

Research article

Open Access

## Characterization of an endogenous retrovirus class in elephants and their relatives

Alex D Greenwood\*<sup>1,2</sup>, Claudia C Enlbrecht<sup>3</sup> and Ross DE MacPhee<sup>2</sup>

Address: <sup>1</sup>GSF-National Research Center for Environment and Health, Institute of Molecular Virology, Ingolstädter Landstr. 1, D-85764 Neuherberg, Germany, <sup>2</sup>Department of Vertebrate Zoology, American Museum of Natural History, Central Park West at 79th Street, New York, New York 10024-5192 USA and <sup>3</sup>GSF-National Research Centre for Environment and Health, Institute of Bioinformatics, Ingolstädter Landstr. 1, D-85764 Neuherberg, Germany

Email: Alex D Greenwood\* - [greenwood@gsf.de](mailto:greenwood@gsf.de); Claudia C Enlbrecht - [c.enlbrecht@gsf.de](mailto:c.enlbrecht@gsf.de); Ross DE MacPhee - [macphee@amnh.org](mailto:macphee@amnh.org)

\* Corresponding author

Published: 11 October 2004

Received: 30 July 2004

BMC Evolutionary Biology 2004, 4:38 doi:10.1186/1471-2148-4-38

Accepted: 11 October 2004

This article is available from: <http://www.biomedcentral.com/1471-2148/4/38>

© 2004 Greenwood et al; licensee BioMed Central Ltd.

This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Endogenous retrovirus-like elements (ERV-Ls, primed with tRNA leucine) are a diverse group of reiterated sequences related to *foamy* viruses and widely distributed among mammals. As shown in previous investigations, in many primates and rodents this class of elements has remained transpositionally active, as reflected by increased copy number and high sequence diversity within and among taxa.

**Results:** Here we examine whether proviral-like sequences may be suitable molecular probes for investigating the phylogeny of groups known to have high element diversity. As a test we characterized ERV-Ls occurring in a sample of extant members of superorder Uranotheria (Asian and African elephants, manatees, and hyraxes). The ERV-L complement in this group is even more diverse than previously suspected, and there is sequence evidence for active expansion, particularly in elephantids. Many of the elements characterized have protein coding potential suggestive of activity.

**Conclusions:** In general, the evidence supports the hypothesis that the complement had a single origin within basal Uranotheria.

### Background

ERV-Ls are retroviral elements (retroelements) lacking the envelope gene (*env*) and exhibiting homology to the class of human endogenous retroviruses designated as HERV-L by [1]. Similar retroelements have been identified in several eutherian groups (see below), but their incidence in metatherians and monotremes is not known at present. They presumably arose from successful germ-line infection by *foamy*-like viruses, but when or how many times this might have occurred during the course of eutherian evolution is unknown. From the perspective of evolution-

ary biology it is of great interest that some classes of ERVs are known to retain original functions, including the capacity to produce infectious viral particles [2]. Others have gained novel regulatory functions in the mammalian genome [3]. Formation of the human placenta may depend on expression of a HERV-W element *env* gene [4]. Human immunodeficiency virus (HIV) shares specific functionally homologous sequences with endogenous retroviruses, suggesting the possibility that recombination with ERVs could change the properties of exogenous retroviruses [5]. Thus, ERVs may serve as a variable pool

from which exogenous viruses may diversify. Exogenous retroviruses may have originated from ERVs and ERV-Ls in particular may represent an intermediate between retrotransposons and exogenous viruses [6].

Comparison of ERV-L polymerase (*pol*) gene sequences from 22 mammalian species revealed ERV-Ls that have expanded in copy number and remained active over long periods of time [1]. Phylogenetic analysis of these sequences demonstrated that primates and rodent ERV-L sequences are both diverse and, with few exceptions, monophyletic, whereas carnivore and ungulate ERV-L sequences were polyphyletic. The phylogenetic picture reflects the particularly robust expansion of the primate and rodent ERV-L complement. Importantly, the primates and rodents were the only groups that included ERV-L sequences with protein coding potential and therefore potential transpositional activity. These points suggest that, if the history of active expansion of retroelements within a group can be deciphered, it might be possible to use this information in the same way that parasite data are conventionally used [7], to perform tests of host phylogenetic relationships that are at least logically independent of other data sources. In this connection, the superorder Uranotheria is of particular interest.

Uranotheria [8] is the most recent nomen for a constellation of relationships that has, in fact, been supported by the majority of ungulate specialists throughout the past century. Simpson [9] grouped proboscideans, hyraxes, embrithopods and sirenians under the group-name Paenungulata, but was not certain of its monophyly. Most other authorities have supported this clade, albeit with some variation in content, in the years since Simpson's [9] publication (e.g., [10-14]). McKenna and Bell [8] divided Uranotheria into three major groups, Hyracoidea, Embrithopoda, and Tethytheria. The last is further subdivided into Sirenia and Behemota; behemotans consist of Proboscidea and Desmostylia. Only Hyracoidea, Sirenia, and Proboscidea possess living members.

