

VI

HIV-2/SIV Proteins

Contents

VI-1	Introduction	361
VI-2	Annotated Features	362
VI-3	Sequences	363
VI-4	Alignments	369
VI-4.1	Gag	369
VI-4.2	Pol	373
VI-4.3	Vif	380
VI-4.4	Vpx	382
VI-4.5	Vpr	383
VI-4.6	Tat	384
VI-4.7	Rev	385
VI-4.8	Env	387
VI-4.9	Nef	393

VI-1 Introduction

There are 104 complete or nearly complete genomes (sequences longer than 4,000 bases) in our database that are from the HIV-2 and SIV-Sooty mangabey group. Twenty-eight sequences are HIV-2 but there are duplicates, only 20 are from unique samples. Seventy-six are SIVs (36 SIVsmm, 37 SIVmac, 2 SIVmne, 1 SIVstm). Although all these viruses have Sooty mangabeys as the natural host, there have been cross-species transfers into humans to create the HIV-2 groups A through G as well as into captive macaques (Rhesus macaques = mac, Stump tailed macaques = stm and Pig tailed macaques = mne) by unintentional interspecies interactions. Intentional cross species transfers of virus remain labeled as being from the original host (e.g., HIV-1 injected into a Chimpanzee is labeled as HIV-1 and not as SIVcpz; SIVsmm injected into a Rhesus macaque is labeled as SIVsmm and not SIVmac). Sixty-four of the 104 complete genomes are presented in the complete genomes alignment in this compendium, the others were not included because they are replicates of the same strain or isolate. These same 64 sequences are present in most of the HIV-2/SIVsmm protein alignments except that in proteins such as Env and Nef, where a large set of sequence entries containing that gene only and not the complete genome were available. For those proteins, we chose a diverse set including the sub-genomic sequences, and when more sequences than would fit on a page were available, highly similar sequences from the complete genome set of 64 were dropped (for example some of the several SIVmac sequences).

VI-2 Annotated Features

Features of HIV-2 annotated in the alignment that follows.

Feature	Protein	Location	Page
Gag p15 start	Gag	1	369
p15	Gag	135	369
p27	Gag	136	369
p27	Gag	364	371
p2	Gag	365	371
p2	Gag	381	371
p8	Gag	382	371
p8	Gag	433	371
p1	Gag	434	371
p1	Gag	447	371
p6	Gag	448	371
PTAP motif	Gag	458-461	371
PSAP in HIV-2 B, U	Gag	476-479	371
p6 end	Gag	510	372
Gag end	Gag	510	372
Pol p15 start (-1 from Gag)	Pol	1	373
p15	Pol	67	373
protease	Pol	68	373
protease	Pol	166	374
p51 RT start	Pol	167	374
D catalytic site	Pol	276	374
DD catalytic site	Pol	351	375
p51 RT end	Pol	605	376
p15 RNase H start	Pol	606	376
p15 RNase H end	Pol	725	377
p31 Integrase start	Pol	726	377
p31 Integrase end	Pol	1018	379
Pol end	Pol	1018	379
Vif start	Vif	1	380
Vif end	Vif	214	381
Vpx start	Vpx	1	382
Vpx end	Vpx	112	382
Vpr start	Vpr	1	383
Vpr end	Vpr	101	383
Tat start	Tat	1	384
exon 1 end	Tat	99	384
exon 2 start	Tat	100	384
Tat end	Tat	130	384
Rev start	Rev	1	385
exon 1 end	Rev	24	385
exon 2 start	Rev	25	385
Rev end	Rev	107	385
Env start	Env	1	387
signal peptide end	Env	22	387
gp120 start	Env	23	387
V3	Env	311-344	388
gp120 end	Env	525	390
gp41 start	Env	526	390

Feature	Protein	Location	Page
Env end	Env	879	392
Nef start	Nef	1	393
R17Y mutation	Nef	17	393
max HIV-1 similarity	Nef	153-182	393
premature stop in Mac239	Nef	93	393
normal Nef end	Nef	262	394

VI-3 Sequences

Sequences included in the HIV-2/SIV protein alignments.

Name	Accession	Proteins	Author	Reference
H201_AE.US.91.PA	L33093	Pol	Gao, F	<i>J Virol</i> 68 (11); 7433-47 (1994)
H2A.CI.88.UC2	U38293	All	Barnett, SW	<i>Virology</i> 222 (1); 257-61 (1996)
H2A.CI.x.IC763124	U76641	Nef	Switzer, WM	<i>J Infect Dis</i> 177 (1); 65-71 (1998)
H2A.DE.91.HOM	U73757	Nef	Fackler, OT	<i>Eur J Biochem</i> 247 (3); 843-51 (1997)
H2A.DE.92.NEP	U73758	Nef	Fackler, OT	<i>Eur J Biochem</i> 247 (3); 843-51 (1997)
H2A.DE.x.BEN	M30502	All	Kirchhoff, F	<i>Virology</i> 177 (1); 305-11 (1990)
H2A.DE.x.PEI2	U22047	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Nef	Talbott, R	<i>PNAS USA</i> 90 (9); 4226-30 (1993)
H2A.FR.x.96206	AF170048	Env	Damond, F	<i>Virology</i> 280 (1); 19-30 (2001)
H2A.FR.x.96226	AF170030	Env	Damond, F	<i>Virology</i> 280 (1); 19-30 (2001)
H2A.FR.x.96330	AF170047	Env	Damond, F	<i>Virology</i> 280 (1); 19-30 (2001)
H2A.GH.x.GH1	M30895	All	Hasegawa, A	<i>ARHR</i> 5 (6); 593-604 (1989)
H2A.GM.87.D194	J04542	All	Kuehnel, H	<i>PNAS USA</i> 86 (7); 2383-7 (1989)
H2A.GM.90.CBL24	U05353	Env	Breuer, J	<i>J Gen Virol</i> 1995 Feb;76(Pt 2):333-45
H2A.GM.x.CBL23	U05352	Env	Breuer, J	<i>J Gen Virol</i> 1995 Feb;76(Pt 2):333-45
H2A.GM.x.ISY	J04498	All	Franchini, G	<i>PNAS USA</i> 86 (7); 2433-7 (1989)
H2A.GM.x.MCN13	AY509259	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev	Schmitz, C	<i>J Virol</i> 78 (4):2006-2016 (2004)
H2A.GW.86.FG	J03654	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env	Zagury, JF	<i>PNAS USA</i> 85 (16); 5941-5 (1988)
H2A.GW.87.CAM2CG	D00835	All	Tristem, M	<i>J Gen Virol</i> 1991 Mar;72(Pt 3):721-4
H2A.GW.x.ALI	AF082339	All	Azevedo-Pereira, JM	Unpublished (1998)
H2A.GW.x.CAM1	U05359	Env	Breuer, J	<i>J Gen Virol</i> 1995 Feb;76(Pt 2):333-45
H2A.GW.x.MDS	Z48731	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Nef	Becker, M	Unpublished (1995)
H2A.PT.x.1069	AJ344389	Nef	Padua, E	<i>J Gen Virol</i> 2003 May;84(Pt 5):1287-99

Name	Accession	Proteins	Author	Reference
H2A.PT.x.1147	AJ344390	Nef	Padua, E	<i>J Gen Virol</i> 2003 May;84(Pt 5):1287-99
H2A.PT.x.1215	AJ344393	Nef	Padua, E	<i>J Gen Virol</i> 2003 May;84(Pt 5):1287-99
H2A.PT.x.1227	AJ344391	Nef	Padua, E	<i>J Gen Virol</i> 2003 May;84(Pt 5):1287-99
H2A.PT.x.1320	AJ344394	Nef	Padua, E	<i>J Gen Virol</i> 2003 May;84(Pt 5):1287-99
H2A.PT.x.1378	AJ344414	Nef	Padua, E	<i>J Gen Virol</i> 2003 May;84(Pt 5):1287-99
H2A.PT.x.1428	AJ344408	Nef	Padua, E	<i>J Gen Virol</i> 2003 May;84(Pt 5):1287-99
H2A.PT.x.1543	AJ344405	Nef	Padua, E	<i>J Gen Virol</i> 2003 May;84(Pt 5):1287-99
H2A.PT.x.1544	AJ344407	Nef	Padua, E	<i>J Gen Virol</i> 2003 May;84(Pt 5):1287-99
H2A.PT.x.1567	AJ344409	Nef	Padua, E	<i>J Gen Virol</i> 2003 May;84(Pt 5):1287-99
H2A.PT.x.268	AJ344410	Nef	Padua, E	<i>J Gen Virol</i> 2003 May;84(Pt 5):1287-99
H2A.PT.x.483	AJ344401	Nef	Padua, E	<i>J Gen Virol</i> 2003 May;84(Pt 5):1287-99
H2A.PT.x.546	AJ344403	Nef	Padua, E	<i>J Gen Virol</i> 2003 May;84(Pt 5):1287-99
H2A.PT.x.794	AJ344388	Nef	Padua, E	<i>J Gen Virol</i> 2003 May;84(Pt 5):1287-99
H2A.PT.x.B1_1	AJ344406	Nef	Padua, E	<i>J Gen Virol</i> 2003 May;84(Pt 5):1287-99
H2A.PT.x.EP	AJ344387	Nef	Padua, E	<i>J Gen Virol</i> 2003 May;84(Pt 5):1287-99
H2A.PT.x.MP1	AJ344385	Nef	Padua, E	<i>J Gen Virol</i> 2003 May;84(Pt 5):1287-99
H2A.PT.x.MP2	AJ344386	Nef	Padua, E	<i>J Gen Virol</i> 2003 May;84(Pt 5):1287-99
H2A.SN.85.ROD	M15390	All	Clavel, F	<i>Nature</i> 324 (6098); 691-5 (1986)
H2A.SN.x.ST	M31113	Env, Nef	Kumar, P	<i>J Virol</i> 64 (2); 890-901 (1990)
H2A.x.x.JAU2	L28936	Vif		<i>ARHR</i> 14 (5); 465-9 (1998)
H2AB.CI.90.7312A	L36874	All	Gao, F	Unpublished
H2B.CI.88.UC1	L07625	All	Barnett, SW	<i>J Virol</i> 67 (2); 1006-14 (1993)
H2B.CI.x.EHO	U27200	All	Rey-Cuille, MA	<i>Virology</i> 202 (1); 471-6 (1994)
H2B.CI.x.IC762993	U76639	Nef	Switzer, WM	<i>J Infect Dis</i> 177 (1); 65-71 (1998)
H2B.FR.x.96200	AF170057	Env	Damond, F	<i>Virology</i> 280 (1); 19-30 (2001)
H2B.FR.x.97227	AF170052	Env	Damond, F	<i>Virology</i> 280 (1); 19-30 (2001)
H2B.GH.86.D205	X61240	All	Kreutz, R	<i>ARHR</i> 8 (9); 1619-29 (1992)
H2B.JP.01.KR020	AB100245	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env	Oka, S-I	<i>ARHR</i> 19 (11); 1045-9 (2003)
H2C.LR.x.2238	M87138	Pol	Gao, F	<i>Nature</i> 358 (6386); 495-9 (1992)
H2D.LR.90.FO784PA	M87110	Pol	Gao, F	<i>Nature</i> 358 (6386); 495-9 (1992)
H2D.LR.90.FORTC2	M87111	Pol	Gao, F	<i>Nature</i> 358 (6386); 495-9 (1992)
H2G.CI.x.ABT96	AF208027	All	Brennan, CA	<i>ARHR</i> 13 (5); 401-4 (1997)
H2U.FR.96.12034	AY530889	All	Damond, F	<i>ARHR</i> 20 (6); 666-72 (2004)

Name	Accession	Proteins	Author	Reference
MAC.US.x.17EC1	AY033233	Vif, Vpx, Vpr, Tat, Rev, Nef	Anderson, MG	<i>Virology</i> 195 (2); 616-26 (1993)
MAC.US.x.17EFR	AY033146	Vif, Vpx, Vpr, Tat, Rev, Nef	Flaherty, MT	<i>J Virol</i> 71 (8); 5790-8 (1997)
MAC.US.x.1937	AY611495	Vif, Vpx, Vpr, Tat, Rev	O'Connor, DH	<i>J Virol</i> 78 (24):14012-14022 (2004)
MAC.US.x.2065	AY611493	Vif, Vpx, Vpr, Tat, Rev	O'Connor, DH	<i>J Virol</i> 78 (24):14012-14022 (2004)
MAC.US.x.239	M33262	All	Kestler, H	<i>Science</i> 248 (4959); 1109-12 (1990)
MAC.US.x.251_1A11	M76764	Vif, Vpx, Vpr, Tat, Rev, Nef	Planelles, V	<i>ARHR</i> 7 (11); 889-98 (1991)
MAC.US.x.251_32H_PJ5	D01065	Vif, Vpx, Vpr, Tat, Rev, Nef	Rud, EW	<i>J Gen Virol</i> 1994 Mar;75(Pt 3):529-43
MAC.US.x.251_BK28	M19499	All	Franchini, G	<i>Nature</i> 328 (6130); 539-43 (1987)
MAC.US.x.270W	AY290712	Env	Buckley, KA	<i>Virology</i> 312 (2); 470-80 (2003)
MAC.US.x.418	AY302466	Env	Mansfield, KG	Unpublished
MAC.US.x.80035	AY611486	Vif, Vpr, Tat, Rev	O'Connor, DH	<i>J Virol</i> 78 (24):14012-14022 (2004)
MAC.US.x.81035	AY599200	Vif, Vpx, Vpr, Tat, Rev	O'Connor, DH	<i>J Virol</i> 78 (24):14012-14022 (2004)
MAC.US.x.85013	AY611490	Vif, Vpx, Vpr, Tat, Rev	O'Connor, DH	<i>J Virol</i> 78 (24):14012-14022 (2004)
MAC.US.x.92050	AY603959	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev	O'Connor, DH	<i>J Virol</i> 78 (24):14012-14022 (2004)
MAC.US.x.92077	AY599201	Vif, Vpx, Vpr, Tat, Rev	O'Connor, DH	<i>J Virol</i> 78 (24):14012-14022 (2004)
MAC.US.x.93057	AY611492	Vif, Vpx, Vpr, Tat, Rev	O'Connor, DH	<i>J Virol</i> 78 (24):14012-14022 (2004)
MAC.US.x.93062	AY607704	Vpx, Vpr, Rev	O'Connor, DH	<i>J Virol</i> 78 (24):14012-14022 (2004)
MAC.US.x.95058	AY611494	Vif, Vpx, Vpr	O'Connor, DH	<i>J Virol</i> 78 (24):14012-14022 (2004)
MAC.US.x.95086	AY607703	Vif, Vpx, Vpr, Tat, Rev	O'Connor, DH	<i>J Virol</i> 78 (24):14012-14022 (2004)
MAC.US.x.95112	AY588946	Vif, Vpx, Vpr, Tat, Rev	O'Connor, DH	<i>J Virol</i> 78 (24):14012-14022 (2004)
MAC.US.x.96016	AY607701	Vif, Vpx, Vpr, Tat, Rev	O'Connor, DH	<i>J Virol</i> 78 (24):14012-14022 (2004)
MAC.US.x.96020	AY611488	Vif, Vpx, Vpr, Tat, Rev	O'Connor, DH	<i>J Virol</i> 78 (24):14012-14022 (2004)
MAC.US.x.96072	AY611491	Vif, Vpx, Vpr, Tat, Rev	O'Connor, DH	<i>J Virol</i> 78 (24):14012-14022 (2004)
MAC.US.x.96081	AY597209	Vif, Vpx, Tat, Rev	O'Connor, DH	<i>J Virol</i> 78 (24):14012-14022 (2004)
MAC.US.x.96093	AY611489	Vif, Vpx, Vpr, Tat, Rev	O'Connor, DH	<i>J Virol</i> 78 (24):14012-14022 (2004)
MAC.US.x.96114	AY588945	Vif, Vpx, Vpr, Tat, Rev	O'Connor, DH	<i>J Virol</i> 78 (24):14012-14022 (2004)

Name	Accession	Proteins	Author	Reference
MAC.US.x.96123	AY611487	Vif, Vpx, Vpr, Tat, Rev	O'Connor, DH	<i>J Virol</i> 78 (24):14012-14022 (2004)
MAC.US.x.96135	AY607702	Vpx, Vpr, Tat	O'Connor, DH	<i>J Virol</i> 78 (24):14012-14022 (2004)
MAC.US.x.97009	AY599199	Vpr, Tat	O'Connor, DH	<i>J Virol</i> 78 (24):14012-14022 (2004)
MAC.US.x.97074	AY599198	Vpr	O'Connor, DH	<i>J Virol</i> 78 (24):14012-14022 (2004)
MAC.US.x.BK28_H824	U86638	Vif, Vpx, Tat, Rev, Env, Nef	Edmonson, P	<i>J Virol</i> 72 (1); 405-14 (1998)
MAC.US.x.BR5	AY290716	Env	Buckley, KA	<i>Virology</i> 312 (2); 470-80 (2003)
MAC.US.x.BR5	AY290711	Env	Buckley, KA	<i>Virology</i> 312 (2); 470-80 (2003)
MAC.US.x.MAC239_87082	AY600249	Vif, Vpx, Vpr, Tat, Rev	O'Connor, DH	<i>J Virol</i> 78 (24):14012-14022 (2004)
MAC.US.x.MM142	Y00277	All	Chakrabarti, L	<i>Nature</i> 328 (6130); 543-7 (1987)
MAC.US.x.SMM142B	BD131285	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Env	Alizon, M	Patent: JP 2002030099-A 2 29-JAN-2002; INSTITUT PASTEUR
MAC.US.x.r80025	AY576480	Vpr	O'Connor, DH	<i>J Virol</i> 78 (24):14012-14022 (2004)
MAC.US.x.r90131	AY576481	Vpr, Rev	O'Connor, DH	<i>J Virol</i> 78 (24):14012-14022 (2004)
MNE.US.82.MNE_8	M32741	All	Kimata, JT	<i>J Virol</i> 72 (1); 245-56 (1998)
MNE.US.x.MNE027	U79412	All	Kimata, JT	<i>J Virol</i> 72 (1); 245-56 (1998)
SMM.SL.92.SL92B	AF334679	All	Chen, Z	<i>J Virol</i> 70 (6); 3617-27 (1996)
SMM.US.02.YNPRC_FAL	AY965498	Gag	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.02.YNPRC_FAL	AY965402	Env	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.02.YNPRC_FBL	AY965462	Pol	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.02.YNPRC_FBL	AY965395	Env	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.02.YNPRC_FCP	AY965499	Gag	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.02.YNPRC_FIP	AY965461	Pol	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.02.YNPRC_FIP	AY965404	Env	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.02.YNPRC_FKL	AY965497	Gag	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.02.YNPRC_FUP	AY965500	Gag	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.02.YNPRC_FUP	AY965405	Env	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.02.YNPRC_FVN	AY965490	Gag	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.02.YNPRC_FVN	AY965431	Env	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.02.YNPRC_FWS	AY965488	Gag	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.02.YNPRC_FWS	AY965394	Env	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.02.YNPRC_FYN	AY965507	Gag	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.02.YNPRC_FYN	AY965410	Env	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.03.TNPRC_D215	AY965511	Gag	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.03.TNPRC_D215	AY965445	Pol	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.03.TNPRC_G932	AY965477	Pol	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.03.TNPRC_M924	AY965481	Gag	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.03.TNPRC_M934	AY965491	Gag	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.77.CNPRC_CFU212	AY965512	Gag	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.79.NIRC_6001_G930	AY965463	Pol	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.79.NIRC_6007_G932	AY965466	Pol	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.79.NIRC_6007_G932	AY965415	Env	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.79.NIRC_CFU212	AY965468	Pol	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.79.NIRC_CFU233	AY965473	Pol	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.79.NIRC_CFU233	AY965422	Env	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.80.NIRC_6001_G930	AY965413	Env	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.81.NIRC_CFU233	AY965515	Gag	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.81.TNPRC_G930	AY965434	Pol	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.82.NIRC_6007_G932	AY965467	Pol	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.84.TNPRC_F104	AY965475	Pol	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)

