C----G-------G--C-G---C-TA--GGC--GGG-AATAACAGTAGT.....................-- 5998
H101_AE.TH.90.CM240     ATA--A...GG--ATATA-C--AT--AG---G------TA----T-----G-C-...---GAAC---C--A----AA------G-TCC-----C-C--------G---------G---A---T--AG--A-G-AGACTAGT.........................-G 6370
H102_AG.NG.x.IBNG       ............GACAT---T--G--A-----------------------A-C-...---GAAG------A----AAAA----GATGC-----C---------G-------G--G---A---TA--G-AA---A-GGT.............................-G 6314
H1N.CM.95.YBF30         GAACCA...GAC--A-GAT-CAA-C-A--G-----T-----A------GCA-C-...--TGAGC---C--A---AAA-A--C---TT--CT-TC-G------GT-GAA---G--G-------CA--GCC.........................TATAAT-A 6383
H1O.BE.87.ANT70         .....................AACCTT--G--G--G--TGAG--T---G-A-C-...--TGTT--C-A--AC--AAA-G-----A--C-G--TC-A--C---GT-TCA---T-G--GGA-C-GA--G-GACA-GCAGCACAAAT...........AAGACAAACAGC-A 6826
H1O.CM.91.MVP5180       CTAT-A...-ATGA-ACA-T-AAT-----G-G---T--TAG---T---G-A-CT...---GT-C-C-C--AC--AAA-G--C--A--C-G--TC-A--C---GT-TCA---C-G-GTAAGG-TA--G-CTCA-A-GCAGTAAAT...................GGA-- 6825
CPZ.CD.90.ANT           ACACCA...---G-A--TGGCATGA--C--C--G-A--TAAC--T---CAG---...---G-AT-T-A--A----AAA--A---ATGA-A--CA-A---------GGA---C-T--GAAGTGTC-GG-CA-C-A-GAGACTAAC.......................... 6244
CPZ.CM.01.SIVcpzCAM13   AAG-AT...GA--CT-GAGTTTT-C-A--G--------TA-A--T--C-CA-CA...---GAAT-G----A----AAA-A-AC-G-TC---T----A------GT-GA----C-GCA-AGCC---GA-G-.....................GGAACTGGTGA 6401
CPZ.CM.05.SIVcpzEK505   GTT-AC...CA--CAA-CC-C-T-C-A--GC----T-----A--T---G-G-C-...--TGAGT------A----AAA-A--C---TC--CT-TC-G-------TGGGA-----------CT----C-C-A---A-AGCAGTGGT.................AATAAC-G 6339
CPZ.CM.05.SIVcpzMB66    AAC--T...-CGG-T-GC-C-AT---C--GT-C-----T---------GCA-C-...---GAAT------AC-GAAA-A-AC-G-TG---T----G-------GG-AA---C-TGAG---C----AGGAA---AGCCACCAGAA...........GGAGATAAGAACG- 6435
CPZ.CM.98.CAM3          CCT-AT...GA--AAAYAG-CA-CACAG----G--T--TA-C--T-----G-CT...--TGAGG------A----A-A-A-A----T----T-CC--------GT-GA----C--G--TTG--T-----C-----CGAT............................. 6231
CPZ.GA.88.GAB2          GAA-A-...GATGG-TT-GC-CA---C--G-G---T--TA----------CA-C-...---GAAT------A----AAAA---C--ATT---T-C--A--C-GG---AA----C-TG--GGT-CTA-C---.............................. 6196
CPZ.CM.98.CAM5          GAGT-A...-A--ATATA----A-ATA--G-----T--TA-A--T-----G-CT...--TGAAG------A----AAA-A--G---T----T-TC-A------GT-GA-------G----T--T-GAGG-A---AGAGTAATGAT....................... 6501
CPZ.CM.05.SIVcpzMT145   TCT-CT...GACGATCGCTTG-----C--G-G---T--TAGC--T---G---CA...--TGAAC------A----AAAAGAC---TG---T-C--A--C---GTGGAG-------C-G------GA---...........................AAT-G 6320
CPZ.TZ.00.TAN1          TTG-GT...TCT--ACCT--CTAT---G-GT-T--T--T--A--T---CAG-CA...--TGAGT-T----A----AAAA-AC--AT----T-C--G-------G-GAA-----TG--AA-GAG---GG-A-C.......................AAT-A 6434
CPZ.CM.05.SIVcpzLB7     ...--T...-AT-GCTCT--T-AGACT--G-----T--TAGC-------CA-CA...---GAAC------A----AA-A-AC-GACT--CT----G------GT-GA----T---CT-AG---A-C--G.....................ACAACCGA 6337
CPZ.US.85.CPZUS         AGTGCT...G---ATACT-CT-ATATA--G-G------T-----T-----A-CT...--TGAAT------AC--AAA-A-AC---TG---T-C--A------GT-GA----C--GCT-AT--CA-----A-............................ 6835
CPZ.GA.88.GAB1          CAG-CA...TCTTCTCC-CCTCTC--A--G--------T--A--T---G-G-CA...---GAAT------A----AA-A-AC-G-TC---T--C-A------GTGGAG---G-GG--AACC---GG-----G-ACAAC........................... 6852
MON.CM.99.L1            GGAGG-...-ATGGAACAGG-CAGCCAG--T-T--T-----A--T--CCAG-CT...---GAAT-C--G-A---AAA-A--C--ATG---AGC--A----GG--GGAA--C-----GAAGGAG--AGGAAGC-A-GGC..........................-G 6735
MON.NG.x.NG1            GGAGGC...-ATGG-AC-GG-CAGCCTC-TT-C--T--T--A--T--CCAA-C-...---GAAT-T----AC--AAR-A-AC--ATG---AGC--G---KGG--GGAG--C--T--G-A-GACA-G......................GATAACACAACTGATGA 5351
GSN.CM.99.CN166         GGGGAA...-ATGGAACAGG-CA-CCCT--T-T-----T--C--T---CAG-C-...---GAAT-T----ACC-AAA-A-AC-GATG---AGCC-C--C-GG-G-GAA--C--T--GGA-GAGACCC---G-.........................AACCAG-G 6724
GSN.CM.99.CN71          GGAGGT...-ATGGAACAGG-CAGCCA---T-T-----T--C--T--CCAG-CT...---GAAT-T--G-AC--AAA-AGAC-GATG---AGCC-C----GG---GA------T--GAGGGC-C-GG---G-.........................AATGAC-- 6715
MUS_1.CM.01.1085        GGC-AT...GAT-GT-CTGG-CA-CCT--GTTC--T-----C------CAG-C-...---GAGT-C--G-A---AAA-AGAC-GATG---T----G--C-T-GTTGA----C-G--GAG-GACAGCTCA--G-G-GAGAATGAT........................ 6689
MUS_1.CM.01.CM1239      GACGCC...-ATGGTACAG-GCA-AGC--GTTC--T--T--A------CAG-CT...---GAAT-C----AC--AAA-A--C-GATG--CT-C--A-----C-CTGA----C-G--G-GTGAG---------........................... 6661
MUS_2.CM.01.CM1246      GAC-AT...-ATGA-ACAC--TTGACC--GT-C--T--T--C------CAG-C-...--CGAGT-C----A-GTAAAAA-AC-GATG--CT-TC-C--C-T-GT-AAA--CC-C--GAA-GGG--G-----G........................... 6796
MUS_2.CM.01.CM2500      GATGGC...-ATGATACA-TGCA--TC--GT-C--T--T--C------CAG-C-...---GAGT-C--G-AC--AAA-A--C--ATG---T-TC-G----T-GTGGAC---T----GAGGGAT-GG-GC---........................GA 6724
RCM.NG.x.NG411          .........-GTG-CCCTG-C-C---TG-C-TG--T--TGAA--T-CGG-GGCA...GGGTTA-A-----A-G-AAAAAT-----T-A---ACACA-GG--CTC-AGA---C-CTGGTGT-AG-CA......................GGTAATAATAATACT-G 6551
RCM.GA.x.GAB1          .........GA-G-TC--G-T-C---TG-G-TGTCT--TGAA--TGC-G-AGCA...GG-CTA-AG----A-G--AAA--C--GT-TA---ATACC-GG---TCTAGA--CC-TTGGTGTGA-A-G..............GAAACAAATTCTACAAATTCTACAAAA-A 6507
DEN.CD.x.CD1            TATGGA...GG--ATA---CTACCATCT-CCTT--G--TAA---------G-C-...---G-A-A-G-CTT--G--ATA---G--T-ATGTATGAAGC-C--TGTAGAATAGA-GATGT---G--A.....................GATAGGGAAACTAACCAA-- 6841
                 V2 loop end                                                                                          V3 loop start
MAC.US.x.239            TGC-CA...GGCT----AC---AGC-A--G-T--G---TAAA-----C--G-CA...GGGTTA-A-----AC---AAAA-AG-GT-CA---A-AC--GG--CTCTGCA---T-GG--TGTGA-C-A.....................GGGAATAACACTGGTAATGAA-G 7214
Env                      C  T  .  G  L  E  Q  E  Q  M  I  S  C  K  F  N  M  T  .  G  L  K  R  D  K  K  K  E  Y  N  E  T  W  Y  S  A  D  L  V  C  E  Q  .  .  .  .  .  .  .  .  G  N  N  T  G  N  E  S
H2A.GW.x.ALI            TGCCCCAGGGG-C-A-G-G-T-A------GGTC--T--TCGG-----C--G-CA...GG-TTACAG---A----CCAA-AC-GT-TA--A-ACA-GG--CTC-AAA---G-GG-TTGTGA-CCA....................TTTAACACCACCACAAACCAG-- 7277
H2A.DE.x.BEN            TGCGCA...GG-T-A-G-T-T-AG-----G-TGC-A--TGAG---------G-AG...GGGTTAGA-CA--A----AAAAG--GGT-TA-G-ACACA-GG---TT-GAA---G-GG-TTGTGACA-C..........................ACAACAGCTGG 7296
H2A.SN.x.ST             TGC-CA...GGGT---GAG---A------GGTCG----TCAG---------G-CA...GG-TTAGAG--G-A----AAAA-ACT-T-TA--A-ACA-GG--CTC-AAA---G--G-CTGTGA-TCA....................AATGACACCAAGAAAGAG-A 6723
H2B.GH.86.D205          TGC-CA...GG-C-A-GAG---AG--A----TGC-A--TAA---T-G---G-CG...GG-CTA-G-----A-G--C-AA-AC--T-TA-A-ACACC-GG--CTC-GAA---T--GAGTGT-ATA--......................ACCAGGAAGTATACC-G 7263
H2B.CI.x.EHO            TGTGCA...GGG--A-GACT--A---A--G-T-G----TCAA-----A--G-CA...GG-TTA-A-----A-G--TCAA-AC--T-TA-G-ACACC-GG-------AA---T--G-GTGTGA-A-GGGGACA-GGAGTAAT......................GAA-G 7258
H2G.CI.92.ABT96         TGT-CA...GGT--AC--TC--A-TCC--GGTTGG----AAA--T--C--G-CA...GGGCTA-A---G-A-C--AGAA--G--T-CA---AGAC--GG---TC--AA---T-GCA-TGYGA-C-G....................TCCAATAAGAGTGAGAATGAG-G 6627
H2U.FR.96.12034         TGCTCA...GG-T-A--AC---A-CCAT-GGTG-GT--TAGA--T-----G-CA...GGCTTA-A---G-A----AAA--AG-GT-TA-A-A-ACA-GG---TCT-A---C---G-TTGTGA-C-G.....................AATACTAGTGAAAGTGAA-G 6756
MAC.US.x.EMBL_3         TGC-CA...GGCT----AC---AGC-A--G-T--G---TAAA----CC--G-CA...GGGTTA-A-----AC---ACAA--G-GT-CA---A-AC--GG--CTCTACA---T-GG-TTGTGA-C-A....................GGGAATAGCACTGATAATGAA-G 6345
SMM.US.x.H9             TGT-CA...GGCT----AC-G-AGCCC--GGT--GT--TAAA--T--C--G-CA...GGGTTA-A-----AC--AAA-AGAG--T-TA---A-ACA-GG---TC-AGA---T--G-GTGTGA-C-A.................AATAGCAATGGAAATGAAACTGAC-G 6691
SMM.US.x.PGM53          TGT-CA...GGCT-A---C-G-A-CCA--GGT--G---TAGA--T--C--G-CA...GGGTTA-A--AG-AC--AAA-A-AG--T-TA---A-ACA-GG---TC-AGA------G--TGTGA-C-A....................AGTACTACTAGTAGAAATGAA-G 7148
SMM.US.x.SIVsmH635F_L3  TGTGCA...GGCT----AC-G-AGCCC--G-T-GGT--TAAA--T--C--G-CA...GGGTTA-AG--G-AC--AAA-ATAG--T-TA---A-ACA-GG---TC-AGA---T----CTGTGAGC-G......................CCAGCAAATGGAAGTGAG-G 7221
SMM.SL.92.SL92B         TGCTCT...GGC--A---C-G-A-CCA--G-T-GGT--TCAA--T--C--G-CA...GG-CTA-A-----ACC--AAAAG-C-GT-CA---A-ACC-GG--CTCCAGA---C--G-CTGTGAGC-A........................GGAGGAAATGAAAGC-G 6650
STM.US.x.STM            TGC-CA...GGCT----AG---A-TCAC--GT-GGT--TAAA--------G-CA...GG-CTA-A-----A---AAA-AGAG-GT-CA---A-ACG-GG--CTCTAGC---C----CTGTGA-C-A.....................AATGTCACTGGCGAAGAG-G 6853
MNE.US.x.MNE027         TGC-CA...GGCT----AC---AGCCA--G-T--G---TAAA-----C--G-CA...GGGTTA-A-----AC---AAAAGAG-GT-CA---A-AC--GG--CTCTGCA---T-GG-TTGTGA-C-A....................GGTAATAGCACTGAAGATGAA-G 6685
DRL.x.x.FAO             GAACCA...G-------A-C-AACAGAG--TGT--A-A-AA-G-G-CAGGATTA...TGC--GGACT-TAAGG-A-A-ATC-C----A-CTTTAGG-A-G--G-TG-A-T-TGT-A-TGTGGCA--G-AAC-GAATCACCAGAAAGCAATAACTGTACTAATAAAAAA-G 6553
MND_2.x.x.5440          AAT-AT...GAG-CAACA-C-CAGC-A-AT-G-GTA--TAAA-----C-CG-CA...GG-TTATGT--G-ACTGCAAATT-G--AT-G-G-A-AAC---AGAT-TGAA---G---CTTGC-C-A-GCT-A---AAACTGGTAGTGCTACTAATAGTACTGAACCTGAATA 6510
MND_2.CM.98.CM16        AAC-CT...GA--CTACA-C-CA-C---AC-GGGT---TAAA-------CT-CA...GG-TTATGT--G-A-TGCAAATT-G-GAT-G-A-A-AA----AGAT-TGAA---G---CTTGT.........................ACAAATGGCACAAATAACACATA 6883
MND_2.GA.x.M14          AAT-AT...GAG-GTACC-C-CA---A-AT-GGGTA--TAAA--T--C-CA-CA...GG-TTATGT--G-A-TGTAAAATAG-GATTA-A-AGAGC---AGAT-TGAG---G-G-CCTGT....................ACTAAAAAGGAAAATGACACTGAAACTGA 6809
OLC.CI.97.97CI12        ACAG-A...-AGCCA--A-C--A-A--C-T---C-A--TAGG-AT-----T-CA...GGGTTATGT----ACTGTAAAATAG--ATGA--CA-ACA---G--TTGGAA---G--GC-TGT-C-A----C.................................... 6373
GRV.ET.x.GRI_677        GAAGGA...GA-----CAG-G-A-CCAGC-TCC------A----TGCA--TGCA...GG-TATCA-----A-GTAAAAA----TT-TAGCATGACC-GG---G-T-AG--GT--G-CTGC-ATA----AACAGGAAGTGAAAAGGGAAGTAAG................. 6774
VER.KE.x.TYO1_patent    GAA-G-...GAG------G-T-AGCCCGCCTCT--T--TA-A---GCA--GGCT...GGCTATG------A-C--AA-A-A--TT-T-C--TGG-G-GGA--G-TGCA--A--CTATTGC-A-A----GAC--A-AGCACTAGCAAA...................... 6882
VER.KE.x.9063           GAAGA-...GA-C-----G---A--CAGCTTC---T--TA-A--TGC---GGCA...GGGTA---T----A-C--AA-A-A--TT-T-CA-T-G-A-GGA--G-TGCA--G--CTATTGT-AGA-C-GCTC--GCTCTAACTCTAGCACAAAA................ 6903
VER.KE.x.AGM155         GAA-A-...GAG---A-TG-C-AG-CAGCGTC-------A-C--TGC---GGCT...GGGTA---T--G-ACC-AAA-A----TT-C-CA-T-G-A-GGA--G-TGCA--A--CT-TTGT-AGCG--G-ACAT-GCATAATGGGACAAAA.................... 6883
VER.DE.x.AGM3           GAA-A-...GA-T-AA-TG---AGCCTGCTTCT--T--TA-A--TGCA--GGCA...GGGTATG------A-C--AAAA----GT-T-CA-TGG-G-GGA--G-TGCA--A--C--GTGT.........................AAGAAGGGTAACAATTCTAAC-G 6400
TAN.UG.x.TAN1           ACA-AT...-GC-----TT-T-AGA-T-GTTCT-TA-----C--TGC---GGCA...GG-TAT-GG----A-GTAAAAA-----T-TA--AGTAC--GG---G-T-AG--GT-GG--TGTGAGA-GG-AA-C-AC...ACCACAGGCACGAGA................. 6866
SAB.SN.x.SAB1C          GAACAA...GAG-----A----A-C--GCC-TG-----T--C--TGC---GGCA...GG-TAT-G---G-A-GTAAA-A-A--TT-C-CAA--G-A-GGG--G-T-AA--GG-GG-TTGTGA---GGGAAGGGAAAAATCAAACGCAACA............CATACTGT 7076
MND_1.GA.x.MNDGB1       AAT-CA...GA-GAAA--GT--AGAG--ATC-TGTA--TAGG-AT--C--A-CA...GG-CTATGC--G-A-TC-AA-G-AG--AT-GTAA--AA----AGAGGGGA----G-G-A-TGTGA-A--...................................-A 6307
DEB.CM.99.CM40          GGGGAA...-AC--AACC-TGAC-C-ATATTGG--T--T--A--T---G--TCA...GGTCCTTAT----A----AAAG-A--GAGC-CA---G-A-GGCTAG-GGA------CCAGTGGGCG------A-AGA-GGTAGTGGGAACAGA..................... 6531
DEB.CM.99.CM5           GGAGA-...-AT-A-ACAGTGAC-C--TATTGG--T--TA-C------G-T-CA...GGGCCATAT--G-AC---AA-G-A-TGC--AG----G-G-GG-TAG-GGA---C---CAGTGGGAGACAG-AA-CCAAACAAATACTAACAAA............AGTATA-G 6481
TAL.CM.01.8023          TGGGGA...-AC-AT-CC-C--A-CCAG-GT-C-----TA-A-GG-----G-CA...GG-G-AT-T--G-A----AA-T-C-CCT-T-T-T-TCA--GG---CT-GCA--CT----GAAGGA-CCAGGCA--GGCAAT.........................GACT- 6125
TAL.CM.00.266           TGGGGA...GAC-ATACA-C--A-CCAG--T-T--T--TA-A-GG--C--G-CA...GG-GCAT-T--G-A----AG-T-T-CCT-C-T-T-TCAC-GG--CCT-GCA--CC----GAAGGA-CCAGGCA--GGATCC.............................. 6582
SUN.GA.98.L14           GAC-CA...GAG-ACATAGCG-AGAGC-AT---GTA--TAAA-AT---G-G-CA...GGGCTTTGT----A-TGTAAAAC-G-G-TGG-ACA-AAC---AGGG-TACA--AG-T-CTTGC-AT-G------C.................................. 7039
WRC.CI.97.97CI14        TCACA-...G-T-ATACTTC--AGATT-ATC-TGTG--TAGA--T--C-C--CA...GG--TGTGT-AG-ACTGTAGAATAG--AT-A-A-A--C----AGAGG-GA----G-GG-TTGT-ATA-----AG-.................................. 6874
WRC.CI.98.98CI04        ACA-CA...G-G-ATACCTCT-AGATA-ACC-TGTA--TAGA--T----C--CA...GG--TGTGC-A--A-TGCAG-ATAG--ATTA-A-A--C----AGAGG-GA----G-GG-TTGC-ACA-A----GG---............................. 6922
WRC.GM.05.Pbt_05GM_X02  ACTTCC...T-G-ATCTC-C--ACACA-ATC-TGT---TAGA--T---G-A-C-...GGGTTATGC----A-TGTAAA--AG--AT-GTA-AG-CC---AGAGG-GA----G---G-TGC-C-C-AC--.........................GGTAATGGTACACA 5817
LST.CD.88.524           TGGT-A...-ATGA-ACT-C--A-ATT-AT-GGG-A--T--A-AT---G-G-CA...GGGCTTTGT----A-TGCAA-G-AG-GAT-A-GCA-AA----AGGT-TGA---G-C--C-TGT-AT--G----G-................................ 5911
LST.CD.88.447           TGGG-A...-ATGA-ACT-CT-AGATT-AT-G-G-G--TAGA-AT---G-T-CA...GG-CTATGT--G-ACTGTAG-ACAG--AT-A-GCAGAG----AGGT-TGA---G-C--C-TGTGATA-G----GA................................. 5908
LST.KE.x.lho7           GGGC-T...-ATGA-ACC-C--AGAG--ATGC-G-A---CAG-AT---G-A-CA...GGGTTATGC----ACTGCAG-AC-G--AT-A-ACA-AGC---AGAT-TGA----G---CGTGCTC--GAG-GAGG..........................GAGAAC-G 6976
SYK.KE.x.SYK173         TGGGGA...GAC-ACAGT-C--AGCCA-G-TTC--T--------T--CT-G-CA...GGGG-GT-C-AG-A---AAAA---C--T-TAGA--T--C--C-----GGA---CT-G--GAA-GAG--A..........................GGCAATAGT-G 6772
SYK.KE.x.KE51           TGGGG-...GAC-ACTCT-C--A-ACT-GGT-C--T--T--A--T--C--G-CA...GGGG-TT-T-AG-A----AAA--TC-GT-TA-------C-----C--GTCA--------GAG-GAG--A..........................GGAAATGAA-G 6468
COL.CM.x.CGU1           GAGTCA...-ATTG-A-AG--TATCC-TGGT-T--T---AGAA-G---TCA-CAGCTTTCTTAC-T-A--A--GAAA-G--TT-----TA-G-----CAGTAG--GA-TTA-CTG-G-TTGG-A----AA--GACAGTAATAGT.......................-T 6197
```

```
                        V2 loop end
H1B.FR.83.HXB2          CAGCTATAAGTTGACAAGTTGTAACACCTCAGTCATTACACAGGCCTGTCCAAAGGTATCCTTTGAGCCAATTCCCATACATTATTGTGCCCCGGCTGGTTTTGCGATTCTAAAATGTAATAATAAGACGTTCAATGGAACAGGACCA......TGTACAAATGTCAGCA 6954
Env                      S  Y  K  L  T  S  C  N  T  S  V  I  T  Q  A  C  P  K  V  S  F  E  P  I  P  I  H  Y  C  A  P  A  G  F  A  I  L  K  C  N  N  K  T  F  N  G  T  G  P  .  .  C  T  N  V  S
H1A1.UG.85.U455         T--T----GAC-A-T--A------T------AC------------T----C----G-----------------------------------A------------------G----GG--CCTGA----------A---G--.......--C-GG--------- 6399
H1B.US.90.WEAU160       A-------C-----T--A------AT-----AC--------------------------------A---------------------------------------------G------G------A-----------AG-----C........--A---------- 6964
H1C.ET.86.ETH2220       TGAT----GA--A-T--A------T------AC---A-----A--------------C--T--G--C--------T-------------T--A-------A-----------G----GAG-------A----C--------------......--CCAT--------- 6357
H1D.CD.84.84ZR085       T-CT----G---A-T--A------T-------C---------A--T--------A----------------------------------A-----------A--------G-----CG----A-G---------------C---.......--C-A---------T 6496
H1F1.BE.93.VI850        TGAA----G-C-A-T--A------T------AC------------T--------------T-GG--T----C--T-------------T--A-------A--A----------------G----A-GA--------G-----G---.......--C-AG--------- 6276
H1G.SE.93.SE6165        T-AT----G-C-A---CA------TGT----AC---C-A------T-----------A-T-----C-------------------------T--AA-------------T----G----GGG-----GA-------------------......----A--------T- 6378
H1H.CF.90.056           TCAG----G-C-A-T--A------T---------------------T--C--------G--------A--T-------------------T--A-----C---------------G-----C-----A--A---------------TT-......---------------T- 6285
H1J.SE.93.SE7887        A-AT----TTC-A-T--AC-----T-------------A---A--T----------G--T---C-A-------------------------A--------------------------G------AT---------------T--......--C-A--------T- 6267
H1K.CM.96.MP535         TGAA--C-G-C-A-T--A------T------AC-----------A-T--------A---A-------------------------------A-----C---------------G-----GG----A-G--------G-------------......----A---------- 6162
H101_AE.TH.90.CM240     TGAG----G---A-T--A------T--T----------AG-----T---------A---------T---------T-----------A-T--A-------A------T----G------G------AT--------G-----G---......----A---------T 6534
H102_AG.NG.x.IBNG       TCAG----G---A-T--A------T-------C------------T---------------------------------------------A-------------------G-----GG----AGG---------------G---......--C-A---------- 6478
H1N.CM.95.YBF30         A-CA----G-C-A-T--A------T---A---CTG-G-----A--T-----T---ACT-----------------A--------C-----A--AC-A--C-----C---A-G---------G-AGGA-AC--T-G-----AT---AGC..............-----G--T- 6547
H1O.BE.87.ANT70         A-TG----CA--A--T-A-------T--A--AC---C--G--A-----------------T-----A--------A-----C--------T--A--A--A-A---T--CT-T--G-----C-GC-CAGAA--T-----------CA--.......--C-G---CA-A-CGG 6990
H1O.CM.91.MVP5180       A-CA----T---A--T-A-------T--A--A-T--C-AG----------G------AGT--------C----------C--------T--AA-A--A-A---C--CT-T--G------G-C-CAGAC--T-----------C-T-.......--CCAC---A-TTCAG 6989
CPZ.CD.90.ANT           .T-----T-C--ATGGCAC--C------A--ACT--C-----AT----GA---TCTA-T-----A-------T--------------T--A--A--A-A---T--AT-G-G----G-AG--G--GAT--T-CA---GT---GATG.........----A-------TCAG 6407
CPZ.CM.01.SIVcpzCAM13   T-CA----CAA---TT-A----------A---C---A--------A--------A--G-----------T----A----C--------A--A--A--A-----A-----C----------G--GTA-AC--TTCA--G-A---AA-..........----G------A---- 6565
CPZ.CM.05.SIVcpzEK505   TCAA----G---A-T--A----------A---CTG-G---------------T--AA-T-------------------A---T----------A--AC-A--A-----C---A-----------G--C-AGAT--T--------T--GGA-.........----AC-----A--T- 6503
CPZ.CM.05.SIVcpzMB66    -TTA----GA--ATAT-A----------A---C---G--T--A--T--CT----A--C--TC----A-------T---------------T--A-----A-----CC-CT-G--G-------G---A-AT--T-C---G-T-----AG.........----A------T--T- 6599
CPZ.CM.98.CAM3          .-CA----GAC-A-T--A---C--T---A---CT-----------G-----T---ACC-----------------A--------C-----AA-CC-A----A---A--CA-G-----C----TGCCA-AC--T-------------A--GGAAGA----AC---A-A---- 6400
CPZ.GA.88.GAB2          .-CG-T--G---A-TT-A---C--T--TA---CA--C-----A--T--------AACC--------A--------A--------C-----A--A-----C-----CC-GT-G-----------G----AGAT-ATCCA----A----AAG..........----A------A--T- 6359
CPZ.CM.98.CAM5          .-C-----GA--A-T--A------T--TA--ACTG-A--------A-----T--AACT-----------------A--------C-----A--CC-A---------A--AA-G-----------G--C-A-A---T---------------.............------A-A---- 6664
CPZ.CM.05.SIVcpzMT145   T-CT--C-G-----TT-A------T---A---C---C--T--A--T--------ACT---------A--------A--------C-----A--A--A--A-----AC-C------------G---TAGAT-A---A--G-AT-A-A-T..........----A------A--T- 6484
CPZ.TZ.00.TAN1          T--T---T-T--ACAT-A---C--T---------------T--A-AA---GAT--ATCTA-T-----A-----------CAGA--C-----T--A--A--C-----CC-GT----------GAG--C---AT----CA--G-A-----A-........--CT-C------TCAG 6598
CPZ.CM.05.SIVcpzLB7     G---------C-T-T--A------T--TA---C---A--C-----T--C-------A-T--T------------A--T-------------T--A-----------TC------GG-----C-----A--A-AT-----T-GT--T---..........----AC-----T--T- 6501
CPZ.US.85.CPZUS         .-C-----GAC---T--A---C------A-C-CT--C---------T-----T---ACC-----------------A-----C--------A--AC-A--C-----C--C--------------G-A--AGAT-----A----AG---GAG..........----A------T--T- 6998
CPZ.GA.88.GAB1          .-CA----G-A-A-TT-A---C--T--TA---C---A-----A--A-------AACG-----------T-----A-----C--------A--A--A--G-----A-----T-----C---G----AGAC--TTCA----A---TAA-.............----T-----A--T- 7015
MON.CM.99.L1            TCA----T-CA-CCTG-A---C-----A---TA---A--C--A--T---GAG---TCCAAT-A-------G----TT-G-----------A--CC-A--C-A---TT-AT-G-G----G--G-CCCCG-C--T-CG---CA---GAGT..........--T-T-----TTCTG 6899
MON.NG.x.NG1            -TAT---T-TC--CT--A------T----Y-TA---Y--C--A-----GA----TCCAAT-A------T-----T--T-----C-----T---C-C--C-A---T--AT-G-G----G--G--CCAK-C----CA--GCAGAACAAT........--CT-R------TCAG 5515
GSN.CM.99.CN166         TG-A---T-TA-T-G--A------T------TAT--C--T--AAAA---GTC---TCC-----CC-A--GG----------C--C-----T--AC-C--C-A---T--G--------C---G--GTT-AT----CA---GT---CA--.........--CTAC-----GTCAG 6888
GSN.CM.99.CN71          TGA----T-TA-T-T--A---C-----A---TATG-C--G--AAAG---GTC---ACC-----CC-A--GG----------C--C-----T---C-C--C-----T--G--------C---G--GCC-AT----CA---GTG--CAAG.........--T-C-----ATCAG 6879
MUS_1.CM.01.1085        .--T---T-TA-C----AC--C-----A---TA---C--C-----A---GAG---AGCAG---CC-A--GG----G-----A--C-----A--C--G--G-ACAGCT-G---------C---G--GCC-AT--T-C---GCA----GA-.........----AT-----G-CTG 6852
MUS_1.CM.01.CM1239      .-CT---T-CA-ATT--AC----------TA---A--T---CAA---GA----AGCAG---CC-A--GG----G-----A--C-----A--T--A--G-A-T-TC-AT-G----------G--GCT-AT--T-CA---CA---AA-..........----AC-----G-CAG 6824
MUS_2.CM.01.CM1246      .-CA---T-CA-T----AC--C-----T--TTAT--C-------CAG---GAG---AGCAG---CC-A--GG-C--A-----A--C--------T--A--A-A-AGTC-CT-------C---G--GCT-AT----CA---CA---CATT..........----AT-----A-C-G 6959
MUS_2.CM.01.CM2500      -G----CT-TA-C----A---C-----A---TAT-----C--ACAG---GAG--AAGCAG---CC-A--GG-------------C--------A--A--A-A-AGCT-GT-----------G--GCT-AT----CA---CA---CAA-..........----GG-----A-C-G 6888
RCM.NG.x.NG411          GCAA-G-TTTA-A-GGCA---------AA--AGT--ACA----TT----GA-CCAAA--A--GG--A---T--AGGT--AGG-------A---C-A----A---C--A--G------CT-G--GCA-AT-AT-CA--CTTT-ATA-C..........--CA------T-CAG 6715
RCM.GA.x.GAB1           G-AA-G-TTTG-A-GGCAC--C--T--AA-TTC---ACA---ATTT---GA-CCAAAG-A--GG--A---T--AGGT--AGA--------T--AC-A--G-----CT-A--GGTC--C--AG----A-AT-AT-CA--CTTT-ATA-C..........--GTT-----T-CTG 6671
DEN.CD.x.CD1            ATAT---T-TA---GGTA---C--TCAA--CAC------C--A--T---GA------TGT------A---T-CAAA---AGGATG-----A--A-----C-----TC-C--T--G------G--C-TC-T-GG-G---GTAT--C-TT..........----A----A-GTCAG 7005

MAC.US.x.239            T--A-G-T-CA---ACCAC--------T--T--T--CCA-G--T-T---GAC--ACAT-AT-GG--TG-T---AGAT-TAGG---------A--TC-A----A---TT-G--T-G-------G-C-CA-AT-ATTCA--CTTTATG--TAAA...---T-T--G--GGTGG 7381
Env                      R  C  Y  M  N  H  C  N  T  S  V  I  Q  E  S  C  D  K  H  Y  W  D  A  I  R  F  R  Y  C  A  P  P  G  Y  A  L  L  R  C  N  D  T  N  Y  S  G  F  M  P  K  .  .  C  S  K  V  V
H2A.GW.x.ALI            ---G-G-T-CA---ACCA---C-----A--------C---G--T-A---GAT---CAC-AT-GG--TG-T--AAGGT-TAGA--C-----A--AC------AC--CC-A---G---CG--G---TC-AT-ATTCA--CTTT-C---CAAT...--CT-T--A--AGTAG 7444
H2A.DE.x.BEN            --CA-G-T-CA---G-CA---C-----A---A----C-A-G--T-A---GAT---CAC-AT-GG--TG-T--GAGGT-TAGA--C-----A--AC-G--C-----CC-AT---G-------G---CC-AC-ATTCA--CTTT-A---TAAG...--C--T--A--AGTAG 7463
H2A.SN.x.ST             A-CA-G-T-CA---ACCAC--C-----A--------C---G--T-A---GAC---CAC-AT-GG--TA-T--GAGGT-TAGA---------A--AC-G---------CC-G----G---C---G---CC-AT-ATTCA--CTTT-AG--CAAT...---T-T--G--AGTAG 6890
H2B.GH.86.D205          ---A-GCT-TA-A-G--CC--C-----AA--A-T--CCA-G--T-A---GAC--ACAT-AT-GG--CAGCT-AAGGT-TAGG---------T--CC-G--G---TTTC-A----G-------G---CC-AC-ATTCA--CTTCATG--CAAC...--C-GT--G--AGTAG 7430
H2B.CI.x.EHO            --AA-G-T-TA-A-A--CC-----T--A---A-T--CCA-G-AT-A---GAC--ACAT-AT-GG--TAGTT-AAGAT-TAGG---------T--CC-A--A-----TT-G----G---C---G---CT-AA-ATTCA--CTTCATG--CAAC...-----GT--G--AGTAG 7425
H2G.CI.92.ABT96         T--A-GCT-TA---GGCAC-----T--A----------CAGG-AT-A---GAT--ACAT-AT-GG--TG-T--AAGAT-TAGA---------T---C-A--G-----TT-G----G-------G---CT-AC-ATTCA---TTT-C---TAAT...--CT-T--A--AGTGG 6794
H2U.FR.96.12034         T-AA-G-T-CA---ACCAC--C---------T--T--CCA-G--T-T---GAC--ACAT-AT-GG--TG-C--AAGAT-TAGG---------A---C-A--C-A---TC-GT---G-------G---CT-AT-A-TCA---TTTATG--TAAT...--CT-T--A--AGTAG 6923
MAC.US.x.EMBL_3         ---A-GCT-CA---ATCAC--------T--T--T--CCA-G--T-T---GAC--ACAT-AT-GG--TA-T---AGAT-TAGG---------A--TC-A----A---TT-G--T-G-------G-C-CA-AT-ATTCA--CTTTATG--TAAA...---T-T--G--GGTGG 6512
SMM.US.x.H9             T-AA-GCT-TA---ACCA-----------AGT--T---CA-G-AT-----GAC---CAT-AT-GG--TG-T---AGAT-YAGA---------A---C-A----A---TT-G--T-GG------G---CC-AT-ATTCA--CTTT-CT--TAAC...-----T--G--AGTAG 6858
SMM.US.x.PGM53          T-AA-G-T-CA---ATCA-----------AGT--T---CA-G-AT-----GAC---CAT-AT-GG--TG-T---AGAT-TAGA--C------A---C-A----A---TT-G--T-GG------G---CA-AT-ATTCA--CTTT-CT--TAAC...---T-T--G--AGTGG 7315
SMM.US.x.SIVsmH635F_L3  T-AA-G-T-CA--CAGCA-----------AGT--T---CAGG-AT-----GAC---CAT-AT-GG--TG-T---AGAT-TAGA--C------A---C-A----A---TT-G--T-GG------G--TCA-AT-ATTCA--CTTT-CT--TAAA...---T-T--G--AGTGG 7388
SMM.SL.92.SL92B         T--A-G-T-CA---ATCA---C--T--AAGT--T--ACA-G--T-A---GAC---CAC-AT-GG--TG-C--CAGATATAGG--C-----A--AC-A--C-A---TT-G--T-G-------G---CA-AT-ATTCA--CTTT-C---TAAT...-----GT--G--AGTAG 6817
STM.US.x.STM            ---A-GCT-CA---G-CA---C--T---AGT--A---CA-G-AT----GAC--ACAT-AT-GG--TG-CT-AAGGT-TAGG---------A--AC------A---TT-G--C-G---C---G---CA-AT-AT-CA--CTTT-C---TAAT...--CT-T--G--AGTAG 7020
MNE.US.x.MNE027         T--A-G-T-CA---ATCAC--------T--T--T---CA-G-AT-T---GAC--ACAT-AT-GG--TG-T---AGAT-TAGG---------A--TC-A----A---TT-G--T-G-------G-C-CA-A--ATTCA--CTTTATG--TAAC...---T-T--G--GGTGG 6852
DRL.x.x.FAO             A-CA-G-T-TG----C-----C---T-AA----G-----C----AT---AAT----CC---ACA--TATG--GAAAT-CAGA-TG-----T--AC-A--A-A-ATAT-G----GG------G-A--ACTAAAT-TAA-T-GTAAG.............----A----A-A-CAG 6714
MND_2.x.x.5440          TGAG-G-T-CA----TTCA------G-AA----A--A-------A-C--AAC--A-CT---ACA--TAG---GA-AT-TAGG-TG-----A--CC-A--G----TCT-A-----------G-A--ATT-AA---GAC--A-TT-............---GGG-----GTCTG 6671
MND_2.CM.98.CM16        TTCT-G-T-TA-----CAA--C--T--A-----A--A--C----A---AAT--A-C----ACA--TGA---AAAAT-TAGACTG-----A--TC-A--A-A--TTC-CT-G-GG--CCTAG-A-GATT-AATGTCTCT-A-AA-..............C---A---CAG 7044
MND_2.GA.x.M14          AGAA-G-T-TA-----CAC-----T--A-----A--A-------AT---AAT----CC--AACA--CAA---GA-AT-TAGA-TA--------TC-A--G-ACATTC-G----G----CG-G-G--ATT-AAT--AACT-A-TT-.............-G-C-----TCAG 6970
OLC.CI.97.97CI12        .-CA-G-T-CA--T--CA-----------T--G--A--T----AT----A-T-T-C-GT--CA--T-AGCAAATGG-GA-G-TG-----A--AC-A---AGG-T-C--A----G------------G---AGAA---CC-..................--C-G---CA-TTCAG 6527
GRV.ET.x.GRI_677        .GAT-G-T-CA---T-CA------TGAT-----G--A-A-G-A--T---GAT--AAC--AT-GG--TA-TT-AAGAG--AGA--C-----A--A--A--G-A---TT-G-----------G-----GAT-AT-GA--CTTT-CT---AAG...--C-AG-----TTCAG 6940
VER.KE.x.TYO1_patent    .GAG-G-T-CA---TTCA------TGA------T--A-A-G-A--A---GAC--AAC--AT-GG--T-AGT-GAGGT--AGG---------T--A--A----A---TT-G-----------G--G-AGAT-AT-----GTATAA--A-AAT...--CT-------ATCAG 7048
VER.KE.x.9063           .GAA-GCT-TA---TTCAC-----TGAT-----T--A-A-G-A------GA---AAC--AT-GG--TGAGC-AAGGC--AGA--C--------A--A--A-----TC-GT-------CC--G--C-TGAT-AT-----TTATAAG-A-AAT...-----GT-----TTCAG 7069
VER.KE.x.AGM155         .GAG-GCT-TA---TCCAC-----TGAT-----T--A-AGG-A--T---GAT---AC--AT-GG--TGA-T-AAGAC--AGA---------T--A--A--A-AC--TT-G--T---------G--TG-GAT-ATGCA---TTTAAG---GAA...---T-T-----TTCAG 7049
VER.DE.x.AGM3           AGAA-G-T-TA---TTCA------TGAT-----T--A-A-G-A------GAT--AAC--AT-GG--TGAGT-AAGAT--AGG--C------------A--G-----TT-AT------C--CG--T-TGAT-ATGCA--GTTTAAGA--AAC...---T-T-----TTCAG 6567
TAN.UG.x.TAN1           .G-T-G-T-CA---TTCAC--C---GA---T--A--A-A-G-A--T---GAG--AACT-AT-GG--TA-CT-AAGAT--AGA--C-----A--A--A--C-----T--CT---------GG---CT-AT-AT-CA---TTT--TGTT..........---G------TTCAG 7029
SAB.SN.x.SAB1C          GG-A-G-T-CA---TCCA------T---------A--C-A-G----T---GAT--AACT-AT-GG--TA-CT--AGGC-TAGG---------T--A--A----A---TT-G----G----GCAG---CAGAT-ATTCA---CATAAGG--...........---GG-----A-CAG 7240
MND_1.GA.x.MNDGB1       T-CT-GCT-TA---ATCA------TGAG-----T-A-----G-A-A-----AG----G-CTT--GATAAG-...TGT--TTTAGG-----TG--TC-A--A-A--TC--GT---G--A----G-G--A----G-G---TTAAAT---AAT-A-TTG.............--GC---A-ATCAG 6465
DEB.CM.99.CM40          .-CAGGGT-CA---A-CAC-----TGA---T-----C--C---T-T---GA--CTAGCCG----A-A--TT--AAG--CAGG--C-----A--A-----A-A--GTC-A----G----G--G----G--G---A-AT--T-----T-----GTT-........--C-AC--C--G-CTG 6694
DEB.CM.99.CM5           ACATGGGT-TA---ACCAC--C----AT-----T--A------T-----GA--CCAGTCG----A----TT--AAG---AGA--C-----A--A-----A-A--GTC-GT---G----G--G-C-GA-AC--T-----T--C--CTTG...........-----C-----G-CTG 6645
TAL.CM.01.8023          ATA----T-TGCA---G-A--C--TTTG-----A--AG-T-----T---GA---AC-CAT-A-C-AG-GT----A--T--G--C-----T--A--A--C-----CC-GA----G------G-C--CC-A-GG-CA--T-G---CAAG.........--CAT--CA--TCAG 6289
TAL.CM.00.266           .TT----T-TGCC---G-A--C--TGTG-----T--AG----A--T---GA---AACTCA--ACC--G--T----T-----G--C-----T--A--A--C-----CC-GA----G------G---GCC-A-GG-CA--T-G----AAG.........--CAC--CA--TCAG 6745
SUN.GA.98.L14           .-CA-G-T-TA-----CA---C--TGAT---A-T--------A-A---C-AT----GTATTA-GC-AAATGC-TA-T-CAGGCT------A--T--G--G-A-AT-T-AT---G----G--G-A--ATTAAATGC-ACT-AGAA-.............---A----A-A-CTG 7199
WRC.CI.97.97CI14        .GAA-GCT-TA-A-ATCAC-----TGAG---------A-AG--A-A-----A---A-G--GGCAG--TGA-CAATAT---AGAATG-----T--AC-A--A-CA-TTT-GT-G---------G-G--ATT-AA-TCCTCT-AGCT-...........---GAG------TCAG 7034
WRC.CI.98.98CI04        .GAA-GCT-TG-A-ATCA------TGAA--T--G--A-A---A-AT----AG--A-GT-GGCAA--TGA-CAATATG--AGAATG-----T--AC-A--A-C--TTT-G--T---------G-GCCCTTAAA-TCCTCT-A-TT-...........---GAG-----ATCAG 7082
WRC.GM.05.Pbt_05GM_X02  A-CT-G-T-CA-A-ATCA------TGAG--T--A--CCAT----AT----A-----C--GGCA---TGAGCAATA----AGAATG-----A--AC-A----A--TAT-AT---G-------G-A--ATTAAAT-G-A-T-AGCT-...........---GAG-----TTCTG 5978
LST.CD.88.524           .-CT-GCT-CA----CCA---C--TGATAGTA-T---CTG--A-A---CAAT----G-GTTA-GTCAAATGC-TATT-TAGGCT------A--A-----C-ACAT-C--T-------C---G-A---CTAAA-TT-TC-G-TAAT.............-----T---A-A-CAG 6071
LST.CD.88.447           .-CT-G-T-TA----CCAC-----TGATAGTA-T---CAG--A-AT--CAAT----G-GTAA-ATCTAATGC-TATT-TAGGCT------A--A--A--C-A-AT-C--T----G--C--CG-A--ATT-AATTTCTC-G-TAAT.............-----C---A-A-CAG 6068
LST.KE.x.lho7           A-C--G-T-CA----TCA------TGAT---A-A--A-----A-AT---AAT---GGGTAA-GC-AAATGC-TATT-TAGGCT-----T--A--A--C-A-AT-C-AT-G-G---C---G-AC-ATTAAA-TT-A-T-A-AA-...........---GA----A-A-CAG 7137
SYK.KE.x.SYK173         -TAT---T-CC--TT-CA------T--AAGT--T--AT-CGCA--A---GAG--ACAGA-----C-A--CT----G--T--A--C-----T--AC-G----ACT-AT-GT-G-----C---G---CA-AT--TG-A---GAC-ATGTC..................-A-CTG 6936
SYK.KE.x.KE51           TTAT--CT-T--ACAGCA------T--TAGT--A--GCGCA-T--T---GA---ACAGA-T---CGC---T-------T--A--C-----T--AC-G--A-ACT-AC-GT-G---------G---CAGAA--TG-A---GAT-CTGAG.........--CA------A-CAG 6632
COL.CM.x.CGU1           ---GGCA-CAA---A-GA----GCA-AT-ATACAG----C--A-TA---GAT-T-ACTATAG-G--T---G--AGG-C-GGG-TC-----AG-TC-A--A-ACAT-C--T---G----G--G----A-A--GGG----G------G--........--C-AC-----G-CAG 6361
```

```
H1B.FR.83.HXB2         CAGTACAATGTACACATGGAATTAGGCCAGTAGTATCAACTCAACTGCTGTTAAATGGCAGTCTAGCAGAAGAAGAGGTAGTAATTAGA.......................TCTGTCAATTTCACGGACAATGCTAAAACCATAATAGTACAGCTGAACACATCTGTA 7100
Env                    T  V  Q  C  T  H  G  I  R  P  V  V  S  T  Q  L  L  L  N  G  S  L  A  E  E  E  V  V  I  R  .  .  .  .  .  .  .  .  S  V  N  F  T  D  N  A  K  T  I  I  V  Q  L  N  T  S  V
H1A1.UG.85.U455        ---------------------C-A------G---------------------------------------AG---AA--AGG------.......................----AA--------AA-------A-------------------TGT--ATC----- 6545
H1B.US.90.WEAU160      ----------------C--------A--------G-----C---T--------------------------------CA-----------.......................----AA----------------------A-----------------TGT---CA-- 7110
H1C.ET.86.ETH2220      ----------C------------AA-----G-------------AT---------T---A--------GT---AC-A--------.......................-T--AA---C-G--AA-------C----TA--------------T--TGA------- 6503
H1D.CD.84.84ZR085      -------------------------G--G------------------------------------------A-C-----A---.......................----AA---C----AA------------T-----------T--T--TCA------- 6642
H1F1.BE.93.VI850       -----------------------AA-----G-----------T--T-----------C-----------GTA-------C---.......................---CAA---A--T-AA-T-----A-----------------C--T--TGA------- 6422
H1G.SE.93.SE6165       -----------------------A------G-------------A--AC-G---------T---------G-A-AA--AA-G-----.......................----AA---------A------A-C---GT---------------T--T-A-A----- 6524
H1H.CF.90.056          ------------------------A-----G----------------A--------A--C------------C--A-CA--------.......................A--AAA---A--T-A------A-C----A---------------T--G---C-A--- 6431
H1J.SE.93.SE7887       ----------------C-------A-----G-------------A---C-----------A---------G---CA--A--------.......................----AA--CA--T-A--------C----A------------A--T--T-A-A----- 6413
H1K.CM.96.MP535        -----------------------AA-----G---------------------------C---------------A--A-------G.......................----AA---A-T--A--T---A-A----A---------------T--TGA-A----- 6308
H101_AE.TH.90.CM240    ----------C------------A------G-----------T-----------------------------A--A----C---.......................----AAG--C----AA-------C-------------G--C--T--T-A------- 6680
H102_AG.NG.x.IBNG      ----------C----------C-A-----------------------------------------G---A-----G--C---.......................----AA---A----AA-------C------------------T--GCT-ATC----- 6624
H1N.CM.95.YBF30        -T--------C-----------A-A------GA----C-----GT-AA-CC-------A--CT--AATAC---T-GAA-T--T-------.......................-AT--T-G-CAC-GT-ATC-GT-G--G--ATG---TGAGA-A--G 6681
H1O.BE.87.ANT70        T---TACT-----------C---------AC----AGT-----G--AA-A--------G-CA--CT-TA---G-A-AA--AG---G-TG.......................G-AAAAG--A-TTT---AGG--GA----AT--C-----GACC--A---T-TA-CC-- 7136
H1O.CM.91.MVP5180      TG--TACT-----------C--C-A----AC----AGT--------AA-AC-G-----G-CA--CT-TAG----A--A--AG-----TG.......................GGAAAA---A-T--A--ATCA--A--G-AT--C------ACC--A-----TC--A-- 7135
CPZ.CD.90.ANT          T------T--C--T--------A--C---A-G--GG----ATGGT-A--A---------A-C-TACCA-AC-A-CACTTC----G-A-TG.......................AA--GTCGCAAA-AT--ATC--TGCTTGTA-G-T-T-G-A-AGAATT-GA-AACT-- 6553
CPZ.CM.01.SIVcpzCAM13  ----C--T---------------C-AA---------A----C---T--A-CA-T-----G---------TAC----A-CA--AC-G-A---.......................GTAAA----GC-T-TA-A---A-CC-TGATTGG--------A--A-G----G-A--G 6711
CPZ.CM.05.SIVcpzEK505  -T-----------C--------A-A------GA----C--A--GT--A-CC-----------CT-----AC-TC-A-TA-C-----A---.......................AAT----CA-AA---..........--AC-CT-G--G--AT-A--TGA-AG-A-C 6637
CPZ.CM.05.SIVcpzMB66   -T--G--T-----------C--C--A--T-----C--------GT-T--T--------A-CCT---A----A-G-T-AC---CC-AGAT.......................AGG---G--T-AA-TG--ATGG-C--T--------GA-AT-A---GA-A----- 6742
CPZ.CM.98.CAM3         ----C-------T-----C--C--A-----G---A-T--A---T--A-CC-T-----A---G----C---A-CA--ACTA-TGC----.......................AGG--C-GGT-C-R-T-CC-GA-N--RT-TC-G-ANN-G--C 6531
CPZ.GA.88.GAB2         ----G--T--C--------------A----ACT---A-T-----GT--A-AC-C--------CT-------------A-AG--GG-GTTC...............AGAACTAAGAA-A-G-CAGCTC-A-GTCTCAG-G-C--AG-T---------T-A--A-GGG-AA-T 6514
CPZ.CM.98.CAM5         ----------C--T-----C--A--A----------A-T--A---G--A-CC-T-----A--C-----T-----G-G-CCC-CTG--GCTAGA.......................-G-GAAGATCTA-GACAC----T---T---A----AT-T--AGA-GGGA-T 6807
CPZ.CM.05.SIVcpzMT145  -------T-----------C----AA-----G-C-A-C--C------A-T--G-----G---AC------CA-TC--ACT--GGCC---.......................A-AGACCCTTCA-AG--TTTAGC----A-------C--AGATC-A--- 6621
CPZ.TZ.00.TAN1         T---T--C-----------G---TAT--TA-GA--G-C--AGC-T-A-AC---------GTCC--G-A---------AAC-AA-GC-TAC.......................T-TGT-AAT--CTCAGT-AA--C-C--T--T-----A-AT-T--TGT---AA-- 6741
CPZ.CM.05.SIVcpzLB7    -------C-----T----------AA-----TA-T--T------T--A-CC-G-----T---------------A-AA-CA-G-----G.......................T-CAAGG-T--T.........-AGG-TC----------C--T---AG-A-- 6632
CPZ.US.85.CPZUS        ----G--G-----T-----C--A-AA-----G--GA-T--A--G--CA-AA-----------T-----ACTA--A-T--TACTG-A---.......................-G-AAA...A--TT---AG-C-TT--TC-------AT-CTCAGAGGGA--C 7135
CPZ.GA.88.GAB1         ----G--T-----------------AA-----G--TA-C-----GT-A--TA-T-----A---T----T----G-A-CA--ACTG-A---.......................GTG-AG---AAGT-AA-A---A--G-TGT-TGG-----------AGTAGA-G-A--T 7161
MON.CM.99.L1           -----AC------T----CT--CCAA--TA-----G-T---TGGT-C-A---G--CA-T-CGGGTAAT-CTCCCA-TAC-AC-G-A-TG.......................ATGAAT--GCAG-AAA-TG-GT---T-GTAG-T-G-T-GGCTAA-C--CT-CA---C 7045
MON.NG.x.NG1           -----AC---C--G-----T--A-T----A-------T---TGGT-T-A-C-T----CT-CCTCT--CAC-A-TR--ACCAC-G---TGATG.......................AATGGGAAA-GAA-TG-AT-C-T-GTA----GGC-TGM-AA--C-CTC--A--G 5661
GSN.CM.99.CN166        -----ACC--C--------C---CAA--CT-----G----ATGG--C-AT--------G-CCTATCAGCC--G-A-TAATACC-GA-T-ATG.......................ATGAATGG-A-G-A-A-TG-AT---T-GTT---GGCT-TGGAGAAG-TTATCA-C-T 7037
GSN.CM.99.CN71         -----AC---C--G-----C---CAA--TT-----G----ATGG--T-ATC-------G-C-TATCAGCC--GGA-CAATACC-GAGT-ATG.......................ATGAATGGCAAA-A-A-TG-AT---T-GT----GGCT-TGG-GAAG-TTATCAGC-T 7028
MUS_1.CM.01.1085       -TAGG--T--C--G-----C--ATTA--CT-G--TG-T---TGG--A-A-C-T-----A-C-TATTT---G-G-A-TAATAC-GCAGT-ATG.......................ATGAATGGGGAT-AAA-TG-AT-A-T-GGA----A-T-TGG-GAA--TTTTAGA--- 7001
MUS_1.CM.01.CM1239     -TACG--T--C--G-----C--ATTA--CT-G--TG-T--CTGGT-A-A-C-T------CATATTT----G-A-TAATAC-GCAGT-ATG.......................ATGAATGGGGA--AAA-TG-AT---T-GGA----AGT-TGG-GAA---TTTAGA--T 6973
MUS_2.CM.01.CM1246     -TAGG--C--C--T----AC--ATTA--GC--A-TG-T---TGG--A-AA--G-----T-CCTACCT-CC--GCA-T-ACAC-GC-GT-ATG.......................ATGAATGG-GAT-AAA-TG-AT---TTG------A-T-TGGAGAA---TTTAGG--- 7108
MUS_2.CM.01.CM2500     -TACG--T--C--G----C--ATT----C----TG-T---TGG--A-A-C-------T-CCTACTT-CCT-GCA-T-ATAC-GCAGT-ATG.......................ATGAATGGCAAG-AAA-TG-GT---T-G-T---CA-T-TGGAGAA--TTTTAGG--C 7037
RCM.NG.x.NG411         --ACGTC---C-----CATG--A-ATA--AC------C--ATC-T-TGGA--------ATC-A-TAGT--GA-TAGAAC-TGG--CTAT.......................CAA-GAAAACAAAGT---AGG-C-GTT---GGGT--A-TAGAC-TTTTAA-T-- 6858
RCM.GA.x.GAB1          -CACTTC---C------ATG----ATA-TACT--GG-CT-AGGGT-TGGA--------ATCAA-TAAT-T-A-T---ACTTGG--ATAT.......................CAG-GAAGGCAAAGT---AGG-C-GTT---GGTC-CA-TAGTTTTTATAA-T-G 6814
DEN.CD.x.CD1           --AG-GTG------G---AG--ACACA-CA-G-CCAGC--CTGGT--A-CA-G--CA-T-CAG---A-CTTCT---TAATCAT-CACA-ATT...............ATAAGAAACACA-GCAAAGATA-TC-GA-A-T-G----TT--T-CAGAAA--GTT-TGTAA-C 7160
                        V3 loop end