Morphologically, there is considerable evidence that supports the association of Proboscidea and Sirenia as sister taxa to the exclusion of Hyracoidea [10], and little that appears to contradict it. Fischer and Tassy [15] take the position that alleged hyracoid morphological resemblances to tethytheres are either convergences or misconstrued, on the argument that hyraxes are in fact perissodactyls or closely allied to them. This aspect of the Fischer-Tassy hypothesis is not supported by molecular data [14,16]. On the other hand, it must also be admitted that sequence data have not provided especially strong support for Tethytheria (and, by extension, the monophyly of Uranotheria) [17]. In the most recent exercise in this arena, Asher et al [18] were able to recover Tethytheria

under certain conditions when fossil and morphological data were combined with sequence information, but not when sequence data were used alone.

To investigate whether ERV-L and other retroelements may be useful in resolving phylogenetic questions involving uranotheres at multiple taxonomic levels, we utilized an ERV-L polymerase gene (*pol*) fragment using degenerate primers tested in other mammalian orders. Extending our previous work [19], we found that ERV-L sequence diversity was high in all members of this group and that phylogenetic analysis of our data to a limited extent supported Uranotheria as a distinct clade when sequences that lack coding potential are used. By contrast, sequences that are potentially active form separate monophyletic groups, indicating a more recent origin. Thus, it appears that ancient ERVs reflect the phylogeny of their host like classic genes and more recently active ERVs will tend to be more similar to one another as opposed to their host.

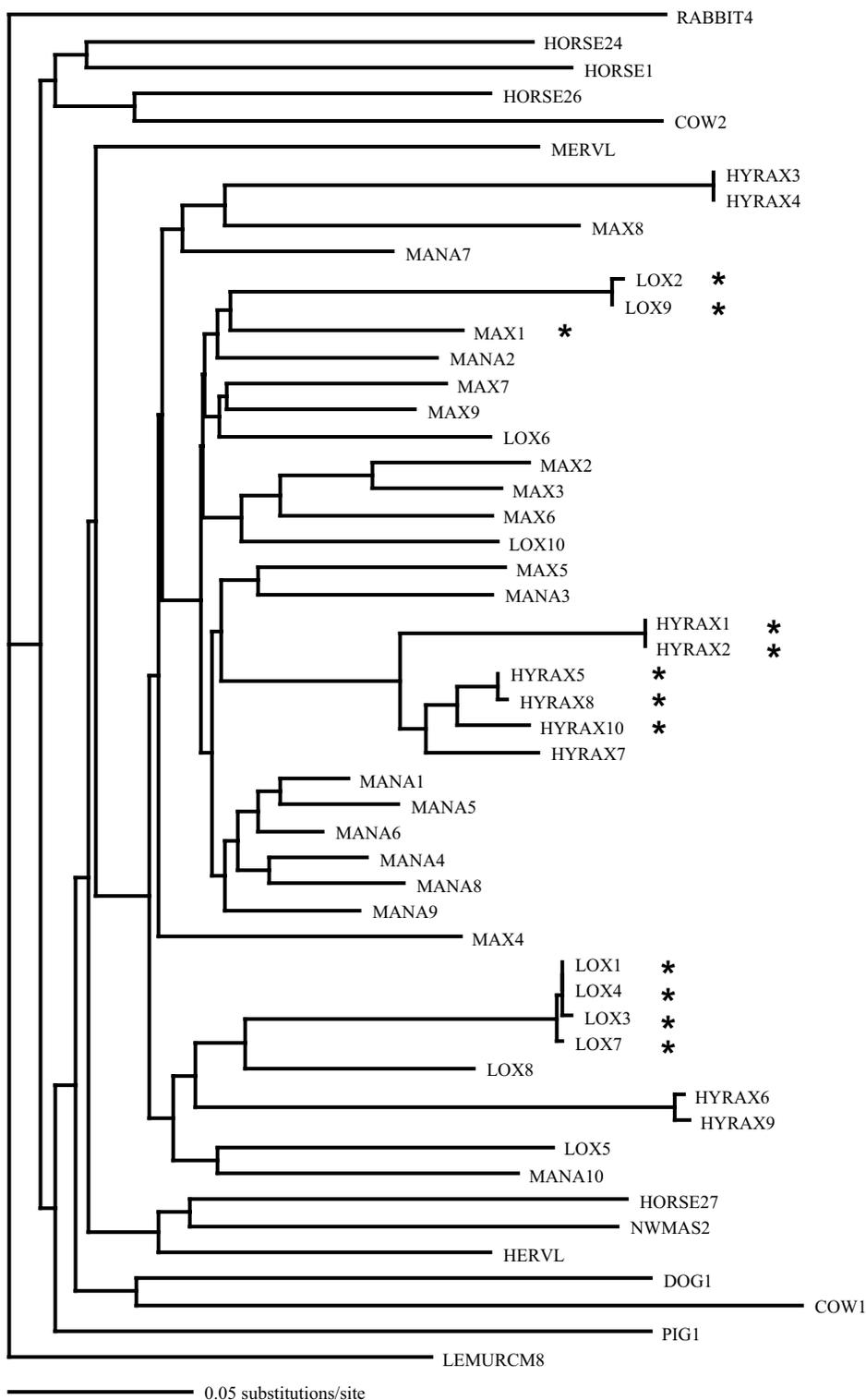
## Results

### Among-clone comparisons

A ~330 bp PCR product was amplified for African elephant, Asian elephant, manatee, and rock hyrax. The products were cloned and 10 clones sequenced for each product. Of the 40 sequences thus developed, only one Asian elephant sequence had no homology to ERV-Ls and was removed from analysis (not shown). No identical sequences were shared among taxonomic groups. All nine Asian elephant and all 10 manatee clones were unique. However, one Asian elephant clone, designated Max3 (accession number AY394573), was a recombination product of clone Max2 (accession number AY394572) and clone Max6 (accession number AY394576). Whether this represents a PCR artifact or is a genomic recombination event is not known. However, it is not expected that recombinational PCR would be observed in modern undamaged DNA [20].

