

# IV

---

## HIV-1/SIVcpz proteins

---

Introduction .....	571
Table of HIV-1/SIVcpz protein alignments .....	574
Gag .....	580
Pol .....	590
Vif .....	608
Vpr .....	612
Tat .....	614
Rev .....	616
Vpu .....	618
Env .....	620
Nef .....	636

### Construction of HIV-1/SIVcpz protein alignments

The number of full-length gene sequences is still growing rapidly for all genes. The envelope master alignment now contains 307 full-length sequences. For the purposes of the printed alignments, we have had to limit the number of sequences dramatically. Here we list the criteria we have followed to make the selection.

First, we have decided to end the supremacy of the B clade sequences. More than half (198, to be precise) of the full-length envelope sequences are still subtype B, though the contribution of other subtypes is increasing. We have tried to balance the number of representatives of all subtypes in these alignments. For this, we had to make a heavy selection on subtype B sequences. We have tried to include as many “classical” sequences as possible. A lot of follow-up work has been done based on lab strains such as HXB2, MN, SF2, and JR-CSF/JR-FL, so these strains are included in the alignments. Furthermore, within subtype B we have tried to represent sequences from diverse geographical origins, so as to represent a broad spectrum of variants. In the case of subtype B, this means that we have included African, Asian and Brazilian variants along with the “Western” strains. For sequences from non-B subtypes, we have selected a few representative sequences from each dataset, again with an eye on maintaining geographical diversity. When possible we have left all representatives of group O in the alignment, as these sequences are much more genetically diverse than any single M-group subtype.

---



---

 Explanation of Symbols in Alignments
 

---

Symbol	Meaning
--------	---------

---

**Alignment symbols**

? in consensus	no majority-rule consensus could be determined at this position
x	nucleotide missing from codon
#	frameshift, or codon contains N or illegal character
\$	stop codon

**Annotation symbols**

-	domain boundaries
/	protein start point
\	protein end point
∨	splice site or exon join
->	start of overlapping coding region
<-	end of overlapping coding region
*	cysteine
^^^ [NxS, NxT]	glycosylation site
^*^ [NCS, NCT]	glycosylation site with cysteine
CD4	residue critical for CD4 binding
cds	coding sequence (indicates regions where two proteins overlap; the overlapping proteins use two different reading frames)
MHR	major homology region
nls	nuclear localization signal
phos site	phosphorylation site
PKC	protein kinase C binding
Zn-motif	Zinc finger binding motif

---



---

Name	Accession	Region	Author	Reference
A.BY.97BL006	AF193275	VPR	Liitsola, K	<i>ARHR</i> <b>16</b> (11):1047–53 (2000)
A.CD.97DC-KMST91	AJ401040	ENV	Vidal, N	<i>J Virol</i> <b>74</b> (22):10498–10507 (2000)
A.CD.HIM401034	AJ401034	ENV	Vidal, N	<i>J Virol</i> <b>74</b> (22):10498–10507 (2000)
A.FI.FIN91121	AF219261	VPU	Liitsola, K	<i>Scand J Infect Dis</i> <b>32</b> (5):475–480 (2000)
A.FI.FIN9199	AF219265	ENV	Liitsola, K	<i>Scand J Infect Dis</i> <b>32</b> (5):475–480 (2000)
A.FR.HIV232956	AJ232956	NEF	Jubier-Maurin, V	<i>ARHR</i> <b>15</b> (1):23–32 (1999)
A.KE.K89	L22943	ENV	Louwagie, J	<i>J Virol</i> <b>69</b> (1):263–271 (1995)
A.RW.PVPI	L07082	ENV	Bex, F	Unpublished (1992)
A.SE.SE7535	AF069671	GAG POL REV TAT VIF VPR	Laukkanen, T	<i>AIDS</i> <b>13</b> (14):1819–1826 (1999)
A.SE.SE8538	AF069669	GAG NEF POL TAT VPR	Laukkanen, T	<i>AIDS</i> <b>13</b> (14):1819–1826 (1999)
A.SE.SE8891	AF069673	GAG NEF VPR	Laukkanen, T	<i>AIDS</i> <b>13</b> (14):1819–1826 (1999)
A.SE.UGSE8131	AF107771	GAG NEF POL REV VIF VPR VPU	Laukkanen, T	<i>AIDS</i> <b>13</b> (14):1819–1826 (1999)
A.UA.ukr970063	AF082486	ENV	Liitsola, K	<i>AIDS</i> <b>12</b> (14):1907–1919 (1998)
A.UG.U13-2	X91354	VIF	Wieland, U	<i>J Gen Virol</i> <b>78</b> :393–400 (1997)
A.UG.UG275A	L22951	TAT	Louwagie, J	<i>J Virol</i> <b>69</b> (1):263–271 (1995)
A1.KE.Q23-17	AF004885	ENV GAG NEF POL REV TAT VIF VPR VPU	Poss, M	<i>J Virol</i> <b>72</b> (10):8240–51 (1998)
A1.SE.SE7253	AF069670	GAG POL REV TAT VIF VPR VPU	Laukkanen, T	<i>AIDS</i> <b>13</b> (14):1819–1826 (1999)
A1.UG.92UG037	U51190	ENV GAG NEF POL REV TAT VIF VPR VPU	Gao, F	<i>J Virol</i> <b>70</b> (3):1651–1657 (1996)
A1.UG.U455	M62320	ENV GAG NEF POL REV TAT VPR VPU	Oram, JD	<i>ARHR</i> <b>6</b> (9):1073–1078 (1990)
A2.CD.97CDKFE4	AF286240	POL VIF	Gao, F	<i>ARHR</i> <b>17</b> (8):675–688 (2001)
A2.CD.97CDKS10	AF286241	ENV REV TAT VIF VPU	Gao, F	<i>ARHR</i> <b>17</b> (8):675–88 (2001)
A2.CD.97CDKTB48	AF286238	ENV GAG NEF REV TAT VIF VPU	Gao, F	<i>ARHR</i> <b>17</b> (8):675–688 (2001)
A2.CY.94CY017-41	AF286237	ENV GAG NEF POL REV TAT VIF VPU	Gao, F	<i>ARHR</i> <b>17</b> (8):675–688 (2001)
B.AU.1052D1	AF133402	VPR	Naif, HM	<i>J Virol</i> <b>73</b> (6):4866–4881 (1999)
B.CN.RL42	U71182	ENV GAG NEF POL REV TAT VIF VPR VPU	Graf, M	<i>ARHR</i> <b>14</b> (3):285–288 (1998)
B.DE.D31	U43096	GAG NEF POL TAT	Kreutz, R	<i>ARHR</i> <b>8</b> (9):1619–1629 (1992)
B.ES.89SP061	AJ006287	GAG NEF TAT	Olivares, I	<i>ARHR</i> <b>14</b> (18):1649–1651 (1998)
B.FR.HXB2	K03455	ENV GAG NEF POL REV TAT VIF VPR VPU	Wong-Staal, F	<i>Nature</i> <b>313</b> (6000):277–284 (1985)
B.GA.OYI	M26727	ENV GAG NEF POL REV TAT VIF VPR VPU	Huet, T	<i>AIDS</i> <b>3</b> (11):707–715 (1989)
B.GB.CAM1	DI0112	POL VPR	McIntosh, AAG	PhD dissertation, University of Cambridge (1989)
B.IN.AF316095	AF316095	VPR	Halani, N	<i>ARHR</i> <b>17</b> (7):637–642 (2001)
B.IN.IND-P35	Y15122	NEF	Ahmad, KM	<i>ARHR</i> <b>14</b> (16):1491–1493 (1998)
B.IT.B-IT-R5	AF147737	NEF	Catucci, M	<i>J Med Virol</i> <b>60</b> (3):294–299 (2000)
B.JP.JH31	M21137	GAG	Komiyama, N	<i>ARHR</i> <b>5</b> :411–19 (1989)
B.JP.JH32	M21138	ENV	Komiyama, N	<i>ARHR</i> <b>5</b> :411–19 (1989)
B.JP.KK51-1	D70887	VIF	Tominaga, K	<i>ARHR</i> <b>12</b> (16):1543–1549 (1996)
B.JP.PT1-01	AB034578	VPU	Yamada, T	<i>Arch Virol</i> <b>145</b> (5):1021–7 (2000)
B.JP.PT7-2	AB034558	VPR	Yamada, T	<i>Arch Virol</i> <b>145</b> (5):1021–7 (2000)
B.KR.WK	AF224507	GAG NEF POL REV TAT VIF VPR VPU	Cho, YK	Unpublished
B.MM.MCSW12	AB043906	VPR	Motomura, K	<i>ARHR</i> <b>16</b> (17):1831–1843 (2000)
B.TH.93TH067	U39258	ENV	Penny, MA	<i>ARHR</i> <b>12</b> (8):741–747 (1996)
B.TH.AF082839	AF082839	NEF	Vallejo, A	<i>AIDS</i> <b>13</b> (4):532–4 (1999)

B.TW.TWCYS	AF086817	GAG NEF POL REV TAT VIF VPR VPU	Huang, LM	Unpublished
B.US.JRCSF	M38429	VPR VPU	O'Brien, WA	<i>Nature</i> <b>348</b> :69–73 (1990)
B.US.JRFL	U63632	ENV GAG NEF POL REV TAT VIF	O'Brien, WA	<i>Nature</i> <b>348</b> :69–73 (1990)
B.US.MNCG	M17449	ENV GAG NEF POL REV TAT VIF VPR VPU	Gurgo, C	<i>Virology</i> <b>164</b> (2):531–536 (1988)
B.US.RF	M17451	ENV GAG NEF POL REV TAT VIF VPU	Starcich, BR	<i>Cell</i> <b>45</b> (5):637–648 (1986)
B.US.SF2	K02007	ENV GAG NEF POL REV VIF VPR VPU	Sanchez- Pescador, R.	<i>Science</i> <b>227</b> (4686):484–492 (1985)
C.BR.92BR025	U52953	GAG NEF POL REV TAT VIF VPR VPU	Gao, F	<i>J Virol</i> <b>70</b> (3):1651–1667 (1996)
C.BR.98BR004	AF286228	ENV NEF POL TAT	Rodenburg, CM	<i>ARHR</i> <b>17</b> (2):161–168 (2001)
C.BW.96BW01B03	AF110959	POL	Novitsky, VA	<i>J Virol</i> <b>73</b> (5):4427–32 (1999)
C.BW.96BW01B21	AF110960	REV VIF VPR	Novitsky, VA	<i>J Virol</i> <b>73</b> (5):4427–32 (1999)
C.BW.96BW01B22	AF110961	GAG TAT	Novitsky, VA	<i>J Virol</i> <b>73</b> (5):4427–32 (1999)
C.BW.96BW0402	AF110962	ENV NEF POL TAT VPU	Novitsky, VA	<i>J Virol</i> <b>73</b> (5):4427–32 (1999)
C.DJ.DJ373A	L23065	REV TAT	Louwagie, J	<i>J Virol</i> <b>69</b> (1):263–271 (1995)
C.ET.ETH2220	U46016	ENV GAG NEF POL REV TAT VIF VPR VPU	Salminen, MO	<i>ARHR</i> <b>12</b> (14):1329–1339 (1996)
C.FI.FIN9126	AF219262	ENV	Liitsola, K	<i>Scand J Infect Dis</i> <b>32</b> (5):475–480 (2000)
C.IL.98IS002	AF286233	ENV GAG NEF POL REV TAT VIF VPR VPU	Rodenburg, CM	<i>ARHR</i> <b>17</b> (2):161–168 (2001)
C.IN.93IN101	AB023804	ENV GAG NEF REV TAT VPU	Mochizuki, N	<i>ARHR</i> <b>15</b> (14):1321–1324 (1999)
C.IN.93IN999	AF067154	POL VPR	Lole, KS	<i>J Virol</i> <b>73</b> (1):152–160 (1999)
C.IN.94IN11246	AF067159	VIF VPR	Lole, KS	<i>J Virol</i> <b>73</b> (1):152–160 (1999)
C.IN.94IN476	AF286223	NEF	Rodenburg, CM	<i>ARHR</i> <b>17</b> (2):161–168 (2001)
C.IN.95IN21068	AF067155	GAG POL REV TAT VIF VPR	Lole, KS	<i>J Virol</i> <b>73</b> (1):152–160 (1999)
C.IN.98IN012	AF286231	GAG NEF POL REV TAT VPU	Rodenburg, CM	<i>ARHR</i> <b>17</b> (2):161–168 (2001)
C.IN.98IN022	AF286232	VIF	Rodenburg, CM	<i>ARHR</i> <b>17</b> (2):161–168 (2001)
C.IN.IND-P728-2	Y15117	NEF	Ahmad, KM	<i>ARHR</i> <b>14</b> (16):1491–1493 (1998)
C.MY.AB043903	AB043903	VPU	Motomura, K	<i>ARHR</i> <b>16</b> (17):1831–1843 (2000)
C.SO.SO145A	L22946	ENV REV VPU	Louwagie, J	<i>J Virol</i> <b>69</b> (1):263–271 (1995)
C.TZ.98TZ013	AF286234	ENV GAG NEF POL REV TAT VIF	Rodenburg, CM	<i>ARHR</i> <b>17</b> (2):161–168 (2001)
C.TZ.98TZ017	AF286235	GAG NEF POL TAT VPR VPU	Rodenburg, CM	<i>ARHR</i> <b>17</b> (2):161–168 (2001)
C.UG.UG268A2	L22948	ENV REV VPU	Louwagie, J	<i>J Virol</i> <b>69</b> (1):263–271 (1995)
C.ZA.97ZA012	AF286227	GAG NEF POL REV TAT VPR VPU	Rodenburg, CM	<i>ARHR</i> <b>17</b> (2):161–168 (2001)
C.ZA.AF325763	AF325763	VIF	Scriba, TJ	<i>ARHR</i> <b>17</b> (8):775–81 (2001)
C.ZM.96ZM651	AF286224	GAG NEF POL REV TAT VPR	Rodenburg, CM	<i>ARHR</i> <b>17</b> (2):161–168 (2001)
C.ZM.96ZM751	AF286225	GAG NEF POL VIF	Rodenburg, CM	<i>ARHR</i> <b>17</b> (2):161–168 (2001)
D.CD.84ZR085	U88822	ENV GAG NEF POL REV TAT VIF VPR VPU	Gao, F	<i>J Virol</i> <b>72</b> (7):5680–98 (1998)
D.CD.ELI	K03454	ENV GAG NEF POL REV TAT VIF VPR VPU	Alizon, M	<i>Cell</i> <b>46</b> (1):63–74 (1986)
D.CD.JY1	J03653	ENV	Yourno, J	<i>ARHR</i> <b>4</b> :165–173 (1988)
D.CD.NDK	M27323	ENV GAG NEF POL REV TAT VIF VPR VPU	Spire, B	<i>Gene</i> <b>81</b> :275–284 (1989)
D.CD.Z2Z6	M22639	GAG POL REV TAT VIF VPR	Srinivasan, A	<i>Gene</i> <b>52</b> :71–82 (1987)
D.CI.CI13	AJ277820	ENV	Beirnaert, E	<i>Virology</i> <b>281</b> (2):305–14 (2001)

D.DE.K23DE	AF233690	VIF	Kuhn, J	Unpublished
D.FI.FIN93167	AF219271	VPU	Liitsola, K	<i>Scand J Infect Dis</i> <b>32</b> (5):475–480 (2000)
D.FI.FIN93178	AF219272	ENV REV VPU	Liitsola, K	<i>Scand J Infect Dis</i> <b>32</b> (5):475–480 (2000)
D.JP.PT14-5	AB034542	VPR	Yamada, T	<i>Arch Virol</i> <b>145</b> (5):1021–7 (2000)
D.SN.SE365A2	L22945	ENV REV TAT VPU	Louwagie, J	<i>J Virol</i> <b>69</b> (1):263–271 (1995)
D.TZ.87TZ4622	U65075	ENV	Robbins, KE	<i>ARHR</i> <b>12</b> (14):1389–1391 (1996)
D.TZ.TZ005	U12406	VPU	Siwka, W	<i>ARHR</i> <b>10</b> (12):1753–1754 (1994)
D.UG.94UG1141	U88824	GAG NEF POL REV TAT VIF VPR VPU	Gao, F	<i>J Virol</i> <b>72</b> (7):5680–98 (1998)
D.UG.U18-0	X91355	VIF	Wieland, U	<i>J Gen Virol</i> <b>78</b> :393–400 (1997)
D.UG.U25-6	X91361	VIF	Wieland, U	<i>J Gen Virol</i> <b>78</b> :393–400 (1997)
D.UG.U36-0	X91363	VIF	Wieland, U	<i>J Gen Virol</i> <b>78</b> :393–400 (1997)
D.UG.UG266A2	L22947	VPU	Louwagie, J	<i>J Virol</i> <b>69</b> (1):263–271 (1995)
D.UG.UG269A	L22949	REV	Louwagie, J	<i>J Virol</i> <b>69</b> (1):263–271 (1995)
D.UG.UG274A2	L22950	REV TAT	Louwagie, J	<i>J Virol</i> <b>69</b> (1):263–271 (1995)
D.UG.WHO15-474	U36886	ENV	Douglas, NW	<i>AIDS</i> <b>10</b> (1):39–46 (1996)
F1.BE.VI850	AF077336	ENV GAG NEF POL REV TAT VPR VPU	Laukkanen, T	<i>Virology</i> <b>269</b> (1):95–104 (2000)
F1.BR.93BR020-1	AF005494	ENV NEF POL REV TAT VIF VPR VPU	Gao, F	<i>J Virol</i> <b>72</b> (7):5680–98 (1998)
F1.BR.BZ162	L11751	GAG	Louwagie, JJ	<i>AIDS</i> <b>7</b> :769–780 (1993)
F1.CD.VII74	L11782	GAG	Louwagie, JJ	<i>AIDS</i> <b>7</b> :769–780 (1993)
F1.DE.AF200475	AF200475	VIF	Kuhn, J	Unpublished
F1.FI.FIN9363	AF075703	NEF POL REV TAT	Laukkanen, T	<i>Virology</i> <b>269</b> (1):95–104 (2000)
F1.FR.MP411	AJ249238	ENV NEF POL REV TAT VIF VPR VPU	Triques, K	<i>ARHR</i> <b>16</b> (2):139–151 (2000)
F1.RW.VI69	L11796	GAG	Louwagie, JJ	<i>AIDS</i> <b>7</b> :769–780 (1993)
F2.CM.CA20	AJ277824	ENV	Nyambi, PN	<i>J Virol</i> <b>70</b> (9):6235–6243 (1996)
F2.CM.CM53657	AF377956	ENV NEF POL REV TAT VIF VPR VPU	Carr, JK	<i>Virology</i> <b>286</b> (1):168–181 (2001)
F2.CM.HIV232985	AJ232985	NEF	Jubier-Maurin, V	<i>ARHR</i> <b>15</b> (1):23–32 (1999)
F2.CM.MP255	AJ249236	GAG POL REV TAT VIF VPR VPU	Triques, K	<i>ARHR</i> <b>16</b> (2):139–151 (2000)
F2.CM.MP257	AJ232986	NEF	Jubier-Maurin, V	<i>ARHR</i> <b>15</b> (1):23–32 (1999)
F2.CM.MP257	AJ249237	ENV GAG POL REV TAT VIF VPR VPU	Triques, K	<i>ARHR</i> <b>16</b> (2):139–151 (2000)
G.BE.DRCBL	AF084936	ENV GAG POL REV TAT VIF VPR VPU	Debyser, Z	<i>ARHR</i> <b>14</b> (5):453–459 (1998)
G.CG.CNG30	AF056186	VIF VPR VPU	Harada, Y	<i>ARHR</i> <b>18</b> (1):79–83(2002)
G.FI.HH8793-1-1	AF061640	GAG POL REV VIF	Salminen, MO	<i>ARHR</i> <b>8</b> (9):1733–1742 (1992)
G.FI.HH8793-12-1	AF061641	TAT	Salminen, MO	<i>ARHR</i> <b>8</b> (9):1733–1742 (1992)
G.ML.HIV232990	AJ232990	NEF	Jubier-Maurin, V	<i>ARHR</i> <b>15</b> (1):23–32 (1999)
G.NG.92NG083	U88826	ENV GAG NEF POL REV TAT VIF VPR VPU	Gao, F	<i>J Virol</i> <b>72</b> (7):5680–98 (1998)
G.SE.SE6165	AF061642	ENV GAG NEF POL REV TAT	Laukkanen, T	<i>Virology</i> <b>247</b> (1):22–31 (1998)
G.TW.TWG1	AF220486	VPU	Lee, CN	<i>J Clin Microbiol</i> <b>38</b> (7):2468–2474 (2000)
H.BE.VI991	AF190127	ENV GAG NEF POL REV TAT VIF VPR VPU	Janssens, W	<i>AIDS</i> <b>14</b> (11):1533–43 (2000)
H.BE.VI997	AF190128	ENV GAG NEF POL REV TAT VIF VPR VPU	Janssens, W	<i>AIDS</i> <b>14</b> (11):1533–43 (2000)
H.CF.90CF056	AF005496	ENV GAG NEF POL REV TAT VIF VPR VPU	Murphy, E	<i>ARHR</i> <b>9</b> (10):997–1006 (1993)
J.SE.SE7022	AF082395	ENV GAG NEF POL REV TAT VIF VPR VPU	Laukkanen, T	<i>ARHR</i> <b>15</b> (3):293–297 (1999)
J.SE.SE7887	AF082394	ENV GAG NEF POL REV TAT VIF VPR VPU	Laukkanen, T	<i>ARHR</i> <b>15</b> (3):293–297 (1999)

K.BE.VI325	L11789	GAG	Louwagie, JJ	<i>AIDS</i> 7:769–780 (1993)
K.CD.EQTB11C	AJ249235	ENV GAG NEF POL REV TAT VIF VPR VPU	Triques, K	<i>ARHR</i> 16(2):139–151 (2000)
K.CM.MP535	AJ249239	ENV GAG NEF POL REV TAT VIF VPR VPU	Triques, K	<i>ARHR</i> 16(2):139–151 (2000)
01_AE.CF.90CF11697	AF197340	ENV GAG NEF POL REV TAT VIF VPR VPU	Anderson, JP	<i>J Virol</i> 74(22):10752–10765 (2000)
01_AE.CF.90CF402	U51188	ENV GAG NEF POL REV TAT VIF VPR VPU	Gao, F	<i>J Virol</i> 70(10):7013–7029 (1996)
01_AE.CF.90CF4071	AF197341	ENV GAG NEF POL REV TAT VIF VPR VPU	Anderson, JP	<i>J Virol</i> 74(22):10752–10765 (2000)
01_AE.CM.CA10	AJ277818	ENV	Beirnaert, E	<i>Virology</i> 281(2):305–14 (2001)
01_AE.CN.97CNGX11F	AY008718	REV TAT VIF VPR VPU	Piyasirisilp, S	<i>J Virol</i> 74(23):11286–11295 (2000)
01_AE.CN.97CNGX2F	AY008714	ENV GAG NEF POL TAT VPR VPU	Piyasirisilp, S	<i>J Virol</i> 74(23):11286–11295 (2000)
01_AE.DE.K08DE	AF215859	VIF	Kuhn, J	Unpublished
01_AE.FI.FIN9379	AF219273	REV	Liitsola, K	<i>Scand J Infect Dis</i> 32(5):475–480 (2000)
01_AE.FR.HIV232982	AJ232982	NEF	Jubier-Maurin, V	<i>ARHR</i> 15(1):23–32 (1999)
01_AE.JP.93JPNH1	AB052995	ENV GAG NEF POL REV TAT VIF VPR	Sato, H	<i>J Virol</i> 75(12):5604–5613 (2001)
01_AE.TH.93TH057	AF197338	GAG NEF	Anderson, JP	<i>J Virol</i> 74(22):10752–10765 (2000)
01_AE.TH.93TH253	U51189	POL REV TAT	Gao, F	<i>J Virol</i> 70(10):7013–7029 (1996)
01_AE.TH.93TH902	AF170549	VPR	Chang, SY	<i>ARHR</i> 15(17):1591–1596 (1999)
01_AE.TH.93TH9021	AF164485	ENV VPU	Chang, SY	<i>ARHR</i> 15(18):1703–1706 (1999)
01_AE.TH.94TH702	AF170545	POL VPR	Chang, SY	<i>ARHR</i> 15(17):1591–1596 (1999)
01_AE.TH.94TH7091	AF170546	GAG	Chang, SY	<i>ARHR</i> 15(17):1591–1596 (1999)
01_AE.TH.95TH022	AB032740	ENV NEF POL VPR VPU	Auwanit, W	<i>ARHR</i> 17(9):867–871 (2001)
01_AE.TH.95TH047	AB032741	GAG POL REV TAT VIF VPU	Auwanit, W	<i>ARHR</i> 17(9):867–871 (2001)
01_AE.TH.CM235-2	AF259954	GAG NEF VPR VPU	Salminen, MO	<i>Virology</i> 278(1):103–110 (2000)
01_AE.TH.CM235-4	AF259955	REV TAT VIF	Salminen, MO	<i>Virology</i> 278(1):103–110 (2000)
01_AE.TH.CM240	U54771	ENV GAG NEF POL REV TAT VIF VPR VPU	Laukkanen, T	<i>J Virol</i> 70(9):5935–5943 (1996)
01_AE.TW.TWE13	AF220479	VPU	Lee, CN	<i>J Clin Microbiol</i> 38(7):2468–2474 (2000)
01_AE.TW.TWE6	AF220478	VPU	Lee, CN	<i>J Clin Microbiol</i> 38(7):2468–2474 (2000)
02_AG.CM.97 CM-MP807	AJ286133	NEF POL REV TAT VIF VPR VPU	Montavon, C	<i>JAIDS</i> 23(5):363–74 (2000)
02_AG.CM.CM52885	AF377954	ENV NEF POL REV TAT VIF VPR VPU	Carr, JK	<i>Virology</i> 286(1):168–181 (2001)
02_AG.CM.CM53658	AF377955	ENV NEF TAT VIF VPR VPU	Carr, JK	<i>Virology</i> 286(1):168–181 (2001)
02_AG.FR.AKON	AF321079	ENV	Lemiale, F	<i>JAIDS</i> 26(5):413–22 (2001)
02_AG.FR.DJ263	AF063223	GAG NEF VPR	Laukkanen, T	<i>Virology</i> 247(1):22–31 (1998)
02_AG.FR.DJ264	AF063224	ENV GAG NEF POL REV TAT VIF VPU	Laukkanen, T	<i>Virology</i> 247(1):22–31 (1998)
02_AG.GH.97GHAG1	AB049811	GAG NEF POL REV TAT VIF VPR VPU	Kusagawa, S	<i>ARHR</i> 17(7):649–655 (2001)
02_AG.GH.AB052867	AB052867	ENV GAG POL VPU	Takahoko, M	<i>ARHR</i> 17(11):1083–1087 (2001)
02_AG.GH.G829	AF184155	GAG POL REV TAT VIF VPR VPU	Candotti, D	<i>J Med Virol</i> 62(1):1–8 (2000)
02_AG.NG.IBNG	L39106	ENV GAG NEF POL REV TAT VIF VPR VPU	Howard, TM	<i>ARHR</i> 10(12):1755–1757 (1994)
02_AG.NG.NG1921	AF069941	ENV	McCutchan, FE	<i>Virology</i> 254(2):226–234 (1999)
02_AG.SE.SE7812	AF107770	ENV GAG NEF POL REV TAT VIF VPR VPU	Laukkanen, T	Unpublished
02_AG.SN.98SE-MP1211AJ251056		ENV GAG NEF POL VPR VPU	Toure-Kane, C	<i>ARHR</i> 16(6):603–9 (2000)

