

Supplementary data 3: Complete nucleotide and translated amino acid sequence of XTERV1 provirus including the host flanking sequence

Host gDNA

1 aagcaTCATAGGGATTAAAATTGCTTAATGCTGAATTTACACAATTTTGTTCATCAATCATGAAGGCCTGTGTATGTGTTACTTCAGT

89 TGTTATGATTATCCTCATTATCTTGTCTTTGTCCAAGTTGCTCTGGCCTGGGTATCTCAGCTAATTTCAAGGCTAAGAAATCATA

177 ATAAGCTTCAGTAACCATTTTAAATGAATTTCTGTTATTTGCTAATAACTGCAAGGCATATGTGCCACCCATGATACGTAAGCATAAT

265 GGGATGTGTTTACCTATTAACATAAAAGTAATAACAAAACAAAATAAAATTTGAACTTGTTCGAATACACATACTGGTTGTATCGTT

TATA box Polyadenylation signal

353 TGTACAAGTTCCTGTATCAGCGCACTTGCAGAGATATATCAGATGTTTTCGCATCATACTTGGATAAAAATCGGACCGGTCCGTGAG

441 ACTGCAGGGACAGAGTGCAGGGAGCACTTTAACGAGGATTTTCTCATCAGTTTGGTGTCAAGAAGTGGGGTCGGCGGGAGAATCCA

Primer Binding Site

529 GCGGTAAGTAGAATTTACATCTTAATTGATGGTTTTCTCTTACCTCTGGCTCTCCCTACAGGCCGATCTAAATAGTTAGCAGTTAA

617 ACAGTTAGCAGTCAGAGATACAGAGATACAGAGTGATAGAGATACAGAGATACAGAGATACAGAAATACAAAGATACAGAGATAAAA

705 GAGATATAGAGATACAGAGATATAAGAGAATTAGGGATTCTCAGGATGTGTATGGGAACATATTTACTGACTGATTCGCTCTCTGC

793 ATGTGCATTTCTGTGTTGTTCTGTCTTTGTTATATGTACCTGTAACGTATTTTATGTATCTGCTCATATAAGCACACCTGGTATTAA

881 GAGTGATCGCAAGTTGTTGTGGGTCTTGTGGGACCTTGATCTCAGTTCAGCCCTAACGATTTTGTGTTACCAAGAGTGCTTACACTC

969 AAGTGTTAACAATTTGAGGATAGACGGTATTAAGAGTTGATCGCGAGAGATTGCCAAAGATATGTTGGGACCGTGATCTCAAAACAA

↗ gag

1057 CCTAGTGTGCTTTCTCTGCTTTTACCGTGTATTTCCCAAACCAAATGGTAAGTATGGGTAACAAAACCTTCTAAAGCTTTGCCTCC

1 M G N K T S K A L P P

1145 ATCATTGGTTAATGAGACATGTAAGCAATATGTGTACCGTTGTAGTCCAGAAGGAGGATTTTTTGTGATCCATGGGTTGATAATATA

12 S L V N E T C K Q Y V Y R C S P E G G F F V D P W V D N I

1233 GCTGGTTGGGGAGAATGTTTAGAATTAGAAAATCCTTTTCCCGTGATGGTGCCAACGATCTTTATTACATGAACATGGTGTACCAGA

5'ORF1 seq 1

41 A G W G E C L E L E N P F P R D G A N D L Y Y M N M V Y Q

1321 TGTGTATTA AAAAGGAGACAGGAGACCCAGATTGGGATTACACATACATGACACGGGGTGGTGAAAATTGGCTTAAAATAGCCCCAGG

70 M C I K K E T G D P D W D Y T Y M T R G G E N W L K I A P G