Among the African elephant runs, four clones differed at 0-1 positions. As PCR errors probably account for these minor differences we assume only 6 unique ERV-Ls were discovered for this individual. Similarly, the hyrax sample yielded 3 groups comprised of 2 identical sequences, while two other sequences differed at 5 positions. Thus, 5 unique ERV-Ls were also obtained for *Procavia*.

Recovered sequences were compared to a mouse element with full coding potential in the *gag* and *pol* genes (MuERV-L, GenBank no. Y12713). Twelve clones were in frame with no stop codons. However, only 6 of the total 12 were unique (Figure 1). This is surprising, as 87 sequences from 22 mammalian species previously revealed only 7 sequences with coding potential [1]. Among the 39 sequences determined here, 6 unique



**Figure 1**  
 Endogenous retrovirus type L (ERV-L) phylogeny. Non-uranothere sequence designations taken from [1]. Uranothere designations are, Max (*Elephas maximus*), Lox (*Loxodonta africana*), Mana (*Manatus trichechus*), and Hyrax (*Procavia capensis*). Neighboring tree of all uranothere sequences including representative ERV-L elements from other mammalian orders. "\*" designates sequences with coding potential.

sequences had coding potential among only 4 species. The observed sequence diversity and frequency of observed coding potential is consistent with active ERV-L expansion in these four species and consistent with results with a smaller internal fragment from the same groups (plus extinct woolly mammoth in the proboscidean sample) [19].

### Phylogenetic analysis

A heuristic search of the entire uranothere ERV-L data set yielded 12 equally parsimonious trees, a strict consensus of which (Figure 2) showed poor recovery of accepted clades within Uranotheria and low bootstrap support at each node. While neighbor-joining analysis produces a tree with uranothere sequences as a monophyletic group, branch lengths in some cases were very short and bootstrap support under any method used was not statistically significant (Figure 1). On the assumption that sequences with potential RT activity may have been under different evolutionary constraints and may differ in their phylogenetic resolution, the sequences with no stop codons in the retrieved *pol* gene were analyzed separately from those with stop codons (Figure 3, 4 and 5). Those with potential RT activity grouped as distinct monophyletic groups, possibly reflecting their more recent activity and thus showing closer affinity to one another as opposed to other related ERV-L sequences (Figure 3). Those with stop codons showed a different picture with modest support for Uranotheria as one might expect for single or low copy sequences that have been transmitted vertically over time (Figure 4). Likelihood analysis of the data produced similar groupings, though with weak support, suggesting the associations found are not an artifact of the phylogenetic analysis methodology (Figure 5). However, ME analysis did not produce statistically significant resolution for any ERV group examined including non-uranothere ERVs (not shown). Poor resolution within Uranotheria with all analyses could be due to several factors, although the likeliest is different ages of individual element copies. This is not unlike the situation with various sequences recovered from primates and rodents: some sequences reconstruct accepted ordinal groupings, while others do not. For example, in the study by Bénit et al. [1], one New World monkey sequence (As2) grouped with two dog sequences and not with other primates. In other mammals, sequences (e.g., those retrieved from cow and horse) were dispersed in no evident pattern.

### Discussion

The most important finding resulting from this study is that elements that have undergone expansion—i.e. have remained transpositionally active—are the ones that are most likely to group monophyletically and those that have not tend to be consistent in their higher-level taxonomic distribution (in this case, at the superordinal level).

Thus, there are different elements within the same family demonstrating different evolutionary trajectories. It is reasonable to suppose that for some groups, such as rodents, primates, and uranotheres, continued expansion of active ERV-L elements was tolerated by the host. By contrast, in other ungulate and carnivore lineages active elements were not inherited or were silenced early during their evolution and ERV-L expansion did not occur. Older elements tend to evolve as typical orthologous sequences. An advantage of the great diversity of elements is that with a single PCR, cloning, and determination of multiple clone sequences one retrieves multiple independent sequences with which to do phylogenetic analysis. In these regards, the uranothere evidence is consistent with results previously reported for primates and rodents in which older elements and elements that have undergone bursts of transposition were found coexisting. In each of these groups there is now good sequence evidence for retroelements that have retained coding capacity, which is of some interest since HERV-L *pol* (for example) is known to be expressed in specific tissue types [21]. This indicates that, in addition to potential transpositional activity, retroelements other than syncytin may have acquired biological functions important for their hosts.