02_AG.SN.MP1213	AJ251057	REV VIF VPR	Toure-Kane, C	<i>ARHR</i> <b>16</b> (6):603–9 (2000)
03_AB.RU.KAL153-2	AF193276	ENV GAG NEF POL REV TAT VIF VPR VPU	Liitsola, K	<i>AIDS</i> <b>12</b> (14):1907–1919 (1998)
03_AB.RU.KAL68-1	AF082485	ENV	Liitsola, K	<i>AIDS</i> <b>12</b> (14):1907–1919 (1998)
03_AB.RU.RU98001	AF193277	ENV GAG NEF POL REV TAT VIF VPR VPU	Liitsola, K	<i>ARHR</i> <b>16</b> (11):1047–53 (2000)
04_cpx.CY.94CY032-3	AF049337	ENV GAG NEF POL REV TAT VIF VPR VPU	Gao, F	<i>J Virol</i> <b>72</b> (12):10234–10241 (1998)
04_cpx.GR.97PVCH	AF119820	ENV GAG NEF POL REV TAT VIF VPR VPU	Nasioulas, G	<i>ARHR</i> <b>15</b> (8):745–758 (1999)
04_cpx.GR.97PVMY	AF119819	ENV GAG NEF POL REV TAT VIF VPR VPU	Nasioulas, G	<i>ARHR</i> <b>15</b> (8):745–758 (1999)
05_DF.BE.VI1310	AF193253	ENV GAG NEF POL REV TAT VIF VPR VPU	Laukkanen, T	<i>Virology</i> <b>269</b> (1):95–104 (2000)
05_DF.BE.VI961	AF076998	ENV GAG NEF POL REV TAT VIF VPR	Laukkanen, T	<i>Virology</i> <b>269</b> (1):95–104 (2000)
06_cpx.AU.BFP90	AF064699	ENV GAG NEF POL REV TAT VIF VPR VPU	Oelrichs, RB	<i>ARHR</i> <b>14</b> (16):1495–1500 (1998)
06_cpx.ML.95ML127	AJ288982	GAG NEF POL REV TAT VIF VPU	Montavon, C	<i>ARHR</i> <b>15</b> (18):1707–1712 (1999)
06_cpx.ML.95ML84	AJ245481	ENV GAG REV VPR	Montavon, C	<i>ARHR</i> <b>15</b> (18):1707–1712 (1999)
06_cpx.SN.97SE1078	AJ288981	ENV GAG NEF POL REV TAT VIF VPR VPU	Montavon, C	<i>ARHR</i> <b>15</b> (18):1707–1712 (1999)
07_BC.CN.97CN001	AF286226	ENV GAG NEF POL REV TAT VIF VPR VPU	Rodenburg, CM	<i>ARHR</i> <b>17</b> (2):161–168 (2001)
07_BC.CN.98CN009	AF286230	ENV GAG NEF POL REV TAT VIF VPR VPU	Rodenburg, CM	<i>ARHR</i> <b>17</b> (2):161–168 (2001)
08_BC.CN.97CNGX6F	AY008715	ENV NEF POL REV TAT VIF VPR VPU	Piyasirisilp, S	<i>J Virol</i> <b>74</b> (23):11286–11295 (2000)
08_BC.CN.97CNGX7F	AY008716	GAG NEF POL REV TAT VIF VPR VPU	Piyasirisilp, S	<i>J Virol</i> <b>74</b> (23):11286–11295 (2000)
08_BC.CN.97CNGX9F	AY008717	ENV	Piyasirisilp, S	<i>J Virol</i> <b>74</b> (23):11286–11295 (2000)
08_BC.CN.98CN006	AF286229	ENV GAG NEF POL REV TAT VIF VPR VPU	Rodenburg, CM	<i>ARHR</i> <b>17</b> (2):161–168 (2001)
10_CD.TZ.96TZBF061	AF289548	ENV GAG NEF POL REV TAT VIF VPR VPU	Kouliniska, IN	<i>ARHR</i> <b>17</b> (5):423–431 (2001)
10_CD.TZ.96TZBF071	AF289549	ENV GAG NEF POL REV TAT VIF VPR VPU	Kouliniska, IN	<i>ARHR</i> <b>17</b> (5):423–431 (2001)
10_CD.TZ.96TZBF110	AF289550	ENV GAG NEF POL REV TAT VIF VPR VPU	Kouliniska, IN	<i>ARHR</i> <b>17</b> (5):423–431 (2001)
11_cpx.CM.MP818	AJ291718	ENV GAG NEF POL REV TAT VIF VPR VPU	Montavon, C	<i>ARHR</i> <b>18</b> (3):231–236(2002)
11_cpx.FR.MP1298	AJ291719	GAG NEF POL REV TAT VIF	Montavon, C	<i>ARHR</i> <b>18</b> (3):231–236(2002)
11_cpx.FR.MP1307	AJ291720	GAG VPR VPU	Montavon, C	<i>ARHR</i> <b>18</b> (3):231–236(2002)
11_cpx.GR.GR17	AF179368	ENV GAG NEF POL REV TAT VIF VPR VPU	Paraskevis, D	<i>ARHR</i> <b>16</b> (9):845–855 (2000)
11_cpx.NG.NG3670b	AF069945	ENV	McCutchan, FE	<i>Virology</i> <b>254</b> (2):226–234 (1999)
12_BF.AR.ARMA159	AF385936	ENV GAG NEF POL REV TAT VIF VPR VPU	Carr, JK	<i>AIDS</i> <b>15</b> :F41–F47 (2001)
12_BF.UY.URTR23	AF385934	ENV GAG NEF POL REV TAT VIF VPR VPU	Carr, JK	<i>AIDS</i> <b>15</b> :F41–F47 (2001)
12_BF.UY.URTR35	AF385935	ENV GAG NEF POL REV TAT VIF VPR VPU	Carr, JK	<i>AIDS</i> <b>15</b> :F41–F47 (2001)
N.CM.YBF106	AJ271370	ENV GAG NEF POL REV TAT VIF VPR	Ayoub, A	<i>AIDS</i> <b>14</b> (16):2623–2625 (2000)
N.CM.YBF30	AJ006022	ENV GAG NEF POL REV TAT VIF VPR VPU	Simon, F	<i>Nature Med</i> <b>4</b> (9):1032–1037 (1998)



O.CM.AF1422	AF316857	VIF VPR VPU	Vallejo, A	Unpublished
O.CM.AFM0	AF316861	VIF	Vallejo, A	Unpublished
O.CM.ANT70	L20587	ENV GAG NEF POL REV TAT VIF VPR VPU	Vanden Haesevelde, M	<i>J Virol</i> <b>68</b> (3):1586–1596 (1994)
O.CM.CM4974	AF009033	ENV	Korber, BT	Unpublished
O.CM.HIV1CA9EN	X96522	ENV	Janssens, W	<i>AIDS</i> <b>13</b> :41–48 (1999)
O.CM.MVP5180	L20571	ENV GAG NEF POL REV TAT VIF VPR VPU	Gurtler, LG	<i>J Virol</i> <b>68</b> :1581–1585 (1994)
O.FR.HIVY16019	Y16019	VPR	Bibollet-Ruche, F	<i>ARHR</i> <b>14</b> (11):951–61 (1998)
O.FR.HIVY16031	Y16031	VPU	Bibollet-Ruche, F	<i>ARHR</i> <b>14</b> (11):951–61 (1998)
O.GA.VI686	X96526	ENV	Delaporte, E	<i>AIDS</i> <b>10</b> (8):903–910 (1996)
O.GQ.193HA	U82990	ENV	Hunt, JC	<i>ARHR</i> <b>13</b> (12):995–1005 (1997)
O.GQ.AF316859	AF316859	VIF VPR	Vallejo, A	Unpublished
O.SN.SE42HALD	AJ300450	GAG POL VIF VPR	Toure-Kane, C	<i>ARHR</i> <b>17</b> (12):1211–1216 (2001)
O.SN.SEMP1299	AJ302646	GAG NEF POL REV TAT VIF VPU	Toure-Kane, C	<i>ARHR</i> <b>17</b> (12):1211–1216 (2001)
O.SN.SEMP1300	AJ302647	ENV GAG NEF POL REV TAT	Toure-Kane, C	<i>ARHR</i> <b>17</b> (12):1211–1216 (2001)
CPZ.CD.CPZANT	U42720	ENV GAG NEF POL REV TAT VIF VPR VPU	Vanden Haesevelde, MM	<i>Virology</i> <b>221</b> (2):346–350 (1996)
CPZ.CM.CAM3	AF115393	ENV GAG NEF POL REV TAT VIF VPR VPU	Corbet, S	<i>J Virol</i> <b>74</b> :529–534 (2000)
CPZ.CM.CAM5	AJ271369	ENV GAG NEF POL REV TAT VIF VPR VPU	Souquiere, S	Unpublished
CPZ.GA.CPZGAB	X52154	ENV GAG NEF POL REV TAT VIF VPR VPU	Huet, T	<i>Nature</i> <b>345</b> (6273):356–359 (1990)
CPZ.US.CPZUS	AF103818	ENV GAG NEF POL REV TAT VIF VPR VPU	Gao, F	<i>Nature</i> <b>397</b> (6718):436–441 (1999)

	membrane binding	p17	nuclear localization	nuclear localization site	
B.FR.HXB2	MGARASVLSGGELDRWEKIRLRPGKKYKXKHIWVWSRELFVAVNPLLETSEGRQLIGLQPSLQTSBELSLNVTATLYCVHORIEIKDTKEALDKIEEENQKSKKA.....QQAAADTGH				124
CONSENSUS A-A1	---k-a	---l	---s	---me	---A-k-T
A.SE.SE7535	---K-A	---Q-R	---L-S	---A-Q	---MK-A-K-T
A.SE.SE8638	---R-A	---RM-L	---D-L	---S-T	---Q-T
A.SE.SE8891	---K-A	---M-R	---KS-L	---S-T	---Q-MDK-A-R-T
A.SE.UGSE8131	---K-A	---N	---K-L	---A-Q	---IE-SAPK-T
AI.KE.Q23-17	---KF-A	---RM-LI	---D-L	---S-AD-Q	---MEK-A-K-T
AI.SE.SE7253	---K-A	---RM-L	---L-S	---T-Q	---ME-SA-K-T
AI.UG.92UG037	---K-A	---R-L	---L-S	---T-Q	---ME-SA-R-T
AI.UG.U455	---KK-S	---N-R	---R-L	---K-TL	---A-Q
A2.CD.97CDKTB48	---K-EA	---R-L	---K-SI	---S-ET	---R-F
A2.CY.94CY017.41	---I	---K-R	---L	---P-IR	---A-E-T
CONSENSUS B	---O	---R	---L	---E-A	---F-L
B.CN.RL42	---I	---G	---R	---Q	---K-DV
B.DE.D31	---I	---Q	---R	---L	---S
B.ES.89SP061	---I	---A	---K-RV	---R-L	---S
B.JP.JH31	---I	---K	---N	---R	---V
B.KR.WK	---I	---Q	---R	---L	---S
B.TW.TWCYS	---I	---A	---K-K	---R-R	---S
B.US.JREFL	---I	---K	---N	---R	---V
B.US.MNCG	---I	---K	---K	---R	---R
B.US.RF	---I	---K	---K	---R	---R
B.US.SF2	---I	---K	---K	---R	---R
CONSENSUS C	---i-r	---k-k	---k-k	---h-m	---l-l
C.BR.92BR025	---I-R	---K-A	---R-K	---K-H	---MM-L
C.BE.96BW01B22	---I-R	---K-Q	---M-L	---L-L	---L-L
C.ET.EHT2220	---I-R	---EK-A	---K	---M-L	---N-K
C.IL.981S002	---I-R	---EK-T	---R-M	---L-L	---N-K
C.IN.93IN101	---I-R	---K-K	---R	---M-L	---L-L
C.IN.95IN21068	---I-R	---K-K	---K	---R-M	---L-L
C.IN.98IN012	---I-R	---K-K	---K	---K	---K
C.TZ.98TZ013	---I-R	---K-K	---K	---K	---K
C.TZ.98TZ017	---I-R	---K-K	---K	---K	---K
C.ZA.97ZA012	---I-R	---EK-K	---R	---H-MI	---L-L
C.ZM.96ZM751	---I-R	---EK-K	---R	---H-MI	---L-L
CONSENSUS D	---K-a	---K-a	---Y-L	---L-l	---L-l
D.CD.84ZRO85	---K-K	---T-R	---A-LI	---L-L	---L-L
D.CD.ELI	---K-A	---R-L	---L-L	---L-L	---L-L
D.CD.NDK	---K-A	---R-L	---L-L	---L-L	---L-L
D.CD.Z226	---K-A	---R-L	---L-L	---L-L	---L-L
D.UG.94UG1141	---K-a	---a	---?	---m-L?	---?
CONSENSUS F1	---I	---K-E	---Q	---R-W	---LI-LD
F1.BE.VI850	---K-A	---R-L	---L	---L	---L
F1.BR.B2162	---K-A	---Q	---R-M	---L-L	---L-L
F1.CD.VI174	---K-A	---R	---M-LI	---L-L	---L-L
F1.RW.VI169	---K-A	---K	---R-R	---L-L	---L-L
F2.CM.MP255	---K-A	---K	---R-R	---L-L	---L-L
F2.CM.MP257	---K-A	---K	---R-R	---L-L	---L-L
CONSENSUS G	---K-a	---a	---?	---l	---?
G.BE.DRCBL	---K-A	---R-RM	---L-L	---L-L	---L-L
G.FI.HH8793-1.1	---K-A	---R-L	---L-L	---L-L	---L-L
G.NG.92NG083	---K-S	---R	---L-L	---L-L	---L-L
G.SE.SE6165	---T-K	---K-A	---R-S	---I-L	---L-L
H.BE.VI991	---K-A	---R	---R-L	---L-L	---L-L
H.BE.VI997	---H	---TL	---R-L	---L-L	---L-L
H.CF.90CF056	---K-A	---K	---A	---L-L	---L-L
J.SE.SE7022	---I	---K-D	---Q-RI	---L-L	---L-L
J.SE.SE7887	---I	---K-D	---Q-RI	---L-L	---L-L
K.BE.VI325	---K-T	---K-Q	---S	---R-L	---L-L
K.CD.F0TB11C	---K-K	---Q	---R-L	---L-L	---L-L
K.CM.MP535	---K-K	---A	---Q	---R-L	---L-L

	p17 membrane binding	nuclear localization site
B_FR_HXB2	MGARASVLSGGEIDRWEKIRLRPGKKYKXKHIWASRELFANPGLLETSEGRQILGQQLPQLQTSBELSRLNVTATLYCVHQRIEIKDTEALDKIEEQNKSKKKA.....QQAAADTGH	
CONSENSUS_01_AE	RM-L	a-Q-LIE
01_AE_CF_90CF11697	RM-L	SN-K
01_AE_CF_90CF402	RM-L	ST-K
01_AE_CF_90CF401	RM-L	IK-F
01_AE_CN_97CNGX2F	RM-L	ST-K
01_AE_JP_93JP-NHL	RM-L	ST-K
01_AE_TH_93TH057	RM-L	ST-K
01_AE_TH_94TH7091	M-L	ST-K
01_AE_TH_95TH047	M-L	ST-K
01_AE_TH_CM235-2	M-L	ST-K
01_AE_TH_CM240	M-L	ST-K
CONSENSUS_02_AG	r-L	a-Q-fme
02_AG_FR_DJ263	R-L	LIE
02_AG_FR_DJ264	R-L	LIE
02_AG_GH_A3	R-L	LIE
02_AG_GH_AB052867	R-L	LIE
02_AG_GH_G829	R-L	LIE
02_AG_NG_IBNG	R-L	LIE
02_AG_SE_SE7812	R-L	LIE
02_AG_SN_98SE-MP1211	R-L	LIE
03_AB_RU_KAL153-2	E-RI-L	L-S
03_AB_RU_RU98001	RI-L	I-S
04_CPX_CV_94CV032-3	R-L	A-Q-LME
04_CPX_GR_97PVCH	R-L	A-Q-LME
04_CPX_GR_97PVNY	R-L	A-Q-LME
05_DF_BE_V11310	R-L	A-Q-LME
05_DF_BE_V1961	R-L	A-Q-LME
06_CPX_AU_BFP90	M-L	A-Q-LIE
06_CPX_ML_95ML127	R-L	A-Q-LME
06_CPX_ML_95ML84	R-L	A-Q-LME
06_CPX_SN_97SE1078	R-L	A-Q-LME
07_BC_CN_97CN001	H-M-L	L-M
07_BC_CN_98CN009	H-M-L	L-M
08_BC_CN_97CNGX7F	H-M-L	L-M
08_BC_CN_98CN006	HCM-L	L-L
10_CD_TZ_96TZBF061	Q-L	L-D
10_CD_TZ_96TZBF071	R-L	L-L
10_CD_TZ_96TZBF110	R-L	L-L
CONSENSUS_CPX	R-L	L-L
11_CPX_CM_MF818	R-L	L-L
11_CPX_FR_MP1298	R-L	L-L
11_CPX_FR_MP1307	R-L	L-L
11_CPX_GR_GRL7	R-L	L-L
12_BF_AR_ARMAL59	R-L	L-L
12_BF_UY_URTR23	R-L	L-L
12_BF_UY_URTR35	R-L	L-L
N_CM_YBF106	R-L	L-L
N_CM_YBF30	R-L	L-L
CONSENSUS_O	R-L	L-L
O_CM_AN770	R-L	L-L
O_CM_MVP5180	R-L	L-L
O_SN_SE42HALD	R-L	L-L
O_SN_SEMP1299	R-L	L-L
O_SN_SEMP1300	R-L	L-L
CONSENSUS_CPZ	M-L	L-L
CPZ_CD_CPFZANT	MI-L	RS-Q
CPZ_CM_CAM3	MM-L	L-D
CPZ_CM_CAM5	MM-L	L-D
CPZ_GA_CPFZGAB	R-R-MM-L	L-L
CPZ_US_CPFZUS	MM-L	L-L





HIV-1/SIVcpz proteins

	major homology region (MHR)	p24 \ / p2
B.FR.HXB2	...NPPIPVGEIYKRWLLGLNKIVRMYSPTSLIDIRQGPKEPFRDYDRFKTLRABQASQEVKNNWMTETLLVQNANPDKTILLKALGPAATLEEMMTACQGVGPFHKARVLAEMSQV	
CONSENSUS_A-A1	V...k	s...f-ag
A.SE.SE7535	D...D	s...f-ag
A.SE.SE8538	V...K	S...R-AG
A.SE.SE8891	V...K	S...R-AG
A.SE.UGSE8131	V...K	S...R-AG
A1.KE.Q23-17	V...K	S...R-AG
A1.SE.SET753	V...K	S...R-AG
A1.UG.92UG037	V...K	S...R-AG
A1.UG.U455	V...K	S...R-AG
A2.CD.97CDKTB48	V...K	S...R-AG
A2.CY.94CY017.41	V...K	S...R-AG
CONSENSUS_B	V...C	S...I
B.CN.RL42	V...C	S...I
B.DE.D31	V...C	S...I
B.ES.89SP061	V...C	S...I
B.GA.OYI	V...C	S...I
B.JP.JH31	V...C	S...I
B.KR.WK	V...C	S...I
B.TW.TWCYS	V...C	S...I
B.US.JRFL	V...C	S...I
B.US.MNCG	V...C	S...I
B.US.RF	V...C	S...I
B.US.SF2	V...C	S...I
CONSENSUS_C	V...k	s...f-ag
C.BR.92BR025	V...k	s...f-ag
C.BW.96BW01B22	V...k	s...f-ag
C.ET.ETH2220	V...k	s...f-ag
C.IL.98ILS002	V...k	s...f-ag
C.IN.93IN101	V...k	s...f-ag
C.IN.95IN21068	V...k	s...f-ag
C.IN.98IN012	V...k	s...f-ag
C.TZ.98TZ013	V...k	s...f-ag
C.TZ.98TZ017	V...k	s...f-ag
C.ZA.97ZA012	V...k	s...f-ag
C.ZM.96ZM651	V...k	s...f-ag
C.ZM.96ZM751	V...k	s...f-ag
CONSENSUS_D	V...d	s...f-ag
D.CD.84ZR085	V...d	s...f-ag
D.CD.FLI	V...d	s...f-ag
D.CD.NDK	V...d	s...f-ag
D.CD.72Z6	V...d	s...f-ag
D.UG.94UG1141	V...d	s...f-ag
CONSENSUS_F1	V...k	s...f-ag
F1.BE.VI850	V...k	s...f-ag
F1.BR.B2162	V...k	s...f-ag
F1.CD.VI174	V...k	s...f-ag
F1.RW.VI69	V...k	s...f-ag
F2.CM.MP255	V...k	s...f-ag
F2.CM.MP257	V...k	s...f-ag
CONSENSUS_G	V...g	s...f-ag
G.BE.DRCBL	V...g	s...f-ag
G.FI.HH8793-1.1	V...g	s...f-ag
G.NG.92NG083	V...g	s...f-ag
G.SE.SE6165	V...g	s...f-ag
H.BE.VI991	V...g	s...f-ag
H.BE.VI997	V...g	s...f-ag
H.CF.90CF056	V...g	s...f-ag
J.SE.SET7022	V...g	s...f-ag
J.SE.SET887	V...g	s...f-ag
K.BE.VI325	V...g	s...f-ag
K.CD.E0TB11C	V...g	s...f-ag
K.CM.MP535	V...g	s...f-ag



HIV-1/SIVcpz proteins

Table with columns: protein name, p2, p7, p1, Pol CDS start, p1, p6, Vpr binding, and sequence alignment. Rows include various HIV-1/SIVcpz proteins like B.FR.HXB2, CONSENSUS A-A1, B.CN.RL42, etc., aligned with their respective amino acid sequences.