Name	Accession	Proteins	Author	Reference
SMM.US.85.TNPRC_F102	AY965370	Gag	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.85.TNPRC_F104	AY965476	Pol	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.86.NIRC_6001_G930	AY965464	Pol	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.86.NIRC_6002_G931	AY965465	Pol	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.86.NIRC_CFU212	AY965513	Gag	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.86.NIRC_CFU226	AY965514	Gag	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.86.NIRC_CFU226	AY965471	Pol	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.86.NIRC_CFU232	AY965472	Pol	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.86.NIRC_CFU233	AY965474	Pol	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.89.TNPRC_G930	AY965509	Gag	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.90.TNPRC_F100	AY965496	Gag	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.90.TNPRC_F100	AY965397	Env	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.91.TNPRC_G931	AY965435	Pol	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.93.TNPRC_D178	AY965505	Gag	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.93.TNPRC_E042	AY965486	Gag	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.93.TNPRC_F102	AY965412	Env	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.93.TNPRC_M920	AY965357	Gag	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.94.TNPRC_G080	AY965455	Pol	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.94.TNPRC_M927	AY965456	Pol	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.94.TNPRC_M927	AY965392	Env	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.95.TNPRC_D171	AY965484	Gag	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.95.TNPRC_D175	AY965377	Gag	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.95.TNPRC_D175	AY965441	Pol	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.95.TNPRC_D175	AY965419	Env	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.95.TNPRC_D176	AY965489	Gag	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.95.TNPRC_D177	AY965443	Pol	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.95.TNPRC_E045	AY965506	Gag	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.95.TNPRC_F102	AY965453	Pol	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.95.TNPRC_G932	AY965508	Gag	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.95.TNPRC_G932	AY965436	Pol	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.95.TNPRC_M939	AY965485	Gag	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.95.TNPRC_M942	AY965501	Gag	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.97.TNPRC_G080	AY965363	Gag	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.97.TNPRC_G080	AY965398	Env	Apetrei, C	<i>J Virol</i> 79 (14); 8991-9005 (2005)
SMM.US.x.62K	U04989	Vpr	Hirsch, VM	<i>J Virol</i> 68 (4); 2649-61 (1994)
SMM.US.x.BPZ_m12	AY603050	Vif, Vpx, Tat, Rev	Glenn, AA	<i>Virology</i> 325 (2); 297-307 (2004)
SMM.US.x.F236_H4	X14307	All	Hirsch, VM	<i>Nature</i> 339 (6223); 389-92 (1989)
SMM.US.x.H445	AY221509	Env	Dehghani, H	<i>J Virol</i> 77 (11); 6405-18 (2003)
SMM.US.x.H9	M80194	All	Courgnaud, V	<i>J Virol</i> 66 (1); 414-9 (1992)
SMM.US.x.P209C15	L20009	Env	Hynes, NA	<i>ARHR</i> 9 (8); 803-6 (1993)
SMM.US.x.PBJ14_15	L03295	All	Dewhurst, S	<i>Nature</i> 345 (6276); 636-40 (1990)
SMM.US.x.PBJA	M31325	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Nef	Dewhurst, S	<i>Nature</i> 345 (6276); 636-40 (1990)
SMM.US.x.PBJC	L03296	Vif, Vpx, Tat, Rev, Nef	Dewhurst, S	<i>Nature</i> 345 (6276); 636-40 (1990)
SMM.US.x.PBJD	L03297	Vif, Vpx, Tat, Rev, Nef	Dewhurst, S	<i>Nature</i> 345 (6276); 636-40 (1990)
SMM.US.x.PBJE	L03298	Vif, Vpx, Tat, Rev, Nef	Dewhurst, S	<i>Nature</i> 345 (6276); 636-40 (1990)

Name	Accession	Proteins	Author	Reference
SMM.US.x.PBJ_143	M80193	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Nef	Courgnaud, V	<i>J Virol</i> 66 (1); 414-9 (1992)
SMM.US.x.PBJ_6P12	L09211	Env, Nef	Novembre, FJ	<i>J Virol</i> 67 (5); 2466-74 (1993)
SMM.US.x.PBJ_6P6	L09212	Gag, Pol, Vif, Vpx, Vpr, Tat, Rev, Nef	Novembre, FJ	<i>J Virol</i> 67 (5); 2466-74 (1993)
SMM.US.x.PBJ_6P9	L09213	Nef	Novembre, FJ	<i>J Virol</i> 67 (5); 2466-74 (1993)
SMM.US.x.PGM53	AF077017	All	Novembre, FJ	<i>J Virol</i> 72 (11); 8841-51 (1998)
SMM.US.x.PT573	AY221511	Env	Dehghani, H	<i>J Virol</i> 77 (11); 6405-18 (2003)
SMM.US.x.PT583	AY221512	Env	Dehghani, H	<i>J Virol</i> 77 (11); 6405-18 (2003)
SMM.US.x.SME543	U72748	All	Hirsch, VM	<i>J Virol</i> 71 (2); 1608-20 (1997)
SMM.US.x.SMP209	L20008	Env	Hynes, NA	<i>ARHR</i> 9 (8); 803-6 (1993)
STM.US.x.STM	M83293	All	Novembre, FJ	<i>Virology</i> 186 (2); 783-7 (1992)

	Gag p15 start		p15_p27	
MAC.US.x.239	MGVRRNSVLSGKKADELEKIRLPNGKKKYML	KHVWAA	NELDRFLGAESLLENKEGCQKILSVLAPLVPTGSENLSKLYNTVCVITWCIHAEKVKHTEEAQIVQRHLVVTGTTETMPKTSRPTAPSSG	RGNGYYPVQQ
H2A.GM.x.MCN13	-A-K-T-G-R-I-S-R-T-G		D-G-AA-A-K-N-P-K-F-V-T-V	166
H2A.CI.88.UC2	-A-R-R-G-R-I-R-N-S-T-D		F-L-D-KL-A-A-K-N-P-K-F-A-V	166
H2A.DE.x.BEN	-A-R-V-G-R-I-K-G-S-T-D		F-L-D-KL-A-A-K-N-P-KR-A-S-V	166
H2A.DE.x.PE12	-A-R-V-G-R-I-K-G-S-T-D		F-L-D-KL-A-A-K-N-P-KR-A-S-V	166
H2A.GH.x.GH1	-A-R-V-G-R-I-K-G-S-T-D		F-L-D-KL-GA-A-K-S-P-R-F-TG-I-V	167
H2A.GM.x.ISY	-AK-R-G-R-I-T-S-I-E		F-L-D-R-G-A-E-A-K-NI-P-K-F	166
H2A.GM.87.D194	-A-R-R-V-R-R-I-S-R-K-E		F-L-D-KLA-A-A-K-NI-P-K-F-A-I-V	166
H2A.GW.x.ALJ	-A-R-R-R-G-Q-I-S-R-K-E		F-V-V-D-G-I-AA-IE-A-K-S-P-E-Q-F-N-T-V-D	166
H2A.GW.x.MJ5	-A-S-R-V-G-R-I-S-T-D		F-F-D-G-KFAR-A-A-K-N-P-K-F-VA-T-I-G	164
H2A.GW.86.FG	-A-R-V-G-K-I-S-R-K-D		F-F-D-G-R-AL-AA-A-K-D-P-K-S-T-V	166
H2A.GW.87.CAM2C9	-A-R-R-G-R-I-K-S-T-D-M		F-F-D-G-R-A-A-A-K-S-E-K-H-V-T-I	166
H2A.SN.85.ROD	-A-R-R-G-R-I-K-S-T-D-M		F-F-D-G-R-A-A-A-K-S-E-K-H-V-T-I	166
H2AB.CI.90.7312A	-A-G-T-V-G-I-N-S-T-M		F-L-K-D-KLA-S-D-KT-TADK-A-T-S-VA-V	161
H2B.CI.x.EHO	-A-G-T-V-G-R-I-V-E-R-GS-R-RK-G		F-F-L-D-K-A-AAD-K-AM-K-SK-T-R-L-A-A-S	161
H2B.CI.88.UC1	-A-S-T-V-G-R-C-I-I-V-S-H-T		F-Y-L-D-K-A-AAD-K-A-P-G-A-M	161
H2B.GH.86.D205	-A-G-T-V-G-K-I-S-H-K		F-I-Y-L-D-K-A-AAD-K-A-NK-P-G-LA	161
H2B.JP01.KR020	-A-G-T-V-G-C-I-V-Y-S-H-T-E		F-Y-L-V-D-K-A-S-AAN-KA-A-K-VA-T	161
H2G.CI.x.ABT96	-A-S-T-V-G-S-L-Q		D-V-S-KV-A-I-P-V	166
H2U.FR.96.12034	-A-X-V-V-A		L-Q-A-E-A-K-A-X-P-G-A	166
MAC.US.x.251_BK28	-A-G		L-A	166
MAC.US.x.MM142	-A-G		M-A-F-T	166
MAC.US.x.SMM142B	-A-G		M-A-F-T	166
MAC.US.x.92050	-X-G-R			166
MNE.US.x.MNE027	-A-G-D-A		A-KR-V-T	166
MNE.US.82.MNE_8	-A-G-D-A		A-KR-V-T	166
SMM.SL.92.SL92B	-A-G-V-G-R-II-R-S-S-R-A-M-FS-V-L-M-D-KT-S-S-A-KL-AQ-P-V-N-T			164
SMM.US.02.YNPRC_FWS	-A-V-S-R-F-I-R-I-I-ADK-V-P-R-V-T			166
SMM.US.95.TNPRC_G932	-A-V-T-I-I-L-V-V-ADK-I-P-K-V-T			166
SMM.US.02.YNPRC_FAL	-A-V-T-R-I-I-L-V-V-ADN-P-R-V-T			166
SMM.US.85.TNPRC_F102	-A-S-V-G-X-R-I-I-ADP-S-P-A-R-V			166
SMM.US.03.TNPRC_D215	-A-S-V-G-II-R-R-I-E-AG-P-M-R-V-I-Q			166
SMM.US.95.TNPRC_D175	-A-V-X-R-R-L-Q-ADK-A-P-R-V-T			139
SMM.US.93.TNPRC_M920	-A-V-X-IFX-R-I-L-R-ADK-A-P-R-V-T			140
SMM.US.97.TNPRC_G080	-A-V-X-F-R-Q-I-L-R-DN-P-R-V-T			139
SMM.US.90.TNPRC_F100	-A-V-R-T-R-I-S-F-L-R-V-ADK-A-P-R-V-T			166
SMM.US.89.TNPRC_G930	-A-V-R-T-R-I-S-F-L-R-V-ADK-A-P-R-V-T			166
SMM.US.03.TNPRC_M934	-A-V-R-L-X-T-R-IA-ADK-A-P-R-V-T			166
SMM.US.95.TNPRC_D171	-A-V-R-L-X-T-R-IA-ADK-A-P-R-V-T			166
SMM.US.95.TNPRC_D176	-A-V-R-L-X-T-R-IA-ADK-A-P-R-V-T			166
SMM.US.93.TNPRC_E042	-A-H-V-S-R-T-R-ADK-I-L-P-R-V-T			166
SMM.US.95.TNPRC_M939	-A-H-K-S-T-R-I-I-V-K-ADK-I-P-R-V-T			166
SMM.US.03.TNPRC_M924	-A-H-K-S-T-R-I-I-V-K-ADK-I-P-R-V-T			166
SMM.US.86.NIRC_CFU226	-A-S-V-G-R-I-ADP-S-P-M-R-V			166
SMM.US.86.NIRC_CFU212	-A-S-V-G-R-I-ADP-S-P-M-K-V			166
SMM.US.77.CNPRC_CFU212	-A-S-V-G-R-I-ADP-S-P-M-K-V			166
SMM.US.95.TNPRC_B045	-A-S-V-R-X-R-I-R-R-ADN-P-T-R-A			166
SMM.US.95.TNPRC_M942	-A-S-V-R-X-R-I-R-R-ADN-P-T-R-A			166
SMM.US.02.YNPRC_FUP	-A-V-R-S-R-I-L-R-V-E-DNL-P-R-V			166
SMM.US.02.YNPRC_FKL	-A-V-R-S-R-I-L-R-V-E-DNL-P-R-V			166
SMM.US.02.YNPRC_FYN	-A-S-V-R-S-R-I-L-R-V-E-AD-P-A-R-V			166
SMM.US.93.TNPRC_D178	-A-S-V-R-S-R-I-L-R-V-E-AD-P-A-R-V			166
SMM.US.02.YNPRC_FCP	-A-V-G-LF-S-R-I-L-R-V-E-ADN-P-R-V			166
SMM.US.02.YNPRC_FVN	-A-V-G-LF-S-R-I-L-R-V-E-ADN-P-R-V			166
SMM.US.x.F236_H4	-A-E-V-D-R-ADK-A-P-R-V-T			166
SMM.US.x.H9	-A-G-R-K-I-R-L-ADK-A-P-R-V-T			166
SMM.US.x.PBJ14_15	-A-G-R-K-I-R-L-ADK-A-P-R-V-T			166
SMM.US.x.PBJA	-A-G-R-Q-I-R-L-ADK-A-P-K-V-T			166
SMM.US.x.PBJ_143	-A-G-R-Q-I-R-L-ADK-A-P-K-V-T			166
SMM.US.x.PBJ_6P6	-A-G-R-K-I-R-L-ADK-A-P-K-V-T			166
SMM.US.x.PGM53	-A-E-V-D-T-I-A-ADK-I-P-R-V-T			166
SMM.US.x.SME543	-A-E-V-D-T-I-A-ADK-A-P-R-V-T			166
STM.US.x.STM	-A-S-E-V-G-S-IT-E-F-V-K-ANK-A-P-R-V			166

