

```

1  tgtggggaaa agcaagagag atcagattgt tactgtgtct gtgtagaaa
51  aagtagacat aggagactcc attttgttat gtactaagaa aaattcttct
101 gccttgagat tctgttaatc tatgacctta cccccaaccc cgtgctctct
151 gaaacatgtg ctgtgtcaac tcagagttaa atggattaag ggcggtgcag
201 gatgtgcttt gttaaacaga tgcttgaagg cagcatgctc cttaagagtc
251 atcaccactc cctaattctca agtaccagg gacacaaaaa ctgcggaagg
301 ccgcagggac ctctgcctag gaaagccagg tattgtccaa ggtttctccc
351 catgtgatag tctgaaatat ggctcgtgg gaagggaaa acctgaccgt
401 ccccagccc gacacccgta aagggtctgt gctgaggagg attagtaaaa
451 gaggaaggaa tgcctcttgc agttgagaca agaggaaggc atctgtctcc
501 tgcctgtccc tgggcaatgg aatgtctcgg tataaaaccc gattgtatgc
551 tccatctact gagataggga aaaaccgcct tagggctgga ggtgggacct
601 gcgggcagca atactgcttt gtaaagcact gagatgttta tgtgtatgca
651 tatctaaaag cacagcactt aatcctttac attgtctatg atgcaaagac
701 ctttgttcac gtgtttgtct gctgaccctc tccccacaat tgtcttga
751 ccctgacaca tccccctctt tgagaaacac ccacagatga tcaataaata
801 ctaagggAAC tcagaggctg gcgggacCt ccatatgctg aacgctggTt
851 ccccgggtcc ccttatttct ttctctatac tttgtctctg tgtcttttC
901 ttttccaaat ctctcgTccc accttacgag aaacaccCac aggtgtgtag
951 gggcaaccCa cccctacatc tggTgcccAa cgtggaggct tttctctagg
1001 gtgaaggTac gctcgagcgt ggtcattgag gacaagTcga cgagagatcc
1051 cgagtacgtc tacagTcagc cttacggtaa gcttgtgcgc tcggaagaag
1101 ctagggtgat aatggggcaa actaaaagta aaattaaaag taaatatgcc
      M G Q T K S K I K S K Y A
1151 tcttatctca gctttattaa aattctttta aaaagagggg gagttaaagt
      S Y L S F I K I L L K R G G V K
1201 atctacaaaa aatctaatac agctatttca aataatagaa caattttgcc
      V S T K N L I K L F Q I I E Q F C
1251 catggtttcc agaacaagga actttagatc taaaagattg gaaaagaatt
      P W F P E Q G T L D L K D W K R I
1301 ggtaaggaac taaaacaagc aggtaggaag ggtaatatca ttccacttac
      G K E L K Q A G R K G N I I P L
1351 agtatggaat gattgggcca ttattaaagc agcttttagaa ccatttcaaa
      T V W N D W A I I K A A L E P F Q
1401 cagaagaaga tagcgTttca gtttctgatg cccctggaag ctgtataata
      T E E D S V S V S D A P G S C I I

