Predicted TgERV-F provirus consensus sequence

5'LTR

TGTGGTAGATAGGGACAGGCGAGCGGAAGATCACGGGATGTCACGGAAAGATAGACCCCTCCCCCCCTCTCTCTCCCCACGTTATCTATTAACCCCAGAA < 100

GCCCCAGAAGAATGTAGCCACACCTGCCCTAGTAAATTTCCACTACCCGACTAACCCCTGAGACCCCCCAACCCCCCTCTGACGTAGCAAAGACCCCCAA < 200

poly-A signal

AACTATTTAAACCCACGAGATAGGATAATAAACGCTTTTCAACCGTCTGCCATATTGGTGTCTGCGTGTGTTGATTAGCCCGAGTGGCCTGGACGAGACC < 300

End of 5'LTR **PBS**

AGGCCGCCGTGCTGCCACCTGAACCAGGTCCCTGGTTGTCTTTCATAAAGGCAACATACTGGTGCCGAAACCCGGGACAAAAGAAGAAAAAAAAAGGAAA < 400

AATTCGCCCGGCTGACGGGATACACCTGGACAGAAGCAGCGGCCGGAACGGTCAGACCGTATCTAGGCGTAGGGGAGACGTCCCGGGATCCGCGGGCGTC < 500

Gag SD

M domain

ATGGACGCAATGGCAAGGATCGTAAGTGCTATTTATTCGCAGTGGGGTATTGGGTGTAAGCTCAAGGATTTTCATCTTGCCATAGCGAGATTGCTTGAGC < 600
M D A M A R I V S A I Y S Q W G I G C K L K D F H L A I A R L L E L

TAGGGGCAATCGAGCGTCCCGTGGATGTATTACATCCGGAAATATGGGAAAAATGCACTGCCGCGCTAGCCGAGGACACAAAATCCTCAGGCAGCGGCAA < 700
G A I E R P V D V L H P E I W E K C T A A L A E D T K S S G S G K

AAATCTTAAAGCATGGGGCAAAGTAGAGAAAGCCCTACGCAAAGCGATAGAAGAGCAGGAGACATGGACCGCGGCGCGCACGTGTTTATTAGTTACTCCC < 800
N L K A W G K V E K A L R K A I E E Q E T W T A A R T C L L V T P

AAACTCGGGGTGGGGGCGGGGACGCAGACTGCCCCTGAGAGCGATCTGCCCGGCAGCGGGGACCCGGGGGGGCTCAGCGCGACACCCGCTCCCCCAGATC < 900
K L G V G A G T Q T A P E S D L P G S G D P G G L S A T P A P P D Q

AGAGCCCACCCCCCCCCGCAGGAACCCCCCGCGACACCCCCGCCGCCGAAAATCCCCAAAAACCCCCCACCGCTCAAACCCCCCGAGAGACCGCAAATTC < 1000
S P P P P A G T P R D T P A A E N P Q K P P T A Q T P R E T A N S

CTCCGAGCCGCCGCCGCCCGCCCGCGATCCGATGTCGGAGACGCAACAGCACGCGGAGCGCTTCTGGCAAGGGCTGCTAAAAGAAGCCCGAGCCGCC < 1100
S E P P P P P A R D P M S E T Q Q H A E R F W Q G L L K E A R A A

L domain

GAAACCGCGGCTTGGGAAGACGGCGCGGCCATGCCGCCTCCCTACCCCTTTGAAAATGGCGCCGGGAGCGGCAGAGAGGGGCGGGGCTCGGGCGGTCCCG < 1200
E T A A W E D G A A M P P P Y P F E N G A G S G R E G R G S G G P G

GCGCGAACACCCGGGAGGAGCGCGATCTCGGGAATGCGCGCGCGCGGGAGAAAGGGGCGGAGAGCGGCACAGAGCGCGGAACCAACCAGCAGCCCCGCCG < 1300
A N T R E E R D L G N A R A R E K G A E S G T E R G T N Q Q P R R

GAAGCTGCCACCATATAAGGGAGAAACCACCCCCCGCGAGAGGCGCGGCGAGCCGGGGGGGCGGGAGCGGCGCCACCCCCACGGGGAGGAGCGAGCTCGG < 1400
K L P P Y K G E T T P R E R R G E P G G R E R R H P H G E E R A R

