bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Phylogenomics reveals an almost perfect polytomy among the almost ungulates (*Paenungulata*)

Jacob Bowman[1], David Enard[2*], and Vincent J. Lynch[1*]

[1] Department of Biological Sciences, University at Buffalo, SUNY, 551 Cooke Hall, Buffalo, NY, USA.

[2] Department of Ecology and Evolutionary Biology. University of Arizona, Tucson, AZ, USA[* Equal contribution

*Correspondence: vjlynch@buffalo.edu, denard@arizona.edu

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

Phylogenetic studies have resolved most relationships among *Eutherian* Orders. However, the branching order of elephants (*Proboscidea*), hyraxes (*Hyracoidea*), and sea cows (*Sirenia*) (i.e., the *Paenungulata*) has remained uncertain since at least 1758, when Linnaeus grouped elephants and manatees into a single Order (*Bruta*) to the exclusion of hyraxes. Subsequent morphological, molecular, and large-scale phylogenomic datasets have reached conflicting conclusions on the branching order within *Paenungulates*. We use a phylogenomic dataset of alignments from 13,388 protein-coding genes across 261 *Eutherian* mammals to infer phylogenetic relationships within *Paenungulates*. We find that gene trees almost equally support the three alternative resolutions of *Paenungulate* relationships and that despite strong support for a *Proboscidea*+*Hyracoidea* split in the multispecies coalescent (MSC) tree, there is significant evidence for gene tree uncertainty, incomplete lineage sorting, and introgression among *Proboscidea*, *Hyracoidea*, and *Sirenia*. Indeed, only 8-10% of genes have statistically significant phylogenetic signal to reject the hypothesis of a *Paenungulate* polytomy. These data indicate little support for any resolution for the branching order *Proboscidea*, *Hyracoidea*, and *Sirenia* within *Paenungulata* and suggest that *Paenungulata* may be as close to a real, or at least unresolvable, polytomy as possible.

Introduction

Systematic studies have successfully resolved most of the phylogenetic relationships within *Eutherian*[1] mammals. Some clades, however, have remained recalcitrant to phylogenetic resolution, including the interrelationships between elephants (*Proboscidea*), hyraxes (*Hyracoidea*), and manatees (*Sirenia*). While these lineages were combined into the superorder *Paenungulata*, the "almost ungulates", by Simpson (1945), who also noted the possibility that *Hyracoidea* might be more closely related to the ungulate order *Perissodactyla*, the recognition that *Proboscideans* and *Sirenians* are closely related to each other is much older (**Figure 1A**). In *Systema Naturae* (1758), for example, Linnaeus classified elephants and sirenians together under the name *Bruta* (Linné, 1758); this group also included sloths, anteaters, and pangolins,

[1] We follow the formatting suggestion of (Thines et al., 2020) and placed scientific names at all taxonomic ranks in italics to facilitate their easy recognition.

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

anticipating the close phylogenetic relationships of *Paenungulates* and *Xenarthrans*. Similarly, de Blainville (1836) classified elephants and manatees together in the group "*les gravigrades*", but he did not formalize this proposal (Blainville, 1839).

The clade uniting *Proboscideans* and *Sirenians* to the exclusion of other taxa was officially proposed and named *Tethytheria* by McKenna (1975), who, like Simpson also supported a close relationship between *Hyracoidea* and *Perissodactyla* (Domning et al., 1986; McKenna, 1975); Although McKenna (1975) does not explicitly describe the etymology of the name *Tethytheria*, it is likely named after the Tethys sea around which these lineages originated (Domning et al., 1986; McKenna, 1975; Prothero and Williams, 2017). Subsequently, Novacek (1982) proposed that *Hyracoidea* was a sibling lineage to the *Tethytheria* within *Paenungulata* (Novacek, 1982). The story of *Paenungulate* systematics, however, is much more complicated than this straightforward narrative suggests (J. H. Shoshani, 1986). Indeed, *Proboscidea*, *Hyracoidea*, and *Sirenia* have been considered closely related to numerous other mammalian orders, most often *Tethytheria* to *Pinnepedia* or *Cetacea*, and *Hyracoidea* to *Perissodactyla*. *Proboscidea*, *Hyracoidea*, and *Sirenia* were even considered to have close affinities to each other by several 19th and early 20th-century morphologists (**Figure 1A**).

Molecular phylogenetic studies have almost uniformly agreed that *Proboscidea*, *Hyracoidea*, and *Sirenia* formed a monophyletic group within *Eutheria*. Still, they disagree on how these lineages are related to each other[2]. Indeed, while there are only three possible resolutions to how *Proboscidea*, *Hyracoidea*, and *Sirenia* can be related to each other (**Figure 1B**), each has received substantial support (**Figure 1C** and **Supplementary Table 1**). Even studies based on large multigene, phylogenomic, and rare genomic events such as chromosome rearrangements, transposable element insertions, gene duplications, and gene losses, infer conflicting relationships within *Paenungulates* (**Supplementary Table 1**). Therefore, we assembled a large dataset of alignments from 13,388 protein-coding genes across 261 *Eutherian* mammals, including an African elephant, rock and tree hyrax, and manatee, and used several complementary methods to infer their phylogenetic relationships. We found that gene trees were almost evenly divided among the three alternative *Paenungulate* resolutions and that relatively

[2] To our knowledge, no in/formal names have been proposed for the alternative *Proboscidea+Hyracoidea* or *Sirenia+Hyracoidea* resolutions of *Paenungulata.* Therefore, we refer to the *Proboscidea+Hyracoidea* clade as *Paratethytheria*, after the Paratethys sea, which was repeatedly disconnected from and reconnected with the Tethys sea, and the *Sirenia+Hyracoidea* clade as *Peritethytheria*, to also reflect the origin of these lineage around the Tethys sea.

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

few genes had enough statistical support to favor one resolution over another. While the species tree favors the *Paratethytheria* tree, there is significant phylogenetic discordance among genes and sites resulting from incomplete lineage sorting, introgression, and gene tree uncertainty. Furthermore, tree topology tests cannot reject a *Paenungulate* polytomy for most genes (>90%). Thus, there is little support for any resolution for the branching order of *Proboscidea*, *Hyracoidea*, and *Sirenia,* and *Paenungulata* is as close to a real polytomy as possible.

Methods

Assembling alignments of orthologous protein-coding genes from 261 Eutherians

The alignment pipeline starts with Ensembl v99 human coding sequences, using the longest isoforms for each gene. We then search for the orthologs of these human coding sequences within the genome assemblies of 260 other mammals (**Supplementary Table 2**) with a contig size greater than 30kb in the NCBI assembly database as of July 2020. We select a minimum 30kb contig size to avoid excessive truncated orthologous coding sequences. To extract the orthologous coding sequences (CDS) to the human CDS, we use the best Blat reciprocal hits from the human CDS to each other mammalian genome and back to the human genome. We use Blat matching all possible reading frames, with a minimum identity set at 30% and the “fine” option activated (Kent, 2002). We excluded genes with less than 251 best reciprocal hits out of the 261 (human+other mammals) species included in the analysis. We find 13,491 human CDS genes with the best reciprocal hits orthologs in at least 250 other mammals; for reference, this corresponds to 68% of all protein-coding genes in the human genome.

We aligned each orthologous gene with Macse v2 (Ranwez et al., 2018). Macse v2 is a codon-aware multiple sequence aligner that can explicitly identify frameshifts and readjust reading frames accordingly. This feature of Macse is particularly important because erroneous indels introduced in coding sequences during genome sequencing and assembly processes can be common and cause frameshifts that many aligners do not consider. These sequencing and alignment errors can result in substantial misalignments of coding sequences due to incomplete codons. We start with the human CDS because they are likely the highest quality regarding sequencing and annotations. We use Macse v2 with maximum accuracy settings.

The alignments generated by Macse v2 were edited by HMMcleaner with default parameters (Franco et al., 2019). HMMcleaner is designed to remove, in a species-specific

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

fashion, “fake” substitutions that are likely genome sequencing errors. HMMcleaner also removes “false exons” that might have been introduced during the Blat search. “Fake exons” are intronic segments that, by chance, are similar to exons missing from an assembly due to a sequencing gap. When looking for the most similar non-human CDS using Blat, Blat can sometimes “replace” the missing exon with a similar intronic segment.

After using HMMcleaner to remove suspicious parts of the CDS alignments in each species, we only select those still complete codons to remain in the alignments. In alignments of coding sequences, the flanks of indels usually include a higher number of misaligned substitutions. In each separate species, we further remove the x upstream or downstream codons from the alignments if more than x/2 of these codons code for amino acids that are different from the consensus amino acid in the whole alignment, with x varying from 1 to 20. For example, if eight of the 12 amino acids on the left side of an indel are different from the consensus in a given species, we remove the 12 corresponding codons. If two of the three amino acids on the right side of an indel are different from the consensus in a given species, we remove the three corresponding codons.

Phylogenomic Analyses

We use IQTREE2 v.2.2.2 COVID-edition (Nguyen et al., 2015) to infer maximum likelihood phylogenetic gene trees from all 13,491 genes (nucleotide sequences) after the best-fitting model was identified for each gene by ModelFinder v.1.42 (Kalyaanamoorthy et al., 2017) with the -m MFP option; note that tree searches failed for 103 genes, thus our final dataset of trees includes 13,388 gene trees. Branch supports were assessed with the Shimodaira-Hasegawa-like approximate likelihood ratio (SH-like aLRT) test with 1000 replicates (Anisimova et al., 2006; Guindon et al., 2010) with the -alrt 1000 option. We also used IQTREE2 to perform tree topology tests with the RELL approximation (Kishino et al., 1990), including the bootstrap proportion, the Kishino-Hasegawa (Kishino and Hasegawa, 1989) and Shimodaira-Hasegawa tests (Shimodaira and Hasegawa, 1999), the weighted Kishino-Hasegawa and Shimodaira-Hasegawa tests, expected likelihood weights (Strimmer and Rambaut, 2002), and the approximately unbiased test (Shimodaira, 2002) with the -zb 10000 -au -zw options; a tree with *Paenungulates* as a polytomy was compared to the three alternate resolutions of this clade (**Figure 3A**). The 13,388 gene trees were used to infer a species tree with ASTRAL-III v.5.6.3 (Zhang et al., 2018); support for the ASTRAL species tree was also assessed with the gene-tree bootstrap approach (Simmons et al.,

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

2019), which generated 100 pseudoreplicate datasets of the 13,388 gene trees followed by ASTRAL species tree inference on each of the 100 datasets using the *msc_tree_resampling.pl* script (Simmons et al., 2019). Gene and site concordance factors (Minh et al., 2020) were inferred with IQTREE2 v.2.2.2 COVID-edition, using the updated maximum likelihood-based method (--scfl option) for site concordance factors (Mo et al., 2022), and the ASTRAL species tree.

Characterizing incomplete lineage sorting and introgression

We used two methods to detect patterns of introgression between *Paenungualte* lineages based on the distributions of gene tree topologies and branch lengths for triplets of lineages. If the species tree is ((A, B), C), or the *Paratethytheria* resolution inferred with the ASTRAL species tree, these tests can detect introgression between A (*Probosceidea*) and C (*Sirenia*), and between B (*Hyracoidea*) and C (*Sirenia*). The discordant-count test (DCT) compares the number of genes supporting each of the two possible discordant gene trees ((A, C), B) and (A, (B, C)); in the absence of ancestral population structure, gene genealogies from loci experiencing ILS will show either topology with equal probability and ILS alone is not expected to bias the count towards one of the topologies (Huson et al., 2005; Suvorov et al., 2022). Introgression, however, will lead to a statistically significant difference in the number of gene trees, which can be evaluated with a X^2–test (Lanfear, 2018; Suvorov et al., 2022); if there is introgression between A and C, there will be an excess of gene trees with the ((A, C), B) topology (Suvorov et al., 2022). The branch-length test (BLT) examines branch lengths to estimate the age of the most recent coalescence event (measured in substitutions per site); introgression leads to more recent coalescences than expected under the species tree topology with complete lineage sorting, while ILS shows older coalescence events (Green et al., 2010; Suvorov et al., 2022).