```

GAG

1451 gattgtaatg aaaacacaag gaaaaaatcc cagaaagaaa cggaagggttt
D C N E N T R K K S Q K E T E G
1501 acattgcgaa tatgtagcag agccggtaat ggctcagtca acgcaaaatg
L H C E Y V A E P V M A Q S T Q N
1551 ttgactataa tcaattacag gaggtgatat atcctgaaac gttaaaatta
V D Y N Q L Q E V I Y P E T L K L
1601 gaaggaaaag gtccagaatt agtggggcca tcagagtcta aaccacgagg
E G K G P E L V G P S E S K P R
1651 cacaagtcct cttccagcag gtcaggtgcc cgtaacatta caacctcaaa
G T S P L P A G Q V P V T L Q P Q
1701 agcagggttaa agaaaataag acccaaccgc cagtagccta tcaatactgg
K Q V K E N K T Q P P V A Y Q Y W
1751 ccgccggctg aacttcagta tcggccaccc ccagaaagtc agtatggata
P P A E L Q Y R P P P E S Q Y G
1801 tccaggaatg cccccagcac cacagggcag ggccgcatat cctcagccgc
Y P G M P P A P Q G R A P Y P Q P
1851 ccactaggag acttaatcct acggcaccac ctagtagaca gggtagtgaa
P T R R L N P T A P P S R Q G S E
1901 ttacatgaaa ttattgataa atcaagaaag gaaggagata ctgaggcatg
L H E I I D K S R K E G D T E A
1951 gcaattccca gtaacgttag aaccgatgcc acctggagaa ggagcccaag
W Q F P V T L E P M P P G E G A Q
2001 agggagagcc tcccacagtt gagggcagat acaagtcttt ttcgataaaa
E G E P P T V E A R Y K S F S I K
2051 atgctaaaag atatgaaaga gggagtataa cagtatggac ccaactcccc
M L K D M K E G V K Q Y G P N S
2101 ttatatgagg acattattag attccattgc tcatggacat agactcattc
P Y M R T L L D S I A H G H R L I
2151 cttatgattg ggagattctg gcaaaatcgt ctctctcacc ctctcaattt
P Y D W E I L A K S S L S P S Q F
2201 ttacaattta agacttggtg gattgatggg gtacaagaac aggtccgaag
L Q F K T W W I D G V Q E Q V R
2251 aaatagggct gccaatcctc cagttaacat agatgcagat caactattag
R N R A A N P P V N I D A D Q L L
2301 gaataggcca aaattggagt actattagtc aacaagcatt aatgcaaaat
G I G Q N W S T I S Q Q A L M Q N
2351 gaggccattg agcaagttag agctatctgc cttagagcct gggaaaaaat
E A I E Q V R A I C L R A W E K
2401 ccaagacca ggaagtacct gccctcatt taatacagta agacaaggtt
I Q D P G S T C P S F N T V R Q G
2451 caaaagagcc ctatcctgat tttgtggcaa ggctccaaga tgttgctcaa
S K E P Y P D F V A R L Q D V A Q
2501 aagtcaattg ccgatgaaaa agcccgttaag gtcatagtgg agttgatggc
K S I A D E K A R K V I V E L M

2551 atatgaaaac gccaatcctg agtgtcaatc agccattaag ccattaaaag
 A Y E N A N P E C Q S A I K P L K
 2601 gaaagggttc tgcaggatca gatgtaatct cagaatatgt aaaagcctgt
 G K V P A G S D V I S E Y V K A C
 2651 gatggaatcg gaggagctat gcataaagct atgcttatgg ctcaagcaat
 D G I G G A M H K A M L M A Q A
 2701 aacaggagtt gttttaggag gacaagttag aacatttggg ggaaaatggt
 I T G V V L G G Q V R T F G G K C
 2751 ataattgtgg tcaaattggg cacttaaaaa agaattgccc agtcttaaac
 Y N C G Q I G H L K K N C P V L N
 2801 aaacagaata taactattca agcaactaca acaggtagag agccacctga
 K Q N I T I Q A T T T G R E P P
 2851 cttatgtcca agatgtaaaa aaggaaaaca ttgggctagt caatgtcggt
 D L C P R C K K G K H W A S Q C R
 2901 ctaaatttga taaaaatggg caaccattgt cgggaaacga gcaaaggggc
 - K W A T I V G K R A K G
 S K F D K N G Q P L S G N E Q R G
 2951 cagcctcagg ccccaaca aactggggca ttcccaattc agccatttgt
 P A S G P T T N W G I P N S A I C
 Q P Q A P Q Q T G A F P I Q P F
 3001 tcctcagggt tttcagggac aacaaccccc actgtcccaa gtgtttcagg
 S S G F S G T T T P T V P S V S
 V P Q G F Q G Q Q P P L S Q V F Q
 3051 gaataagcca gttaccacaa tacaacaatt gtccccgcc acaagcggca
 G N K P V T T I Q Q L S P A T S G
 G I S Q L P Q Y N N C P P P Q A A
 3101 gtgcagcagt agatttatgt actatacaag cagtctctct gcttccaggg
 S A A V D L C T I Q A V S L L P G
 V Q Q -
 3151 gagccccac aaaaaatccc cacaggggta tatggcccc tgccctgagg
 E P P Q K I P T G V Y G P L P E
 3201 gactgtagga ctaatcttgg gaagatcaag tctaaatcta aaaggagttc
 G T V G L I L G R S S L N L K G V
 3251 aaattcatat tagtgtggtt gattcagact ataaaggcga aattcaattg
 Q I H T S V V D S D Y K G E I Q L
 3301 gttattagct cttcaattcc ttggagtgcc agtccaggag acaggattgc
 V I S S S I P W S A S P G D R I
 3351 tcaattatta ctctgccat atattaaggg tggaaatagt gaaataaaaa
 A Q L L L L P Y I K G G N S E I K
 3401 gaataggagg gcttgaagc actgatccaa caggaaaggc tgcattattg
 R I G G L G S T D P T G K A A Y W
 3451 gcaagtcagg tctcagagaa cagacctgtg tgtaaggcca ttattcaagg
 A S Q V S E N R P V C K A I I Q
 3501 aaaacagttt gaagggttgg tagacactgg agcagatgtc tctatcattg
 G K Q F E G L V D T G A D V S I I