1409 ATGGCATAACAACGCCCTCCCTCCCAAACTGAGAATACACATAAGTCACGAGTGTTCAGGATGTTTAAACGTAGCCCTGCAACACCA

5'ORF1 QP2

100 W H T T P S L P K T E N T H K S R V F R M F K R S P A T P

1497 ATTAACAAGTCAGATAATGCCCAATACCAAAAAGAATATCCTGAAAAAGTTTCTACCCAGCTCCTTCCTATACTGCCTTATACCCAA
 ←
 3'ORF1 QP2

129 I N K S D N A Q Y Q K E Y P E K V S T P A P S Y T A L Y P

1585 ATCTGAAACATAAACCTCACAATGAGGAAGACGACGATATCTGTATCGCCGTTGCCTCTACTCCCAGTCTCCACTACCTGCCCTACA

158 N L K H K P H N E E D D D I C I A V A S T P T A P L P A L Q

1673 AGTGCCACAACAATATGTTGACACTGCAGTCCCAGTACAAGCCCAATTATTCTCTCCTTCTTTACCATTAACTCGCCCTGTCCAA

188 V P Q Q Y V D T A V P L Q A Q L F S P S L P L T L A P V Q

1761 ATGCCTATTGCCGCACAACCTGCTGCAACAGTTTCTCTATCAGCTCACCCCTCTCTCGTCCACTCCAGTCTTCCCAAATTTACA

217 M P I A A Q P A A T V S S I S S P P L S S H S Q S S Q I Y

1849 TTGGAGAAGAAGACGAGGAGGGAGAGGATGTAATTTTCCCTTCTCCCGCCACTCAAGACAAATAGATCAAATAAGTGCACATTC

246 I G E E D E E G E D V I F P S P A T Q D K L D Q I S A T F P

1937 TTCTCCCGCTGTTCAGATAAAAATATCCCAAATAGGGTTTGAAGTGGTGTACACCCCTGAAAAAATCATACTTAAAAAGAAGCTA

276 S P A V Q D K I S Q I G F E V V V H T P E K I I L K K K L

2025 CCTGATCCTCTACTCCACCCCTCCATTTAGTCCACCCAGCTCCGATCTCGAAATAAAGCAACACTCCAAAACATAAGAGGCAG

305 P D P L P P P P P F S P P Q L R S R N K A T H S K T K E A

2113 ATAAACATTATCCTTTCCTTACTGTTGGAGATCAACTTATTATGAAACCATTAAACCCCTCTGACTTGCAAATTTACTCACCATGC

334 D K H Y P F L T V G D Q L I M K P L N P S D L Q N L L T N A

2201 TCCTGACCCAGGAAAAATCCTGGTGCCTGTTAATATACCTAAAAAGAATGTGTCAGGGCCAATGCATGCATCAAGCTGACATAAAA

364 P D P R K N P G A C L I Y L K R M C Q G Q C M H Q A D I K

2289 TTAATCGTTGATGGGATCTTGGTTATGACTTAGAATCTGGATGGGATGGACTAAAGTCCCTTCTATTTCCAACCCTGTAGCAGTTG

393 L I V D G I L G Y D L E S G W D W T K V P S I S N P V A V

2377 ATGTTATGAATACTGCCCAATATCCGTGAATACAACGGAGGGTATAAATACAATGTGGGATGAAATAGAAAAGGAAATGAAAAGGTT
 →
 5'ORF1 sequencing ←
 3'ORF1 seq 1