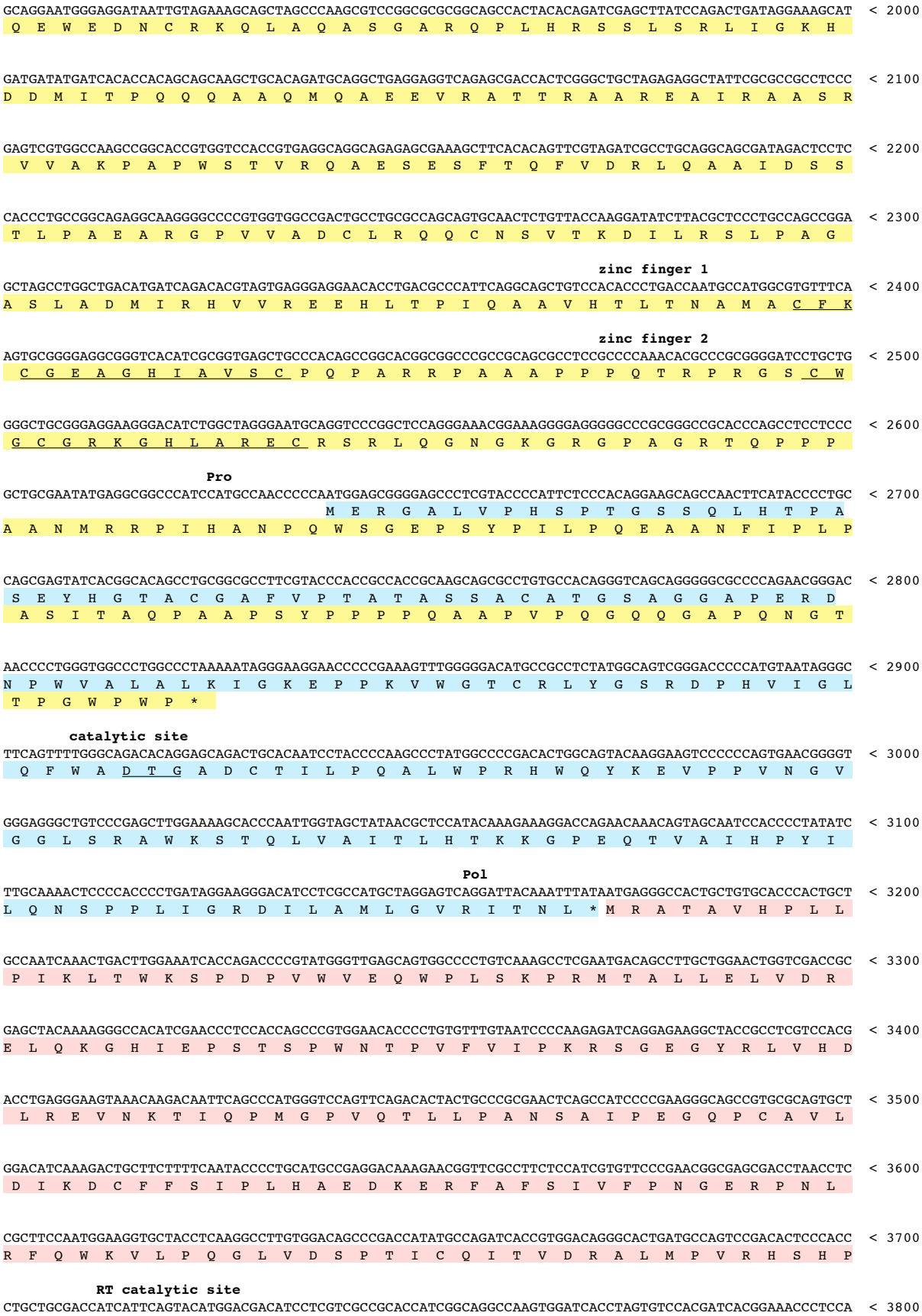
GGCCGAACAAAACGGCACGGAGCCCCGGAAGTGCACTGGCACTCGTCTTCCGGCTCGGAGACCAGCAACAGCTCCACCAGCTCAGAGGAGCTGACGGAGG < 1500
G R T K R H G A P E V H W H S S S G S E T S N S S T S S E E L T E A