MAC.US.x.239	GAEVVPGFQALSEGCTPYDINQMLNCVGDHQAAMQIIRDIIINEEAADWLQH..PQ..PAPQQQLREPSGSDIAGTTSVDEQIQWMYRQNPVPGVNIYRRWIQLGQKCVRMYPNTNILDVKQGPKEPFQSYVDRFYKSLRAEQTDAAVKNMNTQTLLIQNANPDKC	332
H2A.GM.x.MCN13	-----E-V-----V-I.PGPLPA-----R-----T-----F-P-V-----I-----I-----P-----V-----	333
H2A.CL88.UC2	-----Q-----A-I.PGPLPA-----D-R-----T-----P-----V-----I-----K-----V-----S-----A-----P-----V-----	333
H2A.DE.x.BEN	-----E-----S-I.PGPLPA-----D-R-----T-----P-----V-----I-----K-----V-----I-----I-----P-----V-----	333
H2A.DE.x.PE12	-----E-----V-I.PGPLPA-----R-----T-----E-----F-A-V-----I-----I-----K-----V-----I-----I-----P-----V-----	333
H2A.GH.x.GH1	-----E-----D-A-I.PGPLPA-----D-R-----T-----E-----P-----V-----I-----I-----I-----I-----P-----V-----	334
H2A.GM.x.ISY	-----E-----V-I.PGPLPA-----D-R-----T-----E-----E-----E-V-----I-----I-----I-----S-----P-----V-----S-----	332
H2A.GM.87.D194	-----E-----A-I.PGPLPA-----D-R-----T-----E-----P-----V-----I-----I-----I-----S-----P-----V-----	333
H2A.GW.x.AL1	-----E-----VA-I.PGPLPA-----R-----T-----E-----F-PR-V-----I-----I-----I-----S-----P-----V-----	333
H2A.GW.x.M14S	-----E-----A-I.PGPLPA-----R-----T-----E-----F-A-V-----I-----I-----I-----I-----P-----V-----	333
H2A.GW.86.FG	-----E-----VA-I.PGPLPA-----R-----T-----E-----F-P-V-----I-----I-----I-----IN-----P-----V-----	331
H2A.GW.87.CAM2C6	-----E-----AN-I.PGPLPA-----D-R-----T-----E-----F-A-V-----I-----I-----I-----S-----P-----V-----	333
H2A.SN.85.ROD	-----E-----E-V-I.PGPLPA-----R-----T-----E-----F-P-V-----I-----I-----I-----S-----P-----V-----	333
H2AB.CL90.7312A	-----E-----V-E-----Q-V.PGPLPA-----D-R-----T-----E-----G-S-----I-----I-----A-T-RA-E-V-----	328
H2B.CL.x.EHO	-----E-----E-----Q-S.PGMPA-----R-----T-----E-----P-V-----I-----I-----P-----V-----	328
H2B.CL.88.UC1	-----E-----E-----Q-I.PGPLPA-----D-R-----T-----E-----A-V-----I-----I-----P-----V-----	328
H2B.GH.86.D205	-----E-----E-----Q-S.PGMPA-----D-R-----T-----E-----A-V-----I-----I-----P-----V-----	328
H2B.JP01.KR020	-----E-----E-----D-Q-S.PGMPA-----D-R-----T-----E-----P-----I-----I-----P-----V-----	328
H2G.CL.x.ABT96	-----L-----E-----E-----X-----QGP-PA-X-T-----TIE-TH-----X-X-----X-----X-----P-----V-----	333
H2U.FR.96.12034	-----E-----E-----D-T-----N.QGP-PA-----R-----T-----E-----P-----I-----I-----P-----V-----	333
MAC.US.x.251_BK28	-----S-----L-----L-----Q-----T-----E-----P-----V-----	332
MAC.US.x.MM142	-----S-----L-----L-----Q-----T-----E-----P-----V-----	332
MAC.US.x.SMM142B	-----S-----L-----L-----Q-----T-----E-----P-----V-----	332
MAC.US.x.92050	-----S-----L-----L-----Q-----T-----E-----P-----V-----	332
MNE.US.x.MNE027	-----I-----E-----E-----E-----Q-----T-----T-----K-----S-----IR-----P-----V-----	332
MNE.US.82.MNE_8	-----MNE-----E-----E-----E-----Q-----T-----T-----S-----IR-----P-----V-----	332
SMM.SL.92.SL92B	-----E-----E-----E-----PRG-Q-QPA-G-----TPS-E-A-V-D-----P-----V-----	333
SMM.US.02.YNPRC_FWS	-----E-----E-----E-----PGTIPA-----D-R-----T-----K-----I-----M-----IN-----N-----G-----	307
SMM.US.95.TNPRC_G932	-----E-----E-----E-----PGTIPA-----D-R-----T-----E-----I-----M-----IN-----N-----G-----	308
SMM.US.02.YNPRC_FAL	-----E-----E-----E-----PGPLPA-----R-----T-----E-----I-----M-----IN-----N-----G-----	307
SMM.US.85.TNPRC_F102	-----E-----E-----E-----QGPLPA-----R-----T-----E-----V-----I-----M-----IN-----N-----G-----	280
SMM.US.81.NIRC_CFU233	-----E-----E-----E-----PGTIPA-----R-----T-----E-----V-----I-----I-----I-----I-----	303
SMM.US.03.TNPRC_D215	-----E-----E-----E-----PGPLPA-----T-----T-----E-----V-----I-----I-----I-----I-----	293
SMM.US.95.TNPRC_D175	-----E-----E-----E-----PMGTIPA-----D-R-----T-----E-----E-----V-----I-----I-----I-----I-----	268
SMM.US.93.TNPRC_M920	-----E-----E-----E-----PGTIPA-----R-----T-----E-----I-----I-----I-----I-----X-KG-----	282
SMM.US.97.TNPRC_G080	-----E-----E-----E-----V-E-----PGPLPA-----R-----T-----E-----I-----I-----S-----I-----	280
SMM.US.90.TNPRC_F100	-----E-----E-----E-----V-E-----PGPLPA-----R-----T-----E-----I-----I-----I-----I-----	293
SMM.US.89.TNPRC_G930	-----E-----E-----E-----V-E-----PGTIPA-----D-R-----T-----E-----I-----I-----I-----I-----	292
SMM.US.03.TNPRC_M934	-----E-----E-----E-----V-E-----PGPLPA-----R-----T-----E-----I-----I-----I-----I-----	295
SMM.US.95.TNPRC_D171	-----X-----E-----E-----E-----PGTIPA-----D-R-----T-----E-----V-----I-----I-----I-----I-----	295
SMM.US.95.TNPRC_D176	-----X-----X-----E-----E-----PGPLPA-----R-----T-----E-----I-----I-----I-----I-----	296
SMM.US.93.TNPRC_B042	-----E-----E-----E-----E-----PGTIPA-----D-R-----T-----E-----V-----I-----R-----S-----	295
SMM.US.95.TNPRC_M939	-----E-----E-----E-----E-----PGTIPA-----D-R-----T-----E-----V-----I-----S-----	295
SMM.US.03.TNPRC_M924	-----E-----E-----E-----E-----PGPLPA-----R-----T-----E-----V-----I-----I-----I-----I-----	295
SMM.US.86.NIRC_CFU226	-----E-----E-----E-----E-----PGTIPA-----D-R-----T-----E-----V-----I-----I-----I-----I-----	303
SMM.US.86.NIRC_CFU212	-----E-----E-----E-----E-----PGTIPA-----D-R-----T-----E-----V-----I-----I-----I-----I-----	303
SMM.US.77.CNPRC_CFU212	-----E-----E-----E-----E-----PGTIPA-----D-R-----T-----E-----V-----I-----I-----I-----I-----	303
SMM.US.95.TNPRC_B045	-----E-----E-----E-----E-----PGPLPA-----R-----T-----E-----V-----I-----I-----I-----I-----	303
SMM.US.95.TNPRC_M942	-----E-----E-----E-----E-----MV-E-----PGPLPA-----R-----T-----E-----M-----I-----I-----I-----I-----	303
SMM.US.02.YNPRC_FUP	-----E-----E-----E-----E-----V-E-----PGPLPA-----R-----T-----E-----I-----I-----I-----I-----G-----	307
SMM.US.02.YNPRC_FKL	-----E-----E-----E-----E-----E-----PGPLPA-----R-----T-----E-----I-----IN-R-----I-----I-----G-----	307
SMM.US.02.YNPRC_FYN	E-----E-----E-----E-----KE-----PGPLPA-----R-----T-----EK-----I-----I-----I-----I-----C-----	307
SMM.US.93.TNPRC_D178	-----E-----E-----E-----E-----E-----PGPLPA-----R-----T-----V-----I-----I-----I-----I-----G-----	307
SMM.US.02.YNPRC_FCP	-----E-----E-----E-----E-----E-----PGPLPA-----R-----T-----E-----I-----I-----I-----I-----G-----	307
SMM.US.02.YNPRC_FVN	-----E-----E-----E-----E-----E-----PGTIPA-----D-R-----T-----E-----I-----I-----I-----I-----G-----	307
SMM.US.x.F236_H4	-----E-----E-----E-----E-----E-----PGPLPA-----R-----T-----E-----I-----I-----I-----I-----P-----	333
SMM.US.x.H9	-----X-----E-----E-----E-----X-----PGTIPA-----X-----TX-----X-----X-----S-----P-X-----	333
SMM.US.x.PB114_15	-----E-----E-----E-----E-----E-----PGTIPP-----R-----T-----E-----I-----I-----I-----I-----S-----P-----	333
SMM.US.x.PBJA	-----E-----E-----E-----E-----E-----PGTIPP-----R-----T-----E-----I-----I-----I-----I-----S-----P-----	333
SMM.US.x.PBJ_143	-----X-----E-----E-----E-----X-----PGTIPA-----X-----TX-----X-----X-----S-----P-X-----	333
SMM.US.x.PBJ_6P6	-----E-----E-----E-----E-----E-----PGTIPP-----R-----T-----E-----I-----I-----I-----I-----S-----P-S-----	333
SMM.US.x.PGM53	-----E-----E-----E-----E-----E-----PGTIPA-----D-R-----T-----E-----I-----I-----I-----I-----S-----P-----	333
SMM.US.x.SME543	-----E-----E-----E-----E-----E-----PGPLPA-----R-----T-----E-----I-----I-----I-----I-----S-----P-----	333
STM.US.x.STM	-----L-----E-----E-----KE-----M-----PPGPLPA-----PE-----V-----T-----A-PS-----R-----	334

	p27_p2	p2_p8	p8_p1	p1_p6	PTAP motif	PSAP in HIV-2 B. U	
MAC.US.x.239							485
H2A.GM.x.MCN13							496
H2A.CI.88.UC2							496
H2A.DE.x.BEN							496
H2A.DE.x.PE12	I						496
H2A.GH.x.GH1							497
H2A.GM.x.ISY	I						495
H2A.GM.87.D194		S					496
H2A.GW.x.AL1							496
H2A.GW.x.M1S							496
H2A.GW.86.FG							494
H2A.GW.87.CAM2CG							496
H2A.SN.85.ROD							496
H2AB.CI.90.7312A							496
H2B.CI.x.EHO							495
H2B.CI.88.UC1							496
H2B.GH.86.D205							496
H2B.JP01.KR020							496
H2G.CI.x.ABT96							484
H2U.FR.96.12034							497
MAC.US.x.251_BK28							481
MAC.US.x.MM142							481
MAC.US.x.SMM142B							481
MAC.US.x.92050							485
MNE.US.x.MNE027							481
MNE.US.82.MNE_8							481
SMM.SL.92.SL92B							482
SMM.US.02.YNPRC_FWS							307
SMM.US.95.TNPRC_G932							308
SMM.US.02.YNPRC_FAL							307
SMM.US.85.TNPRC_F102							280
SMM.US.81.NIRC_CFU233							303
SMM.US.03.TNPRC_D215							293
SMM.US.95.TNPRC_D175							268
SMM.US.93.TNPRC_M920							282
SMM.US.97.TNPRC_G080							280
SMM.US.90.TNPRC_F100							293
SMM.US.89.TNPRC_G930							292
SMM.US.03.TNPRC_M934							295
SMM.US.95.TNPRC_D171							295
SMM.US.95.TNPRC_D176							296
SMM.US.93.TNPRC_E042							295
SMM.US.95.TNPRC_M939							295
SMM.US.03.TNPRC_M924							295
SMM.US.86.NIRC_CFU226							303
SMM.US.86.NIRC_CFU212							303
SMM.US.77.CNPRC_CFU212							303
SMM.US.95.TNPRC_B045							303
SMM.US.95.TNPRC_M942							303
SMM.US.02.YNPRC_FUP							307
SMM.US.02.YNPRC_FKL							307
SMM.US.02.YNPRC_FYN							307
SMM.US.93.TNPRC_D178							307
SMM.US.02.YNPRC_FCP							307
SMM.US.02.YNPRC_FVN							307
SMM.US.x.F236_H4	M						482
SMM.US.x.H9	XI	X	X				482
SMM.US.x.PBJ14_15	I						482
SMM.US.x.PBJA	I						482
SMM.US.x.PBJ_143	XI	X	X				482
SMM.US.x.PBJ_6P6	I						482
SMM.US.x.PGM53	M						482
SMM.US.x.SME543	M	I					482
SMM.US.x.STM	M						482

	p6 end Gag end ¹
MAC.US.x.239	ESREKPYK.EVTEDLLHLSLFGGDQ*
H2A.GM.x.MCN13	-QG-A-CR--T-----*K---
H2A.CI.88.UC2	-QG-T-RR--A-----K---
H2A.DE.x.BEN	-Q--T-HR--E-----K---
H2A.DE.x.PE12	-QG-T-H--A-----K---
H2A.GH.x.GH1	-QGKA-HR--A-----K---
H2A.GM.x.ISY	-QG-T-HR--T-----N---
H2A.GM.87.D194	-QG-T-HR.GA-----K---
H2A.GW.x.AL1	-QG-T-H-----K---
H2A.GW.x.MDS	-QGKT-C--T-----T---
H2A.GW.86.FG	-QA-T-CR--T-----K---
H2A.GW.87.CAM2CG	-QG-A-CR--T-----K---
H2A.SN.85.ROD	-QG-T--REPP-----K---
H2AB.CI.90.7312A	-KQGR-----E---
H2B.CI.x.EHO	-N--R-----E---
H2B.CI.88.UC1	---R-----E---
H2B.GH.86.D205	---R-----E---
H2B.JP01.KR020	GG-----E---
H2G.CI.x.ABT96	-Q-N-----XX---S---D---
H2U.FR.96.12034	-N--R-----N---
MAC.US.x.251_BK28	-----
MAC.US.x.MM142	---G-----
MAC.US.x.SMM142B	---G-----
MAC.US.x.92050	-----
MNE.US.x.MNE027	--KR-----A-----E---
MNE.US.82.MNE_8	--KR-----E---
SMM.SL.92.SL92B	TR-SR-----E---
SMM.US.02.YNPRC_FWS
SMM.US.95.TNPRC_G932
SMM.US.02.YNPRC_FAL
SMM.US.85.TNPRC_F102
SMM.US.81.NIRC_CFU233
SMM.US.03.TNPRC_D215
SMM.US.95.TNPRC_D175
SMM.US.93.TNPRC_M920
SMM.US.97.TNPRC_G080
SMM.US.90.TNPRC_F100
SMM.US.89.TNPRC_G930
SMM.US.03.TNPRC_M934
SMM.US.95.TNPRC_D171
SMM.US.95.TNPRC_D176
SMM.US.93.TNPRC_E042
SMM.US.95.TNPRC_M939
SMM.US.03.TNPRC_M924
SMM.US.86.NIRC_CFU226
SMM.US.86.NIRC_CFU212
SMM.US.77.CNPRC_CFU212
SMM.US.95.TNPRC_E045
SMM.US.95.TNPRC_M942
SMM.US.02.YNPRC_FUP
SMM.US.02.YNPRC_FKL
SMM.US.02.YNPRC_FYN
SMM.US.93.TNPRC_D178
SMM.US.02.YNPRC_FCP
SMM.US.02.YNPRC_FVN
SMM.US.x.F236_H4	-N--R-----E---
SMM.US.x.H9	-N--R-----E---
SMM.US.x.PBJ14_15	-N--R-----E---
SMM.US.x.PBJA	-N--R-----E---
SMM.US.x.PBJ_143	-N--R-----E---
SMM.US.x.PBJ_6P6	-N--R-----E---
SMM.US.x.PGM53	-N--R-----E---
SMM.US.x.SME543	-N--R-----E---
STM.US.x.STM	--KT-----E---

510
521
522
522
522
522
521
522
522
522
520
522
523
522
521
522
522
510
523
507
507
507
511
507
507
508
307
308
307
280
293
293
268
282
280
280
292
295
295
296
295
295
295
303
303
303
303
303
307
307
307
307
307
307
508
508
508
508
508
508
508
508
508
508

Accession	Pol p15 start (-1 from Gag)	p15 protease	
MAC.US.x.239	FRPWSMGKEAQPFGSSASGADANCS.PRGPSCGSAKELHAVGQAAE.....RKAERKQREALQG.GDRGFAAPQFSLWRRPVVTAHIEGQVPEVLLDGTGADDSIVTGIELGPHYTPKIVGGIGGFINTKEYKNVEIYVGLKRIKGTIMTGDTPINIFGRN		155
H2A.GM.x.MCN13	--DGPP-----L-R-P-PA--NT-ST--SRS-S-PTG-Y-ARKTKRTERETIORSRGLT-P-AG-GTM-R-D--L-----LK-----YV-----A--SN-S-----K--KVRA-----		168
H2A.CL88.UC2	--DGLT-----L-R-P-S---T-ST--SRSGS-PVR-IF-A-EK--GAEGETIIOGGDRLT-P-AG-DTS-R--L-----K-----Y-D-----A--DN-V-----K-N--VRA-----		168
H2A.DE.x.BEN	--VGPT-----L-RDP-P---T-ST--SGRS-S-TVG-IY-AREK--GAEGETIORGGGLA-P-AE-DTS-R--L-----K-----Y-D-----A--DN-----K-N--VRA-----		168
H2A.DE.x.PE12	-D-P-----S-L-RDP-PA--T-ST--SR--SRP-R-VL-AREE--RAENETIIOGGDRLT-P-TR-DTT-R--L-----K-----YV-----A--SN-S-----K-N-KV-A-----		168
H2A.GH.x.GH1	--DG-----L-R-P-S---T-ST--SRS-S--IGKIY-A-ER--GAEGETIORGGRLT-P-AGKSTS-R--L-----K-----Y-V-----A-Q-DN-V-----I-I-K-N--VRA-----		168
H2A.GM.x.ISY	-A-T-----L-R-PKFA--NT-ST--N-S-S-PTG-Y-AREKT-RAETKTIORSDRGLA-S-AR-DTT-R-D--L-----K-----Y-D-----A--SN-S-----D-R-N-KVRA-----		168
H2A.GM.87.D194	--DGPT--A--L-R-P-S---T-ST--NRS-S-PVG-IY-AREK--RAEGETIIOGGDGLT-P-AG-D-P-R--L-T-----K-----F-D-----A--DN-----K-N--VRA-----		168
H2A.GW.x.AL1	-A-P-----S-L-RNP-SA--INT-ST--SRA-S-P-GAVY-A-EK--KRAEREAIORGGGLT-P-AG-DTT-R--L-----K-----Y-----A--SN-----ED-K-N--V-A-----		168
H2A.GW.x.MDS	--D-PL-----L-R-PGA--NT-ST--SRS-S-PTG-IY-AREK--GAETIIOGRDGLA-P-AGKDTM--DN-----N-----N-----A--SN-S-----V-A-----		168
H2A.GW.86.FG	-D-PL--G--L-R-P-PA--NT-ST--I-S-S-PTG-IY-ARKK--KGAERETVQSGDRGLT-F-AG-DTM--D--L-----K-----R-N-----A--SN-S-----V-A-----		168
H2A.GW.87.CAM2C2	-D-PL-----L-R-P-ST--NT-ST--I-S-S-TG-IY-AREK--GAETIIOGRDGLT-P-TR-GPM--DN-L-----K-----Y-----A--SN-S-----V-A-----		168
H2A.SN.85.ROD	--TGPL-----L-R-P-SA--T-ST--S-S-S-TG-IY-AREKT-RAETIIOGRDGLT-P-AGDITI--ATN-----L-----K-----Y-----A--NN-S-----N-KVRA-----		169
H2AB.CL90.7312A	-A-TL-----S-L--P---DS-I-APDEHSRQDTSGSDTICAPCRSSGDAEKLHATREA--EAG-----L-----Q-Y-----A--SN-EAG-----I--V-KVRS-----		169
H2B.CL.x.EHO	--VRPL-----S--R-PGTP-DS-I-APDEPSIRHDTSGDCITPCRSSRQDAKELHATREE--GE--T-----K-T-S-----A--SN-----N--V-VRA-V-----		168
H2B.CL88.UC1	--VRTL-----S-L--DP---S-TI-TPDGPSRGHDTSGGDTICAPCRSSGDAEKLHEDGET--EP-T-----K-C-----A--SN-----D-V-VRA-----		169
H2B.GH.86.D205	--VRTL-----S-L--DP---S-TI-TPDEPSRGHDTSGGDTICAPCRSSGDAEKLHADGETT--EP-T-----K-C-S-----V-----A--SN-----D-V-VRA-----		169
H2B.JP01.KR020	-T-PL-----S-L-CSP-TP-DT-I-TTNEPSRHDTSGCN-ICAPCRSSGDEVLGHAAE--G-G-T-----K-Y-S-----A--SD-----V-K-RA-----		169
H2C.LR.x.2238			0
H2D.LR.90.FO784PA			3
H2D.LR.90.FORTC2			0
H2O1.AE.US.91.PA			0
H2G.CL.x.AB196	-V-TL-----S-L--DP-P--S-SIST.TD--SRPTE-----X-EE-K-----GAEKQT-----XX--L-----Y-S-----A--SN-S-----K-V--QA-V-----		151
H2U.FR.96.12034	-A-PL--K-S-L--P-PA-T-P-SP.SSRT-S--DPDPSGSPSSG--SAKELHATGQETKGE--TI--G-----K-I-Y--K--N-----A--K-E-Y--T-E-E--KFH-N-K--T-R-----		165
MAC.US.x.251.BK28			151
MAC.US.x.SMM142B			151
MAC.US.x.MM142			151
MAC.US.x.92050			155
MNE.US.x.MNE027			151
MNE.US.82.MNE.8			151
SMM.SL.92.SL92B			148
SMM.US.95.TNPRC.G932			0
SMM.US.02.YNPRC.FBL			0
SMM.US.86.NIRC.CFU233			0
SMM.US.03.TNPRC.D215			0
SMM.US.95.TNPRC.D175			0
SMM.US.94.TNPRC.M927			0
SMM.US.94.TNPRC.G080			0
SMM.US.79.NIRC.CFU212			0
SMM.US.79.NIRC.6007.G932			0
SMM.US.95.TNPRC.F102			0
SMM.US.95.TNPRC.D177			0
SMM.US.91.TNPRC.G931			0
SMM.US.81.TNPRC.G930			0
SMM.US.03.TNPRC.G932			0
SMM.US.85.TNPRC.F104			0
SMM.US.84.TNPRC.F104			0
SMM.US.79.NIRC.CFU233			0
SMM.US.86.NIRC.CFU232			0
SMM.US.86.NIRC.CFU226			0
SMM.US.82.NIRC.6007.G932			0
SMM.US.86.NIRC.6002.G931			0
SMM.US.86.NIRC.6001.G930			0
SMM.US.79.NIRC.6001.G930			0
SMM.US.02.YNPRC.FIP			0
SMM.US.x.F236.H4			151
SMM.US.x.H9			151
SMM.US.x.PBJ14_15			151
SMM.US.x.PBJA			151
SMM.US.x.PBJ_143			151
SMM.US.x.PBJ_6P6			151
SMM.US.x.PGM53			151
SMM.US.x.SME543			151
STM.US.x.STM			151