422 D V M N T A Q Y P L N T T E G I N T M W D E I E K E M K R L

2465 ATGGACTGATCGACAATCTATGTCGATAGCCCTTGGATGTAGGCAAAAACCTAACGAGTTACCTAGCGATTTTGTCAAGCGATTTTTT

452 W T D R Q S M S I A L G C R Q K P N E L P S D F V K R F F

2553 ACAATTTGAAAGAAGATGCTGCTTTGTCTTCTGAAGGAAATATGTCTCAACTCAAAGTGCAGACATGTTTAGGGAATATGACGCCTC
 →
 5' shift ORF1

481 T I W K E D A A L S S E G N M S Q L K V Q T C L G N M T P

2641 AGTGTAGTCTAACTACCAACAGTTAATTTCTGAGTGGCCACTCTTACTGCTGGGATTTCAAAAAAAAAAGTTATAGAAAAGAA

510 Q C S L T T K Q L I S E W P T L T A G D F K K K V I E K E

2727 GCCGCAAGTTGTTATGATAACCTTGCCAAAGGAGACAAGAAAGTTTTAGGAACATATGTAGAAAAGTACAAGTGAAGGTCAGAACATA

539 A A S C Y D N L A K G D K K V L G T Y V E S T S A R S E H

2815 ATGTTTTCAGCAACACAACCAACAGTAACAACCCCCACCGTTACCACACTCTCAACAACAACCCCGCCACCAATGACGTGTACTATGC

568 N V S A T Q P T V T T P T V T T L S T T T P A T N D V Y Y A

2903 TACTCCCAATCCTCAACCTTATCAAAATCAGGGATGGGAAAGGGTGGATACAATAGGGGGAGGTGGAACAAAGGCCGTGGTAGGGGA

598 T P N P Q P Y Q N Q G W G K G G Y N R G R W N K G R G R G

2991 CGCCCCAGAAGTTATTATAATAATTGGGGGAGTTTTCCAATTGCCCAAACACGAAACACTCAATGCTTTAACTGTTTCCAACATGGTC
627 R P R S Y Y N N W G S F P I A Q T R N T Q C F N C F Q H G
3079 ATTGGCGCTATCAATGCCTTTCCCTCCACAAAACAGCATTCCTAATAATCAAATTCATTCAGCCTACCGAATCCCAGCGCT
656 H W R Y Q C P F P P Q N S I P N N Q I Q S F Q P T A I P A L
3167 ACCATCACCACAAAACAGTTGTGACATTTCCAACTGTACAACCTGCTCCACAACAAAATAATAGTCCCGCCATGCCATCAATTGGCCA
3'ORF1 bis seq
686 P S P Q T V V T F P T V Q P A P Q Q N N S P R H A I N W P
pol
3255 GGTTCACATAGGGCTGCCCTGAAGGTTGTACCCCGCTCTTCCTGTTCATCTCCTCATCATCTGACGCTGACCCTACATAATGCTAA
715 G S T * G C P E G C T P L F P V I S S S S D A D P Y I M L
3343 AAGTCGGGGGAATAGATTGATGTTTTTAGTAGATACCGGAGCTACCACCTCCTGTATCAGTACAAAACATATGAGGGAGATAGACA
5' Pol seq 1
3 K V G G I D L M F L V D T G A T T S C I S T N S Y E G D R H
3431 TGGATCCACACCTTCAATTGGGATTAATGGGGTACTGACTCCTACTTTCCAAACACCTCCTTTGCCTATTGTGTGTCGGGATCGGAT
33 G S T P S I G I N G V L T P T F Q T P P L P I V C P G S D
3519 CGCATTATATTACATCCATTTAAACTTATCCCAGGTTGCCCTGAAAATTTACTAGGCAGAGATTATTTCCAAAACCTAGGATTTTCAGT
62 R I I L H P F K L I P G C P E N L L G R D L F H K L G F Q
3607 CTATTAGAGTTTCCCTACTGGTGGACTCACTGCTGATATTTTGTTCCTGAAACTAATTCATGTTTCTTAATTTACAAGATTGGCC
5'QP1 xeno
91 S I R V S P T G G L T A D I F V P E T N S M F L N L Q D L P
3695 TGACGATCCATCTTATCTAATGTTTGGGCCAAAACAATATGATGTGGGTTTTGTTGATTGCACACCTTACAAGCCAGGGTTAAA
3'QP1 xeno
121 D D P S L S N V W A Q N K Y D V G F V D C T P Y K A R V K
3783 CCACACTCATACCCATTATACATCAAACAATACCCACTCTCTCAGGAAAAAATAGATGGGATTCGTCCCTCATACAGACTTTCCTTG
150 P H S Y P L Y I K Q Y P L S Q E K I D G I R P L I Q T F L
3871 ACCATGACATACTTACACATGTGGTGTCTCCTTACAACACTCCATTAACCCAGTACTCAAAAAGGACAACCAGTACAGGTTAGTTCA
179 D H D I L T H V V S P Y N T P I N P V L K K D N Q Y R L V Q
3959 AGATTTGCGAGCAATTAACCTCGCTCATAATCCAATAGCACCCATTGTGCCAGACATTACAGTTCTACTTACCTCCATTCCGTGTGAT
209 D L R A I N S L I I P I A P I V P D I T V L L T S I P C D
4047 GCTACAGTATTCACGGTTTTGGACTTGAGCAATGCATACCATTCTGTTCAGTGCATACTGAGACACAGTTGCTTTTGGCTTTTACTT
238 A T V F T V L D L S N A Y H S V P V H T E T Q L L L A F T
4135 TTGAGAGATTTGAGCTGACCTGAAACGACTTGCCCAAGGTTAATCGATTCCGCCGAGCATTCACTTTAGTGTACAGCATACTCT
267 F E R F Q L T W K R L A Q G L I D S A A A F T L V L Q H T L
4223 AAAACATGGGTCCTAGTCATGGTCCGCTCCTGCTTCAGTATATTGATGACATACTGTTGTGCAGTCCAAGTCAAGGATGCTAATTAC
297 K T W V P S H G S V L L Q Y I D D I L L C S P S Q D A N Y
4311 CATGATACTGTTAGCCTATTGCGACATCTAGCTGCAGAAGGACACAAGGTTGCCAAGCATAAGACTCAGATGGGGGCTGAGTCTGTTG
326 H D T V S L L R H L A A E G H K V A K H K T Q M G A E S V