### Conclusions

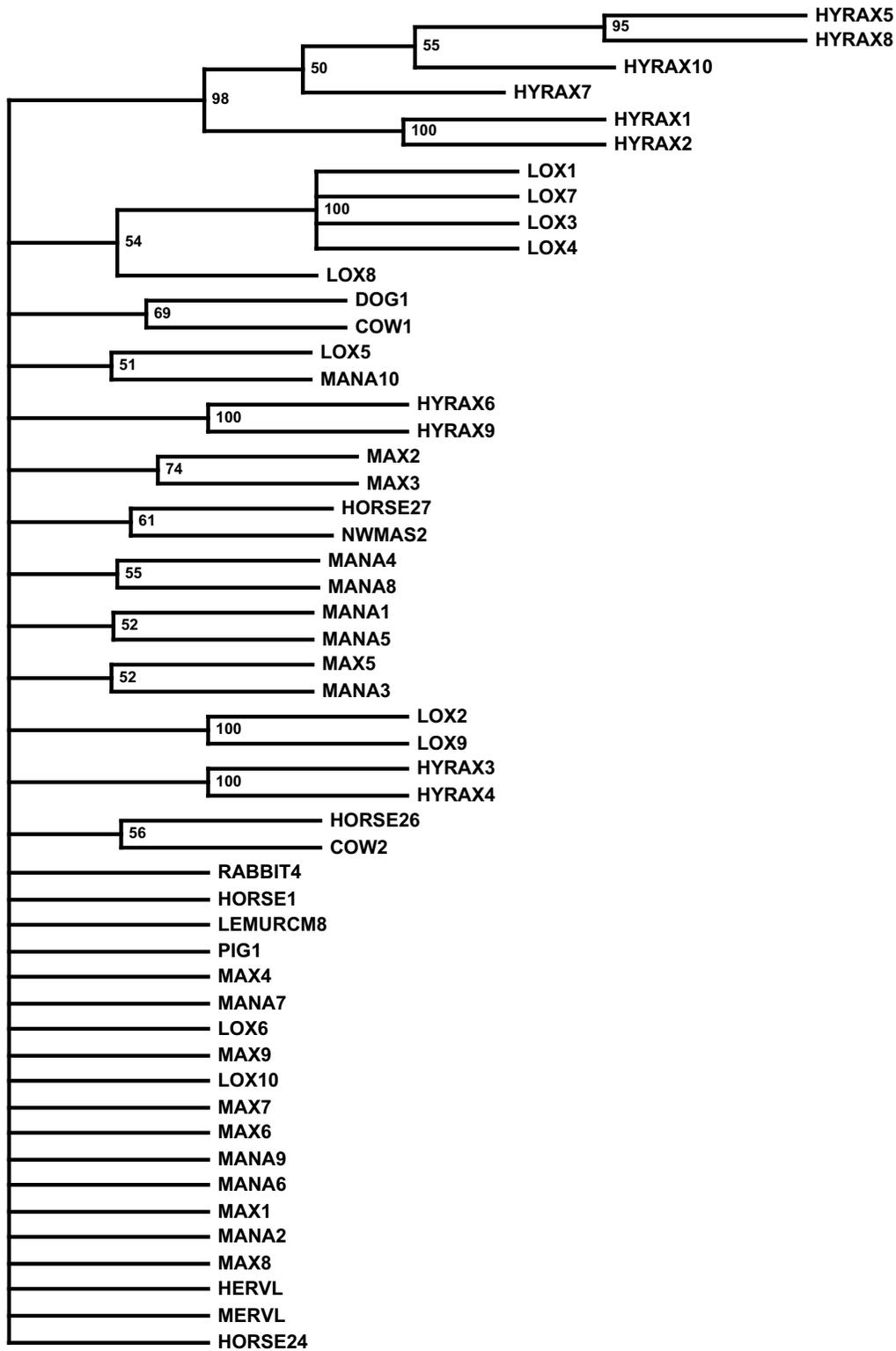
In terms of our general results, elephants and the manatee were found to contain the most diverse sequences, while hyrax showed comparatively less diversity. This finding is consistent with results from our previous study involving extinct elephantids [19]. Elements that have been investigated in other ungulates (bovids and suids) do not yield monophyletic groupings and are represented by low copy numbers [1], suggesting that expansion did not occur (or has not recently occurred) in these taxa. Although critical studies would have to be undertaken to demonstrate the matter conclusively, it appears that amplification and diversification of ERV-L elements were independent events in primates, rodents and uranotheres.

Although this study shows that the value of ERV-Ls for the narrow purpose of phylogenetic reconstruction is limited at higher taxonomic levels, ERV-L is only one class among many different groups of ERVs in mammalian genomes. Some regions of ERVs have been used successfully to reconstruct phylogenies at lower taxonomic levels [22]. Additional transposable elements could serve as phylogenetic markers in a manner similar to ERV-L in the present study, while providing multiple independent sequences to test ordinal level phylogenies.