	p2	p7	p1	p6	p1	p6	p1	p6	p1	p6	p1	p6
	TSNAT	IMQGNFRNRK	IVKFNCKGEGHTANRCRPRKCKCWCCKGEGHQMKDCTE	ROANFLGKIWPSYKG	RPNGFLQSR	ROANFLGKIWPSYKG	RPNGFLQSR	ROANFLGKIWPSYKG	RPNGFLQSR	ROANFLGKIWPSYKG	RPNGFLQSR	ROANFLGKIWPSYKG
	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban
	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban
B_FR_HXB2	TSNAT	IMQGNFRNRK	IVKFNCKGEGHTANRCRPRKCKCWCCKGEGHQMKDCTE	ROANFLGKIWPSYKG	RPNGFLQSR	ROANFLGKIWPSYKG	RPNGFLQSR	ROANFLGKIWPSYKG	RPNGFLQSR	ROANFLGKIWPSYKG	RPNGFLQSR	ROANFLGKIWPSYKG
CONSENSUS_01_AE	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban
01_AE_CF_90CF11697	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban
01_AE_CF_90CF402	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban
01_AE_CF_90CF4071	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban
01_AE_CN_97CNGX2F	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban
01_AE_JP_930P-NH1	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban
01_AE_TH_93TH057	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban
01_AE_TH_94TH7091	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban
01_AE_TH_95TH047	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban
01_AE_TH_CW235-2	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban
01_AE_TH_CW240	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban	gban
CONSENSUS_02_AG	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N
02_AG_FR_DJ263	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N
02_AG_FR_DJ264	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N
02_AG_GH_A3	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N
02_AG_GH_AB052867	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N
02_AG_GH_G829	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N
02_AG_GH_IBNG	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N
02_AG_SE_SE7812	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N
02_AG_SN_98SE-MP1211	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N
03_AB_RU_KAL153-2	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N
03_AB_RU_RU98001	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N	Qc-N
04_CPX_CX_94CY032-3	S-A-AA	S-A-AA	S-A-AA	S-A-AA	S-A-AA	S-A-AA	S-A-AA	S-A-AA	S-A-AA	S-A-AA	S-A-AA	S-A-AA
04_CPX_GR_97PVCH	SSA-AA	SSA-AA	SSA-AA	SSA-AA	SSA-AA	SSA-AA	SSA-AA	SSA-AA	SSA-AA	SSA-AA	SSA-AA	SSA-AA
04_CPX_GR_97PVNY	SSA-AA	SSA-AA	SSA-AA	SSA-AA	SSA-AA	SSA-AA	SSA-AA	SSA-AA	SSA-AA	SSA-AA	SSA-AA	SSA-AA
05_DF_BE_V11310	-G-PA	-G-PA	-G-PA	-G-PA	-G-PA	-G-PA	-G-PA	-G-PA	-G-PA	-G-PA	-G-PA	-G-PA
05_DF_BE_V11361	-G-PA	-G-PA	-G-PA	-G-PA	-G-PA	-G-PA	-G-PA	-G-PA	-G-PA	-G-PA	-G-PA	-G-PA
06_CPX_AU_BFP90	SVG	SVG	SVG	SVG	SVG	SVG	SVG	SVG	SVG	SVG	SVG	SVG
06_CPX_ML_95ML127	SGTL-A	SGTL-A	SGTL-A	SGTL-A	SGTL-A	SGTL-A	SGTL-A	SGTL-A	SGTL-A	SGTL-A	SGTL-A	SGTL-A
06_CPX_ML_95ML84	SGTVAAA	SGTVAAA	SGTVAAA	SGTVAAA	SGTVAAA	SGTVAAA	SGTVAAA	SGTVAAA	SGTVAAA	SGTVAAA	SGTVAAA	SGTVAAA
06_CPX_SN_97SE1078	SGATAA	SGATAA	SGATAA	SGATAA	SGATAA	SGATAA	SGATAA	SGATAA	SGATAA	SGATAA	SGATAA	SGATAA
07_BC_CN_97CN001	NSA	NSA	NSA	NSA	NSA	NSA	NSA	NSA	NSA	NSA	NSA	NSA
07_BC_CN_98CN009	NST	NST	NST	NST	NST	NST	NST	NST	NST	NST	NST	NST
08_BC_CN_97CNGX7F	N-T	N-T	N-T	N-T	N-T	N-T	N-T	N-T	N-T	N-T	N-T	N-T
08_BC_CN_98CN006	NST	NST	NST	NST	NST	NST	NST	NST	NST	NST	NST	NST
10_CD_TZ_96TZBF061	-GGN	-GGN	-GGN	-GGN	-GGN	-GGN	-GGN	-GGN	-GGN	-GGN	-GGN	-GGN
10_CD_TZ_96TZBF071	-KGN	-KGN	-KGN	-KGN	-KGN	-KGN	-KGN	-KGN	-KGN	-KGN	-KGN	-KGN
10_CD_TZ_96TZBF110	-SGNA	-SGNA	-SGNA	-SGNA	-SGNA	-SGNA	-SGNA	-SGNA	-SGNA	-SGNA	-SGNA	-SGNA
CONSENSUS_CPX	qgTh	qgTh	qgTh	qgTh	qgTh	qgTh	qgTh	qgTh	qgTh	qgTh	qgTh	qgTh
11_CPX_CM_MP818	QQTN	QQTN	QQTN	QQTN	QQTN	QQTN	QQTN	QQTN	QQTN	QQTN	QQTN	QQTN
11_CPX_FR_MP1298	ROTN	ROTN	ROTN	ROTN	ROTN	ROTN	ROTN	ROTN	ROTN	ROTN	ROTN	ROTN
11_CPX_FR_MP1307	QQTN	QQTN	QQTN	QQTN	QQTN	QQTN	QQTN	QQTN	QQTN	QQTN	QQTN	QQTN
11_CPX_GR_GRL17	OHTN	OHTN	OHTN	OHTN	OHTN	OHTN	OHTN	OHTN	OHTN	OHTN	OHTN	OHTN
12_BF_AR_ARMAL159	-ST	-ST	-ST	-ST	-ST	-ST	-ST	-ST	-ST	-ST	-ST	-ST
12_BF_UY_URTR23	--PT	--PT	--PT	--PT	--PT	--PT	--PT	--PT	--PT	--PT	--PT	--PT
12_BF_UY_URTR35	--TT	--TT	--TT	--TT	--TT	--TT	--TT	--TT	--TT	--TT	--TT	--TT
N_CM_YBF106	QQT-S	QQT-S	QQT-S	QQT-S	QQT-S	QQT-S	QQT-S	QQT-S	QQT-S	QQT-S	QQT-S	QQT-S
N_CM_YBF30	QQT-S	QQT-S	QQT-S	QQT-S	QQT-S	QQT-S	QQT-S	QQT-S	QQT-S	QQT-S	QQT-S	QQT-S
CONSENSUS_O	qgdkLGGYLAVF	qgdkLGGYLAVF	qgdkLGGYLAVF	qgdkLGGYLAVF	qgdkLGGYLAVF	qgdkLGGYLAVF	qgdkLGGYLAVF	qgdkLGGYLAVF	qgdkLGGYLAVF	qgdkLGGYLAVF	qgdkLGGYLAVF	qgdkLGGYLAVF
O_CM_ANT170	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF
O_CM_MVP5180	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF
O_SN_SEA2HALD	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF
O_SN_SEMP1299	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF
O_SN_SEMP1300	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF	QDLKGGYTAVF
CONSENSUS_CPZ	q-psg??	q-psg??	q-psg??	q-psg??	q-psg??	q-psg??	q-psg??	q-psg??	q-psg??	q-psg??	q-psg??	q-psg??
CPZ_CD_CPZANT	N-AQGTA	N-AQGTA	N-AQGTA	N-AQGTA	N-AQGTA	N-AQGTA	N-AQGTA	N-AQGTA	N-AQGTA	N-AQGTA	N-AQGTA	N-AQGTA
CPZ_CM_CAM3	O-PTG	O-PTG	O-PTG	O-PTG	O-PTG	O-PTG	O-PTG	O-PTG	O-PTG	O-PTG	O-PTG	O-PTG
CPZ_CM_CAM5	O-SN	O-SN	O-SN	O-SN	O-SN	O-SN	O-SN	O-SN	O-SN	O-SN	O-SN	O-SN
CPZ_GA_CPZGB	Q-CGRAD	Q-CGRAD	Q-CGRAD	Q-CGRAD	Q-CGRAD	Q-CGRAD	Q-CGRAD	Q-CGRAD	Q-CGRAD	Q-CGRAD	Q-CGRAD	Q-CGRAD
CPZ_US_CPZUS	K-FSS	K-FSS	K-FSS	K-FSS	K-FSS	K-FSS	K-FSS	K-FSS	K-FSS	K-FSS	K-FSS	K-FSS

HIV-1/SIVcpz proteins

	Vpr binding	p6 \
500	T.....TPFKQEP.....IDK.ELY..PLTSLRSLFGNDFSSQ\$	
494	.....?S-p-g.....k-r.-gn?p-v-k-ll-	
495	A.SE.SE7535	AS-P-O.....GOD.P-V-K-LL-
496	A.SE.SE8538	.....S-P-O.....OV.P-V-K-LL-
497	A.SE.SE8891	.....S-P-K.....K-R.-SEAP-I-K-LL-
492	A.SE.UGSE8131	AS-P-Q.....NN.P-SV-K-LL-
498	AL.KE.Q23-17	VS-L-Q.....K-R.-QA.Q-V-K-LL-
499	AL.SE.SET253	AS-P-Q.....K-R.-QN.S-SV-K-LL-
499	AL.UG.92UG037	VS-P-Q.....N-R.DQN.P-SV-K-LL-
499	AL.UG.U455	.....S-A-L.....K-R.-QT..V-K-LL-
497	A2.CD.97CDKTB48	SSL-N.....R.-PS.T-AI-K-LL-
496	A2.CY.94CY017.41	SSL-L.....ETR.-P.-N-AI-K-LL-
499	CONSENSUS_B	.....S-????.....a
505	B.CN.RL42	.....S.....A-K
500	B.DE.D31	.....E.....A
503	B.ES.89SP061	.....S-L.....SA-K-L
499	B.GA.OYI	.....S.....G
500	B.JP.UH31	.....S.....R.....A-----P-
500	B.KR.WK	.....S.....D.....A
500	B.TW.TWCYS	.....S.....D.....A-E
498	B.US.JRFL	.....S.....M.....
506	B.US.MNCG	.....Y.....KKOET.....D.....A-K-L
501	B.US.RF	.....S.....K.....A-K
502	B.US.SF2	.....S.....S.....
485	CONSENSUS_C	.....?..-ap-.....k-r-????.....k-s-l-
496	C.BR.92BR025	.....SR-T.....K-R.....K-S-L-T-
492	C.BW.96BW01B22	.....AP-L.....K-R.....S-L
504	C.ET.ETH2220	.....SP-L.....K-R.....A-K-HLL-
496	C.IL.98IS002	.....SP-Q.....K-P-KE-SA-K-L
491	C.IN.93IN101	.....AP-L.....K-R.....K-S-LL-
491	C.IN.95IN21068	.....AP-L.....K-R.....K-S-L
492	C.IN.98IN012	.....AP-L.....K-R.....K-S-LF-
499	C.TZ.98TZ013	.....AP-L.....K-R.....I-K-S-L
502	C.TZ.98TZ017	.....LAR-L.....K-R.....T-K-S-L
495	C.ZA.97ZA012	.....L-V-S.....K-R.....A-K-S-L
494	C.ZM.96ZM651	.....AP-S.....K-R.....A-K-S-L
498	C.ZM.96ZM751	.....APR-Q.....K.....A-K-S-L
498	CONSENSUS_D	.....S-OP?.K.....K-K-K-L
502	D.CD.84ZR085	.....S-OKD.K.....A-K-L
500	D.CD.FLI	.....S-O.K.....A-K-L
497	D.CD.NDK	.....S-O.K.....A-K-L
501	D.CD.ZZ26	.....S-Q.Q.....S-A-K-LL-
503	D.UG.94UG1141	.....Q.....K.....K-S-L
473	CONSENSUS_F1	.....SP-Q.....K-e-g--P-A-K
491	F1.BE.VI850	.....SP-O.....K-G--P-A-K-----\$
486	F1.BR.B2162	N.....SP-Q.....KEE.G--P-A-K
487	F1.CD.VI174	.....SP-Q.....KGE.G--P-A-K
486	F1.RW.VI169	.....SP-Q.....K-E-G-S-P-A-K
492	F2.CM.MP255	.....SP-Q.....KGP.-QA.P-V-K-S-\$
496	F2.CM.MP257	A.....SP-Q.....K-K--OV.P-I-K-S-Q\$
486	CONSENSUS_G	.....SP-.....KE.....?..K-s-s-\$
490	G.BE.DRCBL	A.....SP-Q.....KE.....S-K-O\$
497	G.FI.HH8793-1.1	A.....SP-P.....KE.-TH..A-K-S-\$
495	G.NG.92NG083	A.....SP-.....KE.....K-S-\$
498	G.SE.SE6165	A.....SP-M.....KE.....P-K-S-\$
499	H.BE.VI991	.....SPR-L.....OE.P-----O\$
500	H.BE.VI997	SSP-L.....DKE.P-FA-K-----L-
497	H.CF.90CF056	.....SP-Q.....L.DKE.P-A-----S-LL-
497	J.SE.SET7022	.....SP-.....K.....K-----S-L-
497	J.SE.SET7887	.....SP-.....K.....K-----S-L-
486	K.BE.VI325	.....SPR-T.....K-K-OG.S-----K
495	K.CD.EOTB11C	.....SLR-M.....K-K-O.-OG.P-----K-S-L-
495	K.CM.MP535	.....SPR-T.....K-K.-QS.P-----K-L-

```

B_FR_HXB2
CONSENSUS_01_AE
01_AE_CF_90CF11697
01_AE_CF_90CF402
01_AE_CF_90CF4071
01_AE_CN_97CNGX2F
01_AE_JP_93JP-NH1
01_AE_TH_93TH057
01_AE_TH_94TH7091
01_AE_TH_95TH047
01_AE_TH_CW235-2
01_AE_TH_CW240
CONSENSUS_02_AG
02_AG_FR_DJ263
02_AG_FR_DJ264
02_AG_GH_A3
02_AG_GH_AB052867
02_AG_GH_G829
02_AG_NG_IBNG
02_AG_SE_SE7812
02_AG_SN_98SE-MP1211
03_AB_RU_KAL153-2
03_AB_RU_RU98001
04_CPX_CX_94CX032-3
04_CPX_GR_97PVCH
04_CPX_GR_97PVNY
05_DF_BE_V11310
05_DF_BE_V1961
06_CPX_AU_BFP90
06_CPX_ML_95ML127
06_CPX_ML_95ML84
06_CPX_SN_97SE1078
07_BC_CN_97CN001
07_BC_CN_98CN009
08_BC_CN_97CNGX7F
08_BC_CN_98CN006
10_CD_TZ_96TZBF061
10_CD_TZ_96TZBF071
10_CD_TZ_96TZBF110
CONSENSUS_CPX
11_CPX_CM_MP818
11_CPX_FR_MP1298
11_CPX_FR_MP1307
11_CPX_GR_GRI7
12_BF_AR_ARMAL59
12_BF_UY_URTR23
12_BF_UY_URTR35
N_CM_YBF106
N_CM_YBF30
CONSENSUS_O
O_CM_ANT170
O_CM_MVP5180
O_SN_SEA2HALD
O_SN_SEMP1299
O_SN_SEMP1300
CONSENSUS_CPZ
CPZ_CD_CPZANT
CPZ_CM_CAM3
CPZ_CM_CAM5
CPZ_GA_CPZGAB
CPZ_US_CPZUS
T.....TPFKQKEP.....IDK.ELY..PLTSLRSLFGNDFSSQ$
-??.....??s1?..g..K..-?hpb-V-K-K-K-K-1
-SF.....-SPF-O..K..-HPS-V-K-K-K-L-
.....-SFSR-O..K..KO.PP-V-K-K-L-
.....SLL-O..K..-HPP-V-K-K-L-
.....SLP-O..K..-H-YP-K-K-L-
.....SLL-O..K..-HHP-V-K-K-L-
.....SAXNEDE..K..-PP-V-K-K-L-
-SS.....L-Q..K..-PP-P-I-K-K-L-
01_AE_TH_CW235-2..SFL-Q..K..D.HPP-V-K-K-L-
-GE.....ET-SLP-Q..K..-HP-P-V-K-K-L-
-??.....?ssp-???r-..G..P-..K-K-???
-?.....-P-A..R-Q..G..P-A-K-K-$$$
-S.....-R-Q..G..P-A-K-K-$$$
A.....S-P-S..R-G..P-A-K-K-S-$$$
.....SS--KEQEPFR..G..P-A-K-K-O$
02_AG_GH_G829.....SSPQ--R-G..P-A-K-K-$$$
02_AG_NG_IBNG.....P-SPQ--R-G..P-A-K-K-$$$
-SS.....G-G..G..P-A-K-K-$$$
.....SSP--G-G..P-A-K-K-Y-
.....-SL-Q..K-R.-QH.P-SI-K-D-L-
03_AB_RU_KAL153-2..-SL-Q..K-R.GQH.P-SI-K-L-
-SS.....L-..R-..-K-S-L-
-SS.....P-Q..R-..-K-H-L-
04_CPX_GR_97PVNY..P-Q..R-..I-K-S-L-R-
.....ASSP--Q..K-E.GK-.P-A-K-K-L-
-SP--Q..K-E.GK-.P-A-K-K-L-
A.....-SP--SKEK.EE.G-..A-K-S-$$$
A.....A-SP..Q..KPKK-..A-K-$$$
A.....-SL--KE..KE-..A-K-$$$
.....A-SP--..KE-..A-K-S-$$$
-.....-S--T-..-K-
-.....-AP--K-R-..-S-L-
A.....-AP--K-R-..-S-L-
.....-S-Q..K..-PH-..A-K-L-
.....-E-Q..K..-EL.YL-A-K-L-
.....-S-Q..K..-A-K-V-L-
?????????a-SP-..?ke-..?--K-L-?
.....A-SP-..KEKE-..K-S-$$$
.....A-SP-..KE-..K-L-
GPSKQEQSFGIGENVA-SP-..KKE-..K-L-
-.....-SP-..KE-..I-K-S-L-
.....SSP-R-Q..K-E.GQ-.P-A-K-K-$$$
-PS.....-SP-R-Q..K-E.G-.P-A-K-O$
.....QKI--S-E-Q..KGE.G-.P-A-K-$$$
.....KS-QGKMQENQ..FKTETS-.P-----LX-
.....KS-QGKMQENQ..ERTENS-.P-----
CONSENSUS_O
?Kg?gengedfg-gn-..FA-k-T-Q
VKG_OENQOKGGFN-..FA-K-T-Q$
VKE_OENQOKG-OF-..FA-K-T-Q$
KG-NOEO KE-ON-..FA-K-T-Q$
MKG_OENQOKE-QN-..FA-K-T-Q$
MKG_OENQOKE-QN-..FA-K-T-Q
??.....???g-k-ee???-?e-ep-..P-SY-K-K-K-Q$
KR-QKGLKEEEL-..P-SY-K-K-K-Q$
G-ROEBERE KE-EP-..P--K-K-S-E-
ME-KK-..EE-EP-..PO--K-S-LE-
KSQEKKEG-SS-..P-K-K-S-
VT.....KQG-..KE-EP.F..Q--K-S-E-

```



Table with 3 columns: Accession ID, Gag-Pol TF start, Gag-Pol TF end, Gag CDS end. Includes sequences for various HIV-1 and SIVcpz proteins such as Gag, Pol, and RT. The table shows amino acid alignments across multiple sequences, with positions 96 to 103 indicated on the right side.

	Protease	p66	p51	Reverse Transcriptase (RT)	M41L	D67N	K70R				
B.FR.HXB2	RWPKMIGGIGGF	IKVRQYDQILLI	ELICGHKALGTVL	VGPTPWNIIGRNLLFQIGCTLN	FPISPLFI	VPVKLKFQMDGPKVQMP	LTFEEKIKALVLC	TEMEKKGKISK	IGPENP	INTPVFAIKK	KDSTKW
CONSENSUS A-A1	K	k	K	M	T	i	I				
A.SE.SE7535	K	K	K	K	K	K	K				
A.SE.SE8538	K	K	K	K	K	K	K				
A.SE.UGSE8131	K	R	K	S	K	K	K				NR
A1.KE.O23-17	K	K	K	K	K	K	K				N-DR
A1.SE.SET253	K	K	K	K	K	K	K				
A1.UG.92UG057	K	K	K	K	K	K	K				
A1.UG.U455	K	K	I	K	T	S	R	L			
A2.CD.97CDKFE4	K	K	A	K	R	M	V	L			
A2.CY.94CY017.41	K	K	A	K	R	M	V	L			
CONSENSUS B	?										
B.CN.RL42	P		L	L	K						
B.DE.D31											
B.GA.OYI			L	L	V	I					V
B.GB.CAM1			P								
B.KR.WK			VA								
B.TW.TWCYS			P								
B.US.JRFL			T	G							
B.US.MNCG			PV								
B.US.RF											
B.US.SF2											
CONSENSUS C	k	K	m	L		t	a	e			t
C.BR.92BR025	N	X	K	M	L	K	L				T
C.BR.98BR004	K		K	M	L	K	L				
C.BW.96BW01B03	K		K	M	L	K	L				
C.BW.96BW0402	K		E	M	L	K	L				
C.ET.ETH2220	K		I	K	M	L	R				
C.IL.98IS002	K		C	M	L	K					
C.IN.93IN999	K		P	K	L	K	L				
C.IN.95IN21068	K		EE	P	K	L					
C.IN.98IN012	K		P	K	L	K	L				
C.TZ.98TZ013	K		P	K	L	K	L				
C.TZ.98TZ017	K		P	K	L	K	L				
C.ZA.97ZA012	K		R								
C.ZM.96ZM651	K		E	PM	K	L					
C.ZM.96ZM751	K			K	L	K	L				
CONSENSUS D	K	K	K	H							
D.CD.84ZR085	K		P								
D.CD.FLI	K			Y	M						
D.CD.NDK	K										
D.CD.72Z6	K										
D.UG.94UG1141	K		PL								
CONSENSUS F1	K	K	K	N							
F1.BE.VI850	K		K	N							
F1.BR.93BR020.1	K		K	S							S
F1.FI.FIN9363	K		K	H							
F1.FR.MF411	K		I	K							
F2.CM.CM53657	K	R	P	O							
F2.CM.MP255	K		P	O							
F2.CM.MP257	K		VS	O							
CONSENSUS G	k	K	s	K		I					
G.BE.DRCBL	K		S	K	R						
G.FI.HH8793-1.1	K		S	K		X					
G.NG.92NG083	K		VP	S	K						
G.SE.SE6165	K			S	K						
H.BE.VI991	K		E	VA	F	K					
H.BE.VI997	K		VA		K						
H.CF.90CF056	K		E	VA		K					
J.SE.SET022	K		NEVP	E	K						
J.SE.SET887	K		NEVP	E	K						
K.CD.E0TB11C	K		VCM	O							
K.CM.MP535	K		V		O						