ILS alone does not result in different coalescence times between the two discordant topologies, and this forms the null hypothesis for the BLT that the distribution of branch lengths of gene trees supporting the ((A, C), B) and (A, (B, C)) topologies should be similar (Suvorov et al., 2022). In the presence of introgression, these branch length distributions will be skewed such that ((A, C), B) < (A, (B, C)) suggests introgression consistent with discordant topology ((A, C), B) and ((A, C), B) > (A, (B, C)) suggests introgression consistent with discordant topology (A, (B, C)). We implemented two versions of this test, one that does not scale branch lengths by total tree length and one that does (Suvorov et al., 2022); for the former, we tested the statistical significance of differences between the distribution of branch lengths with a Kruskal-Wallis one-

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

way ANOVA while for the latter used a Mann-Whitney U test (Suvorov et al., 2022). *P*-values were corrected for multiple testing within the DCT and BLT with Holm's method (Holm, 1979). For the scaled BLT and DCT with all trios within *Paeunungulates,* we used the *blt_dct_test.r* script (Suvorov et al., 2022).

Results and Discussion

Gene trees

We assembled a dataset of 13,491 orthologous protein-coding gene alignments from the genomes of 261 *Eutherian* mammals (**Supplementary Table 2**), for which we used IQTREE2 to infer the best-fitting nucleotide substtition model and maximum likelihood tree; we thus inferred 13,388 gene trees (103 tree searches failed). We found that 80% of ML gene (nucleotide) trees inferred that African elephant (*Loxodonta africana*), hyraxes (*Procavia* and *Heterohyrax*), and manatee (*Trichechus manatus*) formed a monophyletic clade (*Paenungulata*), similar to nearly all other phylogenetic analyses of *Eutherian* mammals. Relationships within *Paenungulata*, however, were almost evenly divided between the three alternative tree topologies, with the *Paratethytheria* split receiving the most support (**Figure 2A**). Branch supports assessed with the SH-like aLRT test (Anisimova et al., 2006; Guindon et al., 2010) indicate that only 252 of 13,388 genes (1.9%) had SH-like aLRT support values ≥95%, which corresponds to an 5% significance level (Anisimova et al., 2011), and these genes favor the *Tethytheria* resolution (**Figure 2A**). As expected, Pearson's correlation tests revealed that SH-like aLRT support and branch length (substitutions per site) were positively correlated, such that longer branches tended to have higher SH-like aLRT support values (**Figure 2B**). Thus, while *Paratethytheria* received the greatest support among all genes, genes with significant phylogenetic signal (SH-like aLRT≥95%) support the *Tethytheria* resolution.

Multispecies coalescent tree, gene and site concordance

Next, we used the 13,388 gene trees to infer a species tree with ASTRAL-III, which is statistically consistent under multispecies coalescent model (MSC) and is useful for inferring a species tree when there is significant incomplete lineage sorting. While the ASTRAL species tree supported the *Paratethytheria* split with a posterior probability of 1.00 and a gene-tree bootstrap support of 98.7% (**Figure 3A**), there can be significant underlying gene tree discordance even for branches with near 100% support (Jarvis et al., 2014; Pease et al., 2016; Salichos and Rokas,

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

2013; Vanderpool et al., 2020). For example, the internode certainty (IC), which corresponds to the magnitude of conflict between the two most likely splits among gene trees (Salichos et al., 2014) for the *Paratethytheria* split is $6.49x10^{-5}$ (**Figure 3A**); IC values near 1 indicate the absence of conflict for the split, whereas IC values near 0 indicate equal support for each split and hence maximum conflict (Salichos et al., 2014). Note that the maximum IC, corresponding to one more tree supporting *Paratethytheria* over *Tethytheria* is $3.35x10^{-8}$. Similarly, the internode certainty all (ICA), which corresponds to the magnitude of conflict between all splits among gene trees (Salichos et al., 2014) is 0.02 (**Figure 3A**); ICA values at or near 1 indicate the absence of any conflict among splits, whereas ICA values near 0 indicate that one or more conflicting splits have almost equal support (Salichos et al., 2014). Thus, while there is strong support for the *Paratethytheria* resolution in the MSC species tree, there is also significant gene tree discordance for this split.

We next used IQTREE2 (Nguyen et al., 2015) to infer gene (gCF) and site concordance (sCF) factors for each branch in the ASTRAL species tree; gCF and sCF are the fraction of genes and sites that are in agreement with the species tree for any particular branch (Minh et al., 2020; Mo et al., 2022). Consistent with significant phylogenetic uncertainty among gene trees, the gCF was only 34.97% for the *Paratethytheria* split, whereas the discordance factor (the proportion of trees supporting alternative resolutions) was 34.31% for the *Tethytheria* split and 20.94% for the *Peritethytheria* split; 9.77% of gene trees were discordant due to polyphyly (**Figure 3B**). Similarly, quartet support for the *Paratethytheria* was 0.39, 0.37 for the *Tethytheria* split, and 0.24 for the *Peritethytheria* split (**Figure 3B**). Site concordance factors inferred for each gene using the ASTRAL species tree were also very low, with a median of 32% and a mean of 33%, indicating a high rate of discordance among sites (**Figure 3C**). Thus while the species tree strongly supports *Paratethytheria*, concordance factors suggest that most genes and sites support a different relationship. Indeed, there is almost perfect discordance between the *Paratethytheria* and *Tethytheria* splits.

Discordance from incomplete lineage sorting and introgression

If the discordance among gene trees results from incomplete lineage sorting (ILS), the number of gene trees supporting the two discordant topologies should be roughly equal (Huson et al., 2005). In contrast, introgression, or a combination of ILS and introgression, will cause one of the two discordant topologies to be more frequent (Suvorov et al., 2022); the difference between

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

the number of gene trees supporting alternative resolutions can be compared with a X^2–test (Lanfear, 2018; Suvorov et al., 2022), and is called the discordant-count test (DCT)[3]. We found that the number of gene trees supporting *Paratethytheria* and *Tethytheria* was significantly greater than the number supporting *Peritethytheria*, but there was not a significant difference between *Paratethytheria* and *Tethytheria* (**Figure 3D**). These data indicate an excess of gene trees supporting *Tethytheria*, consistent with introgression from *Sirenia* into *Proboscidea* (**Figure 3F**).

ILS alone will not result in different coalescence times between the two discordant trees, while introgression will result in more recent coalescences than expected under the species tree topology with complete lineage sorting (Suvorov et al., 2022); the branch length test (BLT) compares the distribution of branch lengths for each gene tree with discordant topologies, with the expectation that they are similar under ILS alone (Suvorov et al., 2022). The distribution of branch lengths between gene trees supporting *Paratethytheria* and *Tethytheria* was significantly different than *Peritethytheria*, as was the difference between *Paratethytheria* and *Tethytheria* (**Figure 3E**). We also used the scaled BLT of Suvorov et al. (2022), as well as their approach of testing trios of lineages for the DCT, rather than our approach of testing elephant, manatee, and the *Procavia-Heterohyrax* stem-lineage, and found similar results (**Supplementary Table 3**). These data are consistent with introgression from *Sirenia* into *Hyracoidea* (**Figure 3F**)

Other sources of discordance

Technical noise including base composition variation, gene length, low phylogenetic signal, and model misspecification, among many others, also contributes to gene tree discordance. Longer genes tend to have more phylogenetically informative sites than shorter genes, but while there were statistically significant differences in the length of genes that support each of the three splits the effect sizes were small (**Supplementary Figure 1A**). Genes with high GC content can have poor phylogenetic resolution because they can have many homoplasies at hypermutable CpG sites and more frequent biased gene conversion (Romiguier et al., 2013). However, GC-content was similar between genes supporting the three splits (**Supplementary**

[3] As initially proposed (Huson et al., 2005), the significance of the difference in number between the two discordant trees can be compared with the Azuma-Hoeffding inequality (Alon and Spencer, 2023). Using this method, the number of gene trees supporting *Paratethytheria* (Δ=1878, $P_{Holm\text{-}adj.}=4.88\text{x}10^{-26}$) and *Tethytheria* ($\Delta$=1790, $P_{Holm\text{-}adj.}=4.05\text{x}10^{-23}$) was significantly greater than the number supporting *Peritethytheria*, but there was not a significant difference between *Paratethytheria* and *Tethytheria* (Δ=88, $P_{Holm\text{-}adj.}$=0.20).

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 1A). Indeed, a tree inferred from the third of genes with the lowest GC content (n=4,444) concatenated into a supermatrix, with separate a GTR6 model for each gene partition, supported the *Paratethyeria* split (**Supplementary Figure 2A**), as did consensus trees when genes were binned into quartiles based on GC content (**Supplementary Figure 2B**). These data indicate that our results are robust to noise from GC content variation across genes. Genes supporting the *Paratethytheria* split had statistically higher SH-like aLRT scores, sCFs, and longer branch lengths than genes supporting the other splits, again, however, the effect sizes were small (**Supplementary Figure 1A**). There also were no meaningful correlations between sCF (calculated from the ASTRAL species tree with the *Paratethytheria* split) and gene length, GC-content, SH-like aLRT score, or branch length, or between SH-like aLRT score and gene length, GC-content, or branch length (**Supplementary Figure 1B**). Thus, while each of these factors may contribute to gene tree discordance, their effects are likely small (but they may also be additive).

Tree topology tests

Few studies have explicitly tested whether gene trees support alternative resolutions of *Paenungulates* (Liu et al., 2023). Therefore, we used IQTREE2 to perform several tree topology tests, including the Kishino-Hasegawa test (Kishino and Hasegawa, 1989), the Shimodaira-Hasegawa tests (Shimodaira and Hasegawa, 1999), the weighted Kishino-Hasegawa and Shimodaira-Hasegawa tests, the approximately unbiased (AU) test (Shimodaira, 2002), the expected likelihood weights (Strimmer and Rambaut, 2002), and the bootstrap proportion (Kishino et al., 1990). A tree with *Paenungulates* as a polytomy was compared to the three alternate resolutions of this clade (**Figure 4A**); therefore, these methods test if there is an improvement in the likelihood score of the three alternative topologies compared to the polytomy tree (Shen et al., 2017). We found that the likelihood differences (deltaL) between the polytomy and three alternative trees were generally very small for most genes (**Figure 4B**) and that the suite of tree topology tests agreed the worst tree was *Peritethytheria* and the best tree was the *Paratethytheria* (**Figure 4C**). However, only 8-10% of genes statistically rejected the polytomy tree (**Figure 4C**). Thus, very few genes have statistically significant phylogenetic signal to reject the *Paenungulate* polytomy tree.

Caveats and limitations

Inferring reliable gene and species trees from phylogenomic data, and thus sorting among competing scenarios of gene tree inference error, incomplete lineage sorting, and introgression

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(Hibbins and Hahn, 2021) is fraught with *ad hoc* decisions, methodological limitations, and technical challenges. For example, we have focused on inferring phylogenetic relationships from protein-coding genes, for which we could generate high-quality alignments for 261 *Eutherian* mammals. While our dataset is extensive, including 13,388 genes with a total alignment length of 19,968,803 bp, the coding fraction of the genome is relatively small and may have less or different phylogenetic signal than non-coding regions (Foley et al., 2015, 2023; Literman and Schwartz, 2021; William J Murphy et al., 2001). Despite this potential limitation, our observation that a slight majority of gene trees (n=88) support the *Paratethytheria* split is similar to a recent study based on 411 kb of genome-wide nearly neutral sites and non-coding sites (Foley et al., 2023); however, our results differ in supporting *Tethetheria* rather than *Peritehytheria* as the second most likely split. These data suggest that systematic differences in phylogenetic information content between coding and non-coding (or genic and intra/inter-genic) regions are unlikely to bias our results, but it remains a possibility.

