# Placental Mammals - Report

Leann M. Biancani

This study aims to use placental mammal phylogenetic analysis as a case study to investigate the sources of conflicting phylogenetic signal in a genomic dataset. Despite the ever-increasing availability of genomic sequence data, there are still many controversial evolutionary relationships throughout the tree of life. Although genomic datasets have greatly increased the amount of phylogenetic signal present in analyses, they have also increased the amount of information that may support incorrect relationships (Literman and Schwartz 2021). Some sources of such non-phylogenetic signal include alignment errors, inaccurate models of evolution, incorrectly identified orthologs, and homoplasy. This indicates a need to better understand both the sources of conflicting data as well as the systematic and biological biases they represent. This is vital to improving the selection of loci that both maximize phylogenetic signal and minimize these sources of error.

### Objectives and Questions

**Objective 1:** Phylogenetic analysis of placental mammals using publicly available high-throughput genomic sequencing reads and a (nearly) assembly-free approach to identify orthologous loci.

- *Which competing hypotheses are supported by this dataset as a whole?*
- *What characteristics or classifications are exhibited by loci supporting each alternative hypothesis?*

**Objective 2:** Identify and filter loci that provide better phylogenetic signal across the tree and assess their support for competing hypotheses.

- *Which competing hypotheses are supported by loci that also support "control trees" (previously established monophyletic groups)?*
- *What characteristics or classifications are exhibited by loci that support these control trees?*
- *Is there overlap among loci that support different established clades? Is it therefore possible to identify loci that provide reliable signal across the tree?*

**Objective 3:** Annotate loci to classify and filter by type and assess their support for both competing hypotheses and established monophyletic groups.

- *Is there a relationship between loci type and phylogenetic signal across the tree or support for specific clades?*

### Background

Despite a long history of extensive study, mammalian clades have been the subject of many scientific debates (Morgan et al. 2013). Historically, controversial evolutionary relationships were spurred by widespread convergent evolution of mammalian morphological characteristics. The advent of molecular analyses were

very helpful in determining relationships among most living (and some extinct) mammals, but a handful of stubborn polytomies remain unresolved (Foley et al. 2016, Liu et al. 2017). One of the most notable of these is the branching pattern at the root of placental mammals.

Placental mammals are currently divided into three well-established clades: Afrotheria, Xenarthra, and Boreoeutheria (Morgan et al. 2013). Afrotheria includes elephants, aardvarks, manatees, and tenrecs. Xenarthra includes armadillos, anteaters, and sloths. Boreoeutheria contains the remaining placental mammal clades. It is often listed as a pair of sister clades: the Euarchontoglires (primates, rodents, rabbits) and the Laurasiatheria (ungulates, whales, bats, horses, carnivores) (Foley et al. 2016). There are three competing hypotheses (Figure 1) about the evolutionary relationships among these three clades, each of which are supported by numerous studies using diverse datasets, data types and approaches.

The Afrotheria hypothesis (Figure 1a) places Afrotheria as sister to all other placental mammals (Murphy et al. 2001, Asher 2007, Nishihara et al. 2007, Hallstrom and Janke 2010, McCormack et al. 2012, Romiguier et al. 2013). The Xenarthra hypothesis (Figure 1b) places Xenarthra as sister to all other placentals (Kreigs et al. 2006, Churakov et al. 2009, O'Leary et al. 2013). Finally, the Atlantogenata (Figure 1c) places Afrotheria and Xenarthra together (forming clade Atlantogenata) as sister to all other placental mammals (Murphy et al. 2007, Wildman et al. 2007, Prasad et al. 2008, Meredith et al. 2011, Song et al. 2012, Morgan et al. 2013, Tarver et al. 2016).

A number of reasons have been suggested to account for the uncertainty surrounding the the root of placental mammals. Many studies face limitations with respect to gene sampling, taxon sampling and computational limits. Large-scale gene tree heterogeneity due to incomplete lineage sorting may have resulted from the apparent rapid divergence associated with the fragmentation of Pangea and Gondwana. If placental mammals evolved extremely rapidly, it may be theoretically impossible to resolve the note, suggesting a "hard polytomy" and an underlying evolutionary process that was not strictly bifurcating. Other sources of debate have been attributed to examples of non-phylogenetic signal such as incorrectly identified orthologs, long-branch attraction, and evolutionary rate variation (reviewed in Morgan et al. 2013, Tarver et al. 2016).