MAC.US.x.239           TCTCTTC---C---AGGATG--GGA-A--CAGACT--T---TGGT-TGGC--T-----A-C-AG-------A-TAGAACTTAT---TAC..........................TGGC--GGT-G---T---AGG-CT-TA--T-GTT--A-TAA-T-TTATAA-C-- 7524
Env                    V  S  S  C  T  R  M  M  E  T  Q  T  S  T  W  F  G  F  N  G  T  R  A  E  N  R  T  Y  I  Y  .  .  .  .  .  .  .  .  .  W  H  G  R  D  N  R  T  I  I  S  L  N  K  Y  Y  N  L
H2A.GW.x.ALI           -T-CTAC---C---AGGATG--GGA-A-GCA-ACT--T---TGGT-TGGC--T-------C-AGG-----A-TAGAAC-TAT--CTAT..........................TGGC--GGT-GA--T---AGA-CT-T---C-GCT--A-CAAAC-TTATAA-C-T 7587
H2A.DE.x.BEN           -T-CTTC---C---AGGATG--GGAAA-GCA-ACT--T---TGGT-TGGC--T-------C-AG-------A-TAGAAC-TAT--CTAT..........................TGGC--GG--GA--T---AGG-CT-T---T-GCT--A-CAA-T-TTATAA-C-C 7606
H2A.SN.x.ST            -T-CTAC-------AGGATG--GGAAA-GCA-ACC--C---TGGT-TGGC--T-------C-AGG------A-TAGAAC-TAT--CTAT..........................TGGC--GGT-G---T---AGA-CT-T---T-GCT--A-CAA-TTTTATAA-C-C 7033
H2B.GH.86.D205         -GTCCTCC--C---AGAATG--GGAAA--CAGTCC--T--ATGGT-TGGC--C-----T-CAAGG-----GA-CAG-AC-TAT--ATAT..........................TGGC--GAA-AA------AGG-CC-T-----GCT--A-TACAT--TATAA-T-G 7573
H2B.CI.x.EHO           T-TCT-TG...TACAGAATG--GGAAA--CAGACC--T--ATGGT-TGGC--C-----T-CAAGG------A-CAG-AC-TAT--CTAT..........................TGGC--GGT-AA------AGG-CT-T-----GCT--A-TTCAT-TTATAA-C-G 7565
H2G.CI.92.ABT96        TCTCTAC-------AGAATG--GGAAA--CA-ACC--T--ATGGT-CGG---T-----A-CAAGG------A-TAGAAC-TAT--CTAT..........................TGGC--GGG-GA------AGG-CT-TA----GTT--A-TAA-T-TTATAA-T-- 6937
H2U.FR.96.12034        TCTCTAC---C---AGAATG-GGG-AA--CA-ACC--T---TGGT-TGGA--T-------C-AG-T-----A-TAGAACTTAT--ATAC..........................TGGC--GGT-GA------AGA-C--TA--T-GCT--A-TAA-T-TTATAA-T-- 7066
MAC.US.x.EMBL_3        TCTCTTC---C---AGGATG--GGA-A--CAGACT--T---TGGT-TGGC--T-----A-C-AG-------A-TAGAACTTAT---TAC..........................TGGC--GGT-G---T---AGG-CT-TA--T-GTT--A-TAA-T-TTATAA-C-- 6655
SMM.US.x.H9            TTTCTTC---C---AGAATG--GGAAA--CA-ACC--T---TGGT-TGGT--C-----T-C-AG-------A-TAGAAC-TAC---TAT..........................TGGC--RG--GAAG----AGA-CC-TA--T-GCT--A-TAA-T-TTRTAA-C-- 7001
SMM.US.x.PGM53         TTTCTTC---C---AGAATG--GGA-A-GCA-ACC--T---TGGT-TGGT--C-----T-C-AGG------A-TAGAAC-TAC---TAT..........................TGGC--AGT-GAAGT---AGA-CC-TA--T-GCT--A-TAA-T-TTATAA-C-- 7458
SMM.US.x.SIVsmH635F_L3 TTTCTTC---C---AGAATG--GGA-A-GCA-ACC--T---TGGT-TGGC--C-----T-C-AGG------A-TAGAAC-TAC---TAT..........................TGGC--GG--ATAGT---AGA-CC-TA--T-GCT--A-TAA-T-TTATAA-C-- 7531
SMM.SL.92.SL92B        T-TC-TC-------AGGATG--GGAAA--CAGACT--T--ATGGT-TGGC--T-----T-C-AG------GA-TAGAAC-TAT--ATAT..........................TGGC--GG-GGAAGT---AGA-C--TA--T-GCT--A-TAA-T-TTATAA-T-G 6960
STM.US.x.STM           TCTCCTCG-----TAGAATG--GGA-A--CA-ACT------TGGT-TGGC--T-----A-C-AG-------A-TAGAAC-TAT--CTAT..........................TGGC--GG--G---T---AGG-CT-TT--T-GCT-GA-CAA-T-TTACAA-T-- 7163
MNE.US.x.MNE027        TCTCTTC---C---AGAATG--GGA-A--CAGACT--T---TGGT-TGGC--T-----A-C-AG-------A-TAGAACTTAT---TAC..........................TGGC--AG--AA--T---AGG-CT-TA--T-GTT-GA-TAA-T-T-ATAA-C-- 6995
DRL.x.x.FAO            ----G--G-----TAGACAT--GCCTG-CAC-A-C--T-GC--TT-TGGC--C-------CAAAGCAT-GTC-G--T-AGT----AGA-ACA...............AGGGGAAAACCAG-C--TCT-A-TC--AAAT-TGTGT-T-GG--TA-TAA---ATGGAACC-C 6869
MND_2.x.x.5440         ----G--G--C---GCACCGC-GCCAG-CAC-A-C--T--AATGT-TGGC--C-----A-C-AAGCAT--TT-T--T-AGC-T---CAGACA......AACCCACGAAAAGGAAAA-ATG-G--TCAT---C--AAAT-TGTATAT-G---GG-TAAA--ATGGGGAT-- 6835
MND_2.CM.98.CM16       -------G--------GCC-T-GCCAG-CACCACC------ATGTGTGGA-CT-G---G-CAAA-CAT--TT-TA-T-A-T----ACAGACA......AATACAAAGAAAGGCAAA-AGG-G--TCAT---C--AAGT-TGTGTAT-G----G-TGA---GTACAACT-- 7208
MND_2.GA.x.M14         -------G-----TG-CCC---GCCTG-TACTA--------ATGT-TGGC--T-----A-CAAA-CATA-TT-T--T-AGT-----TTGGTG......AACCCTCAA.........AAGG-G--TCAT---C--AAAT-TGTTTAT-GG--GA-CAAA--ATGGGGGT-G 7125
OLC.CI.97.97CI12       ----TACT--C--C-CGCC-C-GTCAG--AC---GAGT-G-TTCT-TGGC--T-----T-C-AG-CATA-G-C-AGCAG-C-G---CCTATA...ACAGGGGATTTGAATCATAAACCA-GAGAGGGAACAGGA-TAT-TGTT--GGC-A--GCTGAT--GTATAAGT-- 6694
GRV.ET.x.GRI_677       T---G--T-----TAGATT---C-ATA-TACTA--A-T--AGGGA-AGGA--------T---AG-T-----A-TAGAAC--AG--AT-G......................CAGAAA-GAGGAAATGATA-TG--A-AGTT-TA----AGT-GA-TAA-TTTTACAACT-G 7089
VER.KE.x.TYO1_patent   T---G--T------GGCTT---G-ATA--AC---GA----AGGGT--T--C-G-----A--CTATCAT--GA-TCGAACCCAG--AT-G......................CAGAAACAT-GGG-A-ATA-T--CA-AGT-TTG--CT-GT-CA-CAA-C--TATAA-C-- 7197
VER.KE.x.9063          T------C-----TGGACTT--G-ATA--AC---T--T--AGG---A-----G-----A--CTATT-T--GA-TCGAAC-CA---AT-G......................CAGAAGCAT-GAG-G-ATA-TG--T--GT--TAG-TC-TT-TA-CAA-T-TTACAA-T-G 7218
VER.KE.x.AGM155        T---G--T--C---AC-TT---G-ATA--AC----A-C---GGT---T-A--G-----A--CTATT-----A-TCGAACCCAG--CT-G......................CAAAAACATGGAG-G-GCA-TG-CT-AGTGTTA--CT-GC-CA-TAA-C-TTATAACC-G 7198
VER.DE.x.AGM3          TG--G--T-----TA-CTTG--A-ATA--AC---GA-T---GG----T----G-----G--CTACT----GA-TCGAACCCAG--AT-G......................CAGAAACAT-GAG-A-GCA-TG-CT-AGTGTTAG-GT--T-TA-TAAAC-TTACAA-C-- 6716
TAN.UG.x.TAN1          T---TAGT-----TGGATTG--G-ATA-TAC---GAGCT-AGC-T-TGGCA------------AG------A-CAGAAC--A---AT-G......................CAAAAGCATGGAG-G-GCA-T---T--GTG-TA----A-C--A-TAAAC-TTATAAGT-G 7178
SAB.SN.x.SAB1C         T-TC-GCT--C--TAGGTT---C-ATA-TACT---AGC--AGG-A-AGGAA-T-----T---TAT-T--C-A-TAGAAC--A---CT-G......................CAGAAAAATGGCAATT-CA-TG--T--GTT-TT--T-GGC-CA-TAGAT-TTTTAA-T-- 7389
MND_1.GA.x.MNDGB1      ----G--G-----T--GCACT-AGTAG-CAC----AGT-GCTTTT-TGGC--T-----A-C-A-GCATA-G...--A-G--A-T-G-T-CCCATAGATGATAAATATAGGGGCC-A-AGG-A--TCATC-A-GGAAGTTTGT-TAT-AG--G-CAGGA--ATATGGCT-- 6632
DEB.CM.99.CM40         ----GGC-------A--CTC--ACACA-CA-G-C---C--ATGGG-A-AA--T-----GTC-GAT-AG--GAGG-CA-A--AGC--CATATA......ATAAGAAAAGAGGTAAAA-GTG-GG-GCA-A-TGGAAGC-TT--T----G---G-CCGCT--ATACAACT-G 6858
DEB.CM.99.CM5          ----GGC-------A--CTC--ACATA-CA-G-C------ATGGA---AA--C-----TTCAGAT-A----AGG-CA-A--AC--ACATATC...............ATAAGGAAG...GGCC-GGAAA-T--GA-C-TT--TG-C-G-A-C-CA-CA--GTATAA-C-- 6797
TAL.CM.01.8023         -------C--C---G--CCC--C-CCA-CA---C-G----ATGGT--G-TC-------T-CAAAG..........--A-TCAGAC-CAA-T-ATA......................CA-CAGGCACC--AAA-TG-AT-AGCC-TA--CC-CT--GG-AA-G-ACACAACA-C 6432
TAL.CM.00.266          -------C------G-CCCT--C-CCA--A---C-G-T--GTGGT-AA-TC-C-----G-CAAAG..........--A-TCACAC-CAGGTCATA......................CA-CAGGC-CCA-A-A-TG-AAG-GCT-TA--CC-CC--GGAAA-G-ACACAACC-C 6888
SUN.GA.98.L14          -TACC-CT-----TA--TAT--G-CAAGTAC---TAGT-GCTTTT-TGGC--T-----A-C-AG-CATA-G-----T-AGT----ACC-ATC............AATAACAAGGTA-GAC-AGGGG-T--GGGA-AAT-TGTATGG-A----GCAGCT--GTGGGGGT-- 7357
WRC.CI.97.97CI14       -----ATG------G-CATG--G-CAG--ACCA-TAGT-GCTTCT-TGGA--T-----A-C-AG-CATA-G--G-GA-A-T-----GA-GTA...ATAGGAAATATAAATCATAAACCA--GGAAGGATCAGGA--ATTTGTGTAT-AG--G-CAGAC--GTATAA---T 7201
WRC.CI.98.98CI04       -----TT---C---G-CATG--G-CAG-TACTA-TAGT-G-TTCT-TGGA--T-----T-CCAG-TATA-GA---G--A-T--G-AGAGATA...ATAGGAAATATAAATCATAGGCCTC--GAAGGTACAGGA--ATTTGTATAT-A---T-CTAGT-GGTATAAG--- 7249
WRC.GM.05.Pbt_05GM_X02 ----TACT-----TG--TT---G-CAG--ACTA-TAGT-GCTTTT-TGG---T-----A-C-AAGTAT--TA---G--A-T-G--AGA-GTA...ATAGGAAATATAAACCATAAACCA--AGATGGAACAGGA--ATTTGTTTAT-A-----CAGCT--ATATAA---G 6145
LST.CD.88.524          -TAC--C---C---G--TAC--G-TAAGTTC---TAGT-GCTTCT-TGGC--T-----A-CCAATCATAC------C-AGT-----CC-CTA.........ACTCCAAAAAAGATG-GAG--C-T-AT-GAGC-AAATTTGTTTAC-AG---GCAGGA--ATGGGGAC-- 6232
LST.CD.88.447          --ACC-CT--C---G-CTAT--G-TTAGTTC---GAGT-G-TTCT-TGGC--T-----A-CCAATCATAC---C--T-A-T-----CC-CTA.........ACTCCAAAAAAGATG-GA---C--CTA--AGCAAAGTTTGTATAT-A----GCAGGA--ATGGGGAC-C 6229
LST.KE.x.lho7          -TAC--C------TGGATAT--GCTTAGTAGT---AGT-G-TTCT-TGGC--T-----A-C-AATCATACCAGG--T-AGC-T---CC-CTA.........ACTCCAAATAAGATG-AAG--C-A-AT-GTGCAAAGTTTGTGTAT-A----GCAGGA--ATGGGG-T-- 7298
SYK.KE.x.SYK173        -----TC---C-----G-A-T-C-ATA--T-G-CTAGC--ATGGT-C-AAC-G-----G-CCTACAA--C-A----TAA---T-GATTTATC......................AAACAA--GGAT-AAA-TG-GT-AGT--TT--TT----T-CAGAAGCACTTAGAT-G 7085
SYK.KE.x.KE51          ----TTC---C------CCTT-C-ACA-TT-G-CTAGT--ATGGT-T-A-----------CCTATAA--C-A----CAA---T-GGTTTATT......................AGACAA--GGAT-AAA-TG-AT-AGTG-TT---C----G-CAGAA--ATT-GGAT-- 6781
COL.CM.x.CGU1          ----GTCC-----C----AGT---ATAT-AC---CATGT----TG-AT-AG-C----C-TCAAA--AGTTGTCT--TTGG-CG-AAGAT...................AGAGAGGGA---TGGAAG-AT---TCA-GG-C--TTGAGTATTATTG-T-TCCA-A-GA-A-- 6513
```

```
                              V3 loop start                                                                                        V3 loop end
H1B.FR.83.HXB2            GAAATTAATTGTACAAGACCCAACAACAAT...ACAAGAAAAAGAATCCGTATCCAGAGAGGACCAGGGAGAGCATTTGTT...........ACAATAGGAAAA...ATAGGAAATATGAGACAAGCACATTGTAACATTAGTAGAGCAAAATGGAATAACACTTTAAA 7252
Env                        E  I  N  C  T  R  P  N  N  N  .  T  R  K  R  I  R  I  Q  R  G  P  G  R  A  F  V  .  .  .  .  T  I  G  K  .  I  G  N  M  R  Q  A  H  C  N  I  S  R  A  K  W  N  N  T  L  K
H1A1.UG.85.U455           A----C------T-C-----TT----T-CAAGA-A--AT-T-C-T-GGTA--GT...-T----T----ACA------CTA-...........GT--C---T---ATA----GG---A-----------------TG-C------AGGG-C-------GA---A--C- 6700
H1B.US.90.WEAU160         ------------------------------..-----------A---AAC-T-A......------------T-C--TA-...........----C----G--ATA------G----A----G-------------C--------A---GT----------------- 7259
H1C.ET.86.ETH2220         -------C------G-------G---T---..------G----T--AA-G--A.....--------ACA-A----CTA-...........G---C----G-CATA------G----A-----------------------GA--A-----------A---C--C- 6652
H1D.CD.84.84ZR085         ----------------G---T----A--G...GA----C----G-CA-C---A..........----A----CA---GC-CTA-...........----C-A-GT-TACA-C-A-G-TA--AG--------T-------T------G---T--------------C---CG 6791
H1F1.BE.93.VI850          C-G----------------------------..-----G---G-T--A-A-T-A......--------ACA-A----CTA-...........G---C----GC-ATA------G-C--A---A-G-------------------G--A--C--------------C-GG- 6571
H1G.SE.93.SE6165          -------------C-------------ACA-TG-A--GG-T--GAATGGGA...-TT---------ACA-A-G--CTA-...........G---C---TGCCATA------G-C--A------------------TG---C--A-AG---------AG-GG-C---C- 6679
H1H.CF.90.056             A-C----CA-----C--G--T-----T---..--G----C---T--A-A-T-A.......--G-----AC-------CTA-...........G---C---TG-CATC------G----A-------------------T---------A--G-C--------G------C- 6580
H1J.SE.93.SE7887          ------GTG---TAC-----T--T-------..------G---G-T--A-AC--G...........--------ACA--TGC-CTAC...........G---C----G--ATA-----------A--G--A-------------------GA-AG-G-T----G----------CG 6562
H1K.CM.96.MP535           C----------------G--A----------..-----------T--A-A---G...........--------A-A------CTA-...........----C---TG-TATA------G----A------G---------C----------G---A----------C-TG-------G 6457
H101_AE.TH.90.CM240       -----C--------C------TC--------..-------C---T--AAC---A...........--------AC---T---CTA-...........-G--C----G-TATA------------A---A-----T-----G-G----A-G--A-------------AGT------ 6829
H102_AG.NG.x.IBNG         A-------------C-----T-----------..----------G-TG-A-A---A...........----------CA------CTA-...........G---C---TG-CATA-----GG---A-----------------TG-C----A-A--G----------A------C- 6773
H1N.CM.95.YBF30           CC---A-----------G--AGGA--T---..---G--GG-CAGG-G-AG--A..........-----T-CT-TGA-----TA-...........-AC----A----ATAG-----G-C--T---------T-C-----TG-CTC--A--A-CT----G-ACCA-TG-GG-- 6830
H1O.BE.87.ANT70           A-C--G-CC---GA------AC-A-TA.....GAC-T-C--GAG--GA-A--A.......--T---AT-GCCTGG-ACAGCATGGGA......-T-GGG----C-GCAGG-AAC-GCTCA--GGC---TT----C--GTA--A-GCCA-TG-T---GGA--A-TA----- 7288
H1O.CM.91.MVP5180         A-C--G-CC--C-T----GAAGGA-TTGCA...GAGGT-C--GAT--ATA--CA.......--T---AT----TGGCGCAG-ATGACACTTAAA-G--GTAAC--TACATC-CC--GATCA--GGT---TT------CATA--A--AGA-TGT----G-A--TG-CC--C- 7296
CPZ.CD.90.ANT             AC-----CA----T------AGGA--T-GGACAGT----...-ATC-A-AA--A...........--------A-TGA-T--CTA-...........-ACG---A--T-GCA-C----G-C-CT--GA----GTTC----CAG-C-A--AGA-GCT----G-AC-AG-ACGT-- 6702
CPZ.CM.01.SIVcpzCAM13     A-TC-G-CG----AG--GGTAGGG--T--C.........------GGC-A-G-A-AG--A...........--------A-TGA-C--CTA-...........-AT--G-ACC-TATAT-T--TG---CA--GA----CTT---CG-AC-C-A-G-GA---C-------G-A--AC-GC- 6860
CPZ.CM.05.SIVcpzEK505     CC---A-----C--------AGGA--T--G.........------GG-CA-G-A-AA--A...........-----T--T-TGA----CTA-...........-AC----A---TATA------G-C-CT--G------T-----G-GG-A-AC--GA--...---G-AC-A-TA-GG-- 6783
CPZ.CM.05.SIVcpzMB66      CGTC-C-----C--T---A-AGGA--T--CACA-T--A-GGT...--A-C---A...........-------A-TCAGAT----TA-...........GG-----A--CTGTA------G-C-CA--G--G--CTT---CC-GT-G-AC-A-A--GT-----C---T--A--T-- 6891
CPZ.CM.98.CAM3            AGC--A---------------AGGA--T-NC...-GC---GG-CA---A-AG--A...........---------C-TGA-C---TA-...........-AC----A---TATTG-----G---CC-----G----T-----C-G--C-A---GA--GTC---G-CG-A-GAC---- 6680
CPZ.GA.88.GAB2            CCC---------C-GT-----AGGA--------...---G---G-GC---AAACC-T......A-C--------C-A----CT-C...........-AC-CG-A-GCCCTA---------CCCT--GA----GTCC--CC--C---A-G-TA--CT-----------T-C---- 6663
CPZ.CM.98.CAM5            A-C--A-T-----TG-----AGGA--T--C...-GC---GG-CA-C-A-AG--A...........----------GT-T-C--CTA-...........-AT----A---CATAG---GG--CCC------------------TG-C--C-AG--CC-G----TG--A-AG----- 6956
CPZ.CM.05.SIVcpzMT145     A-G----CA----GG-----AGGA--T--C.........------GGTCA---A-AA--A...........---------CA-TGA-T---TA-...........-AC----A---TGTAG---GG-C-CT---A-----T-----G-G--A-A-G-GA--C-G---GCA--AG-CC-C-- 6770
CPZ.TZ.00.TAN1            A-TT-A-CG---GA----A-AGGA-------..------GGTCA-G-A-AG--A.......--T-----T-TGA-C---TA-...........-AT----A---TGTAG---GG-C-CC--GA----TT-----TCAG-C-A-GC-A---C----T-C-GG-AC---G- 6890
CPZ.CM.05.SIVcpzLB7       C-C--C---------T--GGTAGGA--T--GACA-T--A-GG-...----CCC-A...........--------ACAGCTG--CTA-...........GG--C--A--CTGTAG-----G-C-CA---G-G--T-------G-A----ACCAGA--C-T---T----A-TAC---- 6781
CPZ.US.85.CPZUS           A-T--G-C-----TT-----AGGA-------..---GTGGG--ATG-A-AAC-A...........--------A-TGA-T---TA-...........-AC---CC---GATAG-----G--G-A---G-----------------CTCA-A-CTG-C----G-G--ACAGAG--- 7284
CPZ.GA.88.GAB1            AGTC-A------CAT-----AGGA-------..------GGG-GAGG-G-AG--A.......--G-----A-TGA-C---TA-...........-AT----A---TGTAG-----G---CC---TCT--CT-C-----G--A-A-G-GA-C-C--------CGA--GG-GG- 7310
MON.CM.99.L1              A-C----C---C-TT--G--AGGA-----A...--T-TT-GG-ATT-A-AG--A.........--G----C-TGA----CTA-...........T-GCA-CTT-T-GTGGG---T----CTC-TA----CT-C--C--AG---A--AGA-CC-----G-G-CTG-CC--C- 7194
MON.NG.x.NG1              A-C--A-C---C-TT-----AGGA--T--G...--CTT--G--ATT-A-AG--T.......--TG----A-TGA----CTA-...........T-GCA-ATC-T-GTGGG------C-CY---A-G---T-C---C-GG-A-A--AGA--C-G---G-CGCTG-C---C- 5810
GSN.CM.99.CN166           A-TC---CA----TC-----TGGA--T--A...--C-T--G--ATT-G-AA--A.........--G-G--A-TGA-----TA-...........T-CCA--TG-T-GTAGG--------CT--GA----CT----CCGAC--GA-CCT-AT-GG--------GGG-AA---G 7186
GSN.CM.99.CN71            ACTC---CA----TC-----TGGA--T--A...--CCT-----ATT-A-AG--A.........--G----A-TGA----CTAC...........T-CCA-ATT-T-GTAGGG---G---CT--GA-G--CT----C--AC-C-ACCATA----G---G---TAG-AA---G 7177
MUS_1.CM.01.1085          A-TC---C-----T------AGG------G.....---T--GG-ACC-T-AA--A.........--G----C-TGA-T--CTA-...........T-GCAGCTT-T-GTAGAT--CG---CAC-CAG---CT----C--A--A-A--AGA--C-G---G--CCAG-A--G-G 7150
MUS_1.CM.01.CM1239        A-CT-A-CC--CTT---G--AGG---T--G...--T-T--GG-ACC---AG--A.......--GG----C-TGA-T--CTA-...........T--CAGCTT-T-GTAGG---GG-C-CAC--AG----T-C--C--G--A-AC-A-A--C-G---G--GTAG-A----G 7122
MUS_2.CM.01.CM1246        A-TC-A-CC----TT--G--AGGA--T--G.....--TT-G--ATC-A-AG--A.........--G----C-TGA-C--CTA-...........T-TCAGCTG-T-GTGGAT---G-C-CCC-CAGG---T----C--G--C-A--A-A-TC-----G--CCTG----G-G 7257
MUS_2.CM.01.CM2500        A-TT-A-CA--CCTG--G--AGGA--T--G...--T-TT-G--ACC-T-AA--A.........--G----C-TGT-T--CTA-...........T-TCAGCT--T-GTAGAT---G---CCC-CAG---CT----C--A--C-AC-AGA-CC-G---G-CCC-G-A----G 7186
RCM.NG.x.NG411            AC-G-A-C-----GG-----A-GT----GA.....--GTG---G-----TCAT-A......-C-A---AGT-TTC...........-TCTCTCT--GGGTGCA-AA--GGCCA-A-GG---TTGG----GAT--...GAT-G-G-T-----AGGTG-G-GG-- 6998
RCM.GA.x.GAB1             TC-G-A-CA--C-GG-----TTCA--T-GA.....--GTG---G-G--ATCGC-A......-C-A---AGTCTTT...........-TCTC-CT--G-GTAGAGAAG-GACCA-A-GG---TTGG----GAT--...GA--GG--T----CGG-TG-A-GG-- 6954
DEN.CD.x.CD1              A--T-G--G-----C--G--AGGG--TGTA.....--CCT---GG--T-AAAC--G......-C-G----A-T--TCA-GAGA...........G-C-CTATT-T-CCA-G-A-GGC---T--GAG----GCA--C--G--AGTAG---GC--C----C--GA--AAGGC- 7309
                              V4 loop start                                                                                        V4 loop end
MAC.US.x.239              AC---G--A----G------AGGA--T--G.....---GTTTT-CC-G--ACC--T.....ATGT-T--ATTG-TT--C..............CACTC-CA-CC-ATC-AT-AT-GGCCA-AG--G---TGG---TGGT--...G---G---------GG-TG-AA---- 7667
Env                        T  M  K  C  R  R  P  G  N  K  .  T  V  L  P  V  T  I  .  .  M  S  G  L  V  F  .  .  .  .  .  .  H  S  Q  P  I  N  D  R  P  K  Q  A  W  C  W  F  .  G  G  K  W  K  D  A  I  K
H2A.GW.x.ALI              ACT--GC------AG--G--AGGA--T--G.....---GTTGT-CC---AACAC-T.....ATGT-----TT-AT---..............CACTCCCAGCC-ATC-ATAA--GACCC---------TGG--CTGGT-C...-A--GCG------GG--AG-CA-GC- 7730
H2A.DE.x.BEN              AC---GCG-----AG-----AGGA--T--G.....---GTTTT-CC---AACAC-T.....ATGT----ATT--TG---..............CACTCTCAGCC-ATC-ACAC--GGCCT--G--G---TGG--CCGGT--...G---GC-G-----GGG-AG-CA-GC- 7749
H2A.SN.x.ST               ACCG-AC------AG--G--AGGA-----G.....---GTTGT-CC---AACAC--.....ATGT-----TT--TG---..............CACTCCCAGCC-ATC-ATA---GACCC--G------TGG--CTGGT-C...-A--GCG-G-----GG-AG-CA-G-- 7176
H2B.GH.86.D205            TC---AC-C----AG--G--AGGA-----G.....--GGTTGT-CC---AA-A-C-......-TGT----ACT-CTT--C..............CATTC-CAGCCTATC-ATAAG-GACCC--------TTGG--CTGGT--...-AG-G---C----CAG-AG-CA---- 7716
H2B.CI.x.EHO              AC---GC-C----A---G--AGGA-----A...-TGGTTGTGCC---AA-A-C-......-TGT----T-TTCTC--C..............CATTC-CAGCCTATC-ATAA--GACCT-A------TTGG--CTGGT--...-A--G---C----CAG-AG-CA--C- 7708
H2G.CI.92.ABT96           AC---A-RA----GG-----AGGG--T--A.....---GT-TT-CC---TAC---T.....ATGT-----CT--TC--C..............CACTCTCA-CC-ATC-AT-A--GACCC-----G--TTGG---TGGT--...G---GC---------GG-AG-CA-GC- 7080
H2U.FR.96.12034           AC---GT-C----G---G--AGGG--T--G.....---GTCTT-CC-G-AACC--T.....ATGT----ACT--TG--C..............CATTCTCA-GCTATC-ATAAC---CCA-A------TTGG---TGGT--...G-T-G---T-----AGGGG-AA-CC- 7209
MAC.US.x.EMBL_3           AC---G--A----G------AGGA--T--G.....---GTTTT-CC-G--ACC--T.....ATGT-T--ATTG-TT--C..............CACTC-CA-CC-CTC-CT-AT-GGCCA-AG--G---TGG---TGGT--...G---G---------GG-TG-AA---- 6798
SMM.US.x.H9               AC---G-GA----G------AGGA--T--G.....---GTCTT-CC-G--ACC--T.....ATGT-----TTG-TC--C..............CATTCGCA-CCCATA-AT-AG-GACCA-A---G--CTGG--CTGGT--...G---G-G-------AG-GG-CA-CCG 7144
SMM.US.x.PGM53            AC---A-GA----G---G--AGGA--T--G.....---GTCTT-CC-G--ACC--T.....ATGT-----TTGATC--C..............CATTCGCA-CCCATA-AT-AG-GACCA-A------CTGG--CTGGT--...G---G--GT----GAG-GG-CA-CC- 7601
SMM.US.x.SIVsmH635F_L3    AC---G--A----G------AGGA--T--G.....---GTTTT-CC-G--ACC--T.....ATGT-----TTG-TC--C..............CATTCGCA-CCCATA-AT-AG-GACCA-A---G--CTGG--CTGGT--...G---G---C----GCG-GG-CA-CC- 7674
SMM.SL.92.SL92B           AC---A--G--C-G------AGGA--T--G.....---GTCCT-CC-G--ACC--T.....ATGT----TTTG-TC---..............CATTCGCA-CC-ATA-AT-AG-GGCCA---------TGG--CTGGT--...G---G--------GAG-AG--A-GC- 7103
STM.US.x.STM              AC---G-GC----G------AGGG-----A.....---GTCTT-CC-G--ACC--T.....ATGT----CTTG-TC--C..............CATTCGCA-CCCATT-AT-AG-GACCA-A------TTGG--CTGGT--...G---G-G------GAGGAG-AA---- 7306
MNE.US.x.MNE027           AC---G--A----G------AGGA--T--G.....---GTTTT-CC-G--ACC--T.....ATGT-T--ATTG-TT--C..............CACTC-CA-CC-ATC-AT-A--GGCCA-A---G---TGG----GGT--...GA--G---T-----GG-GG-AA---- 7138
DRL.x.x.FAO               A------TC----G----AAAGGA--T-GG...-GC-T--T-TCT-CA-CCTCA......-CCT-----TTGTTG---TA-.............CATGGGTT-G--CCAGG-AA----T-A-A-AG--G-ATG-----AT--...GA--G-C-----GGA--GG-----GG 7015
MND_2.x.x.5440            C-GG-A-GG----G---GAAGGGA--T-GA...T-T-TC-TTTCC-CA-CA-GT......TCTA----ATTGCTG--CTA-.............CATGGGTTGG-GCCAGG-AAG---C-T---A-G-GCA-G---C-GT-A...GA--GG------GGAC-GG-----C- 6981
MND_2.CM.98.CM16          C--G-AGTA----G----AAAGGA----GA...T-TGTG-TCTC--CA-C--GT......-C-A-G--ATTGATC---TA-.............-GTGGGTT-G--CCAGGGAA----C-T-AGA---GTATG---C-GT-G...-AG-G-C-----GGA-TAG-CA-GC- 7354
MND_2.GA.x.M14            A-GG-GGTA---GAT--GAAGGGA----GG...T-T-T-GTTTC-G---CA-GT......-C-T----ATT-ATT--CTA-.............CATGGGTT-G--CCAGGCA-G---C---A-A---G-ATG--C-GGT-A...-T--G-C-----GGA-GAG----G-- 7271
OLC.CI.97.97CI12          -----AG-G----TT--GAAGGG---T-GA...-GT-T-GTTTC-G-TAA-TCA......-CTT----TCTG-T--A-TA-.............G--GG-TT-G--CCTTAT--G-----AGATA---GCTTG----GGT--...-AG-G-C-----GGATGG-TGC--T- 6840
GRV.ET.x.GRI_677          AC-G-G-GA--CCG-------TGGT--T--A.....---GTGTTGCC-G-AACA---......ATGG-----TT--T----..............CACTCTCAG---TAT-ATACC-GGT-A-A------GTGG--CC--T-C...CA--G-G-T-----AGGGG-A-GG-- 7232
VER.KE.x.TYO1_patent      TC-G-C-CC----GG-----AGGA-----G.....---GTCCT-CCGG-AACG--A......ATGG-G--ACTG-TT--C..............CACTCTCA----TAC-ACATG--GC-T-----G--TTGG---C--T-C...GA--GC--T----GAGGTG-C-GGCG 7340
VER.KE.x.9063             AC-G-C-C---C-A---G--AGGA--T--G.....---GTCTT-CC-G-AACC--A......ATGG----TTT--TT---..............CATTCGCAG---TAC-ATACG--GC-T-----G--CTGG---C-TT-C...CA--GC-GG----GAG-AG-C-GG-- 7361
VER.KE.x.AGM155           AC-G---CA--C-A---G--AGGG--T--G.....---GTCTTGCC-G-AACG--A......ATGG----ATT--TC--C..............CACTC-CAG--GTAT-ATAC--GAC-A--G--G--CTGG--CC--T-C...CAG-GC--T-----AGGAG---GG-- 7341
VER.DE.x.AGM3             AC-G---C---C-A------AGGA-----A.....---GTCTT-CC-G-AACA---......ATGG-----CT--TG---..............CATTCTCAG-GGTAC-ATAC--GGC-----------TTGG---C--T-C...CAG-GC--C----GAGGAG-C-GG-- 6859
TAN.UG.x.TAN1             A-G--AGTG----G------AGGA--T--G.....---GTCTT-CC-G-AACC---......ATGG----TCTC-TG--C..............CACTC-CA-C--TAT-ATAC---AT-A---------TGG--CC-TT-C...CAG-GTG-T-----AGGAG-C-GG-G 7321
SAB.SN.x.SAB1C            ACT--C-GA----G------AGGA--T--G.....---GT-TTGCCTG-AACC---......ATGG-T--CTTG-TC--C..............CAT-GCCAG---TAT-ACACC-GGT---A------TTGG--CTGGT--...G---GT--T----GAGGAG-C-GG-- 7532
MND_1.GA.x.MNDGB1         A-G--AG-A---CAC---AAAGGA--T-GG...T--GT-GTG--T-CT-CATCA......-CTA----ATT-TT----TA-.............CATGGGTT-G--CCTGG-AAG---T-A-AGA---GCATG--C-C-T-C...-A--G-CGT---GGGTTAG-AC-TTG 6778
DEB.CM.99.CM40            AC-T-A-CA---GT---G--AGGA--T--G.....--CTACCG-GCC--A-AC--G......-CTA----TCTGT-C---TA-.............--T-CCTTT-T-CAA-GGCTG-GG--A-AGAG-----------GAC-A...-AC-G--GC---GCA---G-AAC--- 7004
DEB.CM.99.CM5             TCTT-GG-----GTT--G--AGGG-----G.....--GTAT-G-GCTG-T-A---G......-CCA-----CT-T-C---TA-.............----CCTTC-T-CCA-GGCTG-GG--A-AGAG-----------GG--G...----GT--T----CAGGAG-AG-T-- 6943
TAL.CM.01.8023            ACC--AC-G--C-TT-----AGGA--T-GA.......----TC---GATC-T-AG--A......-C-G----TCT-ATG---CAC.............-GTCAGATC-T-GCAGGTAA-G--T-A-AGAG---CT-------G--A...GA--G---T----CG-GGG-GA-GG- 6578
TAL.CM.00.266             AC---C-CA--C-T------AGGA--T-GG.......----T---GGACT-A-AG--A......-C-G----ACT-ATG---CAC.............-GCCAGAT--T-GCAGGGAAG--CC-C-ACAGG--CT-------GC-A...GAG-GC--T----C---AG-CA-GG- 7034
SUN.GA.98.L14             -T----C-G--C-T---GAAAGGG--T-GG...T--CA-GT---T-CTATA-GT......A-TA----ACTGCTG---TA-.............TATGGGCT-G-GCCAGGTTCT--GT-A----TG-----G-----AT--...GT--G-C-----GGA-GAG-C--TGC 7503
WRC.CI.97.97CI14          A-G--AG-A---CAT--GAAAGGT--T-GA...T-TTT-GT-TC-G-AAA-TCA......-CTT----TTTGAT--A-TA-.............G--GG-TT-G--CCTCATA-G-----T---A---GCTTA--C--AT--...-CT-GGCTT---GGA--A-TGC-TT- 7347
WRC.CI.98.98CI04          --G--AG-A--C-AT--GAAGGGG--T-GA...T-TGT-GTGTC-G-AAACTCA......-CCT----ACTTAT--A-TA-.............G--GGCTT-G--CCACATA----C--A--GA---GCTTA--C--GT--...-A--GGTTG---GGAC---TG---T- 7395
WRC.GM.05.Pbt_05GM_X02    -CT--AG-A---CAT---AAAGGA--T-GA...-GTCT-GTCTC-G-AA--TCA......-CCT-----TTGATT-A-TA-.............G--GG-TTGG--CCTTATAAG-----T--GA-----TTA-C--AT--...-C--G-C-G---GGAC-T-TG---T- 6291
LST.CD.88.524             ATC----GA----T----AAAGGA--T-GG...T--GA-GTCTCC-CAATC-GT......TC-A----ACTTCTG---TAC.............TATGGGTT-G--CATGGGTCT-GAC-A---TTG-----A--C--AT--...GAG-GTC-----GGA-GA-TG--T-- 6378
LST.CD.88.447             ATT----GA----TT--GAAAGGA--T-GG...T--GA-GT-TCT-C-AT--GT......TC-A----ACTTCTG---TAC.............TATGGCTTGG--CATGGTA-C-GGC-----TTG--T--G-----AT--...GA--G-C-G---GGA-GA-TG--TC- 6375
LST.KE.x.lho7             ATT----GA--C-T----AAAGGA--T-GA...T--GA-GT-TCC-CAAT--G-......TCTA----ATATTT----TA-.............TATGGCCT-G--CATGGCA-C-GAT-A---TT------G-----AT--...GA--GCC-----GGA-GA-TG--T-- 7444
SYK.KE.x.SYK173           C-G----TA---GA------AGGA--TG-G...T-T-TC----AT--T-AAT-G......-CTG----CTATTTCC-CCCAGTAATA...CAAGGG-A-TTG---ACCGGGA-GG-CGCA-A--GT--CTTC---CGAG-A...-C--G---T----CAG-ATT---T-- 7240
SYK.KE.x.KE51             C--C-GGTG---GA------AGGA--TG-G...T-T-T-----AC----AAT-G......-C-G----ATTTTTC--GCCAGTGATT...CAAGGG-G-TTG---ACAGGGAA-GC-GC--A-AG----TT---C--AG-C...-A--G-G-----GGA--GTT---TG- 6936
COL.CM.x.CGU1             -CTT-AGGG----T----AGAGGA--T-G-...T-TCAC-G--ATC-AAAC-CT......-CCAAT--AGC-AA---CTAC.............TATGAGCTT-TTCCTTATTC---AGGC-TTT-T-G-AGG---C-GT--GTACC-ATG-CTG--C-G--T-AGAA--- 6662
```

```
H1B.FR.83.HXB2     ACAGATAGCTAGCAAATTAAGAGAACAATTTGGAAATAATAAAACAATAATC.....................................................TTTAAGCAATCCTCAGGAGGGG 7326
Env                 Q  I  A  S  K  L  R  E  Q  F  G  N  N  K  T  I  I  .  .  .
```

```
                                                V4 loop start
H1B.FR.83.HXB2      ACCCAGAAATTGTAACGCACAGTTTTAATTGTGGAGGGGAATTTTTCTACTGTAATTCAACACAACTGTTTAATAGTACTTGGTTTAATAGTACTTGGAGTACTGAAGGGTCAAATAACACT.............................GAAGGAAGTGACACA 7463
Env                 D  P  E  I  V  T  H  S  F  N  C  G  G  E  F  F  Y  C  N  S  T  Q  L  F  N  S  T  W  F  N  S  T  W  S  T  E  G  S  N  N  T  _ __ __ __
```

```
                          V4 loop end                                                                                                  V5 loop start
H1B.FR.83.HXB2            ATCACCCTCCCATGCAGAATAAAACAAATTATAAACATGTGGCAGAAAGTAGGAAAAGCAATGTATGCCCCTCCCATCAGTGGA.........CAAATTAGATGTTCATCAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTAAT..................AG 7606
Env                        I  T  L  P  C  R  I  K  Q  I  I  N  M  W  Q  K  V  G  K  A  M  Y  A  P  P  I  S  G  .  .  .  Q  I  R  C  S  S  N  I  T  G  L  L  L  T  R  D  G  G  N  .  .  .  .  .  .  S
H1A1.UG.85.U455           --A--T----A---T--------G-----------T---------G-------C----------------------CAA---.........GT---A--G---GA------C--------A--A-----------------G----C-AAC................ 7037
H1B.US.90.WEAU160         -----A------------------------------G-------G-----------------------A------GAA---.........-------------T------------------T-A----------------------G-AGT...............GA 7610
H1C.ET.86.ETH2220         -----A----A---------------------------------GG-------CG--------------------TGAA---........AT---A-TG---AG---------C-----A--A---C-G---C-T-----A-CC--A..............GAGCC 6985
H1D.CD.84.84ZR085         --A---A--------------------------------------G-------------------------------TGAA--G.........-------AT----------------------------A-----G---------------GC-...................-A 7127
H1F1.BE.93.VI850          -----T--------TC-G------------G--------------GG-------CG----------A--T-------TGCA---.........A-C----CC---AAC---------------TT-A-----G-----------------..................... 6884
H1G.SE.93.SE6165          -----A--------T-AG------------G-G-GA----------GG--G---C-----------------------GCA---.........A-C---GA----AAC----GC--------A--AA------------------G---AAT......AACAATACA-A 7024
H1H.CF.90.056             --T--A--G---------------------G------------------G------CG-----------------------CAA---.........A-C----TG---GT---------------A--AA----G----TT--C-AG-----C...................GC 6925
H1J.SE.93.SE7887          -----AA-A-----T-A-------------G-G-GA---------A-GGAC----C-------C---------------TGCA---.........A-C----CG--CA---------------AT-A-----G---C-T-----A--A--CAGG......GGGAATGGC-- 6913
H1K.CM.96.MP535           -----A--------T-------G--G------------------------G-----------A--------------TGCA---.........AGT----AC---AGC---------------AA--A----G---------------......................-A 6766
H101_AE.TH.90.CM240       -----A--T-------AG-----G---------------------GG--C----C-------------T------------------.........AG-----AT---GT---------------AA-A-----G---------------GT-AAT................. 7160
H102_AG.NG.x.IBNG         --A--T----A-----------G------G----T---------------------C------------------------CAA---.........AT---A------GAT-----C--------A--A------------------A--G---.................. 7113
H1N.CM.95.YBF30           --T--T--G--T--T-------G---G---G----TT-----ACA-GG----------G---T-----A--A--A--TC-G---.........GTTC---AC---A-C--C--------T--A---G-TC--GA-TATAG------GCC-................... 7152
H1O.BE.87.ANT70           GG---T--A--T----A-C-G-G---GG-GG---GGTCA---ATA-GG-G-CAGTCG-G-C-C-----A----------AA--T.........A-TC-A-C----ATG-----C--A--T--AA--A-CC--CA--TG---AACACATGGAACAGC.............. 7649
H1O.CM.91.MVP5180         GGT--TA-A--T----AGT---G---GC-AG---GATCA---AT---G-G--AGTCGAG---C-----A---------CCC--C.........A-CT-A-C----CAT--C--C--A--T--AA--A-TC--CAGTT----CAACCATGG.................-A 7662
CPZ.CD.90.ANT             ...-TAGCA-AT--------T--G--G--AG-T--TCAT---GGC-T---TTC-----GC--T--CTTAG-C--T-GG--A--G.........A-TG--TCC---A-T--C-GC--A--T--AA-TA-G--G...GA--G-CAAATAT--.................... 7036
CPZ.CM.01.SIVcpzCAM13     --G--AA-A-----------T-G------AG-G--TTCA---AT--G--------GG-G---C-----A-----A--ACCA---.........--T--A-CT---AAC--TCTA-----T-----CA----G---C--...-ACCA-GT-..................... 7182
CPZ.CM.05.SIVcpzEK505     ------T-A-A-----------G-------G----TT-A---ACA-G-----------G---T-----T--G--A--T-AA---.........-CT----AC---CTC--T--C--C--T---A-AA--C-TGAC.........................TATACAAAGTC 7086
CPZ.CM.05.SIVcpzMB66      -----TA-T--T----------GG------G---GGC-A-----A-G---------G--G---A-T-CTT--C-----T--A---.........AC---A-AC---ATC--G-----C------A-AT-G--TG-G...............CAGCAA...AAAACAGAC-- 7227
CPZ.CM.98.CAM3            ...TT-T-------T-------GG------G-----TCC---AT--G---G--G--G-G---T--------A--A--T--A---.........GTT--A--T--CA--------CG----T---A-TA--C-TGAG-C--GGCA---C-TC..............AAC-A 7016
CPZ.GA.88.GAB2            TA--TT----------A---T-GG---G-AG----TTCT---AT--G---G--G----GT--C-T------G--T-GA--A---.........ACT--A-CC---AAT--C-CC---------A--C--C---GA-GT-C-GAA-......................G- 6975
CPZ.CM.98.CAM5            C-A--AA-AT----T-------GG------G-----TCC---AT--G---G--G----G---C--------A--A--A--A--G.........A-C--A-C---CAT---------A--T--A--C--CC-CGA-GC--GA--AAAATGGGAA................-A 7328
CPZ.CM.05.SIVcpzMT145     -CA-TAA----T--T-------G------AG-G--TTCC---AT-CGT-----G----G---T-T---A--A--A--T--A--T.........A-C----CC--CA-T--T-----C-----TA-----C--GAG-T-C--AAGAACCGAGAAGATCAGGGAGAAGATCA 7103
CPZ.TZ.00.TAN1            ...GTTGCA---------C-G-GG--G--AG----TCAT---GGT-T---GTC-----GGG-T--CCTT--C--A-GG--G---.........AC-G-A-A----CAC-----C--C-----A--TA-CA-G---GC---AAAA-AC--CAAT................. 7230
CPZ.CM.05.SIVcpzLB7       T-G--A--A--T--T--G--T-G-------G-----C-A----TA-G---------G-G---A-T---T--A--T--T--A--T.........A-C--CC-C---AAT--C-C------T---T--A-CC--GA-.........................AAACAC-C 7117
CPZ.US.85.CPZUS           ...--A--TAAG--T-------G-------G----TCA----AT--G-----------G---C-T-------A--A----AA--T.........GTGC-A--T---AAC---------A--G--AA--A-TC-TGAC-T-AGCATAA-CGCAGTC............AAT-A 7608
CPZ.GA.88.GAB1            ...-TA--G-----------T-G---G--AG-T-GTTCA---AT--GG-------G--G---A-------A--C--A-----A---.........A-C--C-CT--CAAT--T--C-----T--A--T--C-----TTC---CAC-CC-GTGACT...............-A 7628
MON.CM.99.L1              GGGTGG--AG----T-CC----G-----GG-G-CACAA---GGAT-T--GTC---GAGT--C--CCTA-----ACGG-AA--C.........--TG-GCAG---A----T-----C--T-CC--C--GA-T---G-G--GTTGTA-C-AAAC................ 7579
MON.NG.x.NG1              GAGTGGT-AA-G---C--C-C-GG---C-AG---CTCA----GGAT-CA-CTCC--GAGTR-C---CTG--A--A-GACAA--G.........--TG-GCAG---A-C--T-----C----CC--CT-GG--GATG-G--GTTGTA-GGCAGC................ 6204
GSN.CM.99.CN166           CAATGG--GG-T----A---T-GG---T--G-C-C-CAA---GGAT-T--GTCT---T----C--CTTA-----ACGGCAG--T.........--T--A-AT--CA----C--C-------T---T-GA--GATG-C-C-ATGTA-G-AAGC................ 7553
GSN.CM.99.CN71            CAATGGA-GG-T--------T-GG---T--G-C-CACAA---GGAT-T--GTCT-GGT-G--C--CTTA-----ACGGCAA--T.........--T--A-AT---A----T--C--C--T--C--C----A--GACG---CGATGTA-GGACAT................ 7559
MUS_1.CM.01.1085          CAATGGA-GGTC--T-----T-G----T--G-C----CA---GGAT-T--GTC---GT-T--C--CTTG-----ACG--AG--C.........-TTG--C-----GAG--T-----A---------T--CA-TGACG---CCATGTA--GAGAT................ 7547
MUS_1.CM.01.CM1239        CAATGGA-GGTC--T--G--T-G---GT--G-C-CGCAA---GGCT-T--GTC----T-C--C--CTTG-----ACGAC-G---.........--TG-GCAG---A-T--T---G-G--T-----C--TA-TGATG---C-ATGTA---AGAC................ 7519
MUS_2.CM.01.CM1246        CAATGGA-GGTC----------GG--GT--G---C-CA----GGTT-C---TC---GT-C--A--CCTG--C--TCGA-AA--T.........--TG--CA----A----CC--G-A--------AT--A--GATG-G-CCATGTACGGCAAT................ 7654
MUS_2.CM.01.CM2500        CAGTGGA-GGTT--T--G----G----T--G-T-C-CA----GGTT-TACGTCG--GT-C--C---CTG--C--TCG-CAA--C.........--T--CCAG---A-G--C---G-G--T--T-----C--CA--GATG---CCATGTA-GGCGAC................ 7568
RCM.NG.x.NG411            --GTTTG-T--T--T-TC-CT---ATGG--G-T--TGAT---T-T-C---GTCT-G-AA-G-A--CA-T--A--A-GGCCA-AT.........GCCT-G-----CAACG-C-C-G-G---TACT--T--GC-GAT-TT--CTA-ACAG--AAT................ 7353
RCM.GA.x.GAB1             CCTTTTG-A-------TC-C----ATG--AG-C--TGAT---T-T-C----TCG-GGAAGG-A--CA-G--A--G-GGCCA-AT.........GCGT-A-AG--CAGTG--C-GG-AT-CTAT---T-GGC-GAC-T---CTA-.......................... 7300
DEN.CD.x.CD1              -CATATA-GGTG--T-AC---------T-AG-G----CA-----C-C------C-TG-AT--T--CTTA--A--AGCAGAA---.........G------C----CA-C--C---G-GT----CA-TA-TG--GAC-CG----TCCA-CCGAAC................ 7649
                                                                                                        V5 loop end
MAC.US.x.239              -ATTA-G-G-----TCAT--T-G------A--C----CT-----T---------C---AATG-T---TTG-----A-GAGAG---.........G-CC-C-CG---AAC--C-CAG-G--CA-T--CA--GC--AC-T----T-GAT-G--GGA................. 8028
Env                        N  Y  V  P  C  H  I  R  Q  I  I  N  T  W  H  K  V  G  K  N  V  Y  L  P  P  R  E  G  .  .  .  D  L  T  C  N  S  T  V  T  S  L  I  A  N  I  D  W  I  D  G  .  .  .  .  .  .  .
H2A.GW.x.ALI              -ATTATGCA--G---CAT-----G-----A--T--T--C-----C----C---G---AATG-A---TTG-----T-GGGAA---.........G-GT-G-CC--CAAC----CAG-A--CA-CT--A-TGCT-AC-TT--CACG-A-GGC.................. 8070
H2A.DE.x.BEN              -A-TAT......---CAT-----G--G--A--T--T-CC-----T---------G---AATG-A---TTG-----T-GGGAA--G.........G-GT-GGCC---GA-----CAG-A--CA-CA-AA-TGCT-AC-TT--CATA-A---AAAT................. 8083
H2A.SN.x.ST               -ATTATG-G------CAT-----G-----A--T--T-CC-----C--G-----G---AATG-A---TTG-----T-GGGAA---.........--GT-A-CC--CAAC--T-CAG-G--CA-CA-AA-TGCT-AC-TT--C--A--AG-G.................... 7516
H2B.GH.86.D205            -ATTATGCG-----TCAT--C-G---G--A--C----C------C--G--T--G---AAT--A---TTG-----A-GAGAA--T.........G-GC-CTC----AAT--C-C-G-G--TA-C--CA--GCC-AC-TCA-CTCA-AC..................... 8059
H2B.CI.x.EHO              -ATTATGCGT----CAC--C-G---G--AG-C----C------C--GA-T----G-AATG-----TTG-----A-GAGAG--T.........G--C-CTCC---AAT--C-C-G----CA-C--CA--GCC-AC-TT--CT-GATAG--AAG................. 8048
H2G.CI.92.ABT96           -ATTATG-A--C---CAT--T-GG--G--A--T----CA-----C-------------AATG----CTTG--C----GAGAA--T.........G-TC-G-CG---AAT----CAG-C--CA-C--AA--GCG-AT-T---A-AGAA--GGTCTAAC.............. 7447
H2U.FR.96.12034           -A-TATG-G--T---CAT--T-GG--GG-A--T----CA-----T---------G-G-AATG-----TTG-----A-GAGAA---.........G-TC-C-CC--CAAT--T-CAG-A--TA-CA--A-TGC--AT-TT---C-GAT---CAAA................. 7570
MAC.US.x.EMBL_3           -ATTA-G-G--G--TCAT--T-G------A--C----CT-----T---------C---AATG-T---TTG-----A-GAGAG---.........G-CC-C-CG---AAC--C-CAG-G--CA-T--CA--GC--AC-T----T-GAC-G--GGA................. 7159
SMM.US.x.H9               -ATTATG-G-----TCAT--TRG---G--A--C----C------C---------C---AATG--R-TTG-----T-GGGAA---.........G-CC-G-C----AAC--C-C-G-A--TA-C--CA--GC-GA--T----T-GATC---GGC................. 7517
SMM.US.x.PGM53            -A-TATG-G-----TCAT--T-G---G--A--C----C------C-G------T---AATG-----TTG-----T-GGGAA---.........G-CC-G-C----AAT--C-C-G-A--TA-C--CA--GC-GAT-T---CT-GAT-G--GGC................. 7977
SMM.US.x.SIVsmH635F_L3    -ATTATG-G-----TCAT--T-G------A--C----C------C---------C---AATG-A---TTG-----T-GGGAA---.........G-CC-G-C----AAT--C-C-G-A--TA-T--CA--GC-GAG-T----T-GATC---AAC................. 8050
SMM.SL.92.SL92B           -ATTATG-A--T---CAC--C-GG-----A--T--T-C------C-GG---------AATG-C---TTG-----A-GGGAA---.........G-CT-A-CT---AAC--C-CAG-G--CA-C--AA--GC--AT-TT---T-GAT-G-CAAC................. 7479
STM.US.x.STM              -ATTATG-A------CAC--C-G----G-A--T--T-CA-----T---------G---AATG-T---TTG-----A-GACAA---.........G-CC---CT---AAT--C-C-G-G--CA-TA-AA--GC--AC-TT--CT-GAC---CAAT................. 7682
MNE.US.x.MNE027           -ATTA-G-G-----TCAT--T-G------A--C----CT-----T---------C---AATG-T---TTG-----A-GAGAG---.........G-CC-C-CG---AAC--C-CAG-A--CA-T--CA--GC--AC-T----T-GAT-G--GGA................. 7505
DRL.x.x.FAO               CCATATGCAT-C---CAC----------A---G-TGAT---GCA-G------C-GGAAG--T---ATG-----TG-ATCA---TTTAATAATG----A------A--CA-G--G-C----AAA--T-C--TGAG-TCC-AAAATTGG-GACAGAG......GAAACAG- 7444
MND_2.x.x.5440            CCTTA-A-GT-TC---AT--T-G------AG--GGTGAT-----T---------C--GAA---T---ATG-----TG-GTC---GTTTAATAATG----A------A-TAATG--G-G----AAA--T-C--TGAGGTCC-AAAAAC-G--GATGAC.............. 7390
MND_2.CM.98.CM16          CCATGGA-GT----T-AT--T-G------A---G-TGAT-----T-----------GAAG--T---ATG-----TG-ATC----TTTAATAATG-G--A--------TAATG--G-A----AAA--T-C--TGAGGTTC-AAAGACAG-AGAGGGA.............. 7763
MND_2.GA.x.M14            CCATATA-GT----T-A---C-G---G--AG-GG-TGAT-----T-----G----GGAAG--T---ATG-----TG-GTC----TTTAATAAT--T--A-----CA-TAATG--G-C----A-A--T-T--TGAGGTCC-ACTA-T-G-AGGA................. 7686
OLC.CI.97.97CI12          CCATGGTCA-AG--T---C-G-G-----A-G--G-TGAT---AT-----------G-AG---T---CTT-----ACAACCA---TTCAATAATAG---CC-----...............---CAAAGAG--ACTGA--TGTT-TTG-AA-TACAAAAAGTAGAGAAC..... 7246
GRV.ET.x.GRI_677          -CTTATG-TG-C---CAT---CG----G-AG----TGAT---T-C-CT--CTCT---AAGG-A-----T--A--A-GGGAA--T...........--TT-GGAG---AAC---TCAG-C--G-CA--ATACG-GG---T----TA-AAC--CAAGTCT.............. 7623
VER.KE.x.TYO1_patent      -CGTATG-TG-T---CAT--C-GGTCTG-A-----TGAT---T-T-CCC--TC---GAA--CT-----T--G--A-GAGAA---...........--TT-GCA---CAGG--C-CAG-C--T---A--AC-G-TGAGCTTA--TA-AA--G-AAA................. 7737
VER.KE.x.9063             -CTTATG-TG-T---CAC--C-G-TCTG-C-----TGAT---T-T-CCA--TC---GAA-G-------T--A--A-GGGAA---...........--TC-GGA---CA----C-CGG-C--T--TA--AC-G--GA-CT-A--TA-AAC---AAC................. 7761
VER.KE.x.AGM155           -CTTATG-TG-C---CAT--CCG-TCTG-C-----TGAT---T-T-CTA--TC---GAA--CA-----T--A--A-GAGAA---...........--TT-GCAG--CA-G--C-CAG----T---A--AC-G--GAGCT-A-CTA-AA---CCAG................. 7741
VER.DE.x.AGM3             -CATATG-TG-T---CAT---CG-TCTG-C--T--TGAT---T-C-C-C--TC--GGAA--CC-----A--G--A-GAGAA--G...........--CT-GCA---CA----C-CGG-A--G--TA-TC-G-GGAGCT-A--TACAA--G-AAG................. 7259
TAN.UG.x.TAN1             -CATATG----C---CAT--T-GGTC----G-C--TGAC---T-C-CGC-CA-T--GAA--CC-----A--A--A-GGGAA---...........--CT-AGA---CA------CAG-G---TCTA--A-GG--T--CTG---TA-AAC-GCAAA................ 7721
SAB.SN.x.SAB1C            -C-TATG-T--C---CAC----GG---G-AG----TGAT---T-C-CCT-GTCT---AAG-CA--C----------GAGAA--C...........--CT-GGAG---AAC--G-CAGCC----C----TATG--GA-TTGA-CTA-AAC-G-AAA................ 7917
MND_1.GA.x.MNDGB1         CCTTATGC-T----TCA---T-GG-----AG--G-TGAC---AT-C----------G-AAG--A---TTA--A--T-CATCA---TTTAATAAT--C--A--G---A--CAT-GGG-A----AAA--TAC--TGA--TGC-AAAGATAG--AGTAATGAA........... 7211
DEB.CM.99.CM40            TGGTTGA-G--T--T--G-----T--GT-AG------CT---T-T-CG-----GC--CAT--C---CTA-----G-AAGAA---.........G-GT-G-AG--CAGC---C-C---T-T-CAT-TG-T--TGATGT----CACTA----GGG................. 7350
DEB.CM.99.CM5             -CTTTGA-G--T--T-AG--T--T---T-GG------CA---TTC--G------C--CATG-A---TTG-----A-AGGAA---.........G--T-G-AG---CAT--TTC-G--T-T-CA--TG-T--TGAGGT--C-CA-TA-GGGCAA................. 7271
TAL.CM.01.8023            T-TTT-A-GT-C---CAGT---GG-----G--C--T-G----AGC-------AG--GCT-------CTG--G--A-GAGAG---...........--CC-G--T--CAA-AGC--CG-C-----TA-C--GG-GGAC-TCA-CCACTA---GGGGAAT.............. 6990
TAL.CM.00.266             T--TA-A-GT-----CA-C-G-GG-----G--C----G----AGC-GG--G-AG--GTT------CCT---A----GGGAA---...........--CC-GCAG--CAA-AGT--C--C-----AT--T-GG-GGAT-TCA-CCACCACCGCGGCTCT.............. 7440
SUN.GA.98.L14             CCTTGGGC-AAT---CAC------AGTG-G---G-TGAT---GC-TC---G--C--GAAG--T--CTTG--C--A-CATCA--GTTTAATAACAG---A------G--AAC-GAG-G----AAGCATGG--C---CTG--AA-G-TGG-AGACTGG...AAGGTAAATG- 7929
WRC.CI.97.97CI14          CCATGGAGT-------C---T-G-A-T--AG--G-GGAC---GCAC-T--G----T-AA-G-A---TTG-----AG-ATCA--GTTCAATAAT--T--------C.............--TCAAAGAG--AC-GA--T-TTGTT-AACTGGGAA................ 7738
WRC.CI.98.98CI04          CCTTGGAGT-------CG--T-G-A-T--AG-GG-TGAT---GCAC-T------GT-AA-G-A---TTG-----AG-ATCA---TTCAGGAAT--T--A------...............---CAAAGAG-GACTGA--TTCTATTCAA-TGGGAA................ 7783
WRC.GM.05.Pbt_05GM_X02    CCTTGGGCA--T----CT--C-G-TC---AG-GG-TGAT---GCATCTA----G--GAA---A--CTTG-----GG-GTC----TTTAACAAT--C---------...........--CCAAAG-G---C-GA-GT-CTCTT-AA-TGGGAA................ 6679
LST.CD.88.524             CC-TGGGC-AAT---CAC------TC-------G-TGAT---GCA-CT--T--G--GAA---T---CTT-----A-CATCA---TTTAACAATAG---A------A-TCAC-GAG-G----AAA--TGG--TGA--TG--AAAAT-GG-ACCACAT......GAAGACTT 6786
LST.CD.88.447             CC-TGGGC-AAT---CAC------TC-G-----G-TGAT---GCATCT-----G--GAA---T---CT------A-CATCA--CTTTAACAACAG---A------A-TCAT-GAG-A----AAA--TGG--TCA--TG--AAAGTT-G-ACCAGAA......GAAAACCT 6783
LST.KE.x.lho7             CCTTGGGC-AAC---CAC------AGC--C--TG-TGAT---GCT-C-A-------GAAG--T---CTA--A--T-CTTCA--TTTCAACAATAG---C------A-TCAT-GAG-A----AAA--T-C--TGAG-TG--AAAAT-GG-ACCACAT......GAGGACTT 7858
SYK.KE.x.SYK173           C-TCGT--GT----TGCC--TCGC-----A--T--TGAT---AGAT-T---A-----TTG--A--CCTG--C--T-CGGCA--G...........--T----A----A-T--C---G-C--T-CAG-AT-GACGGAC-T---ATACTACCCAGGA................. 7577
SYK.KE.x.KE51             T-ATAT---T-G--TGCC---CG---G---------TAT---GG-T-T---ACC--GCT-------TTG--A--T-CAGAA--G...........--C----AG--CA-T--T--------T-CTG-AT--AC-GAT-T---GTACTA-CCAAAT................. 7273
COL.CM.x.CGU1             CCAGT-A-GAAC--T-CC--T-G-A-GC-GG-G--T-GT---GTC-CTCAT-CC-GGATCC------GG--G--TCCAG-A--T...........--TC-ACA---AAC-GGG-G-AACAGCCAG-TA-TGCGTTC-TG-GCACCATAG-AGGA.............GAC-A 7001
```