4399 AATACTTAGGATTTATCCTATCACATGGACAACGACAAATTAACCCAAAAAGGGTCCAGGCCTTGACAAAAATCAAAAAACCAGAGAC
355 E Y L G F I L S H G Q R Q I N P K R V Q A L T K I K K P E T
4487 TCCTAAGGAAATGATGTCATTTTTAGGCATAGTGAACATTTGCCGTCAATGGATACCTAACTGTTCTTACTATGACAACATCCTCAGA
5'XLERV seq 1
385 P K E M M S F L G I V N Y C R Q W I P N C S Y Y D N I L R
4575 CAATGTACCAAAAAGGATGCACCACCTTATGTACAATGGACTACTGAAATGCATGATGCTTTTTCTTCTTTGCTACAAATGTTGGTTA
414 Q C T K K D A P P Y V Q W T T E M H D A F S S L L Q M L V
4663 AAGCTCCTGCTCTTGGATTGCCTAATTATGTTAAACCTTTTCATTTGTATGTTAAGGACACCTGTAAAACCATGGCTGCGGTCTGTGC
3'Pol seq 1
443 K A P A L G L P N Y V K P F H L Y V K D T C K T M A A V C A
4751 TCAAGAGTCCGGAGGAAAATTACATCCTGTGGCCTATTTTTCCAAAGTGATGCCTGTTCCGGTACAAGGGATGCCGGCGTGTCTCAGA
473 Q E S G G K L H P V A Y F S K V M P V P V Q G M P A C L R
4839 GCCTTGGCAGTAGTGCCATGGCAGTAGAAATGTCACAGTCTGTCACAATAGGCCATGACACCATTTTACACACCTCGCATCAAGTCA
502 A L A A S A M A V E M S Q S V T I G H D T I L H T S H Q V
4927 CACACCTTTTGAAAAATATCACTACACAACACATGACTTCACAGAGACTGAGTGGATATGAAGTGATTCTTTTAGGAACCACCAATCT
531 T H L L K N I T T Q H M T S Q R L S G Y E V I L L G T T N L
5015 GACAATCCACTATATTGCTGATACACAAGGACCTGTTGCTGTCTACATGCACCTTCTCCATTTGTGAGATCCTGAAAACACTTTCTCTC
561 T I H Y I A D T Q G P V A V L H A L L H L S D P E N T F L
5103 CTGAACTCATGATTGCATTGAATCAATACTACACAACAGCACCTAGATCAGATCTGACTGACATACCCCTTCAGGATGCTCCAA
5' ORF2 sequencing
590 L E P H D C I E S I H Y T T A P R S D L T D I P L Q D A P
5191 GTGTATTGTTGATGGTTCCTGTTCCAGGCCGTCTGATGTTACATATAAAGCTGCCTATGCTGTAGTCCAACACTCTGACATGATTCT
619 S V F V D G S C S R P S D V T Y K A A Y A V V Q L P D M I L
5279 TGAAACTGCAGTAGTCCCTGTTAACTCTGCACAAGCTGCTGAATTAATAGCCCTTACTCGAGCTTGCCAACACTATTTTCTGGCCGTCCA
649 E T A V V P V N S A Q A A E L I A L T R A C Q L F S G R P
5367 GTTAATATTTTTAGAGATAGCAGATACGCTTTTGGTGTGGTTCAATTTGGTAAGATTTGGCAACAGAGAGGCTACCTTACCCTGATG
679 V N I F R D S R Y A F G V V H F G K I W Q Q R G Y L T T D
5455 GTAAATCCATTGCTCATACTGCCCTTATTGAAAAATTATTACAAGCCATCCAACCTCCATCTGCAATTTCTATCATACTTGGCCGTGC
707 G K S I A H T A L I E K L L Q A I Q L P S A I S I I H C R A
5543 ACACACTAACAGAACTGATGAGATTTCTCTTGGCAATGCGTTTGCAGACCAAGTTGCTAAAAACACTGCATCCTCAGCCACTCCTACA
737 H T N R T D E I S L G N A F A D Q V A K N T A S S A T P T
5631 GCTGTCCCATATTTTTATTTACACCACCACAATGCCCTGATAGCCAGATCCTACAACACTTACAAAAATTTGCCACTTCTACAGATT
766 A V P I F L F T P P Q C P D S Q I L Q H L Q K F A T S T D
5719 TATTGTTCTGGCAACAACAAGGCTAATTCTTGATCAATTTGGCCTTTATTCCTTGAAGGTAGAGTTGGCATCCTGAAAATAGCCT
795 L L F W Q Q Q G L I L D Q F G L Y S L E G R V G I P E N S L
5807 ACCTTTATTGATTTCTCAAGCCATGGTGTGGGTCATAAAAGTATAAAATCTACTTTAGCTGAACTACACTCACACTTTGTTGCCAAT
825 P L L I S Q A H G V G H K S I K S T L A E L H S H F V A N