CTAGCGATGGCTCCGAGACGGACGAAACGGAACCGGCACAGCTTGAAACAAAGCCGAGTAAAGCTCTAAGCCGCGCCGAAAAACAACTACAATACGAACC < 1600
S D G S E T D E T E P A Q L E T K P S K A L S R A E K Q L Q Y E P

AGCCCAGTTTACTGACTGGGGGAAAATAAAGATAGCTTGTGCTGAATGGTCCCCAGCGGCCGCCATACAAGCCTTCCCGGTACGGCTTACCGGTCCGGAA < 1700
A Q F T D W G K I K I A C A E W S P A A A A I Q A F P V R L T G P E

GGGAACCAACAAAGGGTATATACCCCGATAAACCCAAAGGATATACAGTCAATTGTCAAAGCCATTGCAGAAAAAGGGATCAACTCGGCTATAGTCACTA < 1800
G N Q Q R V Y T P I N P K D I Q S I V K A I A E K G I N S A I V T T

CGTTAATCGATGGTCTTTTTAGCAACGATGACTTGCTTCCCTTTGATATCGAACGGATAGGGCGTATGATACTTGACGGGGCGGGAATGATCGTCTTTAG < 1900
L I D G L F S N D D L L P F D I E R I G R M I L D G A G M I V F R