```
                                   V5 loop end                                                                                                                        Rev Responsive Element (RRE) region
H1B.FR.83.HXB2              CAACAATGAGTCCGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAAATAT...............AAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAAGAGAAGAGTGGTGCAG...............A 7746
Env                          N  N  E  S  E  I  F  R  P  G  G  G  D  M  R  D  N  W  R  S  E  L  Y  K  Y  .  .  .  .  .  K  V  V  K  I  E  P  L  G  V  A  P  T  K  A  K  R  R  V  V  Q  .  .  .  .  .  .
H1A1.UG.85.U455             .--T-CAA-AAAT----C--------------------------------A---------------G---..............-----------------------C-----------------G-----------------G--...............- 7176
H1B.US.90.WEAU160           AG-G--CC--A-----------------------A------A--T-------------------------..............------------------------------------------------------------...............- 7750
H1C.ET.86.ETH2220           AC-T-GCACAAAA-----A----------A---------------------------------------..............-----G---G----A-G--C---------------T---C----A--G---------G--...............- 7125
H1D.CD.84.84ZR085           T--T-C-C--AAT--C-C---T------------------A-----C----------------------..............-----------------------C---------------------------------G-A...............- 7267
H1F1.BE.93.VI850            .G-G-G-A-CATT----C-----------AG------A-----AA-----------------C-------G---..............---------G-------------------------------A---CA-----------...............- 7023
H1G.SE.93.SE6165            T-CA-G----AGT----------------------------------------------------G---..............---AC---------CA--T--C-----------------G-----G-------------G--...............- 7164
H1H.CF.90.056               GTCTGCA--AAATT-T-C------------------------------------G-------------..............-----------------------C-G---A-------------A---G-------------G--...............- 7065
H1J.SE.93.SE7887            TG-G----GCA-T----C-----------AC-------A-----A-----------------------..............---------G-G-----G--C-----------------------A-------------G--...............- 7053
H1K.CM.96.MP535             T--T-C-C-TAAT----C---------------------C-------------------C-G--------..............---------C-G------------C-----A---------G---C-G-------------...............- 6906
H101_AE.TH.90.CM240         .--T-CG--TAA-----C---------------------A-C--A-A---------------------..............---------C------------C-----A---------G---------------G--...............- 7299
H102_AG.NG.x.IBNG           .--T-G-ACAAAT--A-C------G--------------------------------------G---..............-----------------------C----T-----------CGT-----A-----------G--...............- 7252
H1N.CM.95.YBF30             .G---CCA--GAAACA--AG-ATAT--CTC-------A-C---GTTA-TCTC------CAA--G--G-----G--C..............---------T-GC--A-----CA-------------AGGT--A--T--A---C-CACA---AGT...............- 7291
H1O.BE.87.ANT70             .-G---CA-CAAT-TA-CA--T-----AAT---G-----C----AA--T-TA-------C------G-TC--C--CAAAGTAGTAAGGGTA---CCTT-T-GTG-G-C---CAC-C-TA-T---AGGC-AGTCAT--GC-CT---ACTCAT..................- 7800
H1O.CM.91.MVP5180           TTC--CA-GTGAAA-T-CAC-T-----A-T---G----------AA--T-TA-------C-A----G--C--C--C..............---------C-G--AA----T--TA-T--------T--A--AATGTCA---CC-A-AA-AA-CATTCACACCCCTCAC- 7817
CPZ.CD.90.ANT               ....-----AA-T-TT-AAG-GTC-----CT-C-A---TAGCA.....--C-A------GCG--G--G-CC-GG--C..............C-G--G---G-G----RT--C--GTC------C--A--A-CA-GN--A--GCC--AAA-AA-ACAACACTCC.......- 7178
CPZ.CM.01.SIVcpzCAM13       .--T---ACTAATA-T-CA--T-----AAT---------C----AAA---T------G------C-C--------C..............-----G--GCGC--A---------TCT--G-----A-----A-----A--G...CATACAGTAGGA......GAAAGG- 7327
CPZ.CM.05.SIVcpzEK505       TGGA-C----AAGT-C-C-A-ATAT--CAC---G-----C----CTA--CTC-----GCAG--G--G-----G---..............--G--G--C-GC--A-----CA-------------AGGA--A--T--A---CACACA---ACT...............- 7226
CPZ.CM.05.SIVcpzMB66        G-TG--CA-A--A-CC--G----C-----TT--G-----A-----AA-T---------------------C------..............----------G---------C--------------------------A------AC------T...............- 7367
CPZ.CM.98.CAM3              --G--TCACCAA-ATC-CGC-ATAC---AC-------A-C---GTA---TTG-----GCT------GC-------C..............--G-----T-GC--A-----CA-------------A-GT--A--T--A--GCACACA--ATCA...............- 7156
CPZ.GA.88.GAB2              --CGGGGA-CAA-ACAGAAG-TTAC-TGTC---G--------------A--T-T------G----------------..............---A-------------G---C-G--G--G-----A--A--A--------GTACACA---GCAAAAGCA...CTTGAT- 7127
CPZ.CM.98.CAM5              -......A-TA-TATA-CAG-GTAC---TC---G-----A-----A---TTG-----GCT-------A-C-GG---..............---------T-GC--A-----CA-------------A--T-----T--A--GCACAC----GCA...............- 7462
CPZ.CM.05.SIVcpzMT145       GG-TC-GA-TAA-ACATA-G-TT-T-TAAC---G--TA-C----AA--T-T-------------C----C-----C..............---A---TG-G--AC-----C-------------T--T--AAGT-GA---TAT-CA--AG-AAAACAACACCAT...- 7255
CPZ.TZ.00.TAN1              ..........AATAGTTATAC-CCC-AATTTTCT-CT-TAG-AGAA---T------A-GTA------GCA-G----..............-----G--GG-----C-G--C--GTC---G--T--A-GGCCA-G---A--GCCT-AAA-TA--GCCAATCAT...ACT- 7372
CPZ.CM.05.SIVcpzLB7         A----CA--CAATATC-C----C-T--AAT------G-------CA----TC---------------------C---..............---A-----------------------------G--T----G---G-GA--G--GACT---G-T...............- 7257
CPZ.US.85.CPZUS             -G-T-G-AG-AATATA-CAG-G-TG---AC-------------C--CTTTA---A--A-------C-----G---..............--G--G--C-GC--A-----TA-------G-----AGGT-----C--A--GCATACA---A-A...............- 7748
CPZ.GA.88.GAB1              -----G--GTAA-CT--CA--T------AC-------A-C----AA--T-T------------GC-C-----G---..............---------TCG---A--G--T---TC---G-----A--A--A----GA--GCATACA---GCAAGACAG...AAAGAC- 7780
MON.CM.99.L1                .............A-CG-GAC-CTTGT-CCCTCG-C-C-GG----T---TCA------TCA--GC-CAG-CG----..............--G--G--GG-------C--C---TCTA-G-----T--A-CA--GC-------GACT-G-GTA............CAC- 7709
MON.NG.x.NG1                .............ACAG--ACGTTGAT-CC-TC--C-CGAG--TCT---GCA---C-CTCA--G--GAGC--G--C..............--G--G--CG-G--C--C--C--GTCT--G-----T--AGCA--CC-A--G---CAT-G--TA...............- 6331
GSN.CM.99.CN166             .............TCT---AA--TGA-CCCGTCG-CC---G--GCT--TGCG---ACTA------G-CC-GG--C..............-----G--GG-G--A--T--GC-CTCGA-G--C--A--TCC---T--A--G-A--AACAC-CC............GCTG 7683
GSN.CM.99.CN71              .............TCC---AAT-TGA-CCC-TCG-CC---G-AGCC--TGCA---AGTA---G----CC-GG--C..............--G--G--GG-G--A--T---C-TTCGC-G--C--A---TC---TC-A--G---CCACAT-CAGGGGTG......CATG 7695
MUS_1.CM.01.1085            .............TCA---AAT-TGA--CC-TC--C-A--G--TT----GCC---C---TG--GCCTA-G--GGTACAA..............--G--G--GG-G--C--T---C-TTCCA-G--------GG-A-TGC-A-----GCCT--C-CA............CATG 7680
MUS_1.CM.01.CM1239          .............ACAG--AA--TGA--CC-TCC-CGA--G---T----GCC---C---TG---C--A-A--T--C..............--G-----GG----A--C--GC-TTCCA-G--T-----AG---TGC----G--GCCT--T-CC............CATG 7649
MUS_2.CM.01.CM1246          .............ACA--AAA--TGA-CCCCTC--CCA--G-C..........----CGC---CG-CTAGACTGAGAGAT..............-C---G---G-G--A--C--TC-TTCC--G-----G--AG-A-TGCGC--G---CCT----CC............CACG 7778
MUS_2.CM.01.CM2500          ..........A-T-T--ATCGATTGA-CCC-TC--CCA--G--CTT---GC----C-C-TG------A-G-G---C..............-----G--GG----A--T---C-GTCC--G-----A--GG-A-TGC-------GCCT----C-............CATG 7701
RCM.NG.x.NG411              ..........ATGACA-CAAATGTCA-CTTGTCT-CT--AG-AG-A--T-TA---GC-GCA----GGGC-G---C..............--------GG-G--ATG--A-T---TAT-----A--A--T-TC-GA--GTAC-A-ACA..................- 7480
RCM.GA.x.GAB1               .-TT-----CAGT----CAAA--TCA-CCTCTC--CG--AG--G-T--TT-----GC-GCA-----GGGG-G----..............--G-C-A--G---CAG---A-T--CT-T-----A--AG-AAT---A--GTACCA-AC-..................- 7436
DEN.CD.x.CD1                .......AGAA-AC-T--AACATTGA-GCTCTC--CT---G-----C-AGTC------GC------GGCA-G--GG..............---T-----G-GT-AAC----A-T--CC-T--T--T--TTCA-TG-GAC--TAT-AA-GA-CAGAATCGGCAGAGTCTG 7797
                                                                                                                                                                        Rev Responsive Element (RRE) start
MAC.US.x.239                ..........AA-C-A-CTAAT-TCA-CATGA-T-C---GG--GCA--ACTG-ATC--TTG----GGGAG-T---..............---T-----G-G--CACT---A-T--CT-G--C-----AG-T-TG-----GTACACTACTGGTGGC......ACCTCA- 8167
Env                           .  .  .  N  Q  T  N  I  T  M  S  A  E  V  A  E  L  Y  R  L  E  L  G  D  Y  .  .  .  .  .  K  L  V  E  I  T  P  I  G  L  A  P  T  D  V  K  R  Y  T  T  G  G  .  .  T  S
H2A.GW.x.ALI                ..........AA-C---CAAAT-TTA-CTTTA-T-C---GG--GCA--ACTA-ACC--TTA-----GGGGG-T---..............---T-----G-G--AAC----A-T--CT-C-----T--ATCA-A--G---TACTCCTCTACT.........CCAAGG- 8206
H2A.DE.x.BEN                .......CG-A-TC-T-C-AA--TTA-CTTTA-T-C---AG--GCA--ACTG-ACC--TTA---C-GGGAG-C--C..............---T--A--G----AAC----A-T--CT-C-----T--AG-TCAG-GA--GTACTCCTCAACT.........CCAGTG- 8222
H2A.SN.x.ST                 ..........AA-C---CAAAT-TTA-CTTTA-T-C---GG--GCA--ACTA-ACC--TTA-----GGGGG-T---..............---T-GA--G--G-AAC----A-T--CT-T-----T--ACCA-T---A---TACTCCTCTGCT.........CCAGTG- 7652
H2B.GH.86.D205              .......A-CAGTACC-C-AAT-TTT---TGA-T-C---AG--TCA--ACTG-AT---TTG---C-GGGGG-C---..............---T-----G----AAC----A-T--CT-C-----T--AG-T-T--GA---TACTCCTCAGTA.........AAACCG- 8198
H2B.CI.x.EHO                ..........AA-CTT-CTAAT-TTA---TGA-T-C---AG--TCA--ACTG-AT-A-TTG---C-GGGGG-C--C..............---T-----G-G--AAC----A-T--CT-T-----T--A-GTAT---A---TATTCCTCAGT-..........ACACCG- 8184
H2G.CI.92.ABT96             ..........AA-C-T-CAAAT-TCAT-TTCA-T-C---GG--GCA--ACTG-ATC--TTA-----GGGGG-T---..............---T-----G----AACT---A-T--CT-T--------A--T-T-------TATTCCTCTGT-..........ACACCG- 7583
H2U.FR.96.12034             ..........AATC-A-CAAAT-TCT-CATGA-C-C---AG--GCA--ACTG-AT---TTA--G--GGGGG-T---..............---T-G---G----AAG----A-T--CT-T--T-----ACCA-TG--A---TATTCCTCAGT-..........ACACCG- 7706
MAC.US.x.EMBL_3             ..........AA-C-A-CTAGT-TCA-CATGA-T-C---GG--GCA--ACTG-ATC--TTA--G--GGGAG-T---..............---T-----G-G--CACT--GA-T--CT-G--C-----AG-T-TG-----GTACACTACTGGTGGC......ACCTCA- 7298
SMM.US.x.H9                 ..........AATA---C-AAT-TCA-CATGA-T-C---GG--GCA--ACTG-ATC--TTG--G--GGGAG-T--C..............---T-----G-G---ACT---A-T--CT-C--C-----A-GT-T---A--GTACACCACAACTGGT......GCCTCA- 7656
SMM.US.x.PGM53              ..........AA-C---CTAA--TCA-CATGA-T-C---GG--GCA--ACTG-AC---TTG--G--GGGAG-T---..............---T-----G-G---ACT---A-T--CT-G--C-----A-GT-T--G---GTACACAACTGGTGCC.........TCA- 8113
SMM.US.x.SIVsmH635F_L3      ..........AAT----C-AAT-TCA-CATGA-T-C---GG--GCA--ACTG-ATC--TTG--G--GGGAG-T--C..............---T-----G-G---ACT---A-T--CT-G--C-----AG-T-T--GA--GTACACCACAACTGGT......GCCTCA- 8189
SMM.SL.92.SL92B             ..........AAT----C-AAT-TTA-CATGA-T-C---AG--GCA--ATTG-ATC--TTG-----GGGGG-T---..............---T-G---G-G--CACT---A-T--CA-G--T-----AC-T-TG--A--GTACACCACAAGT.........ACCTCA- 7615
STM.US.x.STM                ..........AAT----CTAA--TCA---C-A-T-C---GG--GCA--ACTG-ATC--TTG-----GGGAG-T--C..............---T-----G----AACT---A-T--CT-G--C--A--G--T-TG------TACACAACTAGT.........ACCTCA- 7818
MNE.US.x.MNE027             ..........AA-C-A-CTAAT-TCA-CATGA-T-C---GG--GCA--ACTG-ATC--TTG-----GGGAG-T---..............---T-----G-G--CACT---A-T--CT-G--C-----A--T-TG-----GTACACTACTGGTGGC......ACCCCA- 7644
DRL.x.x.FAO                 AG-TTCCA-C-ATC----TAAATTTATACCGCA--AT--AG--CAAA-TC-G-TC-C-GCA-T-GG-GCAC-T--C..............---T-G--C--GG-G--G---A-T--CT-T--T--A--TG-T-T-C-T---TACAACC---C-GGA.........CAC- 7590
MND_2.x.x.5440              ....----GA-ATATC--AAAATTTATACC-CA--ATTGG--TCAAA-TC-A-AC-C-GCA-TGGGGGC-C-T--C..............---T-G--G--GG-G--T--GA----TT-T--C-----AG-CAT-C-C---CATCATT---CTAAT.........ACA- 7532
MND_2.CM.98.CM16            ...........-A-ATC---AAATTTGTGCC-CA--AT--GG-ACAAA--CGA-TT-C-GCG-T-GG-GCAC-C---..............C--C---------G-G--C---A-T--CT-T--T--T--AG-A-TGGCA---TATCATC---CAGAA.........GCT- 7899
MND_2.GA.x.M14              ....--AAGA-A-CTA--AAAGTTT-TGCC-CA--AT--GG-ACAAA-TC-A-TC-C-GCA-TGGG-GCAC-T---..............---T-G------G-G--T--GA-----T-C--C--T--TG-T-T-C-C---TATCATC-A-CAGAT.........GCA- 7828
OLC.CI.97.97CI12            ........-GAA-GC-AG-G-CATTG-T-CC-CC-TC---C-ACGT-A-TC-A-TTGTCTCA-TGGC-AG-G-T---..............---T-G---G-GT-G---TTGA-T--CT-T--T--T--AG--A--C-A--GTAT.................CAGCCT- 7376
GRV.ET.x.GRI_677            ..........GG-CCA--AAATGTGA-CCT-A-TCCTC-GG-AC-CAG--TA---GCGTAC---C-GGGAG-C---..............---T-----G-G--AAC----A-T--CT-T--T--T--AG-T-T--GA---TATACT-GC-CCACA.............- 7756
VER.KE.x.TYO1_patent        ..........AA-AGA-C-AA-GTGA-ACT-A-TCCCC-G--AGAATCT-TC---GCGGC------GGGC-G---C..............---T----GG----CAC----A-T--CT-T--------AG-A-T--G-C-TTATAC--GAGGA.........CATGAG- 7873
VER.KE.x.9063               ..........-ATAGA-C-AATGTGA-ATT-A-TCCCC-G--AGAA-GA-TC---GC-GCA-----GGGACG----..............---T-G---G----CAC----A-C---T-T--------G--T-T--GAC-TTATAC--GAGGA.........CAAGAC- 7897
VER.KE.x.AGM155             ..........AA-AG--CAAATGT-A-ATTGA-TCCCC-G--AGAAAC--TC---GCGGCA-----GGGC-G---C..............---T-G---G-G---AC----A-T---T-T--------AG-A-TC-G-C--TACAC--GAGGC.........CAAGAG- 7877
VER.DE.x.AGM3               ..........AA-AG--CTAATGT-A-ATT-A-TCCCC-G--AGAAAC--TC---GC-GCA-----GGGC-GG--C..............---T-----G-----AC----A-T--CT-C--------AG-A-T--GA--GTATAC--GAGGT.........CATGAC- 7395
TAN.UG.x.TAN1               ..........GAAAG--CTAATGTGA-ATTGAC--CTA--T--GA-A---TA---GCTTA------GGGA-G----..............--GC-CA--G-----------A-C--TT-C--G--A--AG--AT--GA---TAT--T-G--CAACT............C 7854
SAB.SN.x.SAB1C              ..........AA-AG--C-AATGTGA-GTT-TC-CCTC-G--TGAAAG--T----GCT-A------GGGAG-T--C..............---T----GG---CA-G--A-T---T-T--T-T--A-----A-T--GA---TACACT-GC...........CCCGAA- 8050
MND_1.GA.x.MNDGB1           ....-CAA-AATGC-A--TAAATTCTTGCCTCCCA-T--A-CCTCCA-TC-A-TTGTTGC-T-TGG-GC-C-T---..............---T---C----AATG--A-T--CA-------T--AG-T-TG--A---CACACTT-A-CTGAA......CATCAT- 7356
DEB.CM.99.CM40              .............AGT--AAC-CTTA-ACC-TC--C---C--C----CAGTA------GCA--CC---T-------..............---A-CA--G--G-AA----TA-----T-C--T----T-C--T---T--GCA-TG-GA--GTAC-A--GA-C-GAG......TCGGTA- 7486
DEB.CM.99.CM5               .............A-C--GAA--T-A-ACCCTCCTC------T-A--CAGT------GGCA--TC----C--G---..............--GA-CA--G--G--A----CA-C--GA-G--C--T--A-GTAT--GA--GTAC-A--GA-C-GAG......TCTGCA- 7407
TAL.CM.01.8023              ..........GA-ATC--AAGTGTGGAACCCTCTTC-C-A--C--A--TGC----C--GCA-CGG--GCACGG--C..............--G--GA--G-G--CAC----A-T--CT-T-----A--G--TATC-GA--GTAT-AT-G--C-..........GCAACA- 7126
TAL.CM.00.266               ..........AGT--C--AACAGT-GAGCCCTCGTC-C-G--AC-T---TCG---C--GCA-C-G-T-CACGG--C..............--G--GA--G-G--C-TG--T--T--GTAT--G--G--A--TATC-GA--GTAC-AT-G--CC.........GCTAAC- 7576
SUN.GA.98.L14               ATCA--CATC--A-TTG-AGC-TTC-AGCC-CCTACCA-C-CCTTAA-TC-A-TTGTG--CAC-GG-GCAC-T--C..............---C------G----AGG---A-T--GT-T--T--A--AG-T-AGC-CC-GTAC-CACCC...............AGG- 8069
WRC.CI.97.97CI14            .......AGAAATC-AG-GA-GTTC-AA-CTCCTTCCA-C--ACA-A--TC--TTGTGGCA-TGGG-C-A-------..............---T-G------T-AA--ATGA---GA-G-----AGGGG-T-G------GGTT--ACCC-CTGAA......CATCGC- 7880
WRC.CI.98.98CI04            .......A-TAAGC-AG-GGAATTC-AA-CTCCTTCCA-C--ACA-A-TTCA-TTGTGGCA-TGGG-C-A--G---..............---C-G--G-G-C-AA--ATGA---GC-G--T--TGGTG-T-GG-GA---GTGACACCAA-AGAA......CATCAT- 7925
WRC.GM.05.Pbt_05GM_X02      .......A-CGGGC--G-GGAGTTC-AA-CTCCTTCCC-C---CAAA--TCC-TTGTGGCA-T-GG-C-A-------..............---C-G------C-AA--TTGA---------T-----TG-T-G------GGTGACACCT-CTGAC......CATCAT- 6821
LST.CD.88.524               AGGAGGAA-TCTGTCT--TAAGTTTTTGCCTCCCTC-TGGGA--CAA-TC-A-TTGTGGCAAC-GG-GC-C-T--C..............---T-GA-C-G-C-GAG----A-T--CT-T--T--A--ATCA-AGC-TC-GTAC-CACCC............AGGGGA- 6929
LST.CD.88.447               GGGAGGAA-T-TGTCTG--AGATTCTTGCCTCCCTCCTGGGAA-CAA-TC-G-TTGTGGCAAC-GG-GC-C-C---..............---T--A---GGC--AG----A-T--CT-T--T--A--AG-T-TGC-TC-GTAC-CACCC............AGTAGT- 6926
LST.KE.x.lho7               AGGAGGCA-CCTAAGT--AAAGTTCTTGCCTCC-TC-TGGGAA-CTA-TC-A-TTGTGGCA---GGG-C--------..............---T-GA-----T-GA-T---A-T--GT-T--T--T--AG-T-AGC-C--GTAC-CACCC............AGGGGA- 8001
SYK.KE.x.SYK173             ..........--TAC-T-GAA-TTTA--CC-AC--C-A--G--GA---TGT-------GCA--TC-G-TC--T---..............---T-GA-TC-----A-----A-T--CT-T--G-----TG-TCA-CGTC-GTAT-A-T---CT.........AATACC- 7713
SYK.KE.x.KE51               ..........--AC--T-GAATTTTG-ACCTACT-CTA--G--GAA--TGTG------GC---T----TC--T---..............---T-GA-TCGG---A----TA-T--CT-T--G--T--A-GTCA--G-C-CTAT-A-C---CT............ACG- 7406
COL.CM.x.CGU1               TG-TGG-A-TGGGTGTGCT-ATCC-G-A-C-CC-AATTT--A-CAT-CATTAA-T-CTTTG-----GGGA-G---C..............---T-G-------GAGGA--ACTACTTAT-TC--A--AG-TAT---A---TC----AATGTAAATTGGCATCATGGG- 7156
```

```
                       Env gp120 end Env gp41 start
H1B.FR.83.HXB2         GAGAAAAAAGAGCAGTG......GGAATAGGAGCTTTGTTCCTTGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCCTCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAGCAGAACAATTTGCTG...AGGGCTATTGAGGCG 7907
Env                    R  E  K  R  A  V  .  .  G  I  G  A  L  F  L  G  F  L  G  A  A  G  S  T  M  G  A  A  S  M  T  L  T  V  Q  A  R  Q  L  L  S  G  I  V  Q  Q  Q  N  N  L  L  .  R  A  I  E  A
H1A1.UG.85.U455        ----------------T.........C---------A-C-------------A--------T--------A--------G--G----A-----------------------------------------A-------G----C-----...---------A-----T 7337
H1B.US.90.WEAU160      -----------------GGA...ATGC--------A---------------------------------------------G------------------------------T--------------------A------------C-----...----------------- 7914
H1C.ET.86.ETH2220      ---------------CA.........C---------------------------------------------------A-----A---------------------------------G-----------A-----A-G-----------...-AA-----A------ 7283
H1D.CD.84.84ZR085      --------------A-A.........C--------A--------------------------------G----------G-----------------------------G-------------------A-----A-------------...---------A------ 7428
H1F1.BE.93.VI850       ---------------C-.........C-------------------TC-T--G-ACAGCA-GGAGCA---------G--G--G--A------------------------------------A-------A-----A-------------...-----------A--- 7184
H1G.SE.93.SE6165       ---------------T.........C-G-----G-C-------------A-------------------------G--G----A-----------------T---------------------C--------A-----AGG---C------...---------A------ 7325
H1H.CF.90.056          -----------------.........-----G-------CT----------------------------------G--G----A--------------------G------------------------------A-G------------...--A-----AC----- 7226
H1J.SE.93.SE7887       -----------------.........-----------G--------------------A---------T-------------G----A-----------------T--------G--------C--------A-----A-G-----------...-A------A------ 7214
H1K.CM.96.MP535        ---------------A.........T--------G-----T----A-------------------------------G--G----A-------------------------------------------A-----A-G-----------...--A-----A--A--- 7067
H101_AE.TH.90.CM240    ----------G------.........---------------A--A--T-----------A--------------------G--G----A-------------------------------------------A-----A-G-----------...---------A------ 7460
H102_AG.NG.x.IBNG      ------------G--T.........C-G-----G-C-------------A--------------------------GCGG----A--------------------------------C--C--------A-------G------------...-AA-----A-----T 7413
H1N.CM.95.YBF30        ---------------CCTTT......C----T--GC----T--------TC-T-----------G-----------------G-----A------------------C-GAC------A-----G--------A--------T-T-C--T--...--A--A--A------ 7455
H1O.BE.87.ANT70        ----------------A.........T-G---ATGC-A---T-G---G-TC-AA-T--------T-----------------GG---CA------G--------A--CACACT--GC--AAG-----------A------G----CC----A...--A--A--AC----C 7961
H1O.CM.91.MVP5180      -G--------------A.........T-G---ATGC-A---T-G---G-GC-AA-T--------T-----------------GG---CAG-----------G-A--CACAGTG--C--AAG-----------A------G----CC-----...--A--G--AC----C 7978
CPZ.CD.90.ANT          --C---------GCA-T..............GCTG--C---T-G--TC-TC-CA-T-----T--C--T--A-----------G-----AG--------C----A----GA-T--GY-CCA------T--A--A------GC----C------...CAA--C--A---A-A 7339
CPZ.CM.01.SIVcpzCAM13  --C-G--G-------CCTTT.....CC-G-----AC--------------TC----T---------------A-----------GG--G-C--------------------C----G---C-A--G--A--------A--A--------CC--T--...--A--A--A-----C 7491
CPZ.CM.05.SIVcpzEK505  --C-------------CCTTT.......C----T--GC----T--------TC-T-----------G-----------------G-----A------------------C-GA------A-----G------------------T---C-----...--A--A--A------ 7390
CPZ.CM.05.SIVcpzMB66   -G-------------CCGTA.....--T-G--------------------------------------------------G---C--------------------------C-GC------------------------G----C-----...--A--C--A------ 7531
CPZ.CM.98.CAM3         ---------------CCTTT.......C-G--T--GC----T-----------A------------------------------A---G--GT-----------------C-G-----------A--G--------A--------------CC-----...C-A--A--A------ 7320
CPZ.GA.88.GAB2         --TCC----------CCTTT.......C-G--G--GC----T-----A---C-T---------------------------A--TG---T-----------A--TC--A-CC-GC-T--G--A--------------A--T--C-------...C-A--C--A--A--T 7291
CPZ.CM.98.CAM5         ----G----------CCTTT.......C-------AC-------C-----TC-A---------------------------A---G--GT-----------------C-G--C--G----A--G--------A--------------CC-----...C-A--A--A------ 7626
CPZ.CM.05.SIVcpzMT145  -------G-----CC-T.........C-------AC----T--------C-----------------------------A---G-CGT-----------A---C-G-----G--AA-G--C------------------T---C----A...--A--A--A--A--T 7416
CPZ.TZ.00.TAN1         -GTC--G-----AT......GTG--C------CTG-----T-----A--TC-TA-T------------T--A-----------G-----AG--------C----------GG----C-C--------T--A--------AC-A--CC----T...CA---C--A--A--- 7533
CPZ.CM.05.SIVcpzLB7    ---------------CA.........--GC-G---TG--T------------------------G-----------------A---T-A-------------------------------A----C--------A--------GT-----------...--A-----A------ 7418
CPZ.US.85.CPZUS        ---------------CCTTC......C----T--AC-------G-----TC-T-----------------------------A---G-AGT-----------------C-------------A--G--------A--------T---C----A...--A--A--A-----C 7912
CPZ.GA.88.GAB1         --C-G----------CCTTC.....--TC-G-----AC-------------C----T-------------------------GG--G-A-------------------C----G--GC---G--------------A-----T---C--T--...-AA--A--A-----T 7944
MON.CM.99.L1           -G-----G-------......ATTACTT-G--TATGGCC--T--G--C---C-CA-CA-G------G----A-----------GG-G-CAG-C------------T-TC-GTC----C--G-------------------G-G--CC-----...-----CG-AACA--C 7870
MON.NG.x.NG1           TGAG-G-----CG--CT...ATCAC-T-G---ATGGCC--T--G--C--TC-CA-CA-G------G----G-----------GG-G-CAG-C------------T--C-GTCT--G---G-------------------G-G--CC-----...-----CG-AACT--C 6495
GSN.CM.99.CN166        TG--G--------C.....CTTTC-T-G---ATA-CC---T-GAC---TC-CA-T--G-----T-C---G-----------GG---CAG---------A----T--C-GTCT--G---G----G---A-------------G-G--CC-----...------G-A-----C 7844
GSN.CM.99.CN71         TGA----G--G--C.....ATTTC-T-G--CATA-CC---T-GAC---TC-CA-T--G-----TGC---G-----------GG---CAG------------AT--C-GTC---G---G----G-----------------G-G--CC-----...------G-A-----C 7856
MUS_1.CM.01.1085       TGC-TGC------C.....ATCTC-T-G--CATAACC--TT-A--A--TC-CA-C----------G----A-----------GG-G-CAG-C------------T-TC-GTCT--G---G-C--G---------------GCA---C-----...--A--CG-A--A--C 7841
MUS_1.CM.01.CM1239     T-A--GC----------.....ATATCTC---CATAACC--T--G--A--TC-CA-C--G------G----G-----------GG-G-CAG-C------------T-TC-GTC----C--G-C--G---------------GCA---C-----...-----CG-A--A--C 7810
MUS_2.CM.01.CM1246     TCA-GGCT--G---.....ATCA-CT-G---ATAACC---T-A--A---C-CA-T--T--G---G----A-----------GG-G-CAG-C------------T-TC-GTCT--G---G-------------A---GCT---C--T--...--A--CG-A--A--C 7939
MUS_2.CM.01.CM2500     TCC--GCT--G---.....ATAACCC-----ATAACC--TT-G--A--TC-CA-T--T------G----A-----------GG-G-CAG-------------T--C-GTCT--G---G-------------A---GCT---C--T--...-----CG-A--AA-C 7862
RCM.NG.x.NG411         A-C----G----T-CCT...TTG-TGC-G--T.......-T--A--T---C-CTCG--------T-CTG-A-----------GG-G-CAG-C-----A--C---T--C-G--T--GC-TG----G---T-------------G--G-------...GACATAG----ACA- 7638
RCM.GA.x.GAB1          A-C-G-----G-T-CCT...TTG-TGC-G--T.......-T--A--T---C-CTC---------T-CTG-A-----------GG-G-CAG-C-----T--C---T--C-G--T--GC-TG-A--G---T-----------A--G--CC-----...GACATAG--A--CG- 7594
DEN.CD.x.CD1           TTTC-CGCC--AGGCG-GAAGTGTCTT---TGCT-GG-C--A-A--A--TC-CA-C----------CTGTC---------G-T-GG--CAGGCT-AG--------T--C-GG-T---C-TG-C--G-----------A---C-AG-GC--T--...--A---G----A-GT 7964
                       Env gp120 end Env gp41 start
MAC.US.x.239           --A-T-------GG--C...TTT-TGC----G......---T-G--T--TC-C-C-A-G-----TTCTG-A--------G--G--GT----------C-CT---T--C--ACT------G----G-------------A---C-AC-GC--T--...GAC-TGG-CA--AGA 8325
Env                    R  N  K  R  G  V  .  F  V  L  G  .  .  F  L  G  F  L  A  T  A  G  S  A  M  G  A  A  S  L  T  L  T  A  Q  S  R  T  L  L  A  G  I  V  Q  Q  Q  Q  Q  L  L  .  D  V  V  K  R
H2A.GW.x.ALI           -GA-T-------GT---...TTC-TGC----G......---T-A--T--TC-C-CGA-------TTCTG-A------A-G--AG-TT-A------T-T-CT---T-TC-GACT------G-C--G-----------A---C-AC-GC--T--...GAC-TGG-CA--AGA 8364
H2A.DE.x.BEN           -GA-C-------GT---...TTC-TGC----G......---T-G--T--TC-C-CGA-------TTCTG-A--------GCGG--CC---------T-A-CC---T--C-GACT---C--G-C--G-----------A---C-AC-GC--T--...GAC-TAG-CA--AGA 8380
H2A.SN.x.ST            -GA-T-------GT--A...TTC-TGC----G......---T-A--T--TC-CACGA--------GCTG-A--------G--G--CT---------T---CT---T-TC-GACT------G-C--G-----------A---C-AC-GC--T--...GAC-TGG-CA--AGA 7810
H2B.GH.86.D205         -GA-T-------GT--A...ATG-TGC-----.........---T-G--A---C-T-C-ATG-----TTCTG-A----------A-G--TT---------T-A-CT---T--C-GACT---C--G-C--G---------------C-AC-GCCTG--...GAC-TGG-CA-AAGA 8356
H2B.CI.x.EHO           -GA-T-------...-GTGTACTT-TGC----G.......--TT-G--A---C-T-CGA-G-----TTCTG-A--------G--G--CT-----T--T---CT---T--C-GACT---C--G----G---------------C-AC-GC-CG--...GAC-TGG-CA-AAGA 8342
H2G.CI.92.ABT96        A-A-T-------GT--A...TTT-TGC----G.......--TT-G--A--TC-T-C-ATG-----TTCTG-A------A----G--GT---------T---CT---T--C-GACT--G---G----G-----------A---C-AC-GC--T--...GAT-TGG-CA--AGA 7741
H2U.FR.96.12034        -GA-T-----G-G---A...TTT-TGC-T--G......---T-G--T--TC-C-CCA-G-----TTCTG-A-----T--G--G--GT---------T-A-CT---T--C-GACT---C--G----G-----------A---C-AC-GC--T--...GAC-TGG-CA-AAGA 7864
MAC.US.x.EMBL_3        --A-T-------GG--C...TTT-TGC----G......---T-G--T--TC-C-C-A-G-----TTCTG-A--------G--G--GT-C-G------C-CT---T--C-GACT------G-----G-----------A---C-AC-GC--T--...G-C-TGG-CA--AGA 7456
SMM.US.x.H9            --A-T-------GG--C...TTT-TGC----G......---T-G--T--TC-C-CGA-------TTCTG-A--------G--G--GT---------T---CT---T--C-GACT--G---G----G-----------A---C-AC-GC--T--...GAT-TGG-CA--AGA 7814
SMM.US.x.PGM53         --A-T--G----GG--C...TTT-TGC----G......---T-G--T--TC-C-CGA-------TTCTG-A--------G--G--CG----T-G-T-C-CT---T--C-GACT--G---G----G-----------A---C-AC-GC--T--...GAT-TGG-CA-AAGA 8271
SMM.US.x.SIVsmH635F_L3 --A-T--G----GG--C...TTT-TGC----G......---T-G--T--TC-C-CGA-------TTCTG-A--------G--G--GC---------T---CT---T--C-GACT--G---G----G-----------A---C-AC-GC--T--...GAT-TGG-CA--AGA 8347
SMM.SL.92.SL92B        AGA-T-------G---C...TTT-TGC----G......---T-G--T--TC-C-CGA-G-----TTCTG-A-----------G--GC-A--------T-CT---T--C-GACT------G----G--------A--A---C-AC-GC--T--...GAC-TGG-CA--AGA 7773
STM.US.x.STM           --ACT--G----GG--C...TTC-TGC----G......---T-G--T--TC-C-CGA-------TTCTG-A-----------G--GC-------------CT---T--C-GACT------A----G-----------A---C-AC-GC--T--...GAC-TGG-CA--AGA 7976
MNE.US.x.MNE027        --A-T-------GG--C...TTT-TGC----G......---T-A--T--TC-C-C-A-G-----TTCTG-A--------G--G--GT----------C-CT---T--C-GACT------G----G-----------A---C-AC-GC--T--...GAC-TGG-CA--AGA 7802
DRL.x.x.FAO            -GC-G--G----GT-CA...GTC-TGC-T--G......A-------CC-----A-CTTG-----TTC-G-A---T--T---TGG-G-C-G-CA----T--C---T--CAGGCT--GC--A----G--G---G-------A--GC-AC-------...C-CCTGG-G--ACA- 7748
MND_2.x.x.5440         -GC---------G--C-...GTCTT-C-T--.......A-G--C--CC--C-A--TTTG-----TTC-G-G------T---TGG-GG--G-C-----T--C---T--CAGACT---C--AA---G--T---G----------GGT-C------...--CCTG--A--TCA- 7690
MND_2.CM.98.CM16       -GC-G--G--G-G--C-...GTCTTGC-T--G......A-GT-C--CC--C-A--TTTG-----TTC---G------T---TGG-GG--G-C-----T--C---T-TCAGGCT--G---AA---G--T---G----------GGT-C------...--CCTG--A--TCA- 8057
MND_2.GA.x.M14         AGC-G--G----G--CA...GTCTTGC-T---.......A-G--C--CC-------CTTG-----TTC-G-G------T---TGG-GG--G-C-----T--C---T--CAGGCT--G---AA---G--T---G----------G-T-C------...--CCTG--A--TCA- 7986
OLC.CI.97.97CI12       --A-T-------G-...GCATTG-TGC-G--G.......--AT-G---G-TC-C--CCT-------TC-G-A------A---TA----C--TA-----T--C---T-TTCG--T--GC-TG-G--G-------G---------GC-CA------...GA-ATG------AA- 7534
GRV.ET.x.GRI_677       ----------G-TGCCATTC....-TGC----G.......--T--A--C-----------T--T----CTG-A-------------GG---CA---------A--C---T-TC-G--T--GC-TG----G---T-------------G--C-------...GC----G-G--ACA- 7914
VER.KE.x.TYO1_patent   -GC----G----TC...CCGTTT-TGC----G......---T-A--C--------G--T--T----CTG-A-----A-----GG-G-CAG-C---------C---T-TCAG--T--GC-TG----G---C-------------G---C------...GC----G-G-----T 8031
VER.KE.x.9063          --TC----C---TC...CCGTTT-TGC----G......-----A--C-----A-----T--T----CTG-A-----A-----GG-G-CAG-C---------C---T-TCAG--T---C-TG----G---A-------------A---C------...GC----G-G-----T 8055
VER.KE.x.AGM155        -GC-----C---TC...CCGTTC-TGC----G......-----A--C------------T--T--G-CTG-A-----A-----GG-G-CAG-C---------C---T-TCAG--T---C-TG----G---T-------------G---C------...GC----G-G-GA--T 8035
VER.DE.x.AGM3          --AC---GC---TC...CCGTTC-TGC----G......-----A--C-----A-----T--T--G-CTG-A-----A-----GG-G-CAG-C---------C---T-TCAG--T---C-TG----G---C-------------G---C------...GC----G-G-----T 7553
TAN.UG.x.TAN1          -------G--G-TGCCCTTC....-TGT-G--G......-----A--C--T-----G-----T---GCTG-A-----T--A-AG-G-CAG--T--------C---T--CAG-------C-TG-A--G---T-------------G---C------...GC---AG-----CA- 8012
SAB.SN.x.SAB1C         --C-------G-TG...CCATTC-TGC----G......-----A--T---------------T---GCTG-A-----T-----GG-G-CAG-C---------C---T-TCAG--G---C-CG----G---C-------------A---C----T...GC----G-A---CA- 8208
MND_1.GA.x.MNDGB1      A---G--G----G--CA...GTAAT-C-T--T......A-------TC-GC-CTCGCTG------TC-G-G------T---TG--GG--G-A-----T--C--AT-TCAGTCT--GG--A----GG--A----G------G-A--A--A--AC-G---T--...-A-CTC--A---CAA 7514
DEB.CM.99.CM40         --C-C--G--G----CAGGAATT-CTT-T--T......C-GG-A-CA--TC-CA-TA-------TGCTG--------------G----CAG---------------T-TC-CTCT--GC--AG---GC--AG---G-----------A---C-AG-GC--T--...-A----G-G-----T 7647
DEB.CM.99.CM5          --C-------G--G-C-GGAATT-CTT-T--T......C-AG-G-CT--TC--A-TA-------TGCTG--------------G----CAG---------------T--C-CTCT--GC--AG---GC--AG---G-----------A---C-AG-GC--T--...-A----G-G-----T 7568
TAL.CM.01.8023         A---G--------C-C-CCTCTG-CTT-G---.......---A-A--A--TC-CA-CA-------C-CTG-C-----------TGG-G-CAG-C---------G---T-TC-TTCT--GC--AGC--GC--AGC--G---------------G-GC-CC------...-----A------CAT 7287
TAL.CM.00.266          AGC-----C----G-CCCCGCTA-CCT----C......---A-A--A--TC-CA-CA-------C-CTG-C-----------TGG-G-CAG-C---------G---T-TC-CTCT--GC--AG---GC--AG---G---------------G-AC-CC------...-----A------CAT 7737
SUN.GA.98.L14          AG-----G--G----CACCTGTT-CCC-----......GCTT-G-CAC-TC-CA-T--T--G--T-CTG-A------TT--TA--G-C--TA--A--T--C-------CA-GC-G-C---CAA--G---T-------------GC-G-------...GT-CT-G-----AA- 8230
WRC.CI.97.97CI14       ----G-------GT...GCCATTATTC-----......A-T--A--A---C-C--GCTG------TCTG-A------A---TA----CAGTA------A-C---T-TCAG--T--GC-TG----G--------------A--G--CC------...GA-CT-G-A---AA- 8038
WRC.CI.98.98CI04       ------------G-...GCTATT-TCT-----......A-T--A--A---C-C--GCTG------TCTG-A------A---TA----CAGTA------A-C---T-TCAG--T--GC-TG----G---T----------A--G--CC------...GA-CT-G-----AA- 8083
WRC.GM.05.Pbt_05GM_X02 --------C---G-...GCAATT-TTT-----......C-T--A--C---C-C--CCTG------TCTG-A------A-G-TA----CAGTA------A-C---T-TCAG--T--GC-TG----G-----------------G--C-------...GA-CT-G-A---AA- 6979
LST.CD.88.524          -GC---C--------CTCCACTA-CTC----T......GC-T-G--AC-TC-CA-T--T-----T-CTG-A------TT--TA--G-C--TA--A--T--C------CA-GT-G--A-ACAA--G---T-------------GC-G-------...GT-CT-G-----AAA 7090
LST.CD.88.447          AGC-G--G-------CACCGCTG-CTC----T......GC-T-G--AC-TC-CA-T--T-----T-CTG-A------TT--TA--G-C--TA--A--T--C--A---CA-GT-G--A-ACAA--G---T-------------GC-G-------...GT-CT-G-----AAA 7087
LST.KE.x.lho7          -GC---CC-------C-CCTCTG-CTT-----......GC-T-G--AC-TC-CA-T--T-----C-CTG-A------TT--TA--G-C--TA--A--T--C------CA-GC-G-C---CAA--G---T-------------GC-G-------...GT-CT-G-----AAA 8162
SYK.KE.x.SYK173        -G--------G----CACCATTG-C-C-T--G......-----G--A--AC-CTCG--T--G--C-CTG-A-----G-GC--TG-G-CAG----A--AC-C---T-TCAGACT--G---G----G---G--------------C-A--G-------...GA---AG-------A 7874
SYK.KE.x.KE51          A-C-G----------CACCATTG-C-C-T--G......-----G--A--AC-CTCG-----G--T-CTG-A-----GAGC--TG-G-CAG----T--AC-C---T--CAGACT--G---G----G---G--------------C-A--G-------...GA----G-A-----A 7567
COL.CM.x.CGU1          --C----G---.........GGCATCT-T-C-TTC-CCA--T-A-CCC-----A-T-GG--C--TGCTG-A------T----G--TG--G--------A-C-----TCA-TCCC-TAATGGAA-GGCTAGTGCCAGCAGC---CGAA--T--CTG-AACTGG-C---A-- 7317
```