Other technical sources of error, such as misalignment, model misspecification, and long branch attraction (LBA), could also bias our gene tree inferences and downstream analyses. For example, alignment error leads to statistical inconsistency of phylogenetic methods, manifesting in nucleotide alignments as gross substitution model misspecification and LBA artifacts (Hossain et al., 2015; Ogden and Rosenberg, 2006). We attempted to mitigate the effects of alignment error through stringent alignment criteria; however, assessing the impact of such error (Landan and Graur, 2008; Sela et al., 2015) in phylogenomic scale datasets is computationally impractical. Similarly, we inferred gene trees with the best substitution model for each gene to minimize the effects of model misspecification. Still, computational limitations prevented us from using more complex models that account for model variation within genes, such as mixture models (Lartillot and Philippe, 2004; Le et al., 2012; Quang et al., 2008; Wang et al., 2008) and site-specific frequency models (Quang et al., 2008; Wang et al., 2018), and processes like heterotachy, i.e., rate variation across sites and lineages (Crotty et al., 2019; Kolaczkowski and Thornton, 2008, 2004), which may be common in coding regions. However, whether model selection is an essential step in phylogenetic inference is debated (Abadi et al., 2019; Doko and Liu, 2023; Gerth, 2019; Wu et al., 2014), and likelihood methods may be robust to all but the most extreme model violation (Naser-Khdour et al., 2021). Indeed, the wrong model may infer the correct tree more often than the true model (Bruno and Halpern, 1999; Yang, 1997). Thus, we do not believe that model misspecification negatively impact our results. Still, we cannot exclude the possibility that

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

extreme convergence following the very rapid diversification of *Paenungulates* has led to long-branch attraction artifacts or that heterotachy has affected gene tree inference for some genes. For example, in a previous study on another rapid radiation event, the branching order of the Hox cluster duplications, we found that genes with internal branch lengths < 0.02 were significantly impacted by LBA artifacts (Lynch and Wagner, 2009); only 5.4% of genes had *Paratethytheria*, *Tethetheria*, and *Peritehytheria* stem-lineages longer than 0.02, and the majority of these (50.07%) support the *Tethytheria* split, thus LBA artifiacts are a possible source of gene tree error. Given that the *Paratethyeria* resolution is supported by only 88 genes more than the *Tethytheria* resolution, incorrect tree inference in even a few genes could adversely affect our results.

Conclusions

Resolving the phylogenetic relationships within *Paenungulates* has been remarkably challenging, with different studies and data supporting different resolutions. Our data indicate that 92.7% protein-coding genes lack significant phylogenetic signal to reject a *Paenungulate* polytomy and that the branching order of *Paenungulates* is obscured by large-scale gene tree error, incomplete lineage sorting, and introgression. This combination of factors suggests that previous molecular phylogenetic studies, even relatively large multigene studies, were biased by stochastic gene sampling, contributing to the many conflicting resolutions of relationships within *Paenungulates*. In contrast, our study includes 13,388 protein-coding genes, representing ~67% of the coding genome and nearly 20 million base pairs, and is unlikely to be biased by gene sampling artifacts. Thus, we conclude that the combination of systematic biases in gene tree estimation caused by very rapid divergence leading to low phylogenetic information content and significant gene tree uncertainty, high levels of ILS, and introgression combine to create a perfect storm scenario (Cai et al., 2020) for inferring the branching order within *Paenungulates*. Therefore, this clade is likely as close to a real, or at least likely unresolvable, polytomy as possible.

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Acknowledgments

VJL is partially supported by NIH award 1R56AG071860-01. Computational support was provided by the Center for Computational Research at the University at Buffalo.

Data availability

Alignments (https://doi.org/10.5061/dryad.5dv41nsbc) as fasta files and IQTREE result output files for each gene (https://doi.org/10.5061/dryad.j0zpc86n3) are available from Dryad.

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

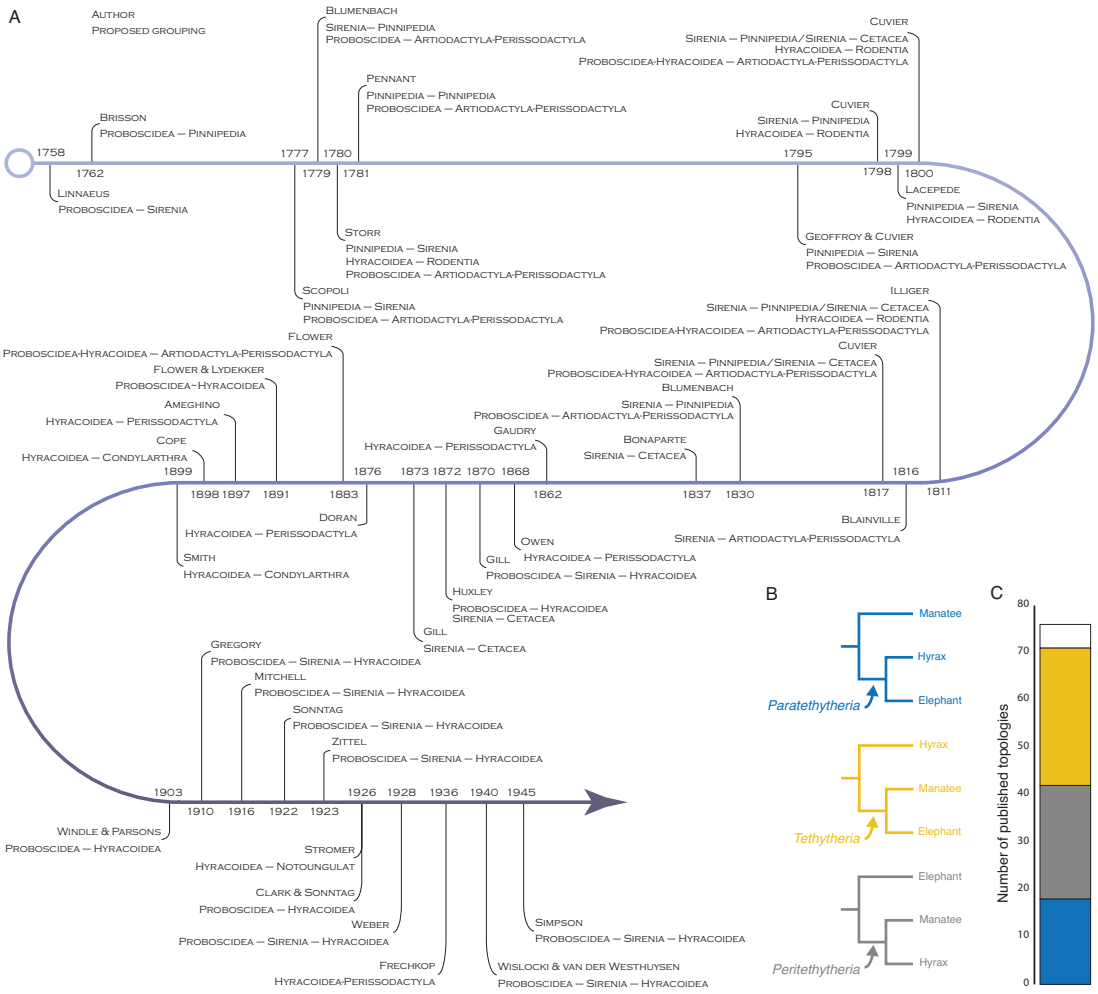


Figure 1. Historical and recent hypothesis of phylogenetic relationships among *Proboscidea*, *Hyracoidea*, *Sirenia*.

A. Timeline (Linnaeus to Simpson) showing proposed relationships between *Proboscidea*, *Hyracoidea*, *Sirenia,* and other mammalian Orders. Modified from (J. H. Shoshani, 1986) Table 1.

B. The three alternative resolutions of *Paenungulata.* We refer to the *Proboscidea+Hyracoidea* clade as *Paratethytheria*, the *Sirenia+Hyracoidea* clade as *Peritethytheria*, and the *Proboscidea+Sirenia* clade as *Tethytheria*.

C. Stacked bar chart showing the number of published molecular phylogenies supporting the three alternative resolutions of *Paenungulata* shown in panel B. Bar colors follow the coloring in panel B. White represents published studies supporting an unresolved relationship.

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

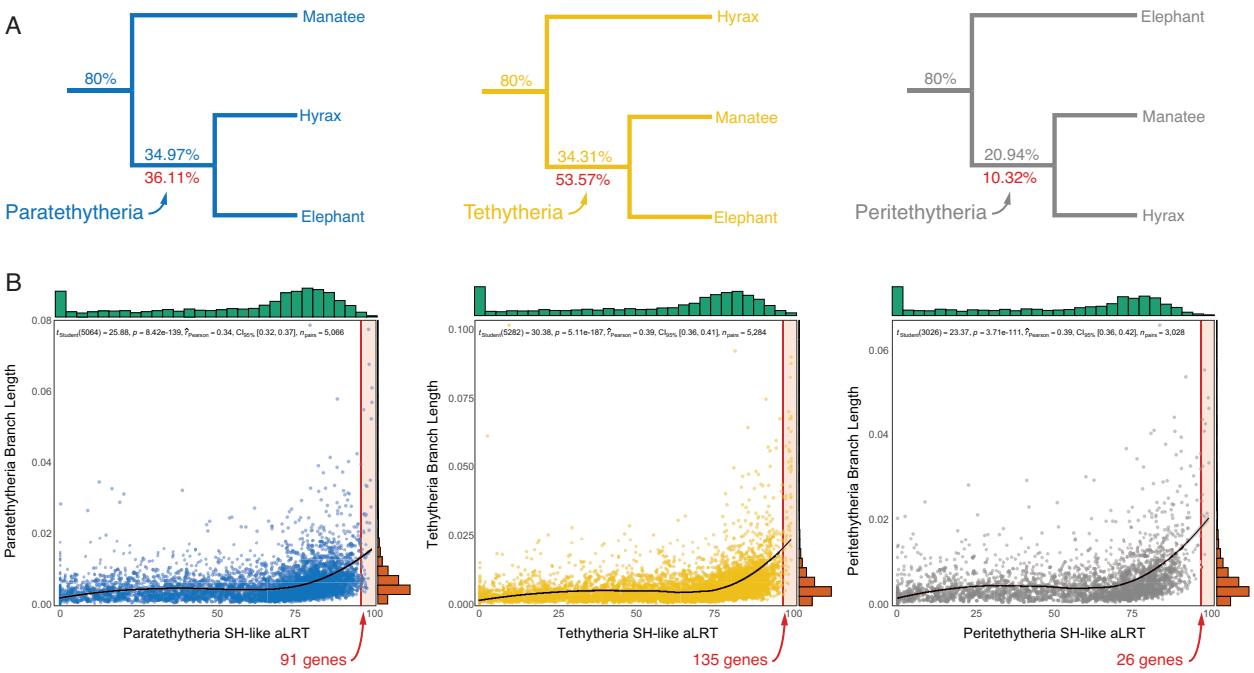


Figure 2. Gene trees are almost evenly divided between the three alternative resolutions of *Paenungulata*.

A. The three alternative resolutions of *Paenungulata*, the percent of genes supporting each resolution are shown above each branch. The percent of genes with SH-like aLRT values ≥ 95% supporting each split is shown below branches in red.

B. Scatterplot showing the correlation between branch length (substitutions per site) and SH-like aLRT support values for each gene supporting *Paratethytheria*, *Tethytheria*, and *Peritethytheria*. Loess regression is a black line (95% confidence interval around repression is shown in gray). The distribution of branch lengths and aLRT values are shown along the top (teal) and side (orange) of the plot, respectively. Vertical red lines indicate SH-like aLRT ≥ 95%, and the number of genes supporting *Paratethytheria*, *Tethytheria*, and *Peritethytheria* among genes with SH-like aLRT ≥ 95% is shown in red.

