

Figure S1

1 R	GCGCCAGTCATCCGATAGACTGAGTCGCCCGGGTACCCGTGTTCCCAATAAAGCCTTTTG ----->	2341 Gag-Pro-Pol	GGACCCCTAAGTGACAACTCTGCCTGGTCCAAGGGGCTACTGGAGGAAAGCGGTATCGC G P L S D K S A L V R Q G A T G G K R Y R
61 R U5	CTGTTTGCATCCGAAGCGTGGCTCGTGTTCCTTGGGAGGGTCTCCTCAGAGTGATTGA -----> ----->	2401 Gag-Pro-Pol	TGGACCA CGGATCGCAAGATACATCTGGCTACCGGTAAGGTACCCACTCTTCTCCCTAT W T T D R K V H L A T G K V T H S F L H
121 U5 tRNA-Pro PBS	CTACCCAGCTCGGGGGTCTTTTCATTTTGGGGCTCGTCCGGGATTCGGAGACCCCGCCCA -----> ----->	2461 Gag-Pro-Pol	GTACCAAGCTGCCCCCTATCTCTGTAGGAAGAGACTTGCTGACTAACTAAAGCCCA V P D C P Y P L R Q D L L T K K A Q
181 splice donor	GGGACCACCGACCCACGTCGGGAGGTAAGCCGGCCGGCGATCGTTTTGTCTTTGTCTCT -----> ----->	2521 Gag-Pro-Pol	ATCCACTTCGAGGGATCAGGAGCTCAGGTTGTGGGACCGATGGGACAGCCCTCGAAGTG I H F E G S G A Q V V G P M G Q P L Q V
241	GTCTTTGTGCGTGTGTGTGTGCGCGCATCTAATCCTCGCGCTCGGTCTGAATCTGTGA	2581 Gag-Pro-Pol	CTGACCCCTAAACATAGAAAATAAGTATCGGTACATGAGACCTCAAAGAGCCAGATGTT L T L N I E N K Y R L H E T S K E P D V
301 g1-Gag start	CTAGTTAGCTAACTAGATCTGTATCTGGCGGTTCCGCGGAAGAACTGACGAGTTCGTATT -----> -->	2641 Gag-Pro-Pol	CCTCTAGGGTCCACATGGCTTTCTGATTTTCCCCAGGCGTGGGCGAAACCGGGGCATG P L G S T W L S D F P Q A W A E T G G M
361	CCCGGCCGACGCCGAGGAGACGTCACGCGGCTCGGGGGCCCGTTTGTGGCCCATTC	2701 Gag-Pro-Pol	GGACTGGCAGTTCGCGAAGCTCCTCTGATCATACCTCTGAAGGCAACCTCTACCCCGTG G L A V R Q A P L I I P L K A T S G L
421	TGTATCAGTTAACTACCGAGTCGGACTCTTTGAGTGGCTTTGTGGGGACGAGAGA -----> ----->	2761 Gag-Pro-Pol	TCCATAAAACAATACCCCATGTCAAGAAGCCAGATCGGGGATCAAGCCCCACATACAG S I K Q Y P M S Q E A R L G I K P H I Q
481	CAGAGACACTTCCCGCCCCGCTGTAATTTTGTCTTTCGGTTTACGCGAAACCGCGCC	2821 Gag-Pro-Pol	AGGCTGTTGGACGAGGAATCTGGTACCTGCCAGTCCCCCTGGAACAGCCCTCGTA R L L D Q G I F L P N P Y N L S G L
541 GAG-OF primer	GCGCGTCTGATTGTGTTTGTGTCTTCTGTCTCTCGTTAGTTTCTCTGTCTTTAAGT -----> ----->	2881 Gag-Pro-Pol	CCCGTTAAGAAACAGGGGACTAATGATTATAGGCTGTCCAGGATCTGAGAGAAGTCAAC P V K K P G T N D Y R P V Q D L R E V N
601 Gag GAG-IF primer	GTCTCGAGATCATGGGACAGACCGTAACCTACCCCTCTGAGTCTAACCTTGCAGCACTGG -----> ----->	2941 Gag-Pro-Pol	AAGCGGTTGAAGACATCCACCCACCGTCCCAACCTTACAACTCTTGAGCGGGCTC K R V E D I H P T V P N P Y N L S G L