zinc finger 1
zinc finger 2
Pro
catalytic site
Pol
RT catalytic site

```
  A  A  T  I  I  Q  Y  M  D  D  I  L  V  A  A  P  S  A  G  Q  V  D  H  L  V  S  T  I  T  E  T  L  Q

GGCCAACGGCTTCGAGATCGCGAACACGAAGATCAAGAGAGGACCGTGCGTGACCTTCCTGGGAGTGGGGATCACAAACTCCTACGTGACCCCACCCAAG  < 3900
  A  N  G  F  E  I  A  N  T  K  I  K  R  G  P  C  V  T  F  L  G  V  G  I  T  N  S  Y  V  T  P  P  K

ATAAAAGTCCGCCGAGACATCAAGACCCTCCACGACATGCAACGACTCGTAGGATCTCTGCAGTGGCTCCGCAACATCATCCTAGTTCCGCCAGAGGTCA  < 4000
I  K  V  R  R  D  I  K  T  L  H  D  M  Q  R  L  V  G  S  L  Q  W  L  R  N  I  I  L  V  P  P  E  V  M

TGGACCCCCTGTATGACCTCCTGAAAGGAAAACACCCGTGGGACCCCAAGGAGCTGACGCCGCAAGCAACGAAATCCCTCGACTTCATCGAACGTCAGAT  < 4100
  D  P  L  Y  D  L  L  K  G  K  H  P  W  D  P  K  E  L  T  P  Q  A  T  K  S  L  D  F  I  E  R  Q  M

GTCCACCAGCCTGCTTGCCAGGTGGAACCCGGGCGTACCACTGGACTTATACGTCCACTTCACGCAGAAGGGAGGAGTGGGAGCACTTGCCCAAGGACCT  < 4200
  S  T  S  L  L  A  R  W  N  P  G  V  P  L  D  L  Y  V  H  F  T  Q  K  G  G  V  G  A  L  A  Q  G  P

TCCGAAAAATCCCAGCCGATCCAGTGGGTGGTCCTCGGAAGACCGACTCACGCATTCTCCCCAGGAGTTGAGTGCATTGCTAACCTCATCACGAAAGGCA  < 4300
S  E  K  S  Q  P  I  Q  W  V  V  L  G  R  P  T  H  A  F  S  P  G  V  E  C  I  A  N  L  I  T  K  G  R

GGAGACTCGCCCTGAGACACCTGGGAACCGAGCCGGCAAGGATCCACCTCCCCTTCCGCAAGCGACCGACCACGGAGTCAACTGCAATATCGGAGCACCT  < 4400
  R  L  A  L  R  H  L  G  T  E  P  A  R  I  H  L  P  F  R  K  R  P  T  T  E  S  T  A  I  S  E  H  L

GGCCCTCGCTCTCACCGGCTTCGGAGGAGAAATCTCCTACGCCACCAAACCACCCTGGACCCAGCTACTGACCATTGTCGACATAGATGTGCCACCGAAG  < 4500
  A  L  A  L  T  G  F  G  G  E  I  S  Y  A  T  K  P  P  W  T  Q  L  L  T  I  V  D  I  D  V  P  P  K

GTCATGGACCGACCGCAACCAGGACCAACGGTCTTTACAGACGCCTCCTCCATGACTTCTACCGCAGCAGCAGTGTGGCAGGCAGGAGAAACATGGCATT  < 4600
V  M  D  R  P  Q  P  G  P  T  V  F  T  D  A  S  S  M  T  S  T  A  A  A  V  W  Q  A  G  E  T  W  H  C

GCGTCAAAACGTGTGACCCCACGCTGTCAGTGCAACAGCTGGAAGCAGCAGCAGTGGTCTTGGCATGCGGACTTTTCCAGGACGAACACCTCAACATCGT  < 4700
  V  K  T  C  D  P  T  L  S  V  Q  Q  L  E  A  A  A  V  V  L  A  C  G  L  F  Q  D  E  H  L  N  I  V

GACAGACTCTATATTCGTGGCAAAGCTCTGCCTAGCCATGTCAAGACCAGGTGTGTCAACATCCACGACGGCCTCCATGCTTGAAGAGGCACTCTCCTCA  < 4800
  T  D  S  I  F  V  A  K  L  C  L  A  M  S  R  P  G  V  S  T  S  T  T  A  S  M  L  E  E  A  L  S  S

CGCCAGGGCACCGTGTCCGTCATCCACGTCAACAGCCATAACCCAGTCAAGGGCTTCTTCCAGACTGGCAACGACAAAGCAGATGCCGCAGCGAAGGGAG  < 4900
R  Q  G  T  V  S  V  I  H  V  N  S  H  N  P  V  K  G  F  F  Q  T  G  N  D  K  A  D  A  A  A  K  G  V

TGTGGACACTGCAGGAAGCTCGTCAGCTGCACGAGTCACTCCACATAGGGGCCAAAGCACTGGCAAAAAGATGCGGGATCTCGACAGCAGACGCGAGACA  < 5000
  W  T  L  Q  E  A  R  Q  L  H  E  S  L  H  I  G  A  K  A  L  A  K  R  C  G  I  S  T  A  D  A  R  H

CGTAGTGGCCACCTGCCCTCACTGCCAGAAGTCACCCCTATGGACCGGTGGAGTCAACCCGAGGGGCCTCAAGGCGTCAGAAATCTGGCAGTCAGACTTC  < 5100