```
                                                                                                                                                                 RRE end
H1B.FR.83.HXB2         CAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAATCCTGGCTGTGGAAAGATACCTAAAGGATCAACAGCTCCTGGGGATTTGGGGTTGCTCTGGAAAACTCATTTGCACCACTGCTGTGCCTTGGAATGCTAGTTGGAGTAATAAATC 8077
Env                     Q  Q  H  L  L  Q  L  T  V  W  G  I  K  Q  L  Q  A  R  I  L  A  V  E  R  Y  L  K  D  Q  Q  L  L  G  I  W  G  C  S  G  K  L  I  C  T  T  A  V  P  W  N  A  S  W  S  N  K  S
H1A1.UG.85.U455        --------------A-------T-----------T--A--------------G------------------------C---------------A--A-------C----------G----C---------A-------C----CT------------------ 7507
H1B.US.90.WEAU160      --------------CG----------------------A---------------G-------------------------------------G------------------------------------------------C---------A--------------------------------G--- 8084
H1C.ET.86.ETH2220      ---------A-------G-----G-----------T------------A-----G--------A-A-----C------G---------------A---------C--------------------------------C-----------------------CT---------------G-- 7453
H1D.CD.84.84ZR085      ---------------------G-----------T--A------------------------A----------G---------------A--A---------------------------A---------A---A--C----CT---------------G--- 7598
H1F1.BE.93.VI850       ----------------G----------------T--A--------------G-------------------------------------A----C-----C---------------C---------AA------C-----CT----------------------- 7354
H1G.SE.93.SE6165       --G------------------------------T--A--------------C---------A----------------------------A----------C----------------C---------AA------C-----C-T--------------------- 7495
H1H.CF.90.056          -G-------A-------G-----G---------T--A--------------G----C-------------------G-----------------------C----------------C---------AA-----------CT----------------------- 7396
H1J.SE.93.SE7887       ----------------A-------T--------T-----------------G-----------------------------A-----A----------C----------------C---------AA------C------------------------------- 7384
H1K.CM.96.MP535        -----A--------A--G-----G---------T--A------G--------------A----------------------------A----------C----------------C---------AA------C-----CT----C-----------.. 7235
H101_AE.TH.90.CM240    --G------------------------------T--A--------------G---------------------------A--T----A--AC-------C------------A-----C------------------C-----CT-C-C-----------G--- 7630
H102_AG.NG.x.IBNG      -----A---------A-------G---------T--A--------------G--------C--------------GA--------------A--A-------C----------------C---------A----A--C-----CT----C-----------A- 7583
H1N.CM.95.YBF30        -----A---T-----------T--A--------T--A-------------A-G----T---A-A-----------T-G------G--AA----AA-TC-A-----C-----A------ACA--A---TAT--CA---------------AG-C------C--C--TA- 7625
H1O.BE.87.ANT70        --G-----AT--C--AGG--AT-TR-A------T----GA--A----GA--TC-CC-G--A--CT-A----CC-TA---C--A----G--A-----AA-CC-A-----C--TAAA-----G--AG-C---TA---AT-A--AAAA------AGA-CA----TAGGA..... 8126
H1O.CM.91.MVP5180      --G--A--CT--C--AGGT-AT-T--A------T--T-GA--A----GA--TC-CC-G-AA--CT-A----CCCTTA--C--A----G--A-G---AAACC-A-----C--TAAA---------A--C--TTA---AT-A--AAAA-----CA-ATCA---TCAGGA..... 8143
CPZ.CD.90.ANT          --G--A------C-A--G---T-G---------AG-A--A--A---------------G--T--A--C--G-AG-------GA---------A--AT--A-CC-C---------TG---AC--GG-G-CC--TCA----A-G------------AA-TCC---GTA--CTTCA- 7509
CPZ.CM.01.SIVcpzCAM13  -----------A---------A-----------AG-A--A--A------A--C---T-A--T-----A------------C-------G---A-T--A--CC-------C------------GTCT--C--TTAT--CA--------------CAAA-CA---TCAGGC..... 7656
CPZ.CM.05.SIVcpzEK505  -----A--CT--C---------T----T----------T--A--------A------G----T---A-A-----------G-G---C--G--AA----A--CC-A-----C-----A------TCTG-G--TTA---CAA--------------A-C-C------C--C..... 7555
CPZ.CM.05.SIVcpzMB66   ----------------------T----------T--A-----T------C-GG-T--T--------G-------------------------T-A--AC-A-----C-------------------------T--------------CA---CC----C----..... 7696
CPZ.CM.98.CAM3         ---------T-----------T----T---------T--A--------A------G-T--T-----A-----------C-GA-----G--AA----A--CC-A-----C-----A------GC------TTAT---A---------------AA--C--------GC---T.. 7488
CPZ.GA.88.GAB2         --G--A--C---C----G--AT-------------A--A-------------G--GG----T--------------T-----------G--AG----A-CAC-A-----C-------------ACTG-G---TA-T--A----A--C------A--TCC----AC-G---CAA 7461
CPZ.CM.98.CAM5         -----A---T-----------T----T-----------A--------A------G-T--T----A-----------------G---A----AA-TC-A-----C-----A-----GC---C--TTAT--CA-G------------------C--------GC---C.. 7794
CPZ.CM.05.SIVcpzMT145  -----A-----------G--TT-T--------A-----A----------G---G-T--T-----------G------G------G--A-----A--TC-A-----C--TA-A------ACT-----TC-----------A-GC-----CAAG-CA---G-A----TT.. 7584
CPZ.TZ.00.TAN1         -----A--CT-------G---T-T--A---------T--------------C-----G--T--A--A--G-A----A---GA--C-----------AA-CC-C------A--TG--AAC---T-GG-G--TCA--G-AG------A-----CCTC-CC---GC-G-AG-T-- 7703
CPZ.CM.05.SIVcpzLB7    --G--A--CT--C---------T--------------A--------A---------A-----G---T--A-C-------------------------A-----A--A--A-------------A-------------C----------T-------A------AGA-C----------C..... 7583
CPZ.US.85.CPZUS        ---------T-A--------AT----T-----------T--A-----T--A--C---G-A--T-----------------T--------G---A----A--TC-G-----C-----A-----ACA-----TTAT--CA---------------A--CC---------C--C.. 8080
CPZ.GA.88.GAB1         --------C--A--------AT--A--------AG-A--A--A--A--A--C---T-G--T-----A-----G-----GC-------G---A-T-----CC-G-----C-----------GGCTG-C--TTAT--CA-G-----------CAACTC----CC-GGG...AG 8111
MON.CM.99.L1           ------AGCT-AC-A--G--T--T--T-----AG------------A-----CC-CC-GAC------A--G-AG-T-A--------C---ACAT-G--AAATGCA-----A--TG-CAAT---GC-G-G--TCA---AA-A-----G-----CAA-TC-.............. 8026
MON.NG.x.NG1           --G---AGCT-A--A--G---T-T-----------------A-AT-A-----CC-CC-GAC------A---AG-T-A--------C---ACAT-GT-AAATTCA-----A--TG-CAATCGCGCAG-R---CA---ACAA----TA-----CAA--CG............. 6651
GSN.CM.99.CN166        ------AGCT-A--A--G-CAT-T--A---------A--A----A--AA-CC-CC-GTCAT-AC-A--G-A---TT-G-G--------ACCA-T--CCAAGCA-----C--TG-GAAC-GG-CA--C--TCA---CATA--A---------CA-CTCA............. 8000
GSN.CM.99.CN71         ------AGCT-A--A--G--AT-T--A---------A--A----A--A--CC-CC-GTCTT-AC-A--G-A------G-G--------ACAA-T--CCAAGC------C--TG-GAATC-G-CA--C--TCA---CATA--A-----------ACTCA............. 8012
MUS_1.CM.01.1085       -----A-----A--A--GT-GT-T--T---------T--A-----A-----TC-CC-GACA--GC-T--G-AG-T-A--------C---GCC--T---AATC-C-----A--TG-AAAC-GG-AG-----TCA---AGA-----A-------A-TCA............. 7997
MUS_1.CM.01.CM1239     --G--A-----------A--GT-GT-T--------T--A--A--A-----A-----CC-CC-GACA--GC-T--G-AG-T-A-------------GCCT-GT--AACC--------A--TG-GAAC-GG-AG--C--TCA---AGG-----A------A-ATCA............. 7966
MUS_2.CM.01.CM1246     -----A------AC-A--GT-GT-T--T---------A--A--A--A-----TC-CC-GACA--AC----G-A--T-G-G-----C---GCCA-A--AAATC-G-----G---G-CAAT-GG-AA--C---CAT--CT-G-----A--------A-TC-............. 8095
MUS_2.CM.01.CM2500     -----A---------A--GT-GT-T--T---------A--A-----A-----CC-CC-GACA--AC----G-A--T-A-----------GCCT-GT-AAATC-G-----A--TG--AAC-GG-AG--C--TCA---CAGG-----A------AACTC-............. 8018
RCM.NG.x.NG411         --G---G-A---C-AA-G-----C---------G-CT--AA-C---------GC-TG--ACT---A-T--G-A-------GC---C---TCT--AT-AAAT-CA-TT--C--TG-AT-G-G--AAG-G--TCA---AA-G---GAG----TATA----CAA........... 7797
RCM.GA.x.GAB1          --G---A-----C-AA-G-----C---------A-CT--AA-T---------GC-TG--ACT---A-T--G-A-------GCA--C---TCT--AT--AAT-CA-TT--G--TG-CT-G-G--AAG-C---CAT--A-TG-----G----CATTC................. 7747
DEN.CD.x.CD1           --CAGTGCA--AC----GT-GT----A---------T--AA-C---A-C--TC-CC-GACT--AA-A--G-A---T--T--A--C---GCA--G--AAATGAA-----A--TG-CT-G--G-AA-----TCA---CT-A--AGAA----GCAA-CAA.............. 8120
                                                                                                                                                                 RRE end
MAC.US.x.239           -----AG-AT------G---G--C--------A-CA---A-C------A-T--GG--ACT--CA-C--G-AG---T-------C--GGC--AG---AATGC------A--TG-GTTT-G--AAG-C---CA----A----A--A---CCAAA-GCAA-T........... 8484
Env                     Q  Q  E  L  L  R  L  T  V  W  G  T  K  N  L  Q  T  R  V  T  A  I  E  K  Y  L  K  D  Q  A  Q  L  N  A  W  G  C  A  F  R  Q  V  C  H  T  T  V  P  W  P  N  A  S  .  .  .  .
H2A.GW.x.ALI           ------G-AA------G---G--C--------A-CG--AA-T------------G--ACT---A-C--G-A---T-------C--GGC--GG--AAATTCA-----A--TG-ATTT-G--AAG-C---CA----A----A--A---GTAAA--AC-CC........... 8523
H2A.DE.x.BEN           -----AG-AA------G---G--C--------A-CG--AA-C------------G--ACT---A-C--G-AG---------C----GGCA-AG--AAATTCA-----A--TG-GTTT-G--AGG-C---CA----A----A--G---GTAAA-GAC-CT........... 8539
H2A.SN.x.ST            -----AG-AA------G---G--C--------A-CA--AA-T------------G--ACT---A-C--G-A---T-------C--GGC--AA--AAATTCA-----A--TG-GTTT-G--AAG-C---CA----A----A--A---GTAAA-GACACC........... 7969
H2B.GH.86.D205         ------G-A-------GG--G--C--------G-CG--AA-C------------G--ACT--CA-C--G-A----------------GGCA-AA--AAATTCA-----A--TG--TTC-G--AGG-C--TCA---GA---A--A---CCAAA-GAAACC........... 8515
H2B.CI.x.EHO           -----AG-A-------GG--G--C--------G-CG--AA-C------------G--ACT--CA-C--G-A---T--C--A--C--GGCA-AA--AAATTCA-----A--TG--TTC-G--AGG-C--TCA---GA---A--A---GTAAA-GAA-CC........... 8501
H2G.CI.92.ABT96        ------G-AA------G---G--C--------G-CG--AA-C------A-T---G--ACT--CA-C--G-A---T---------C--GGCA-GG--AAATTCA-----A--TG-GTTT-G--AGG-C---CA----A----A------G----CCT-G-TGC-------A- 7911
H2U.FR.96.12034        -----AG-AA------GG--G--C--------A-CT--AA-C------A-T---G--ACT--CA-C--G-A----------------GGCAAGT--AAATGCA-----A--TG-CTTT-G--AGG-C---CAT---A--------A----TAAACGACACA........... 8023
MAC.US.x.EMBL_3        -----AG-AT------G-T-G--C--------A-CA---A-C------A-T--GG--ACT--CA-C--G-AG---T--G----C--GGC--AG---AATGC------A--TG-GTTT-G--AAG-C---CA----A----A--A---CCAAA-GCAA-T........... 7615
SMM.US.x.H9            -----AG-AT------G---G--C--------A-CT---A-C------A-T---G--ACT---A-C--G-AG---------------GGC--CRN-AAATTCA-----A--TG--TTT-GG-AGG-C---CA----A-K--A--A---CCAAA-GACACA........... 7973
SMM.US.x.PGM53         -----AG-AT------G-T-G--C--------A-CT---A-C------A-T---G--ACT---A-C--G-AG---G--------GGC--AG--AAATTCA-----A--TG--TTT-GG-AGG-C---CA----A----A--A---CCAAA-GCC-CA........... 8430
SMM.US.x.SIVsmH635F_L3 -----TG-AT------G---G--C--------A-CT---A-C------A-T---G--ACT---A-C--G-AG----G--------GGC--AG--AAATTCA-----A--TG--TTT-GG-AAG-C--TCA----A----A--A---CCAAA-GAC-CA........... 8506
SMM.SL.92.SL92B        -----AG-AT------G---G--C--------A-CT--AA-T--T---A-T---G--ACC--CA-C--G-A----------------GGCA-AG--AAATTCA-----A--TG-ATTT-GG-AAG-C---CA----A----A-TA---CCAAA-GACA-T........... 7932
STM.US.x.STM           -----AG-AT------G---G--C--------A-C----A-C------A-T---G--ACT---A-C--G-A----------------GGCA-AG--AAATTC------A--TG-GTTT-G--AGG-C---CA---CA----A--A---CCAAA-GA--CT........... 8135
MNE.US.x.MNE027        -----AG-AT------G---G--C--------A-CA---A-C------A-T---G--ACT--CA-C--G-AG---T-------C--GGC--AG--AAATGC------A--TG-ATTT-G--AAG-C---CAT---A----A--A---CCAAA-GCAA-T........... 7961
DRL.x.x.FAO            ------G-GT-AC-CA-GT-A--TA--------G-A---A-C--------CC-CC--ACA---C-T--GGA----A--GGA--C---GCAA-GT--A-CC-C-----A-----CTTTGC--AAG-A---CAT--AAA----GTA---CCCAA-GAG-CA........... 7907
MND_2.x.x.5440         --CTCCG-GT-A--AA----A--TA-------TG-A---A-T--T----TCC-CC--ACA--CT----GGA----G--GC---C---TCAAGG--CTC-G-A-----A--T--ATTCTCC-AAG----TCAT---AG---AAAA---CCCAA--A-A-T........... 7849
MND_2.CM.98.CM16       --CTCCG-GT-A--AA----A--TA-------TG-A---A-T--T-----CC-CC-TACA--CT----GGAC--TG--GCT------GCAAGA--ATCC--G-----C-----ATTCGCC-AGG-A---CA---ACA---A--A---CCAAA-GAC-CT........... 8216
MND_2.GA.x.M14         --CTCCG-GT-A--AA----A--TA-------TG-A--AA-T--T-----CC-CC--ACA--CT----GGAC--TG--GC---C---TCAAGA--A-CAG-A-----A-----ATTCTCC-AAG-G--TCAT---AA---T--G---CCAAA-GAAA-T........... 8145
OLC.CI.97.97CI12       --G--AG-AT--C--A-G----TA-------AG-G---A-C--------TC-CC-G----GA--A--G-A------C--A----GGCA--A--AAAACAA-----A--TGAATTT-TG-AAG-G---CAT---A-A-----A-----CTAC-C-.............. 7690
GRV.ET.x.GRI_677       -----A--GT-----A-G--G--CA-T-----TG-G--AA-C---A-T--CC-CG--ACA---C-C--G-AG------G------GGCA-GG--AAATTCA-----A--TG-GT-G----AAG-A--TCA---CA-A-----A-----GTA--A-AAC........... 8073
VER.KE.x.TYO1_patent   --------GA-----A-G--G--CA-T-----TG-T--AA-C---A-T--CC-CG--ACA--CC-T--G-AG------G------GGCA-GA--AAACTCC-----G---G-AT-G----AAG-A--TCAT--CA-A---GAG---CCCTGG-CAAAT........... 8190
VER.KE.x.9063          --------AA-----A-G--G--CA-T-----TG-G--AA-C---A-T--CC-CG--ACA---C-T--G-AG------G------GGCA-GAT--AACG-C-----G---G-AT-G----AGG-C--TCAT--CA-A--A--A---C-ATGGCAAAAC.......... 8214
VER.KE.x.AGM155        --------GA-----A-G--G--CA-T-----TG-G--AA-C---A-T--CC-CG--ACA---C-T--G-AG-----GGC-----GGCA-GGT-AAACGC------G---G-GT-G----AAG-A--TCAT--AA-A--A--C----CGTGG-A-AAT........... 8194
VER.DE.x.AGM3          --------GA-----A-G--G--CA-T-----TG-G--AA-C---A-T--CC-CG--ACA---C-T--G-AG------G----C--GGC--GGT--AATGC------G---G-AT-G--G-AAG-C--TCAT--AA-C--A--G---C-GTGG-A-AAT........... 7712
TAN.UG.x.TAN1          --G-----GA--C-AA-G-----CA-T-----TG-G--AA-C---A-T--CC-CG--ACT--CC-T----A-------G-------GACA-GG--AAATT-G-----A--TG-ATTC----AGG-G--TCA----A-G-----G----C-TTC-ACAAT........... 8171
SAB.SN.x.SAB1C         --G-----AA--C-AA-G-----TA-T-----TG-A--AA-C---A-T--CC-CG--ACT---C-T--G-A-------G-------GGCA-GGT--AAC--------A--TG--TTC-GG-AGG-G--TCA---GA-G--C-T------GTA--A-AAC........... 8367
MND_1.GA.x.MNDGB1      --GTCTG-A--C--AA-------CA-A-----AG-A---A-TT-A---A-TC-CC-GACCAG-T----G-AT--TA-C-----C---GCTT-G---TCTCAA-----G--T--AT-GGC--AGG-G--TCAT---T----AGAG----C-AA--CAA-C........... 7673
DEB.CM.99.CM40         --CGG---G-----AAC-T-A----C------AG-G-GAA-C---A-CA-CC-CC-GACT---A-A--G-AG---T----A--C---GCAAAAT-AAATGAA-----C---G-CTTC----AG--C--TCAT--CA-G--C----------AACTC-CT-.......... 7806
DEB.CM.99.CM5          --CGG----------TC---A----CA-----AG-G-GAA-C---A-CA-CC-CC-GACT---A-A----AG--------A-----GAGCAAGT-AAATGAA-----A--TG-CTTT----AA--C--TCA----A-G--C---------CCA--CA---........... 7727
TAL.CM.01.8023         --G--A--C-----A---T-A-----G-----------A-T---A-T--CC-CC-GACT--GC-T--G-A----T--G----C---GCCAGGT-AAATTCA-----A--TG--T-G----AG--C---TA---CT----A--G------AAG-CA............. 7443
TAL.CM.00.266          --G--A--C-----A---T-A-----G-----------A-T---A-T--CC-CC-GACC--GC-T--G-A----T--G----C---GGCTAAG--AAATTCA-----G---G-AT-G----AG--C---TA---CT-G--A--G-----CAAA-CA............. 7893
SUN.GA.98.L14          --G--AG-G--AC-AAG------CA-A-----AG-G---A-T-----A---C-CC-GACA--CC-A--GGAG--TG-GC-A--C---AGC--TT-A-CTTC------G--TCAAT-G----AAG-A---CAT---AA------C------TA--A-.............. 8386
WRC.CI.97.97CI14       -----AG-G-------G------CA-------GG-G--AA-C--------TC-CC-GTCA--CA-T---G---T-T--G-A--C--GTT-AAAT--AAACAA-----A--TGAATTG-C--AAG-A---CA---C-----T---------CAGTC-.............. 8194
WRC.CI.98.98CI04       -----AG-AT------G------CA-------GG-G--AA-C--------TC-CC-GTCA--CA-T---G---T-T--G-A--C--GTTAAAAT--AAACAA-------TGAATTG-C--AGG-A---CAT--CAA---A-----------AGTC-.............. 8239
WRC.GM.05.Pbt_05GM_X02 -----AG-G-------G------CA-------GG-T--AA-C--------TC-CC-GTCA--A--C---G----T--G-A--C--GGT-AAAT--AAACAA-----A--TGAATTG-C--AAG-C--TCA---A--A--C-----G-G-AGGCA.............. 7135
LST.CD.88.524          --G--AG-G---C-AAG------CA-------AG-G---A-T-----A---C-CC-GACA---A-T--GGAG---T-G--A--C---GCA--A--T-CCTC------G--TCAGT-G----AAG-G--TCAT---AA---A--A------TA--A-.............. 7246
LST.CD.88.447          --G--AG-G--C--AAG------CA-------AG-G---A-T-----A---C-CC-GACA---A-T--GGAG-----------C---ACT--G--T-CTTCC-----G--TCAAT-G----AAG-A--TCAT---AA---AGAA-----CTA--A-.............. 7243
LST.KE.x.lho7          --G--AG-A-----AAG------CA-A-----AG-G---A-T-----A---C-CC-GACA--CC-T--GGAG--TG-G--AC-----GC------T-CCTC------G--TCAAT-G----AGG----TCAT---AA----GAG----CATA--A-.............. 8318
SYK.KE.x.SYK173        --G--A--C--C--AGGCT-A--T--------TG----AA-C--TA-C--CC-GC--ACA---T-A--G-CG--T--C-GA--C--GGCCA--T--TCA-A------A---G-ATTC--G-AG--C---CAT--A--A---A-G---G-GAAAGCA--TG-A--C--C-- 8044
SYK.KE.x.KE51          --------C--C--AGGCT-A--G--------TG-G--AA-C--TA-C--CC--C--ACT---C-A--G-CG--T--C-GA--C---GCCA-TA--TCA-A------A---G-ATTT--G-AA-----TCA---A--A--AA-G---C-GCAAGCA--TG-C-----C-- 7737
COL.CM.x.CGU1          --GTCTGCA---C-A--G-----G--------GG-T--AA-C-------T-C--G-TGCCA--A-A---G-T--TT-GG----A---GCAAAG--T-CC-GCATT--G--TG-AAAC-TG-AA-----T-GA---AT------C-----CAAG-CC---........... 7476
```

```
H1B.FR.83.HXB2           T...............................CTGGAACAGATTTGGAATCACACGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCTTAATACACTCCTTAATTGAAGAATCGCAAAACCAGCAAGAAAAGAATGAACAAGAATTATTGGAATTAGATA 8217
Env                      __.__.__.__.__.__.__.__.__.__._ L  E  Q  I  W  N  H  T  T  W  M  E  W  D  R  E  I  N  N  Y  T  S  L  I  H  S  L  I  E  E  S  Q  N  Q  Q  E  K  N  E  Q  E  L  L  E  L  D
H1A1.UG.85.U455          -...............................-A---GG-C--A------A---T-------C--C-A-----G-A--------GT-G-------G--A-----T-TCAAC-------G--------G--------G-------------T---C-------C---G--C- 7647
H1B.US.90.WEAU160        -...............................-A---TT-C---------A---T------------------G--------------------G--------T--AA-------------------G-----A--------A-----G------------------G---- 8224
H1C.ET.86.ETH2220        -...............................-AA--GG-------G--A---T---------C------T---------GT----T--GA-A-----T--AAT--GC------TC-------------G--C--A------A----T---C-A-C---G--C- 7593
H1D.CD.84.84ZR085        -...............................G-A--GT-T-----GGGA---T---------C------A--------G---------GG-------T--AA---G------------------T-------------------A-----------A-----G--C- 7738
H1F1.BE.93.VI850         -...............................-A----G----------A---T------------------A-AG--G----G------T---A-A-----T--AAA--G-------G--A------G-A----------G--------------------------C------C- 7494
H1G.SE.93.SE6165         -...............................TATA-TG-A-----G--A---T---T-----A--A----A--G-----C-----------TA-CA----T--AG-C--C------------A--G-------G------------------C-------CC--G--CC 7635
H1H.CF.90.056            A...............................-A-AGTG-A--C---G-CA---T---T--------A-----T-A-C------G----------GAGGA----T--AGG--GC------TC---------C------G-------------G--C-----A-C---G--C- 7536
H1J.SE.93.SE7887         -...............................TAT--GG-C------G-GA---T---------AC-A-----A--G------------------G-AA-----T--AGTC-----------G-A-----------G-A----------G-----T------A---C-------C---G--C- 7524
H1K.CM.96.MP535          .............................TCTTG----G---------CA---T------------------A-A---G---GG-------T--GA-AC----T-TAAG--------------A------A---C------G-------------------T-------C---G--C- 7377
H101_AE.TH.90.CM240      -...............................T-T---G---------CA---T--A----A--A----G----------G-----------A-CA----T-TGAGA--C--AC----------G-----A--G--C-G-------A-G--T--G--A------G---- 7770
H102_AG.NG.x.IBNG        -...............................T-TA-TG-C--A---G--A---T--------AC-A----G-A--------G---------GA-A-----T-TAATC---------------------G---G---G--------------------C--G-----C---G--C- 7723
H1N.CM.95.YBF30          CTCT............................TAT--TACA--C------A-TTTA------CAAC-A-----TGAGA--G-A-GA--C--TT--G-TG-C--TTTTGGAC-T--A---C-GG-A---G-A--A--GA-C-CA-----GA--TC-C-C---------G---C 7768
H1O.BE.87.ANT70          ................................AAC---AGC------G-CAC-TTA--A---CA---A-----TC-GC-G--A-G---CATA-GCTC-ACC---T-TGAGGA---AC--A-GG-A---GTA-----G---C-A-----GA--A-G--GC----G------G 8265
H1O.CM.91.MVP5180        .......................AGATATAATGAT--CAGT------G-CA--CTT--A---CA-C-A------CA-C-C--A------GTA-GCTC-A-T---T-TGATGA---AC---C-G-A---G----A--G---------T-A---C---G-----GC-----G 8291
CPZ.CD.90.ANT            GCAAACATGT...GCAAAGAACAGCAGTGATA-AC--TGT------G-AA-T-T---A---CAA--A---------TT-G-AC-G----CA---G-ACAG---T-TAATA-CT-AC--AT-G-A--TG-G--A-----G-GA---A--A-G-------AT---C----C- 7676
CPZ.CM.01.SIVcpzCAM13    ..........................AAGTCTA--AGTG-C--------CA-TTTA--A---CAAC-A-------A-TTG--C-CT---------G-GAC---TTTTGGT---T-A-----GG-A---TCA--A-----G--A------A-G--C--------G--G---C 7801
CPZ.CM.05.SIVcpzEK505    .......................AATAACTCCTAT--CACA--A--GGAA-T-T---A---CAAA-T-----TGAGC--G-A-GA-----TT--G-TG-C--TTTTGGGC-GC-A--GC--G-A---G-G--A---AGC-TA------A-GTC-C-C---------G---C 7703
CPZ.CM.05.SIVcpzMB66     ..........................AAATCTTAT--TG-T------T--A---T---A------C-A-------A----G---G----------GATG-T--CT-TAATC--T-A---A-GG-A----C------G-----T------A-----C-CC-A---------C- 7841
CPZ.CM.98.CAM3           ..........................ACATCCT-T--TG----C------A-TTTA--A---CAA--C-------A-AGGG-A--A-----TT--G-TG-C--TTTTAG-C-T--A---C--G-A---G-G--A---A-C-CA-----GA--AGTC----------G---C 7633
CPZ.GA.88.GAB2           A............................TCTTAT--GG-C-------G-A-TCT-------CA-C-A-----T-AGTT-G-AG-A---C-T---G-AACC--TTTTCAA---T-AC-GAG-G-A--TG-G--A--GA-T-GC------A----GC----------G---C 7604
CPZ.CM.98.CAM5           ..........................ACATCCTAT--TG----C------A-TTTA--A---CAA--T-----T-A-A-GG-A--A------T--G-AG-C--TTTTAG-C-T--A--GC--G-A---G-G--A--GA-C-CA------A---TC-T---------G---C 7939
CPZ.CM.05.SIVcpzMT145    ..........................AGTGACTATC-GGTT---------A-TTAC--A---CAAC-A-----T------G-A--T---------G-AC-C--CT--A-A--GC-A--G--GG-TA-T-CA-----G--G--A------A-G--G-----------G---T 7729
CPZ.TZ.00.TAN1           -ACAAAGTGCAATCACAGTGATGCAAAGTACTAT--CTGT--A-----CA-TTT---T---CA---A-----TC--TT-G-AG-A--C-CT---G-AACC---T-----C-GT-A--GA--G-A----CA--A--G--G--A--CA--------G--G--A--------C- 7873
CPZ.CM.05.SIVcpzLB7      ..........................AAAACATATA-TG-A-----G-CA---T--A----------A---------G---G-A-GA---------GAAA-T--TT-TGGG-----A--GC--G-A---G-------G------T-----GA--A--C-T--A---C-G---C 7728
CPZ.US.85.CPZUS          ..........................CTCTCCTAT--TGCT------GGCA-TCTA--T---CAA--A----------A--G-A-GA--C--TT--G-TACT--TTTTAGTC-T--A---C--G-A---G-A-----GA-T-CA-----A--TC-C-C---------G---C 8225
CPZ.GA.88.GAB1           C.........................AATTCCACA--TG-C-----GGGA-TCTA--A---CA-C-A-----T-A-TT-G-GTCT--C------G-GAA---TTTTGGTC-GT-A--G---G-A---TCA--A-----G--A------AGG--CC-----------G---C 8257
MON.CM.99.L1             ....................TGGGCAAAAGGT-ACTTC-CTGAG---G-CA-T-T---A---CAAC----AGTGAGTT-G-AG-T---G------TGACC--T--GCAGC-CT-G----CTG-----G-G-------GC--A--CC----T--G---A--A-GCCG-GAC 8177
MON.NG.x.NG1             ....................TGGGCTAAAGGC-ACTTC-CCGAG---G-CA---T---A---CAAC-A---AGT-TGCT-G--G------G-T--GGCAC-T--C--GCADC-CT-G--G--GG-T--GG-G--A----G---A----C---T--GC-CA--A----G-GAC 6802
GSN.CM.99.CN166          ....................TGGGCTAATGGCTC-CTT-CAGAC---G-GA---T-------CAAA-----AGT-TGTT-G--G-G---G-----TATAC------GCAG--GT-A--GC--G-AA-CC-G--A----C-TCA--CCTGA-T--GC-GA--A-G--GTCC- 8151
GSN.CM.99.CN71           ....................TGGGCTAAGAACTC-ACG-CAGA----G-G----T-------CA-------AGT-AGTT----G-A---G-----TATACC-----GCAA---T-A---A-TG-AA-CC-T--A---TC------CATGA-T--C------AA-G---TCC- 8163
MUS_1.CM.01.1085         ....................TGGGCAAATCACACCC-G-CAGG----G-AA-T-T---A---CAAC-----AG---GTTGG-AG-----G-T--T-CAACT-----GGAG---T-G---TTGG-A-----GA-----G---G------C-G--TA------CAAA----GTTAG 8148
MUS_1.CM.01.CM1239       ....................TGGGCCAATCATACCTT--CAGGC---G-AA---T-------CA-C------AG--AGTTGG-GG-T---G-T--T-ATACC-----GGG-C-CC-A----C-G-A------G--A--G---G-A--CC----CA-----GCAAA-----TTAG 8117
MUS_2.CM.01.CM1246       ....................TGGGCAAACCACACCC---CAGGA---G-CA---T-------CAAC-----AG----TT-G-GG-----G-----GCAG-C--C--TGGTC--T-G---ATGG-T---C-A--A--G---G-A--CC----CA-G--GCA-A-GC-TTTAG 8246
MUS_2.CM.01.CM2500       ....................TGGGCTAATCATACCC-G-CAGAA---G--A---T-------CA-------AG---GTTGG-GG-----G----TGCAACC--C--AGGTC--T-A--GTTGG-T--G--A-----G--GG-A--CC-G--TA-GC-GCAAA-GC-CTTGG 8169
RCM.NG.x.NG411           ...................................AC--CAGAG-----CA-ACA---A---C-T--------A---A-T---TCA-GA-TAGA-G-AAAC--CTCAGTAGCTT-AC-G--TG-A---G-A-----T---CGT----TG--T--C---GAAA----GA--T 7933
RCM.GA.x.GAB1            .............................AACAAAACG-CTGAG---C-GA-AGAAT-A---T--C-------A---A-T--CTCTT---TAGAGGCTAAC--TACAATAGC-T-AC-G--GG-C--GG-T--A--T--G--A-----TG--T-----GGA-A-----AG-- 7889
DEN.CD.x.CD1             ..........TCAATACCTGGATTGGCATTATACC-G-CTGAC-TCC-GA---T---A---CAAC-A-----G-C---C--AGCAGC-CTATTTG--AATG-GTCAGA-GCCT-GA-GA--G-A--GGCA-----G----GA----TG--C---G--CAATCC---T--G 8280

MAC.US.x.239             ................................--AAC--CA-AG-----CA-TGA---T---CAA--------GC--A-GG--G--TTC-TGGA-GAAAAT---ACAG--C-CC-A--G--GG-A----TT--A-----G-----CATGT-T------CAAA-G--GA--- 8623
Env                      __.__.__.__.__.__.__.__.__.__._ L  T  P  K  W  N  N  E  T  W  Q  E  W  E  R  K  V  D  F  L  E  E  N  I  T  A  L  L  E  E  A  Q  I  Q  Q  E  K  N  M  Y  E  L  Q  K  L  N
H2A.GW.x.ALI             ................................T-AA---CTGA----G-CA---T---G---CAA--------ACA-C--G-CCGTT-CCTAGAGGCAAAT--CAGTGAACAGT-A---CGGG-A----TT--------------ACGT-T---C--CAAA-----A--- 8662
H2A.DE.x.BEN             ................................T-ATCG-CTGAC-----AA-T-T---A---CA----------G-A-C--G-CCG-T-CCTAGAGGCAAAT--CAGTCAAAGTT-A------G-C----TT--A-------------ATGT-T------CAAA-----A--- 8678
H2A.SN.x.ST              ................................T-AACG-CTGA-------CA---T---A---CA---A-----GCA-CG---CCG---CCTAGAGGCAAAT--CAGTGAAAGTT-A---C-GG-A----T-----------------CATGT-T---C--CAAA-----A--- 8108
H2B.GH.86.D205           ................................--CAC--CA-A--------A---T---A---CAAC-------A-AGC--G-CC-TTTC-TGGATGCAAAT---ACGG-TC-GT-A-----GG-T--G-TA--A--G---------CATGT-T-------CA-A--A--A--- 8654
H2B.CI.x.EHO             ................................--TA-G-CAGAC-------A---T---A---CAAC-A-----G--GC--G-CCG-TTC-TGGATGCAAAT---ACAAAA---C-A-----GG-A--G-TA--A------------CATGT-T--G---CA-A-----A--C 8640
H2G.CI.92.ABT96          A...............................T----G-CACAA------A-T-T---A---CAA--A-----A-AGC-G--C---TT--TGGAGGA-AAC--CACACGGC--T-A------G-A----TT-----------A--ATGT-T--G---CA-A-----A--- 8051
H2U.FR.96.12034          ................................T-AAC--CT-A----G--A-T-T---G---CAA--A-----GGA-A-GG-G---T-C-TGGA-GAAAAT--CACACAGC--T-A----C-G-T----TA--A-----G--A--CATGT-T------CA-A-----A--- 8162
MAC.US.x.EMBL_3          ................................--AAC--CAGAC-----CA-TGAT--T---CAA--------GC--A-GG--G--TTC-TGGAGGAAAAT---ACAG--C-CC-A-----GG-A----TT--A-----G-----CATGT-T------CAAA-G--GA--- 7754
SMM.US.x.H9              ................................T--AC--CT-RC-----CA-T-T--YT---CAA--------A-A-C-GG-G---TTCCTAGAGGCAAAT---ACTSAA-Y-T-G------G-A----TT---------M-----ATGT-T-----GCAAA-----A--- 8112
SMM.US.x.PGM53           ................................T---TG-CT-A------CA-T-T---T---CAA--A-----A---C-GG--G--G-CCTAGAGGCAAAT---ACTCAGGC-T-G------G-A----TT----------------ATGT-T-----GCAAA-----A--- 8569
SMM.US.x.SIVsmH635F_L3   ................................T---TG-CT-A----G-CA-T-T---T---CAA--------A---A-GG--G--TTCCTAGAGGCAAAT---ACTCAA---T-A------G-A----TT----------------ATGT-T-----GCAAA--C--A--- 8645
SMM.SL.92.SL92B          ................................--T-TC-CAGAC-----CA-T-T---A---CAA--A-----G-A-A-GG--G-ATTCCTAGAGGCAAAT---ACTCAAA-GT-G--G---G-A--G--G---G-A-G-CTA-------G--A--CATGT-T------CA-A----GA--- 8071
STM.US.x.STM             ................................T---T--C-GA------CA-T-T---A---CA---A-----G---A-GG--G--TT-CTTGAGGCAAAT---ACACAAC-GT-A------G-T---GTT--A--G---------ATGT-T--G--GCAAA--C--A--- 8274
MNE.US.x.MNE027          ................................--AAC--CA-A------CA-TGA---T---CAA--------GC--A-GG--G--TTC-TGGAGGAAAAT---ACGG--C-TC-A-----GG-A----TT--A-----------CATGT-T------CAAA-G--GA--- 8100
DRL.x.x.FAO              ................................G--ACT-CA-A-----CCTCAGA---A-----------C-A-A-AG-G-AG-T-G-ATAT---A-AAC--TACTCTTGATT-AC-GA-GG-AT-TG-A--AG--C--------ATCTTT--G---CAAA------GGG 8046
MND_2.x.x.5440           ................................A-A-TG-CC-A-----CCTCAGA---A---C----A---------AG-G-A----GCATAGTG-CAAAT--GAC-ATAGACT-AC-GAGGG-AT-TG-ATTAG--C-G-G---CATCTTT--G--GCAAA------GGG 7988
MND_2.CM.98.CM16         ................................A-AAC--CA-A-----CCTCAGA---A---C----A-----T-A-AG-G-A-CAGC-CTA--TGA-AAT--GACAGTAAACT-AC-GA-GG-AT-TG-ATT-G-GC--------AT-T-T--GC--GAAA------GAG 8355
MND_2.GA.x.M14           ................................A-AACT-CT-A-----CCTCAGAA--A---C----A-----T---AG-G-G-CAGC-AT---T-A-AAT--GACAAT-GATT-AC-GAGGG-AT-TG-ATT-G--C--------ATGT-T------CAAA----G-GAG 8284
OLC.CI.97.97CI12         ................................-A-TC--CAGA---GGCA-TGAT--T---T--A-----C-A--GA--G-AG-T----TA---CAGGA---G-TGA-AG-T-GAC-AT-G-AT-TC-AG----GC---G----C--T-T--G--GCA-AGG--TC-AG 7829
GRV.ET.x.GRI_677         ...................................ACT-CT-AG---G-CA-T-T---T---T---------G---C------TGCC-TGGA-G--AAC---ACTCAAC--T-G------G-A-----T---G--TC-------CTGG-TCTG-ACCA-A----G---G 8209
VER.KE.x.TYO1_patent     ................................-G-ACT-C-GA----C-AA-T-T---T---T---------A---C---AGCTG---TGGA---AAC--TACGAGACA-T-A-TGA-GG-TAG-G-A--AG-G---------CT-G-T-CC-ATCA-A-G---AC-- 8329
VER.KE.x.9063            ................................A--ACG-CC-A----C-AA---T------T---------A---C-G---GGAG-G-TGGA-G--AAC---ACAGAACAGT-A-T-A--G-AAG-G-A--AG----G--A---TT-G-T-CC-ATCAAAG---GACAT 8353
VER.KE.x.AGM155          ...................................AC--CAGAG------A-T-T-------T----------A-A-C-G--AG-AGGA-TGGAGG--AAC---ACAAAACA-T-G---C-GG-AAGGG-A--AG-G---------TTGG-T-CT-ATCAAA-G--GTCAG 8330
VER.DE.x.AGM3            ................................AG-ACC-CTGA-------A-T-T---T---C----A-----A---C-G--ATCGT---TGGA-G-TAAC---ACAA-ACA-T-A--G---G-CAG-GCA---G-G--G------TTGG-T-C--ACCAAA-----AG-- 7851
TAN.UG.x.TAN1            ...................................AC--CAGAC---G-CA-T-T-------CA---A-----G--CC----A-CTGCC-TGGA-G-AAAT--TAGTA-TACTC---TCA--G-AT-TG-G--AG-GC-G--A---ATGG-TACT-ATCAAA-G----G-G 8307
SAB.SN.x.SAB1C           ...................................ACT-CCGAC---G-AA---T---G---CAA--A-----G--GC----AG-A--G--TGA-GCTAAC--CAGTAGGA-TT-A--GC--G-A--TG-A---G--C-------TTGG-TAGT-ATCAAA----G-TGT 8503
MND_1.GA.x.MNDGB1        ................................A-CACT-CA-A-----CATCAGAA--T----A---A-----G-C-AG--C-G-TT--CTGCA-CAAAAC--TACAGAAA-GT-AA--C-GG-AT-TG-T-GAG-GC---GA--CAC-T-T------CA-A-G----GAG 7812
DEB.CM.99.CM40           ................................GA---T-CAGAC---G-CA-T-T-------CA---------A-TGA-GG-AGCA-----T---GATGAGTGGG-AGGAGCTT-AC-GAG-G-A---G-G-----G--G-G-----T---T-CCC--CAATC-C--C-AG 7945
DEB.CM.99.CM5            ................................GGA----CTGA-------A-T-T-------CAA--------A--GA--G-AGCA--C--T---GATGA-TGGG-AGGAGCCT-AC-GAG-G-A--GG-A--A-----G-GA----TG--T-CC---CA-TC-C-GAC-G 7866
TAL.CM.01.8023           ....................TGGACAAACTCCACAA-C-CTGA----C-GA---T---A---CAA--------G-AGTTGG-AG-----GC-T--GA-ACT--CACAGTA--GT-AC-----G-A---G-A--A--G---GA----TC--T--G--GCAAA-G---A--G 7594
TAL.CM.00.266            ....................TGGTCGAATTACACA--T-CTCAA---C-GA-T-T---A---CAA--A-----G-TGA--G-AG-----C-----G-GC-C---TCACAG--GT-AC-----G-A---G-G--A--G---GA----TC--T--G--GCAAA-G---A--G 8044
SUN.GA.98.L14            ................................A-CACG-CC-A-----C-A-GGAC----------A-----T--GC--G-A--A-TG--TGATGATAAT---ACAG-TC-TC-CC-G--GG-AT-TGTAACAG-GTT-G-----C--A-CA-G--CAAAC-GC--C-AG 8525
WRC.CI.97.97CI14         ..........................TGGAATGCAACT-CA-A----G-AA-TGAA--T---CAAC-T---T-T-A-AG--CAG-TG-ACTGCAGGCAAAT---AGTG-AA--T-A-T-ATTG-C--TG-T--AG-GC--GCA---ATGTTT--G--GCA-A-G---TC-G 8339
WRC.CI.98.98CI04         ..........................TGGAATGCTACT-CA-A-------A-TGA------CAAC-T---T-T--GAG--CAG-TG-A-TGCAGGCAAAT---AGTGTAC-CC-A-T---GG-T--T---ACAG--C-------CATGTTT------CAAA-G--GC-AG 8384
WRC.GM.05.Pbt_05GM_X02   ..........................TGGAATA--AC--CTGC-------A-TGAA--A---CAAC-A---C-TTT-AG--CAG---CC-TG-TG--TAAT--CTCAGATA-GT-A-TGC--G-T--TGCT--AG--C--TCA---ATGTTT--G---CA-A-----AC-G 7280
LST.CD.88.524            ................................G-TACT-CT-A-----CAAGGGAT--------A--A-----A---C--G--GGATC--TAGA-GCAAAC--CAC-A-AC-GT-GC----GG-TT-C-CAACAG-GTT-G-----AGGA-TA-C--TAAAA-G---C-AG 7385
LST.CD.88.447            ................................A-AACT-CT-A-----CAAGGGAT--------T------------C--G--GGAGTA-TAGA-GCAAAT--CTC-A--C-GT-GC-----G-CT-T-C-ACAG-GTT-G-A---AG-A-T-CC--CAA-A-GC--C-GG 7382
LST.KE.x.lho7            ................................A-CAC--CC-A-----CCA-GGAT-------GA--A-----ATC-A-GG-AGCA-T---TGAC-AGAAC--TACT--TC-GT-AC-----G-AT-C-C-ACAG-GCT-G-----C-GA-TA----TAA-A-G---C-AG 8457
SYK.KE.x.SYK173          CAAC......................TTCTGT-CCA---CACAG-----AA-T-T---A---CACAGA-----ACA---GG-AG------TG--TGA-CAT--TG-TGGA--GC-CAG----G-A---G-A--A--G--G-GA----TG--T--TC--ACAA-G---C-GG 8193
SYK.KE.x.KE51            -.........................CGTTGC-CAAC--CCCAA---G-AA---T---A---CATACA-----AC-GC-GG-AG-T----TG---GA-CAC---G-TAAT---C-GAG----G-T--GG-G--------G-------TG--T--TC-TACTA-----C-AG 7883
COL.CM.x.CGU1            .............................GGAGA---GG-TCCA---C-GA---T--------AAC-----C-TGAGCG-G-A-GA---------GATA-----G-GG-AGACT-A-T---GG-AT-TG--TTA-----GG-A------A-GA----GGCA-------GGG 7618
```

```
                                                                                                                                                                      Tat Rev intron end Tat Rev exon 2 start
H1B.FR.83.HXB2            AATGGGCAAGT...TTGTGGAATTGGTTTAACATAACAAATTGGCTGTGGTATATAAAATTATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCACCATTATCGTTTCAGACCCAC 8384
Rev exon 2                                                                                                                                                                             N  P
Tat exon 2                                                                                                                                                                              P  T
Env                       K  W  A  S  .  L  W  N  W  F  N  I  T  N  W  L  W  Y  I  K  L  F  I  M  I  V  G  G  L  V  G  L  R  I  V  F  A  V  L  S  I  V  N  R  V  R  Q  G  Y  S  P  L  S  F  Q  T  H
H1A1.UG.85.U455           -G-------A-...C----A-------------CT------------------T-G-C----TG---A-------------AA---A-----------------A---G---------A-C-------------------C-----TC-G-------------TT 7814
H1B.US.90.WEAU160         -----------...-------C-------G-----T----C-------------------A-C-----------------------A--------------------A-----------C-----A--------------------C--------G--A------------ 8391
H1C.ET.86.ETH2220         ------A--A-...C-----------------------------------------------A----------------G-----TG--A-------------A----------G------------------------------C-------------------------TT 7760
H1D.CD.84.84ZR085         -G-----G---...----------------G-------C-A-------------------A-----------------------A-----------------------A-------------A----------------G--C-----T----------------T- 7905
H1F1.BE.93.VI850          -----------...C--------------G-----T----C-------------------A----------------G---------A---C------------------------------A-----------A------C-----T--G--A--A-------TT 7661
H1G.SE.93.SE6165          -G---------...---------------GG--------GA-----A------------A----T----------------T--AA-----------------------G-----------A--------------------C-----T--G--A-----------TT 7802
H1H.CF.90.056             -----------...C------C-------G-----T--C--------------------A------------------------T--AA-------------A---------G-----------A--------------------C-----T--G--T-----------TT 7703
H1J.SE.93.SE7887          -G---A---A-...---------------------T----C-----------C--C---A-------------A------T--AA-------------A---------G---G-------A-----------------------C-----T--G--A-----------TT 7691
H1K.CM.96.MP535           -----------...C--------------G----------A-----A------------A-------------A----------A-------------C---------G-----G-----A-------C-----------C-----T--G--A-----------TT 7544
H101_AE.TH.90.CM240       -----------...C--------------G-----------------------------A----T-----------------T--AA-------------A---------G-----------A--------------------C-----T--G--T--C-------CT 7937
H102_AG.NG.x.IBNG         -G---------...C-A------------G---------------A---------G-A----T-----------------T--AA-------------------------G---A-----A-A--------------------C-----T--G-----C-------TT 7890
H1N.CM.95.YBF30           ------AC---...C-----GC-------GGT--T-----A------------------A--GCT-------------C----A-T-----CA------CA-AAG-ATA---A-AA-----A-AGCA-------------------T--CC-T-----G-------TT 7935
H1O.BE.87.ANT70           -------CTC-...A-T---------C--G-------T--A---T--------------A--GCA-----C----------CAC-A-----GG-G---G-TA-CA-GATA------AA----------A-ACA--------------CA---CC-C-----A----T--C- 8432
H1O.CM.91.MVP5180         -------CTC-...C-T------------G-------T--A---T--------------A--GCT-----C-----G----CAC-AA-----A-----G-TA--A-GATAA-----AA-C-------G-ACA--------------CA---CC-C-----G----T--CT 8458
CPZ.CD.90.ANT             -----AGCTCA...-A-------------G--------C-A-----A------------A----T--T--------------CTA-T-----AC-------TT-GC---TAT--G--AG-TGCT-A-GA-AG--------------CAT--TC-G--A------...... 7837
CPZ.CM.01.SIVcpzCAM13     ---------TC-...--------C-----G-------C--C-----------C------A-T---C-T---GC---------GA-TA---------------TA-CA-GT-A--TG-C--AG-GA-C-GG-----C---------T--C-----CC-C-----G-------TT 7968
CPZ.CM.05.SIVcpzEK505     -----T------...--------C-----G----T-----A-----------C------A-T----------G------C----A-T-----CA--------A-AAGCATAA--A-G-----G--AGCA-------------------T--CC-C-----G-------TT 7870
CPZ.CM.05.SIVcpzMB66      -G-----G--C...C-C------------G--------TCA------------------A----T--C--A-----------A--AA------------------------T-G-----A-----A--------------------------TC-G--A-----------TT 8008
CPZ.CM.98.CAM3            -----T------...-A------------G-T--T--CCGC-----------C-------G--T-------------C------A-----CA----------AGGG---A-CA-A--AT-T--AGCA-AG-------------C--T--CC-C-----G-------TT 7800
CPZ.GA.88.GAB2            -----T-CTCA...-A-------------G----------C------------------A-G--T-----CT----G-----GG-TA-----------T--AA-A-G----G-AAA----A-A-GA---AG-------------C-----T-------------------TT 7771
CPZ.CM.98.CAM5            -----T------...-A----G------C--T-----TC-G---T--------------A----TT-G--AG------C---T--AA-------T---C--A-AGGAAT---TA-G--AG-TA-AGCA-AG-------------C--T--CC-C-----G-------TT 8106
CPZ.CM.05.SIVcpzMT145     C------T-A-...-A----G--------G------------------T------C-----GA-G---T----AG-----------A-TA----C---------TGC-----AA--GGA--AC--A-C-----G------A-------C-----T--G--A--G-------TT 7896
CPZ.TZ.00.TAN1            -----AGC---...C-T---G--------G-T------C-A------------------A--GCT-----C-------C----A--A-----AC-T-----TC-CA-GTT-A--G--AA-G-----T--GCA------------T---A----CC---TT-CA----T--CT 8040
CPZ.CM.05.SIVcpzLB7       -T---A----C...C--------------G------GTC-C-----------C------A--------T------A-----------AA---TG-GT--G---A-----------G--CG-------A-------C-----------C-----TC-G--A-----------TT 7895
CPZ.US.85.CPZUS           -----T------...C-A-----C-----G-T--T--C--C------------------A----TT----AG------C-A----A-----AA-C-----T--AGG--TGA--T----AC----AGCA-A--------------T-----T--CC-C-----A-------TT 8392
CPZ.GA.88.GAB1            -------C--C...---------------G----------A------------------A-T---C-T---GC----------AA-CA---G--------TA--A-GA-A--TT-C--AG-----C-GG---------------C--C-----TC-C-----G-------TT 8424
MON.CM.99.L1              ------ACTTC...C--------------G----CT-C--A---T-------C--C---A--------T-AG------C--CT---A----C----------C--A-GTTCA---AGGAG-TA-C-G---GT-AG-----------C--T-TTC-T--T-C-----T--CT 8344
MON.NG.x.NG1              ------ACT-G...-A-------------G-T--CT-C--A-----T--------T--GA-C----------AG-G---C--CT--------CC----GG-GC-CA-GTTCA-C--AGGG--CA-TCG---G--AG-------T--C--T-TCC-C--TCCA---...... 6963
GSN.CM.99.CN166           ------ACTCC...C-T---G--------G-T--TT-CG------A-A----C--T--GA-C--TG----AG----T-C--C---AA---C-------------AA-GTT-A-C---AA---GC-T-GA--GA-----------C--C-----TC-C--CCCG---T--TT 8318
GSN.CM.99.CN71            ------ACTCC...C-T---G--------G-T--TT-C---------C-----C--T--GA-T----------G----T-C--CT--A----C---G------A-AA-GTT---T---AA---GC-T-GA--G-----------C--C-----TC-C--CCCG---T--TT 8330
MUS_1.CM.01.1085          -G----ACTCC...C-C---G-A------G-T--CT-C--A-----C--------C---A-C---TGC---G----T-C---TC-A--GCT---T--GC----CA-GTT---GT-AGGA---C-A-GG---------------C--C--T--CC-T--TCC-----T--TT 8315
MUS_1.CM.01.CM1239        ------ACTC-...--A---G-A------G----CT-C--G-----T--------T--GA-C--TTGT---G------C---T--A--GCT---T------T-CA-GTT---G--AGGC---C-T-GG--GA-C---------C-----T--CC-T--TCCA----T---T 8284
MUS_2.CM.01.CM1246        ------ATTCC...C-T---G-A------G----TT-C--A------C------G-C--GA---TTGC---G------C------AA---C---------------A-GTT-A--T-AGGA--GC-T-GA--GA-----------C--C--T--CC-C--TCCA----T---T 8413
MUS_2.CM.01.CM2500        ------ATGCC...C-C---G-A------G-T--TT-C--A-----C-TC---G-T--GA---TTGCT-AG------C------AA---CA--------TA-AA-GTT-C-G--AGGG---C-C-GA-AG------------T-----T--TC-TC-TCCA----T---T 8336
RCM.NG.x.NG411            C-----G-GAC...A--CTC-G----C-A-----GGATTGG---T-AAA--------GGA--GGA--TT-C--TA-TTT---AA-AA---C---------A-C---TTAC-CTGGAGCTGT--T-G--AGC-----GCA--T--CA-G--T--GCTC-C--CCC-A-CT 8100
RCM.GA.x.GAB1             -T----G-GA-...GCA-TC-G----C-G--TC-TGACTGG---A--CAA---------A--GG-T-CT-T------AT---TA-CA----A------G---C--GGCTGT--TGGAA-TGTC-T-G--ATC-------A--G---AGG--TC-C......TCCC-A-C- 8050
DEN.CD.x.CD1              -T----AT-A-...----------------G-TT---GC--A---T-C----GG--T-G---GG-AG-GTATG--A---C---G--A---ATA----------A-AA-GTT-A-CA--AA-G-GT-A-G---GC--T-TAGA--C--C--T--TC-T-T-CAGATCC-TAT- 8447

MAC.US.x.239              GC----ATGTG...--TG-C---------G--C-TG-TTC----A-AAA-------C---ATGGAG-TTAT-----T-T---AG-AA--CTG-----------GA-CTA-A--G-ACAA--GC-AGC--AGT-A--------G---AGG---G-G-TC-C-TCCC-A-C- 8790
Env                       S  W  D  V  .  F  G  N  W  F  D  L  A  S  W  I  K  Y  I  Q  Y  G  V  Y  I  V  V  G  V  I  L  L  R  I  V  I  Y  I  V  Q  M  L  A  K  L  R  Q  G  Y  R  P  V  F  S  S  P  P
H2A.GW.x.ALI              GC----ATGT-...--TACC--C---C--G--T----CGCC---G-CAA------TC---ATGGAG-TTAT---A---T---AA-A----C-C-T--------AA-ATA----G-GCAA--GT-A-G----C-C---A----C---AGG--TG-T-TC-CCTCCC-T-C- 8829
H2A.DE.x.BEN              GC----AT-T-...C-TG-C--C------G--T----CTCC---G-CAA------TC---ATGGAG-GCAT-------TG--AA-AA---C------------CAA-CTA----G-GCAAT-GT-A-G----T----AA----C---AGG--TG-T-TC-C-TCCC---C- 8845
H2A.SN.x.ST               GC----ATGT-...--TG-C--C------G-TT----CTCC---A-CAAA-----TC-G-ATGGAG-TTAT-------T---AA-AA---T------------AA-ATA----G-ACAA--GT-A-G----C----AA----C---AGG--TG-T-TC-C-TCCC---C- 8275
H2B.GH.86.D205            GC----ATGT-...--TG-T---------G--C-C--CTCC---A-AAAA-----CC-T---GGAC--TAT-------C---G--A----TG-----------AG-ATACA-TG-GCAG--GC-AGCA---C-----A----C---AGG---G---TC-CCTCCC-T-C- 8821
H2B.CI.x.EHO              ------AT-T-...--CA-T---------G--T-C--CTCC---A--GCA--C--C-GG---GGAT--TAT------AT---AA-A----TA----------CAA-ATACA-TA-ACAG--GC-AGCA--GC-----A----C---AGG---G---TC-CCTCCC-T-C- 8807
H2G.CI.92.ABT96           GC----ATGTA...--TG-C---------G--C----CTCC---G-TAAA--CG-TT-C---GG-C-TTATG----G-C---AG-AA---TG-----------CA-ATA----G-ACAA---C-AGGA---C-------A--G---AGG--TG-T-YT-C-TCCC-T-C- 8218
H2U.FR.96.12034           -C----AT-TC...--TG-T--C------G-TC-C--CTCA---G-TAAA---G--T-C---GGAC-GTATG------C---AA-AA--ATAC-T---------GA-ATA----G-ACAGC-GT-AGG---TC-----A-A--T---AGG---G-T-TC-C-TCCCA--C- 8329
MAC.US.x.EMBL_3           GC----ATGTG...--TG-C---------G--C-TG-TTC----A-AAA-------C---ATGGA--TTATG----T-T---AG-AA--CTG-----------GA-CTA-A--G-ACAA--GC-AGC--AGT-A--------G---AGG---G-G-TC-C-TCCC-A-C- 7921
SMM.US.x.H9               GSK---ATNT-...--YG-C------K--G--C-T--TTY----A-AAAA------C---ATGGTG--C-------TCT---AG-AA---A-----------GA-ATA----G-GCAG--GT-AGC---GT-A--A-----T---AGG---G-G-TC-C-TCCC-T-C- 8279
SMM.US.x.PGM53            GC----AT-T-...--TG-C---------G--C-T--TTC----A-AAAA------C-G-ATGGTG--C-T-----TCT---AG-A-----G-----------GA-ATA----G-GCAA--GA-AGC---GT-A-----A--T---AGG---G-G-TC-C-TCCC-T-C- 8736
SMM.US.x.SIVsmH635F_L3    GC----AT-TC...--TG-C---------G--C-T--TTC----A-AA-A------C---ATGGTG--C-A-----TTT---AG-A-----G-----------AA-ATA----G-GCAA--GC-AGC---GT-A--A-----T---AGG---G-G-TC-C--CCC-T-C- 8812
SMM.SL.92.SL92B           GC----ATGTG...--TG-A---------G--C-T--CTCC---G-AA-A--C---C-G-ATGGAG-CT-TC----CAT---AA-A---TTG------------A-CTA----G-ACAA--GT-A-G---GT-A--------T---AGG--TG-T-TC-CCTCCC---CA 8238
STM.US.x.STM              GC----ATGTA...--TG-C---------G--C-T--CTC----G-AA-A--C---C---ATGGAG--TATT----TAT---A--A---ATG------G---CAA-CTA-A--A-GCAAT-GT-AGC---GC-----A-A--T---CGG--TG-G-TC-C-TCCC-T-C- 8441
MNE.US.x.MNE027           GC----ATGTG...--TG-C---------G--C-TG-TTC----A-AA--------C---ACGGAG-TTAT-----T-T---AG-AA--CTG-----------CA-CTA-A--G-ACAA--GC-AGC--AGT-A-----A--G---AGG---G-G-TC-C-TCCC-A-CT 8267
DRL.x.x.FAO               -T-TAA--TCA...-G-GCT--C------G-TT-C--TTGG---TCTAAA-----T--GA--GGAT-TT-C------ATG-CAA-AA----A--------TT-AGCA---T--TGGAGC-C---AGG---GT----------T---CGC--TC-T...CC-T-T-T-TTT 8210
MND_2.x.x.5440            -C-TAAACTTCCATGG-CTA-C-G----CG--C-T---TGG----TAA----G-T--GA--GGAC-GT-AG------T--TTA-CA----A--------GT-AGC-TGCT--TGG--AG--T-AGGG-AGT--------------CGC--TC-T...CC-T-TGT-TT- 8155
MND_2.CM.98.CM16          -T---A-TTCC...-G-GCT-GC-----CG--T-T--GTGG-----TAAA---G-T--GA--GGTT--C-A------AT--TGA-TA---TG-----G--TT-AGC-TGCT--TGG--AG-CT-AGGA-AGT----------T---CGC--TC-T.........C-TT-T 8513
MND_2.GA.x.M14            -T-T-A-TTCA...-G-GCT-GC-----CG--C-C--GTGG-----CAAA---G-T--GA--GGA---T-----A--ATG-TAG-AA----AC-T-----TT-AGC-TG-T--TGG----C-A-AGG---GT----------T---CGC--TC-T...CC-T-TGT-TTA 8448
OLC.CI.97.97CI12          ---TAAGTGA-...-G-ACA--A-----C--TT-GG-T------A--ACT---G-----A--GGAC-TT--G----GTT--CA-AT--G--AC-C-AT-----AC-A-GAT--G--AGG-AGA-T-GG-TTA-GC-------T---GTGT-T--GCA-C-C-CA...... 7990
GRV.ET.x.GRI_677          -T---T--G-G...--C---TCA-----CTCAC-GT---C----T-AG-C---G-T---A--GGAT-TT-AG-GA---TGATTA-TC----A------T-T-CA-GG-TAT--TGGGGATGTA-C-GA-ATA--------------AAT--TC-CC-C............ 8364
VER.KE.x.TYO1_patent      GT---T--GA-...--C---TC------CG-TT-CT----A-----TAACAT-T-----A-GGGAT-TT-AG--A---T---AA-AA----G------T--C---ACA-A---TA-GGATGTA-AGTG--G--------------GTT--TC----TCCA......... 8487
VER.KE.x.9063             CT---T---AC...--C---TC------CG-TT-CT----A-----AAACAT-C-T---A--GGAT-CT-AG------T---TA-CA---AC-T--GT--C-G-A-A-AA--TA-AG-TGCA-CGG---G-----------C-----T--C-----TCCA......... 8511
VER.KE.x.AGM155           -C---T-G---...--T---TC------CG-TT-TT----A------AACAT-T---GA--GG-T-TT--GC----AT----G-TA---G------T-GC---ACA-AT--TA-A--TGCA-AGC---G-----------T--C--T--T-----TCC-......... 8488
VER.DE.x.AGM3             GT---T--GA-...--C---TC------CG-TT-CT----G------AACAT-C-----A--GGAT-TT--GAT---CT---TA-TA---A------T-GC---A-A-A---TA----TGCA-AGC---G-----------T--C--T--TC-T--TCCA......... 8009
TAN.UG.x.TAN1             -T---A-TTC-...-G------CATC---G--G-CT--TCA---T-C----GG--T----GGGGAT-TTAT------AT---AC-TA--TTG--T--G--G-C--GGCT-A-TTGGGGATGCA-AGC---G-----------T--C-TT--T--G--TCC-......... 8465
SAB.SN.x.SAB1C            CT---T--GAC...--T---TCG------G--C-C-----A---T-TG-A-GG--G--GA--GC---C---G-TA---C----A-CA---TAGCT---G-TC-GC-A-TAA--A-AGGA--TC-T-GG-AGT----AA-A--G---G----TC-C--C-CC-TCC--TCT 8670
MND_1.GA.x.MNDGB1         -CCTTA--TC-...-G-GCA-G-------G--T-T--TTGG---G-TCAA--CT------GGGGAG-TT-CT----GTT---AA-TA----A--------TT-G--A--CT--TGGAA--C-A-A-G---GT--------------C---CG---TG-C-TT-CA......... 7970
DEB.CM.99.CM40            -T----ACTC-...--A------------G--C-GT-C-GA---T-C----GG--T-G----G-AG--TAC---A---CT-CT--AA--TTGC---------C-A-GTT--G-G-GAA---T-GA-G--AGT-GT-T-GA--------C--T--GCA-ACC-TT----CA 8112
DEB.CM.99.CM5             -T----ATTCA...--A------------G--C--T-C-GA---T-C----GG--T-G----G-AG-GTAT---A---C--CT---A--TTGC-T--G----CAA-GTT--GGG--AA---C-G--C--AGT-GT-T-GA--------C--T--GCA-ACC-TT----CA 8033
TAL.CM.01.8023            -C----AC--C...C-C----GC-----C--TT-GT-TGC----T-TA-A-GGC-T-G-A--GCAG----AG------CTA----AA-TTTGC-T-----T--AA-GTACA-TA-CAA-G--T-A--GTTGC-C---------C--C-T---TC--C-C---A------CT 7761
TAL.CM.00.266             -C----AC---...C-C----GC-----C--TT-GT-TGCC---T-TA-A-GGC-T-GGA--GCAG----AG------CTA-TC-TA-CTTAC-T--G--T--AA-GTACA--G--AA-G-GC-C--GTTGC-C---------C--C-T---TC-TC-----A------CT 8211
SUN.GA.98.L14             ---TTAAC......--C----G-------GG--C-GT-TC-A---T-TCTA--C--------ATGCAG-CC----TA----CATTA-AA---CAGCT-----TC--AG-TT-A-CA-ACAACAGA-CT----GA-GT-T----------AGGGTG--G--CCC-TCTG-TT-T 8689
WRC.CI.97.97CI14          -T---A-TGC-...-G-G-T-GC---C-CC-ATGGTTT---ATCTG-AAC-GGT------GGGGAT-GT-AG-TA-----ATAG-------A---------TT-AGCATG-T--A-GAA--GTA-T-GG-T--C----------T--C--TAAGC-TATA-C----...... 8500
WRC.CI.98.98CI04          -T---A---CA...-G-G-C-G----C-CC-GTGGTTT---ATTTG-AAC-GGT------GGGGAT-GC-T--TA-----ATAG-A-----CA---------T-AGGATGGC--A-GA---GCA-T-GG-T--C----------T--C--TAAGC-TATA-C----...... 8545
WRC.GM.05.Pbt_05GM_X02    -T---A-TCAA...-G-G-TTCA---C-CTCATGGTTT---ATTTG-AAC-GGT------GGG-TC-TT-AG--A-----TTAG-A-----AA-----G-TT-AGGATGCT--T--A---GTC-T-GG-T--C----------T-----TAAGC-TGT--C----...... 7441
LST.CD.88.524             -T-TTAAT........--T----G----A-GG--C-C--T-C----T-CCAA-----T--G-ATGCAG-GC-T---A-CAT---TA-AA--ATAC-------TC-AAGCTT-A-CA-ACAA-GT--AGTG-AGA-GT-TAG---C---AGGGTTC-CG-CCCCTCTG-TT-T 7549
LST.CD.88.447             -G-TCAAT........--T-----------AG----CTT-TCA---T-CCAA-----T--G-ATGCAG--C-A---A--AT---TA-TA---T-------G---AAGCTT-A-TG-GCAA-ATA-AGTA-A-A-GT-TAG-------AGGGTTC-TG-CCCCTCTG-TT-T 7546
LST.KE.x.lho7             ---TTAAC........--T----G------GG-T--CT--C-C---T-TACT---G-T--G-ATGCAG-GT-A--TA--CTT-TAA-CA----G------G--T-GAGCTT-A--A-ACAA-AT--AGTA-A-A-GT-TAG---G---AGGGTGC-C--CCCCTCTGTTT-T 8621
SYK.KE.x.SYK173           ------ACTCG...-------GC------G--T-GT-G--A---T-C-TT---C----GA--GGAT-TTATG--A------CT--A--GCT------GC-----AG-TT-AGTG-AGGA--CA-A--G-ATT-GTT-GGA--C---GTT--TA-T-T-CAGA-CC-TAC- 8360
SYK.KE.x.KE51             ------ATTCC...C------GC------G-TC-GT-G--A---T-CCA---CC-T--GA--GGAT-CT-TGCCA---C--CTA-A--GATA------G--C--AG-TT--CTTGGGGC-----A-GA-ATA-GTT-GGA--C-----T--TC-T-T-CAGA--C-G--G 8050
COL.CM.x.CGU1             -T---A-T-A-...-G--TTTCAG-C---GG-C--TTC--CATCT-TAA----G--CTC-ATGCTGC-TATG-T---------T--AA----AC-T-----TA--A-G-TA---A-AG--TGCA-A-GAG---C-TTCAGA-TTA-GGG-TTTCAGCA-A-AGGC-GGACT 7785
```