661 Gag	GGAGATGTCAGCGCATTCATCCAAACGAGTCTGTGGATGTCAAGAAGAGCGCTGGGTT G D V Q R I A S N Q S V D V K K R R W V	3001 Gag-Pro-Pol	CCACCGTCCCAACGAGTGTACATGTGCTTGAATTTAAAGGATGCCTTTTCTGCCTGAGA P P S H Q W Y T V L D L K D A F F C L R
721 Gag	ACCTTCTGTTCGCGCAATGGCCAACTTCAATGTAGGATGGCCTCAGGATGGTACTTTT T F C S A E W P T F N V G W P Q D G T F	3061 Gag-Pro-Pol	CTCCACCCCAACGATCAGCTCTCTCGCCTTGTAGTGGAGAGATCCAGAGATGGGAATC L H P T S Q P L F A F E W R D P E M G I
781 Gag	AATTTAGGTGTATCTCTCAGGTCAAGTCTAGAGTGTGTTTGTCTGTGTCGCCACGGACAC N L G V I S Q V K S R V F C P G P H G H	3121 Gag-Pro-Pol	TCAGGACAACTGACCTGGACAGACTCCCAAGGGTTTCAAAACAGCTCCCACTGTTT S G Q L T W T R L P Q G F K N S P T L F
841 Gag	CCGGATCAGGTCCCATATATCGTCACCTGGGAGGCACTTGCCTATGACCCCTCCGTGG P D Q V P Y I V T W E A L A Y D P P P W	3181 Gag-Pro-Pol	GATGAGGCACTGCACAGAGACCTAGCAGATTTCCGGATCCAGCACCCAGACTGTATCTG D E A L H R D L A D F R I Q H P D L I L
901 Gag	GTCAAAACCGTTTGTCTCTCTAAACCCCTCTTTACCGACAGCTCCCGTCTCTCCGCC V K P F V S P K P P P L P T A P V L P P	3241 Gag-Pro-Pol	CTACAGTACGTGGATGACTTACTGCTGGCGCCACTCTGAGCAAGACTGCCAACGAGGT L Q Y V D D L L L A A T S E Q D C Q R G
961 Gag GAG-IR primer	GGTCCTTCTGCGCAACCTCCGTCCGATCTGCCCTTTACCTGCCCTTACCTCTCTCTATA G P S A Q P P S R S A L Y P A L T L S I -----> ----->	3301 Gag-Pro-Pol	ACTCGGCCCCATTACAAACCTTAGGGAACCTCGGGTATCGGGCTCGGCCAAGAAAGCC T R A L L Q T L T G N L G Y R A S A K A
1021 Gag	AAGTCCAAACCTCCTAAGCCCGAGTTCTCCCTGATAGCGCGGACCTCTCATTGACCTT K S K P P K P Q V L P D S G G P L I D L	3361 Gag-Pro-Pol	CAAAATTTGCCAGAAACAGGTCAAGTATCTGGGTATCTCCTAAAGAGGGACAGAGATGG Q I C Q K Q V K Y L G Y L L K E G Q R W
1081 Gag GAG-OR primer	CTCACAGAGGATCCCCCGCGTACGAGGTACAACCTTCTCTCTGACAGGAGAAACAT L T E D P P P Y G V Q P S S S A R E N -----> ----->	3421 Gag-Pro-Pol	CTGACTGAGGCCAGAAAGAGACTGTGATGGGCGAGCCCACTCGAAGACCCCTCGACAA L T E A R K E T V M G Q P T P K T P C A
1141 Gag GAG-OR primer	GAAGAAGAGGCGGCCACCACTCCGAGGTTTCCCCCTCTCTCCCATGGTGTCTGAGT E E E A A T T S E V S P P S P M V S R L -----> ----->	3481 Gag-Pro-Pol	CTAAGGGAGTTCTAGGACGCGAGGCTCTGTCGCTCTGGATCCCTGGGTTTGACGAA L R E F L G T A G F C R L W I P G F A E