PRO

GAG

3551 ctttaaatca gtggccaaaa aattggccta aacaaaaggc tgttacagga
 A L N Q W P K N W P K Q K A V T G
 3601 cttgtcggca taggcacagc ctcagaagtg tatcaaagta cggagatttt
 L V G I G T A S E V Y Q S T E I
 3651 acattgctta gggccagata atcaagaaag tactgttcag ccaatgatta
 L H C L G P D N Q E S T V Q P M I
 3701 cttcaattcc tcttaatctg tggggtcgag atttattaca acaatgggg
 T S I P L N L W G R D L L Q Q W G
 3751 gcggaaatca ccatgcccgc tccattatat agccccacga gtcaaaaaat
 A E I T M P A P L Y S P T S Q K
 3801 catgaccaag atgggatata taccaggaaa gggactaggg aaaaatgaag
 I M T K M G Y I P G K G L G K N E
 3851 atggcattaa aattccagtt gaggctaaaa taaatcaaaa aagagaagga
 D G I K I P V E A K I N Q K R E G
 - N K S K K R R
 3901 atagggtatc ctttttaggg gcggccactg tagagcctcc taaaccata
 I G Y P F -
 N R V S F L G A A T V E P P K P I
 3951 ccattaactt ggaaaacaga aaaaccgggtg tgggtaaatc agtggccgct
 P L T W K T E K P V W V N Q W P
 4001 accaaaacaa aaactggagg ctttacattt attagcaaat gaacagttag
 L P K Q K L E A L H L L A N E Q L
 4051 aaaagggcca tattgagcct tcgttctcac cttggaattc tcctgtgttt
 E K G H I E P S F S P W N S P V F
 4101 gtaattcaga agaaatcagg caaatggcgt atgttaactg acttaagggc
 V I Q K K S G K W R M L T D L R
 4151 tgtaaacgcc gtaattcaac ccatggggcc tctccaaccc gggttgcctt
 A V N A V I Q P M G P L Q P G L P
 4201 ctccggccat gatcccaaaa gattggcctt taattataat tgatctaaag
 S P A M I P K D W P L I I I D L K
 4251 gattgctttt ttaccatccc tctggcagag caggattgtg aaaaatttgc
 D C F F T I P L A E Q D C E K F
 4301 ctttactata ccagccataa ataataaaga accagccacc aggtttcagt
 A F T I P A I N N K E P A T R F Q
 4351 ggaaagtgtt acctcaggga atgcttaata gtccaactat ttgtcagact
 W K V L P Q G M L N S P T I C Q T
 4401 tttgtaggtc gagctcttca accagttaga gaaaagtttt cagactgtta
 F V G R A L Q P V R E K F S D C
 4451 tattattcat tatattgatg atattttatg tgctgcagaa acgaaagata
 Y I I H Y I D D I L C A A E T K D
 4501 aattaattga ctgttataga tttctgcaag cagaggttgc caatgtctga
 K L I D C Y T F L Q A E V A N A G
 4551 ctggcaatag catctgataa gatccaaaacc tctactcctt ttcattattt
 L A I A S D K I Q T S T P F H Y

POL

PRO

4601 agggatgcag atagaaaata gaaaaattaa gccacaaaaa atagaaataa
L G M Q I E N R K I K P Q K I E I

4651 gaaaagacac attaaaaaca ctaaattgatt ttcaaaaatt actaggagat
R K D T L K T L N D F Q K L L G D

4701 attaatggga ttcggccaac tctaggcatt cctacttatg ccatgtcaaa
I N W I R P T L G I P T Y A M S

4751 tttgtttctct atcttaagag gagactcaga cttaaatagt aaaagaatgt
N L F S I L R G D S D L N S K R M