	Protease	p66	p51	Reverse Transcriptase (RT)	M41L	D67N	K70R
B_FR_HXB2	RWPKMIGGIGGFIKVRQYDQILLIICGHKALGTIVGPTPWNIIGRNLLFIQIGCTLINFPLISPLIIVVKLKFQMDPKVKQMPLTFEEKIKALVEICTEMEKEGKISKIGPENPINTPFAIKKDKSTKW	M	D	T	K	e	
CONSENSUS_01_AE	K	D	T	K	e		
01_AE_CF_90CF11697	K	D	T	K	e		
01_AE_CF_90CF402	K	D	T	K	e		
01_AE_CF_90CF4071	K	D	T	K	e		
01_AE_CN_97CNGX2F	K	D	T	K	e		
01_AE_JP_93JP-NH1	K	D	T	K	e		
01_AE_TH_93TH253	K	D	T	K	e		
01_AE_TH_94TH702	K	D	T	K	e		
01_AE_TH_95TH022	K	D	T	K	e		
01_AE_TH_95TH047	K	D	T	K	e		
01_AE_TH_CW240	K	D	T	K	e		
CONSENSUS_02_AG	m			td			
02_AG_CM_97CM-MP807	K			td			
02_AG_CM_CW52885	K			td			
02_AG_FR_DJ264	K			td			
02_AG_GH_A3	K			td			
02_AG_GH_AB052867	K			td			
02_AG_GH_G829	K			td			
02_AG_NG_IBNG	K			td			
02_AG_SE_SE7812	K			td			
02_AG_SN_98SE-MP1211	K			td			
03_AB_RU_KAL153-2	K			td			
03_AB_RU_RU98001	K			td			
04_CPX_CV_94CV032-3	K			td			
04_CPX_GR_97FVCH	K			td			
04_CPX_GR_97FVMY	K			td			
05_DF_BE_V11310	K			td			
05_DF_BE_V1961	K			td			
06_CPX_AU_BFP90	K			td			
06_CPX_ML_95ML127	K			td			
06_CPX_SN_97SE1078	K			td			
07_BC_CN_97CN001	K			td			
07_BC_CN_98CN009	K			td			
08_BC_CN_97CNGX6F	K			td			
08_BC_CN_97CNGX7F	K			td			
08_BC_CN_98CN006	K			td			
10_CD_TZ_96TZBF061	K			td			
10_CD_TZ_96TZBF071	K			td			
10_CD_TZ_96TZBF110	K			td			
11_CPX_CM_MP818	K			td			
11_CPX_FR_MP1298	K			td			
11_CPX_GR_GR17	K			td			
12_BF_AR_ARMAL59	K			td			
12_BF_UY_URTR23	K			td			
12_BF_UY_URTR35	K			td			
N_CM_YBF106	K			td			
N_CM_YBF30	K			td			
CONSENSUS_O	I						
O_CM_ANT70	K						
O_CM_MVP5180	K						
O_SN_SE42HALD	K						
O_SN_SEMP1299	K						
O_SN_SEMP1300	K						
CONSENSUS_CPZ	k						
CPZ_CD_CPZANT	k						
CPZ_CM_CAM3	K						
CPZ_CM_CAM5	K						
CPZ_GA_CPZGAB	K						
CPZ_US_CPZUS	T						

	polymerase motif
B.FR.HXB2	RKLVDRELNKRQTQDFWEVQLGIPHPAGLKKKSVTVLVDGDAYFSVPLDDEDFRKYTAFTTIPSNINNETPGIRYQNVVLFQGWKGSFAIFQSSMTKILPEFRKQNDIIVYQYMDLIVGSDLEIGQHRTK
CONSENSUS_A-A1	-----L-----sk-e-I-----
A.SE.SE7535	-----L-----sk-e-I-----
A.SE.SE8538	-----L-----sk-e-I-----
A.SE.UGSF8131	-----L-----sk-e-I-----
A1.KE.O23-17	-----L-----sk-e-I-----
A1.SE.SET253	-----L-----sk-e-I-----
A1.UG.92UG037	-----L-----sk-e-I-----
A1.UG.U455	-----L-----sk-e-I-----
A2.CD.97CDKFE4	-----L-----sk-e-I-----
A2.CY.94CY017.41	-----L-----sk-e-I-----
CONSENSUS_B	-----L-----sk-e-I-----
B.CN.RL42	-----L-----sk-e-I-----
B.DE.D31	-----L-----sk-e-I-----
B.GA.OYI	-----L-----sk-e-I-----
B.GB.CAM1	-----L-----sk-e-I-----
B.KR.WK	-----L-----sk-e-I-----
B.ITW.TWCYS	-----L-----sk-e-I-----
B.US.JRFL	-----L-----sk-e-I-----
B.US.MNCG	-----L-----sk-e-I-----
B.US.RF	-----L-----sk-e-I-----
B.US.SF2	-----L-----sk-e-I-----
CONSENSUS_C	-----L-----sk-e-I-----
C.BR.92BR025	-----L-----sk-e-I-----
C.BR.98BR004	-----L-----sk-e-I-----
C.BW.96BW01B03	-----L-----sk-e-I-----
C.BW.96BW0402	-----L-----sk-e-I-----
C.ET.ETH2220	-----L-----sk-e-I-----
C.IL.98IS002	-----L-----sk-e-I-----
C.IN.93IN999	-----L-----sk-e-I-----
C.IN.95IN21068	-----L-----sk-e-I-----
C.IN.98IN012	-----L-----sk-e-I-----
C.TZ.98TZ013	-----L-----sk-e-I-----
C.TZ.98TZ017	-----L-----sk-e-I-----
C.ZA.97ZA012	-----L-----sk-e-I-----
C.ZM.96ZM651	-----L-----sk-e-I-----
C.ZM.96ZM751	-----L-----sk-e-I-----
CONSENSUS_D	-----L-----sk-e-I-----
D.CD.84ZR085	-----L-----sk-e-I-----
D.CD.FLI	-----L-----sk-e-I-----
D.CD.NDK	-----L-----sk-e-I-----
D.CD.72Z6	-----L-----sk-e-I-----
D.UG.94UG1141	-----L-----sk-e-I-----
CONSENSUS_F1	-----L-----sk-e-I-----
F1.BE.VI850	-----L-----sk-e-I-----
F1.BR.93BR020.1	-----L-----sk-e-I-----
F1.FI.FIN9363	-----L-----sk-e-I-----
F1.FR.MF411	-----L-----sk-e-I-----
F2.CM.CM53657	-----L-----sk-e-I-----
F2.CM.MP255	-----L-----sk-e-I-----
F2.CM.MP257	-----L-----sk-e-I-----
CONSENSUS_G	-----L-----sk-e-I-----
G.BE.DRCBL	-----L-----sk-e-I-----
G.FI.HH8793-1.1	-----L-----sk-e-I-----
G.NG.92NG083	-----L-----sk-e-I-----
G.SE.SE6165	-----L-----sk-e-I-----
H.BE.VI991	-----L-----sk-e-I-----
H.BE.VI997	-----L-----sk-e-I-----
H.CF.90CF056	-----L-----sk-e-I-----
J.SE.SET022	-----L-----sk-e-I-----
J.SE.SET7887	-----L-----sk-e-I-----
K.CD.E0TB11C	-----L-----sk-e-I-----
K.CM.MP535	-----L-----sk-e-I-----



polymerase motif

B_FR_HXB2	RKLVDRELKRTQDFWEVQLGIPHPAGLKKKSVTVLDVGDAYFVPLDDEDFRKYTAFTTIPSNNETPGIRYQYNNVLCQWKGSFAIFQSSMTKLELFFRQNDPIVIYQIMDDLIVGSDLEIGQHRTK	s	ik-Em	356
CONSENSUS_01_AE	V	Y-G	Y-AK-E	354
01_AE_CF_90CF11697		G	AR-E	356
01_AE_CF_90CF402		H-SX	AK-E	356
01_AE_CF_90CF4071			R	356
01_AE_CN_97CNGX2F		S	IK-EM	356
01_AE_JP_93JP-NH1		S	IK-EM	354
01_AE_TH_93TH253		S	IK-EM	348
01_AE_TH_94TH702		S	GIK-EM	356
01_AE_TH_95TH022		R	TK-E	356
01_AE_TH_95TH047		S	IK-EM	356
01_AE_TH_CM240		S	IK-EM	361
CONSENSUS_02_AG		k	tk-E	356
02_AG_CM_97CM-MP807		G	TK-E	356
02_AG_CM_CM52885		K	TK-E	356
02_AG_FR_DJ264		G	IK-E	356
02_AG_GH_A3		KE	IK-E	356
02_AG_GH_AB052867		K	TK-EL	361
02_AG_GH_AB052867		K	AN-E	356
02_AG_GH_G829		K	Y-IK-E	356
02_AG_NG_IBNG		K	TK-E	356
02_AG_SE_SE7812		K	TK-E	356
02_AG_SN_98SE-MP1211		I	TK-EL	356
03_AB_RU_KAL153-2		O	TK-EL	356
03_AB_RU_RU98001		L	O	356
04_CPX_CV_94CX032-3		BE	T	355
04_CPX_GR_97FVCH		PA	T	355
04_CPX_GR_97FVMY		PE	C	355
05_DF_BE_VI1310		I	KE	356
05_DF_BE_VI961		H	F	356
06_CPX_AU_BFP90		I	K	358
06_CPX_ML_95ML127		I	K	355
06_CPX_SN_97SE1078		K	I	360
07_BC_CN_97CN001		I	Y	356
07_BC_CN_98CN009		R	Y	356
08_BC_CN_97CNGX6F		K	V	352
08_BC_CN_97CNGX7F		K	V	352
08_BC_CN_98CNG006		K	V	352
10_CD_TZ_96TZBF061		Y	G	355
10_CD_TZ_96TZBF071		R	R	355
10_CD_TZ_96TZBF110		Y	Y	353
11_CPX_CM_MP818		N	V	357
11_CPX_FR_MP1298		KE	T	355
11_CPX_GR_GR17		L	S	355
12_BF_AR_ARMAL159		A	CV	356
12_BF_UY_URTR23		KE	V	356
12_BF_UY_URTR35		K	V	362
N_CM_YBF106		Q	X	363
N_CM_YBF30		Q	K	363
CONSENSUS_O		G	O	345
O_CM_ANT70		G	O	352
O_CM_MVP5180		G	O	352
O_SN_SE42HALD		G	O	352
O_SN_SEMP1299		G	O	352
O_SN_SEMP1300		G	O	352
CONSENSUS_CPZ		C	??	335
CPZ_CD_CPZANT		I	O	356
CPZ_CM_CAM3		C	N	361
CPZ_CM_CAM5		C	N	357
CPZ_GA_CPZGAB		C	K	354
CPZ_US_CPZUS		R		358

HIV-1/SIVcpz proteins

T215Y	K219Q	IEELRQHLRWGLTTED.KKHOKEPFLWMGYELHPDKWTYQPIVLPEKDSWTWVNDIQKLVGKLNWASQIYPGIKVQRQKLLRGTKALITFVPIPTTEAELELAENRELLKPEVHGVIYDPSKDLIAEIQ	485
B.FR.HXB2		a-s-f-e-e-k-a-DIVT-d	484
CONSENSUS A-A1		E-S-P-V-E-A-DIVT-D-A-V	483
A.SE.SE7535		A-S-E-A-DIVT-D-A-V	485
A.SE.SE8538		A-S-E-A-DIVT-D-A-V	479
A.SE.UGSE8131		A-S-E-A-DIVT-D-A-V	485
A1.KE.O23-17		A-S-I-E-A-DIVT-D-A-V	485
A1.SE.SET253		A-S-F-E-E-A-DIVT-D-A-V	485
A1.UG.92UG037		A-S-F-E-E-A-DIVT-D-A-V	484
A1.UG.U455		A-S-FI-Q-A-K-A-DIVT-D-A-V	477
A2.CD.97CDKFE4		A-P-K-K-R-A-DIVT-K-E-N	485
A2.CY.94CY017.41		A-K-PY-K-A-K-R-A-DIVT-K-E-T	485
CONSENSUS B		f-f-l-m-l-a-k-a	484
B.CN.RL42		F-F-L-M-L-A-KE-A-K-S	485
B.DE.D31		F-F-T-M-T-A-K-A	485
B.GA.OYI		F-F-M-M-A-KN-A-K-L	485
B.GB.CAM1		F-F-M-M-A-KN-A-K-L	485
B.KR.WK		F-F-X-X-A-K-A-V-K-Q-A-V	486
B.ITW.TWCYS		F-F-X-X-A-K-A-V-K-Q-A-V	485
B.US.JRFL		F-F-F-F-A-K-A-L	485
B.US.MNCG		F-F-F-F-A-K-A-L	488
B.US.RF		F-F-F-F-A-K-A-L	484
B.US.SF2		F-F-F-F-A-K-A-L	485
CONSENSUS C		e-k-f-q-o-a-DIV	481
C.BR.92BR025		E-K-F-Q-O-A-DIV	483
C.BR.98BR004		E-K-F-Q-O-A-DIV	481
C.BW.96BW01B03		E-K-F-Q-O-A-DIV	481
C.BW.96BW0402		E-K-F-Q-O-A-DIV	481
C.ET.ETH2220		E-K-F-Q-O-A-DIV	492
C.IL.981S002		E-K-F-Q-O-A-DIV	485
C.IN.93IN999		E-K-F-Q-O-A-DIV	490
C.IN.95IN21068		E-K-F-Q-O-A-DIV	480
C.IN.98IN012		E-K-F-Q-O-A-DIV	481
C.TZ.98TZ013		E-K-F-Q-O-A-DIV	488
C.TZ.98TZ017		E-K-F-Q-O-A-DIV	487
C.ZA.97ZA012		E-K-F-Q-O-A-DIV	483
C.ZM.96ZM651		E-K-F-Q-O-A-DIV	481
C.ZM.96ZM751		E-K-F-Q-O-A-DIV	487
CONSENSUS D		s-k-e-s-t-e-k-k	483
D.CD.84ZR085		S-K-E-S-T-E-K-K	486
D.CD.FLI		S-K-E-S-T-E-K-K	484
D.CD.NDK		S-K-E-S-T-E-K-K	484
D.CD.72Z6		S-K-E-S-T-E-K-K	484
D.UG.94UG1141		S-K-E-S-T-E-K-K	486
CONSENSUS F1		o-d-o-n-o-d-x-ik-a-DIV	478
F1.BE.VI850		O-D-O-N-O-D-X-IK-A-DIV	485
F1.BR.93BR020.1		O-D-O-N-O-D-X-IK-A-DIV	485
F1.FI.FIN9363		O-D-O-N-O-D-X-IK-A-DIV	484
F1.FR.MF411		O-D-O-N-O-D-X-IK-A-DIV	487
F2.CM.CM53657		A-Q-D-S-R-KH-R-A-D-V-V-M	485
F2.CM.MP255		A-Q-D-S-R-KH-R-A-D-V-V-M	485
F2.CM.MP257		A-Q-D-S-R-KH-R-A-D-V-V-M	485
CONSENSUS G		o-n-e-q-d-e-vk-k-i-tvk	478
G.BE.DRCBL		O-N-E-Q-D-E-VK-K-I-TVK	482
G.FI.HH8793-1.1		O-N-E-Q-D-E-VK-K-I-TVK	484
G.NG.92NG083		O-N-E-Q-D-E-VK-K-I-TVK	484
G.SE.SE6165		O-N-E-Q-D-E-VK-K-I-TVK	484
H.BE.VI991		A-F-P-Q-VK-K-X-A-IV-K-A	483
H.BE.VI997		A-F-P-Q-VK-K-X-A-IV-K-A	483
H.CF.90CF056		A-F-P-Q-VK-K-X-A-IV-K-A	484
J.SE.SET022		E-K-FY-Q-ED-IKE-I-A-DIV-R-K	483
J.SE.SET887		E-K-FY-Q-ED-IKE-I-A-DIV-R-K	483
K.CD.E0TB11C		E-K-FY-Q-ED-IKE-I-A-DIV-R-K	485
K.CM.MP535		E-K-FY-Q-ED-IKE-I-A-DIV-R-K	485

	T215Y	K219Q	
B.FR.HXB2	IEELRQHLRWGLTTPD.KKHQKEPFLWMGYELHPDKWTYQPIVLPKSDSWTNDIQKLVGKLNWASQIYPGIKVRQJLCKLRGKTKALITFVPLTEFAELELAENRELLKPEVHGVIYDPSKDLIAEIQ		
CONSENSUS_01.AE	a-s-f	r-e	a-k-a-div
01.AE.CF.90CF11697	D-K-P	E	A-K-A-V
01.AE.CF.90CF402	A-S-P	R	A-A-DIVT
01.AE.CF.90CF4071	S-S-P	R	A-A-DIV
01.AE.CN.97CNGX2F	S-S-P	R	A-K-A-DIV
01.AE.JP.93JP-NH1	S-S-P	R	A-K-A-DIV
01.AE.TH.93TH253	S-S-P	R	A-K-A-DIV
01.AE.TH.94TH702	S-S-P	R	A-K-A-DIV
01.AE.TH.95TH022	S-S-P	R	A-K-A-DIV
01.AE.TH.95TH047	S-S-P	R	A-K-A-DIV
01.AE.TH.CM240	S-S-P	R	A-K-A-DIV
CONSENSUS_02.AG	e-k-f	q	a-k-a-divt
02.AG.CM.97CM-MP807	T-K-F	Q	A-K-A-DIVT
02.AG.CM.CM52885	E-K-F	Q	A-K-A-DIVT
02.AG.FR.DJ264	E-K-F	Q	FA-K-R-A-DIV
02.AG.GH.A3	E-K-F	Q	A-K-A-DIV
02.AG.GH.AB052867	N-K-F	K	A-K-A-DIVT
02.AG.GH.G829	E-K-F	M	A-K-A-DIVT
02.AG.NG.IBNG	E-K-F	VE	A-KE-A-DIVT
02.AG.SE.SE7812	E-K-F	E	A-KE-A-DIVT
02.AG.SN.98SE-MP1211	E-K-F	Q	V-OVR-A-K-A-T-DIVT
03.AB.RU.KAL153-2	E-F	A	A-A-A-DIV
03.AB.RU.RU98001	D-F	A	A-A-A-I-A-V
04.CPX.CV.94CX032-3	E-F	OPA	A-K-A-DIV
04.CPX.GR.97FVCH	E-FY	Q	A-K-A-DIV
04.CPX.GR.97FVNY	E-FF	Q	A-K-A-DIV
05.DF.BE.V11310	E-A-F	P	A-K-A-DIV
05.DF.BE.V1961	E-K-F	Q	N-K-A-DIV
06.CPX.AU.BFP90	E-K-F	Q	D-E-SI-A-DIV
06.CPX.ML.95ML127	E-K-F	Q	N-E-SI-A-DIV
06.CPX.SN.97SE1078	V-A-K-F	Q	D-A-DIV
07.BC.CN.97CN001	F	Q	H-A-DIV
07.BC.CN.98CN009	F	Q	A-DIV
08.BC.CN.97CNGX6F	E-K-F	Q	A-DIV
08.BC.CN.97CNGX7F	E-K-F	Q	A-DIV
08.BC.CN.98CN006	E-K-F	Q	A-DIV
10.CD.TZ.96TZBF061	G-K-F	A	L-Q-K-N-A-V-DIV
10.CD.TZ.96TZBF071	G-K-F	Q	K-Q-E-A-DIV
10.CD.TZ.96TZBF110	E-K-F	SQ	R-A-DIV
11.CPX.CM.MP818	K	Q	D-EC-A-DIV
11.CPX.FR.MP1298	K	TQ	D-E-S-IK-A-DIV
11.CPX.GR.GR17	K	Q	D-E-A-K-A-DIV
12.BF.AR.ARMA159	F		
12.BF.UY.URTR23	A-F		
12.BF.UY.URTR35	A-F		
N.CM.YBF106	G-K-F	K	EV-K-R-I-VTF-Q
N.CM.YBF30	D-K-F	K	V-R-K-I-VNF
CONSENSUS_O	L-E-YQ	F	Q-n-ev-I-S-V-Sk-E-k
O.CM.ANT70	L-E-YQ	F	S-Q-N-V-I-S-V-SR-E-R-O
O.CM.MVP5180	V-L-E-YQ	F	O-R-E-I-S-V-SK-E-K
O.SM.SE42HALD	V-L-E-YQ	F	A-Q-N-EV-I-S-V-SK-E-K
O.SN.SEMP1299	V-L-E-YQ	F	O-R-EE-I-A-S-IV-SK-E-K
O.SN.SEMP1300	V-L-E-YQ	F	Q-N-EE-I-S-S-V-SK-E-K
CONSENSUS_CPZ	L-t-f	q	q-el-Ik-Ik-d-Vm-g-m-e-d
CPZ.CD.CPZANT	K-OV	E	Q-K-P-D-IK-R-I-V-S-DIVOM-R
CPZ.CM.CAM3	L-T	F	Q-O-ET-K-IK-IK-DIVM-O-M-E-X
CPZ.CM.CAM5	L-T	F	Q-O-ET-S-K-IK-IK-VAM-O-M-E
CPZ.GA.CPZGAB	L-T	F	Q-O-ET-IK-IK-K-D-V-P
CPZ.US.CPZUS	L-Y-T	F	E-Q-EI-IK-A-VNF-H-M-E

HIV-1/SIVcpz proteins

Table with columns for protein identifiers (e.g., B.FR.HXB2, CONSENSUS A-A1), amino acid sequences, and alignment markers (p51\_end, Rnase H\_start). The table contains multiple rows of sequence alignments for various HIV-1 and SIVcpz strains.

B.FR.HXB2	KQGQGWYQIYQEPFKNLKTKGYARMGAHTNDYKQLTEAVOKLTTESIVWGKT.PKFKLPIQKETWETWTEYQWATWPEWFEVNTTPPLVKLWYQLEKEPIVGRGETFVVDGAANRETKLKGAGYVT	p51_end \ / Rnase H_start	614
CONSENSUS_01_AE	---d---k---S---R---V---a---I---m---d---S---		614
01_AE.CF.90CF411697	---D---K---S---R---A---V---VA---R---R---A---A---L---S---N---		612
01_AE.CF.90CF402	---D---K---S---R---A---V---VA---R---R---A---A---L---S---N---		614
01_AE.CF.90CF4071	---D---K---S---R---A---V---VA---R---R---A---A---L---S---N---		614
01_AE.CN.97CNGX2F	---D---K---S---R---V---VS---R---R---M---M---D---A---S---O---		614
01_AE.JP.93JP.NH1	---D---K---S---R---V---VA---R---R---M---M---D---A---S---O---		614
01_AE.TH.93TH253	---D---K---S---R---R---V---A---R---R---M---M---D---A---S---O---		612
01_AE.TH.94TH702	---D---K---S---R---R---V---A---R---R---M---M---D---A---S---O---		606
01_AE.TH.95TH022	---D---K---S---R---R---V---A---R---R---M---M---D---A---S---O---		614
01_AE.TH.95TH047	---D---K---S---R---R---V---A---R---R---M---M---D---A---S---O---		614
01_AE.TH.CW240	---D---R---S---R---V---VA---R---R---M---M---D---A---S---O---		619
CONSENSUS_02_AG	---D---k---s---r---v---va---I---m---d---S---		614
02_AG.CM.97CM.MP807	---D---K---S---R---V---VA---I---M---D---A---M---D---A---V---		614
02_AG.CM.CM52885	---D---K---S---R---V---VA---I---M---D---A---M---D---A---V---		614
02_AG.FR.DJ264	---D---K---S---R---V---VA---I---M---D---A---M---D---A---V---		614
02_AG.FR.A3	---D---K---S---R---V---VA---I---M---D---A---M---D---A---V---		614
02_AG.GH.AB052867	---D---K---S---R---V---VA---I---M---D---A---M---D---A---V---		619
02_AG.GH.G829	---D---K---S---R---V---VA---I---M---D---A---M---D---A---V---		614
02_AG.NG.IBNG	---D---K---S---R---V---VA---I---M---D---A---M---D---A---V---		614
02_AG.SE.SE7812	---D---K---S---R---V---VA---I---M---D---A---M---D---A---V---		614
02_AG.SN.98SE.MP1211	---D---K---S---R---V---VA---I---M---D---A---M---D---A---V---		614
03_AB.RU.KAL153-2	---L---L---S---R---V---VA---I---M---D---A---M---D---A---V---		614
03_AB.RU.RU98001	---L---L---S---R---V---VA---I---M---D---A---M---D---A---V---		614
04_CPX.CY.94CY032-3	---H---T---S---R---V---VA---I---M---D---A---M---D---A---V---		613
04_CPX.GR.97PVCH	---Y---K---T---S---R---V---VA---I---M---D---A---M---D---A---V---		612
04_CPX.GR.97PVNY	---Y---K---T---S---R---V---VA---I---M---D---A---M---D---A---V---		613
05_DF.BE.VI1310	---L---A---V---A---A---A---I---M---D---A---M---D---A---V---		614
05_DF.BE.VI1961	---L---A---V---A---A---A---I---M---D---A---M---D---A---V---		614
06_CPX.AU.BFP90	---Y---H---I---K---S---R---V---VA---I---M---D---A---M---D---A---V---		616
06_CPX.ML.95ML127	---Y---H---I---K---S---R---V---VA---I---M---D---A---M---D---A---V---		613
06_CPX.SN.97SE1078	---D---H---T---S---R---V---VA---I---M---D---A---M---D---A---V---		618
07_BC.CN.97CN001	---E---K---T---R---V---VA---I---M---D---A---M---D---A---V---		614
07_BC.CN.98CN009	---D---K---T---R---V---VA---I---M---D---A---M---D---A---V---		614
08_BC.CN.97CNGX6F	---D---K---T---R---V---VA---I---M---D---A---M---D---A---V---		610
08_BC.CN.97CNGX7F	---D---K---T---R---V---VA---I---M---D---A---M---D---A---V---		610
08_BC.CN.98CN006	---R---D---K---T---R---V---VA---I---M---D---A---M---D---A---V---		610
10_CD.TZ.96TZBF061	---D---H---K---T---R---V---VA---I---M---D---A---M---D---A---V---		613
10_CD.TZ.96TZBF071	---D---H---K---T---R---V---VA---I---M---D---A---M---D---A---V---		614
10_CD.TZ.96TZBF110	---D---H---K---T---R---V---VA---I---M---D---A---M---D---A---V---		611
11_CPX.CM.MP818	---L---D---K---R---A---R---V---VA---I---M---D---A---M---D---A---V---		615
11_CPX.FR.MP1298	---L---D---K---R---A---R---V---VA---I---M---D---A---M---D---A---V---		613
11_CPX.GR.GR17	---L---D---K---R---A---R---V---VA---I---M---D---A---M---D---A---V---		613
12_BF.AR.ARMAL159	---F---L---D---K---R---A---R---V---VA---I---M---D---A---M---D---A---V---		614
12_BF.UY.URTR23	---L---D---K---R---A---R---V---VA---I---M---D---A---M---D---A---V---		614
12_BF.UY.URTR35	---L---D---K---R---A---R---V---VA---I---M---D---A---M---D---A---V---		620
N.CM.YBF106	---Y---L---H---K---S---R---V---VA---I---M---D---A---M---D---A---V---		621
N.CM.YBF30	---L---H---K---S---R---V---VA---I---M---D---A---M---D---A---V---		621
CONSENSUS_0	---g---t---o---k---a---s---r---a---v---l---v---s---o---a---l---L---L---V---V---A---D---H---		603
O.CM.ANT70	---G---T---O---K---A---S---R---A---V---L---V---S---O---A---I---L---L---V---V---A---D---H---		610
O.CM.MVP5180	---H---E---V---D---E---H---O---K---A---S---R---A---V---L---V---S---O---A---I---L---L---V---V---A---D---H---		610
O.SN.SE42HALD	---G---D---E---H---O---K---A---S---R---A---V---L---V---S---O---A---I---L---L---V---V---A---D---H---		610
O.SN.SEMP1299	---W---O---S---K---A---S---R---A---V---L---V---S---O---A---I---L---L---V---V---A---D---H---		610
O.SN.SEMP1300	---K---O---S---D---E---H---O---K---A---S---R---A---V---L---V---S---O---A---I---L---L---V---V---A---D---H---		610
CONSENSUS_CPZ	---ns---f---q---h---r---e---r---a---v---a---g---l---E---V---Q---T---R---D---A---S---D---		590
CPZ.CD.CPZANT	---S---N---E---G---L---A---P---T---E---R---A---V---A---A---G---L---E---V---Q---T---R---D---A---S---D---		614
CPZ.CM.CAM3	---NS---F---O---H---O---S---R---A---V---A---A---A---A---R---V---V---S---D---X---		619
CPZ.CM.CAM5	---NS---F---O---H---O---S---R---A---V---A---A---A---A---R---V---V---S---D---X---		615
CPZ.GA.CPZGAB	---NC---F---H---H---O---S---R---A---V---A---A---A---A---R---V---V---S---A---A---I---		612
CPZ.US.CPZUS	---RS---F---R---H---H---O---S---R---A---V---A---A---A---A---R---V---V---S---A---A---I---		616