1201 Gag	CGGGGAAGGAGAGCCCTCCGCGAGCGGACTCCACCACTCCCGGCACTCCCACTCCG R G R R D I A A D S T T S Q A P L R	3541 Gag-Pro-Pol	ATGGCAGCCCCCTTGTAACCTCTTACCAAAACGCGGACTCTGTTTAATTTGGGCGCCAG M A A P L Y P L T K T G T L F N W A G A D
1261 Gag	ATGGGGGAGATGGCAGCTTCAGTACTGGCGGTTTCTCTCTGATTATATAAATGG M G G D G Q L Q Y W P F S S S D L Y N W	3601 Gag-Pro-Pol	CAGCAAAAGGCTATCAAGAAATCAACAGGCTCTTCTAACTGCCCCGCGCTGGGATTG Q Q K A Y Q E I K Q A L L T A P A L G L
1321 Gag	AAAAATAATAACCCCTCTCTTCTGAAGATCCAGGTAATTGACGGGCTTGATTGAGTCC K N N N P S E D P G K L T A L I E S	3661 Gag-Pro-Pol	CCAGATTGACTAAGCCCTTGAACCTTTGTGTCAGAGAGACGAGGCTACGCCAAAGGC P D L T K P F E L F V D E K Q G Y A K G
1381 Gag	GTCTCATCACCCACGACCCACTGGGACGACTGTGAGCAGTTGTTGGGACCCCTGCTG V L I T H Q P T W D D C Q Q L L G T L L	3721 Gag-Pro-Pol	GTCTTAACGCAAAACCTGGGACCTTGGCGTGGCTGTGGCTACCTGTCCAAAAAGCTA V L T Q K L G P W R R P V A Y L S K K L
1441 Gag	ACCGGAGAGAAAGACGCGGGTCTCTAGAGGCTGGAAGGACGAGTCCGGGCAATGAT T G E E K Q R V L L E A G K A V R G N D	3781 Gag-Pro-Pol	GACCCAGTGGCAGCTGGTGGCCCCCTTGCTACGAGTGTAGCAGATTCGCGTCTGTG D P V A A G W P P C L R M V A I A V L T
1501 Gag	GGACGCCCCACTCAGTTGCCTAATGAAGTCAATGCTGCTTTCCCTTGAGCGCCCGAT G R P T Q L P N E V N A A F P L E R P D	3841 Gag-Pro-Pol	ACAAAGGATGAGGCAAGCTAACTATGGGACAGCGCTAGTCATTCTGCCCCCATCTGCG T K D A G K L T M G Q P L V I L A P H A
1561 Gag	TGGGATTACCACTACAGAAGGTAGGAACCACTAGTCTCTACCGCCAGTTGCTCTTA W D Y T T T E G R N H L V L Y R Q L L L	3901 Gag-Pro-Pol	GTAGAAGCACTGGTCAAAACACCCCTGACCGTTGGCTATCCAAATGCCCGATGACCCAC V E A L V K Q P P D R W L S N A R M T H
1621 Gag	GCGGGTCTCCAAAACGCGGGCAGGAGCCCACTAATTTGGCAAAGTAAAGGGATAACC A G L Q N A G R S P T N L A K V K G I T	3961 Gag-Pro-Pol	TATCAGGCAATGCTCTGGATACAGACGCGGTTAGTTTCGAGCCGTTGGTGGCCCTCAAC Y Q A M L L D T D R V Q F G P V V A L N
1681 Gag	CAGGACCTAATGAGTCTCCCTCAGCCTTTTAGAGAGACTCAAGGAGGCTTATCGCAGG Q G P N E S P S A F L E R L K E A Y R R	4021 Gag-Pro-Pol	CCGGCCACCTGCTCCCCCTACCGGAAAGGAGGCCCCCATGACTGCTCGAGATCTTG P A T L L P L P E K E A P H D C L E I L
1741 Gag	TACACTCCTTATGACCTGAGGACCGAGGCAAGAAACCAATGTGTCATGTATTATC Y T P Y D P E D P G Q E T N V S M S F I	4081 Gag-Pro-Pol	GCTGAGACGACGGAACAGACCGGACCTCACGACGACCCATCCAGACGCTGATTAC A E T H G T R P D L T D Q P I P D A D Y
