

11-1-2009

Reading between the LINEs to see into the past

David A. Ray
Mississippi State University

Roy N. Platt
Mississippi State University

Mark A. Batzer
Louisiana State University

Follow this and additional works at: https://digitalcommons.lsu.edu/biosci_pubs

Recommended Citation

Ray, D., Platt, R., & Batzer, M. (2009). Reading between the LINEs to see into the past. *Trends in Genetics*, 25 (11), 475-479. <https://doi.org/10.1016/j.tig.2009.09.006>

This Article is brought to you for free and open access by the Department of Biological Sciences at LSU Digital Commons. It has been accepted for inclusion in Faculty Publications by an authorized administrator of LSU Digital Commons. For more information, please contact ir@lsu.edu.



Published in final edited form as:

Trends Genet. 2009 November ; 25(11): 475–479. doi:10.1016/j.tig.2009.09.006.

Reading between the LINEs to see into the past

David A. Ray¹, Roy N. Platt¹, and Mark A. Batzer²

¹ Department of Biochemistry and Molecular Biology, Box 9650, Mississippi State University, Mississippi State, MS 39762, USA

² Department of Biological Sciences, Louisiana State University, 202 Life Sciences Building, Baton Rouge, LA 70803, USA

Abstract

Transposable elements are an important source of genome diversity and have crucial role in genome evolution.. A recent study by Zhao *et al.* describes novel patterns of transposable element (TE) diversification in the genome of the extinct mammoth, *Mammuthus primigenius*. Analysis of *Mammuthus* has provided a unique genome landscape, a pivotal species for understanding TEs and genome evolution, and hints at the diversity we verge on discovering by expanding our taxonomic sampling among genomes. Strategies based on the work. might also revolutionize investigations of the interface between TE dynamics and genome diversity.

Discovering TEs in the mammoth genome

Transposable elements (TEs, Box 1) have had substantial impacts on eukaryotic genomes throughout life's history, being responsible either directly or indirectly for much of the genomic diversity we see today. Not surprisingly, studies of TE impacts on human and non-human primate genomes are numerous and well developed. We know for example how the movement of TEs has impacted human diseases [1], genome size [2–8], and the transcriptome [9–11]. But how well does our little corner of the genomic world reflect TE diversity and impact in a more general sense? The broader mammalian perspective is only now being investigated, and although we are starting to answer this question (see Ref [12] for an example), many gaps in our knowledge remain.

Recently, Zhao *et al.* [13] applied next-generation sequencing (454) to address the question in a unique way – by investigating TE amplification dynamics in the woolly mammoth (*Mammuthus primigenius*), a species that has been extinct for ~10 000 years. Using the massive data available from the mammoth genome project, they determined likely TE content using an iterative process to identify relatives of known TE families. It was immediately clear that the sheer volume of TEs within the mammoth genome sets it apart from other mammals. The uniqueness of the mammoth genome can be seen in several TE-related areas, including genomic expansion, TE diversity, and a likely case of horizontal transfer of a Class I TE. Such observations hint at tremendous potential for finding a vast array of TE associated diversity in mammals as their genomes are explored. Zhao *et al.* [1] also demonstrate the impressive potential of next-generation sequencing with regard to subgenomic analysis. The increased throughput provided by platforms such as the 454, Illumina, and SOLiD systems has

Corresponding author: MAB; mbatzer@lsu.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

revolutionized numerous aspects of biological analysis from gene discovery, to expression profiling and whole genome sequencing. In particular, the combination of high-throughput DNA sequencing and the repetitive nature of TEs make certain next-generation sequencing platforms ideal for rapid and inexpensive genome wide TE analyses. Here we suggest how researchers can use next-generation sequencing approaches to implement broad genome-wide surveys of TE content and discuss the likely impact this will have for our understanding of the wider role of TEs. Such strategies could radically alter our ability to investigate and understand the complex interface between TE amplification dynamics and genome diversification