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

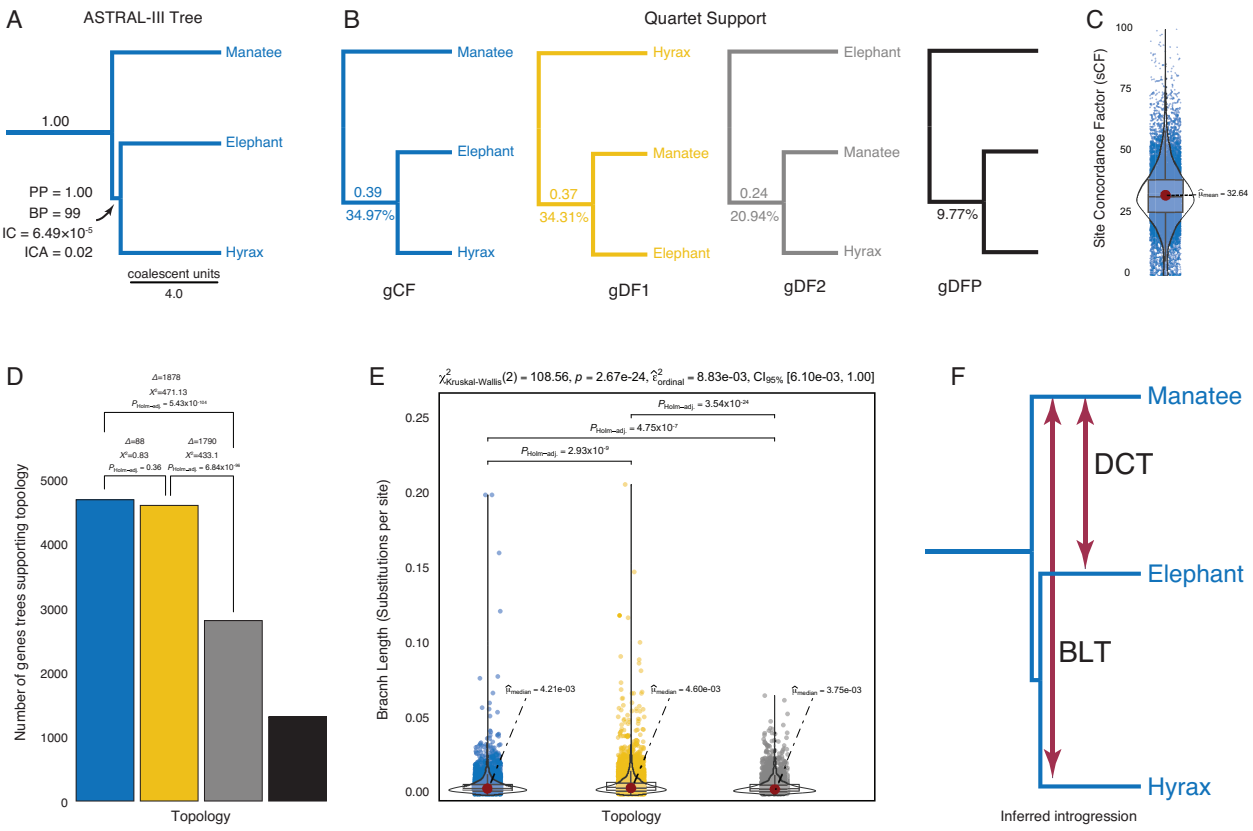


Figure 3. Strong support for the Paratethytheria species tree, but significant gene tree discordance and introgression.

A. ASTRAL-III species tree, with posterior probability (PP), bootstrap proportion (BP), internode certainty (IC), and IC all (ICA) support values for the *Paratethytheria* split. Internal branch lengths in the ASTRAL species tree are shown in coalescent units (i.e., 2*N* generations), terminal branch lengths are arbitrary.

B. Gene concordance and discordance. Gene concordance factor (gCF) for the *Paratethyeria* split, and gene discordance factors (gDF1 and gDF2) for the *Tethytheria*, and *Peritethytheria* splits are shown below each branch and quartet support for each split above. gDFP, gene discordance factor for a tree with a polyphyletic *Paenungulata*.

C. Distribution of site concordance factors among the 13,388 genes inferred using the ASTRAL species tree.

D. Number of gene trees supporting the *Paratethytheria*, *Tethytheria*, and *Peritethytheria* splits. The differences in the number of gene trees supporting each split (Δ), χ^2 values,

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

and Holm-adjusted χ^2 *P*-values for each discordant topology test (DCT) are shown each bar.

E. The distribution of branch lengths in gene trees supporting the *Paratethytheria*, *Tethytheria*, and *Peritethytheria* splits. Statistical significance of the difference in median branch lengths (substitutions per site) from a Kruskal-Wallis one-way ANOVA is shown above each comparison as Holm-adjusted *P*-values for the branch length test (BLT). Descriptive statistics are shown above the plot.

F. The DCT indicates significant introgression between manatee and elephant lineages, while the BLT indicates significant introgression between the manatee and hyrax lineages.

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

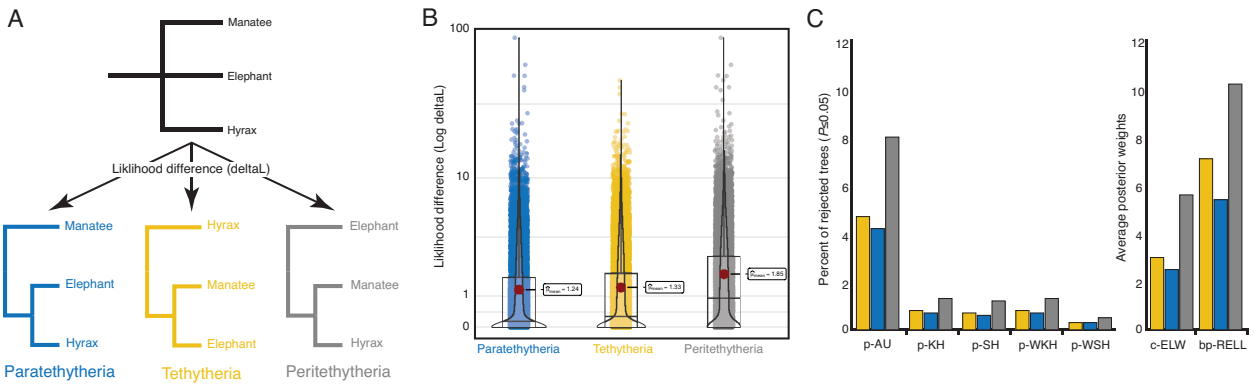


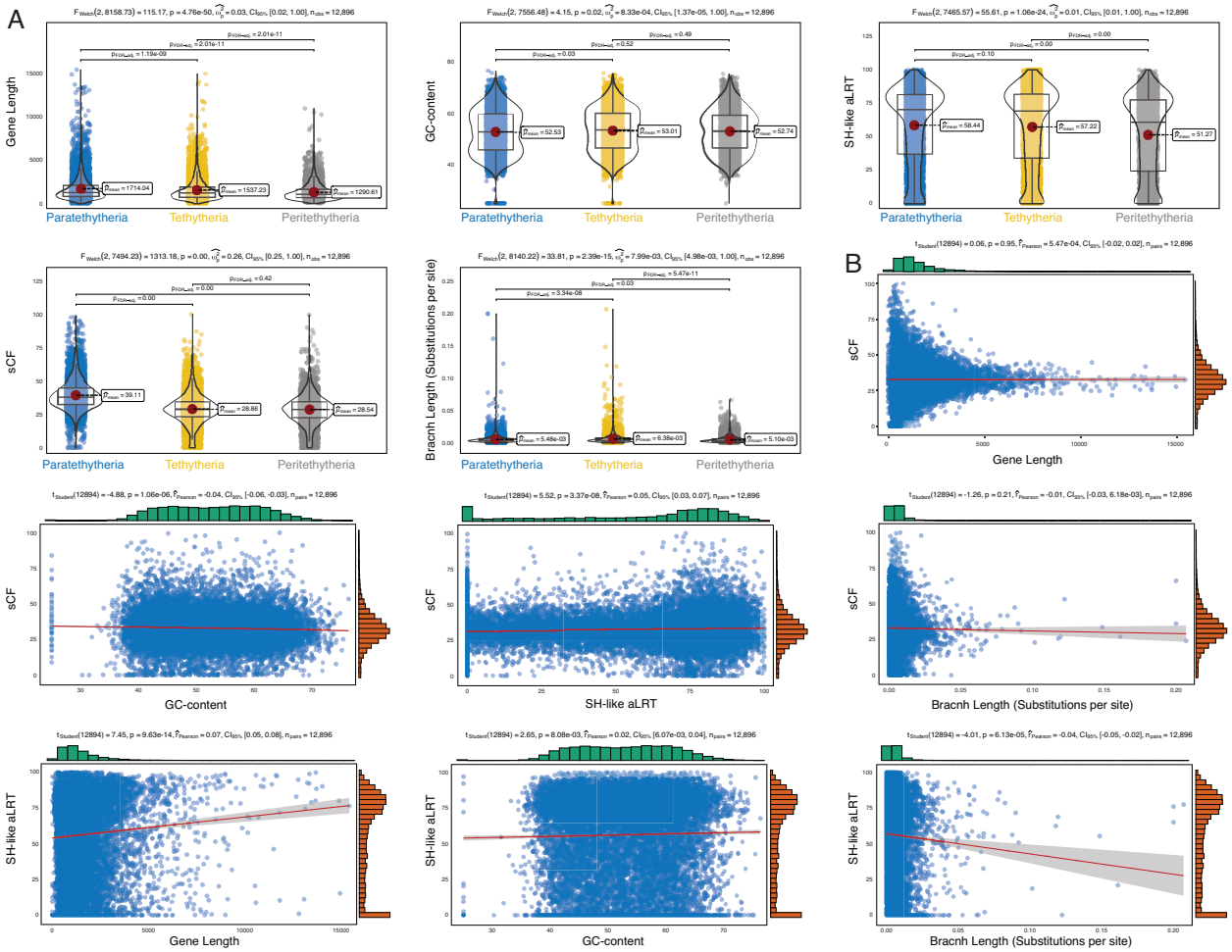
Figure 4. Tree topology tests do not support resolution of the *Paenunulate* polytomy.

A. The reference tree used for topology tests was the ASTRAL species tree with *Paenungulates* collapsed to polytomy, the likelihood of which was compared to the likelihood of the ML tree inferred for each gene.

B. Likelihood difference between the *Paenungulate* polytomy tree and ML gene tree for each gene. The likelihood difference is shown for each gene as jittered dots, overlaid violinplots, and boxplots show the data distribution.

C. Percent of genes that reject each alternative tree, i.e., did not have a statistically significant improvement in likelihood score compared to the polytomy tree ($P \leq 0.05$), inferred from the Kishino-Hasegawa (p-KH), the Shimodaira-Hasegawa (p-SH), the weighted Kishino-Hasegawa (p-WKH), the weighted Shimodaira-Hasegawa (p-WSH), and the approximately unbiased test (p-AU) tests, and the posterior weights from the expected likelihood weight (c-ELW) from the bootstrap proportion (bp-RELL) tests.

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

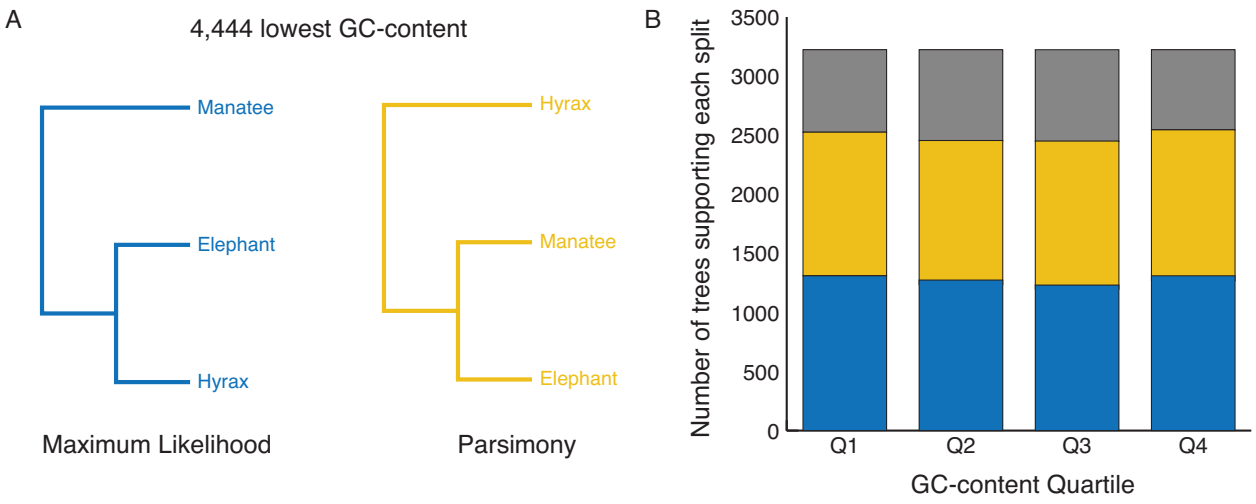


Supplementary Figure 1. Other sources of phylogenetic discordance among *Paenungulates*.