```
                                                                  Tat premature stop in HXB2                                       Tat end
H1B.FR.83.HXB2           CTCCCAACC.....................CCGAGGGGACCCGACAGGCCCGAAGGAATAGAAGAAGAAGGTGGAGAGAGAGACAGAGAC...AGATCCATTCGATTAGTGAACGGATCCTTGGCACTTATCTGGGACGATCTGCGGAGCCTGTGCCTCTTCAGCTACCA 8530
Rev exon 2               P  P  N  .  .  .  .  .  .  .  .  .  P  E  G  T  R  Q  A  R  R  N  R  R  R  R  W  R  E  R  Q  R  .  Q  I  H  S  I  S  E  R  I  L  G  T  Y  L  G  R  S  A  E  P  V  P  L  Q  L  P
Tat exon 2                S  Q  P  .  .  .  .  .  .  .  .  .  R  G  D  P  T  G  P  K  E  *  K  K  K  V  E  R  E  T  E  T  .  D  P  F  D  *
Env                       L  P  T  .  .  .  .  .  .  .  .  .  P  R  G  P  D  R  P  E  G  I  E  E  E  G  G  E  R  D  R  D  .  R  S  I  R  L  V  N  G  S  L  A  L  I  W  D  D  L  R  S  L  C  L  F  S  Y  H
H1A1.UG.85.U455          GC----T-.....................--AGA---T-T--G------A-G-A---C-----------------CA--G--A----...----G----C-------G-----T---A--GA--GC--------C-------A-------------------- 7960
H1B.US.90.WEAU160        ------G--....................-----------------------------C--------------------------------...------GGAA--------G-T----T---AA-----------T---C--A-----------------CT------ 8537
H1C.ET.86.ETH2220        A-----CA-....................-----------------T--G------C-----------------CA--G-----------...--------A--C------------T---A---A--T-T--------C----------------------T----- 7906
H1D.CD.84.84ZR085        ------G--....................------------------A------------------------CA--------GT...-C-----------G---------T--CA------------------A---A------------------- 8051
H1F1.BE.93.VI850         A------G-....................-----------------------------C------G---------CA--G--A----...------G-G--C-------C----T---A--T---GC------------A---A-------------------G 7807
H1G.SE.93.SE6165         AC--ACCA-....................-A-----A-------------------C------G------C--CA--G----G-...------AG-G---------G-----T---AC-G--------------C----------------------------- 7948
H1H.CF.90.056            G----G-A-....................--AC-----------------------C-------G------C--CA-------------...--------G-GA-----------------T---AC--G--G---------C--C-----------CA-------------G 7849
H1J.SE.93.SE7887         A------A-....................--A-C--A-G-----------G-----C-------G---------CA--G---AC-...------G------------------T-C-A--T---GC---------C-------A---------------T----- 7837
H1K.CM.96.MP535          A--------....................T--------G-A-----A----------C-------------------CA----A-A--...------AG-----------G---C-T---A--G---GC-----------------A-------------------G 7690
H101_AE.TH.90.CM240      TC--ATCAT....................-A--A--A---------A----------C-------G-------C--CA--G-------...------AG-G----------G-----T---A------GC-----------A---------------------------- 8083
H102_AG.NG.x.IBNG        AC--ACCA-....................-A-----A---------A------A---C-------G---------CA-----A-----...------G-G-----------G-----T---A------GC------------------------------------------ 8036
H1N.CM.95.YBF30          A-------A....................G-A--------A-------A---A--C----G--GC-T-------CA-------G-...------G-G-----------G-----T--CA--T---G-------G--C--C----A-----TGA-----CT------ 8081
H1O.BE.87.ANT70          AA--ATCA-....................-AAGA--A-G-A-GA-C---A-G-A---C--G--G--G---------AGA--G---GCC-...--G-GG--A-CC-CGCC-C-A--G-T----C----GT-G-ACACG--C--CA-A-CAA-AATAT-G-GG-CT------ 8578
H1O.CM.91.MVP5180        G-----CA-....................-G-CA--A-G-A--A-C---A-G-A---C--G--------------AG-------GCC-...-AG-GG-CAGCC--GCCACCA----T----CA--AGT-G-ACACG-----CA---CAA-AAT-T-G-GG-CT----- 8604
CPZ.CD.90.ANT            A----T---CAAAAC..............-A-CA--AT--A--GCA---A----A----AG------------AGA-A------GAT-...--G-GG-GGGCC--GCA-C----G-T---C-----CT-G----TG--C---AC----A-AAT--AG-GG-T------ 7989
CPZ.CM.01.SIVcpzCAM13    A-----G--....................-A-----A---A-GA-C--T--G--A--C--GC-G-------------CC--G---C-G-...------G-GA-------A-CA--G-G--G-C-G--C------------C--C--C-A----GT-A---GG-------G 8114
CPZ.CM.05.SIVcpzEK505    A-------G....................A-TC-------A--------A---A---C---------C-C-------CT----AC-G-...------G-G--C--------G-----T---A--T---GCG-----G--CT-CA---A---TCTA------CT------- 8016
CPZ.CM.05.SIVcpzMB66     A-----G--....................-G-----AT-GA--------A---A---C-------G-----------CC----AC-T-...------A-----C-------G-----T---A------GC----A----C--A---GA------T------CTT----- 8154
CPZ.CM.98.CAM3           A----G--A....................A-A---A---A--------A-G-----C----------C-T-------CC--G--A--G-...----------------C-------G-----T---A--T------------A-----C-----A---CCTG------T-T----- 7946
CPZ.GA.88.GAB2           A-----G-A....................A-AG-------A--AGT---A---------C-G--G--G---------CA----------G-...------TG-GA-------C-G-----TT--A--T--CTG---------CA-C----A----ACGA-----CT-----G 7917
CPZ.CM.98.CAM5           A----G--A....................A-A---A---A------GA-G-----C----------C----------CC--G-----G-...--------CGA----G--C-G---G-T---A--T---G-------A-----T----A---CCTGA-----TT------ 8252
CPZ.CM.05.SIVcpzMT145    A-----G--....................AATCA------A---G---T--G--A--C----A---G-------------AC-T-----G-...------A--CA----G-----G---G-T----C-T---G-------A-----T----A---CCT-AG----CT------ 8042
CPZ.TZ.00.TAN1           AC--A-G-G....................GA-CA--AT--A--ACA---A-G------C-C--G--G--------GC-------ACAT-...--G-GG-CG-CC-CGCCAGCA----T---CAGTA-CG-------G--C--CA---A---CCT-A---GG-TA------ 8186
CPZ.CM.05.SIVcpzLB7      TC---C-T-....................----------T--CTCA-----G------C------------------CA--G--C----...------AG-GA----GC--G------T---A------GCT---A----C---A--GA--------------CTT----- 8041
CPZ.US.85.CPZUS         T-------A....................A-TC---A---A--------A-------C--------GC-C----A-A-C-----AC-T-...------A-CGA-------C-G------T---A------G-------A-----T----A---CCTGA------CT------ 8538
CPZ.GA.88.GAB1           A-----GT-....................-A-----A--AA-GA---T--G--A--C--C---G-----C---CA-------AG-...------AG-GA--------G-A--G-G----C-----------------C--C--A-A----G-GA-T-GG-------- 8570
MON.CM.99.L1             A-----T--....................-ACGC-----AGCCA-C------C---C--GC-C--GC-----C--CG-CAG--ACTC-...------GCCAGC--ATC----A----TT--TA-CA-----------A-----CA---A---TGT------GG-CA------ 8490
MON.NG.x.NG1             ...ATCC-TACCCAAGTCCCTCGG.........-A---A-GAGC---A----A--C--G----GC-----------CA--G---GA--AGGTCRA-AG-CTATGCGAGAGGATTT-TAACCATTA--TGGACA---TGC-GAGACCTAA-AGT-TGGAC-TT-C-G-T 7118
GSN.CM.99.CN166          A-----CAA....................G-AGT--A--G--GTCC---A--C-C--CC-------GC-T---C-C-CA--G--AG-T-...--G---G-GA---ATC-A-CA----T--C-AGT--CT-G--------C---A---A-----GTGA-T-GG-T------- 8464
GSN.CM.99.CN71           AC----CAG....................-AAGT--A--G---TCC---A--C----C-AG-A----C------C-C-CA--G--AC-T-...--G---G-CA---ATC-C-CA----T--C-AGT--CT-G--------C---A----T---GTAA-T-GG-T------- 8476
MUS_1.CM.01.1085         A---G-C-G....................-A--C------A--AGC---A-------C-----G---C------AGCG--AG--------...-------G-GA---ATC---G-----T---CT-T--CT-----------C--C--C-A---AGTGA-T-GG--------- 8461
MUS_1.CM.01.CM1239       AC--A-C-T....................--AGA--C-G-A--GCC---A-------C-----G--GC-T---CAGC--C-G---GAC-...------G-G----ATC---------T----T-G--GT-T-------C--C--C-A---AGTGA---GGTTT------ 8430
MUS_2.CM.01.CM1246       AC--G-.......................................-AGTC--C-GA-C---GC--CAA-A-A-A-C--A-A--G-...GCTGGA-GAA-CGGCAG-C--A--------T-G--GT-T-CA--A--C--C-AC-A---AATGA-T-GG-CT------ 8538
MUS_2.CM.01.CM2500       AC--G-G--....................---GA------AAGA-C--------A--C---------G------C-CC----G---T-T-...------G-G--C-ATC-----------T----T-T-----------C--C--C-A---ATAG-G-GG-------- 8482
RCM.NG.x.NG411           TGTTATCATCAGCAGATCCCTATCCGCAAT......C-GGAG--ACA---AATCAA-GA--G----A----A--CAGTG----AG-C-GT...G--CA--AATAT-ACAACTGGCAGAGAGGA..................TATTTTTATA-CCTGA-AC--C--CCGAT 8243
RCM.GA.x.GAB1            TCTTATGTTCAGCAGATCCATATCCACAAC...-C---GGAACCGCAAA-TCC----GA-A--AG-----AC--T-GAGA---AG-T-G-.....AA--AGTACAACAATTGGCTGAGAGAA..................TATT-CT-GA-TCAA--GA--CA-CCGTT 8193
DEN.CD.x.CD1             -GA..........................--CCA-AAGG-A-CA-AT-AGCC--A-GA--G-----G--AAA-T-GAGAC-G--ACAGGCTC-AG-GGTGG-A--ATCC--GA----T----AGTA-C--T-----A-----CTCACAG---CTGACG-GG--T----T 8590
                          Tat Rev intron end  Tat Rev exon 2 start                                                                            Tat end
MAC.US.x.239             TCTTATTT-CAGCAGACCCATATCCAACAG......-AC--G-CACT---AACCA--GA--GCA-----A-A-AC-GTG----AG-C-GT...G-CAA--GCTCC-GGCCTTGGCAGATAGAATATA--CAT-TCCTGATC-GC-AACTGA-AC-------G-CT-GG-T 8951
Env                       S  Y  F  Q  Q  T  H  I  Q  Q  .  .  .  D  P  A  L  P  T  R  E  G  K  E  R  D  G  G  E  G  G  .  G  N  S  S  W  P  W  Q  I  E  Y  I  H  F  L  I  R  Q  L  I  R  L  L  T  W  L
Tat exon 2                                  P  I  S  N  R  .  .  .  T  R  H  C  Q  P  E  K  A  K  K  E  T  V  E  K  A  V  .  A  T  A  P  G  L  G  R  *
Rev exon 2                               N  P  Y  P  T  .  .  .  G  P  G  T  A  N  Q  R  R  Q  R  K  R  R  W  R  R  R  .  W  Q  Q  L  L  A  L  A  D  R  I  Y  S  F  P  D  P  P  T  D  T  P  L  D  L  A
H2A.GW.x.ALI             GGTTAC-T-CAACAGATCCATATCCACAAG......-AC-AG--ACA---AACCA--GG----AC-----AA-AC-TTG------AC-TT...G-GGA--GATTG-GGCCCTGGCCGAT-GCATATT-ACAT-TCCTGATC-AC-T-CTAGCTC-------G-T-GGG-T 8990
H2A.DE.x.BEN             GGTTATCT-CAACAGATCCATATCCACAAG......-AC-GG-GACA---A-CCAACGA--G-AC-----AA-AC-TCG---G-GACAGT...G-T-A-GACTTG-GGCCTTGGCC-ATAAACTATG-GCAG-TCCTGATC-AC-TACTGACTC-------G-T-GGG-T 9006
H2A.SN.x.ST              GCTTACTT-CAACAGATCCATATCCACAAG......-AC-GG--ACA---A-CCA--GA----AC-----AA-AC-TTG--A----C-TT...G--GA--A-T-G-GGCCCTGGCCGATAAGATATA-ACAT-TCCTGATC-GC-A-CTGA-TC-------G-A-AGA-T 8436
H2B.GH.86.D205           TCTTAT--TCAACAGATCCCTATCCGCAAG......-AC-GG-GACA---A-CCAACGA----AC-----AA----GTG--A--GAC-GG...GAC-A--GATCC-GGCCCTGGCAGATAGAATACA--CAT-TTCTACTG-GC-A-CTGAG-AA------A-TT-GG-T 8982
H2B.CI.x.EHO             TCTTAT--TCAACAGATCCCTATCCGCAAG......-AC-GG-GACA---A-CCAACGA----AC-----AA----GTG--A---AC--G...G-C-A--GATCT-GGCCCTGGCAGAT-GAATACA-CCA--TCCCGAT--GC-AACTGAG-GA------G-TT-GG-T 8968
H2G.CI.92.ABT96          TCTTATGTTCAGCAGATCCATATCCGCAAG......-AC-AG--ACW---AACCAA-GA----AT------A-A-AGTG--A---A--GT...G-C-A--GATCT-GGCCTTGGCAGATAGAATATA--CAT-TCCTGMTC-GC-A-CTAGG-AA------G-CT-GG-T 8379
H2U.FR.96.12034          TCTTATGTTGAACAGATCCCTATCCAACGG......-AC-AG--ACA---AACCAA-GG----AT-----AA--C-CTG-------C-GT...G-C-A--GATCT-GGCCTTGGCAGATAAGATATA-ACAT-TCCTGATC-AC-AACTGA-TC----G--G-CT-GG-T 8490
MAC.US.x.EMBL_3          TCTTATTT-CAGNNNACTCATACCCAACAG......-AC--G-CACT---AACCA--GA--GCA-------A-AC-GTG---AG-C-GT...G-CAA--GCTCC-GGCCTTGGCAGATAGAATATA--CAT-TCCTGATC-GC-AACTGA-AC-------G-CT-GG-T 8082
SMM.US.x.H9              GCTTRTGTTYAGCAGATCCCTATCYARACG......--C-AG--ACT---AACCAA-GA--G---------A-AC-GTG---R-----GT...G-CAA--GATCT-GGCCTTRGCAGATAGAATATA--CAT-TCCTGATC-GC-A-CTGA-AC-------G-CT-GG-T 8440
SMM.US.x.PGM53           G-TTATGTTCAGCAGATCCCTATCCACAAG......--C-AG--ACA---AACCAA-GA--G---------A-AC-GTG---------GT...G-CAA--GATCT-GGCCTTGGCAGATAGAATATA--CAT-TCCTGATC-GC-A-CTGA-TC-------G-CT-GG-T 8897
SMM.US.x.SIVsmH635F_L3   GCTTATGTTCAGCAGATCCCTATCCACAAG......-AC-AG--ACC---AACCAA-GA--G---------A-A--GTG---------GT...G-CAG--GATCT-GGCCTTGGCAGATAGAATATA--CAT-TCCTAATC-GC-AACTGA-AC-------G-CT-GG-T 8973
SMM.SL.92.SL92B          TCTTATCATCAGCAGATCCATATCCAGCGG......-AC-AG--ACT---A-CCAA-GA--G-AC------A-A--GTG--A--G---GT...G---A--GATTG-GGCCTTGGCAGATAGAATATA--CAT-TCCTGATC-GC-A-CTGA-AC-GA-A--G-CT-GG-T 8399
STM.US.x.STM             TCTTGTCGTCAGCAGATCCCTATCCACAAG......--C-AG--ACA---AACCAA-GA--G-AC-----AA----GTG---------GT...G--AT--ACTCC-GGCCTTGGCA-ATAGAATATAC-CAT-TCCTGATC-GC-A-CTAG-AC-------G-CT-GG-T 8602
MNE.US.x.MNE027          TCTTATTT-CAGCAGACCCATATCCGACAG......-AC-AA-CACT---AACCAA-GA--G-AC------A-AC-GTG-------C-GT...G-CAA--G-TCC-GGCCTTGGCAGATAGA-TATA--CAT-TCCTGATC-GC-AACTGA-AC-------G-CT-GG-T 8428
DRL.x.x.FAO              AAGGACG-ATACCGCCCAGAGACG......A-C-ACTAC-A-AGACA---A--CA--GA-A---G----AGCA---C----G-GAGTGG...-ACAT--GATC-GA-CCCTC-AAGC-AGGAT-CTCC-AAGCTTGGAGCGAAGAA......ACAG-AGGG-C--GGTT 8365
MND_2.x.x.5440           AAGGG-GA-TATCTCCGGCCCCAC............AAC-T-A-AC----A--CA--GA--G--G----AGCC---CTT---A-A-C-G...-ACAT--AGTC-GAGAGCTC-A--CAAGAAT-TAGG-AGCCTTGGA-G-CAGA-..........-AAG----GAG-TG 8301
MND_2.CM.98.CM16         G--TTC-AGGGAGACTATCTCCGTCCCCAC......AAC-T-A-ACA---A-GCAA-GA-AG--------AGCCC--CTC---A-A-C-G...--TA---AATC-GACAGCTC-AA-GTAGAATTTGGG-AGCCTTGGAGCAAAGA-..........-AAA----AG-TTG 8665
MND_2.GA.x.M14           AAGGG-GA-TATCTCCGACCTCGC............AAC-T-AGACA---A--C-A-GA-AG-----G--AGCTCA-CTT---AGA-C-G...-ACAT--AGTC-GAGAGCTC-AAGAAAGAATTTGGC-GACCTTGGAGG-CAGA-..........-AAA-A--GA-TTG 8594
OLC.CI.97.97CI12         ............TTTTCAGATACCCTC......CA--ACAAG--ACA---A--CAACGG-----G-----ACA-C---TTCTT-GAGAG-......A--T-CAACGAGTG----T-TA--ACA-AGTG--CAG--ACATTGAGAGTA-ATTG-A-AGA---ACTA-CA-- 8136
GRV.ET.x.GRI_677         ............CAGATCCATATCCACAGT......TC-G-G--AC----A--CAACGG--G--GGC---ACA---GTG----A--CAG-AGC--CAAAT-GAT-AG-T--C-G-A-GAG-CCT--ACACCT-C-AGGATCAACAACT-GTG-CT-AA-----AGAG-TG 8516
VER.KE.x.TYO1_patent     ............CAGATCCATATCCACCAA...GT---GAAG-GAC----A--CAACGCC--C--GCC---A-A--GTG------ACAG-AGG-TCAAAT-AGAG-CTTG-T-GAA-GA--CCAAGAGC-GG--TATGC-GT--ACCGC-TG-CTGACGAGGCT-A--.. 8640
VER.KE.x.9063            ............CAGATCCATATCCACCCG......T-GAAG-GAC----A-GCAACGC--G---GCC--AA-A--GTG----G------AACGACAG-TCAGAT-CTTG-C-GAA-GGA-CT-G-ACG-GACAAA-GAGAGGCA-TT-GAGCAAA-G---G-C-A-TTG 8663
VER.KE.x.AGM155          ............CAGATCCATATCCATCCG......T-GAAG-GACA---A--CAACGC--G---GCC--AA-A--GTG--AGA-C--G-AAA--CAAATC-ACGCATTA-C-GAA-GAA--T-GGGGACGAGACA-GAGGACCA-TT-GTGCA-G-AA--G-C-A-TTG 8640
VER.DE.x.AGM3            ............CAGATCCATATCCACCCG......T-GAAG-GACA---A--CAACGC-----GGCC---A-A--GTG------AGCG-AAG-ACAG-TCCGAGCCTTG-C-GAA-GAA-CT-GCACAGCAGA-TGGA-GAGCAACT-GTGCAAG-GA--G-C-A-TTG 8161
TAN.UG.x.TAN1            ............CAGATCAATATCCGCCTG......--GAGG--ACA---A--CAACGC--G--G-----ACAA---TTCCAG---CAG-...---GA--AGTCGCCGCCCTCA-TGAAAGAAT-TT-AT-GCCCA--AGAGGAG---T--AAGCGGAGGAG--AGCTTG 8614
SAB.SN.x.SAB1C           TCT-ATTAA...CAGATCCATATCCACCTG......-ACAAG-GACA---A--CAA-GA------GC----CA-C-GTG-GCTA-ACAG-GAC--CAGGTCATACACTTG-C-GA--GAA--CTTG.....................................--AC---T 8795
MND_1.GA.x.MNDGB1         ............CAGGACTGCCAGCAGAAC...CTATAC-G-A-AC----A--CAACGG---------A-CAACAGCTT---ACT--GA...GAGCA--ACTCCGAGAACTTG.........AAGGAAGAA-CCTTAA-CAGATCCTTGA-AGAGGA-C-G-C-AG-TT 8113
DEB.CM.99.CM40           AAG..........................GA-GA--A-GTGCTTC-A-T--C--A-GA--G--GCAG-A------GAGAC-G---G-C-...--G-GG-CAGC---GCC--GA----T---CAGT--C--------G-----C-AACAA---CTGACG-GGCT------ 8252
DEB.CM.99.CM5            AAG..........................GA-GT--A-GTGCTTCAA-T--G----GA--G---CA--A------AG-C-T--AGAG-...--G-GG-CGAC--CGCCC-GA----T----AGG--C-----------CA-CTCACAG--TCTGACG-GGCT------ 8173
TAL.CM.01.8023           A----G-G-....................AAC----A---A-GAGT---A----A-GA--GTAC---------A--CCC-----GAG-...--G-GG-GAAC--CGCCTC-A----T---T--CA----------G--C--CAAC-----CCT-AG--A---T------ 7907
TAL.CM.00.266            A----G-G-....................AAC----AT--A--AGC---A-G----GAC-CTAC---------A--CA--G---CT--...-AG-GG-ACA---CGCAAGCA----T----AGCA----T-----A------ACAG-T---AT-AGA-A--T------ 8357
SUN.GA.98.L14            G-TGA-CAGGATTGGCTCCAGGAAACCTGC...CCCAA---ACAGACAAA----A-GA------GAC--AGAA---A---AT-......TAC-TCAA-T-AGAGCAGTCC--GAA-GAA-CTTTG-CACCACCTTGGACAG-AGACT-GGACGAG-CT--G--AG--TC 8850
WRC.CI.97.97CI14         ............CAGAACCACCTCAACAAC...TT--A-GAG-GACA--AA--CA-CGA-----G-AT--AA-A---TTACA----TATAGAA-TCAA-GA-T-GCAGTAC-GGA--GAA-CT---A-AG---TCTTGA-CAGCTT-CTAACC-CA--A------ATTTG 8655
WRC.CI.98.98CI04         ............CAGAGCCACCTCAACAAC...TT--A-GAG-GACT---A--CA--GA--------C--AA-A---C-ACAGT-CTTTAGACGTCAG-GA-T-GCAGT-C--GAA-GAA-CT---A-A-C--T-TTGA-CAGCTT--TAAAC-T-T-A-CTGCAGT-TG 8700
WRC.GM.05.Pbt_05GM_X02   ............CAGAGTCATATCCTGAAC...CT--A-GAG-GACT---A--CAACGA----AC-----AA-A---TTTCATAGCCATAGAAGA---AGACT-GCAGT-C--GAA-GAA-C-CTGA-AC---TCTGGA-CAGCTT-GCAACC-TA--T-A----ATTTG 7596
LST.CD.88.524            G-TGA-CAGGACTTCAAGTGGGAGAACGGA...GA-A-C-AA--ACA---A--CAA-GA--G---GAC--AGAA---C-----A-ACAT-TAC-TCAA-T-AGAGCAGTGC--GAA-GAA-CTTTCAGGCCACC-TGGA-CG--GACT-GAACGAG-CT--ACAGG--TC 7716
LST.CD.88.447            G-TGA-CAGGACTGCAAGTGGGAGAAAAGA...GA-AAC-AA--ACA---A--CAA-GA--G---GAT--AGAA---C-A---A-ACAT-TAC-TCAA-T-AGA-CAGTAC--GAA-GAA-CTT-CAC-CCACC-TGGA-CG--GACT-GAGCGAG-CT--ACAAG--TC 7713
LST.KE.x.lho7            A-TGA-CAGGACTACAAGTGGGAGAAAGAA...GA-AAC-AA--ACA---A--CA--GA-----GA---A-C---T-C---G-CTAT-TAC-TCAA-T-AGAGCAGTGC--GAA-GAA-CTT-CAGGCCACT-TGGA-CG--GACT-GAACGAG-CT--ACAAG--TC 8788
SYK.KE.x.SYK173          .............................CA---C-GGA-GGAC--A-GCAA-CCC-C---C----AA-A--GC--T-GAGACAGA...GA-GGAC-CAACG-CTCT-C-TTCAG-AGA-A-TC-T-GA--C-ATC-T--GAAGCAGGACAG-AGC-TT-GAGGAC 8494
SYK.KE.x.KE51            -C---T---....................CTCTAC-A-A-GGCC--A-C---GCC--------G-ATA-A-ACAGATACAGACA-T...TACAAGCCCA-CG-CTCT-CGTTCAG-AGA-A-T--T-GA-AC-ATCG--CGCAGAAGC-CAGAAAC-TT-GA--AG 8193
COL.CM.x.CGU1            AATGTGT-TTCA.................-A--TCC-TGG-C-AGAC-GG--GT-GGA-C-GCC---CAAC-C---AGA-ATACACAGA...GAGCAAGAAAACAGCAGC-G-A---TT-G-...........................................-G 7889
```

```
                                                                    Rev end (TAA) in some lineages
H1B.FR.83.HXB2            CCGCTTGAGAGACTTACTCTTGATTGTAACGAGGATTGTGGAACTTCTGGGA.....................CGCAGGGGGTGGGAAGCCCTCAAATATTGG...............................TGGAATCTCCTACAGTATTGGAGTCAGGAACTAA 8646
Rev exon 2                _P__L__E__R__L__T__L__D__C__N__E__D__C__G__T__S__G_ .  .  .  .  .  .  . _T__Q__G__V__G__S__P__Q__I__L_ .  .  .  .  .  .  .  .  .  .  V  E  S  P  T  V  L  E  S  G  T  K
Env                       __R__L__R__D__L__L__L__I__V__T__R__I__V__E__L__L__G_ .  .  .  .  .  .  . _R__R__G__W__E__A__L__K__Y__W_ .  .  .  .  .  .  .  .  .  . _W__N__L__L__Q__Y__W__S__Q__E__L_
H1A1.UG.85.U455           ---A-----------CGC----------G----AGC---------------CGCAGCAGCCTCAAGGGACTGA-ACT---------G-G-----------CT-.............................--------T--G-T-------G---G-------- 8097
H1B.US.90.WEAU160         --------TC----------------C--A---------------------.....................-------------------T-----------T..............................------------G----------C--------- 8653
H1C.ET.86.ETH2220         ----------------A---------C-G------CA---------------CGCAGCAGTCTCAAGGGACTA-AG------------A----T------CT-.............................G-A-GC--TG-G---------G---T---G---- 8043
H1D.CD.84.84ZR085         ------------A---A---A----C-G-------------------------.....................------------------------------CT-.............................------------G--A----------GA-----G- 8167
H1F1.BE.93.VI850          --A-----------CA-A--A----C-G-----------CAGGGGCCTG.....................A-G---------------T---------CT-.............................G--------ACG-G---------------------- 7923
H1G.SE.93.SE6165          --------------CCA-----------G------CA---------------CGCAGCAGCCTCAAGGGACTGA-ACT-----------GT--T-------T-.............................--------T--GTT-------G-CAG-------- 8085
H1H.CF.90.056             --T-------------------A------GT-----C--------------G.....................A-A------A-----------------CTC.............................--------T-----A--C---G-A----------- 7965
H1J.SE.93.SE7887          ---A-----------CG---------C-G------C-----G-AC-------.....................-T------------GAT----------CT-.............................GT---C--TG--TG-------G-G---------- 7953
H1K.CM.96.MP535           --AA------A-----A----A----G--------CC-----AGGGGACTG.....................A-GG----T--------------------CT-.............................-----C--TG------------------------- 7806
H101_AE.TH.90.CM240       ---G------------AC--------C-G------CG---------------CACAGCAGTCTCAAGGGACTGA-AC------------G----------CT-.............................G-------T--GTTA------G-C---------- 8220
H102_AG.NG.x.IBNG         ---A------------A---------C-G------C----------------CACAACTGTCTCAAGGGACTGA-ACT--------G--------------CT-.............................--------TA--TCA--C---GT----------- 8173
H1N.CM.95.YBF30           --------C-----C---------AC-G-G-----C-C-------C------CAGAGTCTCAGCAGGGGACTG-AACTACT-AAT---CT-AGA-C-C-C-T-.............................---GGAA-A--TGCA------G-AA-A--GT--- 8218
H1O.BE.87.ANT70           --T------CA-----GCA-CAGGGA-CCA--A-G-GA-CAGCTA----A-GCTTGGACTGTGGATCCTAGGG-AG-A-ATAATTA-T-TTTG--G-AT---T.............................GCAGC-G-AAC---A--C---CTA--A---T-GC 8715
H1O.CM.91.MVP5180         --T------CA-----A-A-CAGGGA-CCG----C-GA-C--CTAC------CTGGGACTGTGGATCCTGGGA-AA-A-ACAATT-----TTGT-G-CT---T.............................G-AGC-G-AA-G--A------CTA--A---T-G- 8741
CPZ.CD.90.ANT             GAT--GC---AC--GT-------ACC-TTG-GCAG-CC-CC---AC--CT-CAGA.....................ATT-CTTTCA-ACTGTG-AA-C-TCTGGA-.............................AAC------AGCACCCTC---...ACAAT-A-C- 8105
CPZ.CM.01.SIVcpzCAM13     GA------C-AG---GC--GC-----TTG---ACAGT-----AAGT-----CAACTGTTAATCAACATCCTCA-ACT-ATACA---GAAA--G-CTCTACT-.............................A--GGGA-AA-C--A------G-AAGA---T--- 8251
CPZ.CM.05.SIVcpzEK505     --------C-----GT----CA--CC-G-G-----CCT-----------A--CAGAACATCCACAAGGGACTG-AACTACTTAAT---CT-AGA-T-----T-.............................---GGAA-AA--GCT--C---G-AAG---G---- 8153
CPZ.CM.05.SIVcpzMB66      ---A--------------------A--GCT-----CAT-------GG-----CAGACCCTCCTCAAGGGACTGA-AC----AC----------A-T-C-CCT-.............................A--GGCA-T-----A------G-A-----G---- 8291
CPZ.CM.98.CAM3            ---A----------------A-----AC-GGG-----CAT-----AAC-----CAGAGCCTCAACAAGGGACTG-AACA-CT-A--A-CTT-AG--G---CCT-.............................---GGAG-GA--ACC--C---G-AAG---G---C 8083
CPZ.GA.88.GAB2            --T---------T-GT----C-G-CC-TTG---CC-GC------AAT-----CAACAGGTCCTGAGAGGACTGA-ACTACTCA----GCTT--G-GCC-GCT-.............................AAAGGGA-AGG---A------CTG--A------- 8054
CPZ.CM.98.CAM5            ---A----T-----------CA--CC-GTG-----CA---C--A--------CAGAGCATCAACAAGGGACTG-AACA-TT-AA---CTTTAGT-C---CCTA.............................AA-GGAG-A---ACA--C---G-AAGA--GA--C 8389
CPZ.CM.05.SIVcpzMT145     --AA----------GTGCA-CTC-GA-TTG---AC-AC----GA-A------CAATACAGCCTCAGGGGAGTA-AACAAATAG-TACTCT---AC-CC-ACT-.............................A-AGGGAC---G----------CAACA---A-C- 8179
CPZ.TZ.00.TAN1            GAC---TCA-A----CA---G---CC-CTG--TC-GCC--C--GCA---AA-CAGGGGATAATCAGCTTGGCA-A---CCTAGTAAT--TG-AT-G-AC-ATCATAGTAGGAGTTAGACAGATC.............ATTG-GTGGAGCAGTA--ACTTA-GCTAGCT--- 8344
CPZ.CM.05.SIVcpzLB7       ---A----C-------G----A--A-CG-G-----CAT----GA--G-----CAATACACTCTGAAGGGACTGA-ACT--T-------------ACT----CT-.............................CA-GGAA-T------------G-AA----G---- 8178
CPZ.US.85.CPZUS           ---A---GAG--------A--A---C-G-G-----CA--AC--A-C------CAGAACATCAACAAGGGACTG-AACT-TT-AAT---CT-AGAGC-CGC--C.............................---GGGG-AA-CGCT--C---GCAAG---G---- 8675
CPZ.GA.88.GAB1            GA------C-AG---GC--GC-A---GTG---ACAAC-CA-GACGT-----CATCTAATACTTCACAGCCTCA-ACT-CTAC----GAGA--GTGCCTGCT-.............................G--GGAA-TA-T--A------G-AA----G---- 8707
MON.CM.99.L1              GAT-----A-----C-G-A--AG-GA-CTACC----CC-CC-GAGAG--TCCCAGAGGCTGCCACCCCTCCTG-A--TAC-CCT-CTCCAA--GTGGG-AA-CCTTAGACGCTTGCTAGCCTACTGCCAGTAT.....................G---TC--A-----CC 8639
MON.NG.x.NG1              ATTGAG-GACTC-GGCT-GAC---CTACCG---CC-CCA-AG-G-G-GCA-CCACCTGATCCCTCTTCTCCGCGA-CT-T-CC-AC-GCTT-GGG--GCCA-TAGCCGCCTTCTT.........................GCCTA---G------G--CTC--A--GY-CC 7264
GSN.CM.99.CN166           GAT----GCCAG----GC--G--C-C-CCGTG----C-G-A-GT-C---T-GGAACAGCTGCAGAAAGCAATC-AG-AA-CAATCC-GCAAGCG-G-C-GCTCAGGGAAGTAGCAGCCAGAGTG.............ATAGCCTA-A-TAGC---G---TC--A-----CC 8622
GSN.CM.99.CN71            GAT----GCCAC----G---G--C-C-CCA------C-GCA-GT-C---T-GGAGCAG.............CTG-AG-A--C--TCC--CAAGCG-G-C-GCT-GGGGAAGTGGCAGCCAGAGCA.............GCAGCCTATA-TAGC---G---TC--A-----CC 8622
MUS_1.CM.01.1085          GAC---A-T-AG--C-GC--CC-CCC-TTG-T--C-AC-TC-GAGAGCA-TCCACCTGATACAGAAGGGGTCAATA--ATT----TTCA-G----GGGCACT-ATTGAGCCTCTCAGGGAAGCCTTGCAAAGAGCCCTCGC-TG-T-----C--G---TC--A--G--CC 8631
MUS_1.CM.01.CM1239        GACA--AGTCAG--C-G-A-CA-CCC-TCGCT----AC-CC-GAGGGCA-CTCACCTGATACAGAAGTGGTCAGTAC-ACT----ATC-TGA-T-GGG-ACAAGCCCGCCGCTGCAGTGAACTCTTAGCGCGCCTC-TAGCATA-A-T--A---G---TC--A-----CC 8600
MUS_2.CM.01.CM1246        GAT-----CCAG--CCGC--C--ACC-CTG-T--T-GACTC--AGAA-CTACCACCACCTGACAAAGGGGCTAGTA-ATCC----ACTCTG--T----CC-TTAGGGAAGCCTTAGGGCGCTTCATCGCACAC.............GC---A---G---TC--AA----CC 8696
MUS_2.CM.01.CM2500        GAT-------AG--CT----CA-ACC-TTG-T--C-GA-CC-G--GA-A-TCCATCACCTCCAGAAGTGGAGCAT-TCTCT----CTCAGGGCG-G-GCAACCAGCACCCACCTTCTGGAAGTCCTTAGACGCTTGCTTGCATAT--T--A---G--CTC--A---T-CC 8652
RCM.NG.x.NG411            AGAGA-CCTTTG-AGGAGGC-CTA-CA--TCT-C-GCAACTC---CTCA-TGATCTACCAGAGCCTCCAGTCAAT-CCTTCTC-AATCCTT--TC-T-TGC--TGGATAGGAGCT..........................AAG-----AG--G--T-G--A--GT-C- 8386
RCM.GA.x.GAB1             GA--AG--TTTGGACG-AGC-ATCGCAG-TTT-C-GAAGCTGCAGCTCAAT-ATCTTCCAGAGCCTCCGGTGGATTTTA-CT...........................................................AAGA---A--G--T-G--A--GT-C- 8303
DEN.CD.x.CD1              GAC-------A-----G-G--TC-GA-T-G--CCGCCTGTAGT-GGA-TT-GGACTTTGTA.............................-AT-GGC-CA-CTTAATTTTGCTTGGC.......................TA-T-C--A---G--CTC--A-----CC 8706
                                                                                                                   Rev end                          Nef start
MAC.US.x.239              ATT-AGC-ACTG-AG-AC----C-ATCG-GAGTATACCA-ATC--C-AACC-ATACTCCAGAGGCTCTCTGCGAC-CTACA-A--ATTCGAGAAGTCCTCA--ACTGAA..........................CT--CCTA------A---G--T-GAGCT-TT-CC 9094
Env                       __F__S__N__C__R__T__L__L__S__R__V__Y__Q__I__L__Q__P__I__L__Q__R__L__S__A__T__L__Q__R__I__R__E__V__L__R__T__E_ .  .  .  .  .  .  .  . _L__T__Y__L__Q__Y__G__W__S__Y__F_
Nef                                                                                                                                                                  _M__G__G__A__I__S_
Rev exon 2                 I  Q  Q  L  Q  N  L  A  I  E  S  I  P  D  P  P  T  N  T  P  E  A  L  C  D  P  T  E  D  S  R  S  P  Q  D  *
H2A.GW.x.ALI              GTA-AGC-TCTG-AGGGA---AC-ATCC-G--TCTCCCC-ATC--C-AACCGATCTTCCGGAGTCTTCAGAGAGCGCT-ACAACAATCAGGGA-TGGCTGA-ACTTAAA..........................GCAGCCTA---G------G--T-CG--TGGA-CC 9133
H2A.DE.x.BEN              ATA-AAC-TCTG-AG-GA---AC-ATCC-A--ACTCCCC-............ACCCGCCGACTGATCTCCCAGA-TCTAACAGCAATCAGGGA-TGGCTGA-ACTTAAG..........................GC-GCC-AA--G--A---G--T-CG--TGGA-CC 9137
H2A.SN.x.ST               ATA-AAC-TCTG-AGGGA---AC-ATCC-G---CT-CCA-ACC--C-AACT-ATCTCCCAGAGTCTTCGGAGAGCATT-ACAGCA-TCAGAGA-TGGCTGA-ATTTAAC..........................ACAGCCTA---G--A---G--G-CG--TGGA-CC 8579
H2B.GH.86.D205            ATA-AACG-CTG-AG-AC---AC-GT-G-A--CCT-CCAA.......................ATCCTCCAC-AA-TCTCAACCA-TCT--AACC-CTCA--CTCCCG..........................GTTGCGTA---C------G---TCAGCTGGT-CC 9101
H2B.CI.x.EHO              ATA-AGCG-CTG-AG-AC---AT-ATC--A--CCT-CCAAACC--C-AACC-GTACTCCAACCGCTCAGGCTT-CACCT-CA..............................................TA---C-GA---G---TCAGCTGGT-CC 9078
H2G.CI.92.ABT96           ATA-AGC-ACTG-AG-GC---AC-ATCC-G--TCY-CCA-ACT--C-AACCGCTGTTCCAAAGAATCTCCAGGAC-CTCCAAGCAATTAGAGAAC-TCTCC--CTAGAA..........................GCAGC-TATT-CAGC---G--TTC-GCTGG--CC 8522
H2U.FR.96.12034           ATA-AGC--CTG-AGGGA----C-ATCC-G--TTTGCCA-AGC--C-A-CCGTTGTTC....................CA-A-CATCAGAGAACGTCTGCACCTTGAA..........................ATAGC-TA------A---G--T-G---T-TT-C- 8612
MAC.US.x.EMBL_3           ATT-AGC-ACTG-AG-AC----C-ATCG-GAGCATACCA-ATC--C-AACC-ATACTCCAGAGGCTCTCTGCGAC-CTAC-AA--ATTCGAGAAGTCCTCA--ACTGAACTGACC.........................TA------A---G--T-GAGCT-TT-CC 8225
SMM.US.x.H9               ATT-AGC--TTG-AGGGAT-G-C-AT-G-G--NCTGCCAAATC--C-A-CC-GTGCTCCAGAGCCTCTCAAGGACGCT-CA-A-A-CCCGTGAAGTCATCA-AGTTGAGATAACC.........................TA---------G--T-G-GTT-C--CC 8583
SMM.US.x.PGM53            GTT-AAC-ACTG-AGGAC--ATC-CT-G-GA-TATACCA-ATC--C-AACC-GTGTTCCAGAGGCTC.........TTACA-A-AATCCGAGAAGTCGTCA--ACTGAG..........................ATATCCTA---G--A---G--T-C-GCTGGT-CC 9031
SMM.US.x.SIVsmH635F_L3    ATT-AGC--CTG-AGGGAT-G-C-AT-G-G--TCTACCA-ATC--C-AACC-GTGCTCCAGAGACTCTCAAGGACGCT-CAAA-A-TTCGTGAAGTCATCA-AATTGAA..........................ATA-CCTA------A---G--T-GAGCT-TT-CC 9116
SMM.SL.92.SL92B           ATA-AAC-ATTTGACT-G---AGCATCC-GAGCCTACCA-A-C--C-AACA-CTGTGCCAGAGACTCTCAGAGAT---TCA-CC-ATTAGAGAGCTTGTCA-AAGAGAAGCAGGC.........................TATA-C-GT---G--T-GA-TT-CT-C- 8542
STM.US.x.STM              GTA-AAC-ACTT-AGGGCT-GTC-ATCC-G--TCTACCA-ACC--C-ACCC-ACCTTCCAAAGGATCTCCAGGATACT-CA-A-AATTAGAGAAGTCGTCA-ACTAGGA..........................GCAGCCTATT--------G--T-CATCTGGA-CC 8745
MNE.US.x.MNE027           ATT-AGC-ACTG-AG-AC----C-ATCG-GAGCATACCA-ATC--C-AACC-ATATTCCAGAGATTCTCCACGAC-CTACA-A-A-TCCGAGAAGTCCTCA--ACTGAA..........................CTA-CCTA------A---G--T-GAGCT-CT-CC 8571
DRL.x.x.FAO               GAAGGAATCGAGAGG-TA-A-CTGGC-G-A--ACG-GAAA-C-G----T-A-TATGGGTGGTCAGAA..................................................----A-G-AGCGG-A--AGGCAT-TGC. 8459
MND_2.x.x.5440            G-TGAAA--GTCGACG--T-AC---TGGCT--A--ACC-CC--GCAG-AATTGAATATGGGTGGCAAGAG...............................................---AA-GCAGCGG-AGCAGCATT-TATC 8399
MND_2.CM.98.CM16          G-TGAA--CTTCAAGGGG--ACG-GTGGCT--A--A-C-AC--GCAG-AATTGAATATGGGTGGCAAGAG...............................................---AA-ACAGCAG-AG-AA-AAT-T-T- 8763
MND_2.GA.x.M14            GTTGAA-GAGTCAAC-A-T-AC---TGGCT--A--A-C--C-GGCAG-AATTGAATATGGGTGGCAAGAG...............................................---AA-GCAGCGG-AGCAAG-AT-TGCC 8692
OLC.CI.97.97CI12          --TG-G--AG-GAG--G-GAGAG-CACTCGT......................................................................................TGGATC...CA--AGAGCTTG-A-C 8189
GRV.ET.x.GRI_677          -A--------ATAAGGACT-G-TG-TAC-AC-TCTGCC--ACC--C--CAT-TTCATCAGGACAGCAGTGGGA............................................TA---G------G--CTC---C----CC 8617
VER.KE.x.TYO1_patent      .................ACT-G-C-ATAC-AC--TTGCC--ACC--C--CATTCAGCTTAGGAAAGCTTTTCAA............................................TA---G--A---G--CTCGCA-----C- 8725
VER.KE.x.9063             G-T--C--CGTT-AGCAC--G-C-ATAC-AC-CCT-CC--ACC--C--AATTCATCTTAGAGGCGCTTGGCAG............................................TA-A-C--A---G--CT-GG------C- 8764
VER.KE.x.AGM155           GTT-----C-CT-AGCA---G-T-ACAC-AC--CTGCC--ACC--C--TCTCAAGCTTAGAAGCGCCTGGCAG............................................TA-T----A---G--CT-GG---G--C- 8741
VER.DE.x.AGM3             GT---C--TCAG-AGCA---G-C-ATAC-AC--TTGCC--ACC--C--A-TTCATCTTAGGAGCGCTTTCCAG............................................TA-A----A---G--CT-GG------C- 8262
TAN.UG.x.TAN1             G--GCA-CATTTGACCAAT-G-TGCT-G--A-TC-GCAGTTGGT-AT--A--CTTTACCAGATCCTCCGCAGGA--CTCAC-ACTCTTCTT-AACTGCT-A-ACAGGAG..........................--CC-ATA-A-T------G--T-G---C--T-C- 8757
SAB.SN.x.SAB1C            GT-TCAC---TTGA-CACT-G-C-CAGG-AT...T-GACCAGTTGGT-CTCGACAATCTTCAGCAACCTCCAT---T-CCTCCA---CAT--AAC-GAGAACCAGACAGCTAACAGCT.........................-AT--TG-A---G--T-G--A--G--C- 8938
MND_1.GA.x.MNDGB1          TGC-AG-GAGTGACC-A-T-G-C--TGGCT--AA-A-C--AG-GCAGCAATTGAATATGGGTTC...................................CTCGCAGTCCAAGAAGCG.........ATCAGAAGC-T-GG-TCG----TCGTCA-CTT-GC--C-A-T- 8238
DEB.CM.99.CM40            GAC---TC-CA---GCA-T-G-C-GA-T-GA--AGGCT-T---T-A---CA-GAAGGAATATATAGACTGTCATTTGCTCTT---C--AGA--G-G-CTC-T-GGGGAACGCTTTAGAGTGCTT............G--GGGTA-T---------G--T-GA-A---T-C- 8410
DEB.CM.99.CM5             AAC------GA---GCGC--G-C-GA-T-GA--AGGCT-T--G--C---CA-GAAACAATATATAGACTTTCATTTGCATTT---C--AGA--G-G-CTC-T-GGGGAACATTTAAGAGTGGTC............---GG-TATTGT--A---G--T-GA-A---T-C- 8331
TAL.CM.01.8023            GAC-------A-----G-T-G-C-GAGTTG-GACC-CC--A-G-GC---AA-GCGACCACCCAACAGGCAGCG-AAGCCACA-AC-CCTA---G-G-AT-AAATTCACAGCAGCAGTA.....................GCAACAA-C------G--TTCAGT---A-CC 8056
TAL.CM.00.266             GAC---AGCCA-----GCT-G-C-GA-TTG-GACC---C-A--GC-GC-AA-GCCAAGAGCCAACAGGTAGCG-AAGCCACA-AC-CCTT---ACG-CTCCACCTGACAGCAGCAGCAGCGACTCTCCAGTAT.....................G--TTCAGT---A-CC 8506
SUN.GA.98.L14             T-TG--AGTGACTC-T---AA-TGGC-G-AAGCTTGG-GAAT-G-C--A-C-CAAAACATCTACCACCTGCTAA-TTTCTT----C-CCTA--T-CGAC--CCTTCCAT...CATGGGCAACGCCTTTGGCAGACCCTC-GAGGTTGG-T-GG--C-CAC-TTATT-AGG 9017
WRC.CI.97.97CI14          --AACACCTCCTTCACA-TG--TGGAGTCT-CTTTCGTGCCTC--GTG-T-GATAAACGATGGGTGG........................AGTCTTCTCAAAGACATC.....................................AG-G-A--AGCAT-TGG-G 8759
WRC.CI.98.98CI04          --AAACCCTCCTTCA-A-AG--TGGAGTCTCCTTGCGTGCCTCT-CTG-T-GATCCAGCATGGGTGG........................AATCTTCTCAAAAGAATC.....................................AG-G-AGCAGC-C-TGC-T 8804
WRC.GM.05.Pbt_05GM_X02    --TGAGCCTCCTGCAG--AAGCTACAGCCTTCT--GCTATCTGTGGA-A---ATACAGAATGGGTGG........................AGTCTTCTCAAAAGACTC.....................................-G-G-AGCAGC-T-TGCTG 7700
LST.CD.88.524             A-TG--TGTGACTC-C--GAA-TGGC-G-A-GC-TGG-GAAT---G--CCTCACCCTCGTGTGGCAATTCCTCACTT--CT-G-AC-TATAG-G-TCCT--CTTACCAC...CATGGGCAACATCTTTGGCAAACCCTC-GCAGAT-C-T-GTGGA----GCTC-GAA-- 7883
LST.CD.88.447             A-TG--TGTGACTC-C--GAA-TGGA-C-A-GCCT---GAATT-----CCTCAGCCTTGTGTGGCAGTTCCTTTCGT--CTAG-CC-TTTAG-G-TCCT--TCTTCCAA...CATGGGCAACATCTTCGGAAAGCC--C-GCAGA-GGGT-GTGGAA-ACGCTC-GAAG- 7880
LST.KE.x.lho7             G-TG--AGTGACTC-C--GAA-TGGC-C-A-GAAGGG-GAAT---C--CCT-TCACTAGTGTGGCAGAGCCTCA-TT--CT----C-TTTGT-G-TCCT--TCTTCCAA...AATGGGCAACGCCTTTGGCAGACCA-C-GCAGATGGAT-GTGGAA-ACGCTC-GAAG- 8955
SYK.KE.x.SYK173           AGT--GC--CAG---CAGA-CTC-GA-T-GACA-C--ACTATCACCTG----.................................................T--A-CAGC---G--TTCA-C--G--C- 8574
SYK.KE.x.KE51             AGTT-C-GAGAG--GCAGA-CAC-GA-T-GAG-AC----AATTGCCTG----.................................................T--A-CAGT---G--GTCACA--G--C- 8273
COL.CM.x.CGU1             ATTGA--GA---G-GGTGGAGC-GCT-GTT-C-TCA-TGCAGCAA-GG-ATCCTTACT.................................GTCCTGCCTCAACATCCGCCAACAACCTCA............GAC--GAGGAGGAGTGC-----.....---ACT--TC 8008
```


PLV Complete Genomes