TE content and genome dynamics in mammoth

The mammoth genome contains a greater proportion of TEs than any mammal analyzed to date. This led to Zhao et al. to highlight the potential connection between the increased genome size (~50% larger than our own [14,15]) and rapid expansions of particular TEs. Increased genome size has long been considered a potential consequence of TE expansion [16] and many mammals appear to have accommodated massive TE-mediated genome expansions, whereas certain animals (e.g. birds, reptiles and some fish) have had a tendency to eliminate them [17–21]. For example, analysis of the recently sequenced genome of *Anolis carolinensis* revealed that although these lizards have several recently active lineages of LINES, they have essentially reached equilibrium between TE insertion and removal over the past several million years [18]. Observations of the mammoth genome, along with many other comparisons between mammalian and non-mammalian taxa, suggest that we and our hair-bearing relatives share a unique ability to accommodate some TE expansions while repelling others. Several hypotheses have been advanced to explain this observation include the utilization of DNA methylation as a control mechanism [22–24], decreased ectopic inter-element recombination [21,25], and increased permissiveness for some families to allow for inter-element competition and selection [21].

It is important to note, however, that the mammoth genome expansion was probably not the result of the common mammalian LINE, L1, but instead appears to be the result of both L1 and a nearly-parallel RTE (one of eleven well-defined lineages of LINES [26]) element expansion. Zhao *et al.* found that as much as 12% of the mammoth genome consists of RTEs, whereas the mammal with the next highest RTE proportion is the opossum at a mere 2.3% [27]. Thus, the mammoth is the first eutherian genome characterized to have accommodated multiple simultaneous LINE expansions. Most other TE types (SINES, LTRs, LINES (L1 and L2), and DNA transposons) in mammoth DNA were on par with or in fewer numbers than in other mammalian genomes. Significantly, RTE elements are absent in armadillo [13], cetartiodactyls, primates, carnivores and rodents [28] but are found in ruminants and at least two afrotherian clades, tenrec and the modern elephant [29] (Figure 1). This distribution of RTEs lends support to the idea that each repeat is a unique genomic invasion by a lineage of LINES with an unprecedented ability to spread via horizontal transfer [28,30,31].

Although not the main focus of the study, Zhao *et al.* also noted relatively recent activity by Tigger1 (a Class II element or DNA transposon) in the mammoth genome. With the exception of a single bat genome that has experienced multiple massive waves of recent DNA transposon activity [32–34], many mammals including *M. primigenius* have repelled DNA transposons rather successfully. However, this Tigger1 expansion adds to a growing number of studies suggesting that there are rare occasions in which mammalian genomes are impacted by single Class II lineage expansions [35,36] (Ray *et al.*, unpublished). The increasing number of these isolated instances demonstrates that a general shutdown of mammalian Class II TEs, as suggested by numerous studies [37,38], has, at the very least, been subverted by selected elements via multiple instances of horizontal transfer into some mammalian genomes.

These observations suggest that our focus as researchers is to now determine the answers to several likely interrelated questions. What is the mechanism of these horizontal TE transfers, both Class I and Class II? What makes some genomes more susceptible and/or tolerant of TEs and novel genomic invasions by TEs? Why are some TE families, like RTE, better able to ‘jump’ between genomes than others? It is clear that the answers to these questions will not be found by the relatively limited sampling of genomes currently available to us. Instead, extensive study of a wide variety of mammalian and non-mammalian genomes will be necessary to answer any one of these questions. Fortunately, the data presented by Zhao *et al.* provide a start to the process by expanding our knowledge of TE dynamics to a taxonomically important mammalian lineage, Afrotheria,

The importance of Afrotheria

For obvious reasons, most genomic sequencing and analyses tend to be focused on biomedical model species, such as *Rattus*, *Mus*, and various non-human primates [38–41]. However, a thorough knowledge of mammalian TE dynamics is only possible when appropriate outgroups are also examined. One of the more important aspects of the *M. primigenius* analysis is that the authors chose to study a unique and important mammalian lineage that serves as a basal group within Class Mammalia. After the metatherian–eutherian divergence ~105 mya [42], it is estimated that the earliest diverging clade of extant mammals, Atlantogenata, arose quickly (~103 MYA; [42]) almost twenty million years before the next major divergence, Laurasiatheria [43]. As a member of Atlantogenata, Afrotheria is one of the earliest diversifications of Atlantogenata (Figure. 1). As a result, Zhao *et al.*’s analysis of *M. primigenius* is an important contribution to the study of genome dynamics associated with the protherian, metatherian, and eutherian diversifications.