5895 AATCTCAAACGTCACACTGTTCTGACTTTGTCAAACCTGCCTCACCTGCATTGCTGCTAATACCGGGGTACACACAACACCCATCAGC
854 N L K R H C S D F V K T C L T C I R A N T G G T H N T H Q
5983 ATTTACCCCAACCAATTGCCCAATTTTACACTGGCAAGTAGATTTTACTCACATACCCAAAACAGGTAACCTCCAAGAGTATCTCTT
883 H L P P P I A P F S H W Q V D F T H I P K T G K L Q E Y L L
6071 AGTCTTTGTTGACCATTTTTCACGGTGGGTAGAACTTTTCCTTGCCGCATGAAGATGCTAAAACAGTTGTTAAATTTCTTGCCAT
913 V F V D H F S R W V E A F P C R H E D A K T V V K F L V H
6159 GACATTATCCCCAGATATGGCTGCCCTCAACAAATTAGTTCTGATAATGGTCTGCCTTTATTAGCAAAGTGCCTCAATTGCTCATT
942 D I I P R Y G C P Q Q I S S D N G P A F I S K V L Q L L I
6247 CTACGCTAAAAATCACTTGAAATTTAGCATCCCTTATCATCCACAGAGTAATGCTCTTGCTGAAAGATCAAATCGCACAAATTAAGA
971 S T L K I T W K F S I P Y H P Q S N A L A E R S N R T I K D
6335 TAAATTAATTAAGCTACCTCTGCAACATGGACAAATTGGAAGTCATTCCTTCTGCAGTCCTGCAGAAATTCGCATGACACCACAT
1001 K L I K A T S A T W T N W K S F L P A V L A E I R M T P H
6423 TCAGTTACCAAACCTCTCCATTGAAATTGTCATGGGTGCGCCTTTTCCAACCTCCATGGGTAAAGGTCCTTTAGTGGTTATGCCTC
1030 S V T K L S P F E I V M G R P F P T P W V K G P L V V M P
6511 AAGATTTAAATCTCATAACAAGAAGAATATGTA CGTAATCTTGTCTCTGCATTGAATGGCATCTATGGTGATGTTTCTTTGTCAATTC

3'XLERV seq 1

1059 Q D L N L I Q E E Y V R N L V S A L N G I Y G D V S L S F P
6599 TTTACCTTACAGGAACCTACCCATATTTTCCGCCCTGGAGACTCTGTTGTTGTTTCGACAGCTGCATCGCTATCGAAAAGGGGATTT
1089 L P S Q E P T H I F R P G D S V V V R Q L H R Y R K G G F
6687 CCGTTTGGACCCAGAACTACAGTGATTGCCATTACACGTACCTCCATTCTCACGGATTCCAGTGACCAGTGGATTACACGCTCCCGAG
1118 P F G P R T T V I A I T R T S I L T D S S D Q W I H A S R