B. FR. HXB2	NRGRQKVVLTDTTQKTELOAIYLALQDSGLEVNIVTDSQYALGIQAQDPQSESELVNIQIIEQLIKKVKVIWVPAHKIGGNEQDKLVSA. GTRKVLFLDGIKDAQD. EHEKY. HSNWFRAMASDF	741
CONSENSUS A-A1	d-----S-E-E-H-h-----S-----I-----K-e-d-----S-----E-----I-----	740
A. SE. SE7535	D-----S-E-E-H-L-S-----S-----R-----K-G-D-----S-----R-----E-----T-----	739
A. SE. SE8538	D-----S-E-E-VH-H-S-----S-----K-G-D-----S-----K-----E-----R-----	741
A. SE. UGSE8131	D-----S-E-E-H-H-S-----S-----K-----K-E-D-----S-----E-----T-----	735
A1. KE. O23-17	DK-----S-E-E-H-H-S-----S-----K-----I-----K-E-D-----S-----E-----T-----	741
A1. SE. SE7253	D-----S-E-E-H-H-S-----S-----S-----I-----K-G-D-----S-----E-----R-----	741
AI. UG. 92UG037	D-----S-E-E-H-H-S-----S-----R-----I-----K-E-D-----S-----E-----R-----	741
AI. UG. U455	D-----S-E-E-H-H-S-----S-----R-----I-----K-E-D-----S-----E-----D-----C-----	740
A2. CD. 97CDKFE4	D-----I-S-E-E-H-H-----R-----R-----K-E-R-----S-----E-----R-----O-----	733
A2. CY. 94CY017.41	D-----I-S-E-E-H-H-----E-R-----I-----K-E-R-----S-----E-----R-----H-----	741
CONSENSUS B	d-----s-----h-----V-----k-----k-----S-----e-----e-----	740
B. CN. RL42	-K-----S-----H-----S-----I-----K-----K-----S-----E-----C-----	741
B. DE. D31	D-----S-----H-----S-----I-----K-----S-----T-----K-----E-----	741
B. GA. OYI	D-----S-----H-----S-----K-----H-----H-----K-----E-----	741
B. GB. CAM1	D-----K-----P-----P-----K-----K-----S-----E-----E-----	741
B. KR. WK	D-----D-----K-----S-----S-----E-----E-----E-----G-----	741
B. TW. TWCYS	DK-----S-----S-----H-----H-----K-----K-----S-----E-----T-----R-----	742
B. US. JREFL	------S-----S-----H-----H-----K-----K-----S-----E-----E-----D-----K-----	741
B. US. MNCG	------S-----S-----H-----H-----K-----K-----S-----S-----E-----D-----	743
B. US. RF	D-----S-----S-----H-----H-----K-----K-----S-----R-----T-----	740
B. US. SF2	D-----S-----S-----H-----H-----K-----K-----S-----N-----E-----	741
CONSENSUS C	D-----I-s-E-E-g-s-----g-----I-s-s-----I-s-s-----e-----e-----	737
C. BR. 92BR025	D-----I-ST-E-Q-----Q-----S-----K-----R-----S-----E-----N-----E-----	738
C. BR. 98BR004	D-----I-S-E-Q-----Q-----S-----K-----K-----R-----S-----E-----E-----	737
C. BW. 96BW01B03	D-----RI-S-E-Q-----Q-----S-----K-----K-----R-----S-----E-----N-----	737
C. ET. ETH2220	DK-----I-P-E-E-RA-----Q-----S-----L-----K-----T-----R-----S-----E-----C-----	737
C. IL. 98IS002	D-----I-S-E-E-Q-----Q-----S-----L-----K-----I-----S-----R-----S-----E-----NE-----	748
C. IN. 93IN999	D-----I-P-E-Q-----Q-----S-----K-----K-----I-----R-----S-----E-----D-----	741
C. IN. 95IN21068	D-----I-S-E-E-C-----S-----S-----K-----K-----R-----S-----E-----R-----	746
C. IN. 98IN012	DK-N-I-S-E-E-C-----S-----S-----K-----K-----G-----R-----S-----E-----R-----	737
C. TZ. 98TZ013	D-----II-AE-A-Q-----Q-----SK-A-----K-----S-----G-----R-----S-----E-----E-----	744
C. TZ. 98TZ017	D-----I-S-E-S-O-----T-----S-----I-----K-----S-----O-----T-----S-----E-----N-----	743
C. ZA. 97ZA012	D-----I-S-NE-A-Q-----Q-----S-----I-----K-----G-----RI-S-----E-----D-----	739
C. ZM. 96ZM651	D-----I-E-E-A-Q-----Q-----S-----H-----K-----R-----S-----E-----N-----E-----	737
C. ZM. 96ZM751	D-----I-E-E-Q-----Q-----S-----K-----R-----S-----S-----E-----E-----	743
CONSENSUS D	D-----PF-----N-----K-----K-----s-----s-----g-----g-----D-----D-----	738
D. CD. 84ZR085	D-----P-----N-----N-----K-----S-----S-----K-----S-----Q-----E-----	742
D. CD. ELI	D-----P-----N-----N-----K-----K-----P-----Q-----S-----O-----E-----	740
D. CD. NDK	D-----P-----N-----N-----K-----K-----P-----K-----S-----O-----N-----	740
D. CD. Z2Z6	D-----P-----N-----N-----K-----K-----P-----K-----S-----O-----N-----	740
D. UG. 94UG1141	D-----S-----N-----N-----K-----K-----V-----S-----N-----I-----E-----	742
CONSENSUS F1	D-----?-----S-----E-----A-----h-----S-----K-----K-----?-----g-----?-----S-----E-----N-----	729
F1. BE. V1850	DK-K-----S-----E-----A-----A-----O-----S-----K-----I-----O-----R-----S-----V-----I-----E-----	742
F1. BR. 93BR020.1	D-----A-S-E-A-----A-----O-----S-----K-----K-----S-----O-----I-----S-----E-----N-----	741
F1. FI. FIN9363	D-----S-----E-----A-----A-----H-----S-----K-----K-----S-----Q-----I-----S-----E-----N-----	740
F1. FR. MP411	D-----A-I-E-A-----A-----H-----S-----V-----K-----I-----K-----Q-----R-----S-----I-----R-----E-----	743
F2. CM. CM53657	D-----P-----E-----H-----H-----S-----S-----H-----K-----H-----K-----Q-----R-----S-----E-----	741
F2. CM. MP255	D-----P-----E-----H-----H-----S-----S-----H-----K-----K-----Q-----Q-----R-----S-----E-----	741
F2. CM. MP257	D-----S-----E-----Q-----Q-----S-----S-----H-----K-----I-----I-----Q-----R-----S-----E-----	741
CONSENSUS G	DK-k-II-E-E-a-?-----Q-----rs-----R-----R-----R-----S-----E-----I-----	729
G. BE. DRCLB	DK-----III-----E-----A-----H-----Q-----OS-----R-----R-----I-----S-----E-----R-----K-----	738
G. FI. HH8793-1.1	DK-K-II-E-E-Q-----Q-----S-----R-----R-----R-----S-----E-----D-----	740
G. NG. 92NG083	DK-K-II-E-E-A-----A-----H-----Q-----RP-----R-----G-----S-----R-----E-----R-----	740
G. SE. SE6165	DK-K-II-E-E-A-----A-----Q-----RS-----R-----A-----S-----R-----A-----R-----N-----	738
H. BE. V1991	D--K-I-S-E-E-----E-----P-----E-----P-----E-----F-----S-----V-----Q-----	739
H. BE. V1997	DK-K-----A-----E-----E-----E-----E-----E-----S-----S-----A-----R-----N-----E-----	738
H. CF. 90CF056	D--K-----S-----E-----S-----S-----K-----K-----S-----V-----V-----R-----N-----V-----	740
J. SE. SE7022	DK-----I-----H-----R-----I-----K-----K-----E-----E-----S-----E-----E-----	739
J. SE. SE7887	DK-----I-----H-----R-----I-----K-----K-----E-----E-----S-----E-----D-----	739
K. CD. E0TB11C	DK-----ISI-E-----E-----H-----C-----S-----K-----K-----D-----R-----S-----EX-----N-----	742
K. CM. MP535	D-----SI-E-----A-----C-----S-----S-----K-----K-----D-----RI-S-----E-----N-----	741



B.FR.HXB	NLPPVV.AKEIVASCDKQLKGEAMHGQVDCSPGIWQLDCTHLECKVLLVAHVHVASGYIEAEVTPAETGQETAYFLKLA.GRWPYKTIHTDNGSNFTGATVRAACWV.AGLIKQEFGIPTNPQSQGVVES	868
CONSENSUS_A-A1	.....VV--S-A-K--hVq	867
A.SE.SE7535	.....I--S-A-K--NV	866
A.SE.SE8538	.....M--S-A-K--NVO-Y	868
A.SE.UGSE8131	.....A--S-A-K--NVO	862
A1.KE.O23-17	.....V--S-A-K--N-O	868
A1.SE.SE7253	.....IV--S-A-K--S-O	868
A1.UG.92UG037	.....VV--S-A-K--NV	868
A1.UG.U455	.....V--S-A-K-V--N-Q	867
A2.CD.97CDKPE4	.....I--V--P-S-K-K--VQ	860
A2.CY.94CY017.41	.....T--D--I--P--P--IS--K--Q	868
CONSENSUS_B	.....I--St-K--	867
B.CN.RL42	.....I--R--SNS-K--	868
B.DE.D31	.....I--ST-K--V	868
B.GA.OYI	.....I--ST-K--	868
B.GB.CAM1	.....-G--IST-K--	868
B.KR.WK	.....C--SN-K-K--S--R--K--I	867
B.TW.TWCYS	.....I--S-A-K--K--	869
B.US.JREL	.....I--ST-K--	868
B.US.MCG	.....I--P--ST-K--T	870
B.US.RF	.....I--ST-K--	867
B.US.SF2	.....I--ST-K--	868
CONSENSUS_C	.....I--S-a-K--g	864
C.BR.92BR025	.....I--T--I--V--ISN-K--Q	865
C.BR.98BR004	.....I--I--V--SN-K--Q	864
C.BW.96BW01B03	.....I--G--M--V--T--S-A-K--Q	864
C.BW.96BW0402	.....I--YI--V--STA-K--Q	864
C.ET.ETH2220	.....I--P--C--I--N--RV--SNA-K--Q	875
C.IL.98IS002	.....I--YI--V--P--S-A-K--Q	868
C.IN.93IN999	.....I--Q--Q--T--YI--V--S-A-K--Q	873
C.IN.95IN21068	.....I--M--I--V--S-A-K--Q	863
C.IN.98IN012	.....I--NQ--Q--YT-R--I--SN-A-K--Q	864
C.TZ.98TZ013	.....I--I--T--M--SYI--V--SSA-K--Q	870
C.TZ.98TZ017	.....I--I--I--M--I--S-A-K--Q	870
C.ZA.97ZA012	.....I--I--V--M--G--YI--V--S-A-K--Q	866
C.ZM.96ZM651	.....I--Q--T--YI--V--S-A-K--Q	864
C.ZM.96ZM751	.....I--I--T--L--I--V--S-A-K--H	870
CONSENSUS_D	.....I--S-a-K--	865
D.CD.84ZR085	.....I--V--S-A-K--	869
D.CD.ELI	.....I--A--V--S-A-K--	867
D.CD.NDK	.....I--V--S-A-K--	867
D.CD.Z2Z6	.....I--IV--S-A-K--	867
D.UG.94UG1141	.....I--V--G--V--S-A-K--	869
CONSENSUS_F1	.....I--S-a-K--Q	856
F1.BE.VI850	.....I--E--I--S-A-K-S--Q	869
F1.BR.93BR020.1	.....I--L--I--T--S-K--Q	868
F1.FI.FIN9363	.....I--D--I--M--S-A-K--Q	867
F1.FR.MP411	.....I--I--I--SSA-K--Q	870
F2.CM.CM53657	.....I--I--I--STV-K--Q	868
F2.CM.MP255	.....I--I--I--STV-K--Q	868
F2.CM.MP257	.....I--A--I--I--S-V-K--Q	868
CONSENSUS_G	.....I--S-A-K--h-T	855
G.BE.DRCBL	.....I--I--I--V--S-A-K--S-T	865
G.FI.HH8793-1.1	.....I--H--I--V--S-A-K--D-T	867
G.NG.92NG083	.....I--I--I--V--P--IS-A-K--N-T	867
G.SE.SEG165	.....I--I--I--TV--S-A-K--N-T	865
H.BE.VI991	.....I--I--M--S-A-K--D-H	866
H.BE.VI997	.....I--I--XP--M--T--STA-K--D-Q-D	866
H.CF.90CF056	.....I--I--Q--K--S--V--S-A-K--D-Q	867
J.SE.SE7022	.....I--A-F-I--G--A--SGA-K--D	866
J.SE.SE7887	.....I--A-F-I--V--SGA-K--D	866
K.CD.EQTB11C	.....I--I--I--RV--S-V-K--D	869
K.CM.MP535	.....I--I--I--T--STV-K--V	868



B.FR.HXB	NLPVV.AKEIVASCDKCOLKGEAMHGQVDCSPGIWQLDCTHLECKVLLVAHVHVASGYIEAEVIPAETGQETAYFLKLA.GRMPVKTIHTDNGSNFTGATVRAACWV.AGLIKQEFGLIPYNPQSQGVVES	868
CONSENSUS_01_AE	.....n.....V.....S-a-k.....NVR	868
01_AE.CF.90CF11697	.....N.....S-AMK.....NVO	865
01_AE.CF.90CF402	.....RV.....S-A-K.....NVO	868
01_AE.CF.90CF4071	.....N.....S-K.....NVO	868
01_AE.CN.97CNGX2F	.....V.....S-A-K.....NVR	868
01_AE.JP.93JP-NH1	.....V.....S-A-K.....NVR	866
01_AE.TH.93TH953	.....N.....S-A-K.....NVR	860
01_AE.TH.94TH702	.....PX.....S-A-K.....NVR	868
01_AE.TH.95TH022	.....V.....S-A-K.....NVR	868
01_AE.TH.95TH047	.....V.....S-A-K.....NVR	868
01_AE.TH.CM240	.....TN.....S-A-K.....NVO	873
CONSENSUS_02_AG	.....i.....I.....?.....S-a-k.....NVT	868
02_AG.CM.97CM-MP807	.....I.....I-M.....V.....S-A-K.....NVT	868
02_AG.CM.CM52885	.....I.....-G.....S.....S-A-K.....NVT	868
02_AG.FR.DJ264	.....I.....C.....I.....S-A-K.....NVT	868
02_AG.GH.A3	.....I.....I.....I.....S-A-K.....DVT	868
02_AG.GH.AB052867	.....I.....I.....E.....R.....S-A-K.....TNVT	873
02_AG.GH.G829	.....I.....I.....I.....S-A-K.....NVT	868
02_AG.NG.IBNG	.....I.....M.....G.....S-A-K.....NVT	868
02_AG.SE.SE7812	.....I.....I.....I.....S-A-K.....NVT	868
02_AG.SN.98SE-MP1211	.....I.....I.....I.....X.....S-A-K.....XPKLH-L-X	870
03_AB.RU.KAL153-2	.....I.....V.....I.....STA-K.....	868
03_AB.RU.RU98001	.....I.....V.....I.....ISTA-K.....	868
04_CFX.CX.94CX032-3	.....S.....N.....M.....A.....P.....S-A-K.....D-N	867
04_CFX.GR.97FXGH	.....I.....X.....I.....P.....S-A-K.....DVO	866
04_CFX.GR.97FXVNY	.....I.....N.....I.....S-A-K.....N-Q-V	867
05_DF.BE.V11310	.....I.....D.....I.....M.....P.....S-A-K.....Q	868
05_DF.BE.V1961	.....X.....I.....I.....MV.....S-A-K.....	869
06_CFX.AU.BFP90	.....I.....I.....I.....V.....S-A-K.....N-T	870
06_CFX.ML.95ML127	.....I.....I.....I.....N.....S-A-K.....N-T	867
06_CFX.SN.97SE1078	.....I.....P.....I.....V.....S-A-K.....NVT	872
07_BC.CN.97CN001	.....I.....Q.....I.....I.....M.....S-A-K.....Q	868
07_BC.CN.98CN009	.....I.....Q.....I.....I.....M.....S-A-K.....Q	868
08_BC.CN.97CNGX6F	.....I.....Q.....I.....I.....S-A-K.....Q	864
08_BC.CN.97CNGX7F	.....I.....Q.....I.....I.....S-A-K.....Q	864
08_BC.CN.98CN006	.....I.....Q.....I.....I.....M.....S-A-K.....Q	864
10_CD.TZ.96TZBF061	.....L.....L.....L.....V.....S-A-K.....	867
10_CD.TZ.96TZBF071	.....L.....L.....L.....V.....S-A-K.....	868
10_CD.TZ.96TZBF110	.....L.....L.....L.....V.....S-A-K.....	865
11_CFX.CM.MP818	.....I.....I-M.....T.....I.....V.....SS-K.....N-Q	869
11_CFX.FR.MP1298	.....I.....N.....R.....I.....R.....M.....S-A-K.....N-Q	867
11_CFX.GR.GR17	.....I.....N.....I.....I.....SNA-K.....Q	867
12_BF.AR.ARM159	.....I.....V.....L.....L.....I.....P.....S-M-K.....Q	868
12_BF.UY.URTR23	.....I.....V.....L.....L.....I.....T.....SS-A-K.....Q	868
12_BF.UY.URTR35	.....I.....IN.....V.....L.....L.....I.....R.....P.....SS-A-K.....V-Q	874
N.CM.YBF106	.....I.....N.....V.....L.....L.....I.....V.....P.....S-K.....-A	875
N.CM.YBF30	.....I.....N.....V.....L.....L.....I.....V.....S-K.....-A	875
CONSENSUS_0	.....G.....I-n-P-HI.....I.....F.....A.....V.....P.....S-MK.....tn-qH	857
O.CM.ANT70	.....G.....I-P-HI.....I.....M.....I.....F.....A.....V.....P.....S-MK.....N-OH	864
O.CM.MVP5180	.....G.....I-P-HI.....I.....Y-E-M.....I.....DF.....A.....V.....P.....S-MK.....T-OH	864
O.SN.SE42HALD	.....G.....I-N-P-HI.....I.....EV-I.....F.....A.....R.....P.....S-MK.....TN-OH	864
O.SN.SEMP1299	.....G.....I-N-PO-HI.....I.....EV-M.....I.....F.....A.....V.....P.....S-MK.....N-H	864
O.SN.SEMP1300	.....G.....I-N-P-HI.....I.....Y-EV-I.....I.....F.....A.....I.....P.....S-MK.....T-H	864
CONSENSUS_CPZ	.....I.....I.....?.....I.....I.....V.....P.....S?.....k.....q	841
CPZ.CD.CPZANT	.....I-Q.....HV.....R.....V.....I.....S-F.....MAD-KS.....N-Q	868
CPZ.CM.CAM3	.....I-HI.....L-R.....L.....L.....L.....V.....P.....SN-K.....Q	875
CPZ.CM.CAM5	.....I.....H.....V.....I.....L.....L.....V.....P.....SN-K.....Q	869
CPZ.GA.CPZGAB	.....I.....H.....V.....I.....V.....P.....S-A-K.....D	866
CPZ.US.CPZUS	.....I.....P.....I.....I.....L.....L.....V.....P.....SS.....Q	870

B. FR. HXB2	MNKLKLIIGVDDQAEHLKTAVQMAVFTHNPKRGG.IGGYSAGERIVDIIATDIOTKELQKIT.KIQNFRVYVYRSDSNLNPKGPAKLLMKGBGAVVIQDNSDIKVVPRRKAIIIRDYCKQWAGDDCV	/ vif CDS start	996
CONSENSUS A-A1	I-D-I		995
A. SE. SE7535	I-D-I		994
A. SE. SE8538	I-D-I	V	994
A. SE. UGSE8131	I-D-I		990
A1. KE. O23-17	I-D-I	Q	996
A1. SE. SE7253	I-D-I	F	996
A1. UG. 92UG037	I-D-I	N	996
A1. UG. U455	I-S-L	V-K	996
A2. CD. 97CDKFE4	I-D-I		995
A2. CY. 94CY017.41	I-D-I		996
CONSENSUS B	d		995
B. CN. RL42	G-D	V	996
B. DE. D31	D	E	996
B. GA. OYI	E		996
B. GB. CAM1	D		996
B. KR. WK	D		995
B. TW. TWCYS	N-D	V	998
B. US. JREFL	E		996
B. US. MNCG	D	N	998
B. US. RF	D	H	995
B. US. SF2	NKD		996
CONSENSUS C	i		991
C. BR. 92BR025	D-I	?-A-M	993
C. BR. 98BR004	I-D-I	L-V-K	992
C. BW. 96BW0402	I-D-I	V-K	992
C. ET. ETH2220	I-D-V	V	992
C. IL. 98IS002	I-D-I	A	1003
C. IN. 93IN999	I-D-I		996
C. IN. 95IN21068	I-H		1001
C. IN. 98IN012	I-V	K	991
C. TZ. 98TZ013	L-V	K	992
C. TZ. 98TZ017	I-D-I	A	998
C. ZA. 97ZA012	I-D-I	K	994
C. ZM. 96ZM651	MI	K	992
C. ZM. 96ZM751	I-N-K	A	992
CONSENSUS D	I-R	I	998
D. CD. 84ZR085	I-D-I	V	993
D. CD. ELI	I-S		997
D. CD. NDK	I-D-I	K	995
D. CD. Z2Z6	I-D-I	V	995
D. UG. 94UG1141	I-D-I	V	995
CONSENSUS F1	I-I	E	997
F1. BE. VI850	I-S	d-V	982
F1. BR. 93BR020.1	R	E-I	997
F1. FI. FIN9363	I-V	E	996
F1. FR. MP411	I-S	D-V	995
F2. CM. CM53657	I	D-V	998
F2. CM. MP255	I	D-V	996
F2. CM. MP257	I	F-D-V	996
CONSENSUS G	I-S	D?	982
G. BE. DRCLB	I-S	D-I	993
G. FI. HH8793-1.1	T	L-F-D-V	995
G. NG. 92NG083	I-S	I	995
G. SE. SE6165	I-S	D-V	993
H. BE. VI991	R	S-Q-K	994
H. BE. VI997	I	S-N-K	993
H. CF. 90CF056	I	S-N-K	995
J. SE. SE7022	I	D-I	994
J. SE. SE7887	I	R	994
K. CD. E0TB11C	I	L-N-K	996
K. CM. MP535	E-I	X-E	996