4801 taaccccaga ggcaacaaaa gaaattaaat tagtggaaga aaaaattcag
L T P E A T K E I K L V E E K I Q

4851 tcagcgcaaa taaatagaat agatccctta gcccactcc aacttttgat
S A Q I N R I D P L A P L Q L L

4901 ttttgccact gcacattctc caacaggcat cattattcaa aatactgac
I F A T A H S P T G I I I Q N T D

4951 ttgtggagtg gtcattcctt cctcacagta cagttaagac ttttacattg
L V E W S F L P H S T V K T F T L

5001 tacttggatc aaatagctac attaatcggg cagacaagat tacgaataat
Y L D Q I A T L I G Q T R L R I

5051 aaaattatgt ggaaatgacc cagacaaaat agttgtccct ttaaccaagg
I K L C G N D P D K I V V P L T K

5101 aacaagttag acaagccttt atcaattctg gtgcatggca gattggctct
E Q V R Q A F I N S G A W Q I G L

5151 gctaattttg tgggaattat tgataatcat tacccaaaaa caaagatctt
A N F V G I I D N H Y P K T K I

5201 ccagttctta aaattgacta cttggattct acctaaaatt accagacgtg
F Q F L K L T T W I L P K I T R R

5251 aacctttaga aaatgctcta acagtattta ctgatggttc cagcaatgga
E P L E N A L T V F T D G S S N G

5301 aaagcagctt acacagggcc gaaagaacga gtaatcaaaa ctccatatca
K A A Y T G P K E R V I K T P Y

5351 atcggctcaa agagcagagt tggttgcagt cattacagtg ttacaagatt
Q S A Q R A E L V A V I T V L Q D

5401 ttgaccaacc tatcaatatt atatcagatt ctgcatatgt agtacaggct
F D Q P I N I I S D S A Y V V Q A

5451 acaagggatg ttgagacagc tctaattaaa tatagcatgg atgatcagtt
T R D V E T A L I K Y S M D D Q

5501 aaaccagcta ttcaatttat tacaacaaac tgtaagaaaa agaaatttcc
L N Q L F N L L Q Q T V R K R N F

5551 cattttatat tactcatatt cgagcacaca ctaatttacc agggcctttg
P F Y I T H I R A H T N L P G P L

5601 actaaagcaa atgaacaagc tgacttactg gtatcatctg cactcataaa
T K A N E Q A D L L V S S A L I

5651 agcacaagaa cttcatgctt tgactcatgt aaatgcagca ggattaaaaa
K A Q E L H A L T H V N A A G L K

5701 acaaatttga tgtcacatgg aaacaggcaa aagatattgt acaacattgc
N K F D V T W K Q A K D I V Q H C

5751 acccagtgtc aagtcttaca cctgcccact caagaggcag gagttaatcc
T Q C Q V L H L P T Q E A G V N

5801 cagaggtctg tgtcctaata cattatggca aatggatgtc acgcatgtac
P R G L C P N A L W Q M D V T H V

5851 cttcatttgg aagattatca tatgttcatg taacagttga tacttattca
P S F G R L S Y V H V T V D T Y S

5901 catttcatat gggcaacttg ccaaacagga gaaagtactt cccatgttaa
H F I W A T C Q T G E S T S H V

5951 aaaacattta ttgtcttgtt ttgctgtaat gggagttcca gaaaaaatca
K K H L L S C F A V M G V P E K I

6001 aaactgacaa tggaccagga tattgtagta aagctttcca aaaattctta
K T D N G P G Y C S K A F Q K F L

6051 agtcagtgtg aaatttcaca tacaacagga attccttata attcccaagg
S Q W K I S H T T G I P Y N S Q

6101 acaggccata gttgaaagaa ctaatagaac actcaaaact caattagtta
G Q A I V E R T N R T L K T Q L V

6151 aacaaaaaga agggggagac agtaaggagt gtaccactcc tcagatgcaa
K Q K E G G D S K E C T T P Q M Q

6201 cttaatctag cactctatac tttaaatttt ttaaacattt atagaaatca
L N L A L Y T L N F L N I Y R N

6251 gactactact tctgcagaac aacatcttac tggtaaaaag aacagccac
Q T T T S A E Q H L T G K K N S P

6301 atgaaggaaa actaatttgg tggaaagata ataaaaataa gacatgggaa
H E G K L I W W K D N K N K T W E