1801 Gag	TGGCAGTCTGCCCGGATATCGGGCGAAAGTTAGAGCGGTTAGAAGATTAAAGAGCAAG W Q S A P D I G R K L E R L E D L K S K	4141 Gag-Pro-Pol	ACTTGGTACACAGATGGAAGCAGCTTCTTACAAAGGACCAACGAGAGCTGGAGCAGCG T W Y T D G S P F L Q E G Q R R A G A A
1861 Gag	ACCTTAGGAGACTTAGTGAGGGAAGCTGAAAAATCTTTAATAAGCGAGAAACCCCGGAA T L G D L V R E A E K I F N K R E T P E	4201 Gag-Pro-Pol	GTGACTACTGAGACCGAGGTAATCTGGGCGAGGCTCTGCGGCTGGAACCTCGGCCAA V T T E T E V I W A R A L P A G T S A Q
1921 Gag	GAAAGAGAGGAAACGTAACGAGAGAGAAATAGAGAAAAAGAAACCGCGTAGGCGAGAG E R E E R I R R E I E E K E R R R A E	4261 Gag-Pro-Pol	CGAGCCGAACTGATAGCACTCAACCAACCTTAAAGTGGCAGAGAAAGTAAGKCTAAAT R A E L I A L T Q A L K M A E G K A L N
1981 Gag	GATGAGCAGAGAGAGAGAAAGGACCGCAGAAACATAGAGAGATGAGCAAGCTCTTG D E Q R E R E R D R R R H R E M S K L L	4321 Gag-Pro-Pol	GTTTCACTGATAGCGCTATGCTTTCGCAACGCGCCATGTCCATGGAGAAATATATAGG V Y T D S R Y A F A T A H V H G E I Y R
2041 Gag	GCCACTGTAGTTATTGGTCAGAGACGAGTAGACGAGGGGAGAGCGGAGGAGCCGCCAA A T V V I G R Q D R Q G I G E R R R P Q	4381 Gag-Pro-Pol	AGGCGAGGGTGTGTAACCTCAGAAGGACAGAAATTAAGAAACGAGACTCTGGCC R R G L L T S E G R E I K N K N E I L A
2101 Gag	CTTGATAAGGACCAATGCGCTACTGCAAGAAAAAGGACACTGGGCTAAGGACTGCCCA L D K D Q C A Y C K E K G H W A K D C P	4441 Gag-Pro-Pol	TTGCTAAAGCTCTCTTTCTGCCCAACGACTTAGTATAATTCATGCTCCAGGACATCAA L L K A L F L P K R L S I I H C P G H Q
2161 Gag	AAGAAGCCACGAGGCGCCGAGGACCGGAGCCCGGACCTCCCTCTGACTTATGTTGAC K K P R G P R G P R P Q T S L T L G D	4501 Gag-Pro-Pol	AAAGGAACAGTGTGAGGCGAGGCAACCGTATGGCAGATCAAGCAGCCGAGGAGGCA K G N S A E A R G N R M A D Q A R E A
2221 Gag-Pro-Pol	TAGGAGGTGAGGTCAGGAGCCCCCTGAAACCGAGGATAACCTCAAAGTCGGGGG * G G Q G Q E P P P E P R I T L K V G G	4561 Gag-Pro-Pol	GCCATGAAGGAGTCTAGAAACCTCTACACTCTCATAGGAGCTCAACCCCGTATACG A M K A V L E T S T L L I E D S T P Y T
2281 Gag-Pro-Pol	CAACCCGTCACCTTCTGGTAGATCTGGGGCCCAACACTCCGTGCTGACCCAAATCCT Q P V T F L V D T G A Q H S V L T Q N P	4621 Gag-Pro-Pol	CCTCCCCATTTCATTACACCGAAACAGATCTCAAAAGACTACGGGAATCGGAGCCACC P P H F H Y T E T D L K R L R E L G A C