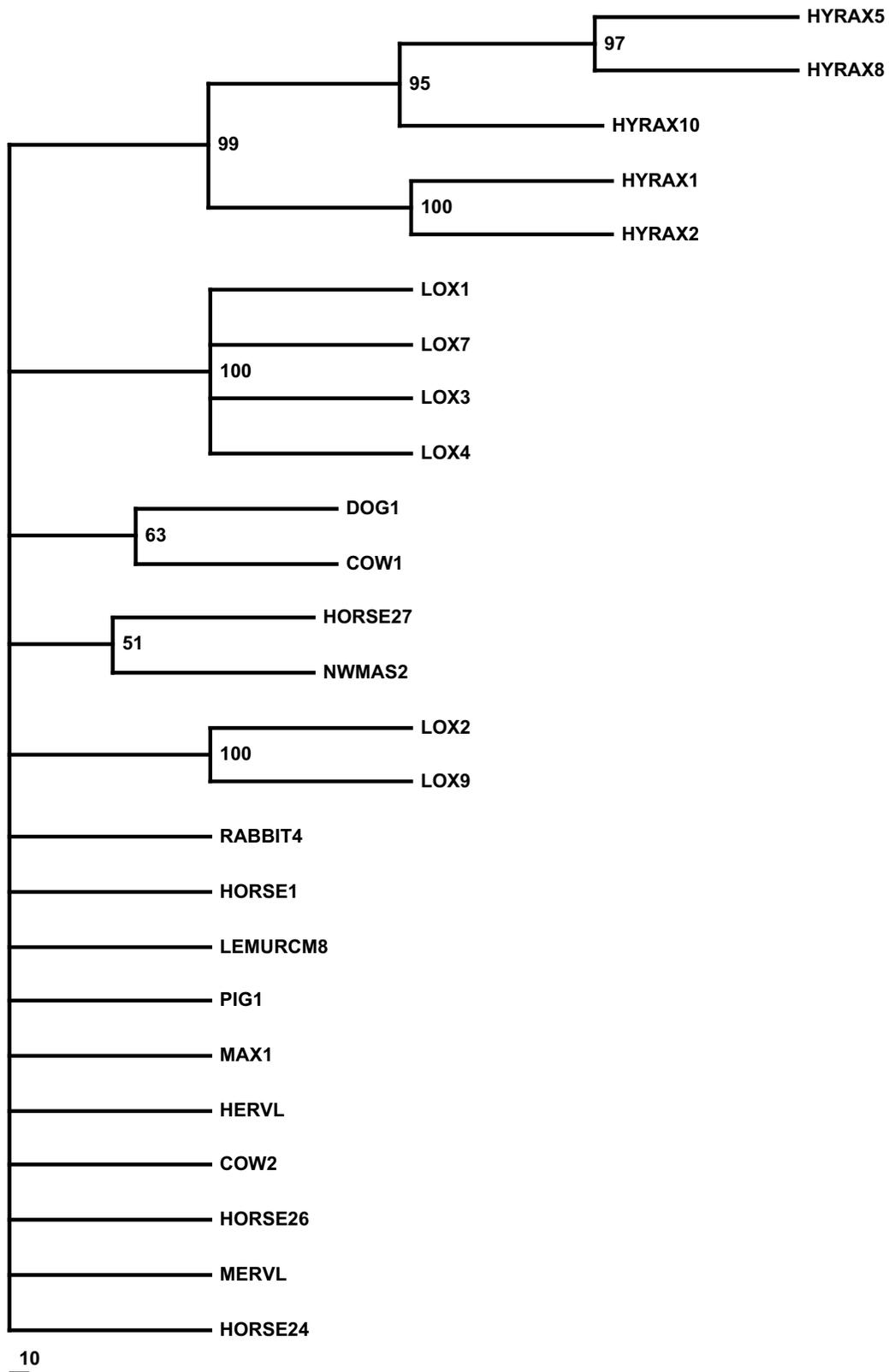
### Methods

#### Samples

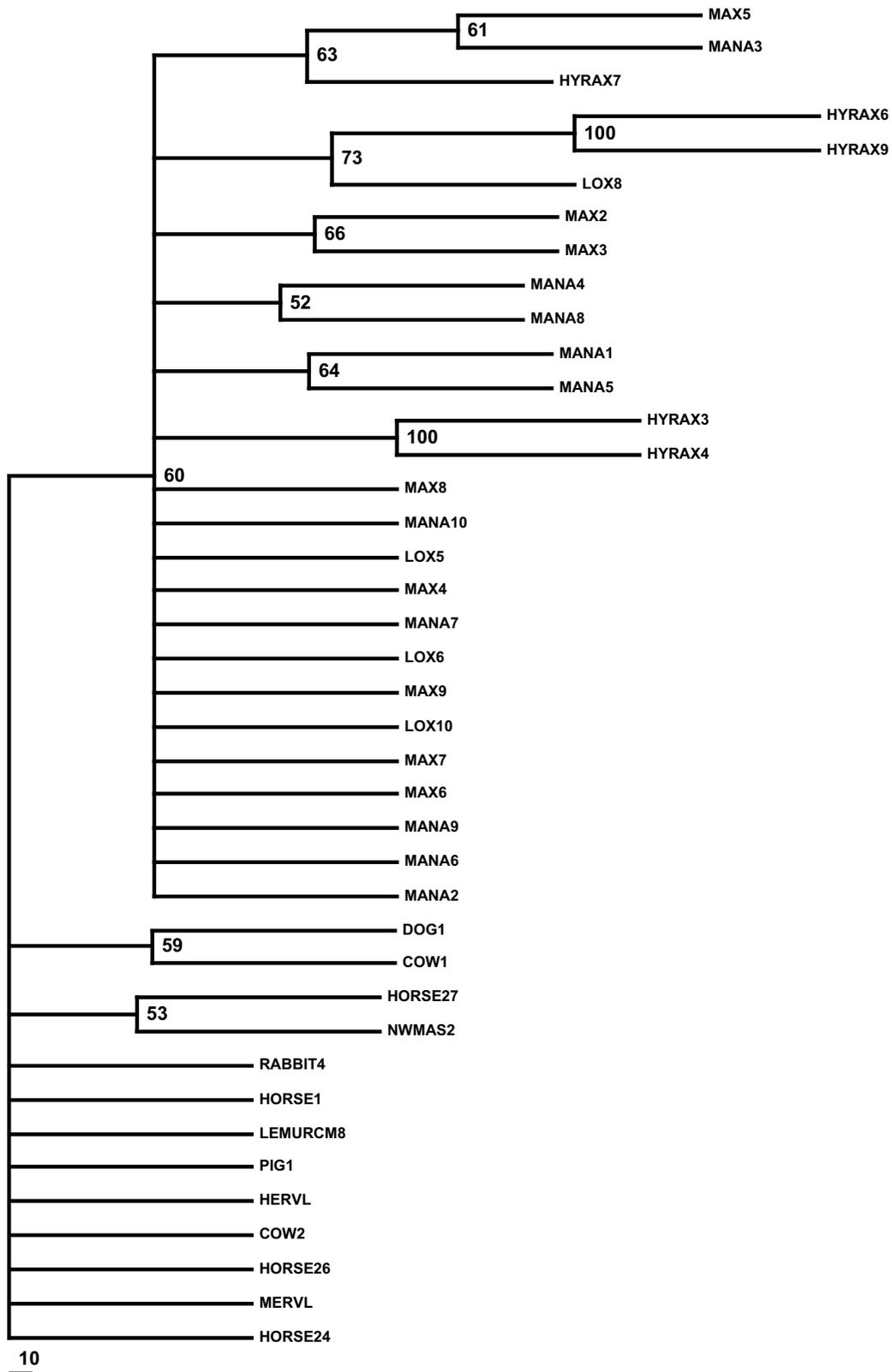
African elephant (*Loxodonta africana*) DNA was supplied by N. Georgiadis of the M'Pala Research Centre, Kenya.