Figure 1: **Figure 1.** Alternative hypotheses for the branching pattern at the root of placental mammals. Adapted from Morgan et a. 2013. Prior Studies, both molecular and morphological, have found support for all three hypotheses. **a. Afrotheria hypothesis** (supported by Murphy et al. 2001, Asher et al. 2007, Nishihara et al. 2007, Hallstrom and Janke 2010, McCormack et al. 2012, and Romiguier et al. 2013). **b. Xenarthra hypothesis** (supported by Keri's et al. 2006, Charakov et al. 2009, and O'Leary et al. 2013). **c. Atlantogenata hypothesis** (supported by Murphy et al. 2007, Wildman et al. 2007, Prasad et al. 2008, Meredith et al. 2011, Song et al. 2012, Morgan et al. 2013, and Tarver et al. 2016).

## Methods

**Taxon Selection** Taxa, for which appropriate genomic data is publicly available via the European Nucleotide Archive, were selected to span the diversity or each clade. To root the analyses, Marsupialia, the sister-group of placental mammals, was selected as the outgroup clade. Ten taxa, representing ten genera and as many families as possible, were identified for each of three in-group and one outgroup clade. A total of 40 genera and 34 families are represented in this dataset. Paired-end Illumina reads with a sequencing depth of at least 10x coverage (estimated from the larger of either the size of an available genome assembly for each taxon or the average mammalian genome size of ~3.5 billion base pairs) were obtained from the European Nucleotide Archive. A taxon selection list with accession numbers is provided in the repository `README` file.

**Alignment and Orthology Assessment** Orthologous loci were identified and aligned using the SISRS v2.1, a pipeline for identifying phylogenetically informative sites directly from high-throughput, whole-genome sequence data (Schwartz et al. 2015; Literman and Schwartz 2021). Briefly, the SISRS pipeline first performs quality and adapter trimming on input reads, then identifies conserved regions by conducting a single, joint de novo assembly using a subset of reads from all taxa. This composite genome, therefore, contains variation from across the tree (Schwartz et al. 2015). Trimmed reads are then aligned back to these contigs to identify variable sites and call genotypes. Finally, sites with missing data are removed from the output alignment. Customized Slurm job files use to run these analyses can be found in the `Scripts\SlurmJobFiles\` directory and were run in numberical order.

**Phylogenetic Inference** Phylogenetic Analyses were performed using RAxML (Stamatakis 2014) and SVDQuartets (Chifman and Kubatko 2014). Slurm job files used to run these analyses an be found in the `Scripts\PhyloJobFiles\`directory and output files are in the `TreeFiles\` directory. Preliminary results of the RAxML maximum likelihood analysis are included in Figure 2.