		4681 Gag-Pro-Pol	TACAAATCAGACAAAGGATATTGGGTCCTACAAGGCAACCTGTGATGCCGATCAGTCC Y N Q T K G Y W V L Q G K P V M P D Q S

Figure S1 (continued)

4741 Gag-Pro-Pol	GTGTTTGAACCTGTTAGACTCCCTACACAGACTCACCCATCCGAGCCCTCAAAGATGAAG V F E L L D S L H R L T H P S P Q K M K	6541 Env	GTCTCTCTCTTCAGGCGCGCCTCTATGGTGCTGGGGCTCCCCGCCTTCTCAACAC R P P P S G A A S M V P G A P P P S Q Q
4801 Gag-Pro-Pol	GCAGTCTCTCGACAGAGAAGACCCCTACTACATGTTAAACCGGACAGAACTTACCTCAG A L L D R E E S P Y Y M L N R D R T I Q	6601 Env	CTGGGACGGAGACAGGCTGTCTAAACCTGGTAGAAGGAGCTACCAAGCCCTCAACCTCA P G T G D R L L N L V E G A Y Q A L N L
4861 Gag-Pro-Pol	TATGTGACTGAGACTGCAACCCGCTGTGCCAAGTAAATGCCAGCAAGACCAAAATGGG Y V T E T C T A C A Q V N A S K A K I G	6661 Env	CCAGTCCCGACAAAACCAAGAGTGTGGCTGTGTCTAGTATCGGGACCCCTACTACG T S P D K T Q E C W L C L V S G P P Y Y
4921 Gag-Pro-Pol	GCAGGGGTGCGAGTAGCGGACATCGGCCAGCACCCATTGGGAAGTTGATTTCAGGAA A G V R V R G H R P G T H W E V D F T E	6721 Env	AAGGGGTGCGCGTCTAGTACTTACTCCAACCATACCTCTGCCCCGGCTACTGCTCCG E G V A V L G T Y S N H T S A P A N C S
4981 Gag-Pro-Pol	GTAAAGCCAGGACTGTATGGGTACAAGTACCTCCTAGTGTGTTGTAGACACCTTCTCTGGC V K P G L Y G Y K Y L L V F V D T F S G	6781 Env	TGACCTCCCAACACAGCTGACCTGTCCGAAGTGACCGGACAGGACTCTGCATAGGAG V T S Q H K L T L S E V T G Q G L C I G
5041 Gag-Pro-Pol	TGGGTAGAGGATTACCGACCAAGCGGAACTGCCAAGTTGTGACCAAAAGCTGTTA W V E A F P T K R E T A K V V T K K L L	6841 Env	CAGTTCCCAAAACCATCAGGCCCTGTGTAATACCCACCAAGACGACGACGCGGGCTCT A V P K T H Q A L C N T T T Q K T S D G S
5101 Gag-Pro-Pol	GAAGACATTTTTCGAGATTGGAAATGCCGACAGTATTGGGATCTGATACCGGCTGCC E D I F P R P F G M P Q V L G S D N G P A	6901 Env	ACTATTGGCTCTCTCCGCGCGGACCATTTGGGCTTCGAGCACCGGGCTACTCCCTGTC Y Y L A S P A G T I W A C S T G L T P C
5161 Gag-Pro-Pol	TTCCGCTCCCAAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTC F A S Q V S Q S V A D L L G I D W K L H	6961 Env	TATCTACTACTGTGCTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT L S T T V L N L T T D Y C V L V E L W P
5221 Gag-Pro-Pol	TGTGCTTATAGACCCAGAGTTCAAGACAGGTAGAAAGAAATGAATAGAACAAATAGAGAG C A Y R P Q S S G Q V E R M N R T I K E	7021 Env	AGGTAACCTACCACTCCCTCAATTTATGTTTATGGCAGTTTGGAAAGAAATTAATATA K V T Y H S P N Y V Y G Q F G K K T K Y