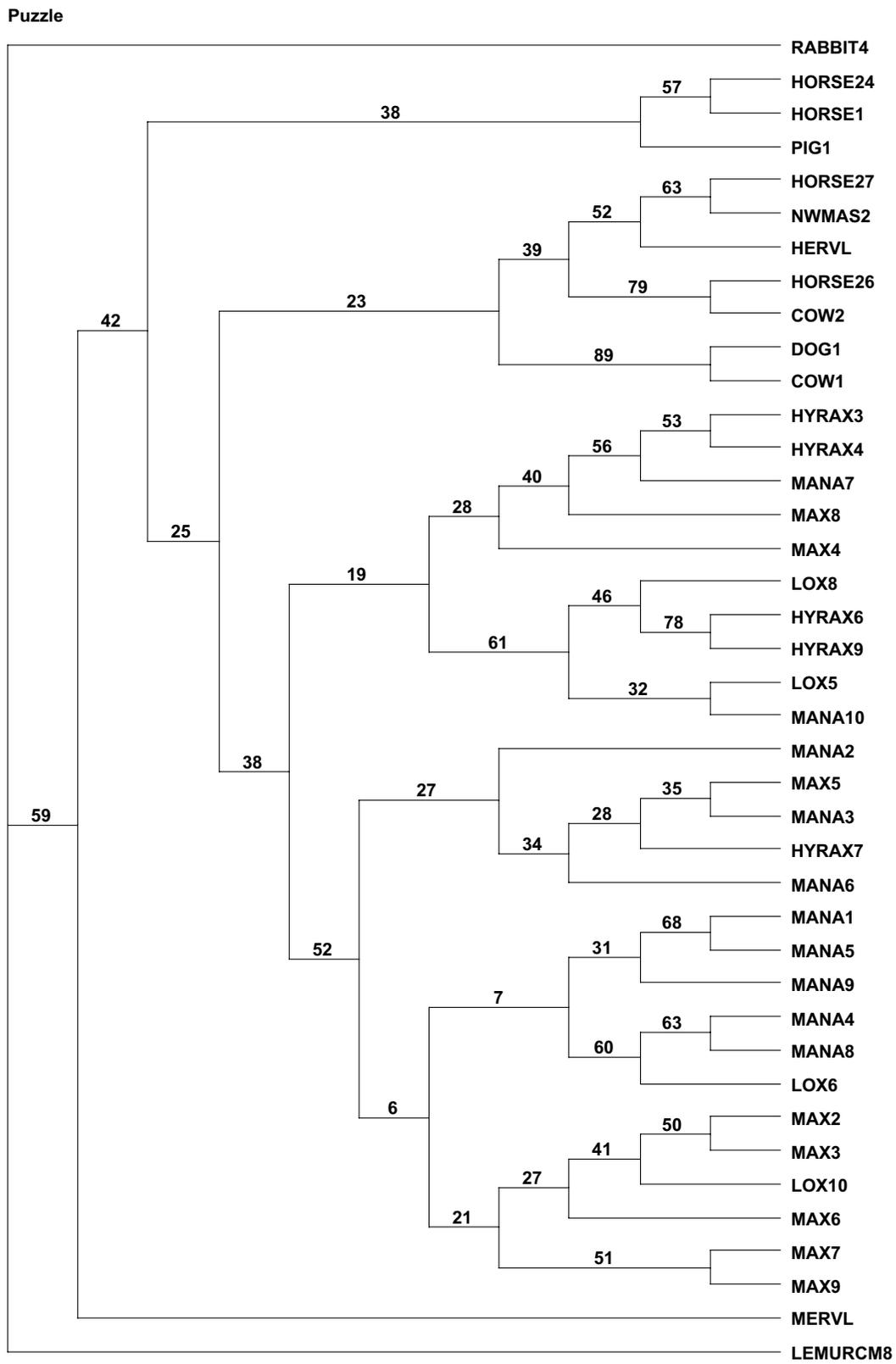
**Figure 2**  
 Maximum parsimony bootstrap consensus tree of all uranothere ERV-L sequences. Bootstrap values over 50% are shown. The scale bar indicates the number of steps.