↪FR47/env

6775 TGAAACTTGCCCCCGATTGCTACCACCACAAGTGCCTTCCAATGTTGACAAGACTACAAAAGATGACCTCTTCTCACCTGATAACAT
1147 V K L A P R L L P P Q V P S N V D K T T K D D L F S P D N I
1 M L T R L Q K M T S S H L I T
6863 CGCCTCAGGACGTCAGGACGGACAGAATGCCAGGATGACCAGGAGGATGACCAGAGAAAACTCAAGAACCAACAACCTACCTGTAAT
1177 A S G R Q D G Q N A Q D D Q E D D Q R K T Q E P T T T C N
16 S P Q D V R T D R M P R M T R R M T R E K L K N Q Q L P V M
6951 GAACAGGATATCGACAAAATATGCACTCTGGTGCAGCCTTCTTACTTTTCTTTGCTGTTTATCCCATGTGGATATTACTTAACTGT
1206 E Q D I D K I C T L V Q P S Y F P F A V Y P I V D I T *1232aa
46 N R I S T K Y A L W C S L L T F L L L F I P L W I L L N C
7039 GTATGTTATCCATCTCAGAATAGCCCTCTTGGCTCTCTTTGCTATAATGCTATGCGCAACAATAGCAATAATACTGAACACGTACGTA

5'FR47 seq 2

75 V C Y P S Q N S P L G S L C Y N A M R N N S N N T E H V R
7127 AGCGCTCTTTATCCACAATGTTAAGCTCTACATTTCATGGTTCATGATGGTACTGACAATTTTCCCTGCTAACATAATGCAGTGGG
104 K R S L S T M L S S T F M V H D G T D N F S P A N I M Q W G

7215 **TTGGCGCAATGCTACAATAGCACGTCAACAAGATAATTTACATGCCCTGCTAATCAACCATGCATTGTGTGAACATGACCACATCA**
134 W R N A T I A R Q Q D N F T C P A N Q P C I V L N M T T S
7303 **TCATCAATTCTTAAAAGTCATAATCATATAGGGTACCTTGCTATCCCTGTTCAACGTTACTGTAATTTTTCCATAATCTCCAATGCAA**
163 S S I L K S H N H I G Y L A I P V Q R Y C N F S I I S N A
7391 **CTGCATTCAATCTACTCTCCTCATGCCTTCTCAAGAGTATGATAAGCCTTTCTTGGTGAACGTTTATCATGATCCAAATAGTTTAA**
192 T A F Q S T L L M P S Q E Y D K P F L V N V Y H D P N S F N
7479 **TTGGACACAAAACACTAATTATCTCTCCTGCAGGATATACTCTGCCTTTTCACTAATGGTACTCCAATTACCAATAGAACATGGTTC**
5'FR47 seq 1
222 W T Q N T N Y S L P A G Y T L L F T N G T P I T N R T W F
7567 **ATGATAGGTATTGTATGTATCATTTCCTTCCACTTCACCTACAGCTCATAATTATTCTCTTATACCAACTCCAGACACATGTGTTAATGTTA**
251 M I G I V C I I S T S P T A H N Y S L I P T P D T C V N V
7655 **CTACTATAAAACAAGGCTGTCAAAGTTATTTTTAAAATCACCACACCTCCTATCCCAGTTGTCTACTAGTAAGCTAGACACAAGCC**
280 T T I N K A V K V I F K I T T P P I P V C S T S K S **R H K R**
7743 **TGCGTGGTATGATACTTTACTAGGAGGTGCTGGCACAGCCATGGGTGTACTCAATGGAGTTGATATGGAAGTCTAAGAAATAAGTTA**
310 A W Y D T L L G G A G T A M G V L N G V D M E V L R N K L
7831 **TCCACTGCAGCATCTCACTTAAAAACAGCTTTTCATGTAAGTGCACAATGGGCTCCTACAGTGGTACAATCTCAACTCATGCAATTAA**
339 S T A A S H L K T A F H V T A Q W A P T V V Q S Q L M Q L
7919 **AATTAGACCAAGAGTTTGTAATAATTTAGGAGACAATACCTTTGCTGTCCTGAACCTCACTGCTGGTTCTGTTAGTTGGACCTTATG**
368 K L D Q E F V K Y L G D N T F A V L N L T A G S V S W T L C
8007 **TGGACTTAGATATGTGTGGTTATCTCAACAACGTTTCAGAATTAGAAAACATGCTACTACTGCCCAAATACTCAAACCTTGCTAAACTA**
5'QPCR FR47 3'QPCR FR47
398 G L R Y V W L S Q Q R S E L E N M L L L P K Y S K L A K L
8095 **TTTAGTTGGAATTCCTCGGATTGGGTACATCTAGATAAAATTAACACTTTCTGCCAAGAAATTCATAATGGTTCCATATCCTTCTTTA**
427 F S W N S R D W V H L D K I N T F C Q E I H N G S I S F F
8183 **CTTGTTACCTATCTATTACATTGTATAAGGCAGATAATGTATCATTAGCTTGTAATTCACTGTATTACCTTTTCCACATTATAATGG**
456 T C Y L S I T L Y K A D N V S L A C K F T V L P F P H Y N G
8271 **CCTTACATCTAACCTTACCTGGTGGGTCCAACCTTTTCTAGTAATTATGTTTTTCTAAGACCAATCGTACTTTCACACTAGAACAT**
486 L T S N L T W W V P T F L G N Y V F P K T N R T F T L E H
8359 **TGCATTTCCACTAATAATGGTTAGCTTGTGCGCCATTCTAGCCCTGTTTATGAACCGTGCTTGCTACATAAAGATTTAAATATTTGCC**
515 C I S T N N G L A C R H S S P V Y E P C L L H K D L N I C
8447 **ATGGGGATGTATATCCAGCAGATTACTCTATCATGCTCGAGATTGGTCCACAATATGTATGTTTGGCAACTAATTTTCCAGCTGGCAT**
544 H W D V Y P A D Y S I M L E I G P Q Y V C F A T N F P A G I
8535 **TCATGCTGTAGATTCTTCACTACCTTTTCACTGCTTGCCTCACTAACGTCACCCTTTTACATTGGTATAATTATACCTATTATTTTCT**
574 H A V D S S L P F S G C L T N V T L L H W Y N Y T Y Y F S
8623 **CCTATCCCCTGTCCATCGTTCTTTGTGATTTCCACCTACCTTACCTGTAGGTGAATATCAGATTCCCAATGCTGTATATCAAT**
603 P I P T V H R S F V Y F P P T L P V G E Y Q I P N A V Y Q