5281 Gag-Pro-Pol	ACTTTGACCAAAATTACGCTTGCATCTGGCACTAGAGACTGGGTACTCTACTCCCTTTA T L T K L T L A S G T R D W V L L L P L	7081 Env	AAAGAGAGCCGCTGTCATTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT K R E P V S L T L A L L L G G L T M G G
5341 Gag-Pro-Pol	GCCCTCTACCGAGCCCGGAACTACTCGGGGCCCAAGGACTGACTCCGATGAAATTCG A L Y R A R N T P G P H G L T P Y E I L	7141 Env	TAGCTGCAGGAGTTGGAACAGGACTACAGCCCTAGTGGCCACCAACATTCGAGCAGC I A A G V G T G T T A L V A T K Q F E Q
5401 Gag-Pro-Pol	TATGGGGCACCGCCGCTGTCAATTTTCATGATCTGAAATGCTCAAGATTAACCTAAT Y G A P P P L V N F H D P E M S K L T N	7201 Env	TCCAGGCAGCATACATACAGACTCTGGGGCTTAGAAAAATCAGTCAGTCAGTCAGTCAGT L Q A A I H T D L G A L E K S V S A L E
5461 Gag-Pro-Pol splice accep	AGTCCCTCTCTCCAGCTCACTTACAGGCCCTCCAGCAGTACAACAGAGGTCTGGAAG S P S L Q A H L Q A L Q A V Q Q E V W K ----->	7261 Env	AGTCTCTGACCTCGTTGTCTGAGGTGGTCTTACAGAACCGGAGGGATAGATCTACTGT K S L T S L S E V V L Q N R R G L D L L
5521 Gag-Pro-Pol	CGCGTGGCGCTGCTTATCAGGACAGCTAGATCAGCCAGTGATACCAACCCCTCCGT P L A A A Y Q D Q L D Q P V I P H P F R	7321 Env	TCCTAAAGAAGGAGGATTATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT F L K E G G L C A A L K K E C C F Y A D
5581 Gag-Pro-Pol	GTCGTTGACCGCGCTGTGGTAGCGGCCACAGACTAAGAACTTAGAAGCTCGCTGGAAA V G D A V W V R R H Q T K N L E P R W K	7381 Env	ACACTGGCGTAGTAAGAGATAGCATGGCAAGCTAAGAGAAAGTTTAAACAGAGACAAA H T G V V R D S M A K L R E R L N Q R Q
5641 Gag-Pro-Pol	GGACCCCTACCGCTCTGCTGACAAACCCCAACCGCTCTCAAAGTAGAGCGGACTCTCTCGG G P Y T V L L T T P T A L K V D G I S A	7441 Env	AATTGTTGGAATCAGGACAGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT K L F E S G Q G W F E G L F N R S P W F
5701 Gag-Pro-Pol Env	TGGATACACCGCGCTCAGCTAAAGCGGGCAGAACCTCTCCGCGGACAGCATGGAAGA W I H A A H V K A A T T P P A G T A W K M E	7501 Env	CGACCTGATATCCCACTTATGGGCCCTCTGATAGTACTTTTATTAATCTACTCTTCG T T L I S T I M G P L I V L L L I L L F
5761 Gag-Pro-Pol Env	GTCCAGCGTTCTCAAACCCCTTAAAGATAAGATTAAACCGTGGGGCCCTGATTAATTA V Q R S Q N P L K I R L T R G A P * S P A F S K P L K D K I N P W G P L I I	7561 Env	GACCCTGTATTCTCAACCGCTTGGTCCAGTTTGGAGGACTGTTTAAACAGGTCCCATGGCAGG G P C I L N R L V Q F V K D R I S V V Q