```
                              Rev end                                                                                                                       Env gp41, gp160 end
H1B.FR.83.HXB2                AGAATAGTGCTGTTAGCTTGCTCAATGCCACAGCCATAGCAGTAGCTGAGGGGACAGATAGGGTTATAGAAGTAGTA.........CAAGGAGCTTGTAGAGCTATTCGCCACATACCTAGAAGAATAAGACAGGGCTTGGAAAGG......ATTTTGCTATAAG..... 8796
Rev exon 2                       E   *
Env                           K  N  S  A  V  S  L  L  N  A  T  A  I  A  V  A  E  G  T  D  R  V  I  E  V  V  .  .  .  Q  G  A  C  R  A  I  R  H  I  P  R  R  I  R  Q  G  L  E  R  .  .  I  L  L  *
H1A1.UG.85.U455               -A-T-------A---CT-----TG-C--TGT---AG-----------G-T---T---------------A---G-.........---AC-AT-G-----------T-A----------------C-----A-----A----------.....GC---------C..... 8247
H1B.US.90.WEAU160             -----------------T-----T-----------T-----------------------------------A----G..........---A--A---------------T-----------------------------------A------......GC-------------..... 8803
H1C.ET.86.ETH2220             -A--G------A---ATC----T---A-------A----T-----G---A--A--------AT--------T--A-..........---A--AT---G------T-CT--A----------------------------------GCA......GC-----A----A..... 8193
H1D.CD.84.84ZR085             -----------A--------G-TG---A-----A--------------------------A--------CA----..........AGGA---------A----G-G-TT--------C-C---------------T--A------......GC----------A..... 8317
H1F1.BE.93.VI850              -----------A-C---C--T-T---A-------A----T---------A-----------AA----------TT-G..........---A-----G--------G---T-A--------------------GCA-------......GC------G---A..... 8073
H1G.SE.93.SE6165              -----------A----T-----TG--A-AGT---A-----------A-TT-----------------------C-..........---A-----------------C-T-A--------C---G--------A-----------A......GC----------A..... 8235
H1H.CF.90.056                 -----------A--GAT-----T--CA-------A-----------------A------G--A------T-A---G..........---A------G---------T-------------------------T-----A......-GC---------TA.... 8116
H1J.SE.93.SE7887              -----------A----T-----T---A-------A--------------A--A--------AA-C------A---C-..........---A-----T----------TT--------------------T--A-----A......GC----------A..... 8103
H1K.CM.96.MP535               ----------CA----------T---A-------A-------------GA--A--------AA--------A---G-..........---A-----T-------T-A-TT--------------------TC-C-----A......GC---A------A..... 7956
H101_AE.TH.90.CM240           -A-T-------A--TCT-----TG----T-----A-------C---G-G-T-------C-----------------C-..........------------G-----C----T-------------C-----------A-------......-C---------C..... 8370
H102_AG.NG.x.IBNG             -----------A--AT------T---A-A-T---A---T-------A-CT----------A-C-------A---G-..........---A---TGG---------------A--------------C-----------C------......GC---------C..... 8323
H1N.CM.95.YBF30               G-G--------A-C---------T---A-A-----T--T-T------A--A--A---------A--------T---C-..........---A--ATAG-A--G-GA--ATTA---------------C-----A---C-A-----A......GCAC--A--------..... 8368
H1O.BE.87.ANT70               -----------ACA--------AG-CA-ACTT--AG-G-------CA-TT----T--CG-CA-A--C-C--GGA--..........---A--ATAG-A-C--GA-----TA---C--A--G-----T-----------A-----A......-G---AT-G---C..... 8865
H1O.CM.91.MVP5180             -A---------ACA-A-C----TG--A-T-TT--AG-GT----T--CA-TT----T--CG-CA-C--CTT--GTC--..........---A--ATAG-ACA--GAT-C-TT-----C--A-------T-----A--TGCA-----A......--C--AG-G---C..... 8891
CPZ.CD.90.ANT                 GA-C-GAGATCA---AGAACA-TG-CAGACTT--T--TTGG----GG--AAAA--------CA-AC-TCT--CTC-C..........---ACTATAGTC---ATC--AA-GG-AG-------GC-C--C-----A--G------TT......GCA--AAAT---AC.... 8256
CPZ.CM.01.SIVcpzCAM13         GA-C-------ACA-----AT-AG----A-----T--T-------GA--A--C--------AA-C---AG--C---G..........---AT--TA-TC---ATC--AG--A-------CC-GC-C--------A---C-A--G--A......-C---A--T---ATCTTG 8406
CPZ.CM.05.SIVcpzEK505         -A-T-------A-A-A------TG--A-T-----TG-T-------A--A-----------AA-C-----GC----G..........---A--ATAG-AC-G-GA--A-TA-----------G-----C-----A--TC-A-----A......GCAC-TT-----A..... 8303
CPZ.CM.05.SIVcpzMB66          -A-C------CA-A--------TG-CA-A-----T-----------A--A--A--------AA--------A---C-..........--GA--TT-G------GC--A-TGA----C-----------C-----A---C-T-----A......GCA--AT-----A..... 8441
CPZ.CM.98.CAM3                -A-C------CA-A-----AT-AG----A--------T-----------A--A----C--AA--C------T-C-..........---AT-ATAG------GC----TG--T-----A-------T--------T--A-----A......-GC---------C..... 8233
CPZ.GA.88.GAB2                GA-C-------A-----C----TG--A-T-----A--TAT------C--AA-A--T----CAA-A---------C-..........ACTA--ATAG-G----GC--ATTG--T-----A--------C--G--A--T--A-----A......GC---A---G--TAAA.. 8207
CPZ.CM.98.CAM5                -A-T-------ACA-----AT-AG--A-A-----A--T---------A-T--A--------AA-------GT-T---..........---AT--TAG-G----GC----TA--T-------G-----T-----A--T--A-----A......-G----T-----C..... 8539
CPZ.CM.05.SIVcpzMT145         ----------CA-A-----AT-AG--A-T--C--T--T-------GA--A--A--T-----AA-CT------C-A--..........AC-C-CTTGG-C--G-GC---TTA--T-------------T-----A-----A-----A......GC------T--------..... 8329
CPZ.TZ.00.TAN1                GAGT-TTGCTAA-ACAAGCCA-AG-CAGACTT--T-ACTTTAC--GGTG-T--------TTAA-C--------G---G..........GTTTACATAGCC--G-GA--CA-AA-T--T------------T--------TC-----CTA......GCC--AAAT---A..... 8494
CPZ.CM.05.SIVcpzLB7           -A-C------CAC---------TG--A-A-----A------------------A--------AA-C------A------..........---A--AT-G------GA--ATTA--T-----------G--T-----A--T--C-------......GC--AT-----A..... 8328
CPZ.US.85.CPZUS               -AGT-------ACA-----AT-AG--A-A-----A--------------A--A--T-----AA--------T--AC-..........AG-A-GCTC-T-CT--G----ATA--------A-------T--G--A---C-A--G--A......-GC--AT---------..... 8823
CPZ.GA.88.GAB1                -A-T-------A-A---C----TG----A-----T--T---------A--A--T--------AA-A-------CTT-T..........----TTA-ACT----AT----A-AA----C-----GC-C-----------A--GC------A......GC---A--T----CACTG 8862
MON.CM.99.L1                  -AGCAGCA-TCACCTCTC-C---G-C---CTT--T-GGTTCAC-ATC-TCT----G---GC-C-AC-TC-T-CT-G-.........GG-C--CTA--G-----A--CGTGGCT--C--C---C-C--CC-------AGC-----TA......T-CC-TAAC-G-...... 8788
MON.NG.x.NG1                  -ARCAGCCTGCACAG-AGCCA-TG-----CTC--ACGGTTCAC--TCATCT--------CGC---CC-TCGCC-G-G-.........GG-C--TTA--GC---GAC-CGTGGCTG-C--C---C-C--CC-------GGC---G-TT......C-CC-TAAC-G-...... 7413
GSN.CM.99.CN166               -AGCAGCA--AACAG-AA-C---G-CAG-CTT--T--CTTCACGTGGA-CT-------GGCC--AC-TC-C-CTTGT.........AG-C---TA--GC---AAT---TGGCT--C--CC--C-C--C-----A--GGC---G-TA......C-A--TAAC-G-...... 8771
GSN.CM.99.CN71                -AGCAGCA-TCTCAG-AA-C---G-CAG-CTT--TG-CTTCAC-TGGA-CT-------GCCCC-CG-TC-GACT--T.........GG-C---TA--GC---AAT-CTTGGCT--C--CC--C-C--C---------GGC---G-TA......C-AC-TAAC-G-...... 8771
MUS_1.CM.01.1085              -A-CAGCCAGCACAGCAG-C--TG--CAGCTC----CCTTCACCTGGA-CT----G--A-CA--AC-TC---CT-GT.........-G-A---TA--GC---AAT-CTTGGCT--C--C---C-CG-CC-G-----GGC---G-TA......C-CC-TAAC-G-...... 8780
MUS_1.CM.01.CM1239            -AGCAGCA-GAACAGCAGCC--TG-CCAGCTC-----CTTCACCTGGA-CT-------GCCCA-AC-TC-GA-T-GT.........AG-C---TA--G----AAT-C-TGGCT--C--C---C-C--CC-------GGC-----TA......C-CC-TAAC-G-...... 8749
MUS_2.CM.01.CM1246            -AGCCGCA-TCTCAG-AGC----G-C---CTT--G--CTTCAC-TGGA-CT----C--ACC-A-AC-TC-GT-TAGT.........---AA-ATA--G---AAAT-C-TGGCT--C--C---C-C--CC-C-----GGC-----TA......C-CC-TAAC-G-...... 8845
MUS_2.CM.01.CM2500            -AG-AG-A-TCACAGCAGCCA-TG---GGCTC-----CTGGACC--AA-CT-------CGCC--AC-TC----TAGT.........A--CA-CTA-AA-A---AT-CTTGGCT--C--C-A-C-C--CC-C-----GGC-----TA......C-CC-TAAC--......... 8801
RCM.NG.x.NG411                -AGGCTTCAGCAGCT-GC-TGCGG-GATGG-GAGG-CTAAT-CTTACT-CACCTGGAGAG-AC-ATGC-CG--TTGC.........AG--ACC--GCAG--TGGCCAGC-ACGT-GTGC-----G--C--G--------C--G--A......T--C-TAAC--G...... 8535
RCM.GA.x.GAB1                 -AG-ATTCAGCAGCT-G--TGCTG-GATGG-GCTTCA-AAT-CTTACT-CACCTGGAGAG--T-ATGT-CG--G-C-.........-G--ACTT-GC-G--TGGCCAGC-ATGG-GTGC-----G--C--G------C-C--G--A......C---GTAAT--G...... 8452
DEN.CD.x.CD1                  -A-CAG-CCTCAGGGAT---GGG-C-AG-G-CAT-CA-CA--G-AGG-C-ACAG-G--AGT---AT-G-CG-CTC-T.........ACCC----AGCA----AAG-GGTGGC---C--C---C-G--C---------C-T----TT......G-CC-CAAC--G...... 8855
                                                                                                                                                            Env gp41, gp160 end
MAC.US.x.239                  -TG-GGCG-TCCAGGC-G-CTGG-GAT-TG-GA-AGAGA-TC-T--G-GC-C-TGG-GAGACT-ATGG--GACTC-T.........AGGA---G-G-A---TGG--A-T-GCA--C--C--G--G--T-----A--GC-T--GCTC......-C-C-CT-G-G-...... 9243
Env                           H  E  A  V  Q  A  V  W  R  S  A  T  E  T  L  A  G  A  W  G  D  L  W  E  T  L  .  .  .  R  R  G  G  R  W  I  L  A  I  P  R  R  I  R  Q  G  L  E  L  .  .  T  L  L  *
Nef                            M  R  R  S  R  P  S  G  D  L  R  Q  R  L  L  R  A  R  G  E  T  Y  G  R  L  L  .  .  .  G  E  V  E  D  G  Y  S  Q  S  P  G  G  L  D  K  G  L  S  S  .  .  L  S  C  E## . .
H2A.GW.x.ALI                  -AG-AGCGTTCCGGGC---TGCA-GGATTG-GAGAGAGA-TC-TA-AA-CACCTGGAGAGACT-GTGG-GG-C---G.........--GT-G-TCG-G--GAGG--A-T-GCAG-C--A--G--G--C--G-----GGCA---TT......GCCC-C--G-G-...... 9282
H2A.DE.x.BEN                  -AG-AGC-TTCCAAGCA--CGCG-GGA-T--GAGAGAGA-TC-T--G-GC-C-TGG-GAT--T-ATGG----C--CG.........-G-C-CATCG-G--G-GA--A-T-GCAG-T--A--------C--G-----AGCA---CTC......GCCC-C--G-G-...... 9286
H2A.SN.x.ST                   -AG-AGCGTTCCGAGC---CGCG-GG--T--G-GAGAGA-TC-TA-AA-C-CCTGGAGAG-CT-CTGG-GGAC-C-G.........GG-CA-AT-G-G--G-GA--A-TTGCAG-C--A-----G--C--G-----GGCA----TC......GCCC-C--G-G-...... 8728
H2B.GH.86.D205                -AG-AGCGCTCCGAGCGGCAGCA-GG--T--G-GAGA-A-TC-T--GAGT-CCGG---G-CCT-ATGG--G-CTC-C.........AGGA----GGCG-----A--CAT-GCG--C--C--G--C--G-----C-----A--AC-T---CTC......-CCC-CT-G-G-...... 9250
H2B.CI.x.EHO                  -AG-AGCAATCCAAGCAGCAGC--GG--TG-G-GAGAGA-TC-T--GAGC-C-G-GAGG-CCTCGTGG-G---CC-C.........AGGA----GGCAG---AG--CATTGCA-----C--G--G--C--G--G--C-----A--GGCT---CTC......GCCC-CT-G-G-...... 9227
H2G.CI.92.ABT96               -AG-AGCGTGCACCGCAGCAACT-GG--TG-GCAAGAGA-TC-TA-AAGCAC-TGGAGAGCCT-ATGGA--ACTC-G.........GG-A-G-T-G-A----GG--A-T-GCA-----G---------C--G--A--GC-T--GCTC......-C-C-CT-G-G-...... 8671
H2U.FR.96.12034               -AG-AGCATTCCAAGCG--CGGG--A--TG-GCGAGAGA-TC-TAGCCGCAC-GGGAGAGAAC-ATGG--GACTC-G.........GGGC-G-T-G-A---TGGC----AGCT-----A--G--G--C--G--A--A--T--GCTC......GCGC-TT-G-G-...... 8761
MAC.US.x.EMBL_3               -TG-GGCG-TCCAAGC-GGCTGG-GAT-TG-GA-AGA-A-TC-T--G-GC-C-TGG-GAGACT-ATGG--GACTC-T.........AGGA---G-G-A---TGG--C-T-GCA--C-----G--T--G--A--AC-T--GCTC......-CGC-CT-G--......... 8374
SMM.US.x.H9                   -AG-AGCA--GCARR---G-TGG--ATTTGYRCGAGAGA-TC-T--AAGC-C-TGGAGAGACT-ATGG--GACTC-G.........GG-A-G-T-G-A--G-GA--A-T-GCA--C--MR--C-C--C--G--A--GC-T--GCTC......-C-C-CT-G-G-...... 8732
SMM.US.x.PGM53                -AG-AGCA-TCCAGGCAGC-TGG-GG--TCGG-GAGA-A-TC-T--AAGC-C-TGGAGAGACTCATGG---ACTC-G.........GG-A-G-T-G-A---TGGG-A-T-GCA--C--G--G--C--G--G----T--G-----AC-----CTT......GCAC-C--T-G-...... 9180
SMM.US.x.SIVsmH635F_L3        -AG-AGCA--ACAGGCG-G-TGG--A-TTG-GCGAGAGA-TC-T--GAGC-C-TGGAGAGACA-ATGG--GACTC-G.........GG-A-G-T-G-A----GG--A-T-GCA--C-----GC-CG-C--G--A--GC-T--GCTC......-C-C-CT-G-G-...... 9265
SMM.SL.92.SL92B               TCG-AGCCTGCCAGGAAGCATGG-GAT-TG-GCAAGA---TA-T-TC-GC-C-TGG-GACTCA-ATGG--GACTC-G.........GG-C-G-T-G-A----GG--CGCAGCA--C--A--G--C--A--G--G--C---------C-T---CTC.......-GC-TAAC--G...... 8691
STM.US.x.STM                  -AG-AGCA--GCAAGCAGCATGG-GA--TG-G-GAGAGA-TC-T--GAGC-C-GGGAGAGACT-ATGG---ACTC-T.........GG-A-G-T-G-G---AGG--CG-GGC---C--A--GC-C--A--GC-C--C-----A--GC-T---CTC......-CAC-CT-G-G-...... 8894
MNE.US.x.MNE027               -AG-AGCG-TCCAAGT-GCCTGG-GAT-TG-GA-AGAGA-TC-T--G-GC-C-TGG-GAGACT-ATGG--GACTC-G.........GG-A-G-T-G-A---TGG--A-T-GCA--C-----G--C-----G--G--C-----A-AGC-T--GCTT......-C-C-CT-G-G-...... 8720
DRL.x.x.FAO                   ..-CGCT-CTACAAGAGG-TGCGC-GAGG-T-TGGCGC-A--GGAGGC-ACTTGG-CTATCATCAGCTAG--CCC-C.........-G--CCCT-GC-CAG-AAG-GGCAGC---T--C--G--G--T--C--G--G--T-----A---GCA--GGTC......C-C--CAAC--G...... 8606
MND_2.x.x.5440                -AGTATTACAA-GCTTTGC--AG-GGCTATGGAG--G--GGTACCAACTT--ACT-TCAG-T----G--G--C--C-............-CCTT-G-C--G-GC--CT-GA----C---C-CC-C---C-CC-C--C-----A--TGCT--GGCC......C-AC-CAAC-G-...... 8545
MND_2.CM.98.CM16              -AGTACTACAA-GCTATGC--AGGGGCTATGGAG--G--GGCACCAATG---ACT-TCA-CA-C-GC-TGCT-CAG-............-CCAT-GC---G-GC---AT-A----C--A-------C--A---------TC-------AGCA---GTC......C-CC-CAAC-G-...... 8909
MND_2.GA.x.M14                -AGTATTACAAAGCTTTGC--AG-GGCTATGGAG-GG--GGCACCAACTT--ACT-TCAGTTA---G--G--C--CG............-CTAT-G-A--G--C---G-GA----C---C-GC-C---C-GC-C--C-----A--AGCT---GTC......C-GC-CAAC--......... 8838
OLC.CI.97.97CI12              -CGTCT-GAGAACCTTAAAA-AG--GAGACA-.................CATGGGATC-ATT-GCAGC-GAAGAA-..........--GA--AGGG-A--TC--C-A-AGA-GGCT-AGGAGCA-T-GG-GAGA-AAGGA-C-GAA...ACA-C-GGAG-C-G-...... 8323
GRV.ET.x.GRI_677              -AG-GGCA--AACAG-GC-TGCTC-A--TCTG--G-GG--T-CGAGG--A-CCTGG-GC--AC-GGGT-CTA-T--C.........-G-TCC----A-C-G--AG-CAT-A---GT--A------G-GC-G--A---C-T----AA......G-CC--GGG--....... 8766
VER.KE.x.TYO1_patent          -A-CCG-C--ACAAGAAA-A---C-AA-TCTG--AGGC-TT-C-CAAA-C-CATGTC-CCA-A-ATGGCTT-CTTGC.........AG-TCC----A---GAAC--CGT-A---GT--A------G-GC----A---C-T--GGAA.......--CC-TAAT--G...... 8874
VER.KE.x.9063                 -A-CCGCA--ACAAGAAGCA-GTCTC--TGTG--ACGCTTT-C-CAAA-C-CAGGTC-CCAAA-ATGGCTT-CTTGC.........AG-TCCA---A---GCAC--CAT-AG--GT--A------G--C----A--GC-T--GGAA.......--CC-TAAT--G...... 8913
VER.KE.x.AGM155               -AGCCGCA--ACAAGAAGCAG-T--CCATCTG--G-GCTTT-C-CGCA-C-C-G--C-CCA-A-ATGGCTT-CTTGC.........AG-TCC----A-C-G--A--CAT-A--TCT--A------G-GC----A--GC-T--GGAA......G-CC-TAAT--G...... 8890
VER.DE.x.AGM3                 -AGCCGCA--ACAAGAAGCAG-TGTC--TTTG--ACGCCTT-C-CAAA-C-C-GGCT-CCA-A-ATGGCTT-CTTGC.........AG-TCC----A---G--A--CAT-A--TCT--A------G-GC----A---C-T---G-A.......--CC-TAAT--G...... 8411
TAN.UG.x.TAN1                 -AG-GG-A--A-CA--G-CTT-TG-G--TTTG--G-GC--T-CGCAAAGC-CC-GTCG--C-C-ATGGA-T-CTTGC.........AG-TCC----A-C-G--A--C-T-G-ACAT--A--------GC----A-AAC-----C--......TGG--CAAC--G...... 8906
SAB.SN.x.SAB1C                -AGCAGCA--A-CGTCAC-CACTGTG--TGTG-T-CA---T-CGA-AAGC-CC-GT---TCA--ACG-C-T-CTTGC.........AG-TCCAT-GT---G--AG-CAT-GCACAT--A--------GC-C--A-AGC-T--G--A......TGG--TAAT--G...... 9087
MND_1.GA.x.MNDGB1             GTTGG--G--C-G-TA-ACCGGATGGCTACA---A-ATAG-A-CTTCACA---TGCAGAGAA-CA--CATT-CT-CG..................G-G-CGTGC--AT-G-------TC-----G--T-----AA-TGCA---C-A......CCCC-TAAC--......... 8378
DEB.CM.99.CM40                -AG---CCATCACGT-GGCAGGTGGG-AGCTTTGGCAGTGGC-CCAG-G-ATTG-T--G-T--CGT--C-G-G-C-T.........GTCA----AG-AG----A----T-AGAG-G---GCT-----C--------AGCA--GCTC.......--CC-CAAC-G-...... 8559
DEB.CM.99.CM5                 -AG-C-CCATCA--T-GGCAGGGC-A-AA-TCTGGGAGTGGC-CCAA-G-ATTG-TC-GGT--CG---C---G-C-T.........GTCT-G-GAG-C---AACC---T-GC---T---GC------C-----A--AGCA--GTT-......T-AC-CAAC-G-...... 8480
TAL.CM.01.8023                -AGC-GC-TGG-GAGA--GCG---CA--GCTT--A-GGTTCACCTGGCGCT----GTC-G-T--AT-G-CG-CT-CG.........-G-T-G--AGTGCAG--A---GCAGCG--C--GC-G----CG--G--A--GC-T---TAT......TGCC-TGCT-G-...... 8205
TAL.CM.00.266                 -AGC-GC-TGGCGAGA-GGCA---CA--GCTT--A-GGTTCAC-TGGTG-T----GA--G-T--AT-G-CG-CT-CG.........-G-T-G--AG--GA---G--CGCAGCT--C--C--GC-C-CG--G--A--AC-T---TAT......TGC--TACT-G-...... 8655
SUN.GA.98.L14                 CT-CG--CAGGA-C--GAACGCGTGCAGA--CAG--GG-AGAG--TATCATC-T-TC-G-A--CA-G-AGTAG-ACCTTTAGTCTCGGCA--AAA--G--GC-C-AAT-GA-T-G-A-A---G---GT---ATTCC-AGT--G-CCACAGAG-CCACTT-G--......... 9181
WRC.CI.97.97CI14              C--TAC-CAAGAAG-TTGCAAAG-GCAGG-AG-AG-ATA-GCAG--A----TAT---CA-CT-GC-GCAG-AAGAC-.........GC--C-AGCA-----ATC-GGATG-TTGCTA-C---CTCTGCGC-AGCTTTGATA--T-CAGAAGA-GC--CAGTG--AGAAGA 8920
WRC.CI.98.98CI04              GC-TAC-CAAGAAG-TTACAAAG-GCAGG-AG-AA-C-A-G-AGAAA---ATAT----A-ATAGC-ACAG--AGAC-.........AC---TAGCA------TCTGGAAA-TT--TA-C---CTCTGCTC-AGCCTTAATA-G--CAGAAGA-GAC-CT-C--------...... 8959
WRC.GM.05.Pbt_05GM_X02        C-CTACAG-AGAAG-TTGCTGCGCG-CGGCAGAAA-G-AAG-AGAAG---ATATACACGCCT-GC-GCAG--AGAC-.........AGCA--AGCA-C--CATC-GGG-AATT-CTG-C---CTCTGC-C-A-CCTTAATCCGT-CAGAAGAGGACA--AG----...... 7855
LST.CD.88.524                 CTGCG--CTGGA-C--GAACAAG-GCAGA-GGAA--G--G--AGCTATCAAAA--TA--G-AACA-ACATTAGGT-G.........GG-CA-AACA-A---AGATGG...AGAT-CAGATTC----GTG-G-TTCC-AGT--G-CCACAGAG-CCACT--G--G...... 8035
LST.CD.88.447                 TTGCG--CTGGAGCT-GAACAAG-TCAGA-GGAA--G-AG--ACATATCA-CAG-TC--GCAAGA-ACCTCAG-C-G.........GG-CA-AAGAAG---TGG-GATT-AGAT-CAGAG------GTG-GTTTCC-AGT--G-CCACAGAG-CCGCT--G--G...... 8035
LST.KE.x.lho7                 TTC-G--CTGGC-C--GGA-AAATGCAGA-GGAA--G--G-CAGTTATC-A-C--T-----AAAA-ACATCCA-T-G.........GGCAA-AAGAAG---TGG-GA-T-AGAT-CGGAG------GTG-GATTTC-AGT---GCCACAGAG-CCGCTT-G--G...... 9110
SYK.KE.x.SYK173               -A-TCGCA--A-C-TCGC-CGGA-GG-AA-T-CG-GATTGG--G--G-CTATTTGGC-AGCAA-ATAC-CC-CTACC.........AG-C---T-GTCGAG--CG--GCAGC-T-G--GC-G--GC-TC----A--------G-TC......TACC-TAAC-G-...... 8723
SYK.KE.x.KE51                 -AG-AGCA--CA-CTCGC--GGACGA-AAGTG--GCACCAGA-G-TG-CCATTTGGC-AGCCC-AC-C-CCTAT-CC.........AG-C---T-GC-GAGAA-G-CGCAGC-T----AC-----C-TC-------GC----G-TC......TACC-TAAC-G-...... 8422
COL.CM.x.CGU1                 G-T--CTA......TTGG-A-C--GG-GTC--AG-C-TCAGA-TCT----ACACTTC-G---TGGC--CG-TCT-CG.........GCGC-T-GC--GC-TCGAGCG-CTG-ATAT-TGC--G--TGG-T-T-T-A-CGTCCCCAA......GG-CCTGCC---...... 8151
```

```
                                                  Nef start
H1B.FR.83.HXB2      .......................ATGGGTGGCAAGTGGTCAAAAAGTAGTGTGATTGGATGGCCTACTGTAAGGGAAAGAATGAGACGA...........................GCTGAGCCAGCAGCAGATAGGGTGGGAGCAGCATCTCGAGACCTGGAAAAACA 8915
Nef                                   M  G  G  K  W  S  K  S  S  V  I  G  W  P  T  V  R  E  R  M  R  R  .  .  .  .  .  .  .  .  .  .  A  E  P  A  A  D  R  V  G  A  A  S  R  D  L  E  K  H
H1A1.UG.85.U455     ...........................................AG--CAGAG-G-A-------GAG--T--A----------GA-..-..-..-..-..-..-..-..-..A--CCTG--------A-AG-A--A-------T-----A---TT-A--T---T- 8366
H1B.US.90.WEAU160   ...............................TA---------C-----GT-G---------G--A--------------A-------------------------------AG------------T----------T---C------ 8922
H1C.ET.86.ETH2220   ...............................G---C-AT------T-----CCAG-A--------G--A----A-------A-------------------------CT-----------GG-A--A--------G---------T-A--C---T- 8312
H1D.CD.84.84ZR085   ...............................A--------------A-AG---------G--A--------------A---AA-...ACTGATCCAAGGGAAAGAAGAAGAC-----------------G-------------T----------------G--- 8460
H1F1.BE.93.VI850    ...............................G----------------C---A-AG----------G-----G--------------A-........................A-CCCTA----------AG------------TG---------T-A--C-G--G 8192
H1G.SE.93.SE6165    ...............................A-----------------CA-AG----------CGAG--------------A---AAC........................A-CCCTA----------AG-A--A-------T-----A---TT-A--T-GG-- 8354
H1H.CF.90.056       ...............................A-----A-------------GA--GG---G---T-----A---------------G-------------------------------A----T------AG---A-------TG--------TT----T-G--G 8235
H1J.SE.93.SE7887    ...............................GAA---A-------------.................CAG----------------G-------------------------GCTC-C-CT------------G-A----------TG----A---TT---CT--G-- 8213
H1K.CM.96.MP535     ...............................------------------------A-AG----------G--A-------------------GCACGACCAGCAGCAGACAGGGTGGGAA-AC-A-----------CG------------T-----A--------CT-G--- 8102
H101_AE.TH.90.CM240 ...............................AA-T---------------CA-AG-G---------CAG--C------A--A-AG-A-........................A--CCT------A----AG-A--A-------T-----A---T--A--T------ 8489
H102_AG.NG.x.IBNG   ...............................-----------------C--CA-AG-G---------AG--T-T-A-----------A-........................A-CCCTA--------ACAG-A--A----------T----A---TT-A--T-G-- 8442
H1N.CM.95.YBF30     ...............................AAAG-TT--------G--C--CC-AG-A--------AGAAA-C--A----------A-------------CAAACGCAAGAACCA--A-TAGAGC------TAG-A-CA-------T----A---T--A-CT--T-G 8502
H1O.BE.87.ANT70     ...............................AAA-GCA-T-AG----G---AAT-TGAG------G-AG-A-----A----------A----------.ACTAGAACTTTCCCTGAGT----A---TGC---CC-G-A--A---CAGATC--CA-G--AT-A-C-GCTAG 9002
H1O.CM.91.MVP5180   ...............................GAATGCA---AGC-----C-AAT-TGCA-----T-AGAA----A--T----------A-----------TCCTCCTCTGATCCTCAAC-A---TGT---CC-G-A--A-----T-TC--CA-G--GT-A-C--CTAG 9025
CPZ.CD.90.ANT       ...............................------TCTGCA-----T----TC-AGTG-G-----GCA.............AC--GC---T--GAA-.......................ATACA---AA--AATC-----GATA-A--GC-CTGTGGAAAC--AT---CC-GTAG 8369
CPZ.CM.01.SIVcpzCAM13 CTTTAAAATAGTCTTGCTTTAAG.....A-CTC------T-------CC-AG-A--------GAAA-C--AA------A-A-GAG.......................----CCAAT----------AG-A--A----AG-TT---AAG-T--A-----T- 8548
CPZ.CM.05.SIVcpzEK505 ...............................AAAAGTT--------G-----CC-TG-G-G-----ACAA--C-TA--T--------A-.......................CAGC-AGATC-------AG----A------TT----A---T--A-CT--T-G 8422
CPZ.CM.05.SIVcpzMB66 ...............................A----------------C---C--G-G--C-----AGAG-----A--C---C-T--GA-G.......................A-AC-AA-T--------AG-A--A---C---T---A--G-----A-C--G--- 8560
CPZ.CM.98.CAM3      ...............................CAA--G----------------CC--G-A--G-----AGA---C--A------C-A--CGA-.......................CTCA-C-TGCTT---CCAG-A--A---C---TC--A-A----T-A--GGC--- 8352
CPZ.GA.88.GAB2      ...............................CAA---A-----T-----C--CA-TG-A--------GCAG--C--A--G---C-T--G--G.......................A--C--GA---------AG-A--A----A--TG----AG--TT-A-C--GG-G 8326
CPZ.CM.98.CAM5      ...............................AA------------------C--G-G-----G-AG----T--AC---G--A----A-.........................A--CCA---------TCAG-A--A---C---TC--A-A---T--A--GGC--G 8658
CPZ.CM.05.SIVcpzMT145 ...............................A-C-----------------CC-AG-A--------AGAAA----A--G-----A-A-A-.......................---CCAA----------G-A--A--G-A--T---A-A----A-A-C--GC-G 8448
CPZ.TZ.00.TAN1      ...............................AAA--TA-TTGGT-G-T-GCC-.......-GGCC-GG-AA-CC-TC---GATC-T.............................CA-A-CA-CT--AGT--GCCT--A---CAG--C--A-A------CC-G--TA- 8601
CPZ.CM.05.SIVcpzLB7 ...............................AA----A-------------CC--G-A--------GAA--T--AAG------T----A-.......................-AAACTGT---------AG-A--A---C----C--A-A------A-CT-GG-- 8447
CPZ.US.85.CPZUS     ...............................-----------AA--------T--G-----CA-AG-A--------AGAA--C--AA-C---T------A-.....................ACTCAGA-AACAG----------AG-A--A---C-T-TC----A----T-A-C-G---- 8948
CPZ.GA.88.GAB1      CATTAAG.......................AAC---A-----T--------C--G-A--------GAG--C--AAG------A--GGA-.......................---CCAA---------GG-A--A----A--TT--GAAG-----A----G--- 8988
MON.CM.99.L1        ..................................................................................................................................................... 8788
MON.NG.x.NG1        ..................................................................................................................................................... 7413
GSN.CM.99.CN166     ..................................................................................................................................................... 8771
GSN.CM.99.CN71      ..................................................................................................................................................... 8771
MUS_1.CM.01.1085    ..................................................................................................................................................... 8780
MUS_1.CM.01.CM1239  ..................................................................................................................................................... 8749
MUS_2.CM.01.CM1246  ..................................................................................................................................................... 8845
MUS_2.CM.01.CM2500  ..................................................................................................................................................... 8801
RCM.NG.x.NG411     ..................................................................................................................................................... 8535
RCM.GA.x.GAB1      ..................................................................................................................................................... 8461
DEN.CD.x.CD1       ..................................................................................................GTTACTG-AGC-TGAT-AG-AC--AGAAACCTTT-T--ATG-AGA-...GAAATC-TC-- 8912

MAC.US.x.239        ..................................................................................................GGGACAGAAAT-CAATCAG-G-C-GTATA--AATA-TC---GGA--A---CA-CTG--G- 9303
Env
Nef                 _.__.__.__
```

```
H1B.FR.83.HXB2         TGGAGCAATCACAAGTAGCAATACAGCAGCTACCAATGCTGCTTGTGCCTGGCTAGAAGCACAA......GAGGAGGAGGAGGTGGGTTTTCCAGTCACACCTCAGGTACCTTTAAGACCAATGACTTACAAGGCAGCTGTAGATCTTAGCCACTTTTTAAAAGAAAAGG 9079
Nef                     G  A  I  T  S  S  N  T  A  A  T  N  A  A  C  A  W  L  E  A  Q  .  .  E  E  E  E  V  G  F  P  V  T  P  Q  V  P  L  R  P  M  T  Y  K  A  A  V  D  L  S  H  F  L  K  E  K
H1A1.UG.85.U455        -------G-------C--T------T--T----T-----CAG-----------G-----G--......-----A-GA--C--A--C--C-----T-GG--A---------------------------T---------T-T-----C---TT---------------- 8530
H1B.US.90.WEAU160      ----------------------GAG......-AT-------A---C-T----T--A-------......-----T-------------------T-G-----------------------------------GCAT---------------.------------- 9079
H1C.ET.86.ETH2220      ---G---C-T-----C-----C---C-C--C-AT---C---A-----------G-----G---GAG...-----A--A-----A--C----------G---------------------------T--------AT-C-----C----T---------------- 8479
H1D.CD.84.84ZR085      ---G----------------------A----T-------C--------G---------GGAA...-----C----------C----------G-----------------------------------------------------------G-A------ 8627
H1F1.BE.93.VI850       G--------T-----C--------TAG-A----T---C---AC-TG--------G-----------......----------A--A--C----------G-------------G-----------C--T-------------C--T------------------ 8356
H1G.SE.93.SE6165       ---------------C-------------C-AT---C---A-----------G---------GAG...-----CTCA--A--A--C----------G---A--A--------G--------------TT----GT---T-T-----C---TT---------------- 8521
H1H.CF.90.056          C--G---G------T--AT---------T----T---CG--A-GCC--------G---------GAG...--C-G-----A--A--C----------GG----------------------------C--T----G----T-T-----C-----T---------------- 8402
H1J.SE.93.SE7887       ---G-----------C-------------------A-------AC-----------G------------......ACA-----------A--A-----------A-------A----A--------------A-------G-----T---------TT---------------- 8377
H1K.CM.96.MP535        ---G---G-------C--------TT--CAC-AT---C---A-----------G--------G......-----A-----A--A------------G------------A---------------T---------T-T-----CG--TT---------------- 8266
H101_AE.TH.90.CM240    ---------A--------T---......ATAGAT-------A-----T------GAG-------......-----C--------A--C----------TG--G--------C---------------T----G----T-T---------TT---------------- 8647
H102_AG.NG.x.IBNG      ---------------C--T---------CAA--T--CC---A-----------G------------......-----T---A-T--A--C---------GG--A-----G----G-------------T----G-----C-----C--------------------- 8606
H1N.CM.95.YBF30        A--G--C-----C-TA--A-----TAG--AC-AT----AAAG-ATA--T--------------GAA...--A--A-----A--A--C--------ACGC----------A-----G-----A--C--T--ACAG---T-T------TC-TT------------T---- 8669
H1O.BE.87.ANT70        A----GG--AC-----TC-C----TC-TCAA-A------A--CCT---A-TC------AGT--C.....C-A-----A--A--A----------AG-------A--G---C----G--------C--T--A-G---AT-T--C--C---TT---------------- 9166
H1O.CM.91.MVP5180      A--G-G---AT-----TC-C-C--TC-TCAA-A------A--CCT---A-TC-----CAGC--C.....A-A--T-----T--A--C--C-----A-G------A--G---C----G--------C-TT--A-----CT-T--C--C---TT---------------- 9189
CPZ.CD.90.ANT          A--G---C-A--CTCCTCT-C--T--GTA-AGAA--A-A--TAATCA---ATTC-----AT--T.....ACA-----A-G-ACA--C-----T--G-G------A--C--CA-G-----C-----AG-A--ATT---------C--GTCATGN--------------- 8533
CPZ.CM.01.SIVcpzCAM13  ------T-----T-----AC-----C--AG---T---CAAA---TA--A---T----T-A-ATGAAACATC-T--A--A--A-----C--------T-G---A-----G--A----------------------A-G---CT-T--C---TC-----------G--G----- 8718
CPZ.CM.05.SIVcpzEK505  A--G--T--T--C-TA--A-----TAAG-AG-AT---CAA--CCTA--T----------AG--......A-A-----A-CA--A--C--------ATGT-----------A-------------------A-----A-A---T-T------TC-TT--------G--------- 8586
CPZ.CM.05.SIVcpzMB66   ---------A--C-----A--C---T-CCAG--A----AGA-CCT-------T------A-GT-......C-AA-T--A--A--A--C--------A-GG-----A-----C------------------C-TT----G---T-T-----C--------------------- 8724
CPZ.CM.98.CAM3         -------G----C-C---GC-C---C--CAG-AT---CAGA--CTG------T------A-ATGCAAGAAC-CA-T--A--A--A--C--------T-G------A------C-T-----T-----C-TT--A------T-T---T-A---------------A------- 8522
CPZ.GA.88.GAB2         ---------A--TTCC--AC----CC-TCAG--T---CAAA--CT---A----------A-ATG.......-AC-------T--A--G-------A-G-----------------------------A--T--A-----CT-T-----C--------------------- 8490
CPZ.CM.98.CAM5         ------------C-C---A---CAC...A-A-AAT--CAGA-CCTG------T------A-ATGGAAAATC-T--A--A--A--A--C--------T-G---C---------CC--------------A------CA---T-T---T-A--------------------- 8825
CPZ.CM.05.SIVcpzMT145  -------G------T---A-----CC--CAA-------AGA---TA--T----------A-ATG.....C----T--A--A-----A----------A---A--A--G--C--------T------C-TT----GT--CT-T-----CTCT------------------ 8612
CPZ.TZ.00.TAN1         A----GTC----T-C--A--CCCT--GTA-CT-AGCA-A--TG-TA-AA-ACTCT-C--AC--T.....ACT--A--A--A--A------------G---AGCA-----CA-G-----C-----AG-G---CT---AA-------GTCATGG--C------------ 8765
CPZ.CM.05.SIVcpzLB7    ---G-----T--------T-----CC--CAA--T----AGA-CAT-------T----G-AG--......A-------A--AA-AT-G---------A-A------------A-----G--------------TT--A-AG---T-T------G--T----------------- 8611
CPZ.US.85.CPZUS        ---G--T-----T-C---A------C--CAG-A---CCAAA--CTG--A---T----T-AGATGACTAATC-T--AAGT--A--A--C--C--T--TCGC-----A-----A-----G--C------------ACA---AT-C---T-AG--TT---------------- 9118
CPZ.GA.88.GAB1         C-----T-----T-----G--C--CC---AG--T---CAAA--CTA--T--------G-A-ATG.....-CA-T--A--A--A--A--C--------A-G---A-----T--AACG--------------T--A------T-T------TCA-------------------- 9152
MON.CM.99.L1           ............GG-C--A---GGCAA-CAACG-CGGCAAC--.........-G--C-AC...GACGAC--C--C--C-GT--C----G-------A-GG--C-GA--T--GC-TC-CGACCCT----GG---CTCATGA-G--C--CTCT--T-A---G---------- 8934
MON.NG.x.NG1           ............GG-GC-G--CCACCGCT-A-A-CGCCGCC-AACCA-TGTCG-T--T-ACGCCGACGAG--CC-A--TTT-CCT--C-G---G--TCGC---AT---G--AC-T---GACCCT--A-GG---ATCATGA-G--C---TCT--T---------------- 7571
GSN.CM.99.CN166        ...........................--AGC-A--AC-CAAGA-A-GAA-C--C...............------T-T--ACCATCA-GC-----A-G---A-G------AC-T---GACCCC--C-----A-TCATG--G-----CTCT--------G---------- 8899
GSN.CM.99.CN71         ...........................--AAT-ACCCT-CAGGCAAAAT--CA-C..................--TT-T--ACCATCA-GC-----G-G---A-----G--GC-T---GACCCC--C-----ACTTATG--G-----CTCG--------G---------- 8896
MUS_1.CM.01.1085       .............-G-AGC--G--A-AG-C--GTG-TA...........................GAA---C-T----...CGTACC-GC--G--GCGT--GAG---G--AA-CC-TGACCCC--C------CTGATG--G---A-GTCG--T--C-------------- 8905
MUS_1.CM.01.CM1239      .............-G-A------GAGCGGC-ACT--TC...........................GAA--AC-T----...CGTTCA-GC----GCGG--G-GA--G--CC--C-GGATCCC--A------ATGATGA-T--C--CTCT--T-A-------------- 8874
MUS_2.CM.01.CM1246      .............--AC-AGGGCCGCAGGCAA-T--TA...........................GAA--AC-T----...TGCTCC-GC----ACGC--GA-A--C--GA-TC-TGATG-C-AC-----ACTCATGA----C--CTCT----A---G---A-----A 8970
MUS_2.CM.01.CM2500     ............GG-A--T--C-A-AA---A-T--TA...........................AAAA-AC-TA-A...CG-TCG-G-----ACGC--G--ACAC--AA-CCAGAACCCC--C--T---TT-ATAA--A-C--CTCT--T-----G---------- 8926
RCM.NG.x.NG411         ..........AGG-G-GTTT-CA-T--A-A--CAGC-GT-A-TGGA---AATAGATTGGTATG--GATACA--T--TACCTT---A--G--------G-G---A--A--C--CA----G------T-------ATT---AA-------ATCT------------------ 8696
RCM.GA.x.GAB1          -AC----GAA--GCAAGCTTCA-T--AT-AC...-T--ACTGG-A--AAGATACT...............--T--CACCTT---A--A--------G-A------A-----AC-T----------G--------CT---AA----CA-GTCT------------------ 8613
DEN.CD.x.CD1           ATC-T-TCCAG--CTGGA-CC-CA-CATC-A-GA.........ATACA-CAATAT-TG-T-G-CCATGAT--T--A--A--A--A--C--------ACGG--A-GACG---AC-GCACG--CC---C--T--A-ATATGA-----A-GTCT--T-----G---------- 9073
                                                     Premature stop in SMM239

MAC.US.x.239           GA---A--AATT-GCATA--GA-A-CA-AA--TGG---A-ATA...ATGA-TA-...............--T--T--CTT---A--GG-AT----G-GG--AA-A--T--CC-----A------G------ATTG--AA----CA-GTCT--T---A------------ 9455
Nef                     R  E  K  L  A  Y  R  K  Q  N  M  D  D  I  .  D  E  *  .  .  .  .  .  .  D  D  D  L  V  G  V  S  V  R  P  K  V  P  L  R  T  M  S  Y  K  L  A  I  D  M  S  H  F  I  K  E  K
H2A.GW.x.ALI           AA---AG-AAGA-TTGTA---GCA-CAGAA--TGG---A--TA...AT-TAGAT--T..............--T--TAGCCTA--A--GG-CT-T--------AAGA---AA-------A-------A----ATTG--A------A-GTCA--T--AA---------G-- 9497
H2A.DE.x.BEN           AA--CAG-AAGATTTGTAT-GGCAGCA-AA--TGG---A--TA...AT-CTGAT...............--T--T--CCTAA-A--AG----T--T-----AAGA-----ACGG---GA------C--T--ATTG--AA-----A-GTCA--T---A----------A- 9498
H2A.SN.x.ST            A--G-AG-AAGGCTCGTA---GCA-CA-AA--TGG---A--TA...AT-CAGAT...............--T--T--CCTA--A--GG-C--T--------AAGA-----A------GA------A--T-G-TTG--AAG----A-GTCA--T--GA------------ 8937
H2B.GH.86.D205         --CCCGGGGGG--GAAG-GGGAGGGCA-CAAGATGCA-A--AA...AGTGATGAG...............--C--T--A-T--GA-CCA-ATGTAAG--C--AATA-----AC-G-----------A-----ACT---A-----CA-GTCT--T---A---------C--- 9411
H2B.CI.x.EHO           GACTTTGGGAG--GAAG-GGGAGG-AA-CAAGATTCA-A-.......-AGGATGAT...............-----CA-T--A--A--GG-C-GT--A-G---CGG---C--AC-G--G---------A-T---ACT---A-----CA-GTCT--T----------------- 9385
H2G.CI.92.ABT96        AA------AGTT-GCATA--GGYAGCA-AA--ATG---A--TA...ATA-TGAT...............--TA-T--CTTA--A--AG-A------TAC--CA-A--C--GC-----GT-----G------ATTG--AA----CA-GTCA--T---A------------ 8883
H2U.FR.96.12034        GAA--AT-AATT-CAATAT--GCA-CA-AAC-TGG---A--TA...ATGAAGA-...............--TA-T--TTTA--A--AG-AG----A-G---AAGA---G-AC-G---A------------ATTG--A------A-GTCA--T---A----------A- 8979
MAC.US.x.EMBL_3        AAAG-A--AATT-GCATA--GA-A-CA-AA--TGG---A-ATA...ATGA-GA-...............--T--T--CTT---A--GG-AT----G-GG--AA-A--T--CC-----G------------ATTG--AA-----A-GTCT--T---A------------ 8586
SMM.US.x.H9            A-------AATT-G--TAT-GACA-CA-AAC-TGG---A--TG...ATAATGA-...............--T......TTA--A----G------G-GC--AAGA--C--CG-----ATC-----A-----ATTG--AA----CA-GTCT--T---A------------ 8938
SMM.US.x.PGM53         GA---A--AGTT--AATAT-GACA-CA-AAC-TGG---A--TA...ATGATGAT...............--T--T---CT-A-A--AG-AT---ACAC--AA----G--A------G-C---T-A-----ATTG--GA----CA-GTCT--T---A------------ 9392
SMM.US.x.SIVsmH635F_L3 AAA-----AATT-G--TAT-GACA-CA-AAC-TGG---A--TG...ATGATGA-...............--T--T--CTTA--A---G-CT----GCAC--AA-A--C--C-----GG-C-----A-----ATTG--AA----CA-GTCT--T---A------------ 9477
SMM.SL.92.SL92B        AA--CA--AGCA-CAGTAT-GGCAGCAGTG--TGG---A-ATA...ATGA-GAT...............--T--T---CT--CT--AG-AGTT---TGG--AA-A-----CC-----G-----T-A-----ATTG--AA----CA-GTCA--T---A------------ 8906
STM.US.x.STM           AA-C--T-AGCT-GAATAT-GACAGCA-AAC-TGG---A--TA...ATGA-GA-...............--T--TA-TCTA--A--AG-AG---ACAT--AAGA--C--AC-T--GGA-------------TTG--AA----C--GTCA--T---A-----AGC---- 9106
MNE.US.x.MNE027        GA-G-A--AATT-GCATA--GA-A-CA-AA--TGG---A-ATA...ATGA-GA-...............--T--T-GCTT---A--GG-A-----G-GG--A-GA--T--C------G-------------ATTG--A------A-GTCT--T---A------------ 8932
DRL.x.x.FAO            -CATATGCAAG--GCAGCAGCAGAGAA--A-GT--CCCT-C-CACA-AAGAAGA-C--C-CTC-GAAGAA--A--A--A--A--A-----------ATGT-----ACGC---------GAGCCA--C-----A-ATCTA-----CT--TC------------G-------- 8791
MND_2.x.x.5440         AA-GCACCAGCAGGA-GTA-C-CTCC-CT-AGAA--A-AGCAACCAAGTGATGA-...............-----A---.....-A--C---------TAC--AAC-CAG--AG-GCCTGA-GCA--C--------ACTTGA----CA-GTC-------------------- 8706
MND_2.CM.98.CM16       .............................................-GAAGAG...............--A--AAC-.......--C-------ATAT--A---TGC--AG-----GA-CCC--C-------ACTTG--T---A-GCC-------------------- 9010
MND_2.GA.x.M14         ......................................AGTA-ATGATGAGC--............A-A--A--A...AG---C--------TTAT--CACACAG---G-GC-TGA-GCA--C-----AAACCTGA-----A-GTC-------------------- 8951
OLC.CI.97.97CI12       G-C-AT-GAGT-TTTA-TG-C-GGG-AC.....................ACTGAC..............AC---T--TCT--AA--A--------G-AG--------G--AG-GC---A-GCC-----T---A-TCTAA-T--G--CTCG--T----------------- 8478
GRV.ET.x.GRI_677       C--CCGCCG-CTGCAAGAAGGAGACA-CTT-...G---AGTGG...-ATGATGAT...............--A--A--A...--A--C--C--T--GCA-----GA--C--C-------AG-----C--T--ATT---A--G--CT--TC-------------TC----- 8930
VER.KE.x.TYO1_patent   G---AGG-GGCTTCAAGAAGGAGACA-CTGG...G-A-AGTGG...T-GGATGAT...............-----A--A..------A--------G-G---AAGA-----C-----G-A---------T--ACTT--A--G---T--TCG------------------- 9047
VER.KE.x.9063          ----AG-CG-CTGCAAGAAGGAGAC-A-TGG...G---AATGG...T-AGATGAT...............--A--T--A..------A-----T--A-A---AAGA-----A--G--G-A---------G--G-A-----------ATT---A--G--CT--TCG------------------- 9086
VER.KE.x.AGM155        A---AG-CGTCT-CAGGA-GGGGACCA-TGG...G---AATGG...T-AGATGA-...............--A--T--A..------A--------A-G---AAGA--G--AC------A---A------A--ACTT--A------T--TCG------------------- 9063
VER.DE.x.AGM3          A---AG--AATT-CAAGAAGGAGATAA-TGG...G---AATGG...T-TGATGA-...............--A--T--A...--A--A--------A-G---AAGA--G--GC------A-------C--T--ATT---G--G--CT--TCG------------------- 8584
TAN.UG.x.TAN1          A---AG--AATTGGCAGCTG--GACA--TGG...G---ACTGG...-AACCAGAG...............--A--A--A...------C--------A-G---A-GA--G---------A-------A--T---TTG--A------T-CTC-------------------- 9073
SAB.SN.x.SAB1C         A--T-T--AA-TG-AACATC-AGG-AG-CAG...CCCT-GTGG-A--ATGA-GAT...............--A--A--A...--A--C---------A-GG---TGCT----AC--C--G--------A--T---TTG--AA-----CT--G------------------- 9263
MND_1.GA.x.MNDGB1      ...............AGATG-G-A-CTT-ACTTA-CACAGCAGGA-C-AGAAGAG...............-----A--A...--T--A-----T--GTGT-GC--A--TT-C------GTGCCAT-A-----A-ATCTGA----CT-CTCT--T---A------------ 8515
DEB.CM.99.CM40         AA-TC-T-GA-G-GAA-TA--GTGCA-CAA--AGC-ACAACAAGACTTAGAT................C--ACA--A-CA--A--C--C-----ACGC--C-------T-CA-CC--GACCCC--A--T---TT-ATGA----CTACTCT--T-----G---------- 8720
DEB.CM.99.CM5          A-ACC-CCA--G-GAG-TA--ATGCA-TCAG-GAGCCCAG-AG---TATGAT................------AC-T-G-A-A--C---G-T----G---AAG---C--AA-T--GGA-CCC---------ATTGATGA----CTACTCT--T-A-------------- 8647
TAL.CM.01.8023         .--GAAGGCACTCTCGGAAGGAGA--TCTGGTATG-AAAG-AGCAA-AAGAAAAT...............--CAT----CTT--A--C---------A-GG-----A--T--A--G--G--CTGT-A------ACTG-G-A----CT-CTCGTT------G---------A- 8393
TAL.CM.00.266          .--CAAGGCATT--CACAAGGAGAG-TTTGGTATG-AAAA-AACAA-AAGAAAAT...............--CAT---ACTA-----C--C--T--G-GG---A-A--T--G--G---A-CTGT-----T--ACT--G-A----CT-CTCGTT------G----------- 8843
SUN.GA.98.L14          ..............................................................................................................GCCA--A--T---CAGCTAA-------CTCA--T---A------------ 9231
WRC.CI.97.97CI14       .......................................-AGTGATGA-...............--T--A-GA...--A-----------A--T-----A--T--CA------A-CCA---TT--ACTGCTA-----C--CTCG--T--A--T---------- 9030
WRC.CI.98.98CI04       ...................................A-AAGA-AGT-CTGAT...............--T--A--A...--A--C---------TCAT---------T--AA------A-GCT--A-TT--ACTGCTA---------CTCG--T--A--T---------- 9072
WRC.GM.05.Pbt_05GM_X02 ...................................A-AAGACT-A-CTGAT...............--T--T-CA...A-A--A--C--T----G---A--A--T----------AGGCA--A-T----ATCTTA-----C--CTCT--T--A--T---------- 7968
LST.CD.88.524          ..............................................................................................................CTCT--C--T--AA-CTTGA----C--CTCA--T---A------------ 8085
LST.CD.88.447          ..............................................................................................................CCCT--C--T---A-TCTGA----C--CTCA--T---A------------ 8085
LST.KE.x.lho7          ..............................................................................................................CCCT--C--T---A-TCT-A----C--CTCA--T---A------------ 9160
SYK.KE.x.SYK173        A---CA--AATTCTCACCTGGAGATAT--TG...C-G.........-ATGAAGG-...............--CACA-GTCTC--A-----C--T--GTGT-----AAC---AC-C---A--T-A--C-----ATTG--AA----C--CTCT--T---A-----A-C---- 8911
SYK.KE.x.KE51          G--GAA--AATGG-A-GCGGGAGACATCTG-...C--.........-ATGA-GG-...............--T-----TCTA--A--C-----T--TTT---A--A-----C--G-----CT-A--------ACTT--CA-------CTC---T---A-----AG----- 8601
COL.CM.x.CGU1          -A--CATCCAGATGTATATT-CCACAAC.........C-A-AACCA-AGGACG--...............--A--T-TA--A-CA--A-G-TAT--G-TGT---GAC-G--AG--C---AGGC---CATT---CT-CTG--G--CA-CTCTTG------G---------- 8318
```