6351 atagggaagg tgataacgtg ggggagaggt tttgcttgtg tttcaccagg
I G K V I T W G R G F A C V S P

6401 agaaaatcag cttcctgttt ggataccac tagacatttg aagttctaca
G E N Q L P V W I P T R H L K F Y

6451 atgaacccat cagagatgca aagaaaagca cctccgcgga gacggagaca
M N P S E M Q R K A P P R R R R
N E P I R D A K K S T S A E T E T

6501 ccgcaatcga gcaccgttga ctcacaagat gaacaaaatg gtgacgtcag
H R N R A P L T H K M N K M V T S
P Q S S T V D S Q D E Q N G D V

6551 aagaacagat gaagttgcca tccaccaaga aggagagcc gccgacttgg
E E Q M K L P S T K K A E P P T W
R R T D E V A I H Q E G R A A D L

6601 gcacaactaa agaagctgac gcagttagct acaaaatc tagagaacac
A Q L K K L T Q L A T K Y L E N
G T T K E A D A V S Y K I S R E H

6651 aaaggtgaca caaacccag agagtatgct gcttgagcc ttgatgattg
T K V T Q T P E S M L L A A L M I
K G D T N P R E Y A A C S L D D

ENV

6701 tatcaatggt ggtaagtctc cctatgcctg caggagcagc tgcagctaac
V S M V V S L P M P A G A A A A N
C I N G G K S P Y A C R S S C S -

6751 tatacctact gggcctatgt gcctttcccg cccttaattc gggcagtcac
Y T Y W A Y V P F P P L I R A V

6801 atggatggat aatcctatag aagtatatgt taatgatagt gtatgggtac
T W M D N P I E V Y V N D S V W V

6851 ctggcccat agatgatcgc tgccctgcc aacctgagga agaagggatg
P G P I D D R C P A K P E E E G M

6901 atgataaata ttccattgg gtatcgttat cctcctattt gcctagggag
M I N I S I G Y R Y P P I C L G

6951 agcaccagga tgtttaatgc ctgcagtcca aaattggttg gtagaagtac
R A P G C L M P A V Q N W L V E V

7001 ctactgtcag tcccatcagt agattcactt atcacatggt aagcgggatg
P T V S P I S R F T Y H M V S G M

7051 tcactcaggc cacgggtaaa ttatttacia gacttttctt atcaaagatc
S L R P R V N Y L Q D F S Y Q R

7101 attaaaattt agacctaaag ggaaaccttg cccaaggaa attcccaaag
S L K F R P K G K P C P K E I P K

7151 aatcaaaaaa tacagaagtt ttagtttggg aagaatgtgt ggccaatagt
E S K N T E V L V W E E C V A N S

7201 gcggtgatat tacaaaacia tgaatttggg actattatag attgggcacc
A V I L Q N N E F G T I I D W A

7251 tcgaggtcaa ttctaccaca attgctcagg acaaactcag tcgtgtccaa
P R G Q F Y H N C S G Q T Q S C P

7301 gtgcacaagt gagtccagct gttgatagcg acttaacaga aagtttagac
S A Q V S P A V D S D L T E S L D

7351 aaacataagc ataaaaaatt gcagtctttc tacccttggg aatggggaga
K H K H K K L Q S F Y P W E W G

7401 aaaaggaatc tctaccccaa gacaaaaat aataagtcct gtttctggtc
E K G I S T P R P K I I S P V S G

7451 ctgaacatcc agaattatgg aggttactg tggcctcaca ccacattaga
P E H P E L W R L T V A S H H I R

7501 atttggtctg gaaatcaaac tttagaaaca agagatcgta agccatttta
I W S G N Q T L E T R D R K P F

7551 tactgtcgac ctaaattcca gtctaacagt tcctttacia agttgcgtaa
Y T V D L N S S L T V P L Q S C V

7601 agccccctta tatgctagtt gtaggaaata tagttattaa accagactcc
K P P Y M L V V G N I V I K P D S

7651 cagactataa cctgtgaaaa ttgtagattg cttacttgca ttgattcaac
Q T I T C E N C R L L T C I D S

7701 ttttaattgg caacaccgta ttctgctggg gagagcaaga gagggcgtgt
T F N W Q H R I L L V R A R E G V