5821 Env	TGGGGATCTTGTGTGAGGGCAGGAGCTCAGTACAGCTGACAGCCCTCAACAGCTCTTTA M G I L V R A G A S V Q R D S P H Q V F	7621 Env	CCCTGGTTCTGACCCACAGTATCCCACTCAATCAATAGATCCAGAGAGAGTGGAAAT A L V L T Q Q Y H Q L K S I D P E E V E
5881 Env	ATGTCACTTGGAAAAATTACCACTAATGACAGGACAAACAGCTAATGACTACTCCCTCC N V T W K I T N L M T G Q T A N A T S L	7681 Env U3	CAGTGAAATAAAGATTTTATTCAGTTTCCAGAAAGGGGGGAATGAAGACCCACCA S R E *
5941 Env	TGGGGACGATGACAGACACTTTCCTAAACTATATTTTGAATTTGTGTGATTAGTTGGAG L G T M T D T F P K L Y F D L C D L V G	7741 U3	TAAGGCTTAGCACGCTAGCTACAGTAACGCCATTTTGAAGGCATGGAAGATACCAGAG -----
6001 Env	ACAACTGGGATGACCCGGAACCGCATATTGGAGATGGTTGCCGCTCTCCCGGGGAAGAA D N W D D P E P D I G D G C R S P G G R	7801 U3	CTGAGTTCTCAAAGTTACAGGAAGTTTAAATAAGAAATAGGCTGAATAACACTGGGA -----
6061 Env	AAAGGACAAGACTATATGATTTCTATGTTTGGCCCGGTCACTAGTATTAACAGGGTGTG K R T R L Y D F Y V C P G H T V L T G C	7861 U3 GRE	CAGGGGCCAAACAGGATATCTGTAGTCAGGCACCTGGGCCCGGCTCAGGGCCAAAGAAC -----
6121 Env	GAGGGCCGAGAGAGGGCTACTGTGGCAATGGGATGTGAGACCACTGGACAGGCATACT G G P R E G Y C G K W G C E T T G Q A Y	7921 U3 GRE	GATGGTCTCAGATAAAGCGAACTAAACAGCTTTCTGGAAGTCCCACTCAGTTTCA ----->
6181 Env	GGAGGCCATCATCATCTGGGACCTAATTTCCCTTAAGCGAGGAACACTCCTAAGGGTC W K P S S S W D L I S L K R G N T P K G	7981 U3 CAAT box	AGTTCCTCCAAAGACCGGGAATACCCAGCCTTATTTAACTAACCATCAGCTCGCT ----->
6241 Env	AGGGCCCTGTTTGTATCTCTAGTGGGTCCGGTAGCATCCAGGGTGCCACACCGGGGG Q G P C F D S S V G S G S I Q G A T P G	8041 U3 TATA box	TCTCGCTTCTGTACCGCGCTTTTGTCTCCCACTGCTAGCCCTATAAAAAAGGGTAAAG ----->
6301 Env	GTGATGCAACCCCTAGTCTAGAAATCACTGACGCGGTAAAGGGCCAGCTGGGATG G R C N P L V L E F T D A G K R A S W D	8101 U3 R	AACTCCACACTCGGCGCGCAGTCATCCGATAGACTGAGTCGCGGGTACCCGTGTTC ----->
6361 Env	CCCCAAAACATGGGGACTAAGACTGTATCGATCCACTGGGGCCGACCCGGTGACCTGT A P K T W G L R L Y R S T G A D P V T L	8161 R polyA signal polyA site	CAATAAGCCTTTTGTGTTTGCAA -----> ----->
6421 Env	TCTCTCTGACCGCGAGTCTCAATGTAGGGCCCCGCGTCCCAATTGGGCTTAATCCCG F S L T R Q V L N V G P R V P I G P N P		
6481 Env	TGATCACTGAACAGCTACCCCTCCCAACCGGTGACATCATGCTCCCAAGGCTCTCTC V I T E Q L P P S Q P V Q I M L P R P P		