```
                           3' LTR U3 start                                                Nef premature end in HXB2
H1B.FR.83.HXB2             GGGGACTGGAAGGGCTAATTCACTCCCAAAGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTAGCAGAACTACACACCAGGGCCAGGG.....................GTCAGATATCCACTGACCTTTGGATGGTGCTACAAG 9228
Nef                        G  G  L  E  G  L  I  H  S  Q  R  R  Q  D  I  L  D  L  W  I  Y  H  T  Q  G  Y  F  P  D  *  Q  N  Y  T  P  G  P  G  . . . . . . .  V  R  Y  P  L  T  F  G  W  C  Y  K
H1A1.UG.85.U455            ----------T---T----------A--G-A---------A---------------G----T-----------A-T----------G-----T-----------------A.....................A------C-----A--A--------------- 8679
H1B.US.90.WEAU160          -----------------T----------A--------------------------G------------------T----------G-------------------------.....................ACT-----C------TG-------------T---- 9228
H1C.ET.86.ETH2220          -------------T-----T-----A-G-A---G--G--G--T------T-----G----TA-----------T----------G--A-------------A-------.....................-----------------------------T---- 8628
H1D.CD.84.84ZR085          ----------------G--T--------G-A--------------------T--G-----------------T----------G--A-------------------------.....................A-----------T-A--------G-------T-G-- 8776
H1F1.BE.93.VI850           -------------T-----T-----A-G-A---GG-----C-------------G-T---------------------------G-------------------------.....................A-----------------G---------T---- 8505
H1G.SE.93.SE6165           ---------T--------T-----A-G-A---------G--------C-----G----A-T---------A--------A---G-----------G----------------.....................ACT----TC--C---------------------T---A 8670
H1H.CF.90.056              ---------T---T-----T-----A-GCA------G--C---------T-A---G---TA-------------------------C-G----------------------------.....................-AG----T---C----------G-------T---- 8551
H1J.SE.93.SE7887           ---------T---T-----T-----A-G-A---------G---------------G-TC-TA---------T-----------C-G--A----------------------------.....................A-------C---T---A----------------- 8526
H1K.CM.96.MP535            ---------T---------T-----A-G-A---------A---------------G-T--T-----------T-T----------G--A------------------------.....................A-------C--------------G------------ 8415
H101_AE.TH.90.CM240        ---------T---------T-----A-G-A---------G--------CT-A---G----TA-T----------T----------G--A-----------------------.....................A------TC-----TGT--------------T---- 8796
H102_AG.NG.x.IBNG          ---------T---T-----T-----A-G-A---------A---------------G----TA----------A-T------A---G--------------------------.....................ACT----TC--------------G-------T---A 8755
H1N.CM.95.YBF30            ---------------G--TGG---AG--A---G--------T--A--C--C-----G--T-------------AT-C-------C-G---T-------------------A.....................A-T-----C--CG-A---------------T---A 8818
H1O.BE.87.ANT70            -A-----------T-----T-------T-A---GC---A----G-----T---G-G--TA----T--G--A-T----------G-------------------A-----A.....................AC---G-TC---------A----------TG-TT--A 9315
H1O.CM.91.MVP5180          -A--------T---T-----T-------T-AG---GC---A----G-----C-----A--T-----T--G--A-T----------G----TGT-------G--A-----A.....................CCT----TC---------A----------TG-TT--A 9338
CPZ.CD.90.ANT              ----------TN--------T-NN-A-TT-A----GC--C------A---A-------G-TTA----C--G--AAT----------G-----T------GA--A------A.....................--T-----------CTGTAGA---------T-TT--- 8682
CPZ.CM.01.SIVcpzCAM13      -------------T--G--T-------G-A---G-----A---G-----A--T-A-----C---G-T--T--T----------T---------C-G-----------C-----A-----A.....................-AG-----C----CCTG--------------TT--A 8867
CPZ.CM.05.SIVcpzEK505      ----------------G--TGG--AACC-AG-------AG--------A--T--A--C--C---G-T--T--T--T----AAT----------C-G-------------T-----A-----C.....................--------C--C-------------------T---- 8735
CPZ.CM.05.SIVcpzMB66       ----------------G--T-----A---AG-------AG--------G-----------T---G----T--T----------A-T-----------G--A--------------------------.....................A---------------------C-------T---A 8873
CPZ.CM.98.CAM3             -------------------T-----AT--NNNNNNNNNNNNNNN-G----ACG--TA---T--------G----T-----------G--------------------------A.....................--------C-----A--A-A---GG------TT--A 8671
CPZ.GA.88.GAB2             -------------T--G--T-T---A-G----------G-----G--C--C---G-T--T--T--C-----A--------------G-A-T-----T--T--T--A-----A.....................--GC-T-----GT------A-----------T---- 8639
CPZ.CM.98.CAM5             -------------T--G--T-----AC--A---G-----G---G-----A-----C---G-T--T---------G--AATT--------C-G-----T--------------------A.....................--------C-----A--A-------------T---- 8974
CPZ.CM.05.SIVcpzMT145      -------------T----CTGG--AAGG-A------G--G---------T-A---G-T--T-----------T-----------C-G--A--------C--G-------A.....................--GC-C--C-----A-----C--G-------TT--A 8761
CPZ.TZ.00.TAN1             ----------T------T--TT---T-C--A----GC--CC-----A--CACC-----G--TA-T-----G--TGT---T--A--C-G-----------C--T--A-----A.....................A-------C------TGTAGG-------TA-TT--- 8914
CPZ.CM.05.SIVcpzLB7        -------------T--G--T----TA---AG--------A--------C--CG--G-T--T-----------AAT---------C-G-----T-----C----------------.....................A-T----------A-A--------CG------T---- 8760
CPZ.US.85.CPZUS            -------------T--G--T-T---AG------------G-----A-----C---G-T----T-------AAT---------C-G-----T-----C--------T--A.....................--A--------C--A--A-A---G-------TT--- 9267
CPZ.GA.88.GAB1             -------------T--G--T-----AGG-----------G-------C--C---G----T------------T---------C-G------------A----A-----A.....................ACA----TC-----CTGT-------------T---A 9301
MON.CM.99.L1               -A-------G--A-A-GT--T---GTG--GAC-----TAGA-AGA-A--G-AA-ATGC---T-TGGA-TGG----TAA----A-GA-G--TAC-A--T--GGA----------.....................-----G--C--GACT-TGCCC--G-TCCTG-GGTGC 9083
MON.NG.x.NG1               -A--------TAA---GT--T---GTGTWGACC-Y--TC-A-AGT-A--G-AA-ATGC---TGT--TGTGG--ACTTG-G----GA-G--TAC-G-TT--CY-C--A-----C.....................--GC-G--C---ACC-T-CC---G-TCCTG-GGTGC 7720
GSN.CM.99.CN166            -A--------G--CA-GT--T---GTG-GGAC--G--CC-G-AGT-G--A-AA-ATGC---T-TGGAGTGG---CTGG-G----GA-G--TCTCA-T---T-----C-----A.....................AC-C-C-----TAC--TACCA----TC---AT-TGT 9048
GSN.CM.99.CN71             -A--------G-CCA-GT--T-T-GTG--GAT-----CC-A-AG--A--GAGC-ATTG---TT-TGAGTGG---ATTG-G----GA-G--TCC-G-GG--T-----A-----A.....................A--C-C--C---ACT-TGCCA--C-TC----GGTGC 9045
MUS_1.CM.01.1085           -A------------A-GT-CGT-AGTAC-GA---G-TCC-G-AGT-G--GACC---GC---TGTGGAGTGG--AATAA-A--A-GA-G--T--TG--T---GAT--C-------.....................ACTC-G-----CAA--TACCA--C-TCCTA-GGTGT 9054
MUS_1.CM.01.CM1239         -A------------A-GT--T-T-GTG-GGA------CC-A-AAT-G--GACT-AT-C---TGTGGACTGG--AATTG-A--A-GA-G--TAC-G-TT---GAG--C--G---.....................AC-C-G--C---ACACTGCCA--G-TCCTG-GGTGC 9023
MUS_2.CM.01.CM1246         -A-----A-----AA-GT--GTG-G-AT-GAC----TGC-A-AA--A--GACT-ATGC---TGTA-A-TC-AAAATAG----AAGA-CA-TA-TA-----GGAT--C-----C.....................AC-C-T--C----G-CAACCA--G-TCCTG-GGTGC 9119
MUS_2.CM.01.CM2500         -A------------A-GT--T--AGTG-GGA------CC-A-AAT-G--GACC-ATGCA--TGTGGAGTGG--TCTTA------GA-G--T--TG------GA---C--T--T.....................--TC-C--C--G--T-TGCCA--TGTC-TG-T-TGC 9075
RCM.NG.x.NG411             --------------A-TTA-TT---AG----G---AGGCGC--TT-A--C--C-AC--GG-GA-TGA---T--AATAA-A--A----G------T-----T-------------.....................A-A---------ACA-TG-----C---CTA-GGC-A 8845
RCM.GA.x.GAB1              --------------A-TTA-TGGAGTATC---------AGA--AT-G---A---ACC-GG-AA-TGAG--T---ATAA-A-------G---A--------T--------A.....................A-A---------ACACTG---------CT--GGC-A 8762
DEN.CD.x.CD1               -A------C-G-ATA-TT--T--AGT-C------G--T-C---TT-A--C--C-ATGC-C-GA-TGA---T-----T-A-AA-A-GA-G-----C-------GAT--------C.....................A-TC-G--C--G--AGAG-----G-TCCT--GG--A 9222
                           3' LTR U3 start
MAC.US.x.239               --------------A-TTA-T--AGTGC---------TAGA---T-A--CA-A-ACT-AG-AA-GGA-G-----AT-A-A--A----G----G-T-----CT----A-----A.....................A-T-----C---AA---A-----C---CTA-GG--A 9604
Nef                        G  G  L  E  G  I  Y  Y  S  A  R  R  H  R  I  L  D  I  Y  L  E  K  E  E  G  I  I  P  D  W  Q  D  Y  T  S  G  P  G  . . . . . . .  I  R  Y  P  K  T  F  G  W  L  W  K
H2A.GW.x.ALI               --------------A-GT--T--AGTG-G-------TAGA---T-A--CA-A-ACT-AG-AA-GGA-G----GATAA-T--A----G---------T--T-AT--------A.....................A-A--G--C--GA--TT-------G---CTG-GG--- 9646
H2A.DE.x.BEN               ---------C-----A-GT--T--AGTAGG--G-----TAGA-----A--CA-A-ACC-AG-AA-AGAGG----GATAA-A--A----G------T--T--T-AT--------A.....................--A--G--C---A--TA---C--G---CTG-GG--- 9647
H2A.SN.x.ST                ----------------GTA-T--AGTG-T--G----GTAGAG----A--CA-A-ACT-AG-AA-GGA-G-G--AATAA-TGGA--C-G---------T--T-AT--A-----A.....................--A--G------AA-TT------G----TA-GG--- 9086
H2B.GH.86.D205             --------------A-GTA-T-TAGTG-G--------CAGA--AT-A--CACA-ATT-TG-GA-TGA-G-----ATTG-GT---GC-G---A-----T--T-AT----------.....................A-A--G-----CAAATA------T---CTG-GG--- 9560
H2B.CI.x.EHO               ---A----------A-TT-CT-TAGTG-G-----G--TA-A--A--A--CACA-ACT-AG-AA-TGA-G-----ATTG-GT---GA-G---------------AT--------A.....................--A--------CAA-TT------C---CTA-GG--- 9534
H2G.CI.92.ABT96            --------------A-TTA-T--AGTG----------TAGA-----A--CA-A-AC--GG-AA-GGAG-----TATAA-T--A----G------T------G----A-----T.....................A-A---------R----G-----C---CTA-GG--A 9032
H2U.FR.96.12034            ----------T---A--T--T-T--T--G--------TAGA---T-A---A-C-ACT-AG-AA-TGAGG-G--AATAA----A----G------T-----CT----A-----A.....................ACT----GG---A--TG-----------CTT-GG--- 9128
MAC.US.x.EMBL_3            --------------A-TTA-T--AGTGC---------TAGA---T-A--CA---ACT-AG-A..........---AT-A-A--A----G----G-T-----CT----A-----A.....................A-T-----C---AA---A-----C---CTA-GG--A 8726
SMM.US.x.H9                --------------G-TTA-T--AGTATT--------TA-A--AT-A-----R-ACT-AG-AA-RGA-G----AATAA-A--A----G---A--T------G----------A.....................A-------C--TA--TT----------CT--GG--A 9087
SMM.US.x.PGM53             --------------A-TTA-T--AGTG-G--------TAGA---T-A--CA-A-ACT-AG-AA-GGA-G-----AT-AC---A----G------T------T----A-----A.....................A-------C---A--TT------C---CTA-GG--A 9541
SMM.US.x.SIVsmH635F_L3     --------------A-TTA-T--AATG-G-A------TAGA--AT-A---A---AC--GG-AA-GGA-G----AATAA-A--A----G---A--T-----TT---------A.....................ACT-----C--TA--TA------G---CT--GG--A 9626
SMM.SL.92.SL92B            ----------T---A-TTA-T--AGTG-G--G-----TAGA-----A---A-A-ACC-AG-AA-AGAGG----TAT-A----A----G------T--T---AGT--------A.....................A-A--G--C---AA-TT------G---CT--GGC-- 9055
STM.US.x.STM               --------------A-TTA-T--AGTG-G--------TAGA--AT-G--CA---ACT-AG-AA-AGA-G----GATAG-T--A----G------T------G----A-------.....................A-A-----C---AA-CAG----------CTA-GG--A 9255
MNE.US.x.MNE027            --------------A-TTA-T--AGTG----------TA-A---T-A--CA-A-ACT-AG-AA-GGA-G-----AT-G-G--A----G----G-T-----CT-G--A-----A.....................CCT-----C---AA---A-----C---CTA-GG--A 9081
DRL.x.x.FAO                --------A-----A-TTGGTGG--T-TC------G--ACA---T-G-----C-ATGCTC-AA-TGA-TGG----T-A-AAAA-GC-G---A-GT-----CAAG--A--T--C.....................A----G--C---AA--TG-----G-TCCTG-T---- 8940
MND_2.x.x.5440             --------------A-TTGGTGG--TA--------G----G---T-G--C--T-ATGCAC-AA--GAGTGG--ACTAA-A---A---G-------------AAG--A--G---.....................A-A--G-----GAA---T--C--G-TC-TG-T---- 8855
MND_2.CM.98.CM16           --------------A-TTGG--T--TA--------G----G---T-G--CT-A-ATGCTC-AA-TGA-TGG--A-T-A-T---ACC-G---ATCT--T--TGAT--C-------.....................A-A-----C--GAA---A---A-G-TC-TA-T---- 9159
MND_2.GA.x.M14             --------------A-TTGGTT---TAG-------G----G---T-AA-C----ATGCTC-GA-TGA-TGG--A-T-A-A--A----G---AG-A-----TT------------.....................A-A-----C--GAA--GA-----G-TC-TA-T---- 9100
OLC.CI.97.97CI12           ----------G---A-GTA-TGG--A-CT-----GG-GA--T-GA--T------ATGCT-TGA-TTGTTGG--AATAA-AGG-A-C-TC---GCAACAGAGGAG-TAGA--A-...............GGCACGT-A--GAGA--TAA---GCCA--G----TA-GGG-T 8633
GRV.ET.x.GRI_677           ----------T---A--TA-T----TG--------G--A-G-----GA--T---ATGC--TGA--GAGTGG--AATAA-AGA-----G---AGCT---T-------C--G---.....................A-A--G--C--GAGAGT------C-TC----TT--- 9079
VER.KE.x.TYO1_patent       -A--------T---A-TTA-T-----G-T--G---A-TA-G-----GA------ATGCTCTTA-TGA-TGG--GATAA-TGA-----G-A-TGC--GGT-GAAG--A-----A.....................A-A----TC--TAAATG------G-TC----TT--- 9196
VER.KE.x.9063              -A--------T---A-TTA-T-----G----G--GA-TAGA-----CA-C----ATGCTCTTA-TGA-TGG--AAT-A-TGA-----G-A-CGC--GGT--A----A-----C.....................--A--------C-GCTG------C-TT----T---- 9235
VER.KE.x.AGM155            -A--------T---A-TTA-T-----G-T--G---A-TA-G------A------ATGC-CT-A-TGA-TGG--AAT-A-TGA-----G-A-CGCA-GGT--AA---A--T---.....................A-A-----C--GAG-TG------C-TC----T---- 9212
VER.DE.x.AGM3              ----------T---A-TTA-T-----G-C--G--GA-TC-G-----AA-C----ACGC-CT-A-TGAGTGG--AAT-A-TGA-----G-A-TGCT-GGT--GA---A-----A.....................A-------C---AGATG---C--C-TC----TT--A 8733
TAN.UG.x.TAN1              -A--------G--CA--TA-TGG--A-C--AG---G-GC-A-----GA----C-ATGC-CTAA-TGA-TGG---ATTA-AGA-----G---AGCA--TT-C--T--------C.....................AC---GA-G--C-GATG---C--C-TT----TTG-- 9222
SAB.SN.x.SAB1C             --------------A-TTA-T--AGTG----G---A-GA-G---T-G-----C-ATGCTCTTA-TGAGTGG---ATAG-AGA--GA-G-------T--T---GAT--------T.....................AC---G--C---AA-TG--------------TT--A 9412
MND_1.GA.x.MNDGB1          ---------G----A--TA-T-TAG-AGG------G----A-----A-----C-ATGCAG-GA-TGAGTGG--A-TTGAA----GA-G---AC-G--T--GA----T-----A.....................AC---------TAA---A-------TCCTG-TT--- 8664
DEB.CM.99.CM40             -T--------G-ATG-TT--T----TGCC-----G--T-C---TT-A--G--TCATGC-C-GA-TGA-TGG--AATTA-A--A-GC-G--TAC-G------AAG--A--T--A.....................--T--G------A-TTA---C--G-TCCT--TT--A 8869
DEB.CM.99.CM5              -T--------G-ATA-TT--T----TGCT--------T-C---AT-A--AA-CCATGC-C----TGA-TGG--AATAA-A--A-G--G-TT---G------GA---C-----A.....................CCA--G------ACATA------G-TCCTA-TT--- 8796
TAL.CM.01.8023             --------A-G---A--T--T--AA-A-C-----G--T-CA---T-AA-C----ATGC-C-TA-TGAGTGG---ATAC-G--A----G----TCA------GA---------A.....................ACA--G--C---T-ATG---C--TATC-TG-GG--- 8542
TAL.CM.00.266              -A------A-G---A--T--T--AA-A--C-T--G--T-CA---T-GA------ATGC-C--A-TGA-TGG---ATAC-G--A--C-G---------------GAG--A-----A.....................ACA-----C---T--TG------TATCCTG-GG--A 8992
SUN.GA.98.L14              -T---------------CTGGT-----AG--CC---G----A---T-G---T---ATGCAG-AA-TGAGTGG----T-A-AA---GC-G---AG-T-----CAA------T--A.....................--G---------AA-G--------C---CTA-GG--- 9380
WRC.CI.97.97CI14           -A--------G-ATA--T--TGG--AGCG------G----A--AG-GA-C--C-ATGCAC--A-TGA-TGG--GCTGA-AGA--GC-G---AGCT------GA---A-------.....................A-A----GG---AA---A-----G-------GG--A 9179
WRC.CI.98.98CI04           -A--------GTATA--T--TGG--T-CT------G----A--AG-GA-C--T-ATGC--T-A-TGA-TGG--ACTGA-AGAA-G--G----GCA------GAG--A-----C.....................A-AC-C-GG---AA---A-----C-----T-GG--A 9221
WRC.GM.05.Pbt_05GM_X02     -A--------G-ATA--T--TGGAGT-CT------G----A--AG-GA-CT-A-ATGC---T--TGA-TGG--TATAG-GGAG-GA-G---AGC------CGAT--A-------.....................---C-C--C---AA---A-----C-------GG--A 8117
LST.CD.88.524              -T---------------TA-TGG-----G----GA-T--C---T-AACA--C-ACTGTG-GA-TGA-TGG--TCT-A-TGGA--C-TCAT----------GAT--A-----A.....................ACTC-T--C-----CG-----------CTG-GGC-- 8234
LST.CD.88.447              -T---------------TA-TGG-----G--------G--C---T-ATG---T-ACTGTG-GA-TGA-TGG--TCT-A-TGGA--C-TCAT--CT------GAT----------.....................ACGC-T-TC---T-AG-A----------TG-GGC-- 8234
LST.KE.x.lho7              -T---------------TA-TGG-----G--------G--C---T-AAT---C-ACTGTG-GA-TGAGTGG--TCT-A-TGGA--C-TCAT----------GAT--A-----A.....................ACA-----C--T-----A----------TG-GGC-- 9309
SYK.KE.x.SYK173            -A------C-G--CA-G-A-T-T-GTG---AG---G-T--A--T---C----C-ACC-AC-GA-TGAG--T--AAT-A-A...--CAG-ATC--T-----CT----------A.....................AC-C-G--C--GT-A-T------------TA-GGG-A 9057
SYK.KE.x.KE51              -A------C-G--CA-GT--TT--GT----AG-----T--G--T--AC-G----ATT-AA-GA-TGAG--T--TGTGA-AGA---C...ATC-C---T--GAGT--------A.....................ACA-----C--G----T---C--G----TA-GGG-A 8747
COL.CM.x.CGU1              -T--------G---A-T---AGAA---CTGAG---G-T--CC-GA-A--G-A--ATGC----ATTGAGTGG----G---AAAA-G--G-TT-G-G--TGA-GAT-A--TG--AGAGGATGGAGCATTAAAAGAG-A----A-G--T---GTTGC-------CTT-GG--A 8488
```

```
H1B.FR.83.HXB2         CTAGTACCAGTTGAGCCAGATAAGATAGAAGAGGCC...AATAAAGGAGAGAACACCAGCTTGTTACACCCTGTGAGCCTGCATGGGATGGATGAC........CCGGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCATTTCATCAC...GTGGCCCG 9383
Nef                    _L__V__P__V__E__P__D__K__I__E__E__A__.__N__K__G__E__N__T__S__L__L__H__P__V__S__L__H__G__M__D__D__.__.__.__P__E__R__E__V__L__E__W__R__F__D__S__R__L__A__F__H__H__.__V__A__R
H1A1.UG.85.U455        --------------T----C-G-AG----G--A--T...-C-GG-----------A---TC-AC--------A-AT---AA-----AG-A-----T........GAA---A----------AT-----A---------TAC---G-----AA-A---...AGA--TTA 8834
H1B.US.90.WEAU160      -----------G--A-----G---G-------A--T...---G-----------A---------------A-----------------------........-AT---A------A---AT-----A---------AAA---------------...--A----- 9383
H1C.ET.86.ETH2220      --------------C---AG-G-AG-G-----AAT-...---G-----------A-T-----C-G-------C-----------A-----G--T........GAA--C-------A---A------A---------T-A-------CGCAGA---...A------- 8783
H1D.CD.84.84ZR085      --------------T-----GGT-G-----A-----...-C-G-------AG---A-T-----C--------A-CT---A------A-----G--........GA----AG----------T------A---A-----A-A---------G-G---...AA----AA 8931
H1F1.BE.93.VI850       --------------T-----AG--G-G---A-----...---G-----------A-T-------------CA------AA-----A-----G--T........GAA--C--------AC-GAG-----A---------A----G---C-GAGA---...A-A---A- 8660
H1G.SE.93.SE6165       ---------A-G--T----CAG--G----G--A---...-----------------A---TC-A--------CA-CT---A------A-----G--........GAA--C---------C-G-T------A--------TA--------CGGAGA---...A-A----- 8825
H1H.CF.90.056          -----------AA-T---C-GG--G-----C-----...---G-----------A-------C-------CA---------------A-----G--T........GAC-G----------C-GAT-----AA--------T--A--------GACA---...T------- 8706
H1J.SE.93.SE7887       --------------T---AGCG-AG----G--A--T...---G-------A----A-T--C---------CA-AT---A------A--A-G--T........GAA--A-----------C------A----------TC-------CGGAGA---...A-A----- 8681
H1K.CM.96.MP535        --------------T---CAG-AG--------A--...-CAG--------G---A-T--C---------CA-A-A--A------A-----G--T........GAAC-C------A-A---A-A--AT----A----------TTC-------CGAAGA---...--A----- 8570
H101_AE.TH.90.CM240    --------------C-A-AGAG-AG----G----A-...--C--------A----A-T--C---------CA------A------A--A-G--........GAA--A---------C-GAT-----A---------TGC-------CGAA-A---...--A----- 8951
H102_AG.NG.x.IBNG      ---------A-G--T----CAG--G----G--A---...---G-----------A---T--A--------CA-CT-T-AA-----A-----G--........GAC--T---------C-GATC-----A---------TA-A--G-----CAGA---...ACA--TA- 8910
H1N.CM.95.YBF30        ---------T-GTCAG-T--AG-AG----------T...---G-------C----ATGC-C-C--------CA-AT-T-AA-----AGCA-----T........GATC-T-A---------G-T----C-A---------TC-------AGAAGA--T...--A--AA- 8973
H1O.BE.87.ANT70        -----------GTCAGA---AG--GC----AGACTAGGA----C-T-T----GGG-T-ATC-CC-G--T--A-CAT-TGCC-----AT-T--A--T........A-AC-T-A----A-AC-GAT-----A------T--ATCT----GCAACACC--T...--T--TAT 9473
H1O.CM.91.MVP5180      --G--------GTCAG----AG--GC---GAGACTGGGT----C-AAT--AG-TG-T--TC-TC----T--A-CTT-TAAT-----AGCT--G--T........G-AC-CG-G--GA-AC--A-A---CA------T--ATCAT---GC--AACA--T...A-A----T 9496
CPZ.CD.90.ANT          T----------A--AT--CCACCTGAT...............-AT----GG-A--TAC-CC-C-------CTT-TACA-----AGAT-GA---.........--AC-T-A---GA-C---AGA---GAA-----TGCAA----GATGAGAAGA---...A-T--AA- 8822
CPZ.CM.01.SIVcpzCAM13  --------TT-GACAA-T--GC-AG-G----CA---...---G-G-----T----A-T----AC-G------A-AT-T-A------A-----G--T........GAAAGT-A------C---ATA---C-C----------A-A--G--T--AAGA--T...A-A--AA- 9022
CPZ.CM.05.SIVcpzEK505  ---------C-GTCA--T--GG--G-------A--T...---C-------C----ATGT-C-------T--CA--T---AA------A-----A--T........--A--C-A-------C-G-T----C-C----------TC--------AGAGTA---...AGA--AA- 8890
CPZ.CM.05.SIVcpzMB66   ---------T-A--A-----AG-AG-G---AGA--A...---G-------C----A--TA--------T--CA-CT---A---C--ACAA--G--T........GAA-CA---------C---TC----CA-----------G--T---C-GA-G---...AGA--T-- 9028
CPZ.CM.98.CAM3         --G--T---C-GACA-----AG-AG-G---C-----...---G-------T--T-A--T-C-AC-G-----CA-CT---A------A-----A--T........G-AC-TG-------C--G-TC-----A----------A-G--G--C--AAGA---...A-A---A- 8826
CPZ.GA.88.GAB2         -----G-----ATCA-----GGCAG-------A---...--C--G------G-T-ATCT-C-A-----T--CC--T--ACT-----AT-T--G--T........GAA----AG--G--TC-GATA----A---------T-AA-----CC-G-GG--T...C-T--AA- 8794
CPZ.CM.98.CAM5         --G-----TC-GACA-----GG-AG-G--GA-A---...--CG-------C-C--AT-TAC--C-T-----CA-AT---AA-----A-----G--T........GAAC-TG-------C----TC-----A----------A-G--G--TC-GA-A---...C-A---A- 9129
CPZ.CM.05.SIVcpzMT145  --G--C--TT-GCCA--T--GG--G-G--GA-A---...---G-------A----A-T--C--C----T--CA--T---A------A-----G--T........GA---A--------AC-G-TA----CT-----T---AAG--G--TC-CAGA--T...C-A--AA- 8916
CPZ.TZ.00.TAN1         T-G-----G--A--C...............CCAC-A...G-AG-T-AT-----G-A--T----C----T--A-CCT-TAGC-----A-CTACC--T........--A--TG----GACTC-GATC---C-C----------A---------AGAAGG---...A-A---A- 9054
CPZ.CM.05.SIVcpzLB7    ---------C-AACC-----AG--G-G--G-----T...-----------------ATCTAC-AC-G-----AA-AT---A------A-----A--T........GAA--C-A-------C-GATA----AA-A---------AG--C--CC-AAGA---...A-A--T-- 8915
CPZ.US.85.CPZUS        T----C--TC-CACAGA---GG-AG----GC-----...---------------C--A--TAC--C-G-----CA--T---A------A-----A--T........GAAC-TG-C-----C---ATC---CA--------CTGAA--G--TCGGAGA---...AGA--TAA 9422
CPZ.GA.88.GAB1         -----G--CC-GACAGAG--GC--G-----C-A---...---G-------T----A-T--C-----G--T--CA-AT-T-A------------A--T........GA---C-A---G--C--G-TC---C-C----------A-G--G--C--AAGA--T...A-T---A- 9456
MON.CM.99.L1           T-GCGC--G--A-CC...ACC-C-GAG--TTCC-AG...G-GGG--AT---G--TT-CT-C-AACC--T--G-CCTAT-AAGGCC-------G---.........--CC-T---C-GT-C--G-TC-TCTCA--CTG-TC-AAG--G---G-CA-GAG-GGCAGACAG-T 9238
MON.NG.x.NG1           --GCGT--G--A-CC...ATCC--GA---CTC--AG...G-CGG--AT--YG-GTT-CT-C--ACC-----G-CCTA--A-GG-C--GAT--A---.........--CC-T-AGC-GT-C--G-TC-TYTCC--CTGTTCTAAG--G-G-G-CA-GAG-GGCCCTCAG-T 7875
GSN.CM.99.CN166        --GCGG-----A-CC...ACA-CTGAG--CTCT-AG...CC-GGG-AT--TG-ATATCTGC--ACC--------CCTAT-A-GGAC---GT--A---.........-A-C-T-AG--GT-C--G-TC-TCTCC--CTG-TC-AAG--G--CA-AA-GTCAGGCA-ACAG-T 9203
GSN.CM.99.CN71         T--CGG-----G-CC...ATG-CTGAG--CTCT-AG...CCCGGG-AT--TC-ATATCTG---AAC--T--C-CCTAT-A-GGACA-CAA--G---.........-ACC-T--G--GA-C--G-TC-TCTCC--CTG-TCTA-G-----C--GA-ATCAGGCTG-CAGAT 9200
MUS_1.CM.01.1085       --GAG-GA---G-CC...ATG-C-GAG-GCTCT-AG...G-GGG--AT--CC-ATA-CT-C-AGAC-------CCTTT-A-GGGA-AGAA--A---.........-ACC-T-AGC-GT-CC-G-TC-TCTCC--CTG-TC----T-G--CC-GA-GTCAGGCT-ACAA-T 9209
MUS_1.CM.01.CM1239      ---AGGGA---C-CC...ATA-CTGAG--CTCT-AG...G-GGGG-AT---G-ATATCT-C-CACC--TG-A-CCTA--AAGGCAA-GA---G---.........--CC-T-AGC-GT-C--G-TC-TCTCA--CTG-TCT-----C--CA-GA-GTCAGGCAG-CAATT 9178
MUS_2.CM.01.CM1246     T-GAG---G--A-CC...ATG-CAGAG--GTC--AG...CCAGG--AT--CC-GTA-CT-C--AC------G-CCTATG--GG-C--GAT--A---.........-ACC-C-A---GT-CC-G-TC-TCTCC---TG-TCTA-G--G--CC-AA-GTCAGGCCC-CAG-T 9274
MUS_2.CM.01.CM2500     T--AGG--T--GC-C...ATA-CTGAGA-CTCA-AA...CTGGGG-AT--TG-ACATCT-C---GC--T--C-CCTT-A-TGGGA--GAA--A---.........--CC-T-AG--GT-CC-G-TC-TCTCC--CTG-TC------G--TC-CA-GTCAGGCCGCCAGTT 9230
RCM.NG.x.NG411         T----------A--T...-TC-GTGAT----CAA-A...G-G--T...-----GGCATT-TC-CC-G--T--A-CAGAAACTAG---------C---.........--CTG-G-G--G--A----CA----A----A-ACC--T-T-----G-AG--T--...-CA-GGTA 8994
RCM.GA.x.GAB1          T-G--G-----A--T...-TATCTGAT----CAAGA...G-AG-T...--AG-GCAT--T---C----T--A-CAGAAACAAG---------A---.........--ATG-G-G--G--C--G-CC----A----A-TCCTATG--G---G-AG--T--...A-A-G-TA 8911
DEN.CD.x.CD1           T-G--C--G--GACT...ATAG--GAG---T-T-A-...-----G...---G-A-ATT-TC--C-G--TGACAGATATGA-GG-CAACAA-C----.........--TTG------GACCC-G-TA-----A------CCTGAAT----C-GGTG-T-T...AAA...GC 9368

MAC.US.x.239           T----C--T--AA-T...-TATCAGAT--G-CACAG...G-GG-T...---G-GCATTAT--AA-G--T--A-CTCAAACTTCCCA-TG-------.........--TTG-G----G--TC---CA----A------TCCAACT--G--C-ACAC-T-T...-A---ATA 9753
Nef                    _L__V__P__V__N__.__V__S__D__E__A__Q__.__E__D__.__E__E__H__Y__L__M__H__P__A__Q__T__S__Q__W__D__D__.__.__.__P__W__G__E__V__L__A__W__K__F__D__P__T__L__A__Y__T__Y__.__E__A__Y
H2A.GW.x.ALI           -----------A--T...-TCCCACA-----G--AG...G-C-CT...----CTCA-T--C--C-------A--ACAAACAAGCA--CAT-------.........A--C-TG-G--GACA----TT-----A------CCTAAG--G--TCA-G--T--...AAA---TT 9795
H2A.DE.x.BEN           ------T----A--A...CTCTCACA---G-CA-AG...G-AG-T...----GC--A-T----AG-------A-CACAAACAAGCA-ACAT-----T.........GA-C-TG-G--GACA----T----CA-------TC-ATG--G--C-A-A-CT--...AA----TT 9796
H2A.SN.x.ST            -----------A--T...-TCCCACA---G-GA-AT...G-C-GT...----CTCA-T----AG-G--T--A-CACAAACAAGCA--T-T-------.........---C-TG-----ACA----TT-----------CC-ACG-----T---AGCT--...-A----TT 9235
H2B.GH.86.D205         --G--------A---...-TGCCAGC--CGACCCGA...G-GG-G-AG--A-C-CATT--C-AA-G-----G-CACAGA-CTCCTCATG-------.........ATCC-TG-G--GACTC-TATC---CA------TTC--T---G----A-G--T-T...-----TTT 9712
H2B.CI.x.EHO           -----G---A-AA-T...ATG-TAGC----CCA-AG...G-CG-G...--A-C-CATT-TC--G-G--T--A-CACAGACCTCC-CATG-------.........--CC-CGA---GACCC-T-TC---CA-------TC--T--------A-G-CT-T...------TT 9683
H2G.CI.92.ABT96        -----------CAC-...-TCTCTGAT--G-CACAG...G-GG-C...----CTCATT-----G----T--A-CACAAACTAGCCCATG-------.........-------CC--------G-C-----A-------CCTACAT----C-A-G-CT--...AG---ATT 9181
H2U.FR.96.12034        ----A------A--T...-CA--TGATA---CACAG...G-GG-T...----GGCTGTATC--G-GGGTT-A-CACAGACCAGCT-CGA---G---.........-ACTG-G----G-CC----TA----A------TTC-TCT--G----A-AGCT-T...CA---ATT 9277
MAC.US.x.EMBL_3        T----C--T--AA-T...-TATCAGAT--G-CACAG...G-GG-T...----GGCATTAT--AA-G--G--A-CTCAAACTTCCAA-TG-------.........--TTG-.----G--TC-GAC-----A------TCCAACT-----C-ACAC-T-T...-A---ATA 8874
SMM.US.x.H9            T---WC--G--AA-T...-TCTCAGAT----CTCAG...G-AG-C...----CACATT-----A-G--T--A-CACAAACTTC-CA-TG-------.........--CTG-G----G--AC-G-CA----A------TCCARAGT----T-A-R-CT-T...AA---ATT 9236
SMM.US.x.PGM53         T-G--C-----AAGT...-CCTCAGAT----CTCAA...G-AG-C...----CACATTA--AG-G--T--A-CTCAGA-ATCCCA-TG------T.........--CTG-G----G--TC-G-CA----A------TTCA-AA-----C-ACAGAT-T...-A---TTT 9690
SMM.US.x.SIVsmH635F_L3 T----C-----A--T...-TCTCAGAT----CCCAG...G-AG-C...----CACATT--C--G-G--T--A-CACAGACT---CA-TG-------.........--CTG-G----G--AC-G-CA----A------TCCAGAAT-G--T-A-AGCT-T...AA---ATT 9775
SMM.SL.92.SL92B        --G--G-----A--C...-TCTC-GAT--G-CAAA-...---G-T...----C-CATT-----G----T--A-CCCAGACAT--CA-CATTCA---.........--TTG-G--A-G--CC----CC----A----A-TCCT-A------C-ACACAT-T...-AA--ATT 9204
STM.US.x.STM           -----------A--T...ATGTCA-AT--G-CACAG...G-GG-T-AT-G--CACATTATC--G-G--T--A-CACAGACA---CA-TG-------.........--TTG-G-------AC-G-TT----AA-----TCCA-TAT----CCA-AC-T-T...-A----TT 9407
MNE.US.x.MNE027        T----C--T--AA-T...-TATCAGAT--G-CACAG...G-GGGT...---G-G-ATTAT--AC-G--T--A-CTCAAACTTCCCA-TG-------.........--TTG-G----G--TC---TA----A------TCCAACT-----C-ACAC-T-T...-A---ATA 9230
DRL.x.x.FAO            --G--C--T--G---...-TAG-TGA---TCT--TA...---C-G...C--TGT-A---GC----GA--T-CAGCCAGA--GG-CCT--T------.........--A--AG-----AGAC-CAT----CA------TCCAG----G----A-ACATTT...CA---TT- 9089
MND_2.x.x.5440         --G--C--G--G-CA...-TGTC-CCCCCCCT--AG...G-GG-T...--ATG--A---GC--C--A--T--TCTCAAA--GGAATCCA---A---.........--A---G----GAG---GAT-----A------TTCAG-G--T--T-A-ACGTT-...TAT--T-C 9004
MND_2.CM.98.CM16       --TTGC--G--G-CA...-TGCCTCC---CC---AG...--C--T...---TGT-A--AA---C-C--GT-CTCTCAG---GGAATCCA---A--A........T--TTG-G----GAG-C----T-----A-------TCTG----T--T-ACACCTTT...TAT----C 9309
MND_2.GA.x.M14         T--A-C-----AC-A...-TACCACC---CC---AG...--CC-G...---TGT-A---GC--C-TA--T-CTCTCAG--AGGAATCCA---A--A........T--ATG-G-G--GAG-C-GAT-----A-------CCTGAAT-G--T-A-ACATTT...TAT----C 9250
OLC.CI.97.97CI12       --TAGGAG---G--A---TTGCCTGAGACCCTT-AAGAG-GAGGG.................................................................................................................................... 8678
GRV.ET.x.GRI_677       -----C-----G--C...CTGC-TGAG--G-CACG-...--CTGT...----GACA-T-TC--A-G--T--A-CACAGA--GGG-AAGATCC---TGGAATA...GATC-TG-------C--G-TC----A-------CCGAAGT-G--GG-GG-GT--........... 9226
VER.KE.x.TYO1_patent   -----G-----G--C...TTAC-TGAG---CACAA...-CATGT...--A-GACATT--C-AG-C--T--A-C-CAGA--GGA-AAGATCCA--TGGTATC...AGCC-TG----GA-C--G-T-----A------TCCTATGT-G---A-A--GT--.......... 9343
VER.KE.x.9063          --------G--A---...ATGC-TGA---G-CA-AG...-CCTGT...----GACATTA----GCG--T----CACAAA-AA-A-A-GACCCC--TGGCATC...AGTC-TG----GACC--G-TC----A-------CCTATGT-----G-G--GT-T.......... 9382
VER.KE.x.AGM155        --------G----CC...CTGC-TGAG----CA-AA...-CATGT...--A-GGCATT-----G-------A-CACAA-------AAGACCC---TGGTATA...AATC-TG-----A-A--G-CA----A------TCCAATGT-G--TG----GT--.......... 9359
VER.DE.x.AGM3          T-G--------A--C...CTGC-TGAG----CA-AG...-C-TGT...----GACATT--C--G-G--T--A-CACAAG--AGG-AAGACCC---TGGAATC...AACC-TG-------C--G-TC----A------TCC-ATGT-----G-C--AT-T...-ACC-TAA 8888
TAN.UG.x.TAN1          T----G--G--A--T...-TCTCCCAG----CCCAG...G--.........----GACA-T-T---C-C-------CCCAGA-TG-AT--GAAAGC---.........--TTG--A---GACA--G-T-----AA---A-TCCA-T-------G----GT--.......... 9360
SAB.SN.x.SAB1C         --G-----T--G--C...CTGTCAGAG--G-CTAAG...----GT...------CA-T--C--C-G--T----CTCAAG--GCCTATGA-------.........G-CTG--A---GACCC-G-T-----A-------CCA-TA------G-GG-CT-T...--T--TT- 9561
MND_1.GA.x.MNDGB1      --G-A------GAGCAG--C--TAGG---T---TAT...GCAGCT..........----ATCATC--------T-CTCCCAGT-ATG-CCTCA---A--T.........--A--AG---GACCC-CAT----TCTGGGAC-CT-AT-T-GC-TA-GAC-TTG...CA-CATTA 8813
DEB.CM.99.CM40         --G--G-----A--A...ATAGCAGACCC---TTAT...G-G--T-AT----GG-A--T---------TGA--CCCA--A-GGGAT------A---.........--CC-T-----GAG---G-TT----A-----------ACT--G--C-A-TG-T-T...AA---AG- 9021
DEB.CM.99.CM5          -----G-----A---...ATAGCAGATCCT--CTAT...G-G--T-AT----GA-AT-T-C-C-----TGA--CCCA--AAGGACAAGCA--A---.........--CC-T-AG--GAG-C---TA----A------TTC-TC---T--C-A-TGCT-T...AAA--AG- 8948
TAL.CM.01.8023         --GTGC--C--G---...ATCC-CGAG--C--CA-G...G-GG-C...-GAC--...CT-C-CA-G-----A-CATATGACGG-CA-CA---A---.........--CTG-G-C----CCC-G-TC---GTC-----TGAGAAG--G--C-ACACC-CAGGA-C-AAAAT 8691
TAL.CM.00.266          --CTGC-----A---...ATCC-TGAG--T--CA-G...G-GG-C...-GAC-T...CT-C-CCCG--T--A-CATATGATGGGCA-GCA--A---.........--ATG-G-----TCCC-G-TC---GTC-----TGAGAAG--G--C-ACACC-C-GGA-C-AAGAT 9141
SUN.GA.98.L14          ----C------CACC...AT-G-TGA---CAG--AT...CC---TCATCCTTGTCAAGC-C-AC-G--TT-CAGCCAG-AAGGG-TT-AT--A---.........--TTG-G----GAGAC-TATC----CA-----TCC-ACT--G--C-A-G--TTT...AG----AT 9532
WRC.CI.97.97CI14       T-G--T-----AC-C...TTC-GTGA----TCA-AT...-----C-...--CCCATA-TATGA-AG-A-TAAGT-ACTG-AT-CA-CATGT-C---GCCAGAGGAT--CTG-G-----CATC-G-TC----A-------CCTAAG--G--C-G-G--TTTGTA-CA-GAA- 9340
WRC.CI.98.98CI04       T----G--T--G--A...TT--GTGA---GTCTAAT...G-C-GC...TCTC-GTA-TA-GA-AAGA--AAGC--TTA-AT-CA--CTGCC-A--ATCAGAGGAT--TTG-G-----CAC--G-TC----A-------CCT--A-----C-G-G-CTTTGTG-CAAGTA- 9382
WRC.GM.05.Pbt_05GM_X02 T-G--T-----GC-C...TTC-GTGA---GTCTC-T...--CTC-...--CC-AGG-TTTGA-AAGA---TGC--CTA-AT-CA-CAG-A-C---GTCAGAAGAT--ATG--AG---TAC--G-TC----A------TCCTA-G--G--T-G-G--TTTGTG-CT-GGA- 8278
LST.CD.88.524          T-G-AG-----A-CA...TG-G-TGAGTTTA-A-AT...CCCTC--AT----GACAAT-T---C----TT-AAGCCAGT--GGG-TCC-A--A---.........--CTG-G-----AG-C-CATC---CAC---A-TCC-ATGT-----G-GG-CT-T...--A----T 8386
LST.CD.88.447          --G-A------G-CC...TG-G-TGAGT-CACA-AT...CCCTC--ATT-C-GACAGT--C--C-G---T-AAGTCAG---GGG-TCCA---G---.........--CTG-G-----AG-C-TATC---CAC---A-TCCTATG------G-GG-CT-T...--A----T 8386
LST.KE.x.lho7          T---A------G-CT...TG-G-TGAGT-TA-A-AT...CCCTC--AT----CACAAT--C-T--G--TT-AAGCCAG---GGG-TCC----A---.........--CTG-G----GAG-C-TATC---CAC--CA-TCCAATGT-G---G-GG-CTTT...A-A----T 9461
SYK.KE.x.SYK173        -----T---AA-----...ATAG--GG-TGCTT--AG...T--G-G...---C-T--ACT-C-CC-C--T--A-CCTCTGGA--G----GTTCGAG-ATGGGAGAA--TC-TGTG---C-TCA-CCCCCACCTGGCT-T-CA-CAGGTTGGGAGATGGC-...CGCTTG-A 9215
SYK.KE.x.KE51          T-G-CC--TAA---A...ATAG-AGGGT-CTT--TA...G--G-G...--AG----ACTGA--A-G-----A-CAGCAGGAGTG---GCTTCA--AGAC.......--AC-T--G--GAATC-CAT-----AC--CA-TCCA-A-T----C-A-ACA-CAGGCTG--AAAT 8902
COL.CM.x.CGU1          T----CTACA-A---...C-ACTT.............................................................................................G-AGAGTAT----ACAGCT-TGACT--AGT-T 8541
```

```
                                   TCF-1 alpha binding
                                                    Nef end
H1B.FR.83.HXB2           AGAGCTGCATCCGGAGTACTTCAAGAACTGC...TGA.....................................................................................CATCGAGCTTGCTACAA......... 9434
Nef                       _E__L__H__P__E__Y__F__K__N__C__.__*_
H1A1.UG.85.U455          T---------------T--AT--AG--...---....................................................................AAGACTGCTGA--CA--AG-----GAC-......... 8893
H1B.US.90.WEAU160        --------------------T--AG----...---...................................................................................-------T--T-----.......... 9434
H1C.ET.86.ETH2220        C-----A-----------T-A---AG----...---...................................................................................--CA--CGGGA--TTCCGCC...... 8837
H1D.CD.84.84ZR085        ---AAA-T-----------------------...---...................................................................................---------C................ 8975
H1F1.BE.93.VI850         ----AGA---------T--A-C-AG--....---.....................................................GACTGCTGACACAGAGATTGCTGA--CA--AGAAT---A--.......... 8732
H1G.SE.93.SE6165         -------------------A---AG----...---...................................................................................--CA--AG-----GAC-AG....... 8878
H1H.CF.90.056            --TAAA-------------...--AG----...---...................................................................................--CA--AGA-.--GA--.......... 8753
H1J.SE.93.SE7887         ----------------TT-A---AG----...---...................................................................................--AA--AG--T---GCG......... 8732
H1K.CM.96.MP535          --------------C----A---AG----...---.............................................................................................................. 8604
H101_AE.TH.90.CM240      ---A-A------A-------AT--AG----...---...................................................................................--AG--AG--T--.ACT......... 9001
H102_AG.NG.x.IBNG        -------------------A---AG----...---...................................................................................--CA--AG-----GAC-......... 8961
H1N.CM.95.YBF30          ----------------TT-A----------...---...................................................................................--AG-GA---TACTGCTGACAA.... 9029
H1O.BE.87.ANT70          GATAACT--C--A---CT----C----GGA-...-A-..............................................................AAACTGCTGACCT--AGA-TGCTG-CACTGT.... 9538
H1O.CM.91.MVP5180        GC-AAA---C--A---CT----CCC--G......-A-..........................................................................CTGACACTGC.......... 9537
CPZ.CD.90.ANT            ---AAGA--C--T--A--T----GAG--......-A-..................................GACTGCTAACAGGCGGATATAAAACCGCTGACAG-G-AGA----TGGTCT.......... 8903
CPZ.CM.01.SIVcpzCAM13    ----AAA-----------T-A---AG--.......---.........................................................CTGGCGCATGCG-C-AAGAAG--..CTG-CTGTGC... 9086
CPZ.CM.05.SIVcpzEK505    ----------------T--A-C----------...---...................................................................................--A......................... 8927
CPZ.CM.05.SIVcpzMB66     ------A------------A---AG-T---...---...................................................................CAA-GGACTTG--..CTG.............. 9078
CPZ.CM.98.CAM3           ---A-AA-----------T-A-C-----.......---.........................................................TGCCCAGCGCCTGCG-AGTAGAAA.CTGCTGACAA... 8891
CPZ.GA.88.GAB2           ---AAAA-----T-----T-A--GAG--.......---...............................................AACCGCAACTGGCGCATGCG-TCT..A-AACTGCTGAGCCTGA 8866
CPZ.CM.98.CAM5           ---A-AA-----------T-ATGGAG--.......---..................................................CACAACCCAGCGCCTGCG-AGTAGAA..CTGCTGACAA... 9193
CPZ.CM.05.SIVcpzMT145    ----AAA--------A--T-A--GAG--.......---...................................................CCACAGCGGCGCATGCG-ACT.AA-AACTGCTGATCAACA 8986
CPZ.TZ.00.TAN1           ---AAGAT---------------AT-A........................GGACTTCCGGGTGCCATGACTCAGAACTGC-GACAGAG-ACTTTTGGACTC... 9133
CPZ.CM.05.SIVcpzLB7      ----------T---------A-----------...---...................................................................................--A......................... 8952
CPZ.US.85.CPZUS          ------------A---------CG---.......---................................................TGGCTGTAACCGCGCAGGCG-A-TAAAA..CTGCTGACTG... 9491
CPZ.GA.88.GAB1           ---A-AA------------A---AG--........---................................................CTTTAGACTGGCGCATGCG-ACAAGAA..CTGCTGACTCTGC 9527
MON.CM.99.L1             --CA-A--T--A-C-AG-GGAGCG---GAAGCGC-T-GCCGCAAACCGCATCCTCTAG..........................................................-ACTCTGCCC... 9306
MON.NG.x.NG1             ---TGA-TTA-A-C-AG-AGAACG---GA-GCGC-T-ACCGCAAACCGCATCCTATAG.....................AACTCTGCCC.................................... 7943
GSN.CM.99.CN166          ---C-A--TG-A-C--G-GGAGCG---GATGCGCCT-ACCGCAAACCGCTTCCTATAG..........................................................-ACTCTGCCC... 9271
GSN.CM.99.CN71           GA-C-A--TG-A-C--G-GGAGCG---GAAGCGCCT-ACCGCAAACCGCTTCCTATAG..........................................................-ACTCTGCTC... 9268
MUS_1.CM.01.1085         ---C-AAATG-A-C--G-GGAGCG---GAAGCGC-T-GCCACAAACCACATCCTATAG..........................................................-AC-TTGCCC... 9277
MUS_1.CM.01.CM1239       ---T-AAATA-A-C-AG-AGAACG---GA-GCGC-T-ACCGCAAACCGCATCCTATAG..........................................................-ACTTTGCCC... 9246
MUS_2.CM.01.CM1246       TA-T-AAATA-AAC-AA-AGAGCG---GA-ACGC-T-ACCGCAAACCGCATCCTCTAGAACTCT......GCTA................................................... 9342
MUS_2.CM.01.CM2500       -A-C-A-ATG-A-C--G-GGAGCGA--AC-GCGC-T-GCCGCAAACCGCATCCTCTCGCAGTCT......CCATAA...................................CTAC.............. 9304
RCM.NG.x.NG411           TAGA--------T----T---TGGAG-GA-A...AAGAACAAGACCAAGTAA..................CCACTTCCTCTGGAGTTGCCTTGGAAACAGCG-AT--C-A--GAG-AGATGCTGA 9098
RCM.GA.x.GAB1            TAGA--------A----T---TGG-G-AA-GAAGAACAAGACCCAGTAA..........................CCACATCCTCTGGGGTTGC-T-G-TAACCAGGCAG-AGAATCTGC 9005
DEN.CD.x.CD1             --GC-ATA-GAA-CTTC-AACTTCAC-GCATAATAACTACTGCCAGAGAGCTAAGCCCAGGAAG......TAG...........CCTAACCGCAAACCACATCCTACTGCAGAACTTGCTGAGAGGCTCA........................... 9481
                                   TCF-1 alpha binding                                                                          Nef end

MAC.US.x.239             T-TTAGAT-C--A--AG-G--TGGA-G-AAGTCAG-CCTGTCAGAGGAAGAGGTTAGAAGAAGG......CTAACCGCAAGAGGCCTTCTTAACATGGCTGACAAGAAGGAAACTCGCTGA................................ 9868
Nef                       _V__R__Y__P__E__E__F__G__S__K__S__G__L__S__E__E__E__V__R__R__R__.__.__L__T__A__R__G__L__L__N__M__A__D__K__K__E__T__R__*_
H2A.GW.x.ALI             TATT--A--C--A---G-A--TGG-T--AAGTCAG-CCTGCCAGAAGATGAGTGGAAGGCAAGA......CTGAAAGCAAGAGGGATACCATTTAGTAAGAACAGGAACAGCTGA.................TTTGGTCAGGG--GGA--TAACTACT.GAAAACAGCT 9940
H2A.DE.x.BEN             CACT---T-C--A--AG-G--TGG-C--AAGTCAG--TTGCCAGAGAAAGAATGGAAGGCAAAA......CTGAAAGCAAGAGGGATACCATATAGTGAATAA...............CAGGAACAACCATACTTGGTCAAGG--GGA--TAGCTACT--GAAACAGCT 9944
H2A.SN.x.ST              TATT-GAT-C--A---G-G--TGG-T--AAGTCAG-CCTGCCAGAGGATGAATGGAAGGCAAGA......CTGAAAGCAAGAGGGATACCGTTTAGCTAA...............AAACAGGAACAGCTATACTTGGTCAGGG--GGA--TAACTA---GAAAACAGCT 9383
H2B.GH.86.D205           CA-TAG-TT---A--AG-G---GG-T--CAGTCAG-GCTACCAGAGGAGGAGTGGAAGGCTAGA......CTAAAAGCAAGAGGGATACCTACAGATTAG..........GCAAGAGACAGCAGCATAAACAGGAACTAGCTGACA-TGCAC.AAG-AGGAAACTAGCA 9864
H2B.CI.x.EHO             CAGCAG-TTC--A---G-G--TGG-T-TCAGTCAG--ATGCCAGAGAAAGAGTGGAAGGCTAAA......CTGAGAGCAAGAGGAATACCTACAGAGTAG..........ACAGGAAATAACAGCACAAAGAGGAACTAGCTAACA-TGCA-A-AG-AGGAAACTAGCT 9836
H2G.CI.92.ABT96          TATAT-------T---G-G----G-TGGAAGTCAG-GCTACCAGAAGCAGTATGGAAAGAGAAA......CTAAAACAGAGAGGACTGCCTATAGAATAA..................................................GCAGA-AG-GGACAAGCT 9293
H2U.FR.96.12034          TATTAA-TGC--A--AG-G--TGG--GTAAGTCAG-TCTGTCAGAGGAAGAGGTTAAGAGGAGG......CTAACCGCAAGAGGCCTCCCAGTAAAGAATTGCTGA.............................................CAGA-AGGAAACAAGCT 9394
MAC.US.x.EMBL_3          T-TTAGAT-C--A--AG-G--TGGA-G-AAGTCAG-CCTGTCAGAGGAAGAGGTTAGAAGAAGG......CTAACCGCAAGAGGCCTTCTTAACATGGCTGACAAGAGGGAAACTCGCTGA...G..ACAGCA.......................... 8996
SMM.US.x.H9              T-TTGA---C--A--AG-G--TGGT-GTCAGTCAG-CTTGTCAAAGGAAGAGGTGCAGAGAAGG......CTAACCGCAAGAGGCCTTTTAAAAATGGCTGACAAGAAGAAAACAAGCTAG................................. 9351
SMM.US.x.PGM53           CATTAGA--C--A--AG-G--TGGC-GTAAGTCAG-CTTGTCAGAGGAAGAGGTAAAGAGAAGG......CTAACCGCAAGAGGCCTTTTAAAAATGGCTGACAAGAAGGAAACAAGCTGA................................. 9805
SMM.US.x.SIVsmH635F_L3   TATTAA-T-C--A--AG-G--TGGT-GTAAGTCAG-CTTGTCAGAGGAAGAGGTAAAGAGAAGG......CTAACCGCAAGAGGCATTTATAAAATGGCTGACAAGAGGGAAACAAGCTGA................................. 9890
SMM.SL.92.SL92B          T-TAAG---C--A--AG-G--TGG-TGGAAGTCAG-CTTGACAGAGGAAGAGGTTGAGAGAAGGTTAGCTAACAAGCCAAAACCGCAGAAGAAGATGGCGGATAAGAAGGAAACAAGCTGA................................. 9325
STM.US.x.STM             T-TTAGA--C--A--AG-G--TGGTTCTAAGTCAG-CTTGCCAAAGGAAGAGGTTGAGAGAAGG......CTAACCGCAAGAGGCCTTCTAAAGATGGCTGACAAGAAGGAAACAAGCTGA................................. 9522
MNE.US.x.MNE027          TATTAGAT-C--A--AG-G--TGGA-G-AAGTCAG-CCTGTCAGAGGAAGAGGTTAGAAGAAGG......CTAACCGCAAGAGGCCTCTTAAAAATGGCTGACAAGAAGGAAACTAGCTGA................................. 9345
DRL.x.x.FAO              TATT-AC-----A---G-G--TGGAC-TGTTAAAA-TATAAAGTCAGTAGAAGAGCCAGGT.................................TGTTGCAAGAGGAAGTGGTGGCAGTTCAAGCCAACCACAGAAGGTTGTC-TGGA-AC-ACCTTCAGAAGT 9220
MND_2.x.x.5440           TATAA-TAGG--A---G-A-A----TGTGTG..................ACAAGTCTGAGCTAT......GAAGCCTACAAGAAGGAGGAAAAGCCTGACTGCTGCAAGAGGAAGTGGTGGCAGTTCTAGCTAACCACAAAGGGTTG-T-TAAGCAACTGCTAACTCTGG 9150
MND_2.CM.98.CM16         CATAAA-AGA--A-GTG-A------C-TGTG...CA-AGCATCAGCCATATTGCTTATGCTAAG......GAGCATAAGCCAGAGTGCTGCAAGAGGAAGTGGTGGCAGTTCTAG...............CCGACCACAGAAGGTTG-C-TGGA-A-GACCTGCAGAAAT 9455
MND_2.GA.x.M14           CATAAAATGC--AA--G-A-A--G-C-TGTT..................ACAAGCCTCTCCTAT......ACTGCCTACCAGGAG......AAATCAGACTGCTGCAAGAGGAAGTGGTGGCAGTTCTAGCTAACCACAGAAGGTTG-T-TGGA-A-G-TCTTTAATTCA 9390
OLC.CI.97.97CI12         ........................................................................................................................................................... 8678
GRV.ET.x.GRI_677         ........-GC-----CATG--T---G--ATGCACGA-CATGCAAAGCGCTAG.......................................TGTCAGCACTTTGCGGTT........................................ 9289
VER.KE.x.TYO1_patent     .......G-C--CA-TCGGGAGT-CTTTACTGACATGCATGGGCTGGTGAAGAGGAAGTAG....................CCAGACCGCAAGCCTGCGGTTAGAACATCACCATGGAGATGAC-TTAAAAACTGCTGAC..... 9456
VER.KE.x.9063            .......G-C--AA-CAGACAGT-TTTTGAAGACATGCATGCAATAGTGAAGAGGAAGTAG....................CTAACCGCAGGCTTGTGGTTAAGCCGTTGCCAGGGAGATGAC-TT-GAAACTGCTGACAG... 9496
VER.KE.x.AGM155          .......G-C--CTCAAGGGAGT-CTTTACAGAC-T-TATTCAACAGTTGGTACAGGAAACTAG.................CCGACCACAGGCTTGCGGTTTCCTGGTTGCCTAGGAGATGAC-TTAA-AACTGCTGAC..... 9474
VER.DE.x.AGM3            CAGAAAAT---TCACTG--A-GC-TG-TCTT...G-CAAGAGGAAGTAG.......................................CT.AACCGCA.GGCTTGTGG-TA---ACATC--C-TGGTGATGA 8976
TAN.UG.x.TAN1            .......A-C-----C......-GCTTTAAGGACATGCATGGTCTTGTTAAGAGGAAGTAG....................CCTAACCGCAGGCTTGCGGTTAAGCTAACTAGGCCGTTGCCTAG-AGATGACAT-AAGAAGTGC 9472
SAB.SN.x.SAB1C           GAG------C--A---C-GG-GCC-TCAGCA...CA-GGCTAA..........................................CCACAAAACCACATCCTATGGAGTTGTCATG-TGATGACATTAAGAACTGC 9652
MND_1.GA.x.MNDGB1         -C-TA-CACC-T-AGTGTT-CA-TA-GGCTAAGAGT-TTGAACATCTGCCATTTTGGAAGAGGA......AGTAG........................CCTAACCGCAAAACCACATCCTAC---AGA-CTGTAGTTGC 8923
DEB.CM.99.CM40           CC-TGCTG-A-AAAGAG-GCAT-CT-GAA-ATGTATGTTCCCTAAGAGGAAGTAG..........................CTAGCTTAACCGCAAACCACATCCTA--G-AGCGCCCG.............. 9114
DEB.CM.99.CM5            -C-TGAAGCC-ATA-ACCA...G--C-TAA-AGAA--TGCATGTTCCCTAAGAGGAAGTAG....................CTAGCTTAACCGCAAACCGCATCCTCTTG-AGAGAGC-C............. 9045
TAL.CM.01.8023           G-CTGA-T--GACAG--TGGAG-G---GAAGAAGGAGCTGCTAGCTCCACCGCAAACCGCATCC......TCTTGA..................................................AAG 8764
TAL.CM.00.266            G-CCGA-TGGGACAG--TGGAG-GAG-GAAGCGGATGCTGCTAGCTCCACCGCAAACC..........................................GCATCCTCTTGAAG 9213
SUN.GA.98.L14            -C-AAA---C--T--AG-G------C-TGTTACTA-TCTGCAGTGGGAGGTGTAG..............CTTAACCGCAAACCGCA.....TCCTCTTGCATCGCC..TAGGCAACGGGGCTAGCGCATGCGCGCTAGAGTA........................ 9653
WRC.CI.97.97CI14         -CTC-CTGGA-A-TCAAT-GC--CTGGGAAAGAGGCCATTGAGAGTCTCTATAAG...AACAAG......AAGCAGTGCAAGTAG.........................................-GACATG... 9423
WRC.CI.98.98CI04         -CTC-CTGGA-ATATAATAGC-C-AGGAGAAGATGCTATTGAGACCCTCCATAAAGAGAAGCTG......AAGCGGCACAAGTAG.................................................. 9461
WRC.GM.05.Pbt_05GM_X02   -CT-GCCG-CGGTC--ATAGC--CTGGTATTGAGACCCTC......TATGAGATG...AAGAAG......AGAAACAACAAGTAG.........................................-GACTTG... 8355
LST.CD.88.524            GAGAAA---A--A-CAAGTG-TC-AGCAACTGCTATGAGATTTAACTGCGAGAGGAAGTAG.........CCAAACCGCAAACCGCAGAACATCCTCTTGCAGCCCGGTTGCCCTAGCAACCGGGCTAGCGCATGCGC.......................... 8515
LST.CD.88.447            CAGAAAA--A--TA-AAG-A-AC-AGCA-CAGCTATTGGCTTG......CAGAGGAAGTAG.........CCAAACCGCAGACCGCAACACATCCTCTTGCAGCCTGGTTGCCAGGGCAACCAGGCTAGCGCATGCGC.......................... 8509
LST.KE.x.lho7            CA-AAA---G--T-C-A-GA-AC-A--TACTGCT-TCGCTTTTGACTGCAAGAGGAAGTAG.........CCAAACCGCAAACCGCAGCACATCCTCTTGCAGCCCGGTTGCTAAGGCAAC...CGGGCTAGCGCATGCGC....-AT-GCT-GCTG-...CGAGCA... 9609
SYK.KE.x.SYK173          GTTAGA-AGA-AAACAGGAAAACCAC-AGAG.........CTGCAGTCAGCACTTAGTAAGAAC......ATCTCCTAG..................GAGACTCCATGGTGACAAGCTCGGCCCAGGGGA.................................. 9312
SYK.KE.x.KE51            G-CT-GT--G-AATTAG-GAGAC-A-CAG-AAAGC-CTAGCTCACCGCAGAAACCACAGGGTAC......ATCTCCTAG..................GAGACACCTAGGAGACAGAGACCGCCCAGGGGAGGAGTCGGGGCGGGGGAG-CT-ACC-A--CTGCATAAAAG 9048
COL.CM.x.CGU1            GTT-AGCGT-A-AAGCAGGAAG-----GAAACCACA-CAGGTTGCCATAGAGATGGTGGACTGA.........................................................GTTGCT 8611
```