  V  V  A  T  C  P  H  C  Q  K  S  P  L  W  T  G  G  V  N  P  R  G  L  K  A  S  E  I  W  Q  S  D  F

ACCCTCTGCGAACTGCTGAAGCCCCGAGCATGGCTTGCAGTGACAGTGGACACCTACAGCGGAGTGATCATAGCGACACAGCATCTCAAACCCAACTCGA  < 5200
T  L  C  E  L  L  K  P  R  A  W  L  A  V  T  V  D  T  Y  S  G  V  I  I  A  T  Q  H  L  K  P  N  S  K

AGGCCACGATCCAGCACTGGCTGACAGTTATGGCATGGCTTGGTATCCCCAAGCAAATTAAAACTGACAATGCTTCCAATTTTATCTCCAAATCAGTGCG  < 5300
  A  T  I  Q  H  W  L  T  V  M  A  W  L  G  I  P  K  Q  I  K  T  D  N  A  S  N  F  I  S  K  S  V  R

AGAATTCGCCTCAGTGTGGGGTATCACCTTAGCACAGGGAATCCCATATAACAGCACCGGACAGGCCATTGTCGAGCGAGCAAATCAGACCCTAAAAGCC  < 5400
  E  F  A  S  V  W  G  I  T  L  A  Q  G  I  P  Y  N  S  T  G  Q  A  I  V  E  R  A  N  Q  T  L  K  A

AAGTTAGAAGTGTTGGCAAAGGCAGAGGGCTTTGCCAATTCCATCCCCTCAGGAGACCAGACACGCATGCTAGCAACCGCGCTGCTAGCACTGAACCAAT  < 5500
K  L  E  V  L  A  K  A  E  G  F  A  N  S  I  P  S  G  D  Q  T  R  M  L  A  T  A  L  L  A  L  N  Q  F

TCCCTAGGGGAGATGAAGCAAACAGTCCCATTCGAAAACACTGGGCCACCCAGACACTAGAGGAGGGCCCACAGGTCATGGTCAAAAACGAGCTAGGCGA  < 5600
  P  R  G  D  E  A  N  S  P  I  R  K  H  W  A  T  Q  T  L  E  E  G  P  Q  V  M  V  K  N  E  L  G  E

GTGGGAACGGGGCTGGAGACTGGTGCTTACGGGACGGGGGTACGCGGCAGTTAAAAAGGAGGGCAAGATCAGGTGGTGTCCACTCAGGTCGATCAAACCT  < 5700
```

```
W  E  R  G  W  R  L  V  L  T  G  R  G  Y  A  A  V  K  K  E  G  K  I  R  W  C  P  L  R  S  I  K  P

                    Env -- SU    SA
GACCTTAAGAATGAAACTAATGGAAAACTGTGAGTTTTCGTTCGCAGGACACGCTCGTGGACCGTCCCCGCGACACACACACCCCTGCTCCAGAGAGAGA  < 5800
D  L  K  N  E  T  N  G  K  L  *        R  S  Q  D  T  L  V  D  R  P  R  D  T  H  T  P  A  P  E  R  D

TAACGGACCACCAAGCCACCAGACAGAACCTGAGCGTCCCACGGCGGTGAGACCCAGACATGCACAGTGTCTCTGTGCAATCCTGCTGCTGGGGTTTGTG  < 5900
 N  G  P  P  S  H  Q  T  E  P  E  R  P  T  A  V  R  P  R  H  A  Q  C  L  C  A  I  L  L  L  G  F  V

GCCGGGGGGCAAGCCGACCCAGGCCACTACCCTCACCAGCCGTTCAGATGGGTCATGCAACATCTTTCAAGTGACAAGGTGTTCAAAGAGGTCACCACCG  < 6000
A  G  G  Q  A  D  P  G  H  Y  P  H  Q  P  F  R  W  V  M  Q  H  L  S  S  D  K  V  F  K  E  V  T  T  A

CAAACGCTCCATCCTTCGTATTCCACATAGCCGATCTGTTTCCAGGGCAACCGAAAATACGGCCCTCAAGCCCACACATCATACTCATGTACATATCTTA  < 6100
  N  A  P  S  F  V  F  H  I  A  D  L  F  P  G  Q  P  K  I  R  P  S  S  P  H  I  I  L  M  Y  I  S  Y

CTGGTGCCCAGCGTCCAACCCAGGGAAAAGGTACTGTGACTACCCGGGGTGGGGACATTGTGGATATTGGGGCTGCGAAACCATTGTTACAGATGCCAGA  < 6200
 W  C  P  A  S  N  P  G  K  R  Y  C  D  Y  P  G  W  G  H  C  G  Y  W  G  C  E  T  I  V  T  D  A  R

CCATGGGGAGACGGGTGGCAACCGCAGGAGCCCGACAAATTCTTACAGTTCACCTGGGCACCCTTTGGCTGCGGAGACCACAATGCACAGCCTAGACAAC  < 6300
P  W  G  D  G  W  Q  P  Q  E  P  D  K  F  L  Q  F  T  W  A  P  F  G  C  G  D  H  N  A  Q  P  R  Q  R

GGGGATGCGTAAGTTATAACATGACTGTCCTACAGCCAGATCACCCTAGCTGGGCCACGGGTAGAACATGGACAGTGGTCCTCAGAGGACCGAGGAGGTG  < 6400