	protease_p51 RT start		D catalytic site	
MAC.US.x.239	LLTALGMSLNFP	IAKVEPVKVALKPGKDGPKLQKWPLSKEKIVALEICEK	MEKDGQLEEAPPTNYNTPTFAIKKDKNKWRMLIDFRELNRVTQDFTEVQLGIPHPAGLAKRKRTVLDIGDAYFSIPLDEEFQRYTAFTLPSVNNNAEPGKRYIYKVLPGQWKGSPA	324
H2A.GM.x.MCN13	I-----L	V-I-I-IM-----R-T-E-K-----RE-----K-I-----K-----V-----H-D-----I-V-----		337
H2A.CL.88.UC2	I-----L	V-RI-I-IT-----R-T-VE-K-----E-----K-I-----K-----V-----Y-D-----P-----A-----M-----		337
H2A.DE.x.BEN	I-----L	V-I-I-T-----R-T-E-K-----E-----K-I-----K-----SI-----V-----H-D-----I-----		337
H2A.DE.x.PE12	I-----L	V-D-I-I-----VR-T-E-K-----RE-----K-E-I-----K-----I-----H-D-----T-----		337
H2A.GH.x.GH1	I-----L	V-I-I-T-----R-R-T-E-----E-----K-I-----K-----V-----H-D-----		337
H2A.GM.x.ISY	I-----L	V-I-I-T-----QR--TR-E-----RE-----K-----K-----V-----Y-D-----		337
H2A.GM.87.D194	I-AT-----L	V-LD-I-T-----R-T-E-K-----RE-----K-I-----K-----V-----H-D-----E-V-----		337
H2A.GW.x.A1J	I-----L	V-I-IE-R-----R-T-E-K-----T-RE-----R-----K-I-----K-----V-----H-S-----		337
H2A.GW.x.M1S	I-----L	V-I-I-T-----R-T-E-K-----E-----K-I-----K-----V-----H-D-----		337
H2A.GW.86.FG	I-----L	V-I-I-IM-----R-T-E-K-----E-----K-I-----K-----V-----H-D-----		337
H2A.GW.87.CAM2C2G	I-----L	V-I-I-IM-----R-R-T-E-K-----E-----K-I-----K-----V-----H-D-----		337
H2A.SN.85.ROD	I-----L	V-I-I-IM-----R-T-E-K-----E-----K-I-----K-----V-----H-D-----P-----		338
H2AB.CL.90.7312A	I-NT--T--V	I-Q-----IR-----L-----E-----S-----K-E-----EKR--I-V--V--PD-----I-----		338
H2B.CL.x.EHO	I-NS--T--V	RI--Q--E-----IR-----L-K-----E-----S-----K-E-----SK--V--V--PD-----A-----L-----		337
H2B.CL.88.UC1	I-NT--T--V	I-K-----IR-----L-K-----E-----R-----K-E-----EKR--V--V--PN-----I-----		338
H2B.GH.86.D205	I-NT--T--V	I-E-----IR--R--L-K-----E-----R-----K-----NWVF--TRQV--EKR--I-V--PN-----		337
H2B.JP.01.KR020	I-NT--T--V	I-I-K-----IR-----L-K-----X-----E-----R-----K-----E-----PN-----		339
H2C.LR.x.2238				0
H2D.LR.90.FO784PA				0
H2D.LR.90.FORTC2				172
H201_AE.US.91.PA				117
H2G.CL.x.ABT96	I-VK-----L	R-M-IR-----Q-D--QE-----T-----EKR-----V-----VD-----I-----Q-----		320
H2U.FR.96.12034				334
				T
MAC.US.x.251.BK28				320
MAC.US.x.SMM142B				320
MAC.US.x.MM142				320
MAC.US.x.92050				324
MNE.US.x.MNE027				320
MNE.US.82.MNE_8				320
SMM.SL.92.SL92B	I-A--V--Y	V--YT--K-E-M--R-----Q-T-----E--R-----N-D-----KM-E-----KEME-----V--P-----I--Q-----		317
SMM.US.95.TNPRC_G932				0
SMM.US.02.YNPRC_FBL				0
SMM.US.86.NIRC_CFU233				0
SMM.US.03.TNPRC_D215				0
SMM.US.95.TNPRC_D175				0
SMM.US.94.TNPRC_M927				0
SMM.US.94.TNPRC_G080				0
SMM.US.79.NIRC_CFU212				0
SMM.US.79.NIRC_6007_G932				0
SMM.US.95.TNPRC_F102				0
SMM.US.95.TNPRC_D177				0
SMM.US.91.TNPRC_G931				0
SMM.US.81.TNPRC_G930				0
SMM.US.03.TNPRC_G932				0
SMM.US.85.TNPRC_F104				0
SMM.US.84.TNPRC_F104				0
SMM.US.79.NIRC_CFU233				0
SMM.US.86.NIRC_CFU232				0
SMM.US.86.NIRC_CFU226				0
SMM.US.82.NIRC_6007_G932				0
SMM.US.86.NIRC_6002_G931				0
SMM.US.86.NIRC_6001_G930				0
SMM.US.79.NIRC_6001_G930				0
SMM.US.02.YNPRC_FIP				0
SMM.US.x.F236_H4				0
SMM.US.x.H9				320
SMM.US.x.PBJ14_15				320
SMM.US.x.PBJA				320
SMM.US.x.PBJ_143				320
SMM.US.x.PBJ_6P6				320
SMM.US.x.PGM53				320
SMM.US.x.SME543				320
STM.US.x.STM				320

Accession	Sequence	Position
MAC.US.x.239	IFQYTMRHVLEPFRKANPDVTLVQYMDLILIASDRDLEHDRVVLQSKELLNSIGFSTPEEKFKQDPPFQWNGYELWPTKWLQKIELPQRETWTVNDIQKLVGLNWAQIYPGIKTKHLRGLRGMKTLTEEVQWTEMAEAEYEENKIILSQEQEGCYOEGKPLEAT	494
H2A.GM.x.MCN130.....L.....III.....K.....L.....NL.....D.....YH.....Q.....KDV.....R.....K.....L.....L.....R.....D.....H.....E.....E.....E.....	507
H2A.CL.88.UC2OI.....I.....I.....G.....K.....L.....GL.....D.....C.....Q.....K.....I.....S.....L.....L.....R.....D.....H.....E.....E.....E.....	507
H2A.DE.x.BENH.....0.....I.....I.....G.....K.....L.....GL.....D.....YK.....C.....Q.....KDV.....L.....KDI.....S.....L.....L.....R.....D.....H.....E.....E.....E.....	507
H2A.DE.x.PE12H.....0.....I.....I.....G.....K.....L.....GL.....D.....YK.....C.....Q.....KDV.....L.....KDI.....S.....L.....L.....R.....D.....H.....E.....E.....E.....	507
H2A.GH.x.GH1H.....0.....I.....I.....G.....K.....L.....GL.....D.....YK.....C.....Q.....KDV.....L.....KDI.....S.....L.....L.....R.....D.....H.....E.....E.....E.....	507
H2A.GM.x.ISYH.....0.....I.....I.....G.....K.....L.....GL.....D.....YK.....C.....Q.....KDV.....L.....KDI.....S.....L.....L.....R.....D.....H.....E.....E.....E.....	507
H2A.GM.87.D194FM.....OI.....I.....I.....G.....K.....L.....GL.....D.....YK.....C.....Q.....KDV.....L.....KDI.....S.....L.....L.....R.....D.....H.....E.....E.....E.....	507
H2A.GW.x.ALH.....OI.....I.....I.....G.....K.....L.....GL.....D.....YK.....C.....Q.....KDV.....L.....KDI.....S.....L.....L.....R.....D.....H.....E.....E.....E.....	507
H2A.GW.x.MDSH.....OI.....I.....I.....G.....K.....L.....GL.....D.....YK.....C.....Q.....KDV.....L.....KDI.....S.....L.....L.....R.....D.....H.....E.....E.....E.....	507
H2A.GW.86.FGH.....OI.....I.....I.....G.....K.....L.....GL.....D.....YK.....C.....Q.....KDV.....L.....KDI.....S.....L.....L.....R.....D.....H.....E.....E.....E.....	507
H2A.GW.87.CAM2C8H.....OI.....I.....I.....G.....K.....L.....GL.....D.....YK.....C.....Q.....KDV.....L.....KDI.....S.....L.....L.....R.....D.....H.....E.....E.....E.....	507
H2A.SN.85.RODH.....OI.....I.....I.....G.....K.....L.....GL.....D.....YK.....C.....Q.....KDV.....L.....KDI.....S.....L.....L.....R.....D.....H.....E.....E.....E.....	507
H2AB.CL.90.7312AS.....K.....D.....R.....N.....II.....V.....S.....K.....S.....L.....DM.....K.....QI.....E.....V.....LF.....R.....I.....K.....L.....M.....Q.....G.....S.....K.....V.....	508
H2B.CL.x.EHOAK.....D.....N.....II.....V.....S.....S.....L.....NM.....K.....K.....Q.....EK.....V.....LF.....R.....I.....K.....L.....F.....Q.....E.....S.....K.....V.....	507
H2B.CL.88.UC1S.....K.....D.....S.....III.....S.....S.....L.....DM.....K.....K.....Q.....EK.....V.....LF.....R.....I.....K.....L.....L.....Q.....E.....S.....K.....V.....	508
H2B.GH.86.D205C.....S.....K.....D.....S.....III.....S.....S.....L.....DM.....K.....K.....Q.....EK.....V.....A.....LF.....R.....I.....K.....L.....L.....Q.....E.....S.....K.....RV.....	507
H2B.JP.01.KR020SVGR.....D.....D.....I.....V.....S.....L.....DM.....K.....K.....Q.....EK.....R.....A.....LF.....R.....I.....K.....L.....L.....Q.....E.....S.....K.....T.....	509
H2C.LR.x.2238NI.....S.....I.....K.....L.....DL.....L.....YH.....D.....K.....L.....L.....R.....D.....H.....E.....E.....E.....	0
H2D.LR.90.F0784PANI.....S.....I.....K.....L.....DL.....L.....YH.....D.....K.....L.....L.....R.....D.....H.....E.....E.....E.....	342
H2D.LR.90.FORTC2NI.....S.....I.....K.....L.....DL.....L.....YH.....D.....K.....L.....L.....R.....D.....H.....E.....E.....E.....	0
H201.AE.US.91.PA	V.....E.....0.....S.....L.....I.....L.....G.....N.....SMT.....M.....V.....L.....DM.....NL.....D.....RE.....LK.....L.....Y.....K.....EK.....R.....I.....V.....N.....K.....LS.....LA.....G.....R.....R.....EEN.....	287
H2G.CL.x.AB196H.....K.....D.....E.....XIXI.....V.....Q.....L.....KX.....D.....Y.....X.....K.....N.....XEK.....X.....L.....N.....XK.....X.....L.....LQ.....X.....X.....A.....K.....XE.....X.....	490
H2U.FR.96.12034NL.....E.....AVII.....N.....Q.....L.....S.....L.....KT.....D.....HK.....K.....E.....DV.....QL.....V.....K.....K.....FA.....A.....R.....EEN.....	504
MAC.US.x.251.BK28L.....S.....S.....L.....R.....D.....H.....E.....E.....E.....	490
MAC.US.x.SMM142BL.....S.....S.....L.....R.....D.....H.....E.....E.....E.....	490
MAC.US.x.MM142L.....S.....S.....L.....R.....D.....H.....E.....E.....E.....	490
MAC.US.x.92050L.....S.....S.....L.....R.....D.....H.....E.....E.....E.....	494
MNE.US.x.MNE027H.....N.....V.....L.....K.....L.....R.....D.....H.....E.....E.....E.....	490
MNE.US.82.MNE_8H.....N.....V.....L.....K.....L.....R.....D.....H.....E.....E.....E.....	490
SMM.SL.92.SL92BA.....Q.....L.....L.....G.....N.....GLT.....KM.....T.....LRDM.....NL.....D.....N.....L.....L.....Y.....K.....EK.....R.....N.....KM.....L.....LA.....R.....N.....R.....R.....DE.....	487
SMM.US.95.TNPRC_G932L.....S.....S.....L.....R.....D.....H.....E.....E.....E.....	0
SMM.US.02.YNPRC_FBLL.....S.....S.....L.....R.....D.....H.....E.....E.....E.....	0
SMM.US.86.NIRC_CFU233L.....S.....S.....L.....R.....D.....H.....E.....E.....E.....	0
SMM.US.03.TNPRC_D215L.....S.....S.....L.....R.....D.....H.....E.....E.....E.....	0
SMM.US.95.TNPRC_D175L.....S.....S.....L.....R.....D.....H.....E.....E.....E.....	0
SMM.US.94.TNPRC_M927L.....S.....S.....L.....R.....D.....H.....E.....E.....E.....	0
SMM.US.94.TNPRC_G080L.....S.....S.....L.....R.....D.....H.....E.....E.....E.....	0
SMM.US.79.NIRC_CFU212L.....S.....S.....L.....R.....D.....H.....E.....E.....E.....	0
SMM.US.79.NIRC_6007_G932L.....S.....S.....L.....R.....D.....H.....E.....E.....E.....	0
SMM.US.95.TNPRC_F102L.....S.....S.....L.....R.....D.....H.....E.....E.....E.....	0
SMM.US.95.TNPRC_D177L.....S.....S.....L.....R.....D.....H.....E.....E.....E.....	0
SMM.US.91.TNPRC_G931L.....S.....S.....L.....R.....D.....H.....E.....E.....E.....	0
SMM.US.81.TNPRC_G930L.....S.....S.....L.....R.....D.....H.....E.....E.....E.....	0
SMM.US.03.TNPRC_G932L.....S.....S.....L.....R.....D.....H.....E.....E.....E.....	0
SMM.US.85.TNPRC_F104L.....S.....S.....L.....R.....D.....H.....E.....E.....E.....	0
SMM.US.84.TNPRC_F104L.....S.....S.....L.....R.....D.....H.....E.....E.....E.....	0
SMM.US.79.NIRC_CFU233L.....S.....S.....L.....R.....D.....H.....E.....E.....E.....	0
SMM.US.86.NIRC_CFU232L.....S.....S.....L.....R.....D.....H.....E.....E.....E.....	0
SMM.US.86.NIRC_CFU226L.....S.....S.....L.....R.....D.....H.....E.....E.....E.....	0
SMM.US.82.NIRC_6007_G932L.....S.....S.....L.....R.....D.....H.....E.....E.....E.....	0
SMM.US.86.NIRC_6002_G931L.....S.....S.....L.....R.....D.....H.....E.....E.....E.....	0
SMM.US.86.NIRC_6001_G930L.....S.....S.....L.....R.....D.....H.....E.....E.....E.....	0
SMM.US.79.NIRC_6001_G930L.....S.....S.....L.....R.....D.....H.....E.....E.....E.....	0
SMM.US.02.YNPRC_FIPL.....S.....S.....L.....R.....D.....H.....E.....E.....E.....	0
SMM.US.x.F236_H4X.....H.....X.....N.....I.....L.....G.....N.....S.....L.....R.....D.....H.....E.....E.....E.....	490
SMM.US.x.H9H.....N.....I.....L.....S.....L.....R.....D.....H.....E.....E.....E.....	490
SMM.US.x.PB14_15H.....N.....I.....L.....S.....L.....R.....D.....H.....E.....E.....E.....	490
SMM.US.x.PBJAX.....H.....X.....N.....I.....L.....S.....L.....R.....D.....H.....E.....E.....E.....	490
SMM.US.x.PBJ_143H.....N.....I.....L.....S.....L.....R.....D.....H.....E.....E.....E.....	490
SMM.US.x.PBJ_6P6H.....N.....I.....L.....S.....L.....R.....D.....H.....E.....E.....E.....	490
SMM.US.x.PGM53H.....N.....G.....I.....L.....G.....N.....S.....L.....R.....D.....H.....E.....E.....E.....	490
SMM.US.x.SME543N.....I.....I.....L.....G.....N.....S.....L.....R.....D.....H.....E.....E.....E.....	490
STM.US.x.STMNI.....R.....I.....L.....NL.....D.....V.....K.....A.....A.....Q.....R.....D.....	490