B.FR.HXB2	MNKLKLIIGVDPDQAEHLKTAVQMAVFIHNFPRKGG.IGGYSAGERIVDIIATDIOTKELQKIT.KIQFRVYVYRSDSNPLMKGPAKLLMKGEGAVVIQNSDIKVVVPRKAKIIRDYCKQWAGDDCV	/ vif CDS start	996
CONSENSUS_01_AE	-----I-----D-I-----		996
01_AE.CF.90CF411697	-----I-----D-I-----		993
01_AE.CF.90CF402	-----I-E-X-----D-I-----		996
01_AE.CF.90CF4071	-----I-A-----D-I-----		996
01_AE.CN.97CNGX2F	-----I-----D-I-----		996
01_AE.JP.93JP.NH1	-----I-----D-I-----		994
01_AE.TH.93TH253	-----I-----D-I-----		988
01_AE.TH.94TH702	-----I-----D-I-----I-----		996
01_AE.TH.95TH022	-----I-----D-I-----		996
01_AE.TH.95TH047	-----I-----D-I-----		996
01_AE.TH.CW240	-----I-----D-I-----		1001
CONSENSUS_02_AG	-----I-S-----D-I-----		996
02_AG.CM.97CM.MP807	-----I-S-----D-I-----V-----		996
02_AG.CM.CM52885	-----I-S-----D-I-----I-----		996
02_AG.FR.DJ264	-----I-S-----D-I-----K-----		996
02_AG.GH.A3	-----I-S-----D-I-----K-----		1001
02_AG.GH.AB052867	-----I-S-----D-I-----L-----		996
02_AG.GH.G829	-----I-S-----D-I-----L-----		996
02_AG.NG.IBNG	-----I-S-----D-I-----V-V-----		996
02_AG.SE.SE7812	-----I-S-----D-I-----V-V-----		998
02_AG.SN.98SE.MP1211	-----I-S-----N-K-----D-I-----T-L-----		996
03_AB.RU.KAL153-2	-----I-----D-I-----N-----		996
03_AB.RU.RU98001	-----I-----D-I-----N-----		996
04_CPX.CY.94CY032-3	-----I-S-----E-I-----N-----		995
04_CPX.GR.97PVCH	-----I-S-----D-I-----K-----		994
04_CPX.GR.97PVNY	-----I-S-----D-I-----K-----		995
05_DF.BE.VI1310	-----G-I-T-----I-----D-V-----E-----		996
05_DF.BE.VI961	-----G-I-S-----X-----D-V-----E-----		998
06_CPX.AU.BFP90	-----I-S-----D-I-----E-----		998
06_CPX.ML.95ML127	-----I-S-----D-I-----E-----		998
06_CPX.SN.97SE1078	-----I-S-----R-----E-----		1000
07_BC.CN.97CN001	-----I-----D-S-----K-----A-----		996
07_BC.CN.98CN009	-----I-----D-S-----K-----A-----		996
08_BC.CN.97CNGX6F	-----R-I-----D-I-----K-----A-----		992
08_BC.CN.97CNGX7F	-----R-I-----D-I-----K-----A-----		992
08_BC.CN.98CN006	-----R-I-----D-I-----K-----A-----		992
10_CD.TZ.96TZBF061	-----I-----G-I-----E-----V-K-----A-----		995
10_CD.TZ.96TZBF071	-----I-----S-----E-----V-K-----A-----		996
10_CD.TZ.96TZBF110	-----I-----D-I-----E-----V-K-----A-----		993
11_CPX.CM.MP818	-----L-----D-I-----		997
11_CPX.FR.MP1298	-----L-----D-I-----		995
11_CPX.GR.GR17	-----L-----D-I-----I-----		995
12_BF.AR.ARMA159	-----I-S-----R-----D-V-----		996
12_BF.UY.URTR23	-----I-S-----D-V-----		996
12_BF.UY.URTR35	-----I-----N-D-V-----E-----		1002
N.CM.YBF106	-----XT-----I-----TK-T-L-V-X-----D-I-----G-----		1003
N.CM.YBF30	-----T-----I-----TN-T-L-V-----D-I-----G-----		1003
CONSENSUS_0	-----I-L-SQ-----T-----f-----k-g-----D-I-----Q-----K-----G-----h-----T-SM		985
O.CM.ANT70	-----I-L-SQ-----T-----L-----XHK-----D-I-----Q-----K-----G-----E-----T-SM		992
O.CM.MVP5180	-----I-L-SQ-----T-----L-----N-----D-I-----Q-----K-----G-----E-----T-SM		992
O.SN.SE42HALD	-----I-L-SQ-----T-----F-----K-Q-----D-I-----Q-----K-----G-----H-----T-SM		992
O.SN.SEMP1299	-----I-L-SQ-----T-----F-----K-Q-----D-I-----Q-----K-----G-----H-----T-SM		992
O.SN.SEMP1300	-----I-L-SQ-----T-----F-----K-Q-----D-I-----Q-----K-----G-----H-----T-SM		992
CONSENSUS_CPZ	-----I-----L-----V-----tn-----L-----v-----ke?ev-----sm		968
CPZ.CD.CPZANT	-----TP-Q-----L-----L-----Q-----H-----D-V-----Q-----K-----QEE-----KE-R-KLEDR-DL		996
CPZ.CM.CAM3	-----I-----I-----TN-----L-----V-----D-I-----R-----KEREV-----I-----SM		1003
CPZ.CM.CAM5	-----I-----I-----TN-----L-----V-----D-I-----R-----KEREV-----I-----SM		997
CPZ.GA.CPZGAB	-----I-----I-----S-----L-----V-----K-----D-I-----T-----OGEL-----GSM		994
CPZ.US.CPZUS	-----I-----I-----SEL-DL-----L-----V-----D-I-----K-----KE-EV-----A-SM		998

B.FR.HXB2	ASRQDED\$	1003
CONSENSUS_A--A1	-g-----	1002
A.SE.SE7535	-G-----	1001
A.SE.SE8538	-G-----	1003
A.SE.UGSE8131	-G-----	997
A1.KE.O23-17	-G-----	1003
A1.SE.SE7253	-G-----	1003
A1.UG.92UG037	-G-----	1003
A1.UG.U455	-G-----	1002
A2.CD.97CDKFE4	-C-----	995
A2.CY.94CY017.41	-G-----	1003
CONSENSUS_B	-----	1002
B.CN.RL42	-----	1003
B.DE.D31	-G-----	1003
B.GA.OYI	-----	1003
B.GB.CAM1	-----	1003
B.KR.WK	-----	1002
B.TW.TWCYS	-G-----	1005
B.US.JRFL	-----	1003
B.US.MNCG	-----	1005
B.US.RF	-----	1002
B.US.SF2	-----	1003
CONSENSUS_C	-g-----Q	999
C.BR.92BR025	-----	1000
C.BR.98BR004	-----	999
C.BW.96BW01B03	-G-----	998
C.BW.96BW0402	-G-----	999
C.ET.ETH2220	-G-----	1010
C.IL.98IS002	-----Q	1004
C.IN.93IN999	-G-----	1004
C.IN.95IN21068	-G-----	1008
C.IN.98IN012	-G-----	998
C.TZ.98TZ013	-G-----	999
C.TZ.98TZ017	-G-----Q	1005
C.ZA.97ZA012	-G-----	1006
C.ZM.96ZM651	-G-----	1001
C.ZM.96ZM751	-G-----	999
CONSENSUS_D	-----	1005
D.CD.84ZR085	-----	1000
D.CD.ELI	-----	1004
D.CD.NDK	-----	1002
D.CD.Z2Z6	-----	1002
D.UG.94UG1141	-----	1002
CONSENSUS_F1	-G-----	1004
F1.BE.VI850	-G-----	989
F1.BR.93BR020.1	-G-----	1004
F1.FI.FIN9363	-G-----	1003
F1.FR.MP411	-G-----	1002
F2.CM.CM53657	-G-----	1005
F2.CM.MP255	-G-----	1003
F2.CM.MP257	-G-----	1003
CONSENSUS_G	-G-----	989
G.BE.DRCBL	-G-----	1000
G.FI.HH8793-1.1	-G-----	1000
G.NG.92NG083	-G-----	1002
G.SE.SE6165	-G-----	1002
H.BE.VI991	-G-----	1000
H.BE.VI997	-G-----	1001
H.CF.90CF056	-----	1000
J.SE.SE7022	-G-----	1002
J.SE.SE7887	-G-----	1001
K.CD.EQTB11C	-G-----	1001
K.CM.MP535	-G-----	1003
CONSENSUS_01_AE	-G-----	1003
O1_AE.CF.90CF11697	-G-----N	1000
O1_AE.CF.90CF402	-G-----N	1003
O1_AE.CF.90CF407	-G-----	1003
O1_AE.CN.97CNGX2F	-G-----	1003
O1_AE.JP.93JP-NH1	-G-----	1003
O1_AE.TH.93TH253	-G-----	1001
O1_AE.TH.94TH702	-G-----	995

```

01_AE.TH.95TH022      -G-----
01_AE.TH.95TH047      -G-----
01_AE.TH.CM240        -G-----
CONSENSUS_02_AG      -g-----
02_AG.CM.97CM-MP807   -G-----
02_AG.CM.CM52885      -G-----
02_AG.FR.DJ264        -G-----
02_AG.GH.A3           -G-----
02_AG.GH.AB052867     -D-----
02_AG.GH.G829         -G-----
02_AG.NG.IENG         -G-----
02_AG.SE.SE7812       -G-----
02_AG.SN.98SE-MP1211 -GG-N---
03_AB.RU.KAL153-2    -G-----
03_AB.RU.RU98001     -G-----
04_CPX.CY.94CY032-3  -G-----
04_CPX.GR.97PVCH      -G-----
04_CPX.GR.97PVMY      -G-----
05_DF.BE.VI1310      -G-----
05_DF.BE.VI961       -G-----
06_CPX.AU.BFP90       -G-----
06_CPX.ML.95MLJ127    -G-----
06_CPX.SN.97SE1078    -G-----
07_BC.CN.97CN001     -G-----
07_BC.CN.98CN009     -G-----
08_BC.CN.97CNGX6F    -G-----
08_BC.CN.97CNGX7F    -G-----
08_BC.CN.98CN006     -G-----
10_CD.TZ.96TZBF061   -----Q
10_CD.TZ.96TZBF071   -----Q
10_CD.TZ.96TZBF110   -----Q
11_CPX.CM.MP818       -G-----
11_CPX.FR.MP1298      -G-----
11_CPX.GR.GR17        -GS-----
12_BF.AR.ARMA159     -G-----
12_BF.UY.URTR23      -G-----
12_BF.UY.URTR35      -G-----
N.CM.YBF106          --GX--NQ
N.CM.YBF30           --G---NQ
CONSENSUS_0          --g-t-se
O.CM.ANT70           -g-t-se
O.CM.MVP5180         -n-t-se
O.SN.SE42HALD        -g-t-se
O.SN.SEMP1299        -g-t-se
O.SN.SEMP1300        -g-t-se
CONSENSUS_CPZ       -g-n--Q
CPZ.CD.CPZANT        -g-n--
CPZ.CM.CAM3          -GG---SQ
CPZ.CM.CAM5          -G--N-Q
CPZ.GA.CPZGAB        ---N--
CPZ.US.CPZUS        -G---S-

```









B.FR.HXB2	VSPRCEYOAGH.NKVGSIQVLAALALITP.KKIKKPP...LPSVTKLITEDRWNKPKQTKYGHGSSHTMNGH\$	/ Vpr CDS start	192
CONSENSUS_01_AE	-----K-T-I-I-I-----k-----k-----I-I-Enp-----		191
01_AE.CF.90CF11697	-----K-A-----PR-----R-----K-----R-----Enp-----		192
01_AE.CF.90CF402	-----K-TKT-----T-----R-----R-----E-----P-----		191
01_AE.CF.90CF4071	-----K-A-----R-----R-----K-----R-----Enp-----S-----		192
01_AE.CN.97CNGX11F	-----K-T-----R-----R-----K-----E-----IR-----Enp-----		192
01_AE.DE.K08DE	-----K-S-----TA-----R-----R-----K-----S-----IRD-----Enp-----		192
01_AE.JP.93JP-NH1	-----K-V-T-----R-----R-----K-----R-----Y-----Enp-----		191
01_AE.TH.95TH047	-----K-T-----R-----R-----K-----H-----GD-----Enp-----		191
01_AE.TH.CM235-4	-----K-T-----R-----R-----K-----R-----R-----Enp-----		192
01_AE.TH.CM240	-----K-T-----R-----R-----EI-----R-----Enp-----		190
CONSENSUS_02_AG	-----k-v-ttt-----k-a-a-e-e-r-----Rp-----W-----		193
02_AG.CM.97CM-MP807	-----k-VI-----TRTR-----k-----K-----R-----K-----RS-----W-----		193
02_AG.CM.CM52885	-----K-K-V-----T-----k-----K-----A-----E-----R-----K-----RP-----		192
02_AG.CM.CM53658	-----K-V-----RRT-----R-----R-----A-----E-----E-----RP-----		192
02_AG.FR.DJ264	-----O-----K-V-----V-----T-----k-----K-----A-----R-----NRS-----O-----		192
02_AG.GH.A3	-----SR-----K-V-----TRR-----R-----R-----A-----E-----R-----RPT-----		192
02_AG.GH.G829	-----K-V-----TK-----R-----R-----A-----E-----R-----RP-----R-----		192
02_AG.NG.IBNG	-----N-VA-----T-----T-----R-----A-----KE-----R-----RP-----		192
02_AG.SE.SE7812	-----K-V-----TRT-----R-----K-----A-----E-----R-----RS-----		192
02_AG.SN.MP1213	-----K-V-----IRR-----R-----K-----E-----R-----RS-----		192
03_AB.RU.KAL153-2	-----R-----R-----R-----R-----R-----D-----S-----		192
03_AB.RU.RU98001	-----R-----R-----R-----R-----R-----D-----S-----		192
04_Cpx.CY.94CY032-3	-----S-----T-----k-----K-----V-----N-----R-----R-----ENQI-----		191
04_Cpx.GR.97PVGH	-----S-----T-----k-----K-----V-----N-----R-----R-----ENQI-----		192
04_Cpx.GR.97PVNY	-----S-----T-----k-----K-----V-----N-----R-----R-----E-----I-----		192
05_DF.BE.VI1310	-----T-----T-----Q-----R-----R-----R-----R-----N-----Y-----		192
05_DF.BE.VI1961	-----T-----Q-----R-----R-----R-----R-----R-----C-----Y-----		192
06_Cpx.AU.BFP90	-----T-----K-----E-----R-----O-----V-----R-----E-----		192
06_Cpx.WI.95M1127	-----T-----K-----R-----Q-----V-----RD-----E-----		192
06_Cpx.SN.97SEL1078	-----K-----VKT-----RR-----Q-----V-----D-----E-----I-----D-----		192
07_BC.CN.97CN001	-----T-----K-----K-----IK-----V-----N-----E-----IR-----R-----N-----		193
07_BC.CN.98CN009	-----T-----K-----K-----IK-----V-----N-----IR-----R-----N-----		193
08_BC.CN.97CNGX6F	-----T-----K-----R-----IK-----V-----N-----IR-----R-----N-----		193
08_BC.CN.97CNGX7F	-----T-----K-----R-----IK-----V-----N-----IR-----R-----N-----		193
08_BC.CN.98CN006	-----T-----K-----R-----IK-----V-----N-----IR-----R-----N-----		192
10_CD.TZ.96TZBF061	-----T-----N-----KO-----T-----T-----R-----R-----		192
10_CD.TZ.96TZBF071	-----O-----Q-----T-----TR-----R-----R-----E-----R-----		192
10_CD.TZ.96TZBF110	-----O-----Q-----T-----T-----R-----R-----R-----E-----I-----		192
11_Cpx.CM.MP818	-----O-----K-----V-----TRA-----R-----A-----R-----N-----A-----C-----		192
11_Cpx.FR.MP1298	-----K-----VA-----TRT-----R-----R-----A-----R-----N-----L-----C-----		192
11_Cpx.FR.GR17	-----K-----V-----RT-----R-----R-----R-----R-----R-----N-----V-----C-----		192
12_BF.AR.ARMA159	-----T-----A-----T-----K-----V-----N-----R-----R-----		192
12_BF.UY.URTR23	-----T-----T-----T-----V-----V-----E-----R-----EN-----		192
12_BF.UY.URTR35	-----T-----A-----T-----Q-----V-----R-----R-----RD-----		192
N.CM.YBF106	-----XO-----T-----XT-----XVGV-----RR-----ER-----Q-----NPI-----P-----		192
N.CM.YBF30	-----KO-----T-----TWGA-----R-----EH-----MQ-----NPI-----		192
CONSENSUS_0	-----Sq-----T-----I-----Y-----VVK-----rd-----hlxIRDOle-----S-----W-----		187
O.CM.1422	-----S-----T-----L-----R-----VKA-----RSR-----Q-----K-----R-----LRIDPE-----S-----		192
O.CM.ANY70	-----SO-----T-----L-----R-----VKA-----RSR-----Q-----HLKIRDOLE-----S-----		192
O.CM.MD	-----SO-----T-----L-----R-----VKE-----RH-----Q-----HLKIRDOLE-----S-----		192
O.CM.MVP5180	-----SO-----T-----L-----R-----VVK-----RN-----QR-----W-----IRDOL-----S-----		192
O.GQ.1159	-----SO-----T-----L-----R-----VKA-----RRS-----Q-----L-----R-----HPIRDOL-----S-----		193
O.SN.SE42HALD	-----SO-----T-----L-----RVVKE-----RN-----Q-----SRYLIRDOLE-----S-----		192
O.SN.SEMP1299	-----SO-----T-----L-----RVVKE-----RN-----Q-----SRHLIRDOLE-----S-----		192
CONSENSUS_Cpz	-----Yk-----pt-----?-----?-----?-----?-----?-----?-----h-----x-----qEn-----		181
Cpz.CD.CPZANT	-----LITY-----H-----KK-----SQ-----T-----FKLLEFRGYP-----G-----RRQF-----LSI-----RRMR-----ENQ\$-----		192
Cpz.CM.CAM3	-----Y-----K-----P-----OO-----R-----WVR-----G-----K-----IX-----A-----RH-----IRV-----OE-----		192
Cpz.CM.CAM5	-----A-----KE-----RO-----F-----K-----SE-----RRHR-----A-----H-----R-----V-----QENL-----R-----		193
Cpz.GA.CPZGAB	-----I-----K-----PT-----F-----K-----VGQ-----S-----RR-----C-----OPEN-----S-----		192



B_FR_HXB2	MEQAPEDQG	POREPHNWTLELLELNKNEAVRHPFR	WLHGLGQHIYETVYGTWAGVEALIRLQQLLFIHFRIGCRHSRI	GV	...TRQRRAR	...NGASR\$S	/ Tat_CDS end
CONSENSUS_01_AE	Y	Y	Y	Y	Y	Y	ipg
01_AE_CF_90CF11697	Y	Y	Y	Y	Y	Y	ipg
01_AE_CF_90CF402	Y	Y	Y	Y	Y	Y	ipg
01_AE_CF_90CF4071	Y	Y	Y	Y	Y	Y	ipg
01_AE_CN_97CNGX11F	Y	Y	Y	Y	Y	Y	ipg
01_AE_CN_97CNGX2F	Y	Y	Y	Y	Y	Y	ipg
01_AE_JP_93JP-NH1	Y	Y	Y	Y	Y	Y	ipg
01_AE_TH_93TH902	Y	Y	Y	Y	Y	Y	ipg
01_AE_TH_94TH702	Y	Y	Y	Y	Y	Y	ipg
01_AE_TH_95TH022	Y	Y	Y	Y	Y	Y	ipg
01_AE_TH_CM235-2	Y	Y	Y	Y	Y	Y	ipg
01_AE_TH_CM240	Y	Y	Y	Y	Y	Y	ipg
CONSENSUS_02_AG	h	h	e	e	e	e	g
02_AG_CM_97CM-MP807	h	h	e	e	e	e	g
02_AG_CM_CM52885	h	h	e	e	e	e	g
02_AG_CM_CM53658	h	h	e	e	e	e	g
02_AG_FR_DJ263	h	h	e	e	e	e	g
02_AG_GH_A3	h	h	e	e	e	e	g
02_AG_GH_G829	h	h	e	e	e	e	g
02_AG_NG_IBNG	h	h	e	e	e	e	g
02_AG_SE_SE7812	h	h	e	e	e	e	g
02_AG_SN_98SE-MP1211	h	h	e	e	e	e	g
02_AG_SN_MP1213	h	h	e	e	e	e	g
03_AB_RU_KAL153-2	h	h	e	e	e	e	g
03_AB_RU_RU98001	h	h	e	e	e	e	g
04_CPX_CV_94CV032-3	h	h	e	e	e	e	g
04_CPX_GR_97PVGH	h	h	e	e	e	e	g
04_CPX_GR_97PVNY	h	h	e	e	e	e	g
05_DF_BE_VI1310	h	h	e	e	e	e	g
05_DF_BE_VI1961	h	h	e	e	e	e	g
06_CPX_AU_BFP90	h	h	e	e	e	e	g
06_CPX_WL_95WL84	h	h	e	e	e	e	g
06_CPX_SN_97SEL078	h	h	e	e	e	e	g
07_BC_CN_97CN001	h	h	e	e	e	e	g
07_BC_CN_98CN009	h	h	e	e	e	e	g
08_BC_CN_97CNGX6F	h	h	e	e	e	e	g
08_BC_CN_97CNGX7F	h	h	e	e	e	e	g
08_BC_CN_98CN006	h	h	e	e	e	e	g
10_CD_TZ_96TZBF061	h	h	e	e	e	e	g
10_CD_TZ_96TZBF071	h	h	e	e	e	e	g
11_CPX_CM_MP818	h	h	e	e	e	e	g
11_CPX_FR_MP1307	h	h	e	e	e	e	g
11_CPX_GR_GRI17	h	h	e	e	e	e	g
12_BF_AR_ARMAL159	h	h	e	e	e	e	g
12_BF_UY_URTR23	h	h	e	e	e	e	g
12_BF_UY_URTR35	h	h	e	e	e	e	g
N_CM_YBF106	h	h	e	e	e	e	g
N_CM_YBF30	h	h	e	e	e	e	g
CONSENSUS_O	h	h	e	e	e	e	g
O_CM_1422	h	h	e	e	e	e	g
O_CM_AN770	h	h	e	e	e	e	g
O_CM_MVP5180	h	h	e	e	e	e	g
O_FR_HIVV16019	h	h	e	e	e	e	g
O_GQ_11159	h	h	e	e	e	e	g
O_SN_SE42HALD	h	h	e	e	e	e	g
CPZ_CD_CPZANT	h	h	e	e	e	e	g
CPZ_CM_CAM3	h	h	e	e	e	e	g
CPZ_CM_CAM5	h	h	e	e	e	e	g
CPZ_GA_CPZGAB	h	h	e	e	e	e	g
CPZ_US_CPZUS	h	h	e	e	e	e	g

HIV-1/SIVcpz proteins

Table with columns for protein names (e.g., B.FR.HXB2, CONSENSUS A-A1), alignment positions (101-103), and amino acid sequences. Includes sub-headers for intramolecular disulfide bonding, Rev CDS start, and nuclear localization signal.

Table with 4 columns: Accession ID, protein name, alignment sequence, and residue number. Includes entries like B.FR.HXB2, CONSENSUS\_01.AE, and various CPZ sequences.

HIV-1/SIVcpz proteins

HIV-1/SIVcpz proteins

Accession	Region	High affinity binding site	Leu-rich effector domain	Accession
B.FR.HXB2	exon 1 \ / exon 2	high affinity binding site	Leu-rich effector domain	116
CONSENSUS A-A1		high affinity binding site	Leu-rich effector domain	116
A.SE.SE7535		high affinity binding site	Leu-rich effector domain	116
A.SE.UGSE8131		high affinity binding site	Leu-rich effector domain	116
AL.KE.Q23-17		high affinity binding site	Leu-rich effector domain	116
AL.SE.SET253		high affinity binding site	Leu-rich effector domain	116
AL.UG.92UG037		high affinity binding site	Leu-rich effector domain	116
AL.UG.U455		high affinity binding site	Leu-rich effector domain	116
A2.CD.97CDK610		high affinity binding site	Leu-rich effector domain	116
A2.CD.97CDKTB48		high affinity binding site	Leu-rich effector domain	116
A2.CY.94CY017.41		high affinity binding site	Leu-rich effector domain	116
CONSENSUS B		high affinity binding site	Leu-rich effector domain	115
B.CN.RL42		high affinity binding site	Leu-rich effector domain	116
B.GA.OYI		high affinity binding site	Leu-rich effector domain	116
B.KR.WK		high affinity binding site	Leu-rich effector domain	116
B.TW.TWCYS		high affinity binding site	Leu-rich effector domain	116
B.US.JRFL		high affinity binding site	Leu-rich effector domain	116
B.US.MMCG		high affinity binding site	Leu-rich effector domain	116
B.US.RF		high affinity binding site	Leu-rich effector domain	116
B.US.SF2		high affinity binding site	Leu-rich effector domain	116
CONSENSUS C		high affinity binding site	Leu-rich effector domain	107
C.BR.92BR025		high affinity binding site	Leu-rich effector domain	107
C.BW.96BW01B21		high affinity binding site	Leu-rich effector domain	107
C.DJ.DJ373A		high affinity binding site	Leu-rich effector domain	100
C.ET.ETH2220		high affinity binding site	Leu-rich effector domain	107
C.IL.98ILS002		high affinity binding site	Leu-rich effector domain	109
C.IN.93IN101		high affinity binding site	Leu-rich effector domain	100
C.IN.95IN21068		high affinity binding site	Leu-rich effector domain	107
C.IN.98IN012		high affinity binding site	Leu-rich effector domain	109
C.SO.S0145A		high affinity binding site	Leu-rich effector domain	107
C.TZ.98TZ013		high affinity binding site	Leu-rich effector domain	109
C.UG.UG268A2		high affinity binding site	Leu-rich effector domain	107
C.ZA.97ZA012		high affinity binding site	Leu-rich effector domain	109
C.ZM.96ZM651		high affinity binding site	Leu-rich effector domain	102
CONSENSUS D		high affinity binding site	Leu-rich effector domain	115
D.CD.84ZRO85		high affinity binding site	Leu-rich effector domain	116
D.CD.ELI		high affinity binding site	Leu-rich effector domain	117
D.CD.NDK		high affinity binding site	Leu-rich effector domain	116
D.CD.Z2Z6		high affinity binding site	Leu-rich effector domain	117
D.FI.FI893178		high affinity binding site	Leu-rich effector domain	114
D.SN.SE365A2		high affinity binding site	Leu-rich effector domain	116
D.UG.94UG1141		high affinity binding site	Leu-rich effector domain	116
D.UG.UG269A		high affinity binding site	Leu-rich effector domain	116
D.UG.UG274A2		high affinity binding site	Leu-rich effector domain	116
CONSENSUS F1		high affinity binding site	Leu-rich effector domain	113
F1.BE.VI850		high affinity binding site	Leu-rich effector domain	116
F1.BR.93BR020.1		high affinity binding site	Leu-rich effector domain	116
F1.FI.FI893363		high affinity binding site	Leu-rich effector domain	116
F1.FR.MP411		high affinity binding site	Leu-rich effector domain	116
F2.CM.CM53657		high affinity binding site	Leu-rich effector domain	117
F2.CM.MP255		high affinity binding site	Leu-rich effector domain	116
F2.CM.MP257		high affinity binding site	Leu-rich effector domain	116
CONSENSUS G		high affinity binding site	Leu-rich effector domain	121
G.BE.DRCBL		high affinity binding site	Leu-rich effector domain	123
G.FI.HH8793-1.1		high affinity binding site	Leu-rich effector domain	123
G.NG.92NG083		high affinity binding site	Leu-rich effector domain	123
G.SE.SE6165		high affinity binding site	Leu-rich effector domain	122
H.BE.VI991		high affinity binding site	Leu-rich effector domain	116
H.BE.VI997		high affinity binding site	Leu-rich effector domain	116
H.CF.90CF056		high affinity binding site	Leu-rich effector domain	116
J.SE.SET022		high affinity binding site	Leu-rich effector domain	116
J.SE.SET7887		high affinity binding site	Leu-rich effector domain	116
K.CD.F0TB11C		high affinity binding site	Leu-rich effector domain	116
K.CM.MP535		high affinity binding site	Leu-rich effector domain	116