7751 ggatccctgt gtccatggac cgaccgtggg aggcctcacc atccatccat
W I P V S M D R P W E A S P S I H

POL

7801 attttgactg aagtattaaa aggtgtttta aatagatcca aaagattcat
I L T E V L K G V L N R S K R F

7851 ttttacttta attgcagtga ttatgggatt aattgcagtc acagctacgg
I F T L I A V I M G L I A V T A T

7901 ctgctgtagc aggagttgca ttgcactctt ctgttcagtc agtaaacttt
A A V A G V A L H S S V Q S V N F

7951 gttaatgatt ggcaaaaaaa ttctacaaga ttgtggaatt cacaatctag
V N D W Q K N S T R L W N S Q S

8001 tattgatcaa aaattggcaa atcaaattaa tgatcttaga caaactgtca
S I D Q K L A N Q I N D L R Q T V

8051 tttggatggg agacagactc atgagcttag aacatcgttt ccagttacaa
I W M G D R L M S L E H R F Q L Q

8101 tgtgactgga atacgtcaga tttttgtatt acaccccaaa tttataatga
C D W N T S D F C I T P Q I Y N

8151 gtctgagcat cactgggaca tggtagacg ccatctacag ggaagagaag
E S E H H W D M V R R H L Q G R E

8201 ataatctcac tttagacatt tccaaattaa aagaacaaat tttcgaagca
D N L T L D I S K L K E Q I F E A

8251 tcaaaagccc atttaaattt ggtgccagga actgaggcaa ttgcaggagt
S K A H L N L V P G T E A I A G

8301 tgctgatggc ctcgcaaadc ttaaccctgt cacttgggtt aagaccattg
V A D G L A N L N P V T W V K T I

8351 gaagtactac gattataaat ctcatattaa tccttgtgtg cctgttttgt
G S T T I I N L I L I L V C L F C

8401 ctgttgtagt tctgcagggt tacccaacag ctccgaagag acagcgacca
L L L V C R C T Q Q L R R D S D
S A G V P N S S E E T A T

8451 tcgagaacgg gccatgatga cgatggcggg tttgtcgaaa agaaaagggg
H R E R A M M T M A V L S K R K G
I E N G P -

8501 gaaatgtggg gaaaagcaag agagatcaga ttgttactgt gtctgtgtag
G N V G K S K R D Q I V T V S V -

8551 aaagaagtag acataggaga ctccattttg ttatgtacta agaaaaattc

8601 ttctgccttg agattctgtt aatctatgac cttaccccca acccgtgct

8651 ctctgaaaca tgtgctgtgt caactcagag ttgaatggat taagggcggg

8701 gcaggatgtg ctttggttaa cagatgcttg aaggcagcat gctccttaag

8751 agtcatcacc actcccta atcaagtacc caggacaca aaaactgcgg

8801 aaggccgcag ggacctctgc ctaggaaagc caggatttgt ccaaggtttc

8851 tccccatgtg atagtctgaa atatggcctc gtgggaaggg aaagacctga

8901 ccgtcccca gcccgacacc cgtaaagggt ctgtgctgag gaggattagt

8951 aaaagaggaa ggaatgcctc ttgcagttga gacaagagga aggcattctg

REC

REC

ENV

9001 ctctgcctg tccctgggca atggaatgtc tcggtataaa acccgattgt
9051 atgctccatc tactgagata gggaaaaacc gccttagggc tggaggtggg
9101 acctgcgggc agcaatactg ctttgtaaag cattgagatg tttatgtgta
9151 tgcatatcta aaagcacagc acttaatcct ttacattgtc tatgatgcaa
9201 agacctttgt tcacgtgttt gtctgctgac cctctcccca caattgtctt
9251 gtgaccctga cacatcccc tctttgagaa acaccacag atgatcaata
9301 aatactaagg gaactcagag gctggcggga tcctccatat gctgaacgct
9351 ggttccccgg gtccccttat ttctttctct atactttgtc tctgtgtctt
9401 tttcttttcc aaatctctcg tcccaccta cgagaaacac ccacaggtgt
9451 gtaggggcaa cccacccta ca

SUPPLEMENTARY FIGURE 1

This figure provides the nucleotide sequence of the *Phoenix* provirus, generated from the 9 human-specific HERV-K(HML2) copies listed in the Supplementary Table 1, together with the amino acid sequence of the encoded retroviral proteins (Gag, Pro, Pol and Rec).