  G  C  V  S  Y  N  M  T  V  L  Q  P  D  H  P  S  W  A  T  G  R  T  W  T  V  V  L  R  G  P  R  R  W

GGTGAATGTGAGAATTATCAGGCTCCAGCCGCCAACACCTCGACCAGTGGGACCCAACAAAATTATCAAGAATGTGCTGAGAGGGAAAAACACAACCCAC  < 6500
 V  N  V  R  I  I  R  L  Q  P  P  T  P  R  P  V  G  P  N  K  I  I  K  N  V  L  R  G  K  N  T  T  H

CCCAAAACCCTGACCCCAAAAGCCACTGACACCCCGACCAGCCTTGCAGATACCCCCCAGATAGGCCGCACGGCTGAGTCAGACCCAAACCCAATCTTTC  < 6600
P  K  T  L  T  P  K  A  T  D  T  P  T  S  L  A  D  T  P  Q  I  G  R  T  A  E  S  D  P  N  P  I  F  R

                                    isomerization motif
GTATGCTAGAAGCTACCTTCCTAACCCTAAACGAAACCAAACCAAACCTAACTAACTCCTGTTGGCTTTGCTATGATATCAAACCCCCTTTTTATGAAGG  < 6700
  M  L  E  A  T  F  L  T  L  N  E  T  K  P  N  L  T  N  S  C  W  L  C  Y  D  I  K  P  P  F  Y  E  G

CATTGCTTTGGACACCCCCTTCAGTTACTCCACAGCCAGCGCCCCCCACCAGTGCAGATGGGACACTCCCCGCAGAGGAATCACCCTGAGTCAAATCACA  < 6800
 I  A  L  D  T  P  F  S  Y  S  T  A  S  A  P  H  Q  C  R  W  D  T  P  R  R  G  I  T  L  S  Q  I  T

GGACAGGGCAGATGTTTTGGGAATGCAACCTTAGCAAAGCAGAAAGGCAACTTCTGCACTAAAGTTGTCAAACCCAACAGAAAAACTAACAAGTGGGTGA  < 6900
G  Q  G  R  C  F  G  N  A  T  L  A  K  Q  K  G  N  F  C  T  K  V  V  K  P  N  R  K  T  N  K  W  V  I

TCCCATCTGCGTCTGGGATGTGGGTTTGCCAAAGGTCCGGAGTGAGTCCTTGTGTGTTCCTTGCCAAATTCAATGACTCTACCGATTTCTGTGTCCAAGT  < 7000
  P  S  A  S  G  M  W  V  C  Q  R  S  G  V  S  P  C  V  F  L  A  K  F  N  D  S  T  D  F  C  V  Q  V

                                                cleavage site    Env -- TM
TCTGATTGTCCCCAGGGTCTTGTACCACTCAGACGAAGAGATATACCATCTTCTCGAAGAACCTAACAGACTCCACAAAAGGGAAATCATCACAGGTATA  < 7100
 L  I  V  P  R  V  L  Y  H  S  D  E  E  I  Y  H  L  L  E  E  P  N  R  L  H  K  R  E  I  I  T  G  I

ACTATCGCAATGCTGCTCGGCCTGGGAGCAGCTGGCACGGCCACGGGTGTCTCAGCCATCGCAACCCAACAGCACGGACTCTCTCAGCTGCAAATGACCA  < 7200
T  I  A  M  L  L  G  L  G  A  A  G  T  A  T  G  V  S  A  I  A  T  Q  Q  H  G  L  S  Q  L  Q  M  T  I

                                                                immunosuppressive domain
TCGACGAGGACCTGCAGAGGATCGAGAAATCCATCTCCTATCTAGAGAAATCAGTCTCTTCGCTTTCAGAAGTAGTTTTACAAAATAGGCGAGGACTGGA  < 7300
 D  E  D  L  Q  R  I  E  K  S  I  S  Y  L  E  K  S  V  S  S  L  S  E  V  V  L  Q  N  R  R  G  L  D

                    CX6CC motif
CCTTTTGTTCATGCAGCAGGGAGGACTGTGTGCAGCCTTGAAGGAGGAATGCTGCTTTTATGCAGATCATACGGGAGTCGTTAAAGACTCCATGGCAGAA  < 7400
 L  L  F  M  Q  Q  G  G  L  C  A  A  L  K  E  E  C  C  F  Y  A  D  H  T  G  V  V  K  D  S  M  A  E

CTCCGAGACAGACTGGCTCAGAGAAAGAGAGACAGGGAAACCCAACAGAGCTGGTTTGAATCCTGGTTCAATCAATCACCTTGGCTCACCACTTTAATTT  < 7500
L  R  D  R  L  A  Q  R  K  R  D  R  E  T  Q  Q  S  W  F  E  S  W  F  N  Q  S  P  W  L  T  T  L  I  S

CCGCCCTGGTAGGTCCACTGGCAATACTGCTTTTAGCTATTACCATAGGACCATGCCTGCTGAACAAACTAGTCTCGTTTGTTCAGGCCCGTCTGGAAAG  < 7600
```