Accession	Protein	Position	Sequence	Position	Sequence	Position	Sequence
B.FR.HXB2	MAGRSGSDDEE	116	L.IR.TVRLIKLKYQSNPPP	116	L.IR.TVRLIKLKYQSNPPP	116	L.IR.TVRLIKLKYQSNPPP
CONSENSUS_01	AE	116	ra	116	ra	116	ra
01.AE.CF.90CF11697	AE	116	ra	116	ra	116	ra
01.AE.CF.90CF402	AE	116	ra	116	ra	116	ra
01.AE.CF.90CF401	AE	116	ra	116	ra	116	ra
01.AE.CN.97CNGX41F	AE	116	ra	116	ra	116	ra
01.AE.FI.FIN9379	AE	116	ra	116	ra	116	ra
01.AE.JP.93JP.NHI	AE	116	ra	116	ra	116	ra
01.AE.TH.93TH253	AE	116	ra	116	ra	116	ra
01.AE.TH.95TH047	AE	116	ra	116	ra	116	ra
01.AE.TH.CM235-4	AE	116	ra	116	ra	116	ra
01.AE.TH.CM240	AE	116	ra	116	ra	116	ra
CONSENSUS_02	AG	117	ra	117	ra	117	ra
02.AG.CM.97CM.MP807	AG	117	ra	117	ra	117	ra
02.AG.CM.CM52885	AG	117	ra	117	ra	117	ra
02.AG.FR.DJ264	AG	117	ra	117	ra	117	ra
02.AG.GH.A3	AG	117	ra	117	ra	117	ra
02.AG.GH.G829	AG	117	ra	117	ra	117	ra
02.AG.NG.IBNG	AG	117	ra	117	ra	117	ra
02.AG.SE.SE7812	AG	117	ra	117	ra	117	ra
02.AG.SN.MP1213	AG	117	ra	117	ra	117	ra
03.AB.RU.KAL153-2	AB	117	ra	117	ra	117	ra
03.AB.RU.RU98001	AB	117	ra	117	ra	117	ra
04.CPX.CV.94CX032-3	CPX	117	ra	117	ra	117	ra
04.CPX.GR.97FVCH	CPX	117	ra	117	ra	117	ra
04.CPX.GR.97FVWY	CPX	117	ra	117	ra	117	ra
05.DF.BE.V11310	DF	117	ra	117	ra	117	ra
05.DF.BE.V1961	DF	117	ra	117	ra	117	ra
CONSENSUS_06	CPX	117	ra	117	ra	117	ra
06.CPX.AU.BFP90	CPX	117	ra	117	ra	117	ra
06.CPX.ML.95ML127	CPX	117	ra	117	ra	117	ra
06.CPX.ML.95ML84	CPX	117	ra	117	ra	117	ra
06.CPX.SN.97SE1078	CPX	117	ra	117	ra	117	ra
07.BC.CN.97CN001	BC	117	ra	117	ra	117	ra
07.BC.CN.98CN009	BC	117	ra	117	ra	117	ra
08.BC.CN.97CNGX6F	BC	117	ra	117	ra	117	ra
08.BC.CN.97CNGX7F	BC	117	ra	117	ra	117	ra
08.BC.CN.98CNG006	BC	117	ra	117	ra	117	ra
10.CD.TZ.96TZBF061	CD	117	ra	117	ra	117	ra
10.CD.TZ.96TZBF071	CD	117	ra	117	ra	117	ra
10.CD.TZ.96TZBF110	CD	117	ra	117	ra	117	ra
11.CPX.CM.MP818	CPX	117	ra	117	ra	117	ra
11.CPX.FR.MP1298	CPX	117	ra	117	ra	117	ra
11.CPX.GR.GR17	CPX	117	ra	117	ra	117	ra
12.BF.AR.ARMA159	BF	117	ra	117	ra	117	ra
12.BF.UY.URTR23	BF	117	ra	117	ra	117	ra
12.BF.UY.URTR35	BF	117	ra	117	ra	117	ra
N.CM.YBF106	CM	117	ra	117	ra	117	ra
N.CM.YBF30	CM	117	ra	117	ra	117	ra
CONSENSUS_0	CM	117	ra	117	ra	117	ra
0.CM.AN170	CM	117	ra	117	ra	117	ra
0.CM.MVP5180	CM	117	ra	117	ra	117	ra
0.SN.SEMP1299	SN	117	ra	117	ra	117	ra
0.SN.SEMP1300	SN	117	ra	117	ra	117	ra
CONSENSUS_CPZ	CPZ	117	ra	117	ra	117	ra
CPZ.CD.CPZANT	CPZ	117	ra	117	ra	117	ra
CPZ.CM.CAM3	CPZ	117	ra	117	ra	117	ra
CPZ.CM.CAM5	CPZ	117	ra	117	ra	117	ra
CPZ.GA.CPZGAB	CPZ	117	ra	117	ra	117	ra
CPZ.US.CPZUS	CPZ	117	ra	117	ra	117	ra

HIV-1/SIVcpz proteins

	transmembrane domain	-  -	cytoplasmic domain	phos	phos	-	alpha helix	-	
	domain		/ Env	CDS	start		-		
B. FR. HXB2	TQPI.....	PIVAIVALVAIIIAIVVSVIIIE	YRKILRQ.RK...	IDRLIERRAEDSGNESEG...	EISALVEMVEMGHHA.PWD.VDDL\$				82
CONSENSUS A-Al	M2-L.....	?E-W--G-I-L-L-L	T-G--K-L-k-	e-lr	d-dTe-L-?				78
A. FI. FIN91121	MT-L.....	E-S--G-I-L-L-L	T-G--K-LOK	E-lr	G-DTE-L				81
A. SE. UGSE8131	MS-L.....	E-W--G-I-L-L-L	T-GU--K-LK	E-lr	D-DTE-L-T				81
A1. KE. Q23-I7	MS-L.....	E-S--G-I-L-L-L	T-G--K-RK	E-lr	D-DTE-L-T				81
Al. SE. SET253	ML-L.....	ST-W-IG-I-L-L-L	T-G--K-L-K	K-VE	D-DTN-LA				82
Al. UG. 92UG037	M-L-L.....	E-C-V-G-L-L-L	T-G--K-L-K	V-IR	D-DRE-L-L				81
Al. UG. U455	MT-L.....	E-W-TG-I-L-L-L	T-G--\$XK-L-K	LN-IR	D-DTE-L-L				81
A2. CD. 97CDK510	MSL.....	E-A-G-L-A	T-F--RE	W-Q	D-DTD-L				81
A2. CD. 97CDKTB48	MS-L.....	A-LS--G-S-L	T-V-F--KK	W-LE	D-DTE-L-KM				81
A2. CY. 94CY017.41	ML-L.....	V-L--G-I-L-L-L	T-F--K-KS	W-K	D-DTE-L				81
CONSENSUS B	m-sl.....	q-l.....	a-t-l-?	lr	dge-l				81
B. CN. RL42	M-L.....	T-L.....	VA	IR	D-DOE-L-F				81
B. GA. OYI	IYSL.....	Q-L.....	VT	IR	D-DOE-L				81
B. JP. PT1-01	M-SL.....	Q-L.....	F	IR	G-DOE-L				80
B. KR. WK	M-L-L.....	A-L--G-A-L	F-F--K-K-K	IR	DOE-L				81
B. TW. TWCSY	M-L.....	H-L.....	L-L-L	IR	DOE-L				81
B. US. JRC5F	M-L.....	Q-L.....	G-I	KIR	DOE-L				81
B. US. MNCG	M-L.....	A-L.....	\$-F	IS	DOE-L				80
B. US. RF	M-SL.....	E-L.....	A-L	IR	D-DEE-L				82
B. US. SF2	M-SL.....	Q-L--S	T-L	X	DOE-L				81
CONSENSUS C	MLDla???pvdYrlgVg- i- l	t-y-y	l	w-k	ir				84
C. BR. 92BR025	MLEHG...RIDYRLGVG-I-L-V-I	T-AY	L-V	W-VK	IK				86
C. BW. 96BW0402	MISLAA...RDYRIGVG-FI-L-L-I	T-Y	L-V	W-VK	IR				84
C. ET. ETR2220	MVLLAA...KVDYRIV-FI-L-L	T-AY	L-L	R	D-DTE-L-TM				86
C. IL. 98LS002	MVDLLELLKVDYRL-VI-I-VA	T-Y	L-L	R	D-DTE-L-TM				89
C. IN. 93IN101	MIDL.....DYRLGVG-I-L-L	T-Y	L-L	W	DAE-L-TM				82
C. IN. 98IN012	MLDL.....DYKL-VG-FI-L-L	T-Y	L-L	W	DTE-L-TM				82
C. MY. MIDU15	M-L.....	T-L.....	VA	IR	D-DOE-L-GF				81
C. SO. S0145A	MLNLLA...GVDYRIV-G-FS-L-L	T-Y	L-L	W-VK	IR				86
C. TZ. 98TZ017	MLNLLA...RDYRLGVG-L-L-L	T-Y	L-L	W	D-DTE-L-TM				86
C. UG. UG268A2	MLNLLA...GVDYRIG-G-LI-L-L	I-Y	L-L	M	D-DTE-L-TM				86
C. ZA. 97ZA012	MPSLLE...KVDYRLGVG-I-L-L-LT-I	I-AY	L-V	W	DTE-LATM				86
CONSENSUS D	m-l.....	?-l	t-f	w	dke-l-?				80
D. CD. 84ZR085	M-SL.....	Q-L-L-L-L	T-F	R-K	DKE-L-T				82
D. CD. ELI	M-L.....	G-I-A	L-L	R-KK	R				81
D. CD. NDK	M-L.....	V-I-A	L-L	R-K	ERE-L-K				81
D. FI. FIN93167	MSL.....	K-I	L-L	Y	IR				81
D. FI. FIN93178	MXL.....	V-LG-P	XA-L-V	T-Y	K-E-X				81
D. SN. SE365A2	M-SL.....	E-L-A-I	L-L-L	T-F	V				81
D. TZ. TZ005	M-SL.....	V-L--V-A-L	L-L-L	T-F	C-RLS				81
D. UG. 94UG1141	M-L.....	E-L.....	L-L-L	T-F	CK-IR				81
D. UG. UG266A2	IN-L.....	E-W--I-L-L	A-S	C-IK	W				81
F1. BR. V1850	MSYL.....	LAIG-A-I-L-L	T-Y	K-LV	DAE-LA				81
F1. BR. 93BR020.1	MSNL.....	LAIG-A-I-L-L	T-AY	K-LV	DAE-LA				81
F1. FR. MF411	MSNL.....	YVLS--FII-L-L	T-F	N	YE				81
F2. CM. CM53657	MSSL.....	LTI--YI-L-L	I-T-Y	K-L	KR				81
F2. CM. MP255	MPSL.....	LA-G-S-I-L-L	T-YL	K-L	DAE-LA				81
F2. CM. MP257	MSSL.....	L-V-A-YI-VL-L	I-T-Y	K	DAE-LA				81
CONSENSUS G	M-L.....	E-?--g-I-a-A-?	t-f	Rk	K?				76
G. BE. DRGBL	M-L.....	E-S--G-I-S-A	T-F	RK	KR				81
G. CG. CNG30	M-L.....	EVAS--G-I-A-A-I	I-F	\$	E-R				80
G. NG. 92NG083	M-AL.....	E-S..X-I-F-A-TI	-F	RK	K				78
G. TW. TWG1	M-SL.....	E-A--G-I-A-A	-T-F	RK	KR				81
H. BE. V1991	MNLL.....	GIG-G-F	T-AY	LXK	W				80
H. BE. V1997	M-Y.....	IGIG-G-I-F	T-Y	LVK	K				77
H. CF. 90CF056	MYLL.....	GLG-G--TF	-VI-T-Y	K-LV	K				80
J. SE. SE7022	MVSL.....	O-----I-FFL-C-T-Y	-Y	K-L	N				81
J. SE. SE7887	MI-L.....	Q-A-----FI-FL-GM-T-Y	-Y	K-L	N				81
K. CD. E0TB11C	MV-L.....	T-G-I--A-L-L-I	I-T-AYL	NW	F				80
K. CM. MP535	MVSL.....	A-S-----L-L-I	I-T-Y	LVK	KR				80



Table with 3 columns: protein ID, cytoplasmic domain, and alpha helix. Rows include protein IDs like B\_FR\_HXB2, CONSENSUS\_01\_AE, and various MSL, MSH, and MSHL entries. The cytoplasmic domain column contains amino acid sequences with annotations like 'phos' and 'phos start'. The alpha helix column contains sequences with annotations like 'alpha helix' and 'phos'. Line numbers 80-84 are on the right side of the table.

HIV-1/SIVcpz proteins

signal peptide \ / gp120
Vpu CDS end \*

Table with 3 columns: Accession/Label (e.g., B.FR.HXB2, CONSENSUS A-A1, A.CD.97DC.KWST91), Protein Sequence (e.g., MRVKE...KYQHLRWG...WRWGTWMLGMLMICSAT...), and Line Number (124-123).

signal peptide / gp120
Vpu CDS end

Table with columns for accession numbers (e.g., B\_FR\_HXB2), protein names (e.g., CONSENSUS AE), and amino acid sequences. The sequences are aligned across multiple rows, with gaps represented by dashes.

HIV-1/SIVcpz proteins

HIV-1/SIVcpz proteins

Table with 5 columns: Protein ID, Amino Acid Sequence (V1-V4), Conserved Residues (marked with \*), and Reference Number. The table lists various HIV-1 and SIVcpz proteins such as B.FR.HXB2, CONSENSUS A-A1, A.CD.97DC.KMST91, etc., and their corresponding amino acid sequences with conserved residues highlighted by asterisks.

Table with columns for protein names (e.g., B\_FR\_HXB2, CONSENSUS AE, 01\_AE\_CF\_90CF11697) and their corresponding amino acid sequences. The sequences are aligned across multiple lines, with some gaps indicated by dashes. The table is numbered 228 to 233 on the left and 228 to 233 on the right.

HIV-1/SIVcpz proteins



Main alignment area containing protein sequences for HIV-1/SIVcpz, including identifiers like B\_FR\_HXB2, CONSENSUS\_AE, and various amino acid residues aligned in columns.

HIV-1/SIVcpz proteins







	gp120	gp41	fusion peptide	
B.FR.HXB2	NS.....NNESEIFRPGGDMRDNWRSSELYKVKVVKIEPLGVAPTAKRRVV.....QREKRAV.LG.IGALFLG.FLGAAGSTWGAASMTLITVQARQLLSGIYQQNNLLRAIEAQQHLLQLITVWGIKQ			575
CONSENSUS A-A1	vn...??-d-t-?	?	v-?	546
A.CD.97DC.KMST91	N.....D-VT-S-I-E-	?	N...V-A-S	577
A.CD.HIM401034	TN...ETR-N-T-I-E-	EG	M-F-A	574
A.FI.FIN9199	VT...S--T-G		I--V-I	573
A.KE.K89	DN...S--T		I--V	567
A.RW.VPFI	VG...DNST-GT-T-N		I--V	578
A.UA.ukc970063	SG...T-SSN-T-I-N		I--L-A	573
AI.KE.Q23-17	KD...-VN-T		E--V	565
AI.UG.92UG037	VN...SSD-T-R		T-I-V-I	569
AI.UG.U455	AI...-TKN-T-K		E--L-I	568
A2.CD.97CDKSL0	S...-TN	SR	E--V	573
A2.CD.97CDKTB48	KN...SIN-T		I--L-V	562
A2.CY.94CY017.41	-N...GIN-T		E--L-V	577
Consensus B	?n...??-t-?	?	?	0
B.CN.RL42	N...ESKPT-T		R--M	548
B.GA.OYI	KN...TT-GI-A		R--ML-M	564
B.JP.JH32	DN...O--T-N	L	I--V	566
B.TH.93TH067	N...GS-TT-T-N	R	I--T-M	558
B.US.JRFL	IN...E-GT		E--V	566
B.US.MNCG	KD...TDT-DT		A--V	576
B.US.RF	ED...TT-TT-L-N		R--T-M	564
B.US.SF2	TN...VT-DT-V	I	I--IV-M	574
CONSENSUS C	-?????et-?t-?		E-K-I	535
C.BR.98BR004	EL...-DT-R-N-\$K		E--R-V	580
C.BW.96BW0402	TT...R--G-I-E-R		E--V	571
C.ET.ETH2220	KE...PHSTK-E	P	E--A-L	563
C.IL.98LS002	IN...NDT-T-T		E--A-L-M	563
C.IN.93LN101	IKE...NDTE-KT		E--V	576
C.FI.FIN9126	-D...NST-GT-T		E--A-L-M	571
C.SO.S0445A	-DKRTEQ-NT-R		E--I-G	573
C.TZ.98TZ013	-A...GN-TGN-T-RN		E--L-V	574
C.UG.UG268A2	ET...SET-ST-T		E--V	568
CONSENSUS D	an...??-ssp-t-	t	e--I-L-m	544
D.CD.84ZR085	AN...-TONDT		E--I-L-m	575
D.CD.ELI	IN...SIN-T		E--I-L-m	573
D.CD.JY1	VN...-STN-T-K-N		E--I-L-V	582
D.CD.NDK	AN...SSH-TI	I	E--I-L-V	565
D.CI.C113	VN...HNSN-T-I		E--I-L-M	560
D.FI.FIN93178	TN...SSH-T		R--L-L	554
D.SN.SE365A2	IH...NTS-T-T	T	K--I-L-VM	570
D.TZ.87TZ4622	AN...SSD-T		E--I-L-V	570
D.UG.WHO15-474	AN...NGSNN-T-I-T		E--I-L-M	579
F1.BE.VI850	-E...SNI-T-E-N-K	Q	E--L-X-T-X	552
F1.BR.93BR020.1	LN...STN-T-N-K	Q	E--L	565
F1.FR.MF411	Q...NDT-RT-T-E-N-K-N		E--R-R-NI	562
F2.CM.CA20	KN...G-TL-R		K--R	550
F2.CM.CM53657	KN...-TNDTL	Q	K...RE	568
F2.CM.MF257	EG...-TTL	Q	I--M-M	569
G.BE.DRCBL	DN...ST-KN		E--V-I	566
G.NG.92NG083	N...NDSTE-T		E--L-V	561
G.SE.SE6165	-NN.NTNTS	KS	E--L-V	578
H.BE.VI991	-T...NNVT		E--M-F	581
H.BE.VI997	HG...DNVT		E--M-F	574
H.CF.90CF056	-A...SA-NVT		E--M-S	567
J.SE.SE7022	-R...NGSE-GT-T-T-N-K	EL	E--V	572
J.SE.SE7887	-RG.NGSE-GT-T-T-N-K		E--V	570
K.CD.F0TB11C	-D...NNTRTE-T		E--V	570
K.CM.MF535	-N...THN-T	Q	I--L-V-F	561

	-	V5	^	gp120	gp41	fusion peptide		
B_FR_HXB2	NS.....NNESEIFRPGGDRMDNRSELKYKVKVIEPLGVAPTKAKRRVV.....QREKRAV.LG.IGALFLG.FLGAAGSTWGAASMTLTVQARQLLSGIYVQQNNLLRAIEAQQHLLQLTVWGIKQ						575	
CONSENSUS AE	an?..??-tdt-T	NiK	g	i	r	i	MF	561
01 AE.CF.90CF11697	-A.....TN-T	N-K	O	I	R	N	K	575
01 AE.CF.90CF402	INO...NOT-KN-T	NiK	O	I	R	R	S	583
01 AE.CF.90CF4071	KI...E-FIT-T	N-K	O	I	R	R	S	570
01 AE.CM.CA10	GN...STDT-T	N-K	Q	I	R	R	M	581
01 AE.CN.97CNGX2F	AD...N-TTN-T	NiK	E	I	R	R	S	566
01 AE.JP.93UP-NHL	AN...KTNN-T	NiK	Q	I	R	R	S	569
01 AE.TH.95TH022	DN...N-TIN-T	NiK	Q	I	R	R	S	569
01 AE.TH.93TH9021	AN...MSTN-T	NiK	Q	I	R	R	S	568
01 AE.TH.CW240	VN...T-DN-T	NiK	Q	I	R	R	S	569
CONSENSUS 02 AG	-n.....?..-stb-T	e	?	l	v	s	k	554
02 AG.CM.CM53658	-N.....NSSTN-T	L	L	V	F	S	K	567
02 AG.FR.AKON	-N.....SAN-T	R	L	V	F	S	K	581
02 AG.CM.CM52885	SN.....STN-T	R	L	V	F	S	K	574
02 AG.FR.DJ264	IN.....STN-T	L	L	V	F	S	K	567
02 AG.GH.AB052867	IN.....STN-T	L	L	V	F	S	K	572
02 AG.NG.IBNG	IN.....STN-T	L	L	V	F	S	K	566
02 AG.NG.NG1921	KH.....STN-T	A	Y	L	V	S	K	585
02 AG.SE.SE7812	-D.....NT-T	SH	L	V	F	S	K	578
02 AG.SN.98SE-MP1211	DN.....T-GD-T	SH	L	V	F	S	K	572
03 AB.RU.KAL153-2	-Q.....S-VT-T	R	L	V	F	S	K	559
03 AB.RU.KAL168.1	YQ.....S-VT	R	L	V	F	S	K	558
03 AB.RU.RU98001	-Q.....VT	R	L	V	F	S	K	557
04 CPX.CV.94CF032-3	TN.....NT	L	N	R	M	S	R	573
04 CPX.GR.97FVCH	I-N..NNEF-DN-TL	V	I	G	R	L	S	573
04 CPX.GR.97FVNY	-R.....SDTGN-T	Q	V	P	L	S	V	580
05 DF.BE.V11310	-GX..XDSS-DT	E	O	I	R	Q	A	581
05 DF.BE.V1961	VN.....IT-DT-T	E	R	Q	V	L	Q	572
06 CPX.AU.BFP90	-E.....V-T	K	I	W	R	S	S	578
06 CPX.ML.95ML84	TN.....TT-I	N	K	I	E	R	S	559
06 CPX.SN.97SE1078	-D.....T	N	K	I	R	V	S	565
07 BC.CN.97CN001	TE.....P-DT-T	N	E	K	T	M	S	571
07 BC.CN.98CN009	TE.....T-DT-T	N	E	K	A	V	S	571
08 BC.CN.97CNGX6F	RT...EP-NT	N	N	A	L	V	M	576
08 BC.CN.97CNGX9F	RT...ES-DT	N	N	A	L	V	M	570
08 BC.CN.98CN006	TE.....SNKTKV	N	N	A	L	V	S	578
10 CD.TZ.96TZBF061	AN...N-SON-T	X	I	L	V	V	S	559
10 CD.TZ.96TZBF071	GGX..XNNTSON-T	E	I	L	M	E	K	572
10 CD.TZ.96TZBF110	AN...NS-QSG	R	L	L	V	L	M	567
11 CPX.CM.MP818	LN.....STN-T	T	E	K	R	V	L	564
11 CPX.GR.GR17	EG...NDTIGK-T	N	E	K	I	V	K	561
11 CPX.NG.NG3670b	KN.....THN-T	K	E	K	R	V	L	574
12 BF.AR.ARMAL159	LN...ET-QT-T	N	K	R	Q	T	Q	561
12 BF.UY.URTR23	-N...NAT-KT-T	W	N	K	R	Q	S	566
12 BF.UY.URTR35	HPD.XIMTG-RT-T	N	K	R	Q	M	S	576
N.CM.YBF106	PD.....GNKTVY-S-N-VNL-Q	S	I	G	T	S	I	546
N.CM.YBF30	PD.....TKET-VY-S-N-VNL-Q	S	I	G	T	S	I	563
CONSENSUS O	wN...??s-??t	i	k	i	l	fn	svk	544
O.CM.ANT70	WN...SS-NNVT	I	K	I	T	FN	R	573
O.CM.CM4974	WN...KS-PNVT-G-T	I	A	CM	VS	PII	LNPH	582
O.CM.HV1CA9EN	WN...SSRDNTL-V	K	I	T	F	VK	FS	584
O.CM.MVP5180	WN...STGNTL-V	K	I	TK	N	MS	PII	583
O.GA.VI686	WN...KSNVT	V	K	I	T	F	R	584
O.GQ.L193HA	WN...TSTNEAK-I	I	K	FN	R	VK	FS	570
O.SN.SEMPI300	WN...RTI-SEST	I	K	I	T	F	K	581
CONSENSUS CPZ	?e...n?s?nitvvt-n-?l-l-b-s-i	ht-??	af-??	af-??	af-??	af-??	af-??	479
CPZ.CD.CPZANT	-E...TVKYS-AARVA-O-A-SR-O	E	X	S	TX	PEIK	QHS	568
CPZ.CM.CM3	IN...NSITNITLY-T-N-V-L	L	H	S	O	GI	LF	567
CPZ.CM.CM5	WE...NNTIVY-S-E-N-L	L	NR	S	I	S	HT	573
CPZ.GA.CPZGBAB	VT...N-SGNLT	I	N	K	I	A	AP	565
CPZ.US.CPZUS	AV...NDSRNITVM-T	TAL	KN	H	H	G	HT	547





HIV-1/SIVcpz proteins

HIV-1/SIVcpz protein alignment: ENV

Multiple sequence alignment of HIV-1/SIVcpz ENV proteins. Columns represent amino acid positions from 781 to 824. Rows list protein identifiers such as B.FR.HXB2, A.AL, A.CD.97DC.KMST91, etc. The alignment shows conserved regions and variable loops. A shaded box highlights a region from position 781 to 824 in the first column.