```
                              NF-κ-B-II          NF-κ-B-I                       Sp1-III                        Sp1-II     Sp1-I                    TATA box
H1B.FR.83.HXB2            ...................GGGACTTTCCGCTG...................GGGACTTTCCAGGGAGGCGTG.....................................GCCTGGGCGGGACT.GGGG.AGTGGC.GAGCCCTCAG.ATCCT....GCATATAAGCAG..C 9521
H1A1.UG.85.U455           ...................----------A---...................----------GA--GA-G.--T.................................G-TT-------AGT-.----.G------.T---------.--G--....-------------..- 8981
H1B.US.90.WEAU160         ...................----------A---...................----------G----------.................................AT----------A.----.------.----------.--G--....-------------..- 9521
H1C.ET.86.ETH2220         ...................----------A---...................----.-G-------AG-AG-G.................................-T-----------.----.------.C-A-------.--G--....-------------..- 8921
H1D.CD.84.84ZR085         ........................................................................................................................................................................ 8975
H1F1.BE.93.VI850          ...................----------A---...................------------A-G-CG--CC.................................AGAG--.-------.----.------.TCA-------.--G--....-------------..- 8819
H1G.SE.93.SE6165          ...................----------A-CT...................----------G-------C-G.................................------A--AGT-.----.------.T-A-------.AG--....-------------..- 8965
H1H.CF.90.056             ...................-------------...................----------G---------A.................................T---------GA-.----.------.C-A-------.--G--....-------------..- 8839
H1J.SE.93.SE7887          ...................-------------...................-------------GA-GAGTG.................................-----------GT-.----.------.T-A-------.--G--....-------------..- 8820
H1K.CM.96.MP535           ........................................................................................................................................................................ 8604
H101_AE.TH.90.CM240       ...................A-A----.-------...................-------------GA-GTGTG.................................GC-G.------AGT-.----.---A--.T-A-------.--G--....-----A-------..- 9088
H102_AG.NG.x.IBNG         ...................----------A---...................-----------.---GA-GTGT.................................G-TT----A--AGT-.----.------.T-A-------.C-G--....-------------..- 9048
H1N.CM.95.YBF30           ...................--------ATA--TG....GGGACTTTCCGCCA.--------------.A-GTGT.............................GGTT..GGG--AGT--CT-.-...................G--....-----A-------..- 9120
H1O.BE.87.ANT70           ...................--A-------AGCAAAGACTGCTGACACTGCG..------------T-G-AG-GA.................................-A---G-C-GT-C.---G-----TA-.-------A-.G--....-------------..- 9642
H1O.CM.91.MVP5180         ...................----------AGAC.....TGCTGACACTGCG..------------C-T--GAG-GA.................................TAA--G-C-GT-C.---G-----TA-.--------.--G--....-------------..- 9638
CPZ.CD.90.ANT             ...................----------A--AT...................----.........................................GT-A--AT-G-C-GGCA-T--G-----TTTC----G-CT..G--....-------------..- 8978
CPZ.CM.01.SIVcpzCAM13     ...................----------AA...................----------G-------G-.................................TC.-AG---A-T-GT-.-----.-----TTTC---------.AG--....-------------..- 9172
CPZ.CM.05.SIVcpzEK505     ...................---------TA--TG....GGGACTTTCCAGCA.----------G------.....................................T-T-GT-T----.-----TTG.--------.AG--....-------------..- 9018
CPZ.CM.05.SIVcpzMB66      ...................----------A...CA...................---GA-----G-------.................................GT-A-CAT-G-C-GGGGA-A-G------TT.---------.AG--....-------------..- 9165
CPZ.CM.98.CAM3            ...................---------TAA...................-----------A--G--G..TG.................................GT-CA-G-C-GGACT.--GC----TTTTA-------.-AG--....-------------..- 8976
CPZ.GA.88.GAB2            ACTTTTACAA.........----------AA.......................................GG.........................GGGGTGG-TTCAA-G-A-GAGTC---GT-----TTT--------A..G--....-------------..- 8951
CPZ.CM.98.CAM5            ...................---------TAA...................-----------A--G--A..-G.................................-TCA--GC--GACA..--GC----TTTCAGTT-TCAG-.......................... 9261
CPZ.CM.05.SIVcpzMT145     TTTGCTGACAA........----------A....................A-------------G--TTG-GT.........................TTGGGGCGTAACT--G-C-GAACT.--G-----TTG..--CA-CA-AG--...-------------CG 9094
CPZ.TZ.00.TAN1            ...................----------AA.......................................................................TGTGGGTG-TTACT-G-C-GGACA---G-----TTTT.G--.-GCTGAG--....-------------..- 9207
CPZ.CM.05.SIVcpzLB7       ...................-A------T-TGACCA...................-----------CCA---G.....................CTTCCGGGGAGGCGT-T-AA--GA--GT-T----.-----TTT-.--------.AG--....-------------..- 9054
CPZ.US.85.CPZUS          ...................A--------TAA...................-----------A--GACGT-C.C.........................AAGGGGGT-GG-CA-G-C-GAACA.--GC----TTTTA------A.-GA--....-----A-------..- 9587
CPZ.GA.88.GAB1            ...................----------AA...................----------G-----AG-.T.................................G-T-G--AG-C-GGGTT---GT-----TTT---------.-.G--....-------------..- 9614
MON.CM.99.L1              ...................----------A......................-T-ACCATGACT-C-G-GCCCGGTTGCTGCGGCAACC.GGGCGGTCCAAAGGACTTG-T-G-T-GACCC---G------TTG.-----TCGT-T.....CA---A-------..A 9423
MON.NG.x.NG1              ...................----------A......................-TTGCC-TGACT-C-G-GCCCGGTTGCCGAGGCAACCGGGGCGGACTCGTGGGAGGGACTG--AGG-AC---G-----......-T-ACCCTCAGATCCT-----A-------..A 8061
GSN.CM.99.CN166           ...................----------A......................-T-ACCATGACT-C-G-GCCCGGTTGCTGAGGCAACCGGGGCGGACTCGTGGGAGGGACTG--AGG-.C---G------TCTA---CGCTAT.--A....-----A-----A..A 9389
GSN.CM.99.CN71            ...................----------A......................-TTACCATGACT-C-G-GCCCGGTTGCTGAGGCAACCGGGGCGGACTCGTGGGTGGGACTG--AGGC.C.--G------TCTA---CGCTAT.--A....-----A-------..A 9385
MUS_1.CM.01.1085          ...................----------A......................-T-ACCATGACT-C-G-GCCCGGTTGCTGAGGCAATCGGGGCGGACTCGTGGGTGGGACTG--CGG-ACT--G-----......-T-ACCCTCAGATTCC-----A-------..A 9395
MUS_1.CM.01.CM1239         ...................----------A......................-T-ACCATGACT-C-G-GCCCGGTTGCTGAGGCAACCGGGGCGGACTCGTGGGTGGGACTG--AGG-ACT--G-----......-T-GCCCTCAGATTTC-----A---CA..G 9364
MUS_2.CM.01.CM1246         ...................-A---C----A......................-T-ACCATGACT-C-G-GCCCGGTTGCCGCAGCAACCGGGGCGGACTC.AGGGCGGACT-G-A-GGACCT-AG------TTG.-------ATC-A.........----A-------..A 9459
MUS_2.CM.01.CM2500         ...................------A-AA-.............................GCCAGGTTGCCTGGGCAACCTGGGCGAACTCGTGGGTAAAACTG--AGGAGC---A------TT-.-------A--T-.........----A-------..- 9405
RCM.NG.x.NG411            ..CACAGAA..........----------A---GCGCCTGCGCGCTGGTGTAA-----------.................................AGACT-A--T-GGAG----GT-----TTG.-----TCCT.G--...-------------..- 9203
RCM.GA.x.GAB1             TGATGCAAAA.........----------A---GTGCATGCGCACTGGGGAA.-----------.................................GGGAT-AC-TGGGA----G-----TCAG.-----TCCT.G--...-------------..- 9111
DEN.CD.x.CD1              ...................----------A.-.............GTTGCCTA-CA--CC-GGGC---C.........................TCAGGGGA-GGCCT-G-A-GTCT.CT-GGAGT-GCTT.A--CGCTTC-A--........-A---T-..A 9579
                              NF-κ-B-II                          NF-κ-B-I                                                                                         TATA box
MAC.US.x.239              ..AACAGCA..........----------A-AAG...................---.A-G-TA.................................CGGGGAGG-ACTG----GGA-CC-GTCG--AACG.---A-TTTC.TTG....ATG-----AT-TCA- 9960
Nef
H2A.GW.x.ALI              GAGACTGCA..........----------AGAAG...................--G--G-AACC.................................AGGGGAAGGACAT----GGA-CT-GT-G--AACG.----.TCAT-.-TC....CTG-----ATGTAC-10035
H2A.DE.x.BEN              GAGGCTGCA..........----------AGAA...................---G--G-AACC.................................AAGGGAGGGACAT----GGA-CT-GT-G--AACG.------TAC.TTA....CTG-----ATGTAC-10040
H2A.SN.x.ST               GAGACTGCA..........----------AGAA...................---G--G-TACC.................................AGGGGAGGGACAT----GGA-CC-GT-G--AACG.------TAC.TT-....CTG-----ATGTAC- 9479
H2B.GH.86.D205            GACACTGCA..........----------AAAA...................-A-G--G-AACA.................................TGGGAGGA-CAA----GGG-TT-GTTG--AACG.---ATTAA-.T--....CTG-----ATGTAC- 9959
H2B.CI.x.EHO              GATACTGCA..........----------AGAA...................---G--G-AACA.................................GTGGGAGGGACAT----GGA-CA-G.-A--AACG.---ATCAAG.A--....CTG-----ATGTAC- 9930
H2G.CI.92.ABT96           GACACAGCA..........----------AG-A..................T--G--G-AACA--G-AG.............................GTACTAT----GGGACT-GTTG--AACG.---C.-AA-.T--....CTG-----AT-TAG- 9388
H2U.FR.96.12034           GAGACAGCA..........----------A-CA..................GA--GAGG-AACCA-G-AG.............................GA-CAA----GGA-CCA.T-G--AACG.---.TCAA--GA-....CTG-----ATGTAC- 9489
MAC.US.x.EMBL_3           ...................----------A-AA.......................................................GGGGATG-TATG----GGA-CC-GTCG--AAC-.---A-TTTC.TTG....ATG-----AT-TCA- 9070
SMM.US.x.H9               ..GACAGCA..........----------A-AAA...................---GA-G-TAC.................................GGGGAGG-TCTG----GGA-CT-G.C---AACG.---A-TTAT.T--....CTG-----AT-CAA- 9443
SMM.US.x.PGM53            ..GACAGCA..........----------A-AAA...................---GA-G-TA..................................CGGGGAGG-ACTG----GGA-CT-G.C---AACG.---A-TTAT.TT-....CTG-----AT-CAA- 9897
SMM.US.x.SIVsmH635F_L3    ..GACAGCA..........----------A-AA...................-----------CAAG--GC--TCA.........................TGGGGAGG-ACTG----GGA-CT-G.C---AACG.---A-TTAT.T--....CTG-----AT-CAA- 9995
SMM.SL.92.SL92B           ..GACAGCA..........----------A-AA..................G----G-G-A--A--G.............................A-GAGTCTG--CGGGG--AA..CG--AACG.---CT-AAC--TG....TAG-----AT-TAC- 9418
STM.US.x.STM              ..GACAGCA..........----------A-AA...................---G--G-AACA.................................GGGGAGG-ACT.----GGAACT-GT-G--AACG.----TTTAC.TT-....CTG-----AT-CAA- 9614
MNE.US.x.MNE027           ..GACAGCA..........----------ATAA...................---GA-G--.................................ATGGGG-GG-ACTG----GGA-CT-GTCG--AAC-.---A-TTTC.TTG....ATG-----AT-TCA- 9437
DRL.x.x.FAO               GCTGACTGAGAA.......----------A-A-GAGCATGCGCTCTGTGGC.A-----------.................................AGGGTGAC-T-G-TTGGGG-A.........-TCTG.-------CT.G-A.........----A-------.CT 9330
MND_2.x.x.5440            A..................----------A---GTGCATGCGCACTGGACT.G----------.A-ACT-ACGTCGGG.........................................T----G------CAG.---AA...C.-TGGC.T------------.CT 9249
MND_2.CM.98.CM16          GCTGACTCTGGA.......----------A--AGTGCATGCGCACTAGACT.G--------T-CA-ACT-A.................................................................................................. 9518
MND_2.GA.x.M14            GCTGACTCTGGA.......----------A--AGTGCATGCGCACTGGACT.G--------T---ACT-A---CGGG.........................T----G-----TCAG.---CA..T..-TGAC.T------------.C- 9500
OLC.CI.97.97CI12          ........................................................................................................................................................................ 8678
GRV.ET.x.GRI_677          ...................------------CA...................-----------CA-T--G.........................TGGATCGGAGGCG-TACA--G-C-GTAC.T--G------TTTC-----.AG-G-TG.........----A-------..A 9387
VER.KE.x.TYO1_patent      ...................----------AGC-AA...................-----------A--C--G.................................ACAT-G-C-GTCC.---G------TTTA--------.-.A--.........----A-------..A 9541
VER.KE.x.9063             ...................----------AA...................-----------.......................GGGCG-G-CAT-G-C-GTAC.---G-----TTTTA--------.-.G--.........----A-------..A 9577
VER.KE.x.AGM155           ...................----------AGCAC...................-----------A--C--.................................GACAT-G-C-GTAC.---G------TTTA--------.-.G--.........----A-------..A 9558
VER.DE.x.AGM3             CATTAAGAACTGCTGAC..----------AGCAA...................-------------.................................C-GG-CATG--CGG-AC---G------TTTA--------.-.G--.........----A-------..A 9077
TAN.UG.x.TAN1             TGATCAGCA..........----------AG.......................................................GGTG-TGC-A-G-C-GTACTT--G------TT.C--------.-.G--.........----A-------..A 9550
SAB.SN.x.SAB1C            TGACTGAGA..........----------AG.......................................................GGT-GAGACTG--CGG-ACT--G------TTG.---CGT-CT.G--.....CA-----------..- 9730
MND_1.GA.x.MNDGB1         TTGGCAACCTGCTT.....A-C-ACC-GGA---GCGCTTGCGCGCTAGGA..A-----------AA...............................................CAGG-A--GG-A---TCG.---CA..TG.-TG....CT-----------..- 9023
DEB.CM.99.CM40            ...................----------A.................................GTTGCCTAGGAAATATCATGGCAACCAGGGGA-GGCCT-G-T-GTAC.---G------CA..G--CGCTG-A---.....A----A----C-..A 9210
DEB.CM.99.CM5             ...................--T-------A.................................GTAGTCAAGGAGACTCCATGGTTACAGGGGAG.GGACT-G-C-GTAC.---G-----TTTG.----GCTGT-A--.....-----------T-..A 9141
TAL.CM.01.8023            GACTC..............------------................................CGCTGCTAAGGCAACGGGGGACGGACTGGGGCG-GGA-C-G-A-GAGCT.--G------TA..-----.......................... 8845
TAL.CM.00.266             GACTC..............----------A.................................CGTTGCTAAGGCAACGGGGGACGGACTGGGGCG-GGA-C-G-A-GAGTT.--G------TTG.-------A-.--A....-------------..A 9315
SUN.GA.98.L14             ...................-----------GAC............................................................AGGG..A--GG-A---TT-.---A-..TT.A--....-------------..- 9707
WRC.CI.97.97CI14          .......TTGCAGTAGGAA-T-G----G--G-TAAGCTACA.........................TCCTGAGACTGTGTACCAACGGACTTA.......................................................................... 9484
WRC.CI.98.98CI04          ........................................................................................................................................................................ 9461
WRC.GM.05.Pbt_05GM_X02    .......TCAGA...GGAA-CTGG---G--G-TAGGCTAGTA.........................TCCGCTAGCTGTGTACCAACGGCCTTAGCAA--AT-CTAATGAACT---ACTCT-GGAG-AGACTT-TGGTT-TGCCCT---A---T--ATG 8479
LST.CD.88.524             ...................ACA-GC--G-TGAACAGC..............A-----------..............................--A----CGGG--A-G.T-T-.........---A-..TT.G--....-------------..- 8588
LST.CD.88.447             ...................A-------G-TGATGAGC..............A-----------..............................--A----CGGG--A-G.T-T-.........---A-..TT.G--....-------------..- 8582
LST.KE.x.lho7             ...................----------AGGACGGGCGG...................----GG--TG........................................---A-..TT.G--....-------------..- 9663
SYK.KE.x.SYK173           ....................................................--AG-C-GGGC---G-AA-GCTCA.........................................----......A--.........----A----C-..A 9356
SYK.KE.x.KE51             CCGATGCTCTGTAGCTATCC.................................................................................................................................................... 9068
COL.CM.x.CGU1             GAGACAGCA..........-----------T-C...................-----------A-TT--GAG-.............................-A-CT--GC---GGGAA--GC---.........----GCTA-A.--GCTT-T---A---TCAG.A 8709
```

```
                                    TAR element start
                          3' LTR U3 end 3' LTR repeat start
H1B.FR.83.HXB2           TGC....TTTTTGC.CTGT..ACTGGG..TCTCTC...TGGTTAGA.CCAGAT.CTGAGCCTGGGAG.CTCTCTGGC.....TAA...........................CTAG..................................................GG 9592
H1A1.UG.85.U455          ---......----C--.----.------.------...-T------.------.C-----------.---------......--G...........................-G--..................................................-- 9051
H1B.US.90.WEAU160        ---......-------.----.------.------...--------.------.------------.---------......--G...........................----..................................................-- 9592
H1C.ET.86.ETH2220        ---......----C--.T---..--C---.------...-A-G----.------.------------.---------......--T...........................--G-..................................................-- 8992
H1D.CD.84.84ZR085        ...................................................................................................................................................................... 8975
H1F1.BE.93.VI850         C--......----C--.----.------.------...-A------.------.T---------G-.---------......--G...........................--G-..................................................-- 8890
H1G.SE.93.SE6165         C--......--C-C--.----.------.------...-T------.------.T-----------.---------......--G...........................-AG-..................................................-- 9036
H1H.CF.90.056            ---......----C--.T---.------.-----T...--------.------.------------.---------......--G...........................---A..................................................-- 8910
H1J.SE.93.SE7887         ---......----C--.T---.------.------...-T------.------.------------.---------......---............................----..................................................-- 8891
H1K.CM.96.MP535          ...................................................................................................................................................................... 8604
H101_AE.TH.90.CM240      C--......----C--.T---.------.------...-T-----------G-.-----C-------.---------......--G...........................-A--..................................................-- 9158
H102_AG.NG.x.IBNG        ---......--C-C--.----.------.------...-T-C----.------.------------.---------......--G...........................-GGA..................................................-- 9119
H1N.CM.95.YBF30          C--......--C-C--.T---.------.------...-T-C-G--.------.TA---T-------.-ATAT---G.......................................................................................... 9182
H1O.BE.87.ANT70          C--......---C---.T---..--C---.----GG...-TAGAG--.----G-.-------C-----.---C-----....CTC...........................---C.................................................T- 9713
H1O.CM.91.MVP5180        ---......---CC--.T---..--C---.---TAG...-TAGAG--.----G-.-------C-----.---C-----....CTC...........................---C.................................................T- 9709
CPZ.CD.90.ANT            ---......---GC--.T---..TAA---.C-----G...CC.-AG--G-------.-A-----------.---------......---TTTTT.....................GGCAG.................................................-A 9055
CPZ.CM.01.SIVcpzCAM13    C--......---C---.T---.------TC--..-T...CAC.-G--.-------.------C-----.-------C-......-G...........................-GA..................................................-- 9240
CPZ.CM.05.SIVcpzEK505    ---.....-GC-C--.T---.------TC-----T....-.----------.TA------------.---------......--G...........................-A-..................................................-- 9088
CPZ.CM.05.SIVcpzMB66     C--......---GC--.T---.------TC-----T....-.---------T-------------.---------......--G...........................--A..................................................-- 9236
CPZ.CM.98.CAM3           C--.....-C-C---.T---.------.-------..T--C.------.------------.---------T.....C-G...........................G---.................................................A- 9047
CPZ.GA.88.GAB2           ---.....-C-C---.T---..T------.-----T...A--CAGA..------.------------.---------......-G-...........................----..................................................- 9020
CPZ.CM.98.CAM5           ...................................................................................................................................................................... 9261
CPZ.CM.05.SIVcpzMT145    C--.....-C-C--.T---..CAA---TC-----T......------.------------.---------......---CAAG.................................................................................-- 9164
CPZ.TZ.00.TAN1           ---......---GC--T-----..AAA--..----TG.CCTAA-CTG.------.------------.---------......--G...........................-G-CTGGC.......................................TA-A 9284
CPZ.CM.05.SIVcpzLB7      C--.....-C-GC--.T---.------TC-----T.....-----------T-------------.---------......--G...........................--A..................................................-- 9125
CPZ.US.85.CPZUS          C--.....-C-C---.T---.------.------...TC-C-AGA..------.AA-----C------.---------......--G...........................-GTA..................................................- 9656
CPZ.GA.88.GAB1           ---......---C---.----.------.-----T..TCAC.-G--.------.T-----------.------AC-.....AGT...........................G-A..................................................-- 9684
MON.CM.99.L1             C--.....CC..........----.AG--ATCCT--GGAGTC........................................................................................................................ 9448
MON.NG.x.NG1             C--.....-C..........----.AGM-ATCCGC-GGAGTC...........................--.-----GTCT...........................CCAC--CGGCCT..............................GGGAGCCCTTGGA- 8123
GSN.CM.99.CN166          C--.....-C..........----.AG-GACTCT--GGAGTC........................................................................................................................ 9414
GSN.CM.99.CN71           C--.....-C..........----.AG-GACTCT--GGAGTC........................................................................................................................ 9410
MUS_1.CM.01.1085         C--.....-C..........----.AG--ATCCC--GGAGTC........................................................................................................................ 9420
MUS_1.CM.01.CM1239       C--.....-C..........----.AG--ATCCC--GGAGTCAGTTC---.CT-C-CA.GC-GC.-------C-CTCGGAA-......--G.GTAA.AGCGGCCTGAGAGCCCTTGCCTAGGAGAACTG..............................GTTCTCGGCTTGCT 9479
MUS_2.CM.01.CM1246        C--.....-C..........----.AG--ATCCAA-GGAGTC........................................................................................................................ 9484
MUS_2.CM.01.CM2500       ---.....-C...........G--ATCCAGACGGAGTC........................................................................................................................ 9430
RCM.NG.x.NG411           ---.....A---C--T----.A--CA-..------CCTG--AGGCTA------.----------TTGT-------T.....A-GTCTCTAGGAGACT.AGCTTGAGCCTGGGTGTTCGCT.....................................A-T 9311
RCM.GA.x.GAB1            ---.....-C-GC--T.---AA-AC---..------CCTG--AGGCTA--G---.----------T-.T-------T.....A-GTCTCTAGGAACTCCAGCTTGAGCCTGGGTGTTCGCT...................................GGT- 9221
DEN.CD.x.CD1             ---...........T-----.A--A-TCGG-T...........................-C---.-----CCAGCGGCCTGG..........................................................................-A 9622
                                     TAR element start
                          3' LTR U3 end 3' LTR repeat start

MAC.US.x.239             ---...A---CGCTC.---.ATTCA-TC.G----GC.GGA-AGGCTGG-----.TGAGC--------GT-----CCAGCACTA.......GCAGGTAGAGCCTGGGTGTTCC--........................................GCTA-10064
H2A.GW.x.ALI             C--...TGC--GCAT.---.ATTCA-TC.G----GC.GGA-AGGCTGG-----.TGAGC--------GT-----CCAGCACTA.......GCAGGTAGAGCCTGGGTGTTCC--........................................GCTA-10139
H2A.DE.x.BEN             C--...T-C--GCAT.---.ATTCA-TC.G----GC.GGA-AGGCTGG-----.-GAGC----A---GT-----CCAGCACTA.......GCAGGTAGAGCCTGGGTGTTCC--........................................GCT--10144
H2A.SN.x.ST              C--...TAC-CGCAT.---.ATTCA-TC.G----GC.GGA-AGGCTGG-----.TGAGC--------GT-----CCAGCACTA.......GCAGGTAGAGCCTGGGTGTTCC--........................................GCTA- 9583
H2B.GH.86.D205           C--...T-C-CGCAT.---.ATTCA-TC.G----GC.GGA-AGGCTGG-----.TGAGC--------GT-----CCAGCACTA.......GCAGGTAGAGCCTGGGTGTTCC--........................................GCTA-10063
H2B.CI.x.EHO             C--...T--G-GCAT.---.ATTCA-TC.G----GC.GGA-AGGCTGG-----.-GAGC--------GT-----CCAGCACTA.......GCAGGTAGAGCCTGGGTGTTCC--........................................GCTA-10034
H2G.CI.92.ABT96          ---...A---CGCTC.---.ATTCA-TC.G----GC.GGA-AGGCTGG-----.TGAGC--C-----GT-----CCAGCACTA.......GCAGGTAGAGCCTGGGTGTTCC--........................................GCTA- 9492
H2U.FR.96.12034          C--...TC---GCAT.---.ATTCA-TC.G----GC.GGA-AGGCTGG-----.TGAGC--------GT-----CCAGCACTA.......GCAGGTAGAGCCTGGGTGTTCC--........................................GCTA- 9593
MAC.US.x.EMBL_3          ---...A---CGCTC.---.ATTCA-TC.G----GG..CA-AGGCTGG-----.TGAGC--------GT-----CCAGCACTA.......GCAGGTAGAGCCTGGGTGTTCC--........................................GCTA- 9173
SMM.US.x.H9              ---...A---CGCTC.---.ATTCA-TC.A----GC.GGA-AGGCTGG-----.TGAGC--------GT-----CCAGCACTA.......GCAGGTAGAGCCTGGGTGTTCC--........................................GCT-- 9547
SMM.US.x.PGM53           ---...A---CGCTC.---.ATTCA-TC.G----GC.GGA-AGGCTGG-----.TGAGC--------GT-----CCAGCACTA.......GCAGGTAGAGCCTGGGTGTTCC--........................................GCTA-10001
SMM.US.x.SIVsmH635F_L3   ---...A---CGCTC.---.AGTCA-TC.G----GC.GGA-AGGCTGG-----.TGAGC--------GT-----CCAGCACTA.......GCAGGTAGAGCCTGGGTGTTCC--........................................GCTA-10099
SMM.SL.92.SL92B          C--..TT-GCCGCTC.---.ATTCA-TC.G----AC.GGA-AGGCTG------.AGAGC--C-A---GT------CAGCACT-........AGTAGGTGAGCCTGGGTGTTCT--........................................ACTAA 9523
STM.US.x.STM             C--...T-A-AGCTC.---.ATTCA-TC.A----GC.GGA-AGGCTGG-----.TGAGC--------GT-----CCAGCACTA.......GCAGGTAGAGCCTGGGTGTTCC--........................................GCTA- 9718
MNE.US.x.MNE027          ---...A---CGCTC.---.ATTCA-TC.G----GC.GGA-AGGCTGG-----.TGAGC--------GT-----CCAGCACTA.......GCAGGTAGAGCCTGGGTGTTCC--........................................GCTA- 9541
DRL.x.x.FAO              ---...T-AC-GCATG-...A-AC--TC.----C-..TG--AGGCTA------.----------T-.T-------TG....AGTCTGCTGAAGGAGGCTACCTGCTCGAGCC-CG-GTGGTTCGCAGGGCCTTGGCCTTGGGCTCTGGGTAGCTCGGTCAGGAGGCTTC 9484
MND_2.x.x.5440           ---...T-GCCGC-..---.A-G--A-TC----C-T..GA-AGGCTG------.TGAGC.------..TTGT---CT.GG.TGAGTTCTTG.AAGGAG..TGCCTGCTTGTAGC-CT-GGCGGT...............................TCGCAGGCCCCT-- 9369
MND_2.CM.98.CM16         ...................................................................................................................................................................... 9518
MND_2.GA.x.M14           C--...T-GCCGCT..---.ATT.-A-TC----C-T...................................................................................................................................... 9529
OLC.CI.97.97CI12         ...................................................................................................................................................................... 8678
GRV.ET.x.GRI_677         ---.....-CGC--GCT---..-ACTCAG.-----T.AC-A-GAGA.-----C-..A---------T-.T--G----T.......-GCCTA...ACCCGGT.TGGCCA.CCGGGGG--A.......................................-- 9483
VER.KE.x.TYO1_patent     ---.....-CGC--GCT---..-ACTCAG.-----T.AC-A-GAGA.-----C-.-----------T-.T--G----T.......-GCCTA...ACCTGGT.TGGCCA.CCAGGGG--A.......................................-- 9637
VER.KE.x.9063            ---.....-CGC--GCT---..-ACTCAG.-----T.AC-A-GAGA.-----C-.-----------T-.T--G----T.......-GCCTA...ACCTGGT.TGGCCA.CCAGGGG--A.......................................-- 9673
VER.KE.x.AGM155          ---.....-CGC--GCT---..-ACTCAG.-----T.AC-A-GAGA.-----C-.-----------T-.T--G----T.......-GCCTA...ACCTGGT.TGGCCA.CCAGGGG--A.......................................-- 9654
VER.DE.x.AGM3            ---.....-CGC--GCT---..-ACTCAG.-----T.AC-A-GAGA.-----C-.-----------T-.T--G----T.......-GCCTA...ACCTGGT.TGGCCA.CCAGGGG--A.......................................-- 9173
TAN.UG.x.TAN1            ---.....-CGC--GCT---..-ACTCAG.-----T.AC-A-GAGA.-----C-.-------------.T--G----T.....G-GCCTA...ACCTGGTCTGGCCATCCAGGGG--A.....................................GA- 9649
SAB.SN.x.SAB1C           ---.......---.GCGC---AA-ACA-.......----GC.T---AGGCTG------.----------T-.T-------T.....A-GTCTG..AACCAGCT.TGAG...CCTGGGTGTTCGCT...............................GGG-A 9833
MND_1.GA.x.MNDGB1        ---...A---CGCT..----.T-G--AGTCTCTA-TACA-AGGCT-A..............----..TTGTATCTCTGA.GC.................................................................................. 9080
DEB.CM.99.CM40           ---..........T-----..-GCTATCCA........................................................................................................................... 9227
DEB.CM.99.CM5            ---..........T-----..-GCTATCCA........................................................................................................................... 9158
TAL.CM.01.8023           ...................................................................................................................................................................... 8845
TAL.CM.00.266            ---.....-C..........----.AG--ATCCT-T..G-TC....................................................................................................................... 9338
SUN.GA.98.L14            ---.....-C--C--C.----..T-G--AG.------..GC-CAGAG-..AT--GCGAC--C----A---.G----ACCT.....AG-GCTGGCTTGTCGCTCCTGAGTGAGCACTTGCTAAG...............................CACCTCAC 9817
WRC.CI.97.97CI14         ...................................................................................................................................................................... 9484
WRC.CI.98.98CI04         ...................................................................................................................................................................... 9461
WRC.GM.05.Pbt_05GM_X02   CT-TGTAGC-A-CTGG-TGCAGTCTTCTGC--GAGGCTA---GCCTGAGT-G-GG-CCCA--T--CCTAG----CTTGGTGCCTGAGTGGTGGTCCCACTTGGCCAAGAGACT-AG-TAGGAG...............................CTGACCACGCTT-C 8616
LST.CD.88.524            ---.....-C-........................................................................................................................................................ 8594
LST.CD.88.447            ---.....-C-........................................................................................................................................................ 8588
LST.KE.x.lho7            ---.....-C-CC--T.---TCGGGA-..TC--TG-TCTG--GA...C-AGC-G.TC-CT------T-AGCACACT-GT...CAGCTAAGC.AGTCAGCCCTGGGAGGCTACA.....CTTAGC..............................ACCCACCTGGGCCT 9780
SYK.KE.x.SYK173          ---.....-C..........---AGCTATCCGCGGAG-TGCAACTA........................G------CCAGCGGCCT...GGG.AGCCCTGAGAGACTGGTAGAGCGGCCC.GGGA...............................GCCCTTGCCAGTAA 9452
SYK.KE.x.KE51            ...................................................................................................................................................................... 9068
COL.CM.x.CGU1            GC-.....-CGAGAGCT-CGAGG-............................................................................................................................................. 8728
```

```
                         TAR element end┘                                                                                                                     Poly-A signal
H1B.FR.83.HXB2           AACCC.....ACTGCTTAA.....................................................................................GCCTC...................................AATAAAG...CTTGCC.......TT 9626
H1A1.UG.85.U455          -----.....---------.....................................................................................-----...................................-------...------.......-- 9085
H1B.US.90.WEAU160        -----.....---------.....................................................................................-----...................................-------...------.......-- 9626
H1C.ET.86.ETH2220        -----.....---------.....................................................................................-----...................................-------...------.......-- 9026
H1D.CD.84.84ZR085        ......................................................................................................................................................................... 8975
H1F1.BE.93.VI850         -----.....--------....................................................................................................................................................... 8903
H1G.SE.93.SE6165         -----.....---------A....................................................................................-----...................................-------...------.......AA 9071
H1H.CF.90.056            -----.....---------.....................................................................................-----...................................-------...------.......-- 8944
H1J.SE.93.SE7887         -----.....---------.....................................................................................-----...................................-------...------.......-- 8925
H1K.CM.96.MP535          ......................................................................................................................................................................... 8604
H101_AE.TH.90.CM240      -----.....---------A....................................................................................-----...................................-------...------.......-- 9193
H102_AG.NG.x.IBNG        -----.....---------.....................................................................................-----...................................-------...------.......-- 9153
H1N.CM.95.YBF30          ......................................................................................................................................................................... 9182
H1O.BE.87.ANT70          -----.....G--CG----.....................................................................................CG---...................................-------...------.......-- 9747
H1O.CM.91.MVP5180        -----.....G--------.....................................................................................CG---...................................-------...------.......-- 9743
CPZ.CD.90.ANT            G----.......-----........................................................................................-T-............................................................. 9068
CPZ.CM.01.SIVcpzCAM13    -----.......-----.....................................................................................AA-----...................................-------...------.......-- 9274
CPZ.CM.05.SIVcpzEK505    -----.....---------.....................................................................................-----...................................-------...------.......-- 9122
CPZ.CM.05.SIVcpzMB66     -----.....---------.....................................................................................-----...................................-------...------.......-- 9270
CPZ.CM.98.CAM3           -----.....---------.....................................................................................-----...................................-------...------.......-- 9081
CPZ.GA.88.GAB2           -----.....---------.....................................................................................-----...................................-------...------.......-- 9054
CPZ.CM.98.CAM5           ......................................................................................................................................................................... 9261
CPZ.CM.05.SIVcpzMT145    -----.....---------.....................................................................................-----...................................-------...------.......-- 9198
CPZ.TZ.00.TAN1           G----.....G--------.....................................................................................CG-.TC..................................-------...-C---.........- 9316
CPZ.CM.05.SIVcpzLB7      -----.....---------A....................................................................................-----...................................-------...------.......-- 9160
CPZ.US.85.CPZUS          -----.....---------.....................................................................................-----...................................-------...------.......-- 9690
CPZ.GA.88.GAB1           -----.....---------.....................................................................................-----...................................-------...------.......-- 9718
MON.CM.99.L1             ......................................................................................................................................................................... 9448
MON.NG.x.NG1             CTAGGTAAAGCGGC--GGGAGCCCTTGCCTAGAGAGGTGCCGGCTCTCGGCTCGCTTGCTGAATAGGCTTTACC..............................AG-C-...................................-------...G-GCTTGAAAGCAAA 8223
GSN.CM.99.CN166          ......................................................................................................................................................................... 9414
GSN.CM.99.CN71           ......................................................................................................................................................................... 9410
MUS_1.CM.01.1085         ......................................................................................................................................................................... 9420
MUS_1.CM.01.CM1239       TTG-TTCGAAG--TA--G-GCCCACAATA............................................................................................................................................ 9508
MUS_2.CM.01.CM1246       ......................................................................................................................................................................... 9484
MUS_2.CM.01.CM2500       ......................................................................................................................................................................... 9430
RCM.NG.x.NG411           TCT-TGAAGGG--TGC-GGGAGCGCCTTGGCTGGGTCTTGGTAGACCTTCAGCTGGGTACTTGGCCAGTCCCAGTGGGAGTAGGCTCCGCTTGCT..TTCT.GGTTAA-TGCTC..............................------C....-.--ATGCTCTTAG 9442
RCM.GA.x.GAB1            TCT-TGAACAGGCTTGCTGGGGTGCCTCTCGCTCTTCGGGTAGACCGCCAGTTGAGGCTCGGCCGGCCTCAA..CGGGAGAGATCACCGCTTGC...TTATAGCCTTGAAGCTC..............................-------C......ATGCCAG..-- 9347
DEN.CD.x.CD1             GC--TGAG..-GGC----GCAGCGGCCTGGGAGCCCTCTAAGGTATTGCAACAAGTACC.............................................................................................................. 9679
                                                                              TAR element end┘                                                        Poly-A signal
MAC.US.x.239             -CT-TC.ACC-GCA---GGCCGGTGCTGGGCAGAG.......................................TGACTCCACGCT.TGCTTGC...TTAA.AG.C----TTC...............................-------...--.---AT..T..-- 10148
H2A.GW.x.ALI             -CT-TC.ACC-G-----GGCCGGCACTGGGCAGA........................................CGGCTCCACGCT.TGCTTGC...TTAA..AGA----TT................................-------...--.---AA.....-- 10220
H2A.DE.x.BEN             -CT-TC.ACC-G-A---GGCCGGTACTGGGCAGA........................................CGGCTCCACGCT.TGCTTGC...TTAA..AGA----TTC...............................-------...--.---AG.....-- 10226
H2A.SN.x.ST              -CT-TC.ACC-G-----GGCCGGCACTGGGCAGA........................................CGGCTCCACGCT.TGCTTGC...TTAAA.AGA----TT................................-------...--.---AG.....-- 9665
H2B.GH.86.D205           -CT-TC.ACC-GCA--AGGCCAGTGTTGGGCAGA........................................CGGCTCCACGCT.TGCTTGC...TTAA..AAC----TT................................-------...--.---AA.....-- 10143
H2B.CI.x.EHO             -CT-TC.ACC-GCA---GGCCGGTGCTGGGCAGA........................................CGGCTCCACGCT.TGCTTGC...TTAA..AAC----TT................................------A...--.---TA.....-- 10115
H2G.CI.92.ABT96          -CT-TC.ACC-GCA---GGCCGGTGCTGGGCAGAG.......................................CGGCTCCACGCT.TGCTTGC...TTACAACGA----TT................................-------...--.---AAA....-- 9577
H2U.FR.96.12034          -CT-TC.ACC-AAAG.......................................................................................................................................................... 9607
MAC.US.x.EMBL_3          -CT-TC.ACC-GCA---GGCCAGTGCTGG.CAGAG.......................................TGGCTCCACGCT.TGCTTGC...TTAA.AG.A----TTC...............................-------...--.---CA..T.T-- 9257
SMM.US.x.H9              -CT-TC.ACC-GCA---GGCCGGTGCTGGGCAGAG.......................................TRGCTCCACGCT.TGCTTGC...TTAA.AG.A----TTC...............................-------G..--.---AG.....-- 9631
SMM.US.x.PGM53           -CT-TC.ACC-GCA---GGCCGGTGCTGGGCAGAG.......................................TGGCTCCACGCT.TGCTTGC...TTAA.AG.A----TTC...............................-------...--.---GC.....-- 10084
SMM.US.x.SIVsmH635F_L3   -CT-TC.ACC-GCA---GGCCGGTGCTGGGCAGAG.......................................TGGCTCCACGCT.TGCTTGC...TTAA.AG.A----TTC...............................-------...--.---AT.....-- 10182
SMM.SL.92.SL92B          -CT-TCAACC-GCAACAGGCCATTGCTGGGTAGAC........................................GGCTCCACGCT.TGCTTGC...TTGA.......---TC................................-------..G--.--AAC.....-- 9602
STM.US.x.STM             -CT-TC.ACC-GCA---GGCCAGTGCTGGGTAGAG.......................................TGGCTCCACGCT.TGCTTGC...TTAT.AT.AT---TTC...............................-------...--.---AA..T..-- 9802
MNE.US.x.MNE027          -CT-TC.ACC-GCA---GGCCGGTGCTGTGCAGAG.......................................TGGCTCCACGCT.TGCTTGC...TTAA.AG.A----TTC...............................-------...--.---AT.....-- 9624
DRL.x.x.FAO              TGGAGGTAGAT-ACTAGGGGCTGCTCTGGTCTTGGTCTGGTAGACCTTCGCTACTTCTAGTGGCCATAGAAGTAGGGGAGTAGGCACCGCTTGC...TTAT..CCTAA-TGCTC..............................-------C..TA..--CA.....-- 9609
MND_2.x.x.5440           CTTGTGGCTCTGG-TAGCTCG.TCAGGTG..TTCTGGAAAGGCTTACGGGAGGACACTCTTGCTTGGTCTTG..ATAGACCTCTAGCAATCCCTTAGGCCTGGAGAT.-GTGGGATTGACTACCGCTTGCTTGCTATTATTGCCC-------...C-TA-CTGATTAGA 9529
MND_2.CM.98.CM16         ......................................................................................................................................................................... 9518
MND_2.GA.x.M14           ......................................................................................................................................................................... 9529
OLC.CI.97.97CI12         ......................................................................................................................................................................... 8678
GRV.ET.x.GRI_677         -CT--.....TTG----C-TAT..................................................................AG---...................................------C....C---TCGC....-- 9522
VER.KE.x.TYO1_patent     -CT--.....TTG-----GAA...................................................................AG--....................................-------...------TG...CA-- 9675
VER.KE.x.9063            -CT--.....TTG----GGAA...................................................................AG--....................................------AC..T-----TG...CA-- 9713
VER.KE.x.AGM155          -CT--.....TTG-----GAA...................................................................AG--....................................-------...------TG...CA-- 9692
VER.DE.x.AGM3            -CT--.....TTG-----GAA...................................................................AG-.T...................................------T...---CG-TG...CA-- 9212
TAN.UG.x.TAN1            CCT--......T------GAA...................................................................GGCTATT.........G--....................................-------...G-----TG....C-- 9691
SAB.SN.x.SAB1C           -GG-AAGCCT-AGAGCACTCTGGTGAGGTCTTGGTAGACCCTCGCCTGGCGACTGGCCATTGCCAGTAGCAG..AGACTCC......GCTTGC...TTGCTT.GATTG-CTC...............................-------T...............-- 9949
MND_1.GA.x.MNDGB1        ................................AGATCCCCTTAGAGCAAGGACCAGAGTCCTGAGTGACTGG..GTCTGAGCACCTCACTCGGGGCTGATCACCTCGAGGTAGTGGAACTCCTTGCTTGCTTGCTATTGTCTTC.-------T.AAC-TAGAA..TTAGA 9212
DEB.CM.99.CM40           ......................................................................................................................................................................... 9227
DEB.CM.99.CM5            ......................................................................................................................................................................... 9158
TAL.CM.01.8023           ......................................................................................................................................................................... 8845
TAL.CM.00.266            ......................................................................................................................................................................... 9338
SUN.GA.98.L14            CCGAGCCTGAT-A----GGGGAGCTAGAGG............................................CTCCTTGCTTTGCAGTACAAG.........C-T--...................................-------...-C----.......-- 9888
WRC.CI.97.97CI14         ......................................................................................................................................................................... 9484
WRC.CI.98.98CI04         ......................................................................................................................................................................... 9461
WRC.GM.05.Pbt_05GM_X02   TTGTTAGGGTGT-TGCAG-T.............................................................................................................................-------...AG-CTTTTGCTTAGC 8658
LST.CD.88.524            ......................................................................................................................................................................... 8594
LST.CD.88.447            ......................................................................................................................................................................... 8588
LST.KE.x.lho7            G-T-AGCCTAGGGAGC---G.................................GGCTCCTAGCTTGCTACTT..TAAA...GCCTTC.........................................................-------....AA-TTGCTCATTG- 9849
SYK.KE.x.SYK173          -GG-TAGTGAGG-C-GGGTTTGGCCGCCCGGTCTTCGGCAGCTCTCCGTTGCTT............................GCTTTGCTTACGC.TGATCAGCC-.A...................................-------T.GTGGTAAACGCAAG-A 9554
SYK.KE.x.KE51            ......................................................................................................................................................................... 9068
COL.CM.x.CGU1            ......................................................................................................................................................................... 8728
```

```
                                Extensive secondary structure in this region?
                     3' LTR R repeat end 3' LTR U5 start                                                                                       3' LTR U5 end
H1B.FR.83.HXB2           GAG..TG.CTTCAA.GTAGTG...TGTGCCCGTC.TG.TT..GTG.TGACTCT...GGT..AA.CTAGAGA..............TCCCT..CAGA...................CCCTTTTAGTCAGTGTGGAAAATCTCTAGCA..   9719
H1A1.UG.85.U455          ---..GTG---G--.------..----------.--.--..A-T.-------..---..--.-------..............-----.----...................--AC-A---A-T.GTGTA-------------..   9178
H1B.US.90.WEAU160        ---.--.------.------..----------.--.--..---.-------..---..--.-------..............-----.----...................--A----------------------------..   9719
H1C.ET.86.ETH2220        ---.--..............................................................................................................................................   9031
H1D.CD.84.84ZR085        ....................................................................................................................................................   8975
H1F1.BE.93.VI850         ....................................................................................................................................................   8903
H1G.SE.93.SE6165         -CC.................................................................................................................................................   9074
H1H.CF.90.056            ---.--.---T.........................................................................................................................................   8953
H1J.SE.93.SE7887         ---..G-CGCATGC.AAGCC-...............................................................................................................................   8943
H1K.CM.96.MP535          ....................................................................................................................................................   8604
H101_AE.TH.90.CM240      ---.--.---A-........................................................................................................................................   9203
H102_AG.NG.x.IBNG        ---.--.------.------..----------.--A--.---.-------..---.--.-........................................................................................   9201
H1N.CM.95.YBF30          ....................................................................................................................................................   9182
H1O.BE.87.ANT70          ---.---.AG..........................................................................................................................................   9754
H1O.CM.91.MVP5180        ---.---.AGAAGC.AGT---...--CT-A..--TG..--CAACC.CTGG-G-...CTA..G-.......--TC..........................................................................   9793
CPZ.CD.90.ANT            ....................................................................................................................................................   9068
CPZ.CM.01.SIVcpzCAM13    ---..--CT-CA........................................................................................................................................   9284
CPZ.CM.05.SIVcpzEK505    ---..--CT-AA........................................................................................................................................   9132
CPZ.CM.05.SIVcpzMB66     ---..--A---A-.......................................................................................................................................   9281
CPZ.CM.98.CAM3           ---..--A---G..A-C----..------TA--CA..--C..A-.-A-----..---..--.-------TCC.................................CTCAGAA.-TA--A--A..GTA-GT-----------.....   9170
CPZ.GA.88.GAB2           ---..--CTC-A..A------..-----T-A--TG..--C..A-.-A-----..---..GT.-------TCC.................................CTCAGAC.TAAA.CATAA.G-A-----------------..   9146
CPZ.CM.98.CAM5           ....................................................................................................................................................   9261
CPZ.CM.05.SIVcpzMT145    ---..--CTGCA........................................................................................................................................   9208
CPZ.TZ.00.TAN1           ---..A-.TG-T.-......................................................................................................................................   9326
CPZ.CM.05.SIVcpzLB7      ---..--CTGCA........................................................................................................................................   9170
CPZ.US.85.CPZUS          ---..--CT--A..A------..------TA--.C..--C..A-.--------..---..--.-------TCC.................................CTCAGAA.-T---AA-A..GTA-GT--------------..   9781
CPZ.GA.88.GAB1           ---.TGTAT--G..A-C----..------TA--TA..--C..A-.AC-----..---..--.-------TCC.................................CTCAGAT.TAAA--ATA..GTCAA---------------..   9811
MON.CM.99.L1             ....................................................................................................................................................   9448
MON.NG.x.NG1             A-AAAAAAAAAA--AAA-AAAAAAAAAAAAAAAAAAA...............................................................................................................   8260
GSN.CM.99.CN166          ....................................................................................................................................................   9414
GSN.CM.99.CN71           ....................................................................................................................................................   9410
MUS_1.CM.01.1085         ....................................................................................................................................................   9420
MUS_1.CM.01.CM1239       ....................................................................................................................................................   9508
MUS_2.CM.01.CM1246       ....................................................................................................................................................   9484
MUS_2.CM.01.CM2500       ....................................................................................................................................................   9430
RCM.NG.x.NG411           TT-AACCAGAATC-.AGT---TG..CCCATTT-A.....TCTCTCA-TGAAACGACTCTGGGGT-G-GA..............TC--TCAG.......ATATTTGTGGCAGAAGT.AC-.G-AGGCTAAGAA------C-A-C--G.   9557
RCM.GA.x.GAB1            AGTTTAC.TG-A-GC.A----TGTGCCTGTTT-A...CCTCTCAGCAGT-AACGACTCTGGGGTAG-GA..............TC--TCAG.......ATTCTTGTGGCAGAAGAG-C--G..GGCTAAGAA--.T--C---C--GT   9465
DEN.CD.x.CD1             ....................................................................................................................................................   9679
                 3' LTR R repeat end 3' LTR U5 start                                                                                           3' LTR U5 end
MAC.US.x.239             AGA..A-TAAG-T-..GT---TGT-CCCAT-TCT.CCTAGCC-CC.GCCTGG-...CAAC.TCGG-..-CTCAATA....ATAAGAAGA-CCTGGTCTGTTAGGACCCTTTCTGCTT.TGGGAA-CCG-AGCA--------C------..  10279
H2A.GW.x.ALI             AGA..A-CAGGTT-.AAG---TGT-CCCAT-TCT.CCTAGTC-CC.GCCTGG-...CA-..TCGG-GTTC-CCTGA....GTAACAAGA-CCTGGTCTGTTAGGACCCTTTCTGCTT.TGGGAA-CGG--GCA--------C------..  10353
H2A.DE.x.BEN             AGA..A-CAAGTT-.AGT---TGT-CCCAT-TCT.CCTAGTC-CC.GCCTGG-...CA-..TCGG-GTTC-TCTGA....GTAACAAGA-CCTGGTCTGTTAGGACCCTTCTCGCTT.TGGGAATCCA--GCA--------C------..  10359
H2A.SN.x.ST              AGA..A-CA...........................................................................................................................................   9672
H2B.GH.86.D205           AGA..A-CAAGT-..AGT---TGT-CCCAT-TCT.CCTAGTC-C..GCCTGG-...CAC..TCGG-..GCTCCAC.....TGA.TAGA-CCTGGTCTGTTAGGACCCTTCTTGCTT.TGGGAA-CCG-AGCG--------C------..  10269
H2B.CI.x.EHO             AGA..A-CAAGAC-.AGT---TGT-CCCAT-TCT.CCTAGTC-CC.GCCTGG-...CA-..TCGG-..GCTCCAC......TA..GAAA-CCTGGTCTGTTAGGACCCCTTCTGCTT.TGGGAA-CCA--GCA--------C------..  10242
H2G.CI.92.ABT96          AGA..A-TAAG---GTGT---TGT............................................................................................................................   9599
H2U.FR.96.12034          ....................................................................................................................................................   9607
MAC.US.x.EMBL_3          AGA..A-TA...........................................................................................................................................   9264
SMM.US.x.H9              AGA..A-TA...........................................................................................................................................   9638
SMM.US.x.PGM53           AGA..A-TAAG---.GT---TGT-CCCAT-TCT.CCTAGTC-CC.GCCTGG-...CA-C.TCGG-..-CTCGAAT..CATAAGAAGA-CCTGGTCTGTTAGGACCCTTTCTGC....................................  10184
SMM.US.x.SIVsmH635F_L3   AGA..A-TAAG---.GT---TGT-CCCAT-TCT.CCTAGTC-CC.GCCTGG-...CA-C.TCGG-..-CTCGACA..CATAAAGAAGA-CCTGGTCTGTTAGGACCCTTTCTGCTT.TGGGAA-CCG--GCA--------C------..  10315
SMM.SL.92.SL92B          AGA..A-CAAGTC.......................................................................................................................................   9613
STM.US.x.STM             AGA..A-TAAG---.GT-G-TGT-CCCAT-TCT.CCTAGTC-CC.GCCTGG-...CAAC.TCGG-..-CTCAA............GA-CCTGGTCTGTTAGGACCCTTTCTGC....................................   9892
MNE.US.x.MNE027          AGA..A-TAAG-C-..GT---TGC-CCCAT-TCT.CCTAGTC-CC.......................................................................................................   9664
DRL.x.x.FAO              T--ATG-ATGAGT-.AGT--ACTC-TGT-TTT-G........T-CT-AGTGAGACTCTGGTT-CTGGAGA..............TC--TCAGACT-GGTGAGAGCTATCCTGAGCTTTAG-..GAGAGTAAA-A--------C----..   9731
MND_2.x.x.5440           ATAGCAAGTAAG-G.TACTC-TCCAT-TTGTC-TCAATAGAA...ACTCTGG-TACC-GAG-........................-----.T----TCTGTGGCAGAGATTTGATAT-TAG-G--AGTAGAGAA--------C----..   9647
MND_2.CM.98.CM16         ....................................................................................................................................................   9518
MND_2.GA.x.M14           ....................................................................................................................................................   9529
OLC.CI.97.97CI12         ....................................................................................................................................................   8678
GRV.ET.x.GRI_677         AGT..C-CTA-ATTG.G---CAAG--CT-ATTG-TGCGCCGA-CCT.CTAGAGG.....TG-A-CTCTCTTACTGGGTTCTCCTG-A--CAGGT.................................G-G..-G--A---C----..   9623
VER.KE.x.TYO1_patent     AGA..G.CT-ATCTGAGTCAAGTG-CCT-ATTGACG.CC-CAC-CTCTTGAACG.....GG-ATCTTCCTTACTGGGTTCTCTCTCT........................GACCCAGGCGAG--...........--A---C----..   9778
VER.KE.x.9063            AGA..G.CT-ATCTGAGTCAAGTGCCCT-ATTGACG.CC-CAC-C..A-GCAGG....GG--C-GTTCCTTACTGGGTTCTCTCTCTGACCCAG.................................GCG---G--A---C----..   9815
VER.KE.x.AGM155          AGA..G.CT-ATCTGAGTCAAGTGCCCT-ATTAATG.CC-CAC-C..TTGAACG....GG-GAAGTTCCTTACTGGGTTCTCTCTCAAACCCAG.................................GCGA--G--A---C----..   9794
VER.DE.x.AGM3            AG-.CA-AGACTT..--GAGAGA....CATT-C-TG-TGCATCCAGCACAAG..................TGAGGGAAGACCCTGATGGAAT--ACCATGGAGAAGTCTTGGTCTGGAAG.........-T--ATCCCATGT-----..   9325
TAN.UG.x.TAN1            A--..A-T--ATCTGAG.CAAGTGCC-CATT-CGCGCCC-TCC-CAG-............................GAACCCTCGTTACTGGG-...........TTTCTTATC-AGGC-G--..........-------C----..   9784
SAB.SN.x.SAB1C           AGTTTACAAG-A--GCA----TGTGCCCATTTATTCCTCAG.........................................GACAA-CCTGGTTACTAAGGATCCC.........TGA.........-AGAA------CT-----..  10036
MND_1.GA.x.MNDGB1         -CA.................................................................................................................................................   9215
DEB.CM.99.CM40           ....................................................................................................................................................   9227
DEB.CM.99.CM5            ....................................................................................................................................................   9158
TAL.CM.01.8023           ....................................................................................................................................................   8845
TAL.CM.00.266            ....................................................................................................................................................   9338
SUN.GA.98.L14            AG...-TTTACTGCAAGCAA-TGT-CCCGTGT--TCTC--TCTCC..............TTCTAA-CCCTGTTCGACTGCTCTCAG-TAGGGAGC-A...........TTACTGG--GCC-AGTGATCC-G--CTG-CGG------..  10006
WRC.CI.97.97CI14         ....................................................................................................................................................   9484
WRC.CI.98.98CI04         ....................................................................................................................................................   9461
WRC.GM.05.Pbt_05GM_X02   A--AAGCGTG-GGTTAGTT-TTGT---AAGAAGTAGAACCCT-G-AG-CT-.................................................................................................   8709
LST.CD.88.524            ....................................................................................................................................................   8594
LST.CD.88.447            ....................................................................................................................................................   8588
LST.KE.x.lho7            T--AAA-TACAAGCTAGT--CTCA-AGTAGTC--T-TC-CCTTC-CCC........TGGTTCAG-GATC.................T--CA-TAGAGAGATTGGAGCCTTGTGAT--GGGAC--GGCT-CCA-...............   9957
SYK.KE.x.SYK173          .....ACGTC-GCT.C-CA-CTGT-C---TG-GT.CACC-CCCGA.A-GTG.................................................................................................   9597
SYK.KE.x.KE51            ....................................................................................................................................................   9068
COL.CM.x.CGU1            ....................................................................................................................................................   8728
```