A L V G P L A I L L L A I T I G P C L L N K L V S F V Q A R L E R

GGCAAACATTCTGTTCATAGGCCACCAACAAATGCTATAAACCAAAAACTGCAAACACAGTCAGTAGCTAAAGCCTTCAGCACCTGCCTTGAAAAATCTA < 7700
A N I L F I G H Q Q M L *

polypurine tract

CCCAGATCTACCAAACCACCCTTTCCTTAACAAGTTACAAGTTTGTACCTCACTCCAGTGCCTATACCTACGACTACCTCATTTTGTATGTGATAAGGAG < 7800

3'LTR

GGGGGAGATGTGGTAGATAGGGACAGGCGAGCGGAAGATCACGGGATGTCACGGAAAGATAGACCCCTCCCCCCCTCTCTCTCCCCACGTTATCTATTAA < 7900

CCCCAGAAGCCCCAGAAGAATGTAGCCACACCTGCCCTAGTAAATTTCCACTACCCGACTAACCCCTGAGACCCCCCAACCCCCCTCTGACGTAGCAAAG < 8000

ACCCCCAAAACTATTTAAACCCACGAGATAGGATAATAAACGCTTTTCAACCGTCTGCCATATTGGTGTCTGCGTGTGTTGATTAGCCCGAGTGGCCTGG < 8100

ACGAGACCAGGCCGCCGTGCTGCCACCTGAACCAGGTCCCTGGTTGTCTTTCATAAAGGCAACA < 8164

Table S1. Other taxa in which *pol-env* sequences highly similar to TgERV-F were found

Species	**% similarity (amino acids)**		**accession number**
	Pol	Env	
Geospiza fortis (medium ground finch)	87	72	AKZB01067212.1
Zonotrichia albicollis (white-throated sparrow)	86	59	ARWJ01028302.1
Pseudopodoces humilis (ground tit)	82	49	ANZD01009211.1

Table S2. Retroviral reference sequences

Virus name	Abbrev.	Accession
Avian leukemia virus	ALV	AB522906.1
Avian endogenous retrovirus	EAV-HP	AJ623291.1
Human endogenous retrovirus-K (HML2)	HERV-K(HML2)	Y10392.1
Koala retrovirus	KoRV	AF151794.2
Jaagsiekte sheep retrovirus	JSRV	AF136225.1
Mouse mammary tumor virus	MMTV	D16249.1
Mason Pfizer monkey virus	MPMV	AF033815.1
Python endogenous retrovirus	PyERV	AAN77284.1
Human T-cell leukemia virus 1	HTLV-1	AF033817.1
Baboon endogenous retrovirus	BAEV	P10269.1
Feline leukemia virus	FeLV	M18247.1
Murine leukemia virus (Moloney)	MoMLV	J02255.1
Reticuloendotheliosis virus	REV	FJ439119.1
Equine infectious anemia virus	EIAV	M16575.1
Feline immunodeficiency virus	FIV	M25381.1
Human immunodeficiency virus 1	HIV-1	FJ405080.1
Simian immunodeficiency virus (African green monkey)	SIVagm	M30931.1
Caprine arthritis encephalitis virus	CAEV	M33677.1
ChiRV1	ChiRV1	DQ280312.2
Human endogenous retrovirus W	HERV-W	AF135487
Black duck-RV	-	AY820069.1
Blue tit-RV	-	AY820077.1
Guineafowl-RV	-	AY820119.1
Hermit thrush-RV	-	AY820099.1
Toucanette-RV	-	AY820092.1
Tragopan-RV	-	AY820072.1
Taeniopygia guttata alpha-1	Tg alpha1	AC188312.1
Taeniopygia alpha-2	Tg alpha2	ABQF01114967.1
Taeniopygia alpha-3	Tg alpha2	ABOF01089922.1