Subgenomic targeting using next-generation sequencing technology

Of broader significance, Zhao *et al.* have successfully harnessed the power of next-generation sequencing to target a particular genome component. Indeed, the power of next-generation sequencing for understanding the evolutionary patterns of repetitive sequences and their impact on genome evolution has recently been shown in work in the pea (*Pisum sativa*) genome [28] and in the soybean genome [29]. The use of 454 sequencing technology seems particularly well suited to genome-wide TE analysis. [44]. As seen from Zhao *et al.*’s analysis in the mammoth genome, the relatively long average read-lengths of 454 technology (~250 bp for FLX chemistry and 400 bp or more for Titanium chemistry) enabled the identification of full insertions of smaller, nonautonomous elements, such as SINEs and MITEs (Minature Inverted-repeat Transposable Elements): a task that would be difficult for other technologies that have higher throughput but considerably shorter reads (Illumina and SOLiD) on the order of 75+ bp. It is now clear that in complex genomes (e.g. the plants examined by Macas *et al.* [45] and Swaminathan *et al.* [44], and now, the extinct *M. primigenius*), TE content can be accurately surveyed using the random genomic fragments targeted by 454 pyrosequencing.

What is the best way to use the sequencing tools available? For example, can we examine multiple taxa in a single 454 run? We suspect that this might be an efficient strategy. Using the older FLX chemistry, Macas *et al.* [45] obtained ~33 Mb of data from the 4.3 Gb pea genome (~ 0.77% of the genome), with the data consisting of 319 402 reads with an average length of 104 bp. Using this data, they identified what are likely to be all of the major TE families in the genome. The new Titanium 454 chemistry from Roche promises one million reads averaging 400 bp, or 0.4 Gb of data. Assuming an average mammalian genome size of 3.3 Gb (<http://www.genomesize.com>), a single Titanium run would provide a researcher with a random sample of 12% of a single genome, much more than is necessary for surveying mammalian TEs. It would therefore be more cost effective and efficient for a researcher to

subdivide the run among 10 taxa (1.2% of the genome for each), thus generating the data for a taxonomic survey of TE dynamics in a cluster of taxa rather than just one. As a result, we would be making large-scale genome-wide comparisons and investigations of TE dynamics in a broad range of taxa the norm rather than the exceptional study.

Concluding remarks

By providing a thorough analysis of TEs within the mammoth genome Zhao *et al.* [13] have advanced the quest to understand TE dynamics in mammalian genomes. Although the mammoth has some typical mammalian genome qualities, such as a relative lack of recent DNA transposon activity, other characteristics, such as the increased activity of RTEs and larger overall size, make this genome unique, further highlighting its significance. The data also point to the potential variety of mammalian TE dynamics that might be just around the corner given our rather limited sampling to date.

Still, several questions remain. For example, how does TE diversity impact species diversity? In some cases, we see a correlation between the two – consider the bat genus *Myotis* with its massive Class II TE activity within the same historical period of its world-wide diversification into 100+ species [33]. In other cases, there is no obvious connection – the bat genus *Pteropus* is nearly as species-rich but appears to have experienced a shutdown of all TE activity [46]. Obviously, mere TE activity is not enough to ensure diversification. Under what ecological and genomic conditions would such activity contribute to adaptive radiations?

What is the mechanism through which the observed instances of horizontal transfer might occur? There must be a vector by which transferred elements are moved from genome to genome. The most obvious place to look would be among the blood-borne parasites and the pathogens (viruses in particular) that they harbor. However, probing random parasite and pathogen genomes would be a rather inefficient methodology.

Depending on taxon selection, the goals of both studies could be easily addressed. For example, examinations of taxa sharing similar ecological niches but distinct taxonomic distributions and levels of species diversity might be a way to investigate whether TE expansions have impacted species diversity. Alternatively, targeted 454 sequencing of parasites that are shared among organisms known to have participated in horizontal transfer events parasites might yield results of interest to researcher attempting to identify the mechanism of horizontal transfer of elements.

Regardless, this study and others make it clear that we are entering a new phase in genomic research. By utilizing high-throughput genome sequencing and available computational tools efficiently there is little reason for us not to gather genome-scale data in an effort to investigate the interface among TE dynamics, genome change and species diversification.