p51 RT end_p15 RNase H start

MAC.US.x.239	VIKSQDNQWSYKIHQEDKILKVG.KFKAIKNTHTNGVRLLAHVVIQKIGKEAIVIGWQPKFHLPVKEKDVWEQWTDYVQVWTWIPEDWFISTPPLVRLVFNLVKDPTEGEEETYYTDGSCNKQSKQKAGYITDRGKDKVKVLEQTTNQAELEAFMLALTDGSPKANIIVD	663
H2A.GM.x.MCN13	Q-D---T-V-GE-----Y-----Q-V---L-I---RI---RET---D-----D-V---A---PRT-F-----R-----V-R-RM-----A-----A-----V-----	676
H2A.CL.88.UC2	IQ---GH---T---E-----Y-----Q-V---L-I---RI---RET---DN-----V-----T-G-P-T-F-----R-----V-R-R-----A-A-----V-----	676
H2A.DE.x.BEN	Q-D---T---E-----Y-----Q-V---L-I---RI---RET---DN-----D-V---A---P-A-F-----R-----V-R-RI-----A-----A-----V-----	676
H2A.DE.x.PE12	IQ---N---T---E-----Y-----Q-V---L-I---RI---RET---DN-----D-V---T-G-P-A-F-----R-----R-V-R-R-----A-T-----V-----	676
H2A.GH.x.GH1	Q-D---T-V-GE-----Y-----Q-V---L-I---RI---RET---DN-----D-V---A---P-A-F-----R-----R-RI-----A-V-----V-V-----	676
H2A.GM.x.ISY	Q-D---A---GERV-----Y-----Q-V---L-I---R---R-T---DN-----V-----T-G-P-T-F-----R-----V-R-R-R-----S-----A-A-----V-----	676
H2A.GM.87.D194	Q-D---T---GE-T---E-Y-M-----Q-V---L-I---RI-R---RET---D-----D-V---A---L-A-F-----P-----R-----V-K-R-----A-L-V-----	676
H2A.GW.x.A1J	Q-D---T-V-GE-T---E-Y-V-----Q-V---L-I---RI---REI---D-----D-V---A---L-A-F-----R-----V-K-R-----A-----A-----V-----	676
H2A.GW.x.MDS	Q-D---T-V-GE---G-X-Y-----X---Q-V---L-I---RI---R-T---DN-----D-V---A---GE-VP-A-F-----R-----V-R-R-----A-----A-----V-----	677
H2A.GW.86.FG	Q-D---T---E-----Y-H-----X---Q-V---L-I---RI---REI---DN-----D-V---A---G-P-T-F-----R-----V-R-I-----A-----A-----V-----	675
H2A.GW.87.CAM2CQ	Q-D-E---T---E-----Y-V---I---Q-V---L-I---RI---REI---DN-----D-V---A---G-P-A-F-----R-----V-R-K-----A-----A-----V-----	677
H2A.SN.85.ROD	Q-NLA---T---GN-V---Y-V-----I---V---L-I---M---RET-D-----V-----AY---L-KV-----RT-----V-R-P-----A-L-Q-----QV-----	677
H2AB.CL.90.7312A	Q-NLA---T---G-----Y-V-----V---L---EI-M---RET-D-----V---I-AY---L-V-----A-----V---P-----A-L-Q-----QV-----	676
H2B.CL.x.EHO	Q-NLA---T---GNR-----Y-V-----V---L---EI-V---RET-D-----V---AY---L-KV-----RA-----V-----A-L-Q-----QV-----	677
H2B.CL.88.UC1	Q-NLA---T---GN-V---Y-V-----V---L---EI-V---RET-D-----V---I-AY---L-R-----RT-----V-----A-L-Q-----E-QV-----	676
H2B.GH.86.D205	Q-NLA---T---G-R-R---Y-V-----L-T---EI-M---RET-D-----V---I-AY---L-V-----RT-----V-----A-L-Q-----QV-----	678
H2B.JP.01.KR020	0
H2C.LR.x.2238	0
H2D.LR.90.FO784PA	0
H2D.LR.90.FORTC2	V-D-----Y-----V-----R-S-I-R-----S-----V-----E-L-KT-F-V---HR-----V-DRE-T-A-Q-----L-Q-----	511
H201_AE.US.91.PA	L-N-----G-R-----PQ-----S-V-----	332
H2G.CL.x.ABT96	Q-ML---T-V-GNR-----Y-----V-X---L---KL-F---R-T-----D-V-H-Y---E-L-QT-----T-X-V---X-----X-----A-Q-----V-X-----	659
H2U.FR.96.12034	L-N---T---G-RT-----V-----V-----KI-M---RE-----VLD-V---AY---E-LL-----RA-Q-V---R-P-----A-E-----V-VT-----	672
MAC.US.x.251_BK28	659
MAC.US.x.SMM142B	659
MAC.US.x.MM142	659
MAC.US.x.92050	663
MNE.US.x.MNE027	---N-----	659
MNE.US.82.MNE_8	---N-----	659
SMM.SL.92.SL92B	L-N-----G-R-----I---N-V---SL---KT-F---RE-D-----A-----I-----K-V-I---RM-----V---E-LP---A-----Q-L-L-K-PS-V-VVT-----	656
SMM.US.95.TNPRC_G932	0
SMM.US.02.YNPRC_FBL	0
SMM.US.86.NIRC_CFU233	0
SMM.US.03.TNPRC_D215	0
SMM.US.95.TNPRC_D175	0
SMM.US.94.TNPRC_M927	0
SMM.US.94.TNPRC_G080	0
SMM.US.79.NIRC_CFU212	0
SMM.US.79.NIRC_6007_G932	0
SMM.US.95.TNPRC_F102	0
SMM.US.95.TNPRC_D177	0
SMM.US.91.TNPRC_G931	0
SMM.US.81.TNPRC_G930	0
SMM.US.03.TNPRC_G932	0
SMM.US.85.TNPRC_F104	0
SMM.US.84.TNPRC_F104	0
SMM.US.79.NIRC_CFU233	0
SMM.US.86.NIRC_CFU232	0
SMM.US.86.NIRC_CFU226	0
SMM.US.82.NIRC_6007_G932	0
SMM.US.86.NIRC_6002_G931	0
SMM.US.86.NIRC_6001_G930	0
SMM.US.79.NIRC_6001_G930	0
SMM.US.02.YNPRC_FIP	0
SMM.US.x.F236_H4	V-----V-----S---V---L---E---REI-----D-V-----E-Q-A-F-V---R---R---V---R-A-L-----YL-A-----	659
SMM.US.x.H9	V-----V-----S---V---I---R---XEI-----V-----E-Q-A-F-V-R---R---V---R-A-L-----YL-A-----	659
SMM.US.x.PB114_15	V-----V-----R---REI-----V-----E-Q-A-F-V---R---R---V---R-A-L-----YL-A-----	659
SMM.US.x.PB1A	V-----V-----R---REI-----V-----E-Q-A-F-V---R---R---V---R-A-L-----YL-A-----	659
SMM.US.x.PB1_143	V-----V-----R---XEI-----V-----E-Q-A-F-V-R---R---V---RG-T-L-----YL-A-----	659
SMM.US.x.PB1_6P6	V-----V-----R---REI-----V-----E-Q-A-F-V---R---R---V---R-A-L-----YL-A-----	659
SMM.US.x.PGM53	V-----VE-----R---REI-----V-----E-Q-A-F-V---R---R---V---R-A-P-----YL-A-----	659
SMM.US.x.SME543	V-----V-----V---L---E---REI-----D-V-----E-Q-A-F-V---R---R---V---R-A-L-----YL-A-----	659
STM.US.x.STM	V-D-----I-----RE-----A-----V-----E-L-T-F-V---R---N---A-----A-----A-----V-----	659

p15 RNase H end_p31 Integrase start

MAC.US.x.239	SQYVMGIITGCPTESESRVNIQIIEEMIKKSEIYVAWVPAHKGGIGNQEDHLVSGIRQVLFLEKIEPAQEEHDKYNSVKELVFKFGLPRIVARQIVDTCDKCHQKGEAIHQANSDLGTWQMDCTHLEGGKIIIVAVHVASGFIAEAVIPQETGRQTAFLFLKLAGR	832
H2A.GM.x.MCN13	VA-Q-I-EA-V-Q-E-SH-KL-N-AQ-Q-VDAE-S	845
H2A.CI.88.UC2	VASQ-I-D-EAV-I-V-E-IH-I-QL-N-AQ-Q-V-AEI-V-S	845
H2A.DE.x.BEN	VA-Q-N-T-EAV-V-E-II-TH-I-LL-NS-AQ-Q-V-AEI-V-Y-S	845
H2A.DE.x.PE12	VA-Q-K-ETL-V-V-SH-KL-N-AQ-Q-VDAE-S	845
H2A.GH.x.GH1	VV-Q-I-D-EAV-V-R-E-M-TH-I-QL-N-AQ-Q-V-AEI-V-S	845
H2A.GM.x.ISY	V-Q-A-I-K-EA-V-R-G-AH-NL-N-AQ-Q-V-AE-S	845
H2A.GM.87.D194	VA-Q-N-I-D-EAV-V-E-I-TH-I-QL-N-AQ-Q-V-AEI-V-S	845
H2A.GW.x.AL	VA-Q-N-I-EA-V-V-E-SH-NL-N-AQ-Q-V-AE-V-I-S	845
H2A.GW.x.MDS	VV-Q-N-T-EA-V-V-E-II-TH-I-QL-N-AQ-Q-V-AE-V-S	845
H2A.GW.86.FG	VA-Q-N-I-EA-V-V-E-I-SH-I-KL-N-AHVO-V-AE-V-S	846
H2A.GW.87.CAM2C2	VA-Q-N-T-EA-V-V-E-T-CH-DI-QL-N-AQVO-V-AEV-S	844
H2A.SN.85.ROD	SASQ-KT-EA-V-V-E-SH-I-NL-NS-AQ-Q-V-AE-S	846
H2AB.CI.90.7312A	VA-Q-T-PI-K-EA-G-L-V-E-G-H-I-QL-K-NS-Q-V-V-AE-V-S	846
H2B.CI.x.EHO	VAAQ-T-PI-RE-EK-G-L-V-E-N-H-I-QL-NS-Q-V-E-V-S	845
H2B.CI.88.UC1	VA-Q-T-P-EA-G-R-L-V-E-G-H-QL-K-NS-Q-V-V-AE-V-S	846
H2B.GH.86.D205	AAQ-T-PI-AK-EAV-G-L-V-E-G-H-I-QL-K-NS-Q-V-A-S	845
H2B.JP01.KR020	AAQ-T-PI-KT-V-EA-G-L-V-E-G-H-I-QL-K-NS-Q-V-AE-S	847
H2C.LR.x.2238	XL-A-Q-T-P-R-L-R-EAV-IS-L-V-ERF-A-TH-I-NL-K-N-HV-Q-V-A-V-S	166
H2D.LR.90.F0784PA	XI-VA-Q-K-EA-I-R-V-D-E-I-KL-I-R-V-AE-V-S	166
H2D.LR.90.FORTC2	VA-Q-K-EA-I-R-V-D-E-I-KL-I-R-V-AE-VT-S	657
H201_AE.US.91.PA		332
H2G.CI.x.ABT96	X-AXQ-T-P-K-X-EXL-G-V-X-X-EXF-TH-I-QL-K-NSYHS-Q-V-AE-V-V-S	828
H2U.FR.96.12034	I-A-Q-T-P-SK-L-R-EAV-G-V-V-E-QH-I-QL-K-N-HV-Q-M-V-TEV-VV-S	841
MAC.US.x.251.BK28		828
MAC.US.x.SMM142B		828
MAC.US.x.MM142		828
MAC.US.x.92050		832
MNE.US.x.MNE027		828
MNE.US.82.MNE_8		828
SMM.SL.92.SL92B	LN-Q-S-D-DI-A-QLVQ-EAV-IG-N-V-R-S-D-AQ-YNI-QL-K-NA-N-Q-T-AEV-V-R-I-S	825
SMM.US.95.TNPRC_G932		123
SMM.US.02.YNPRC_FBL		123
SMM.US.86.NIRC_CFU233		123
SMM.US.03.TNPRC_D215		123
SMM.US.95.TNPRC_D175		123
SMM.US.94.TNPRC_M927		123
SMM.US.94.TNPRC_G080		123
SMM.US.79.NIRC_CFU212		123
SMM.US.79.NIRC_6007_G932		123
SMM.US.95.TNPRC_F102		123
SMM.US.95.TNPRC_D177		123
SMM.US.91.TNPRC_G931		123
SMM.US.81.TNPRC_G930		123
SMM.US.03.TNPRC_G932		123
SMM.US.85.TNPRC_F104		123
SMM.US.84.TNPRC_F104		123
SMM.US.79.NIRC_CFU233		123
SMM.US.86.NIRC_CFU232		123
SMM.US.86.NIRC_CFU226		123
SMM.US.82.NIRC_6007_G932		123
SMM.US.86.NIRC_6002_G931		123
SMM.US.86.NIRC_6001_G930		123
SMM.US.79.NIRC_6001_G930		123
SMM.US.02.YNPRC_FIP		123
SMM.US.x.F236_H4		820
SMM.US.x.H9	VA-Q-K-EA-X-E-V-E-XK-E-E-K-X-L-K-L-X-V-XX-G-X-S	828
SMM.US.x.PB14_15	VA-Q-EA-V-E-E-E-K-X-L-K-L-X-V-XX-G-X-S	828
SMM.US.x.PB1A	VA-Q-EA-V-E-E-E-K-X-L-K-L-X-V-XX-G-X-S	828
SMM.US.x.PB1_143	VA-Q-EA-X-V-E-E-E-K-X-L-K-L-X-V-XX-G-X-S	828
SMM.US.x.PB1_6P6	VA-Q-EA-V-E-E-E-K-X-L-K-L-X-V-XX-G-X-S	828
SMM.US.x.PGM53	VA-Q-EA-V-E-E-E-K-X-L-K-L-X-V-XX-G-X-S	828
SMM.US.x.SME543	VA-Q-EA-V-E-E-E-K-X-L-K-L-X-V-XX-G-X-S	828
STM.US.x.STM	Q-K-EA-V-E-E-E-K-X-L-K-L-X-V-XX-G-X-S	828

MAC.US.x.239	WPITLHTDNGANFASQEVKMAVWAGIEHTFGVYPNPQSGVQVEAMNHLKNQIDRIREQANSVETIVLMAVHCMNFKRRGGIGDMTPAERLINMITTEQEQIQFQSKNSKFKNFRVYYREGRDQLWKGGPELLWKGGEGAVILKVGTDIKVPPRRKAKIKDYGGGKEV	1002
H2A.GM.x.MCN13	-----T-----V-0S-----S-----T-----T-----LHA-L-L-F-----D-V-----RQ-L	1015
H2A.CI.88.UC2	-----P-T-----V-0S-----S-----T-----LHA-L-N-K-Q-----V-A-----R-RO-L	1015
H2A.DE.x.BEN	-----P-T-----V-0S-----S-----T-----LHA-L-N-Q-----V-A-----R-RO-L	1015
H2A.DE.x.PE12	-----P-T-----T-0S-----S-----TM-----LHA-L-L-F-----D-V-----I-----R-RR-L	1015
H2A.GH.x.GH1	-----S-T-----I-0S-----S-----TI-----L-R-N-Q-----D-V-A-I-----R-RO-L	1015
H2A.GM.x.ISY	-----T-----V-0S-----E-----TM-----V-V-----L-A-L-F-----N-Q-----D-V-A-I-----R-PRO-M	1015
H2A.GM.87.D194	-----P-T-----I-0S-----S-----TI-----L-R-N-K-Q-----D-V-A-I-----R-RO-L	1015
H2A.GW.x.AL1	-----S-V-T-----V-0S-----S-----D-----S-----S-----L-T-L-P-----D-V-----I-----R-RO-L	1015
H2A.GW.x.MDS	-----T-----V-0S-----S-----D-----S-----I-----L-A-L-F-----D-V-E-I-----R-RO-M	1016
H2A.GW.86.FG	-----T-----V-0S-----S-----T-----S-----I-----L-A-L-F-----D-V-E-I-----R-RO-M	1016
H2A.GW.87.CAM2CG	-----T-----V-0S-----S-----T-----S-----I-----L-A-L-F-----D-V-E-I-----R-RO-M	1016
H2A.SN.85.ROD	-----T-----I-0S-----S-----TI-----I-----S-----L-A-L-D-F-----LV-----II-----R-RO-M	1016
H2AB.CI.90.7312A	-----T-D-----V-0-----D-V-I-V-A-----V-----F-A-L-Q-Q-----I-E-I-----RH-----L	1016
H2B.CI.x.EHO	-----T-D-A-I-0-----E-----D-V-I-V-T-----IV-----L-T-L-Q-----I-E-I-----RN-----L	1015
H2B.CI.88.UC1	-----T-D-A-I-0-----V-----D-V-I-V-A-----V-----F-A-L-Q-Q-----LI-----E-I-----RH-----L	1016
H2B.GH.86.D205	-----T-PS-----V-0-----L-D-V-I-V-T-----V-----F-A-L-Q-Q-----I-E-I-----RH-----L	1015
H2B.JP01.KR020	-----T-D-A-I-V-0-----D-V-I-V-A-----IV-A-----L-T-KL-Q-----F-----I-E-I-----N-----L	1017
H2C.LR.x.2238	-----T-----V-V-----TM-----	236
H2D.LR.90.FO784PA	-----T-----V-0-----K-D-----	236
H2D.LR.90.FORTC2	-----T-----V-0-----	657
H201_AE.US.91.PA	-----T-----V-0A-----S-----K-K-D-I-----I-----T-----I-E-----	332
H2G.CI.x.ABT96	-----T-----V-0A-----S-----K-K-D-I-----I-----S-----L-----AD-----LI-----I-----R-----L	998
H2U.FR.96.12034	-----T-----V-0A-----E-----TM-----A-----I-----S-----L-----AD-----LI-----I-----R-----L	1011
MAC.US.x.251_BK28	-----T-----V-0A-----E-----TM-----A-----I-----S-----L-----AD-----LI-----I-----R-----L	998
MAC.US.x.SMM142B	-----T-----V-0A-----E-----TM-----A-----I-----S-----L-----AD-----LI-----I-----R-----L	998
MAC.US.x.MM142	-----T-----V-0A-----E-----TM-----A-----I-----S-----L-----AD-----LI-----I-----R-----L	998
MAC.US.x.92050	-----T-----V-0A-----E-----TM-----A-----I-----S-----L-----AD-----LI-----I-----R-----L	1002
MNE.US.x.MNE027	-----T-----M-----L-----	998
MNE.US.82.MNE_8	-----T-----M-----L-----V-----	998
SMM.SL.92.SL92B	-----K-----T-----L-V-0S-----DL-KN-K-E-L-A-----IV-----L-T-YLN-Q-Q-----A-----VI-E-----RQ-M	995
SMM.US.95.TNPRC_G932	-----T-----	198
SMM.US.02.YNPRC_FBL	-----T-----	203
SMM.US.86.NIRC_CFU233	-----T-----TN-----	203
SMM.US.03.TNPRC_D215	-----T-----	203
SMM.US.95.TNPRC_D175	-----T-----	203
SMM.US.94.TNPRC_M927	-----T-----	203
SMM.US.94.TNPRC_G080	-----T-----	203
SMM.US.79.NIRC_CFU212	-----T-----TN-----	198
SMM.US.79.NIRC_6007_G932	-----T-----	198
SMM.US.95.TNPRC_F102	-----T-----	198
SMM.US.95.TNPRC_D177	-----T-----DQ-----	198
SMM.US.91.TNPRC_G931	-----T-----	198
SMM.US.81.TNPRC_G930	-----T-----	198
SMM.US.03.TNPRC_G932	-----T-----	203
SMM.US.85.TNPRC_F104	-----T-----I-----R-----	203
SMM.US.84.TNPRC_F104	-----T-----	203
SMM.US.79.NIRC_CFU233	-----T-----	203
SMM.US.86.NIRC_CFU232	-----T-----TA-----	203
SMM.US.86.NIRC_CFU226	-----T-----	203
SMM.US.82.NIRC_6007_G932	-----T-----TN-----	203
SMM.US.86.NIRC_6002_G931	-----T-----	203
SMM.US.86.NIRC_6001_G930	-----T-----	203
SMM.US.79.NIRC_6001_G930	-----T-----	203
SMM.US.02.YNPRC_FIP	-----T-----	203
SMM.US.x.F236_H4	-----T-----	999
SMM.US.x.H9	-----T-----X-0-X-X-----T-----I-----I-----V-----	998
SMM.US.x.PBJ14_15	-----T-----	998
SMM.US.x.PBJA	-----T-----	998
SMM.US.x.PBJ_143	-----T-----X-0-X-X-----T-----I-----I-----V-----	998
SMM.US.x.PBJ_6P6	-----T-----	998
SMM.US.x.PGM53	-----T-----	998
SMM.US.x.SME543	-----T-----	998
STM.US.x.STM	-----V-----T-----Q-----T-----D-T-V-----K-L-----V-----R-----V-----	998