A. Stripchart/violin/boxplots showing differences in mean gene length, GC-content, SH-like aLRT score, sCF, and branch length between genes with gene trees that support the *Paratethytheria*, *Tethytheria*, and *Pertethytheria* splits. Summary statistics and FDR-adjusted *P*-values from pair-wise Games-Howell tests are shown.

B. Scatterplots showing the correlation between variables in panel A; sCF, SH-like aLRT, and branch length were calculated from the ASTRAL species with the *Paratethytheria* split. Summary statistics from Pearson's correlation coefficients are shown above reach split, and side histograms show the distribution of values for each gene.

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



Supplementary Figure 2. GC content variation across genes is an unlikely source of gene tree discordance.

A. Maximum likelihood and parsimony trees inferred from a concatenated supermatrix of the third of genes with the lowest GC content (n=4,444). A GTR substitution model was used for each gene partition in the supermatrix.

B. Number of gene trees supporting each split when genes are binned by GC content quartiles. The *Paratethytheria* split is supported by the greatest number of gene trees in each quartile.

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References

Abadi S, Azouri D, Pupko T, Mayrose I. 2019. Model selection may not be a mandatory step for phylogeny reconstruction. *Nat Commun* **10**:934. doi:10.1038/s41467-019-08822-w

Alon N, Spencer JH. 2023. The Probabilistic Method. doi:10.1002/0471722154

Álvarez-Carretero S, Tamuri AU, Battini M, Nascimento FF, Carlisle E, Asher RJ, Yang Z, Donoghue PCJ, Reis M dos. 2022. A species-level timeline of mammal evolution integrating phylogenomic data. *Nature* **602**:263–267. doi:10.1038/s41586-021-04341-1

Amrine-Madsen H, Koepfli K-P, Wayne RK, Springer MS. 2003. A new phylogenetic marker, apolipoprotein B, provides compelling evidence for eutherian relationships. *Mol Phylogenetics Evol* **28**:225–240. doi:10.1016/s1055-7903(03)00118-0

Anisimova M, Gascuel O, Sullivan J. 2006. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Systematic Biol* **55**:539–552. doi:10.1080/10635150600755453

Anisimova M, Gil M, Dufayard J-F, Dessimoz C, Gascuel O. 2011. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. Systematic Biol 60:685–699. doi:10.1093/sysbio/syr041

Asher RJ. 2007. A web-database of mammalian morphology and a reanalysis of placental phylogeny. *BMC Evol Biol* **7**:108. doi:10.1186/1471-2148-7-108

Bininda-Emonds ORP, Cardillo M, Jones KE, MacPhee RDE, Beck RMD, Grenyer R, Price SA, Vos RA, Gittleman JL, Purvis A. 2007. The delayed rise of present-day mammals. Nature **446**:507–512. doi:10.1038/nature05634

Blainville H-MD de. 1839. Ostéographie, ou, Description iconographique comparée du squelette et du système dentaire des Mammifères récents et fossiles : pour servir de base à la zoologie et à la géologie . Paris: J.B. Baillière et fils.

Bruno WJ, Halpern AL. 1999. Topological bias and inconsistency of maximum likelihood using wrong models. *Mol Biol Evol* **16**:564–566. doi:10.1093/oxfordjournals.molbev.a026137

Cai L, Xi Z, Lemmon EM, Lemmon AR, Mast A, Buddenhagen CE, Liu L, Davis CC. 2020. The Perfect Storm: Gene Tree Estimation Error, Incomplete Lineage Sorting, and Ancient Gene Flow Explain the Most Recalcitrant Ancient Angiosperm Clade, Malpighiales. *Syst Biol* **70**:491–507. doi:10.1093/sysbio/syaa083

Crotty SM, Minh BQ, Bean NG, Holland BR, Tuke J, Jermiin LS, Haeseler AV. 2019. GHOST: Recovering Historical Signal from Heterotachously Evolved Sequence Alignments. Systematic

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Biol 69:249–264. doi:10.1093/sysbio/syz051

Damas J, Corbo M, Kim J, Turner-Maier J, Farré M, Larkin DM, Ryder OA, Steiner C, Houck ML, Hall S, Shiue L, Thomas S, Swale T, Daly M, Korlach J, Uliano-Silva M, Mazzoni CJ, Birren BW, Genereux DP, Johnson J, Lindblad-Toh K, Karlsson EK, Nweeia MT, Johnson RN, Consortium Z, Lewin HA, Andrews G, Armstrong JC, Bianchi M, Birren BW, Bredemeyer KR, Breit AM, Christmas MJ, Clawson H, Damas J, Palma FD, Diekhans M, Dong MX, Eizirik E, Fan K, Fanter C, Foley NM, Forsberg-Nilsson K, Garcia CJ, Gatesy J, Gazal S, Genereux DP, Goodman L, Grimshaw J, Halsey MK, Harris AJ, Hickey G, Hiller M, Hindle AG, Hubley RM, Hughes GM, Johnson J, Juan D, Kaplow IM, Karlsson EK, Keough KC, Kirilenko B, Koepfli K-P, Korstian JM, Kowalczyk A, Kozyrev SV, Lawler AJ, Lawless C, Lehmann T, Levesque DL, Lewin HA, Li X, Lind A, Lindblad-Toh K, Mackay-Smith A, Marinescu VD, Marques-Bonet T, Mason VC, Meadows JRS, Meyer WK, Moore JE, Moreira LR, Moreno-Santillan DD, Morrill KM, Muntané G, Murphy WJ, Navarro A, Nweeia M, Ortmann S, Osmanski A, Paten B, Paulat NS, Pfenning AR, Phan BN, Pollard KS, Pratt HE, Ray DA, Reilly SK, Rosen JR, Ruf I, Ryan L, Ryder OA, Sabeti PC, Schäffer DE, Serres A, Shapiro B, Smit AFA, Springer M, Srinivasan C, Steiner C, Storer JM, Sullivan KAM, Sullivan PF, Sundström E, Supple MA, Swofford R, Talbot J-E, Teeling E, Turner-Maier J, Valenzuela A, Wagner F, Wallerman O, Wang C, Wang J, Weng Z, Wilder AP, Wirthlin ME, Xue JR, Zhang X. 2022. Evolution of the ancestral mammalian karyotype and syntenic regions. *Proc Natl Acad Sci* **119**:e2209139119. doi:10.1073/pnas.2209139119

Delsuc F, Scally M, Madsen O, Stanhope MJ, Jong WW de, Catzeflis FM, Springer MS, Douzery EJP. 2002. Molecular Phylogeny of Living Xenarthrans and the Impact of Character and Taxon Sampling on the Placental Tree Rooting. Mol Biol Evol 19:1**65**6–1**67**1. doi:10.1093/oxfordjournals.molbev.a003989

Doko R, Liu KJ. 2023. The Impact of Multiple Sequence Alignment Error on Phylogenetic Estimation under Variable-Across-Phylogeny Substitution Models. *bioRxiv* 2023.01.24.5**25**451. doi:10.1101/2023.01.24.525451

Domning DP, Ray CE, McKenna MC. 1986. Two New Oligocene Desmostylians and a Discussion of Tethytherian Systematics. *Smithson Contrib Paleobiology* 1–56. doi:10.5479/si.00810266.**59**.1

Douady CJ, Douzery EJP. 2003. Molecular estimation of eulipotyphlan divergence times and the evolution of “Insectivora.” Mol Phylogenetics Evol 28:285–**29**6. doi:10.1016/s10**55**-7903(03)00119-2

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Foley NM, Mason VC, Harris AJ, Bredemeyer KR, Damas J, Lewin HA, Eizirik E, Gatesy J, Karlsson EK, Lindblad-Toh K, Consortium‡ Z, Springer MS, Murphy WJ, Andrews G, Armstrong JC, Bianchi M, Birren BW, Bredemeyer KR, Breit AM, Christmas MJ, Clawson H, Damas J, Palma FD, Diekhans M, Dong MX, Eizirik E, Fan K, Fanter C, Foley NM, Forsberg-Nilsson K, Garcia CJ, Gatesy J, Gazal S, Genereux DP, Goodman L, Grimshaw J, Halsey MK, Harris AJ, Hickey G, Hiller M, Hindle AG, Hubley RM, Hughes GM, Johnson J, Juan D, Kaplow IM, Karlsson EK, Keough KC, Kirilenko B, Koepfli K-P, Korstian JM, Kowalczyk A, Kozyrev SV, Lawler AJ, Lawless C, Lehmann T, Levesque DL, Lewin HA, Li X, Lind A, Lindblad-Toh K, Mackay-Smith A, Marinescu VD, Marques-Bonet T, Mason VC, Meadows JRS, Meyer WK, Moore JE, Moreira LR, Moreno-Santillan DD, Morrill KM, Muntané G, Murphy WJ, Navarro A, Nweeia M, Ortmann S, Osmanski A, Paten B, Paulat NS, Pfenning AR, Phan BN, Pollard KS, Pratt HE, Ray DA, Reilly SK, Rosen JR, Ruf I, Ryan L, Ryder OA, Sabeti PC, Schäffer DE, Serres A, Shapiro B, Smit AFA, Springer M, Srinivasan C, Steiner C, Storer JM, Sullivan KAM, Sullivan PF, Sundström E, Supple MA, Swofford R, Talbot J-E, Teeling E, Turner-Maier J, Valenzuela A, Wagner F, Wallerman O, Wang C, Wang J, Weng Z, Wilder AP, Wirthlin ME, Xue JR, Zhang X. 2023. A genomic timescale for placental mammal evolution. *Science* **380**:eabl8189. doi:10.1126/science.abl8189

Foley NM, Thong VD, Soisook P, Goodman SM, Armstrong KN, Jacobs DS, Puechmaille SJ, Teeling EC. 2015. How and Why Overcome the Impediments to Resolution: Lessons from rhinolophid and hipposiderid Bats. Mol Biol Evol **32**:**31**3–333. doi:10.1093/molbev/msu329

Franco AD, Poujol R, Baurain D, Philippe H. 2019. Evaluating the usefulness of alignment filtering methods to reduce the impact of errors on evolutionary inferences. *Bmc Evol Biol* **19**:**21**. doi:10.1186/s12862-019-1**35**0-2

Gadagkar SR, Rosenberg MS, Kumar S. 2005. Inferring species phylogenies from multiple genes: Concatenated sequence tree versus consensus gene tree. *J Exp Zoöl Part B: Mol Dev Evol* **304B**:64–74. doi:10.1002/jez.b.21026

Gerth M. 2019. Neglecting model selection alters phylogenetic inference. bioRxiv 849018. doi:10.1101/849018

Graur D, Gouy M, Duret L. 1997. Evolutionary Affinities of the Order Perissodactyla and the Phylogenetic Status of the Superordinal Taxa Ungulata and Altungulata. Mol Phylogenetics Evol 7:195–200. doi:10.1006/mpev.1996.0391

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH-Y, Hansen NF, Durand EY, Malaspinas A-S, Jensen JD, Marques-Bonet T, Alkan C,

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Prüfer K, Meyer M, Burbano HA, Good JM, Schultz R, Aximu-Petri A, Butthof A, Höber B, Höffner B, Siegemund M, Weihmann A, Nusbaum C, Lander ES, Russ C, Novod N, Affourtit J, Egholm M, Verna C, Rudan P, Brajkovic D, Kucan Ž, Gušic I, Doronichev VB, Golovanova LV, Lalueza-Fox C, Rasilla M de la, Fortea J, Rosas A, Schmitz RW, Johnson PLF, Eichler EE, Falush D, Birney E, Mullikin JC, Slatkin M, Nielsen R, Kelso J, Lachmann M, Reich D, Pääbo S. 2010. A Draft Sequence of the Neandertal Genome. Science **328**:710–722. doi:10.1126/science.1188021

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic Biol 59:307–321. doi:10.1093/sysbio/syq010

Hibbins MS, Hahn MW. 2021. Phylogenomic approaches to detecting and characterizing introgression. *Genetics* **220**:iyab173. doi:10.1093/genetics/iyab173

Holm S. 1979. A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics* **6**:65.