Box 1

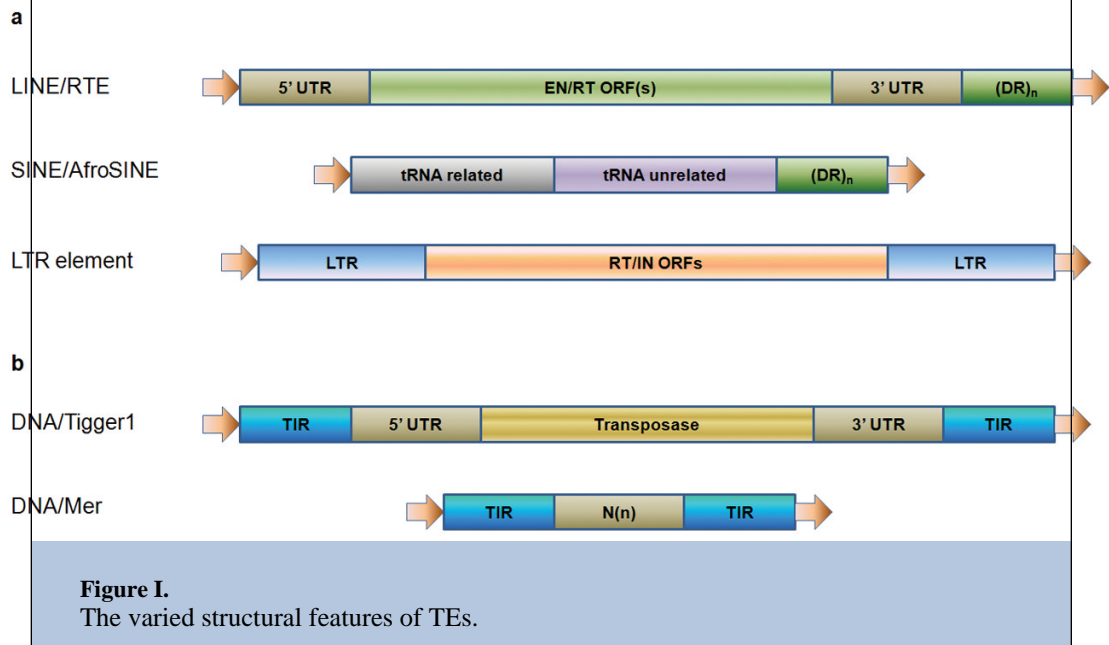
Transposable element structure

Transposable elements are repetitive DNA sequences that accumulate in genomes via multiple mechanisms. Class I elements (a), the retrotransposons, utilize a 'copy and paste' method referred to as retrotransposition. With these elements, the original DNA copy in the genome is first transcribed to mRNA. This transcript is often used as a template by reverse transcriptase to form a DNA molecule that is then inserted into a new location in the genome via a process known as target-primed reverse transcription [47]. There are several subgroups of Class I elements including the LTR (long terminal repeat) and non-LTR retrotransposons (including the Long INterspersed Elements, LINEs). Autonomous elements within these

categories encode much of the enzymatic machinery required for mobilization. In the case of LINE/RTE, the principle Class I element described by Zhao *et al.*, one or two open reading frames (ORFs) provide endonuclease (EN) and reverse transcriptase (RT) activity. LTR element structure differs from LINEs in the identity and organization of ORFs. Often, they include a virus-like integrase (IN) coding sequence, and do not contain a tract of direct repeats (DR).

Class II elements (b), the DNA transposons and their derivatives, are common in many organisms from bacteria to humans. First discovered by Barbara McClintock in maize, DNA transposons differ from Class I elements in that they often utilize a 'cut and paste' mechanism. In other words, the entire DNA segment of autonomous elements is excised from where it resides and is reinserted into the genome at a different location. This is accomplished via an encoded transposase. Surrounding the transposase open reading frame are 5' and 3' untranslated sequences and the terminal inverted repeats (TIRs), which are the recognition target of the transposase during mobilization. The Tigger1 transposon was mentioned in Zhao *et al.* as potentially recently active in the mammoth.

Non-autonomous elements exist within both classes. These typically short elements rely on an autonomous partner to provide the enzymatic machinery for their mobilization. Two SINEs (Short INterspersed Elements), common non-autonomous partners of LINEs, were discussed in detail by Zhao *et al.* – AfroSINE and AfroLA. These tRNA-derived SINEs are similar in structure but differ in their temporal distribution with AfroSINEs spreading throughout the genome much earlier than AfroLA. Non-autonomous DNA transposons (various MERs) were also recovered but appear to be ancient and inactive. In all cases, the arrows represent target site duplications (TSD) generated on insertion.