	p31 Integrase end Pol end	
MAC.US.x.239	DSSSHMEDTGEAR.EVA*	1018
H2A.GM.x.MCN13	--G-L-GAR-DG-M-	1032
H2A.CI.88.UC2	--P-L-GAR-DG-M-CPCQVPEIQ	1032
H2A.DE.x.BEN	--L-G-R-DG-	1040
H2A.DE.x.PE12	--HLEGARE-DG-	1032
H2A.GH.x.GH1	--G-L-GAR-DG-M-	1032
H2A.GM.x.ISY	--L-GAR-DG-	1032
H2A.GM.87.D194	--GP-L-GAR-DG-	1032
H2A.GW.x.AL1	--GP-L-G-R-DG-	1032
H2A.GW.x.MDS	--G-L-GAR-DG-M-	1032
H2A.GW.86.FG	--L-GAR-NG-	1033
H2A.GW.87.CAM2CG	--G-L-GAR-DG-M-	1031
H2A.SN.85.ROD	--C-TDV--RQ--M-QSSQVSEA*	1033
H2AB.CI.90.7312A	--C-ADV--M0--QSN+IPEI*	1040
H2B.CI.x.EHO	--CGTD---RQ---M-QSGQVPEA*	1038
H2B.CI.88.UC1	--C-AD---RQ---M-QSD*	1040
H2B.GH.86.D205	--C-AN----Q---M-QPSQISET*	1034
H2B.JP.01.KR020		1041
H2C.LR.x.2238		236
H2D.LR.90.FO784PA		236
H2D.LR.90.FORTC2		657
H201_AE.US.91.PA		332
H2G.CI.x.ABT96	--TN---RQTG----	1015
H2U.FR.96.12034	---NL-GAEKV-.M-LPDQTPPEYX	1036
MAC.US.x.251_BK28		1015
MAC.US.x.SMM142B		1015
MAC.US.x.MM142		1015
MAC.US.x.92050		1019
MNE.US.x.MNE027		1015
MNE.US.82.MNE_8		1015
SMM.SL.92.SL92B	G-AS--Q+KNSGAL--	1012
SMM.US.95.TNPRC_G932		198
SMM.US.02.YNPRC_FBL		203
SMM.US.86.NIRC_CFU233		203
SMM.US.03.TNPRC_D215		203
SMM.US.95.TNPRC_D175		203
SMM.US.94.TNPRC_M927		203
SMM.US.94.TNPRC_G080		203
SMM.US.79.NIRC_CFU212		198
SMM.US.79.NIRC_6007_G932		198
SMM.US.95.TNPRC_F102		198
SMM.US.95.TNPRC_D177		198
SMM.US.91.TNPRC_G931		198
SMM.US.81.TNPRC_G930		198
SMM.US.03.TNPRC_G932		203
SMM.US.85.TNPRC_F104		203
SMM.US.84.TNPRC_F104		203
SMM.US.79.NIRC_CFU233		203
SMM.US.86.NIRC_CFU232		203
SMM.US.86.NIRC_CFU226		203
SMM.US.82.NIRC_6007_G932		203
SMM.US.86.NIRC_6002_G931		203
SMM.US.86.NIRC_6001_G930		203
SMM.US.79.NIRC_6001_G930		203
SMM.US.02.YNPRC_FIP		203
SMM.US.x.F236_H4	--G-L-----	1016
SMM.US.x.H9	--G-L-----	1015
SMM.US.x.PB14_15	--G-L-----	1015
SMM.US.x.PB1A	--G-L-----	1015
SMM.US.x.PB1_143	--G-L-----	1015
SMM.US.x.PB1_6P6	--G-L-----	1015
SMM.US.x.PGM53	--GP-L-----	1015
SMM.US.x.SME543	--G-L-----	1015
STM.US.x.STM	--G-L--R--G----	1015

Vif start

MAC.US.x.239	MEEEKRWIAVPTWRIPER.L	ERWHSLIKLYKYTKDLQKVCYVPHFKVGMWMTCSRVI	FLPQEGSHLEVOGYWHLTPEKGLSTYAVRITWYSKNFWDVTPNYADILLHSTYFPCFTAGEVRRRAIRGEQLLSCCRFPRAHKYQVPSLQYLAKVVS	.DVR	SQGENP	166
H2A.GM.x.MCN13	--G-N-V--V-G-M-K--V--R-ED--H-----KGN-I-A-N-----S-S-M-TER-----DC-S-I-S-----K-NY-Q-RA-----F-V-00-ND-P-RNGT					167
H2A.CL88.UC2	--G-S-V--V-G-M--V--R-EG-R-H-----KGE-I-A-N-----SHS-L-TEK-----DC-I-C-A-----K-NY-Q-A-----F-V-00-ND-P-RKG-					167
H2A.DE.x.BEN	--DRH-V--V-G-M-K-A-V--R-EG-R-H-----GK-I-A-N-----SH-L-TEK-----DC-S-I-S-----K-NY-Q-A-----F-V-00-ND-P-RKGT					167
H2A.DE.x.PE12	--GE--V--V-G-M-K--V--HR-EG-R-H-----GN-I-A-N-----S-S-M-TER-----DC-S-I-S-----K-NY-Q-RSK-L-F-V-00-ND-P-KNST					167
H2A.GH.x.GH1	--G-N-V--V-G-RM--V--R-EE-R-H-----KGE-I-A-N-----SHS-TER-----D-I-S-----K-NY-Q-V-----F-V-00-ND-P-RKGT					167
H2A.GM.x.ISY	--DQG--V--V-G-M-K--V--R-EO-R-H-----KGN-I-A-N-----S-S-M-EG-----DC-T-I-S-----KS-NY-Q-SK-----F-V-00-NDPK-RD-T					167
H2A.GM.87.D194	--G-N-V--V-G-M--V-H-EE-R-H-----EGE-I-A-N-----SHS-L-TEK-----DC-S-I-S-----K-NY-Q-A-----F-V-00-ND-P-RKGA					167
H2A.x.JAU2	--D--V--V-G-M-K--V--R-EE-H-----K-I-A-N-----S-TER-----DC-S-I-G-S-----K-NY-Q-S-----F-V-00-ND-P-RKYT					167
H2A.GW.x.ALI	--G-S-V--V-G-M-K--V--R-EE-H-----GR-I-A-N-----S-TER-----DC-S-I-G-S-----K-NY-Q-S-----F-V-00-NDPK-RNST					167
H2A.GW.x.MDS	--G-T-V--V-G-M-K--V--R-EE-R-H-----KGN-I-A-N-----S-L-TEK-----DC-S-I-T-----A-K-NY-Q-S-T-----F-V-00-ND-P-RDST					167
H2A.GW.86.FG	--G--V-I--V-G-M--V--R-EE-H-----K-N-I-A-N-----SHS-TER-----DC-T-I-S-----K-KY--RS-----F-V-00-ND-P-NSA					167
H2A.GW.87.CAM2CG	--G-S--V--V-G-M-K--V--G-EO-H-----RGD-R-I-A-N-----S-S-M-TEK-----DC-T-I-S-----K-KY--RS-----F-V-00-ND-P-RDRT					167
H2A.SN.85.ROD	--D--V--V-G-M-K--V--E-H-----KGN-I-A-N-----S-S-TEK-----DC-V-I-----K-NY--RA-----F-V-00-ND-P-RDST					167
H2AB.CL90.7312A	--G-N-V--K-G--K--V-HR-GE-Q-S-H-----KG-AC--N-R-F-S-L-EES-Y-DV-Q-G-S-N-KI-Y-NY-S-EG-F-R-OEGKNG--SA					168
H2B.CL.x.EHO	--N--N--C--R--Q-S-H-----K-A--N-R-F-S-L-ERS-Y-DV-R-GS-SS-N-KI-H-NY-S-TG-F-R-OEGKDG--ST					168
H2B.CL88.UC1	--G-N-V--G--K--V-HR-E-Q-S-H-----K-EAY--N-R-F-S-L-KRS-Y-DV-Q-GS-S-N-KI-Y-NY-S-EG-F-R-OEGKDG--SA					168
H2B.GH.86.D205	--D--V--G--K--V-HR-GE-Q-S-H-----I-NK-AW--N-R-F-S-L-ER-Y-DV-Q-GS-F-S-N-KI-Y-NY-S-EG-F-R-OEGKNG--SA					168
H2B.JP01.KR020	--N--V--G--K--V-HR-E-Q-T-H-----K-A-Y-I--N-R-F-A-L-K-S-Y-DV-H-GS-F-N-KI-Y-NY-S-RG-F-R-OEGGNGP--SA					168
H2G.CL.x.ABT96	--G--K--F-F--AV-H-----TKEA-I--N-Q-L-TRK-Y-ET-Q-GS-D-----I-NY-T-R-F-Q-QK-GHG-K-SX					167
H2U.FR.96.12034	--N-V--V-R--H-X-N-E-H-----KK--A-N-S-L-TRG-DC-Q-GS-S-----N-S-RQ-F-RALQ--G-KR--					166
MAC.US.x.251.1A11						166
MAC.US.x.251.32H.PJ5						166
MAC.US.x.251.BK28						166
MAC.US.x.BK28.H824						166
MAC.US.x.MM142						166
MAC.US.x.SMM142B						166
MAC.US.x.1937						166
MAC.US.x.2065						166
MAC.US.x.95112						166
MAC.US.x.96114						166
MAC.US.x.80035						166
MAC.US.x.81035						166
MAC.US.x.85013						166
MAC.US.x.MAC239_87082						166
MAC.US.x.92050						166
MAC.US.x.92077						166
MAC.US.x.93057						166
MAC.US.x.95058						166
MAC.US.x.95086						166
MAC.US.x.96016						166
MAC.US.x.96020						166
MAC.US.x.96072						166
MAC.US.x.96081						166
MAC.US.x.96093						166
MAC.US.x.96123						166
MNE.US.x.MNE027						166
MNE.US.82.MNE_8						166
SMM.SL.92.SL.92B	--G--V--V--SR-IV--C-FH--RE-E-A-H--Y-A-E--V-N--S--EK--DV-Q-A--A-HA-Q--V-Y-GYAV-HSS-Q--L--LQ-ND-PK-K--					168
MAC.US.x.17EC1						166
MAC.US.x.17EFR						166
SMM.US.x.F236_H4	--N-V--G--K--H-N--MA-H-----RDET--N-A--R--D-T--SE--K--K--K--K--N--T--H--D--					166
SMM.US.x.H9	--N-V--G--K--H-N--MA-H-----RDEA--N-X--R--A-D-T-G--SE--K--K--K--N--T--H-X-X--					166
SMM.US.x.PBJ14_15	--N-V--G--K--H-N--A-H-----KDEA--N--R--D-T-G--SE--K--K--K--N--T--H--					166
SMM.US.x.PBJA	--N-V--G--K--H-N--A-H-----KDEA--N--R--D-T-G--SE--K--K--K--N--T--H--					166
SMM.US.x.PBJC	--N-V--G--K--H-N--A-H-----KDEA--N--R--D-T-G--SE--K--K--K--N--T--H--					166
SMM.US.x.PBJD	--N-V--G--K--H-N--A-H-----KDEA--N--R--D-T-G--SE--K--K--K--N--T--H--					166
SMM.US.x.PBJE	--N-V--G--K--H-N--A-H-----KDEA--N--R--D-T-G--SE--K--K--K--N--T--H--					166
SMM.US.x.PBJ_143	--N-V--G--K--H-N--A-H-----RDEA--N-X--R--A-D-T-G--SE--K--K--K--N--T--H-X-X--					166
SMM.US.x.PBJ_6P6	--N-V--G--K--H-N--A-H-----KDEA--N--R--D-T-G--SE--K--K--K--N--T--H--					166
SMM.US.x.PGM53	--N-V--K--G--K--FR--A-H-----RDE--N-A--R--D-T-G--SE--K--K--K--N--T--H--R-D--					166
SMM.US.x.SME543	--N-V--K--G--K--FR--A-H-----RDET--N-A--H--R--DC-T-G--SE--Q--K--K--K--N--T--H--R-D--					166
SMM.US.x.BFZ_m12	--N-V--K--G--K--FR--A-H-----RDE--N-A--R--D-T-G--SE--K--K--K--N--T--H--R-D--					166
STM.US.x.STM	--V--G--H--N-E-S-A-H-----GEA--N--E--TR-S-DC-Q-G-----K--TK--N--E--H--R--T					166

	Rev start	exon 1 end	exon 2 start	Rev end	
MAC.US.x.239	MSNHEREEELRKRRLIHLHQTNPYPTGPGTANQRQRKRWRRRRQQLALADRIYSFPDPTDPLDLAIQQLNLAIIESIPDPPTNTPEALCDPTEDSRSPQD*				107
H2A.GM.x.MCN13	TGRAD	---RK---	Q-L-S-N-R--KQ-R-V-K-T-A-S-GRT-H-E-T-QDL-SF-SSEGTDSNO-LAKN		108
H2A.CI.88.UC2	CEKAD	---ORE---	YR-G-N-R-R-Q-LR-V-KLHTA-A-SS-W-H-G-T-REL-DL-S-DSNOGLAET		104
H2A.DE.x.BEN	ERAD	---G-ORK-LR---	Q-L-S-N-R-R-Q-LR-V-NKLCAV-A-S-R-H-G-T-QEL-DL-S-NSNOGLAET		104
H2A.DE.x.PE12	NGRAD	---G-ORQO---	Q-L-R-N-R-RKQH-R-V-NS-T-A-S-R-H-G-T-QEL-L-SSESTNNQGLAETYNLSLPAIIVRVDPRAAPGPKDYERDSYERVLVGGNGTDRQGTCSSSKKDQAGGRTCPVVR		170
H2A.GH.x.GH1	HEKADG	---OE---	H-S-N-R-Q-LR-V-KL-T-S-R-D-R-T-HEL-DL-S-NSNOGLAET		104
H2A.GM.x.ISY	TERAD	---GV-RK---	Q-L-R-N-R-E-K-I-T-A-P-QT-G-T-QTL-TQ-LAETQGLPAVWVVDPRSPVPGREGYKRDSYERGEELVGGSGTNRKGDTRSSTKDQAGSRNCPVVR		159
H2A.GM.87.D194	RDRAD	---G-ORK---	I-H-Q-S-N-R-R-Q-FR-V-TKLHTI-S-R-G-T-QEL-DL-S-NSNOGLAET		104
H2A.GW.x.AL1	TERAG	---D-ORR---	Q-S-N-R-Q-G-IV-FT-ASS-R-V-H-G-T-QDL-DL-SSEADNNOGLAET		108
H2A.GW.x.MDS	NERAD	---G-ORR---	Q-S-N-R-Q-R-I-S-T-A-S-QTV-G-T-QDL-HL-S-Q-PAAT		101
H2A.GW.86.FG	TERAD	---G-ORR---	Q-S-N-R-KQ-R-I-S-T-A-S-R-H-G-T-QDL-HL-SPESTNSNQ-LAEA		108
H2A.GW.87.CAM2CG	TERAD	---G-ORR---	Q-S-N-R-Q-R-I-S-T-A-P-RT-H-E-T-QDL-HP-S-Q-LAEA		101
H2A.SN.85.ROD	NERAD	---G-ORR---	Q-S-N-R-KQ-R-I-S-T-A-S-QT-H-G-T-QEL-HL-S-Q-LAEA		101
H2AB.CI.90.7312A	TAR-K.	D-Q-GI-L	Q-S-N-R-K-GL-I-HPL-NS-AEE-RR-TV-DL-N-SS-TT-APVCVPIWDLVPRSPSSGGYGRDSCHEREDLMGGSQEDGEGNHRSPQKNTGTA*		150
H2B.CI.x.EHO	NAR	---D-Q-G-L---	Q-S-N-R-KQ-GL-I-HPL-S-EG-R-I-KDL-N-S-TA-ASTCIPPIWDLVPRSNPSSQGGCRDSCERGEDLVGSPQESGRRDHCNTQEDOTRG*		150
H2B.CI.88.UC1	TTR-K.	D-Q-G-L---	Q-S-N-R-K-GL-I-R-LS-S-EE-V-R-E-TV-DL-N-S-TA-AFTCIPPVWDLVPRSNPSSNEGCRDSCERKSPMESSQKDSGNSHRDPOEDOTRT*		150
H2B.GH.86.D205	TAR-G.	D-Q-E-L---	Q-S-N-R-K-GL-I-H-STA-AEE-N-R-TV-DL-N-P-LNOSP-TT-APGCVPPVWDLVPRSPSGSGYGRNSCECRDLMGGSQESGESNHRDPOENQTRT*		153
H2B.JP01.KR020	TTR-K.	D-Q-G-LR---	Q-S-N-R-GL-I-Q-HPL-S-EE-N-K-TV-DL-T-ST-TA-TSTCIPPIWDLVPRSNPSSDEGCRDYSYKHKRKGPMGGSQKNSGNHRDPOEDQTRTRTRPLVR		157
H2G.CI.x.AB196	L-E	---F---	Q-X-N-R-K-KQ-L-I-X-ARE-S-T-QDX-S-AV-KN-Q-PSN		101
H2U.FR.96.12034	P-E.	GVOR	N-R-Q-L-I-K-T-S-V-G-QDL-E-AVV-H-Q-TSAP		99
MAC.US.x.251_1A11	S		R-K		108
MAC.US.x.251_32H_PJ5	S		R-N		108
MAC.US.x.251_BK28	S		R-KG		108
MAC.US.x.BK28_H824	S		R-RG		108
MAC.US.x.MM142	R5-TG	---R---	S-S-KR-Q-I-N-A		108
MAC.US.x.SMM142B	R5-TG	---R---	S-S-KR-Q-I-LRRIR-A		108
MAC.US.x.1937	XX	---R---	X-I-X		108
MAC.US.x.2065			R-X		108
MAC.US.x.95112			R-X		108
MAC.US.x.96114			R-X		108
MAC.US.x.80035			R-X		108
MAC.US.x.81035			R-X		108
MAC.US.x.85013	X		R-X		108
MAC.US.x.MAC239_87082			R-X		108
MAC.US.x.92050			R-I		108
MAC.US.x.92077			R-I		108
MAC.US.x.93057			R-X		108
MAC.US.x.93062			R-V		108
MAC.US.x.95086			R-V		108
MAC.US.x.96016			R-V		108
MAC.US.x.96020			R-X		108
MAC.US.x.96072	X		R-X		108
MAC.US.x.96081			R-X		108
MAC.US.x.96093			R-V		108
MAC.US.x.96123			R-V		108
MAC.US.x.90131			R-V		108
MNE.US.x.MNE027	S-AE	---R---	S-N-R-Q-F-V-I-I-H-SP		108
MNE.US.82.MNE_8	S-AE	---R---	S-N-R-Q-N-V-I-I-H-NP		108
SMM.SL.92.SL.92B	Q-P-E.	---R---	I-A-S-N-R-K-I-IV-A-DI-FDS-S-Q-L-E-TV-T-R-QSA		100
MAC.US.x.17EC1	N		R-T		108
MAC.US.x.17EFR	N		R-T		108
SMM.US.x.F236_H4	ST-E.	---R---	F-Q-R-R-Q-I-V-G-EL-SA-P-N-VAKSP		101
SMM.US.x.H9	SN-E.	---R---	F-XX-R-R-XO-I-X-A-G-XL-N-ASA-P-K-AA-SP		101
SMM.US.x.PBJ14_15	T-SNGE	---R---	F-S-D-R-R-Q-I-X-V-R-EL-N-ASA-P-K-AA-SP		101
SMM.US.x.PBJA	SN-E.	---R---	F-D-R-R-Q-I-V-R-EL-N-ASA-P-K-AA-SP		101
SMM.US.x.PBJC	SN-E.	---R---	F-D-R-R-Q-I-V-G-EL-N-ASA-P-K-AA-SP		101
SMM.US.x.PBJD	SN-E.	---R---	F-D-R-R-Q-I-V-R-EL-N-ASA-P-K-AA-SP		101
SMM.US.x.PBJE	SN-E.	---R---	F-D-R-R-Q-I-V-R-EL-N-ASA-P-K-AA-SP		101
SMM.US.x.PBJ_143	SN-E.	---R---	F-XX-R-R-XO-I-X-A-G-XL-N-ASA-P-K-AA-SP		101
SMM.US.x.PBJ_6P6	SN-E.	---R---	F-D-R-R-Q-I-V-R-EL-N-ASA-P-K-AA-SP		101
SMM.US.x.PGM53	SSAG	---R---	F-Q-R-R-Q-I-A-S-V-D-SL-N-SV-NP-R		105
SMM.US.x.SME543	ST-E.	---R---	F-Q-R-R-Q-I-N-G-DL-SA-T-K-AAKS		101
SMM.US.x.BFZ_m12	SNAG	---R---	F-Q-R-R-Q-I-A-V-D-SL-N-SV-T-LT-SP-H		108
STM.US.x.STM	DQ-E.	---QF---	Q-Q-N-R-Q-N-N-AS-V-G-S-QDL-P-L-KD-Q-TA-N		100