**Figure 3**  
Bootstrap consensus tree of Uranothere ERV-L sequences with coding potential only.



**Figure 4**  
Bootstrap consensus tree of Uranothere ERV-L sequences without coding potential.



**Figure 5**  
 Quartet Puzzle maximum likelihood tree of sequences without coding potential. Puzzle support for each node is indicated.

Asian elephant (*Elephas maximus*) blood was provided by J. Hektor of the Tierpark Hellabrunn, Munich. Manatee (*Trichechus manatus*) blood was provided by D. Murphy of the Lowry State Park, Florida. Hyrax (*Procavia capensis*) muscle was provided by G. Amato of the Bronx Zoo, New York.

#### DNA extractions

One ml of blood or approximately one gram of tissue was incubated in 1–2 ml 10 mM Tris-/Cl (pH 7.5), 10 mM EDTA (pH 8.0), 50 mM NaCl, 2% SDS, and 0.6 mg/ml Proteinase K overnight at 37 C, extracted with phenol and chloroform, and subsequently concentrated with 50 µl Millipore Ultrafree MC 30,000 NMWL columns or precipitated in 2.5 volumes ethanol and 1% NH<sub>4</sub> Oac.

#### PCR, cloning, and sequencing

PCR primers for the ERV-L *pol* gene are described in [1]. Three µl of extract was added to 50 µl PCR containing standard buffer supplied by Boehringer Mannheim and 30 PCR cycles performed. PCR products were cloned using the pGEM-T cloning system (Promega). After heat shock into bacteria, ampicillin and blue/white selection, colonies were picked with a sterile pipette tip and added to 30 µl PCR reactions where M13 forward and reverse primers were used to amplify inserts for 25 cycles using the same buffer system described for ERV-L amplifications and as described in [19]. Five µl of the colony PCR products were visualized on ethidium-stained gels. Insert positive PCR reactions were purified with QIAquick columns and sequenced with T7 and SP6 primers using an ABI 377 sequencer.

#### Phylogenetic analysis

##### Alignment

Representatives of each clade determined by Bénit et al. [1] were included in an alignment with the elephant, manatee, and hyrax sequences determined. HERV-L, X89211; MERV-L, Y12713; NWM (AS2), AJ233633; Lemur CM8, AJ233645; horse1, AJ233650; horse24, AJ233654; horse26, AJ233655; horse27, AJ233656; pig1, AJ233661; cow1, AJ233662; cow2, AJ233663; dog1, AJ233665; rabbit4, AJ233627. Alignments were performed using ClustalX [23] and adjusted where necessary. Elephant, manatee and hyrax sequences have been deposited in GenBank (accession numbers AY394571-AY394609)

##### Phylogenetic methods

Maximum parsimony and neighbor joining analysis was performed using PAUP 4.0b [24]. Heuristic searching including all uranothere ERV-L sequences yielded 12 maximum parsimony trees (MPTs). 100 bootstrap replicates were performed to test MPT robustness, the strict consensus of which is shown in Figure 2. A bootstrap consensus

tree using only sequences with no stop codons yielded 2 trees. The strict consensus of these trees following 1000 bootstrap replicates is shown in Figure 3. A consensus tree of 1000 bootstrap replicates, this time excluding sequences with coding potential, is shown in Figure 4. Maximum likelihood was performed using quartet puzzle in PAUP 4.0b after determining the evolution model as HKY +G using Modeltest 3.5 [25]. 10,000 puzzling steps were employed to determine the tree topology. Minimum evolution trees were generated using the program Mega2 [26].

#### Authors' contributions

ADG oversaw the molecular genetic studies, coordinated the study, participated in the sequence alignment and phylogenetic studies and co-drafted the manuscript. CCE participated in the phylogenetic analysis. RDEM participated in the design and coordination of the study and co-drafted the manuscript.

#### Acknowledgments

The authors wish to thank F. Lee and P. Marx (Tulane University and Aaron Diamond Aids Research Center in New York City) for technical assistance. P. Wynne helped execute the figures. We also thank N. Georgiadis of the M'Pala Research Center (Kenya) and J. Hektor of Tierpark Hellabrunn (Munich) for African and Asian elephant samples respectively. We wish to thank D. Murphy of the Lowry State Park Zoo (Florida) for the manatee specimen and G. Amato of the Bronx Zoo (New York) for the hyrax specimen used in this study. D. Ho (ADARC) is thanked for his support of this project. The work was supported by grants from the National Science Foundation, the Irene Diamond Fund, and the Adler Fund.