Acknowledgments

D. G. Peterson contributed valuable comments to earlier versions of this manuscript. Transposable element research in the Batzer Laboratory is supported by the Louisiana Board of Regents Governor's Biotechnology Initiative GBI (2002–005), and National Institutes of Health RO1 GM59290.

References

1. Callinan PA, Batzer MA. Retrotransposable elements and human disease. *Genome Dyn* 2006;1:104–115. [PubMed: 18724056]
2. Callinan PA, et al. Alu retrotransposition-mediated deletion. *Journal of Molecular Biology* 2005;348:791–800. [PubMed: 15843013]
3. Gilbert N, et al. Genomic deletions created upon LINE-1 retrotransposition. *Cell* 2002;110:315–325. [PubMed: 12176319]
4. Han K, et al. Alu recombination-mediated structural deletions in the chimpanzee genome. *PLoS Genet* 2007;3:1939–1949. [PubMed: 17953488]
5. Han K, et al. Genomic rearrangements by LINE-1 insertion-mediated deletion in the human and chimpanzee lineages. *Nucleic Acids Res* 2005;33:4040–4052. [PubMed: 16034026]
6. Huie ML, et al. A large *Alu*-mediated deletion, identified by PCR, as the molecular basis for glycogen storage disease type II (GSDII). *Human Genetics* 1999;104:94–98. [PubMed: 10071199]
7. Sen SK, et al. Human Genomic Deletions Mediated by Recombination between Alu Elements. *Am J Hum Genet* 2006;79:41–53. [PubMed: 16773564]
8. Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. *Nature Reviews Genetics*. (in press)
9. Cordaux R, et al. Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. *Proceedings of the National Academy of Sciences of the United States of America* 2006;103:8101–8106. [PubMed: 16672366]
10. Xing J, et al. Emergence of primate genes by retrotransposon-mediated sequence transduction. *Proc Natl Acad Sci U S A* 2006;103:17608–17613. [PubMed: 17101974]
11. Lowe CB, et al. Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. *Proceedings of the National Academy of Sciences of the United States of America* 2007;104:8005–8010. [PubMed: 17463089]
12. Mikkelsen TS, et al. Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* 2007;447:167–177. [PubMed: 17495919]
13. Zhao F, et al. Tracking the past: Interspersed repeats in an extinct Afrotherian mammal, *Mammuthus primigenius*. *Genome Res* 2009;19:1384–1392. [PubMed: 19508981]
14. Redi CA, et al. Genome size: a novel genomic signature in support of Afrotheria. *J Mol Evol* 2007;64:484–487. [PubMed: 17479346]
15. Miller W, et al. Sequencing the nuclear genome of the extinct woolly mammoth. *Nature* 2008;456:387–390. [PubMed: 19020620]
16. Kumar A, Bennetzen JL. Plant retrotransposons. *Annu Rev Genet* 1999;33:479–532. [PubMed: 10690416]
17. Shedlock AM, et al. Phylogenomics of nonavian reptiles and the structure of the ancestral amniote genome. *Proceedings of the National Academy of Sciences of the United States of America* 2007;104:2767–2772. [PubMed: 17307883]
18. Novick PA, et al. The Evolutionary Dynamics of Autonomous Non-LTR Retrotransposons in the Lizard *Anolis Carolinensis* Shows More Similarity to Fish Than Mammals. *Mol Biol Evol* 2009;26:1811–1822. [PubMed: 19420048]
19. Dequeant ML, Pourquie O. Chicken genome: new tools and concepts. *Dev Dyn* 2005;232:883–886. [PubMed: 15736170]
20. Wicker T, et al. The repetitive landscape of the chicken genome. *Genome Research* 2005;15:126–136. [PubMed: 15256510]
21. Furano AV, et al. L1 (LINE-1) retrotransposon diversity differs dramatically between mammals and fish. *Trends Genet* 2004;20:9–14. [PubMed: 14698614]
22. Yoder JA, et al. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet* 1997;13:335–340. [PubMed: 9260521]
23. Bird A. Does DNA methylation control transposition of selfish elements in the germline? *Trends Genet* 1997;13:469–472. [PubMed: 9433135]