Hossain ASMM, Blackburne BP, Shah A, Whelan S. 2015. Evidence of Statistical Inconsistency of Phylogenetic Methods in the Presence of Multiple Sequence Alignment Uncertainty. *Genome Biol Evol* **7**:2102–2116. doi:10.1093/gbe/evv127

Huson DH, Klöpper T, Lockhart PJ, Steel MA. 2005. Research in Computational Molecular Biology, 9th Annual International Conference, RECOMB 2005, Cambridge, MA, USA, May **14**-18, 2005. Proceedings. *Lect Notes Comput Sci* 233–249. doi:10.1007/**114**15770_18

Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SYW, Faircloth BC, Nabholz B, Howard JT, Suh A, Weber CC, Fonseca RR da, Li J, Zhang F, Li H, Zhou L, Narula N, Liu L, Ganapathy G, Boussau B, Bayzid S, Zavidovych V, Subramanian S, Gabaldón T, Capella-Gutiérrez S, Huerta-Cepas J, Rekepalli B, Munch K, Schierup M, Lindow B, Warren WC, Ray D, Green RE, Bruford MW, Zhan X, Dixon A, Li S, Li N, Huang Y, Derryberry EP, Bertelsen MF, Sheldon FH, Brumfield RT, Mello CV, Lovell PV, Wirthlin M, Schneider MPC, Prosdocimi F, Samaniego JA, Velazquez AMV, Alfaro-Núñez A, Campos PF, Petersen B, Sicheritz-Ponten T, Pas A, Bailey T, Scofield P, Bunce M, Lambert DM, Zhou Q, Perelman P, Driskell AC, Shapiro B, Xiong Z, Zeng Y, Liu S, Li Z, Liu B, Wu K, Xiao J, Yinqi X, Zheng Q, Zhang Y, Yang H, Wang J, Smeds L, Rheindt FE, Braun M, Fjeldsa J, Orlando L, Barker FK, Jønsson KA, Johnson W, Koepfli K-P, O'Brien S, Haussler D, Ryder OA, Rahbek C, Willerslev E, Graves GR, Glenn TC, McCormack J, Burt D, Ellegren H, Alström P, Edwards SV, Stamatakis A, Mindell DP, Cracraft J, Braun EL, Warnow T, Jun W, Gilbert MTP, Zhang G. 2014. Whole-genome analyses resolve

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

early branches in the tree of life of modern birds. Science **346**:1320–1331. doi:10.1126/science.1253451

Jong WW de, Goodman M. 1982. Mammalian phylogeny studied by sequence analysis of the eye lens protein a-crystallin. *Z Säugetierkunde* 257–276.

Jong WW de, Zweers A, Goodman M. 1981. Relationship of aardvark to elephants, hyraxes and sea cows from α-crystallin sequences. Nature **292**:538–540. doi:10.**103**8/292538a0

JONG WW, ZWEERS A, VERSTEEG M, NUY-TERWINDT EC. 1984. Primary structures of the α-crystallin A chains of twenty-eight mammalian species, chicken and frog. *Eur J Biochem* **141**:**131**–140. doi:10.1111/j.1432-1033.1984.tb08167.x

Jong WWD, Nuy-Terwindt EC, Versteeg M. 1977. Primary structures of α-crystallin a chains of elephant, whale, hyrax and rhinoceros. *Biochim Biophys Acta (BBA) - Protein Struct* **491**:573–580. doi:10.1016/0005-2795(77)90303-8

Kalyaanamoorthy S, Minh BQ, Wong TKF, Haeseler A von, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature methods* **14**:587–589.

Kent WJ. 2002. BLAT—The BLAST-Like Alignment Tool. *Genome Res* **12**:656–664. doi:10.1101/gr.229202

Kjer KM, Honeycutt RL. 2007. Site specific rates of mitochondrial genomes and the phylogeny of eutheria. BMC Evol Biol 7:8. doi:10.1186/1471-21**48**-7-8

Kleinschmidt T, Czelusniak J, Goodman M, Braunitzer G. 1986. Paenungulata: a comparison of the hemoglobin sequences from elephant, hyrax, and manatee. Mol Biol Evol 3:427–35. doi:10.1093/oxfordjournals.molbev.a040411

Kolaczkowski B, Thornton JW. 2008. A Mixed Branch Length Model of Heterotachy Improves Phylogenetic Accuracy. Mol Biol Evol 25:1054–**106**6. doi:10.1093/molbev/msn042

Kolaczkowski B, Thornton JW. 2004. Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. Nature **431**:980–984. doi:10.1038/nature02917

Kuntner M, Collado LJM, Agnarsson I. 2011. Phylogeny and conservation priorities of afrotherian mammals (Afrotheria, Mammalia). *Zool Scr* **40**:1–15. doi:10.1111/j.1463-6**409**.2010.00452.x

Landan G, Graur D. 2008. Local reliability measures from sets of co-optimal multiple sequence alignments. *Pac Symp Biocomput Pac Symp Biocomput* 15–24.

Lanfear R. 2018. Calculating and interpreting gene- and site-concordance factors in phylogenomics.

Lartillot N, Philippe H. 2004. A Bayesian Mixture Model for Across-Site Heterogeneities in the

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Amino-Acid Replacement Process. Mol Biol Evol 21:1095–1109. doi:10.1093/molbev/msh112

Lavergne A, Douzery E, Stichler T, Catzeflis FM, Springer MS. 1996. Interordinal Mammalian Relationships: Evidence for Paenungulate Monophyly Is Provided by Complete Mitochondrial 12S rRNA Sequences. Mol Phylogenetics Evol 6:245–258. doi:10.1006/mpev.1996.0074

Le SQ, Dang CC, Gascuel O. 2012. Modeling Protein Evolution with Several Amino Acid Replacement Matrices Depending on Site Rates. Mol Biol Evol 29:2921–2936. doi:10.1093/molbev/mss112

Linné C von. 1758. Systema naturae per regna tria naturae :secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. doi:10.5962/bhl.title.542

Literman R, Schwartz R. 2021. Genome-Scale Profiling Reveals Noncoding Loci Carry Higher Proportions of Concordant Data. Mol Biol Evol 38:msab026. doi:10.1093/molbev/msab026

Liu F-GR, Miyamoto MM. 1999. Phylogenetic Assessment of Molecular and Morphological Data for Eutherian Mammals. Syst Biol 48:54–64. doi:10.1080/106351599260436

Liu G-M, Pan Q, Du J, Zhu P-F, Liu W-Q, Li Z-H, Wang L, Hu C-Y, Dai Y-C, Zhang X-X, Zhang Z, Yu Y, Li Meng, Wang P-C, Wang X, Li Ming, Zhou X-M, China KL of AE and CB Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China U of CA of S Beijing 100049, China S of LS University of Science and Technology of China, Hefei, Anhui 230026,, China C of LS Nanjing Normal University, Nanjing, Jiangsu 210023,. 2023. Improved mammalian family phylogeny using gap-rare multiple sequence alignment: A timetree of extant placentals and marsupials. *Zoöl Res* **44**:1064–1079. doi:10.24272/j.issn.2095-8137.2023.189

Liu L, Zhang J, Rheindt FE, Lei F, Qu Y, Wang Y, Zhang Y, Sullivan C, Nie W, Wang J, Yang F, Chen J, Edwards SV, Meng J, Wu S. 2017. Genomic evidence reveals a radiation of placental mammals uninterrupted by the KPg boundary. Proc Natl Acad Sci 114:E7282–E7290. doi:10.1073/pnas.1616744114

Lynch VJ, Wagner GP. 2009. Multiple chromosomal rearrangements structured the ancestral vertebrate Hox-bearing protochromosomes. *PLoS genetics* **5**:e1000349.

Madsen O, Scally M, Douady CJ, Kao DJ, DeBry RW, Adkins R, Amrine HM, Stanhope MJ, Jong WW de, Springer MS. 2001. Parallel adaptive radiations in two major clades of placental mammals. Nature 409:610–614. doi:10.1038/35054544

Maswanganye KA, Cunningham MJ, Bennett NC, Chimimba CT, Bloomer P. 2017. Life on the rocks: Multilocus phylogeography of rock hyrax (Procavia capensis) from southern Africa. Mol Phylogenetics Evol 114:49–62. doi:10.1016/j.ympev.2017.04.006

McKenna MC. 1992. The alpha crystallin A chain of the eye lens and mammalian phylogeny.

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Ann Zool Fennici 349–360.

McKenna MC. 1975. Phylogeny of the Primates, A Multidisciplinary Approach. pp. 21–46. doi:10.1007/978-1-4684-2166-8_2

Meredith RW, cka JEJ, Gatesy J, Ryder OA, Fisher CA, Teeling EC, Goodbla A, Eizirik E, Simão TLL, Stadler T, Rabosky DL, Honeycutt RL, Flynn JJ, Ingram CM, Steiner C, Williams TL, Robinson TJ, Burk-Herrick A, Westerman M, Ayoub NA, Springer MS, Murphy WJ. 2011. Impacts of the Cretaceous Terrestrial Revolution and KPg Extinction on Mammal Diversification. Science **334**:521–524.

Minh BQ, Hahn MW, Lanfear R. 2020. New methods to calculate concordance factors for phylogenomic datasets. Mol Biol Evol 37:2727–2733. doi:10.1093/molbev/msaa106

Miyamoto MM, Goodman M. 1986. Biomolecular systematics of eutherian mammals: phylogenetic patterns and classification. *Systematic Biology*.

Mo YK, Lanfear R, Hahn MW, Minh BQ. 2022. Updated site concordance factors minimize effects of homoplasy and taxon sampling. *Bioinformatics* **39**:btac741. doi:10.1093/bioinformatics/btac741

Morgan CC, Creevey CJ, O'Connell MJ. 2014. Mitochondrial data are not suitable for resolving placental mammal phylogeny. *Mamm Genome* **25**:636–647. doi:10.1007/s00335-014-9544-9

Murphy William J, Eizirik E, Johnson WE, Zhang YP, Ryder OA, O'Brien SJ. 2001. Molecular phylogenetics and the origins of placental mammals. Nature 409:614–618.

Murphy William J., Eizirik E, O'Brien SJ, Madsen O, Scally M, Douady CJ, Teeling E, Ryder OA, Stanhope MJ, Jong WW de, Springer MS. 2001. Resolution of the Early Placental Mammal Radiation Using Bayesian Phylogenetics. Science **294**:2348–2351. doi:10.1126/science.1067179

Murphy WJ, Pringle TH, Crider TA, Springer MS, Miller W. 2007. Using genomic data to unravel the root of the placental mammal phylogeny. *Genome Research* **17**:413–421.