#### References

1. Bénit L, Lallemand JB, Casella JF, Philippe H, Heidmann T: **ERV-L elements: a family of endogenous retrovirus-like elements active throughout the evolution of mammals.** *J Virol* 1999, **73**:3301-3308.
2. Patience C, Takeuchi Y, Weiss RA: **Infection of human cells by an endogenous retrovirus of pigs.** *Nat Med* 1997, **3**:282-286.
3. Löwer R: **The pathogenic potential of endogenous retroviruses facts and fantasies.** *Trends Microbiol* 1999, **7**:350-356.
4. Mi S, Lee X, Li X, Veldman GM, Finnerty H, Racie L, LaVallie E, Tang XY, Edouard P, Howes S, Keith JC Jr, McCoy JM: **Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis.** *Nature* 2000, **403**:785-789.
5. Yang J, Bogerd HP, Peng S, Wiegand H, Truant R, Cullen BR: **An ancient family of human endogenous retroviruses encodes a functional homolog of the HIV-1 Rev protein.** *Proc Natl Acad Sci USA* 1999, **96**:13404-13408.
6. Temin HM: **Origin of retroviruses from cellular moveable genetic elements.** *Cell* 1980, **21**:599-600.
7. Page RDM: **Parallel phylogenies: reconstructing the history of host-parasite assemblages.** *Cladistics* 1994, **10**:155-173.
8. McKenna M, Bell SK: *Classification of mammals above the species level* New York: Columbia University Press; 1997.
9. Simpson GG: **The principles of classification and a classification of mammals.** *Bull Amer Mus Nat Hist* 1945, **85**:1-350.
10. McKenna MC: **Toward a phylogenetic classification of the Mammalia.** In *Phylogeny of the primates, a multidisciplinary approach* Edited by: Luckett PW, Szalay FS. New York: Plenum; 1975:21-46.
11. Novacek MJ: **The skull of lepidictid insectivorans and the higher-level classification of eutherian mammals.** *Bull Amer Mus Nat Hist* 1986, **183**:1-111.

12. Domning D, Ray CE, McKenna MC: **Two new Oligocene desmostylians and a discussion of tethythere systematics.** *Smithsonian Contrib Paleobiol* 1986, **59**:1-56.
13. Shoshani J: **Hyracoidea-Tethytheria affinity based on myological data.** In *Mammal phylogeny, Placentals Volume 2*. Edited by: Szalay FS, Novacek MJ, McKenna MC. New York: Springer-Verlag; 1993:235-256.
14. Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryder OA, O'Brien SJ: **Molecular phylogenetics and the origin of placental mammals.** *Nature* 2001, **409**:614-618.
15. Fischer MS, Tassy P: **The interrelation between Proboscidea, Sirenia, Hyracoidea, and Mesaxonia: the morphological evidence.** In *Mammal phylogeny, Placentals Volume 2*. Edited by: Szalay FS, Novacek MJ, McKenna MC. New York: Springer-Verlag; 1993:217-234.
16. Springer MS, Cleven GC, Madsen O, de Jong WW, Waddell VG, Amrine HM, Stanhope MJ: **Endemic African mammals shake the phylogenetic tree.** *Nature* 1997, **388**:61-4.
17. Amrine HM, Springer MS: **Maximum likelihood analysis of the tethythere hypothesis.** *J Mammal Evol* 1999, **6**:161-176.
18. Asher RJ, Novacek MJ, Geisler JH: **Relationships of endemic African mammals and their fossil relatives based on morphological and molecular evidence.** *J Mamm Evol* 2003, **10**:131-194.
19. Greenwood AD, Lee F, Capelli C, DeSalle R, Tikhonov A, Marx PA, MacPhee RDE: **Evolution of endogenous retrovirus-like elements of the woolly mammoth (*Mammuthus primigenius*) and its relatives.** *Mol Biol Evol* 2001, **18**(5):840-847.
20. Pääbo S, Irwin DM, Wilson AC: **DNA damage promotes jumping between templates during enzymatic amplification.** *J Biol Chem* 1990, **265**:4718-4721.
21. Seifarth W, Spiess B, Zeilfelder U, Speth C, Hehlmann R, Leib-Mösch C: **Assessment of retroviral activity using a universal retrovirus chip.** *J Virol Methods* 2003, **112**:79-91.
22. Johnson WE, Coffin JM: **Constructing primate phylogenies from ancient retrovirus sequences.** *Proc Natl Acad Sci USA* 1999, **96**:10254-10260.
23. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nucleic Acids Res* 1997, **24**:4876-4882.
24. Swofford DL: *PAUP\*: Phylogenetic analysis using parsimony (\*and other methods), version 4.0b\** Sunderland, MA: Sinauer Associates; 2002.
25. Posada D, Crandall KA: **MODELTEST: testing the model of DNA substitution.** *Bioinformatics* 1998, **14**:817-818.
26. Kumar S, Tamura K, Jakobsen IB, Nei M: **MEGA2: molecular evolutionary genetics analysis software.** *Bioinformatics* 2001, **17**:1244-1245.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