24. Suzuki S, et al. Retrotransposon silencing by DNA methylation can drive mammalian genomic imprinting. *PLoS Genet* 2007;3:e55. [PubMed: 17432937]
25. Cooper DM, et al. Factors affecting ectopic gene conversion in mice. *Mamm Genome* 1998;9:355–360. [PubMed: 9545491]
26. Malik HS, et al. The age and evolution of non-LTR retrotransposable elements. *Mol Biol Evol* 1999;16:793–805. [PubMed: 10368957]
27. Gentles AJ, et al. Evolutionary dynamics of transposable elements in the short-tailed opossum *Monodelphis domestica*. *Genome Res* 2007;17:992–1004. [PubMed: 17495012]
28. Zupunski V, et al. Evolutionary dynamics and evolutionary history in the RTE clade of non-LTR retrotransposons. *Mol Biol Evol* 2001;18:1849–1863. [PubMed: 11557792]
29. Gilbert C, et al. Target site analysis of RTE1_LA and its AfroSINE partner in the elephant genome. *Gene* 2008;425:1–8. [PubMed: 18796327]
30. Kordis D, Gubensek F. Unusual horizontal transfer of a long interspersed nuclear element between distant vertebrate classes. *Proc Natl Acad Sci U S A* 1998;95:10704–10709. [PubMed: 9724768]
31. Kordis D, Gubensek F. Horizontal transfer of non-LTR retrotransposons in vertebrates. *Genetica* 1999;107:121–128. [PubMed: 10952205]
32. Pritham EJ, Feschotte C. Massive amplification of rolling-circle transposons in the lineage of the bat *Myotis lucifugus*. *Proceedings of the National Academy of Sciences of the United States of America* 2007;17:422–432.
33. Ray DA, et al. Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*. *Genome Res* 2008;18:717–728. [PubMed: 18340040]
34. Ray DA, et al. Bats with hATs: Evidence for Recent DNA Transposon Activity in Genus *Myotis*. *Molecular Biology and Evolution* 2007;24:632–639. [PubMed: 17150974]
35. Novick PA, et al. Jumping Genomes: Independent and parallel transfer of DNA transposons in tetrapod genomes. *GENE*. (in press)
36. Pace JK 2nd, et al. Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. *Proc Natl Acad Sci U S A* 2008;105:17023–17028. [PubMed: 18936483]
37. Pace JK 2nd, Feschotte C. The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. *Genome Research* 2007;17:422–432. [PubMed: 17339369]
38. Waterston RH, et al. Initial sequencing and comparative analysis of the mouse genome. *Nature* 2002;420:520–562. [PubMed: 12466850]
39. Gibbs RA, et al. Evolutionary and biomedical insights from the rhesus macaque genome. *Science* 2007;316:222–234. [PubMed: 17431167]
40. Gibbs RA, et al. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 2004;428:493–521. [PubMed: 15057822]
41. Lander ES, et al. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860–921. [PubMed: 11237011]
42. Murphy WJ, et al. Using genomic data to unravel the root of the placental mammal phylogeny. *Genome Res* 2007;17:413–421. [PubMed: 17322288]
43. Prasad AB, et al. Confirming the phylogeny of mammals by use of large comparative sequence data sets. *Mol Biol Evol* 2008;25:1795–1808. [PubMed: 18453548]
44. Swaminathan K, et al. Global repeat discovery and estimation of genomic copy number in a large, complex genome using a high-throughput 454 sequence survey. *BMC Genomics* 2007;8:132. [PubMed: 17524145]
45. Macas J, et al. Repetitive DNA in the pea (*Pisum sativum* L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*. *BMC Genomics* 2007;8:427. [PubMed: 18031571]
46. Cantrell MA, et al. Loss of LINE-1 activity in the megabats. *Genetics* 2008;178:393–404. [PubMed: 18202382]
47. Luan DD, et al. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* 1993;72:595–605. [PubMed: 7679954]
48. Kriegs JO, et al. Retroposed elements as archives for the evolutionary history of placental mammals. *PLoS Biol* 2006;4:e91. [PubMed: 16515367]

49. Seiffert ER. A new estimate of afrotherian phylogeny based on simultaneous analysis of genomic, morphological, and fossil evidence. *BMC Evol Biol* 2007;7:224. [PubMed: 17999766]