8711 CACTTATTATCCATAACACTAACATGATTGCTGCTTATCGCAAGCATGATACCCAATTGCATGACACTGCTATTAAGACTGTTGCTAC
←
3'FR47 seq 1

632 S L I I H N T N M I A A Y R K H D T Q L H D T A I K T V A T

8799 CAAGAACAATTGCTCTCTCTGGGATGAACTGCAACACCATTGGTATGATGTCTTCACTGGCTGGTCTCCTACAGGTTCTGCTATTAT

662 K N K L L S L ? D E L Q H H W Y D V F T G W S P T G S A I M

8887 GCAAAGTTTTTTTTATCCACTTCTTGCTGTTATGCTGTGCTTGTCTTACTTGATTATTCAATTTATATCTCATGTACTTTTCGT

691 Q S F F Y P L L A V I A V L V L L V L F N L Y L S C T F R

8975 AAACGTATACAAAAATGATTTACATGCTCAGTCTGAACAACCTAAGCTTATACATGCTTCACATATG TAGTCACTTTTGT GACTAA

720 K R I Q K M I S H A Q S E Q L K L I H A S H M * 742aa PPT

9063 AAGGGGATTCATAGGGATTTAAAATTGCTTAATGCTGAATTTACACAATTTTGTCTTCAATTATGAAGCCTGTGTATGTGTTACTT
→
5' AseI LTR3

9151 CAGTTGTTATGATTATCCTCATTATCTTGTCTTTGTCCAAGTTGCTCTTGGCCTTGGGTATCTCAGCTAATTTTCAAGGCTAAGAAAT

9239 CATAATAAGCTTCAGTAACCATTTTAATGAATTTCTGTTATTGCTAATAACTGCAAGGCATATGTGCCACCCATGATACGTAAGCA

9327 TAATGGGATGTGTTTACCTATTAAC TATAAAAAGTAATAACAAAACAAAATAAATTTGAACTTGTTCGAATACACATACTGGTTGTAT

9415 CGTTTGTACAAGTTCCTGATCAGCGCACTTG CAGAGATATATCAGATGTTTTCGATCATACTGGATAAAATCGGACCGGTCCG
→
5' LTR interne

9503 TGAGACTGCAGGGACAGAGTGCAGGGAGCACTTAAACGAGGATTTTCCTCATCA agcaAGCTTTGTATCACCTCCAAATCTTATGAA
←

Host gDNA 3' Locus Xterv

9591 TTTGCCATATAGGTGATATC