Naser-Khdour S, Minh BQ, Lanfear R. 2021. The Influence of Model Violation on Phylogenetic Inference: A Simulation Study. bioRxiv 2021.09.22.461455. doi:10.1101/2021.09.22.461455

Nguyen L-T, Schmidt HA, Haeseler A von, Minh BQ. 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol 32:268–274. doi:10.1093/molbev/msu300

Nishihara H, Hasegawa M, Okada N. 2006. Pegasoferae, an unexpected mammalian clade revealed by tracking ancient retroposon insertions. Proc Natl Acad Sci 103:9929–9934. doi:10.1073/pnas.0603797103

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Nishihara H, Maruyama S, Okada N. 2009. Retroposon analysis and recent geological data suggest near-simultaneous divergence of the three superorders of mammals. Proc Natl Acad Sci 106:5235–5240. doi:10.1073/pnas.0809297106

Nishihara H, Satta Y, Nikaido M, Thewissen JGM, Stanhope MJ, Okada N. 2005. A retroposon analysis of Afrotherian phylogeny. *Molecular biology and evolution* **22**:1823–1833.

Noro M, Masuda R, Dubrovo IA, Yoshida MC, Kato M. 1998. Molecular Phylogenetic Inference of the Woolly Mammoth Mammuthus primigenius, Based on Complete Sequences of Mitochondrial Cytochrome b and 12S Ribosomal RNA Genes. *J Mol Evol* **46**:314–326. doi:10.1007/pl00006308

Novacek MJ. 1982. Macromolecular Sequences in Systematic and Evolutionary Biology 3–41. doi:10.1007/978-1-4684-4283-0_1

Ogden TH, Rosenberg MS. 2006. Multiple Sequence Alignment Accuracy and Phylogenetic Inference. Syst Biol 55:314–328. doi:10.1080/10635150500541730

O'Leary MA, Bloch JI, Flynn JJ, Gaudin TJ, Giallombardo A, Giannini NP, Goldberg SL, Kraatz BP, Luo Z-X, Meng J, Ni X, Novacek MJ, Perini FA, Randall ZS, Rougier GW, Sargis EJ, Silcox MT, Simmons NB, Spaulding M, Velazco PM, Weksler M, Wible JR, Cirranello AL. 2013. The Placental Mammal Ancestor and the Post–K-Pg Radiation of Placentals. Science **339**:662–667.

Ozawa T, Hayashi S, Mikhelson VM. 1997. Phylogenetic Position of Mammoth and Steller's Sea Cow Within Tethytheria Demonstrated by Mitochondrial DNA Sequences. J Mol Evol **44**:406–413. doi:10.1007/pl00006160

Pardini AT, O'Brien PCM, Fu B, Bonde RK, Elder FFB, Ferguson-Smith MA, Yang F, Robinson TJ. 2007. Chromosome painting among Proboscidea, Hyracoidea and Sirenia: support for Paenungulata (Afrotheria, Mammalia) but not Tethytheria. *Proc R Soc B: Biol Sci* **274**:1333–1340. doi:10.1098/rspb.2007.0088

Pease JB, Haak DC, Hahn MW, Moyle LC. 2016. Phylogenomics Reveals Three Sources of Adaptive Variation during a Rapid Radiation. *PLoS Biol* **14**:e1002379. doi:10.1371/journal.pbio.1002379

Phillips MJ. 2016. Geomolecular Dating and the Origin of Placental Mammals. Syst Biol 65:546–557. doi:10.1093/sysbio/syv115

Porter CA, Goodman M, Stanhope MJ. 1996. Evidence on Mammalian Phylogeny from Sequences of Exon 28 of the von Willebrand Factor Gene. Mol Phylogenetics Evol 5:89–101. doi:10.1006/mpev.1996.0008

Poulakakis N, Stamatakis A. 2010. Recapitulating the evolution of Afrotheria: 57 genes and rare

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

genomic changes (RGCs) consolidate their history. *Syst Biodivers* **8**:395–408. doi:10.1080/14772000.2010.484**43**6

Poux C, Madsen O, Glos J, Jong WW de, Vences M. 2008. Molecular phylogeny and divergence times of Malagasy tenrecs: Influence of data partitioning and taxon sampling on dating analyses. BMC Evol Biol 8:102. doi:10.1**18**6/1471-2148-8-102

Prothero DR, Williams MP. 20**17**. Elephants, Hyraxes, Sea Cows, Aardvarks, and Their RelativesThe Princeton Field Guide to Prehistoric Mammals. Princeton University Press. p. 58.

Quang LS, Gascuel O, Lartillot N. 2008. Empirical profile mixture models for phylogenetic reconstruction. Bioinformatics 24:2317–2323. doi:10.1093/bioinformatics/btn445

Rainey WE, Lowenstein JM, Sarich VM, Magor DM. 1984. Sirenian molecular systematics--including the extinct Steller's sea cow (Hydrodamalis gigas). *Die Naturwissenschaften* **71**:586–8. doi:10.1007/bf01189187

Ranwez V, Douzery EJP, Cambon C, Chantret N, Delsuc F. 2018. MACSE v2: Toolkit for the Alignment of Coding Sequences Accounting for Frameshifts and Stop Codons. Mol Biol Evol 35:2582–2584. doi:10.1093/molbev/msy159

Roca AL, Bar-Gal GK, Eizirik E, Helgen KM, Maria R, Springer MS, O'Brien SJ, Murphy WJ. 2004. Mesozoic origin for West Indian insectivores. Nature **429**:649–651. doi:10.1038/nature02597

Rogaev EI, Moliaka YK, Malyarchuk BA, Kondrashov FA, Derenko MV, Chumakov I, Grigorenko AP. 2006. Complete Mitochondrial Genome and Phylogeny of Pleistocene MammothMammuthus primigenius. PLoS Biol 4:e73. doi:10.1371/journal.pbio.0040073

Romiguier J, Ranwez V, Delsuc F, Galtier N, Douzery EJP. 2013. Less Is More in Mammalian Phylogenomics: AT-Rich Genes Minimize Tree Conflicts and Unravel the Root of Placental Mammals. Mol Biol Evol 30:2134–2144. doi:10.1093/molbev/mst116

Salichos L, Rokas A. 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. Nature **497**:327–331. doi:10.1038/nature12130

Salichos L, Stamatakis A, Rokas A. 2014. Novel Information Theory-Based Measures for Quantifying Incongruence among Phylogenetic Trees. Mol Biol Evol 31:1261–1271. doi:10.1093/molbev/msu061

Seiffert ER. 2007. A new estimate of afrotherian phylogeny based on simultaneous analysis of genomic, morphological, and fossil evidence. BMC Evol Biol 7:224. doi:10.1186/1471-2148-7-224

Sela I, Ashkenazy H, Katoh K, Pupko T. 2015. GUIDANCE2: accurate detection of unreliable

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Research* **43**:W7-14.

Shen X-X, Hittinger CT, Rokas A. 2017. Contentious relationships in phylogenomic studies can be driven by a handful of genes. *Nat Ecol Evol* **1**:0126. doi:10.1038/s41559-017-0126

Shoshani J. 1986. Mammalian phylogeny: comparison of morphological and molecular results. Mol Biol Evol 3:222–42. doi:10.1093/oxfordjournals.molbev.a040389

Shoshani JH. 1986. On the phylogenetic relationships among paenungulata and within elephantidae as demonstrated by molecular and osteological evidence. Wayne State University.

Simmons MP, Sloan DB, Springer MS, Gatesy J. 2019. Gene-wise resampling outperforms site-wise resampling in phylogenetic coalescence analyses. Mol Phylogenetics Evol 131:80–92. doi:10.1016/j.ympev.2018.10.001

Springer MS, Amrine HM, Burk A, Stanhope MJ, Waddell P, Hasegawa M. 1999. Additional Support for Afrotheria and Paenungulata, the Performance of Mitochondrial versus Nuclear Genes, and the Impact of Data Partitions with Heterogeneous Base Composition. Syst Biol 48:65–75. doi:10.1080/106351599260445

Springer MS, Foley NM, Brady PL, Gatesy J, Murphy WJ. 2019. Evolutionary Models for the Diversification of Placental Mammals Across the KPg Boundary. *Front Genet* **10**:1241. doi:10.3389/fgene.2019.01241

Springer MS, Kirsch JAW. 1993. A molecular perspective on the phylogeny of placental mammals based on mitochondrial 12S rDNA sequences, with special reference to the problem of the Paenungulata. *J Mamm Evol* **1**:149–166. doi:10.1007/bf01041592

Springer MS, Signore AV, Paijmans JLA, Vélez-Juarbe J, Domning DP, Bauer CE, He K, Crerar L, Campos PF, Murphy WJ, Meredith RW, Gatesy J, Willerslev E, MacPhee RDE, Hofreiter M, Campbell KL. 2015. Interordinal gene capture, the phylogenetic position of Steller's sea cow based on molecular and morphological data, and the macroevolutionary history of Sirenia. Mol Phylogenetics Evol 91:178–193. doi:10.1016/j.ympev.2015.05.022

Stanhope MJ, Madsen O, Waddell VG, Cleven GC, Jong WW de, Springer MS. 1998. Highly Congruent Molecular Support for a Diverse Superordinal Clade of Endemic African Mammals. *Molecular Phylogenetics and Evolution* **9**:501–508.

Stanhope MJ, Smith MR, Waddell VG, Porter CA, Shivji MS, Goodman M. 1996. Mammalian evolution and the interphotoreceptor retinoid binding protein (IRBP) gene: Convincing evidence for several superordinal clades. J Mol Evol 43:83–92. doi:10.1007/bf02337352

Suvorov A, Kim BY, Wang J, Armstrong EE, Peede D, D'Agostino ERR, Price DK, Waddell PJ,

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Lang M, Courtier-Orgogozo V, David JR, Petrov D, Matute DR, Schrider DR, Comeault AA. 2022. Widespread introgression across a phylogeny of 155 Drosophila genomes. *Curr Biol* **32**:111-123.e5. doi:10.1016/j.cub.2021.10.052

Tabuce R, Marivaux L, Adaci M, Bensalah M, Hartenberger J-L, Mahboubi M, Mebrouk F, Tafforeau P, Jaeger J-J. 2007. Early Tertiary mammals from North Africa reinforce the molecular Afrotheria clade. Proc R Soc B: Biol Sci 274:1159–1166. doi:10.1098/rspb.2006.0229

Thines M, Aoki T, Crous PW, Hyde KD, Lücking R, Malosso E, May TW, Miller AN, Redhead SA, Yurkov AM, Hawksworth DL. 2020. Setting scientific names at all taxonomic ranks in italics facilitates their quick recognition in scientific papers. *IMA Fungus* **11**:25. doi:10.1186/s43008-020-00048-6

Upham NS, Esselstyn JA, Jetz W. 2019. Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. PLoS Biol 17:e3000494. doi:10.1371/journal.pbio.3000494

Vanderpool D, Minh BQ, Lanfear R, Hughes D, Murali S, Harris RA, Raveendran M, Muzny DM, Hibbins MS, Williamson RJ, Gibbs RA, Worley KC, Rogers J, Hahn MW. 2020. Primate phylogenomics uncovers multiple rapid radiations and ancient interspecific introgression. PLoS Biol 18:e3000954. doi:10.1371/journal.pbio.3000954

Vazquez JM, Lynch VJ. 2021. Pervasive duplication of tumor suppressors in Afrotherians during the evolution of large bodies and reduced cancer risk. *Elife* **10**:e65041. doi:10.7554/elife.65041

Waddell PJ, Shelley S. 2003. Evaluating placental inter-ordinal phylogenies with novel sequences including RAG1, γ-fibrinogen, ND6, and mt-tRNA, plus MCMC-driven nucleotide, amino acid, and codon models. Mol Phylogenetics Evol 28:197–224. doi:10.1016/s1055-7903(03)00115-5

Wang H-C, Li K, Susko E, Roger AJ. 2008. A class frequency mixture model that adjusts for site-specific amino acid frequencies and improves inference of protein phylogeny. BMC Evol Biol 8:331. doi:10.1186/1471-2148-8-331

Wang H-C, Minh BQ, Susko E, Roger AJ. 2018. Modeling Site Heterogeneity with Posterior Mean Site Frequency Profiles Accelerates Accurate Phylogenomic Estimation. Systematic Biol 67:216–235. doi:10.1093/sysbio/syx068

Weitz B. 1953. Serological Relationships of Hyrax and Elephant. Nature **171**:261–261. doi:10.1038/171261a0

Wu J, Hasegawa M, Zhong Y, Yonezawa T. 2014. Importance of synonymous substitutions under dense taxon sampling and appropriate modeling in reconstructing the mitogenomic tree of

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Eutheria. *Genes Genet Syst* **89**:237–251. doi:10.1266/ggs.89.237

Yang Z. 1997. How often do wrong models produce better phylogenies? Mol Biol Evol 14:105–108. doi:10.1093/oxfordjournals.molbev.a025695

Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *Bmc Bioinformatics* **19**:153. doi:10.1186/s12859-018-2129-y

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Supplementary Table 1. Molecular phylogenetic studies that have included *Paenungulata*. If the study has more than 10 genes, total alignment length (TAL) is shown.