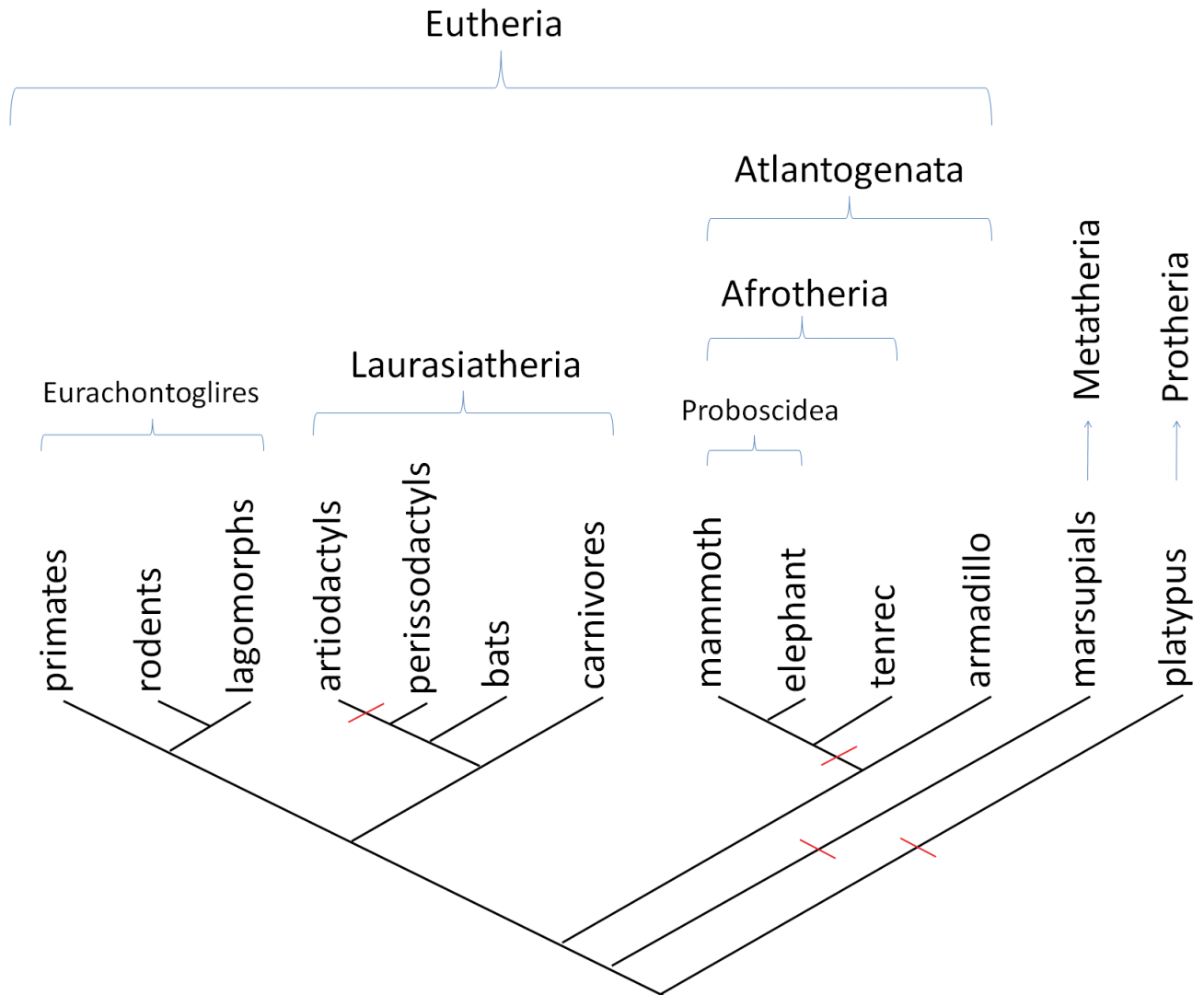


Figure 1.

Phylogenetic relationships of selected mammalian clades with the hypothesized position of mammoth included. The phylogeny reveals the basal position of Afrotheria (and the mammoth) within Mammalia, specifically Eutheria. With morphologies as diverse as that seen in the fossorial, golden mole to the slow swimming manatee, many studies have had difficulty deducing interordinal relationships within Afrotheria due to rapid evolution and cladogenesis. Studies incorporating TEs as phylogenetic markers [42,48] within Afrotheria have augmented large-scale DNA sequence studies [49] to better understand this rapid diversification. An understanding of the unique TE dynamics within this group might eventually enable a deeper understanding of the unique evolutionary pressures that allowed it to diversify so successfully. The red lines indicate the likely points of invasion of RTE elements based on their taxonomic distribution.