Rev CDS end \

Tat CDS end \

\*\*\*

IVFAVLIVNRVQGYSPLSFQTHLP.TPRGDRPEGIEEGEGRDRSIRLVNGSLALWDDLRSLCLFSYHRIRLDLLIVTRIVELGR.....RGWEALKKWNLLQYWSQELKNSAVSLINAT

Main alignment area containing amino acid sequences for various HIV-1/SIVcpz strains, including B.FR.HXB2, A.AL, A.CD.97DC.KMST91, A.CD.HIM401034, A.FI.FIN91199, A.KE.K89, A.RW.PVPI, A.UA.ukz970063, A1.KE.Q23-17, A1.UG.92UG037, A1.UG.U455, A2.CD.97CDKSL0, A2.CD.97CDKTB48, A2.CT.94CY017.41, B.CN.RL42, B.GA.OYI, B.JP.JH32, B.TH.93TH067, B.US.JRFL, B.US.MNCG, B.US.RF, B.US.SF2, C.BR.98BR004, C.BW.96BW0402, C.ET.ETH2220, C.IN.98IS002, C.IN.93IN101, C.FI.FIN91126, C.SO.S0145A, C.TZ.98TZ013, C.UG.UG268A2, CONSENSUS\_D, D.CD.84ZR085, D.CD.ELI, D.CD.JY1, D.CD.NDK, D.CI.CI13, D.FI.FIN93178, D.SN.SE365A2, D.TZ.87TZ4622, D.UG.WH015-474, F1.BE.VI850, F1.BR.93BR020.1, F1.FR.MP411, F2.CM.CA20, F2.CM.CM53657, F2.CM.MP257, G.BE.DRCBL, G.NG.92NG083, G.SE.SE6165, H.BE.VI991, H.BE.VI997, H.CF.90CF056, J.SE.SE7022, J.SE.SE7887, K.CD.EQTB11C, K.CM.MP535

IVFAVLIVNRVQGYSPLSFQTHLP.TPRGPDREGEIEGRRDRSIRLVNGSLALWDDLRSLCIFSFRHRLDLLLLIVTRIVELGR.....RGWEALKWNNLLQYWSQELKNSAVSLLNAT
Tat CDS end
Rev CDS end
CONSENSUS\_AE
01 AE.CF.90CF11697
01 AE.CF.90CF402
01 AE.CF.90CF4071
01 AE.CM.CA10
01 AE.CN.97CNGX2F
01 AE.JP.930P.NH1
01 AE.TH.95TH022
01 AE.TH.95TH9021
01 AE.TH.CM240
CONSENSUS\_02 AG
02 AG.CM.CM53658
02 AG.FR.AKGN
02 AG.CM.CM52885
02 AG.FR.DJ264
02 AG.FR.AB052867
02 AG.NG.IBNG
02 AG.NG.NG1921
02 AG.SG.SE7812
02 AG.SN.98SE.MP1211
03 AB.RU.KAL153-2
03 AB.RU.KAL68.1
03 AB.RU.RU98001
04 CPX.CX.94CX032-3
04 CPX.GR.97PVCH
04 CPX.GR.97PVNY
05 DF.BE.V11310
05 DF.BE.V1961
06 CPX.AU.BFP90
06 CPX.ML.95ML84
06 CPX.SN.97SE1078
07 BC.CN.97CN001
07 BC.CN.98CN009
08 BC.CN.97CNGX6F
08 BC.CN.97CNGX9F
08 BC.CN.98CN006
10 CD.TZ.96TZBF061
10 CD.TZ.96TZBF071
10 CD.TZ.96TZBF110
11 CPX.CM.MP818
11 CPX.GR.GR17
11 CPX.NG.NG3670b
12 BF.AR.ARMA159
12 BF.UY.URTR23
12 BF.UY.URTR35
N.CM.YBF106
N.CM.YBF30
CONSENSUS\_0
O.CM.ANT70
O.CM.CM4974
O.CM.HV1CA9EN
O.CM.MVP5180
O.GA.VI686
O.GQ.193HA
O.SN.1SEMP1300
CONSENSUS\_CPZ
CPZ.CD.CPZANT
CPZ.CM.CAM3
CPZ.CM.CAM5
CPZ.GA.CPZGAB
CPZ.US.CPZUS

B.FR.HXB2	ATAVAEGRDRIEVVQACRAIRHPRPRRQGLERILL\$	856
CONSENSUS_A-AL	---gw---ig-rig---ln---f---A---	830
A.CD.97DC.KMST91	---T-G-R-I---LN---F---A---	865
A.CD.HIM401034	---T---W---A---G-RIG-G-LN---F---A---	862
A.FI.FIN9199	---V---GW---DIG---IG---L---A---	861
A.KE.K89	---V---GW---IA---IG---L---F---A---	855
A.RW.PVPI	---V---GW---G-RIG-FLN---A---A---L---	866
A.UA.ukz970063	---GW---IG-RF---N---X---A---KA-Q---	861
AL.KE.Q23-17	---GW---IA-RIG---L---V---A---	853
AL.UG.92UG037	---GW---T---RLG---LN---F---A---	857
AL.UG.U455	---V---GWI---IG-TIG---LN---A---	855
A2.CD.97CDKS10	---W---W---IG-R---LN---F---A---	861
A2.CD.97CDKTB48	---V---W---IG-R---N---A---	850
A2.CY.94CY017.41	---V---W---IG-R-F---LN---A---	865
CONSENSUS_B	---?---R-Y---L---L---t---a---	824
B.CN.RL42	---R-Y---L---L---T---A---	845
B.GA.OVI	---I---R-Y---FLN---A---	855
B.JP.JH32	---L---KIL-R-F---L---T---A---	867
B.TH.93TH067	---L---R-Y---L---T---FK-A---	837
B.US.JREFL	---I---AL-RTY---L---T---A---	847
B.US.MNCG	---L---R-G---L---T---A---	856
B.US.RF	---I---A---RIL-FL---A---	865
B.US.SF2	---T---A---R-Y---L---H---L---	855
CONSENSUS_C	---i-l?-xi?-cp-t---f-aa-Q---	817
C.BR.98BR004	---I---II-RIW---FCNV-T---AA-Q---	866
C.BW.96BW0402	---I---IA-RI---CN-T---F-AA-Q---	859
C.ET.ETH2220	---V---G---F-LI-RW---FCN---AA-Q---	851
C.II.98IS002	---I---OL-NIW-T-CN---F-AA-Q---	851
C.IN.93IN101	---I---L---F---N---T---F-AA-Q---	857
C.FI.FIN9126	---I---LA-RIW---CN-T---F-AA-Q---	859
C.SO.S0145A	---T---I---LA-RI---GV-N-T---F-AA-Q---	861
C.TZ.98TZ013	---I---II-RI---CN---F-TA---	862
C.UG.UG368A2	---S---I---G---IG---L---F-AA-Q---	856
CONSENSUS_D	---i-?-R---vl---a---	820
D.CD.84ZR085	---I---DI-RR-K-VL-T---A---	856
D.CD.ELI	---II-R---VLN---S---	854
D.CD.JY1	---I---LIRR-F-VL---V---A---	863
D.CD.NDK	---R---R---R---LNV---L---	846
D.CI.CI13	---V---IL-A-RVM-VL-T---A---	841
D.FI.FIN93178	---G---A---R---VL---GA---	835
D.SN.SE365A2	---I---D---R---L---T---A---	851
D.TZ.87TZ4622	---T---DI-RTV-VLN-T---S---A---	851
D.UG.WH015-474	---T---I---G-R---N---A---	860
F1.BE.VI850	---V---I---L-R-G-VLN---A---A---	833
F1.BR.93BR020.1	---W---AL-R-G-LN---A---	846
F1.FR.MP411	---V---L---L-R-G-VLNV---S---	843
F2.CM.CA20	---V---IV---L-R-G-VL---F---A---	831
F2.CM.CM53657	---V---I---L-R-G-L---A---A---	849
F2.CM.MP257	---I---IL-R-G-VLN---A---	850
G.BE.DRCBL	---NW---A-R-G-VLN---A---	854
G.NG.92NG083	---T-N---A-R-Y-LNV-T---A---	849
G.SE.SE6165	---NW---A-R---LN-T---A---	866
H.BE.VI991	---I---L---R-W---L---F---A---	862
H.BE.VI997	---V---I---R-W-VL---F---A---	855
H.CF.90CF056	---GI-VI-R-W-L---F---S---	848
J.SE.SE7022	---I---IA-R-F-L---A---	853
J.SE.SE7887	---I---IA-R-F-L---A---	851
K.CD.EQTB11C	---I---I-YR-F-LL---F---L---	851
K.CM.MP535	---G---I---IG-R-F-LL---A---	842



B.FR.HXB2	ATAVAGETDRVIEVQACRAIRHPRRIQGLERILL\$	856
CONSENSUS_AE	-----a-I-W-L-----a	848
01_AE.CF.90CF11697	-V--GW-----R-W-LI-----A	863
01_AE.CF.90CF402	-T--GW-----I-R-W-L-----A	871
01_AE.CF.90CF4071	-T--GW--A--IA-R-----L-A	858
01_AE.CM.CA10	-T--GW-----GA-R-Y-FL-----F-A	869
01_AE.CN.97CNGX2F	--GW-----A-R-W-LI-----A	854
01_AE.JP.93JP-NHI	--GW-----A-W-L-----A-V	857
01_AE.TH.95TH022	-V--GW-----A-W-L-----A	857
01_AE.TH.93TH9021	-V--GW-----A-W-L-----A	856
01_AE.TH.CM240	-A--GW-----A-W-L-----T	857
CONSENSUS_02_AG	-----nW-----iG-R-G--D-----a	842
02_AG.CM.CM53658	--NW-----IG-R-G-VLN-----S	855
02_AG.FR.AKGN	--NW-----IG-R-G-CN-----A	862
02_AG.CM.CM52885	-----NW-----IG-R-G--N-----A	855
02_AG.FR.DJ264	-----NW-----IG-RVG--N-V-----A	855
02_AG.GH.AB052867	--W-----G-R-G-L-----A	860
02_AG.NG.IBNG	-V--NW--A--IG-RVG--N-----F-A	854
02_AG.NG.NG1921	--FT-NW-----IG-R-G--N-----S	866
02_AG.SE.SE7812	-V--NW-----L-R-G-LN-----F-A	859
02_AG.SN.98SE-MP1211	-V--NW-----I--RTG--CN-----A-Q	861
03_AB.RU.KAL153-2	-----GW-----IG-RF--M-N-----A-KA-Q	839
03_AB.RU.KAL68.1	-----GW-----IG-RF--N-----A-KA-Q	839
03_AB.RU.RU98001	-----GW-----IG-RF--N-----A-KA-Q	838
04_CPX.CY.94CY032-3	-----I--A--R-----CN-----A	855
04_CPX.GR.97PVCH	-V-----IL-A-R-----N-----F-KA	876
04_CPX.GR.97PVNY	-V-----I--A--RI-----LN-----A	862
05_DF.BE.VI1310	-VV-----AL-R-G--LN-----A	862
05_DF.BE.VI961	-VV-----IL-AL-R-G-VLN-----A	853
06_CPX.AU.BFP90	-----NW--A--RIF--FLNV-----F-A	859
06_CPX.ML.95ML84	--NW-----R-F-VLN-----A-A-I	840
06_CPX.SN.97SE1078	--NW-----I--RVF--FLNV-----F-A	846
07_BC.CN.97CN001	-I--L--L-----YN-----F-AA-Q	859
07_BC.CN.98CN009	-I--L--L-----YN-----F-AA-Q	859
08_BC.CN.97CNGX6F	-I--NR--I-----HN-----F-AA-Q	864
08_BC.CN.97CNGX9F	-I--NI--I-----HN-----F-AA-Q	858
08_BC.CN.98CN006	-----IVNI--I-----N-----F-AA-Q	866
10_CD.TZ.96TZBF061	-E--GW-----I--R-V--LN--T-----A	840
10_CD.TZ.96TZBF071	-V-----A--I--R-V-TVIN-----A-AF	852
10_CD.TZ.96TZBF110	-V-----A--I--R-V-VLN--T-----A	848
11_CPX.CM.MP818	-I--I--I--RVL-G-L-----A	845
11_CPX.GR.GR17	-I--AHR-L-----LN-----F-A	844
11_CPX.NG.NG3670b	-I--I--I--VL-----I-----A	855
12_BF.AR.ARMA159	--V--G-----IL-AL-RIG--VLNA--V-----A	842
12_BF.UY.URTR23	-V-----AL-RVG--VLNA-----A	847
12_BF.UY.URTR35	-V-----I--AL-R-G--LN-----A	857
N.CM.YBF106	-V-----I--LA-RIG-G-L-----A-I	836
N.CM.YBF30	-V-----I--LA-RIG-G-L-----A-I	852
CONSENSUS_0	-V--NW--GI--LGI--RIG-G--ln-----?	821
O.CM.ANT70	-V--NW--GI--AGI--RIGTG--N-----S	859
O.CM.CM4974	-VS--NW--GI--LGI--RIGOG--LN-----A-F	869
O.CM.HIVICA9EN	-V--NW--SI--LGI--RIG-G--LN-----V	872
O.CM.MVP5180	-VS--NW--GI--LGI--RIGOGFL-----A--G	872
O.GA.VI686	-VT--NW--GI--LGI--RIG-G--N--T-----S	875
O.GQ.193HA	-V--NW--GI--LGI--RIG-G--WN-----S	875
O.SN.SEMP1300	-V--NW--GI--LGI--RIG-G--LN-----A	872
CONSENSUS_cpz	-I--a?--iig-g?-----s	745
CPZ_CD.CPZANT	--W-G-K--SILLAL-TIV-I--EV-----IA-N	858
CPZ_CM.CAM3	-----II--A--IIG-G-L-----S	856
CPZ_CM.CAM5	-----N-----I--F--IVG-G-L-----S	862
CPZ_GA.CPZGAB	-----I--AF-VTI--I--N-----A	854
CPZ_US.CPZUS	-----I--LTRLFLG-I-----S	856

\*  
ATAVAGETDRVIEVQACRAIRHPRRIQGLERILL\$

HIV-1/SIVcpz proteins

Table showing protein alignment for NEF. Columns include domain annotations (anchor domain, alpha 2 helix, core domain, alpha 3 helix, alpha 4 helix) and amino acid sequences for various protein variants (e.g., B.FR.HXB2, CONSENSUS\_A-1, F1.BR.93BR020.1, etc.).





	alpha 4 helix	beta 2	beta 3	beta 4	beta 5	alpha 5 helix	alpha 6 helix	
		D123 essential for dimerization						
		beta-COP recruitment ?						
		di-leucine based AP recruitment						
		V-ATPase and Raf-1 recruitment						
		endocytic signal						
		C-terminal flexible loop						
		recruitment						
B_FR_HXB2	HTQGYF.PP\$QNYTPGG.VRYP	LPFGMCYKIVPPDKIEEA.NK	GENTSLHPVSLHGMDFR	REVLWFRDLSRAFHVHARELHPEYFK.NG\$				205
CONSENSUS_01_AF	n---f---w---	i---c---f---	d---r---e---v---d---	nc---nc---ie---e---	m---q---v---e---	m---k---a---rk---i---	a---rk---i---	204
01_AE_CF_90CF11697	W---W---	I---I---	D---REV---D---	D---SNC---	M---Q---V---E---	M---K---S---RR---I---	R---Y---D---	204
01_AE_CF_90CF402	W---W---	I---I---	D---REV---D---	D---SNC---	M---Q---V---E---	M---K---S---RR---I---	R---Y---D---	211
01_AE_CF_90CF4071	W---W---	I---I---	D---REV---D---	D---SNC---	M---Q---V---E---	M---K---S---RR---I---	R---Y---D---	200
01_AE_CN_97CNGX2F	F---W---	T---C---	D---REV---D---	NC---MNO---	Q---IE---E---	M---K---S---RR---I---	Y---D---	206
01_AE_CN_HIV232982	F---W---	T---C---	D---REV---D---	NC---MNO---	Q---IE---E---	M---K---S---RR---I---	Y---D---	204
01_AE_JP_93JP_NH1	F---W---	T---C---	D---REV---D---	NC---MNO---	Q---IE---E---	M---K---S---RR---I---	Y---D---	204
01_AE_TH_93TH057	F---W---	T---C---	D---REV---D---	NC---MNO---	Q---IE---E---	M---K---S---RR---I---	Y---D---	204
01_AE_TH_95TH022	F---W---	T---C---	D---REV---D---	NC---MNO---	Q---IE---E---	M---K---S---RR---I---	Y---D---	201
01_AE_TH_CM235-2	F---W---	T---C---	D---REV---D---	NC---MNO---	Q---IE---E---	M---K---S---RR---I---	Y---D---	204
01_AE_TH_CM240	F---W---	T---C---	D---REV---D---	NC---MNO---	Q---IE---E---	M---K---S---RR---I---	Y---D---	204
CONSENSUS_02_AG	f---w---	t---c---	d---rev---d---	nc---nc---	ie---e---	m---k---a---rk---	a---rk---i---	206
02_AG_CM_97CM_MP807	F---W---	T---C---	D---REV---D---	NC---ICQ---	E---ED---V---	H---LK---R---I---	F---Y---D---	206
02_AG_CM_CM52885	I---I---	F---F---	MD---AEV---TE---N---	ICQ---	LE---AD---V---	T---KV---	F---Y---D---	206
02_AG_CM_CM53658	W---W---	T---F---	MD---AEV---TE---N---	ICQ---	LE---AD---V---	T---KV---	F---Y---D---	206
02_AG_FR_DJ263	F---W---	T---F---	MD---AEV---TE---N---	ICQ---	LE---AD---V---	T---KV---	F---Y---D---	206
02_AG_FR_DJ264	F---W---	T---F---	MD---AEV---TE---N---	ICQ---	LE---AD---V---	T---KV---	F---Y---D---	206
02_AG_GH_A3	F---W---	T---F---	MD---AEV---TE---N---	ICQ---	LE---AD---V---	T---KV---	F---Y---D---	216
02_AG_NG_I18NG	F---W---	T---F---	MD---AEV---TE---N---	ICQ---	LE---AD---V---	T---KV---	F---Y---D---	206
02_AG_SE_SE7812	F---W---	T---F---	MD---AEV---TE---N---	ICQ---	LE---AD---V---	T---KV---	F---Y---D---	206
02_AG_SE_SE7812	F---W---	T---F---	MD---AEV---TE---N---	ICQ---	LE---AD---V---	T---KV---	F---Y---D---	206
02_AG_SE_SE7812	F---W---	T---F---	MD---AEV---TE---N---	ICQ---	LE---AD---V---	T---KV---	F---Y---D---	209
02_AG_SE_SE7812	F---W---	T---F---	MD---AEV---TE---N---	ICQ---	LE---AD---V---	T---KV---	F---Y---D---	209
03_AB_RU_KAL153-2	W---W---	I---I---	D---REV---D---	NC---ICQ---	E---EDK---V---	R---K---I---	Y---D---	202
03_AB_RU_RU98001	W---W---	I---I---	D---REV---D---	NC---ICQ---	E---EDK---V---	R---K---I---	Y---D---	207
04_CPX_CY_94CY032-3	F---F---	E---F---	D---REV---D---	NC---ICQ---	E---EDK---V---	R---K---I---	Y---D---	219
04_CPX_GR_97PVGH	W---W---	T---F---	D---REV---D---	NC---ICQ---	E---EDK---V---	R---K---I---	Y---D---	215
04_CPX_GR_97PVNY	W---W---	T---F---	D---REV---D---	NC---ICQ---	E---EDK---V---	R---K---I---	Y---D---	210
05_DF_BE_VI1310	N---S---	I---I---	D---REV---D---	NC---MNO---	Q---IE---E---	M---K---S---RR---I---	Y---D---	206
05_DF_BE_VI1961	W---W---	T---F---	D---REV---D---	NC---MNO---	Q---IE---E---	M---K---S---RR---I---	Y---D---	209
06_CPX_AU_BFP90	N---F---	T---F---	D---REV---D---	NC---MCO---	AE---E---G---M---K---S---RR---I---	K---Y---D---	205	
06_CPX_WL_95WL127	W---W---	I---I---	D---REV---D---	NC---MCO---	AE---E---G---M---K---S---RR---I---	K---Y---D---	209	
06_CPX_SN_97SEL078	F---F---	I---I---	D---REV---D---	NC---MCO---	AE---E---G---M---K---S---RR---I---	K---Y---D---	216	
07_BC_CN_97CN001	W---W---	T---F---	D---REV---D---	NC---CO---	E---EH---K---K---Q---HR---R---	F---Y---D---	206	
07_BC_CN_98CN009	W---W---	T---F---	D---REV---D---	NC---CO---	E---EH---K---K---Q---HR---R---	F---Y---D---	206	
08_BC_CN_97CNGX6F	W---W---	T---F---	D---REV---D---	NC---CO---	E---EH---K---K---Q---HR---R---	F---Y---D---	206	
08_BC_CN_97CNGX7F	W---W---	T---F---	D---REV---D---	NC---CO---	E---EH---K---K---Q---HR---R---	F---Y---D---	206	
08_BC_CN_98CN006	W---W---	T---F---	D---REV---D---	NC---CO---	E---EH---K---K---Q---HR---R---	F---Y---D---	206	
10_CD_TZ_96TZBF061	N---F---	T---F---	D---REV---D---	NC---M---	E---KHG---V---K---T---HK---I---	Y---D---	207	
10_CD_TZ_96TZBF071	N---F---	T---F---	D---REV---D---	NC---M---	E---KHG---V---K---T---HK---I---	Y---D---	212	
10_CD_TZ_96TZBF110	N---F---	T---F---	D---REV---D---	NC---M---	E---KHG---V---K---T---HK---I---	Y---D---	207	
11_CPX_CM_MP818	F---F---	C---C---	D---REV---D---	NC---M---Q---	IE---ED---R---K---S---RR---I---	DFY---D\$	207	
11_CPX_FR_MP1298	F---F---	C---C---	D---REV---D---	NC---M---Q---	IE---ED---R---K---S---RR---I---	DFY---D\$	207	
11_CPX_GR_GR17	F---F---	C---C---	D---REV---D---	NC---M---Q---	IE---ED---R---K---S---RR---I---	DFY---D\$	207	
12_BF_AR_ARMAL159	W---W---	L---L---	D---REV---D---	NC---M---Q---	IE---EDK---V---	LR---L---K---WY---D\$	205	
12_BF_UY_URTR23	W---W---	L---L---	D---REV---D---	NC---M---Q---	IE---EDK---V---	LR---L---K---WY---D\$	209	
12_BF_UY_URTR35	W---W---	L---L---	D---REV---D---	NC---M---Q---	IE---EDK---V---	LR---L---K---WY---D\$	205	
N_CM_YBF106	F---F---	X---C---	D---REV---D---	NC---ICQ---	V---D---HKQ---V---	S---RR---K---DFY---	207	
N_CM_YBF30	I---I---	V---V---	D---REV---D---	NC---ICQ---	V---D---HKQ---V---	S---RR---K---DFY---	212	
CONSENSUS_O	?---F---	LF---	Se2EA-2LG-?c-rA-aCh-fe-2H?-i-k-q-rs-Gst-?it-L-?kd					199
O_CM_ANP70	F---F---	LF---	SEEEA-RLG-TC-RAN-ACN-AE-THK-I-M-K-RS-GNT--MIT-L-O-KD					213
O_CM_MVP5180	F---F---	LF---	SEEEA-RLG-TN-DA-ACN-AE-THK-I-M-K-RS-GNT--MIT-L-O-KD					214
O_SN_SEMP1299	F---F---	LF---	SEAAA-LG-C-RA-ACN-FE-NHGOI-K-Q-RS-GST--MVTN-L-N-KD					211
O_SN_SEMP1300	F---F---	LF---	SEAAA-LG-C-RA-ACN-FE-NHGOI-K-Q-RS-GST--MVTN-L-N-KD					211
CONSENSUS_CPZ	i---i---	?---?---	lt?eev-q-e-d?Ni-iCq-e-ehg-v-yr-d					198
CPZ_CD_CPZANT	N---I---	CR---	SPDD-R-NI--ACT-DG-HK-I-R-E-AS-MRR-I-R--R-D\$					207
CPZ_CM_CM3	F---F---	Y---G---F---	LT-EV-Q--E-D-NI-ICQ-E-ahg-v-yr-d					207
CPZ_CM_CM5	I---I---	W---W---	LT-EV-Q--E-D-NI-ICQ-E-ahg-v-yr-d					208
CPZ_GA_CPZGAB	F---F---	T---T---C---	LTEEQ-Q--E-D-NC-ICQ-E-EDK-V--LR-L-Q--YG.DH\$					205
CPZ_US_CPZUS	I---I---	W---W---	LTEEQ-Q--E-D-NC-ICQ-E-EDK-V--LR-L-Q--YG.DH\$					209