MAC.US.x.239	107
H2A.GM.x.MCN13	108
H2A.Cl.88.UC2	104
H2A.DE.x.BEN	104
H2A.DE.x.PE12	GSGINRÉTL*	179
H2A.GH.x.GH1	104
H2A.GM.x.ISY	DRDISKETL*	168
H2A.GM.87.D194	104
H2A.GW.x.AL1	108
H2A.GW.x.MDS	101
H2A.GW.86.FG	108
H2A.GW.87.CAM2CG	101
H2A.SN.85.ROD	101
H2AB.Cl.90.7312A	150
H2B.Cl.x.EHO	150
H2B.Cl.88.UC1	150
H2B.GH.86.D205	153
H2B.JP.01.KR020	DAVLQEHKGEGRV*	171
H2G.Cl.x.ABT96	101
H2U.FR.96.12034	99
MAC.US.x.251_1A11	108
MAC.US.x.251_32H_PJ5	108
MAC.US.x.251_BK28	108
MAC.US.x.BK28_H824	108
MAC.US.x.MM142	108
MAC.US.x.SMM142B	108
MAC.US.x.1937	108
MAC.US.x.2065	108
MAC.US.x.95112	108
MAC.US.x.96114	108
MAC.US.x.80035	108
MAC.US.x.81035	108
MAC.US.x.85013	108
MAC.US.x.MAC239_87082	108
MAC.US.x.92050	108
MAC.US.x.92077	108
MAC.US.x.93057	108
MAC.US.x.93062	108
MAC.US.x.95086	108
MAC.US.x.96016	108
MAC.US.x.96020	108
MAC.US.x.96072	108
MAC.US.x.96081	108
MAC.US.x.96093	108
MAC.US.x.96123	108
MAC.US.x.90131	108
MNE.US.x.MNE027	108
MNE.US.82.MNE_8	108
SMM.SL.92.SL.92B	100
MAC.US.x.17EC1	108
MAC.US.x.17EFR	108
SMM.US.x.F236_H4	101
SMM.US.x.H9	101
SMM.US.x.PBJ14_15	101
SMM.US.x.PBJA	101
SMM.US.x.PBJC	101
SMM.US.x.PBJD	101
SMM.US.x.PBJE	101
SMM.US.x.PBJ_143	101
SMM.US.x.PBJ_6P6	101
SMM.US.x.PGM53	105
SMM.US.x.SME543	101
SMM.US.x.BFZ_m12	108
STM.US.x.STM	100

Env start	signal peptide end_gp120 start	
MAC.US.x.239	MGCLGNQLLIAILLSSVYGIYCT	LYVTVFYGVPAWRNATIPLF*.ATKNRDTWGTQCLPDNGDYSEVALNVTESFDANNVTVEQAIEDVWQLFETSIKPCKVLSPICITMRCKNSETDRWGLTKSI.....TTTASTTSTTASAK.....VDMVNETSSCIAQDN* 157
H2A.GM.x.CBL23	MGR	V-A-TCL-N-I-S-I-D-Q-IT-A-D-N-T-VA-NNTDARN--PT-A-PRTIKP.....TEIS-N--RAN-- 150
H2A.CL88.UC2	EPGR	AV-T-ACL-K-Q-I-V-S-I-D-Q-IP-A-D-R-NPVTGNNTATAKPTAARP-NPSY.....LTIII-S-T-VGA-- 154
H2A.DE.x.BEN	EPGR	FVV-T-ACLV-S-Q-I-K-S-I-D-Q-II-A-D-V-H-RVQNTI-PNPR-SSTTSRPPSTAASII--W-FNNT-- 155
H2A.FR.x.96226		SAIPSSITTPGSTM-TASKAST-PT-I-A-VGK-- 84
H2A.GH.x.GH1		GKSL-CV-S-A-LV-Q-V-S-I-K-D-Q-IT-A-D-V-S-T-VA-SNS-TNN-T-G-TT.....GMEINET-PSYS-- 142
H2A.FR.x.96330		XKEI--SDT-C-- 16
H2A.FR.x.96206		XNLI--NP-KN-S- 16
H2A.GM.87.D194	EPGR	V-T-ACL-K-Q-I-S-I-D-Q-IT-A-D-R-NITSG--ATPSPPN.....ITIID-N-T-GDN-- 143
H2A.GM.90.CBL24	GR	VTT-A-ACL-S-K-I-S-I-D-Q-I-A-D-V-S-MTA-PEGEAKA-SITTAATAGAMEI--P-VTN-- 151
H2A.GW.x.AL1	MSSR	VT-A-ACLV-K-Q-I-K-S-I-D-Q-I-A-D-V-R-NISTES--SPSPGS-L.....KPLI--SDP-KA-- 147
H2A.GW.x.CAM1	AYKR	T-ACL-KRKO-I-S-I-D-Q-IP-A-D-N-DGN-TSTGN--STRARI.....SEEIK-N--SA-- 150
H2A.GW.86.FG	KGSK	V-A-A-L-H-K-OF-I-S-I-D-Q-IT-A-D-V-N-TRNM-TW-GR-DTON.....ITII-D-H-ARA-- 143
H2A.GW.87.CAM2CG	ERGR	A-ACL-R-Q-Q-K-S-I-D-Q-IP-A-D-I-NIST-DTMTIR--P-AK.....EAPISDN-P--RTN-- 149
H2A.SN.x.ST	KM-GR	FV-S-A-ACLV-Q-V-S-I-D-Q-I-A-D-V-S-NS-KN-TS-PTITTTA..NTTIG-N--RT-- 147
H2A.SN.85.ROD	M	A-ACLV-Q-T-K-S-I-D-Q-IT-A-D-T-VA-K-SSTESSTGN-SKS---TTPTDQ..EQEIS-DTP-ARA-- 152
H2A.GM.x.ISY	SGKI	V-F-T-ACL-K-V-K-S-I-D-Q-IP-A-D-I-V-N-NASTESAVA--PSG.....P-I-D-DP-QLN-- 143
H2B.CL.x.EHO	AHVN-Y-VTL-I-I-YMGK-NF	I-K-S-R-V-D-T-IQ-I-A-D-D-TK-S-T-V-K-K-T.....WSSASKE--S-ASLRST.....QTLI--D-K-ON-S- 151
H2B.CL.88.UC1	AHTS-H-F-LL-I-FLGHKKN	I-K-V-T-V-D-T-IV-I-A-D--VD-S-T-VA--NTGTN--KPI-TPITTT-PS..ENLL-D-P-KN-T- 153
H2B.FR.x.96200		DOEG-KGSV-PTT.....D---P-VA-N-S-N..... 53
H2B.FR.x.97227		NTTEGGNKNE-TT-ATI-PTSTAKSAEPTSMIGG-A-G-- 87
H2B.GH.86.D205	AYFSSR-P-L-IGIS-FV-K-Q-I-V-I-T-V-T-IR-I-A-D-Q-VD-R-T-VA-N-S-T-N.....PGNAS--T-KPTTSRG..LKTI--DP-KN-S- 153	
H2B.JI.01.KR020	AHINKY-FA-L-I-FMGE-DF	I-K-V-I-V-D-D-I-I-VA-N-SNL-N-AWT-AANLTK-SKN..IS-L-GND--RO-- 149
H2G.CL.x.AB796	AY	TX-T-F-K-I-SV-D-L-A-D-V-N-T-T-A-T-NK.....G-VTPATVPTTKMVT..AEL-S-Q-LMY-- 157
H2U.FR.96.12034	A	L-C-NACV-Q-I-N-D-L-A-V-S-T-A-S-T-N-GTA...A-T-PAQOES-K.....VI--EP-RNNS- 154
H2AB.CL.90.7312A	GK-L-FV-S-A-A-L-K-V-S-I-D-Q-I-A-V-S-NSTTA--PPS-TNNTTTEPTTGGPEI--FP-MRT-- 152	
MAC.US.x.BR5		M-TTAPAKVDMV..... 145
MAC.US.x.270W		TSTTASVDMV--N-- 143
MAC.US.x.BR5		TTTASVDMV--P-- 144
MAC.US.x.418	X	M-D-G-R-TTASAKVDMV..... 146
MAC.US.x.BK28_H824		E-E-IDMV--N-- 156
MAC.US.x.MM142		D-L-E-K-S-T-TAKSVET..R-I--P-VVH-- 160
MAC.US.x.251_BK28		S-TTI-AP-APV-E...I--N-- 159
MAC.US.x.SMM142B		CSKCLW-I-I-Q-D-L-E-TAS-T-TT-SVE-TR-I--P-VVH-- 159
MNE.US.x.MNE027		F-A-I-Q-V-R-D-L-I-E-H-T-K-K-S-AP-TK-TTT...EIEV--N-T-VNR-- 158
MNE.US.82.MNE_8		V-R-D-L-I-E-TAP-AIPTK-E..ALKV--N-P-NH-- 157
SMM.SL.92.SL92B	A-P-LH-D-F-L-TW-A-Q-I-I-Q-V-D-L-V-I-A-D-N-T-A-K-N--RAATTTSSP-T-PLTASPSG..EET-D-M-TKIN-- 162	
SMM.US.x.PT583		L-V-LE-C-V-Q-R-D-L-V-I-A-N-T-T-A-T-GRATTTAKS-T-TTTTTVTPKVI--GN--KNNS-- 161
SMM.US.02.YNPRC_FWS		0
SMM.US.02.YNPRC_FAL		0
SMM.US.02.YNPRC_FBL		0
SMM.US.93.TNPRC_F102		0
SMM.US.79.NIRC_Cf1233		0
SMM.US.95.TNPRC_D175		0
SMM.US.79.NIRC_6007_G932		0
SMM.US.94.TNPRC_M927		0
SMM.US.97.TNPRC_G080		0
SMM.US.90.TNPRC_F100		0
SMM.US.80.NIRC_6001_G930		0
SMM.US.02.YNPRC_FVN		0
SMM.US.02.YNPRC_FYN		0
SMM.US.02.YNPRC_FUP		0
SMM.US.02.YNPRC_FIP		0
SMM.US.x.PT573		0
SMM.US.x.H445		L-V-LE-C-V-Q-V-K-R-D-L-V-I-A-N-T-A-T-GRATTTAKS-T-TTTTTVTPKVI--GD-KNNS-- 161
SMM.US.x.P209C15		L-V-AL-V-Q-X-K-V-X-R-D-L-V-I-A-N-T-A-T-GRATTTAKS-T-TTTTTVTPKVI--GD-KNNS-- 161
SMM.US.x.H9		LS-X-AS-V-Q-V-R-N-T-A-D-XX-GXP-AP-TTO--PPSP..IIAKV--DSDP-RSN-- 159
SMM.US.x.SME543		L-V-LE-C-V-Q-K-R-D-L-V-I-A-N-T-A-T-GRATTTAKS-T-TTTTTVTPKVI--GD-KNNS-- 161
SMM.US.x.PBJ_6P12		L-AS-V-Q-I-V-D-L-I-A-D-N-T-A-T-GTPAPITIQ-TTQA--PTSP..ITAKV--DSDP-KIN-- 164
SMM.US.x.P236_H4		L-V-LE-C-V-Q-K-V-D-L-I-A-D-N-T-A-T-GNA-GT-I-TA-P-VAENVI-SNP-KNNS-- 159
SMM.US.x.PBJ14_15		L-AS-V-Q-I-V-D-L-I-A-D-N-T-A-T-GTPAPITIQ-TTQA--PTSP..ITAKV--DSDP-KIN-- 164
SMM.US.x.PGM53		L-V-AL-V-Q-K-R-D-L-V-I-A-N-T-A-T-GTKNEIPT--TT-K-SPK-EA..ITAKV--SDP-SNN-- 164
SMM.US.x.SMP209		L-V-AL-V-Q-K-R-D-L-V-I-A-N-T-A-T-RGA...E--T-PAG-P-G...EKVI--GDP-RNNS-- 158
STM.US.x.STM	A-P-ACLT-Q-L-I-A-D-N-T-A-N-K-GKTVT.....TV-PT-AAATKPEL--VSNN-- 157	

gp120_end_gp41_start

Strain	Sequence	Position	
MAC.US.x.239	GNQTNITMSAEVAELRYLELGDYKLVETPIGLAPTVDVKRYTTGGTSRNKRGVFVLGFLGFLATAGSAMGAASLTLTAQSRLLAGIVQ00Q0LLDVVKRQ0EQLRLTVWGTKNLQTRVTAIEKYLKDQA0LNAWGCAFRQVCHTTPVW...	640	
H2A.GM.x.CBL23	--N-D--F--F--F--AE--SSTP.M--S--M--A--S--S--V-DT--E--M--	620	
H2A.CI.88.UC2F--I--F--SE--SSAP.A--L--S--I--A--S--V-D--R--M--	630	
H2A.DE.x.BEN	RTH--F--I--F--OR--SSTP.V--M--M--A--H--S--V-D--S-D-K-M--	625	
H2A.FR.x.96226	E-R--F--F--AQ--SSAP.V--M--M--A--S--V-D--Q-D-I-M--	560	
H2A.GH.x.GH1	S--F--I-V--F--RE--SSAP.V--S--M--A--S--D--S-D--M--	617	
H2A.FR.x.96330	E--F--I-V--F--SE--SSAP.G--M--T-A--S--M--A--S--V-DT--D-M--K--	491	
H2A.FR.x.96206	T-R--F--I--F--SE--SSAP.P.H--M--TV--S--M--A--S--V-D--N-E-M--	491	
H2A.GM.87.D194	H2A.GM.90.CBL24F--I-V--PF--KE--SSAP.V--T-A--G--S--M--A--S--V-D--D--M--	616
H2A.GM.90.CBL24	N--F--I-V--F--SE0--SSAP.K--T-A--S--M--A--S--V-D--Q-D--M--	627	
H2A.GW.x.AIF--F--SER--SSTP.R--M--T-A--S--M--A--S--V-N-K-D-D-M--	623	
H2A.GW.x.CAM1TE--F-A--F--TE--SSAH.G.Q--T-V-T--S--M--A--S--V-D--D-A-M--	626	
H2A.GW.86.FGD--F-A--F--S--SSAH.Q.HT--M--M--L--S--V-DT--D-M--	617	
H2A.GW.87.CAM2C6	NQT--F-D--F--SQ--SSAH.G.P--A--T-V-T--S--M--I-A--S--V-E--D-M--	627	
H2A.SN.x.ST	E--F--I-V--F--P--SSAP.V--T-A--S--M--A--S--V-DT--D-M--	621	
H2A.SN.85.ROD	W--F--I-V--F--KE--SSAH.G.HT--L--T-A--S--FR--M--A--S--V-D--D-D-M--	626	
H2A.GM.x.ISY	D-R--F--V--F--AE--SSAP.G.H--L--T-A--S--FR--M--A--S--V-DT--E-M--	618	
H2B.CI.x.EHO	K-L--V--S--K--F--EI--SSVT.P--L--M--T--S--V--A--S--V-E-K-D--M--Q--	624	
H2B.CI.88.UC1	DTK--G--G--F--EI--SSVT.P--M--L-M-T--S--L--S--ET--D-E-M--Q--	629	
H2B.FR.x.96200	D-A--SV--S--F--S--SSEK.P--M--M--T--S--L--S--DTFK-D--M--Q--	535	
H2B.FR.x.97227	NDTA--S--S--F--I--SSVT.P--L--M--T--S--L--S--DTFM-R-D-M--	566	
H2B.GH.86.D205	NST--SV--S--F--R--SSVK.P--M--M--T--S--PV--M--A--S--ET--N--M--Q--	628	
H2B.JP01.KR020	N-R--F--S--F--S--SSAT.P--M--M--T--S--L--S--KNF--N-D-M--Q--	617	
H2G.CI.x.ABT96	N-H--IF--S--F--N--SSVT.PK--M--T--S--M--A--S--DALGA-KT-E-Q--M--	645	
H2U.FR.96.12034	K--S--R--F--P--SSVT.P--S--M--S--I-DT--N-D-M--	635	
H2AB.CI.90.7312A	N-R--F--I-V--F--SE--SSTP.G.H--T-A--S--M--A--S--V-D--D-D-M--Q--	629	
MAC.US.x.BR5G--H--S--I--R--K--Y--E--R-D-K--	626	
MAC.US.x.270WK--I--N--S--I--R--K--Y--E--R-D-K--	623	
MAC.US.x.BR5K--I--N--S--I--R--K--Y--E--R-D-K--	629	
MAC.US.x.418N--N--D-T--	639	
MAC.US.x.BK28_H824S--N--V--S--D-D--	643	
MAC.US.x.MM142S--N--V--S--D-D--	642	
MAC.US.x.251_BK28S--N--V--S--D-D--	642	
MAC.US.x.SMM142BS--N--V--S--D-D--	642	
MNE.US.x.MNE027N--P--N--N--	643	
MNE.US.82.MNE_8N--P--N--N--	642	
SMM.SL.92.SL92B	N-E--F--M--R--ST-K--S--H--I--S--L--D-V-D-M--	647	
SMM.US.x.PT583	N-E--F--M--R--T-A--S--H--I--S--L--D-V-D-M--	648	
SMM.US.02.YNPRC_FWSX--X--R--S--D-V-D-M--	73	
SMM.US.02.YNPRC_FALX--X--R--S--D-V-D-M--	73	
SMM.US.02.YNPRC_FBLX--X--R--S--D-V-D-M--	73	
SMM.US.93.TNPRC_F102X--X--R--S--D-V-D-M--	73	
SMM.US.79.NIRC_CFI233X--X--R--S--D-V-D-M--	73	
SMM.US.95.TNPRC_D175X--X--R--S--D-V-D-M--	73	
SMM.US.79.NIRC_6007_G932X--X--R--S--D-V-D-M--	73	
SMM.US.94.TNPRC_M927X--X--R--S--D-V-D-M--	73	
SMM.US.97.TNPRC_G080X--X--R--S--D-V-D-M--	73	
SMM.US.90.TNPRC_F100X--X--R--S--D-V-D-M--	73	
SMM.US.80.NIRC_6001_G930X--X--R--S--D-V-D-M--	73	
SMM.US.02.YNPRC_FVNX--X--R--S--D-V-D-M--	73	
SMM.US.02.YNPRC_FYNX--X--R--S--D-V-D-M--	73	
SMM.US.02.YNPRC_FUPX--X--R--S--D-V-D-M--	73	
SMM.US.x.PT573X--X--R--S--D-V-D-M--	73	
SMM.US.x.H445X--X--R--S--D-V-D-M--	73	
SMM.US.x.P209C15X--X--R--S--D-V-D-M--	73	
SMM.US.x.H9X--X--R--S--D-V-D-M--	73	
SMM.US.x.SME543X--X--R--S--D-V-D-M--	73	
SMM.US.x.PBJ_6P12X--X--R--S--D-V-D-M--	73	
SMM.US.x.F236_H4X--X--R--S--D-V-D-M--	73	
SMM.US.x.PBJ14_15X--X--R--S--D-V-D-M--	73	
SMM.US.x.PGM53X--X--R--S--D-V-D-M--	73	
SMM.US.x.SMP209X--X--R--S--D-V-D-M--	73	
STM.US.x.STM	N-E--A--N--ST--T--S--T--D-V-D-M--	647	