Study	Data type	Method	Phylogeny
(Weitz, 1953)	Albumin	Antibody cross-reactivity	Paenungulata[4]
(Jong et al., 1977)	Protein (n=1)	Overall similarity	Paenungulata[5]
(Jong et al., 1981)	Protein (n=1)	Parsimony	Peritethytheria
(Jong and Goodman, 1982)	Protein (n=1)	Parsimony	Peritethytheria
(McKenna, 1992)	Protein (n=1)	Parsimony	Peritethytheria
(Rainey et al., 1984)	Albumin	Immunological distances	Unresolved
(JONG et al., 1984)	Protein (n=1)	Parsimony	Peritethytheria
(J. H. Shoshani, 1986)	Albumin	Immunological distances	Paratethytheria
(J. H. Shoshani, 1986)	Whole sera	Immunological distances	Tethytheria
(J. Shoshani, 1986)	Albumin, whole sera	UWPGN	Paratethytheria
(J. Shoshani, 1986)	Morphology	Parsimony	Tethytheria
(Kleinschmidt et al., 1986)	Protein (n=2)	Concatenation, Parsimony	Tethytheria
(Miyamoto and Goodman, 1986)	Protein (n=7)	Concatenation, Parsimony	Paratethytheria
(Miyamoto and Goodman, 1986)	Protein (n=7)	Concatenation, Parsimony	Unresolved
			Tethytheria
(Springer and Kirsch, 1993)	Mitochondrial gene (n=1)	Parsimony	Tethytheria
(Porter et al., 1996)	Nuclear gene (n=1)	Parsimony, Neighbor-joining	Peritethytheria
(Lavergne et al., 1996)	rRNA (n=1)	Parsimony, Neighbor-joining	Tethytheria

[4] Included ox, African elephant, rock hyrax, white rhinoceros, and minke whale but did not include manatee, thus it could not resolve relationships within Paenungulata but noted that African elephant and rock hyrax had several amino acid changes in common compared to the other species.
[5] Included Asian elephant, yellow-spotted rock hyrax (*H. b. prittwitzi*),human, ox, sheep, horse, pig, dog and cat but did not include manatee, thus it could not resolve relationships within Paenungulata but noted that African elephant and rock hyrax anti-elephant and anti-hyrax sera cross-reacted with each other but not the other species.

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Study	Data type	Method	Phylogeny
(Stanhope et al., 1996)	Nuclear gene (n=1)	Parsimony, Neighbor-joining	Paratethytheria
(Graur et al., 1997)	Mitochondrial and nuclear genes (n=36, TAL=7.9kb)	Concatenation, Maximum likelihood	Unresolved
(Graur et al., 1997)	Mitochondrial and nuclear genes (n=36, TAL=7.9kb)	Concatenation, Neighbor-joining	Peritethytheria
(Graur et al., 1997)	Mitochondrial and nuclear genes (n=36, TAL=7.9kb)	Concatenation, Parsimony	Unresolved
(Ozawa et al., 1997)	Mitochondrial gene (n=1)	Neighbor-joining (nucleotide and amino acids)	Tethytheria
(Ozawa et al., 1997)	Mitochondrial gene (n=1)	Parsimony (amino acids)	Peritethytheria
(Stanhope et al., 1998)	Nuclear gene (n=1)	Parsimony	Tethytheria
(Stanhope et al., 1998)	Mitochondrial gene (n=2)	Parsimony	Peritethytheria
(Stanhope et al., 1998)	Nuclear gene (n=1)	Parsimony	Paratethytheria
(Stanhope et al., 1998)	Nuclear gene (n=1)	Parsimony	Peritethytheria
(Stanhope et al., 1998)	Nuclear gene (n=1)	Parsimony	Peritethytheria

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Study	Data type	Method	Phylogeny
(Noro et al., 1998)	Mitochondrial genes (n=2)	Neighbor-joining, Parsimony	Tethytheria
(Springer et al., 1999)	Mitochondrial and nuclear genes (n=8, TAL=7.3kb)	Concatenation, Maximum likelihood	Paratethytheria
(Springer et al., 1999)	Mitochondrial genes (n=4, TAL=3.3kb)	Concatenation, Maximum likelihood	Tethytheria
(Springer et al., 1999)	Mitochondrial and nuclear genes (n=4, TAL=4kb)	Concatenation, Maximum likelihood	Paratethytheria
(Liu and Miyamoto, 1999)	Nuclear genes (n=3, TAL=3.6kb)	Concatenation, Parsimony	Peritethytheria
(William J. Murphy et al., 2001)	Mitochondrial and nuclear genes (n=22, TAL=16.4kb)	Concatenation, Bayesian	Peritethytheria
(William J. Murphy et al., 2001)	Mitochondrial and nuclear genes (n=22, TAL=16.4kb)	Concatenation, Maximum likelihood	Paratethytheria
(Madsen et al., 2001)	Mitochondrial and nuclear genes (n=6, TAL=2.9kb)	Concatenation, Maximum likelihood	Peritethytheria
(Madsen et al., 2001)	Mitochondrial and nuclear genes (n=6, TAL=5.7kb)	Concatenation, Maximum likelihood	Paratethytheria
(Madsen et al., 2001)	Mitochondrial and nuclear genes (n=7, TAL=8.6kb)	Concatenation, Maximum likelihood	Peritethytheria
(William J Murphy et al., 2001)	Mitochondrial and nuclear genes (n=18, TAL=9.8kb)	Concatenation, Neighbor-joining	Tethytheria
(William J Murphy et al., 2001)	Mitochondrial and nuclear genes (n=18, TAL=9.8kb)	Concatenation, Parsimony	Paratethytheria
(William J Murphy et al., 2001)	Mitochondrial and	Concatenation, Maximum	Tethytheria

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Study	Data type	Method	Phylogeny
	nuclear genes (n=18, TAL=9.8kb)	likelihood	
(Delsuc et al., 2002)	Nuclear genes (n=3, TAL=5.1kb)	Concatenation, Maximum likelihood	Peritethytheria
(Waddell and Shelley, 2003)	Mitochondrial and nuclear genes (n=8, TAL=?kb)	Concatenation, Maximum likelihood	Peritethytheria
(Douady and Douzery, 2003)	Nuclear genes (n=3, TAL=1367 AAs)	Bayesian	Tethytheria
(Amrine-Madsen et al., 2003)	Nuclear gene (n=1)	Maximum likelihood, Bayesian	Paratethytheria
(Amrine-Madsen et al., 2003)	Mitochondrial and nuclear genes (n=23, TAL=17.7kb)	Concatenation, Maximum likelihood, Bayesian	Paratethytheria
(Roca et al., 2004)	Mitochondrial and nuclear genes (n=19, TAL=13.9kb)	Concatenation, Maximum likelihood	Peritethytheria
(Nishihara et al., 2005)	Retroposons	Parsimony	Peritethytheria
(Gadagkar et al., 2005)	Mitochondrial and nuclear genes (n=448, TAL=?kb)	Majority rule among gene trees, Neighbor-joining	Paratethytheria
(Nishihara et al., 2006)	Retroposons	Parsimony	Peritethytheria
(Rogaev et al., 2006)	Complete mitochondrial genomes	Bayesian, Parsimony, Neighbor-joining	Peritethytheria
(Murphy et al., 2007)	Mitochondrial and nuclear genes (n=19, TAL=13.9kb)	Concatenation, Maximum likelihood	Peritethytheria
(Kjer and Honeycutt, 2007)	Complete mitochondrial genomes	Concatenation, Bayesian	Tethytheria
(Seiffert, 2007)	Morphology, Nuclear genes (n=23, TAL=17.7kb),	Concatenation, Parsimony	Tethytheria

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Study	Data type	Method	Phylogeny
	chromosome painting, and retroposons		
(Asher, 2007)	Morphology, Nuclear genes (n=23, TAL=17.7kb), indels	Concatenation, Parsimony	Tethytheria
(Tabuce et al., 2007)	Morphology	Parsimony	Tethytheria
(Pardini et al., 2007)	Chromosome painting	Parsimony	Unresolved
(Bininda-Emonds et al., 2007)	Supertree	Matrix Representation with Parsimony	Peritethytheria
(Poux et al., 2008)	Nuclear genes (n=4, TAL=4.3kb)	Concatenation, Maximum likelihood	Peritethytheria
(Nishihara et al., 2009)	Retroposon insertions	Parsimony	Peritethytheria
(Poulakakis and Stamatakis, 2010)	Mitochondrial and nuclear genes (n=57, TAL=32.2kb), chromosome rearrangements, and retroposons	Concatenation, Maximum likelihood, Bayesian	Tethytheria
(Kuntner et al., 2011)	Mitochondrial and nuclear genes (n=9, TAL=?kb)	Concatenation, Bayesian	Tethytheria
(Meredith et al., 2011)	Mitochondrial and nuclear genes (n=26, TAL=35.6kb)	Concatenation, Maximum likelihood, Bayesian	Paratethytheria
(O'Leary et al., 2013)	Morphology and nuclear genes (n=27, TAL>35.6kb)	Concatenation, Parsimony	Tethytheria
(Morgan et al., 2014)	Mitochondrial protein coding genes (n=13, TAL= 3906 AA)	Concatenation, Maximum likelihood	Tethytheria

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Study	Data type	Method	Phylogeny
(Wu et al., 2014)	Mitochondrial protein coding genes (n=12, TAL=10.7kb)	Concatenation, Maximum likelihood	Tethytheria
(Springer et al., 2015)	Nuclear genes (n=26, TAL=30kb)	Concatenation, Maximum likelihood	Tethytheria
(Phillips, 2016) replication of Meredith et al. (2011)	Mitochondrial and nuclear genes (n=26, TAL=35.6kb)	Concatenation, Bayesian	Paratethytheria
(Liu et al., 2017)	Nuclear genes (n=4388, TAL=13,040kb)	Multispecies coalescence (STAR, NJist)	Tethytheria
(Maswanganye et al., 2017)	Mitochondrial gene (n=1)	Bayesian	Tethytheria
(Springer et al., 2019)	Morphology (4,541 characters)	Parsimony on a molecular backbone	Tethytheria
(Upham et al., 2019)	Nuclear genes (n=31, TAL=39,099bp)	“backbone-and-patch”	Paratethytheria
(Vazquez and Lynch, 2021)	Gene duplications	Maximum likelihood	Tethytheria
(Souza et al., 2021)	Mitogenome	Maximum likelihood	Tethytheria
(Álvarez-Carretero et al., 2022)	Nuclear genes (n=15,268, TAL=33.2Mb)	Bayesian	Tethytheria
(Damas et al., 2022)	Reconstructed ancestral karyotypes	Parsimony	Peritethytheria
(Foley et al., 2023)	411,110 genome-wide nearly neutral sites; 100-kb alignment windows along human Chr1, Chr21, Chr22, and ChrX	Multispecies coalescence (SVDQuartests)	Paratethytheria
(Liu et al., 2023)	Dataset A: Nuclear genes (n=26, TAL=35,063bp) Dataset B: Nuclear	Maximum likelihood, Bayesian, Multispecies coalescence (ASTRAL-III)	Paratethytheria

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.07.570590; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Study	Data type	Method	Phylogeny
	genes (n=98, TAL=67, 080bp)		
Lynch Unpublished	Gene losses	Dollo parsimony	Paratethytheria