Nef start	R17Y mutation	premature stop in Mac239	max HIV-1 similarity	
MAC.US.x.239	MGAIAIMRRSRPSGLRQRLRLRARGETYGRLLGEVDEGYSQSPGLDKGLSSLSLC	.EGOKYNOGYMNTPWNPAAERKELAYRKONDDIDE+DDD.LVGVSVR.PKVPLRMTYSYKLAIDMSHFKEKGGLEGITYSARRHRILDIYLEKEEGITPDWDQYTSVGPGR	165	
H2A.CL88.UC2	--ASG-KKL-KQ-QG-E	---GG-V-QCSASGGES---Q-SGREOK-P	0-0-R-D---T-M-GQ-ES-KQ---V-SD-N-Q---	T-SR---P-T---V-----MF-RD---L---V---N-H---V
H2A.DE.x.BEN	--ASG-KKL-KH-RG-E	---DG-KORDASGGE---FOEESGREON-P	0-0-E-S---T-Q-DL-Q0---V-SD---	I-P-T-R-R-E-T-----Q-MF-R-----N-H---V
H2A.DE.x.PE12	--ASG-KK-SLQG-E	---C-GWDSGAGE---LFOESGR-ONLP	0-0-R-DF---T-AG-GTL-KQ---V-ADN-N---	I-P-T-R-R-A-T---V-T-LN---D-M-F-----N-H---V
H2A.GH.x.GH1	--ASG-KKH-KH-OR-E	---H-GG-WOOCASGGE---LEFG-SG-QK-P	0-0-R-DF---T-I-GQ-KL-KQ---V-SD---	S-S---P-T-R-A-T---V-T---KR-D-M-F-RD---L---M-----N-H---V
H2A.GM.x.ISY	--ASG-KK---RG-QE	---GAC-G-WD-S-G---FHE-SGREOKLP	0-0-DF---T-T-K-ES-Q0---V-SD---	DT-SR---A-T-RM-V---DL-D---M---E-----N-H-L-V---166
H2A.GM.87.D194	--ASG-KK-EH-QG-E	---GG-VKORNASGGES---Q-SGREOK-P	0-0-EF---T-AIGO-NS-KQ---V-SD---	P-M-R-E-T-----F-RE---LF-----N-H-T---166
H2A.GW.x.ALI	--ASG-KK-G-LOG-E	---QTP-C-GOCSSGSG---Q-SGR-OKLP	0-0-DF---T-T-EL-KQ---V-LD-S---	T-R-Q---T---V---L-R---MF-E-----N-H---167
H2A.GW.x.MDS	--ASG-KK---RG-QE	---AGAC-GHWD-LGGE---R-OES-OR-P	0-0-DF---T-T-G-DS-Q0---V-SD---	Y-T-RT-A-T---V-----MF-E-----N-H-T---166
H2A.CL.x.IC763124	I-S-G-K-P-R-PGX-KK	---KKG-R-HRNR-KNCF---R-OESGR-NAP	0-0-KFT---K-T-E-KV-Q0---ANN-NIDMNN-QIK-H-S	R-T-R-KA-T---N---D---MF-PK-----N---G-HN-H-E
H2A.GW.87.CAM2CG	--ASG-KK---LOG-QE	---AG-C-EYNALD-GESLR-Q-S-REON	0-0-DF---A-A-GK-N-Q0---V-SD---	PAT-R-T---V-----LF-E-----A-N---V---166
H2A.PT.x.MP1	--ASG-K-P-SLOE-E	---F-A-C-GRCNKS-G-L-FHE-S-REOR-P	0-0-EF---GT-V-ELHKQ---V-LD-N-Q0-FP-T	P-T-F-M-V---L---LF-Q-E-L---V---N-H-T---166
H2A.PT.x.MP2	--ASG-K---SLOG-QE	---C-GRCN-S-G-L-FHE-S-REOR-P	0-0-EF---T-V-EL-KQ---V-LD-N-Q0-FP-T	P-T-F-M-V---L---LF-Q-K-L---VA---N-H-T---166
H2A.PT.x.EP	--ASG-KT---RG-E	---Q-C-S-GRW-GS-G-L-OE-SGREONLP	0-0-DF---T-TDE-RS-Q0---V-LD---	A-T-----T-R-V---E-N---M-D-----N-H---166
H2A.PT.x.794	--ASG-KK---QG-Q	---AG-C-EYWN-LGGE---L-OESGREOR-P	0-0-DF---D-A-G-NS-Q0---V-SD---	P-T-RTQ-E-T-R-----D---MF-E-----N-H---166
H2A.PT.x.1069	--ASG-KKH-K-LOG-E	---Q-G-C-GRSN-S-GEFL-Q-ESGREON-P	0-0-DF---T-G-RAM-KQ---V-ADN-N---	I-P-T-R-R-A-T---V-T-LN---D-M-F-----N-H---166
H2A.PT.x.1147	--ASG-KKP---PRG-QE	---AGAC-GHWEOL-G-L-OE-SGREOKLP	0-0-EF---T-T-G-LN-KQ---V-SD-Q-A-T	Q-A-T-RM---L-D---MF-EK-----N---N-H-V---166
H2A.PT.x.1227	--ASG-KKQ-K-RG-E	---S-EGWR-LGGE-L-OEESGREOR-R	RR-0-EF---KD-P-GKEN-Q0---V-SD---	P-T-R-E-T-RR---D---MF-E-K-QN---N-H---166
H2A.PT.x.1215	--ASG-KK---QG-Q	---AGAC-GYWN-LGGE---L-OESGREOK-P	0-0-DF---T-A-G-K-Q0---V-SD-E-Q---	T-R-Q-G-T-----MF-D-K-----V---N-H---166
H2A.PT.x.1320	--ASG-K---K-QG-Q	---AGAC-GHWD-SGVY-R-OESGREOK-P	0-0-DF---T-T-RD-KQ---V-SD-Q---	T-R-G-T---V-----MF-E-K-----V---N-H-V---166
H2A.PT.x.483	--ASG-KKP---PRG-QE	---AG-C-GYWE-LGEE-L-OE-SGR-OK-P	0-0-DF---A-I-G-RS-KQ---V-SD-Q-P-T	R-A-THR---L-D---MF-E-----N-H---166
H2A.PT.x.546	--ASG-KK-K-QG-QE	---QTH-GAC-GHCNGS-G-FL-Q-ESGR-ON-P	0-0-DF---T-A-AL-Q0---V-SD-Q-P-T	R-P-T-R---D---MF-CE-----N-H---166
H2A.PT.x.1543	--ASG-KK---LOG-QE	---G-C-GOCSSG-G-L-OE-SG-EON-P	0-0-DF---T-A-K-NOL-KQ---V-SD-Q-P-T	Q-Q-T---D---MF-E-----F-K---N-H---166
H2A.PT.x.B1_1	--ASG-KK-K-RG-QE	---AGAC-GHWD-LGGK-OEESGREOK-P	0-0-DF---H-A-G-NS-Q0---V-SD---	T-T-Q-E-T-R-V---D---MF-E-----A-N-H-V---166
H2A.PT.x.1544	--ASG-KKH-K-RG-Q	---AGAC-GHWD-LGGE-LRFOE-SGREOK-P	0-0-DF---T-K-T-E-Q0---V-SD-Q-H-K	R-E-TT-----MF-E-K-T-N---N-H---166
H2A.PT.x.1428	--ASG-KT---RG-QE	---AGAC-GHW-LG-GE-FO-ESGREOK-P	0-0A-DF---T-T-G-NP-KQ---V-SD-Q-P-T	R-R-E-T-R---V-----MF-E-L---V---N-AH-T---166
H2A.PT.x.1567	--ASG-KKQ---RG-E	---T-EG-C-GSHGS-G-L-Q-ESGR-Q-P	0-0-DF---T-A-G-AL-Q0---V-SD-I-P	R-Q-T-F-M-V-----MF-E-----A-N-H-T---166
H2A.PT.x.268	--ASG-KKP---RG-QE	---AG-C-GYWO-LGGE-R-OEESGREOR-P	0-0-DF---KD-SAV-KENS-Q0---V-SD---	T-R-E-T-RR-V---L-D---MF-E-----QN---N-H---166
H2A.PT.x.1378	--ASG-KK-T-LRG-QE	---QTG-RDLWGRCRNSGE-S-RFO-ESGREON-P	0-0-P-DFV-Q-T-P-G-SKL-Q0---A-AD---	R-R-A-T-RM---L-QE-MF-E-----M-N-H-V---166
H2A.SN.x.ST	--ASG-KK-E-RG-E	---QTP-AS-GHWDKLGGE-L-OE-SGR-OK-P	RR-0-DF---A-G-GS-KQ---V-SD-P-T	R-E-T-R-R---L---L-D-R-V---G-N-H-V---165
H2A.SN.85.ROD	--ASG-KKH---PRG-QE	---AGAC-GYWN-SGGE-RFOE-S-REOK-P	0-0-DF---KD-A-NL-Q0---V-SD-Q-R-T	P-THR---L-TR---MF-E-K-N---A-N-H-V---166
H2A.DE.92.NEP	--ASG-KTH---RG-Q	---AG-C-GHWD-LGGE-R-OEES-OK-P	0-0-DF---T-S-G-N-Q0---V-LD---	T-A-THR-RN---L-D---MF-E-----M-N-A---160
H2A.DE.91.HOM	--ASG-KKP-K-RG-E	---QTP-S-GHWRHLGGE-L-OESGREON-P	RR-0-DF---A---XX-DL-KQ---V-SD-N-IE	P-T-RTQ-V-THR-R---L-D---MF-E-----M---V---N-H-V---166
H2B.CL.x.EHO	--S-G-KKQ-KOORP-E	---RGRP-ESS-RROERSL-Y-S-N-N-P	D-TLGAE	GGG-DS-E-DE-NE-R-G-P-TF-V-V-L-E-E-F-E-K-T-N-VSG-N-H-V
H2B.CL88.UC1	--S-G-KKQ-KOORG-E	---TOE-P-K-SEGQRKOS-R-S-D-N-P	RNAPRAE	GGGQ-DT-DS-DE-NE-Y-NR-S-T-M-----E-K-T-N-VSG-N-Y-----148
H2B.CL.x.IC762993	--ES-S-KKQ-KORRE-EK-QTOEGP-RK-SE-QOQSLR-EKS-NSK-P	---KKNRPRAE	VMNSNKNK-KAY-K-NR-KPIT	KE-IKK-X-NT-FKNKK-VSK-LN-H-----148
H2B.GH.86.D205	--S-G-KK-ERQOG-EK-VPERP	---S-RREOS-R-ES-D-N-P	NARGAE	GGGQ-DA-ESDE-E-AICKT-I-P-T-V-Q-M-E-T-F-N-VSG-N-H-----149
H2G.CL.x.ABT96	--S-G-KK-APQQLG-K-Q	---P-K-WEGL-E-YREESG	Q-T-F	T-A-X-N-V-SD-N-P-Y-V-----E-M-Q---N-A-----166
H2U.FR.96.12034	--S-KKH-KR-K-E	---A-N-W-GL-FELYQ-SG-D-RF-EAQ	G-SE-EF	T-TGKD-Q-KQ-V-E-N-A-R-A-T-V-D-F-Q-----N-N-T---168
H2AB.CL.90.7312A	--S-G-KKQ-KOORG-E	---Q-GP-S-GRGES-EES-R-N-P	TAPGAE	GGGQ-DIEE-DE-NE-Y-Q-P-T-----F-E-----T-F-N-R-VGG-N-Y-----148
MAC.US.x.251_1A11	--T-----ST	---E-S-L-E	E	E-----E-----166
MAC.US.x.251_BK28	---K-A	---K-L-G-R	K	K-----A-T-----M-----166
MAC.US.x.MMI42	---KK-K-PR	---N-FKG-S-L	E	E-----K-P-I-E-IP-E-AR-----I-H-----165
MAC.US.x.251_32H_P15	---R-KSA	---S-L-G-R	E	E-----V-E-P-M-R-T-V-----M-----166
MAC.US.x.BK28_H824	---R-A	---S-L-G-R	E	E-----E-----M-----N-----166
MNE.US.82.MNE_8	---T-K-KSP	---WE-L-L-S-S-L	F-K-G-I-E-N-P	R-R-II-V-E-E-K-M-M-N-P-----166
MNE.US.x.MNE027	---T-K-KSP	---WE-L-L-S-S-L	F-K-G-I-E-N-P	R-R-A-T-V-E-E-K-M-M-N-P-----166
SMM.US.x.H9	---VT-KKORXAG-N-XE-Q	---WEGL-GE-XDASG	P-CE-F-T-GA-G-Q-V-NE	CP-S-R-V-I-T-V-I-K-X-X-----N-A-----164
SMM.US.x.PBJC	---VA-KKORKHG-N-E-Q	---WEGL-GE-XDASG	P-CE-F-TK-A-D-Q-V-SA	CP-S-R-V-I-T-V-I-K-L-----N-A-----164
SMM.US.x.PBJ_143	---VT-KKORXAG-N-XE-Q	---WEGL-GE-XDASG	P-CE-F-T-A-D-Q-V-SA	CP-S-R-V-I-T-V-I-K-X-X-----N-A-----164
SMM.US.x.PBJD	---VT-KKORKHG-N-E-Q	---WEGL-GE-XDASG	P-CE-F-TK-A-D-Q-V-SA	CP-S-R-V-I-T-V-I-K-L-----N-A-----164
SMM.US.x.PBIE	---VT-KKOR-RG-N-YE-Q	---WEGL-GE-XDASG	P-CE-F-T-A-D-Q-V-SA	CP-S-R-V-I-T-D-K-L-----N-A-----164
SMM.US.x.PBJ14_15	---VT-KKOR-RG-N-YE-Q	---WEGL-GE-XDASG	P-CE-F-T-A-D-Q-V-SA	CP-S-R-V-I-T-D-K-L-----N-A-----164
SMM.US.x.PBIA	---VT-KKOR-RG-N-YE-Q	---WEGL-GE-XDASG	P-CE-F-T-A-D-Q-V-SA	CP-S-R-V-I-T-D-K-L-----N-A-----164
SMM.US.x.PBJ_6P2	---VT-KKOR-RG-N-YE-Q	---WEGL-GE-XDASG	P-CE-F-T-A-D-Q-V-SA	CP-S-R-V-I-T-D-K-L-----N-A-----164
SMM.US.x.PBJ_6P12	---VT-KKOR-RG-N-YE-Q	---WEGL-GE-XDASG	P-CE-F-T-A-D-Q-V-SA	CP-S-R-V-I-T-D-K-L-----N-A-----164
SMM.US.x.PBJ_6P21	---VT-KKOR-RG-N-YE-Q	---WEGL-GE-XDASG	P-CE-F-T-A-D-Q-V-SA	CP-S-R-V-I-T-D-K-L-----N-A-----164
SMM.US.x.PBJ_6P12	---VT-KKOR-RG-N-YE-Q	---WEGL-GE-XDASG	P-CE-F-T-A-D-Q-V-SA	CP-S-R-V-I-T-D-K-L-----N-A-----164
SMM.US.x.SME543	---KKQY-RG-N-E	---WEGL-E-L-ASG	P-SE-TA-KA-G-KQ-V-DE	H-A-T-NEK-M-M-----N-L-T---166
MAC.US.x.17EC1	---S	---K	Q	I-----I-----K-----166
MAC.US.x.17EFR	---S	---K	Q	I-----I-----K-----166
SMM.US.x.PGM53	---A-G-KKQ-OR-G-GEK-Q	---H-K-WEGL-R-E-GRDWNH-L-F	G-SE-F-R-Q-K-Q-V-DD-E-I-H-A	A-----D-E-----T-N-----166
SMM.SL.92.SL92B	---IT-SKPA-KH-KK-S-DS	---WDGL-E-SQ-Q-S-RA-N-CLT-RD-Y	V-EF-T-Q-QQ-C-D-E-A-V-W	A-----D-----N-----167
SMM.US.x.F236_H4	---KKQYKRG-N-E-Q	---WEGL-E-L-ASG	P-SE-T-A-G-Q-V-DE-I-H-R	A-T-----NE-M-----N-----166
STM.US.x.STM	---ASG-KKQKQH-E-E	---K-EGLGE-SGP-Q-AS-N-H	P-R-E-F-K-A-SA-E-Q-V-E-N	A-H-R-E-T-L-S-E-M-V-N-A-----166