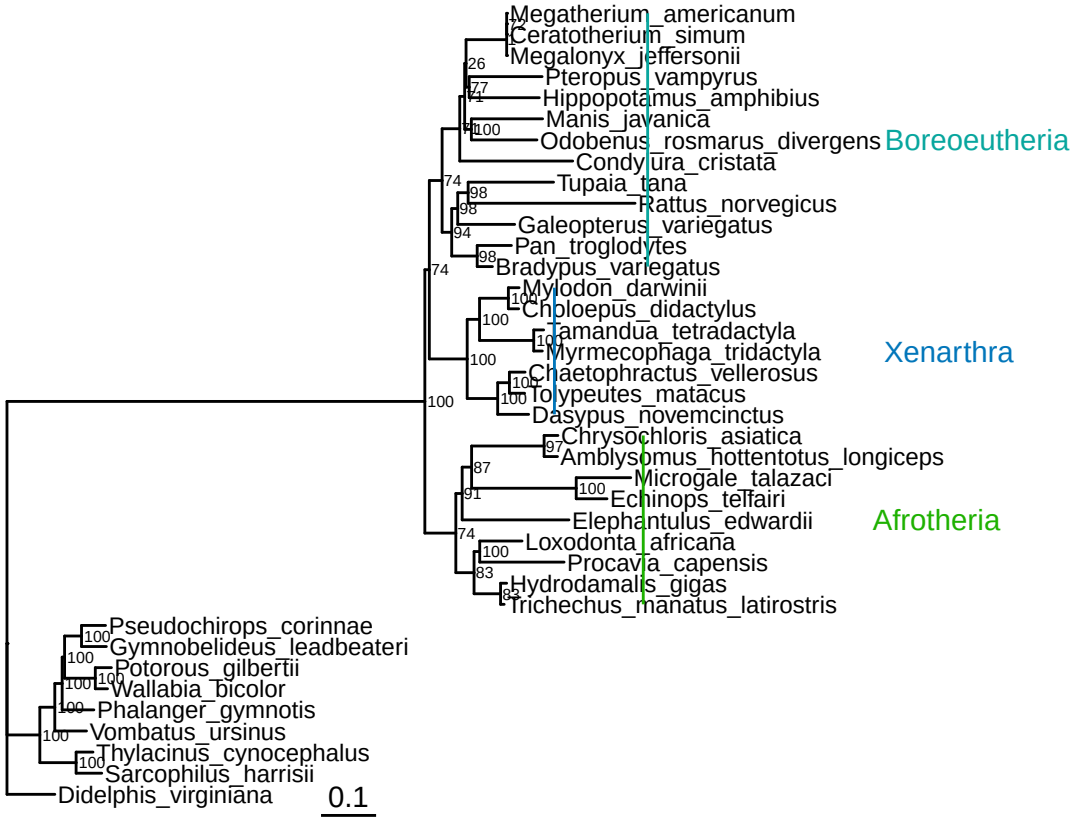


**Figure 2.** Maximum Likelihood Phylogenetic Analysis performed in RAxML for 30 placental mammal and 10 marsupial outgroup taxa. Orthologous loci were obtained from raw genomic reads using the SISRS pipeline. Support values represent 100 bootstrap replicates. Placental mammal clades are indicated by color-coded vertical lines. These preliminary results appear to support the Afrotheria hypothesis (Figure 1a) for the branching pattern at the root of placental mammals.

## References

R. J. Asher. A web-database of mammalian morphology and a reanalysis of placental phylogeny. BMC Evol Biol. 7:108, 2007.

Chifman, J. and L. Kubatko. Quartet inference from SNP data under the coalescent, Bioinformatics, 30(23): 3317-3324, 2014.

N.M. Foley, M.S. Springer, E.C. Teeling. Mammal madness: is the mammal tree of life not yet resolved? Philosophical Transactions, Royal Society B, 371: 20150140, 2016.

B. M. Hallstrom and A. Janke. Mammalian evolution may not be strictly bifurcating. Mol Biol Evol. 27:2804-2816, 2010

J. E. McCormack et al. Ultraconserved elements are novel phylogenetic markers that resolve placental mammal phylogeny when combined with species tree analysis. Genome Res. 22:746-754, 2012

C. Morgan, P. Foster, A. Webb, D. Pisani, J. McInerney, and M. O'Connell. Heterogeneous models place the root of the placental mammal phylogeny. Molecular Biology and Evolution, 30:2145–2156, 2013.

W. J. Murphy et al. Molecular phylogenetic and the origins of placental mammals. Nature 409:614-618, 2001

W. J. Murphy, T.H. Pringle, T.A. Crider, M.S. Springer, W. Miller. Using genomic data to unravel the root of the placental mammal phylogeny. Genome Res. 17:413-421, 2007

R. Literman and R. Schwartz. Genome-scale profiling reveals non-coding loci carry higher proportions of concordant data. Molecular Biology and Evolution, 38:2306-2318, 2021.

L. Liu, J. Zhang, F. E. Rheindt, F. Lei, Y. Qu, Y. Wang, Y. Zhang, C. Sullivan, W. Nie, J. Wang, F. Yang, J. Chen, S. V. Edwards, J. Meng, and S. Wu. Genomic evidence reveals a radiation of placental mammals uninterrupted by the KPg boundary. Proceedings of the National Academy of Sciences of the United States of America, 114:E7282–E7290, 2017.

H. Nishihara, N. Okada, M. Hasegawa. Rooting the eutherian tree: the power and pitfalls of phylogenomics. Genome Biol. 8:R199, 2007

J. Romiguier, V. Ranwez, F. Delsuc, N. Galtier, E.J.P. Douzer. 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, 2013

R. Schwartz, K. Harkins, A. Stone, and R. Cartwright. A composite genome approach to identify phylogenetically informative data from next-generation sequencing. BMC Bioinformatics, 16:193, 2015.

A. Stamatakis. RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. Bioinformatics, 2014

J.E. Tarver , M. Dos Reis, S. Mirarab, R.J. Morgan, S. Parker, J.E. O'Reilly, B.L. King, M.J. O'Connell, R.J. Asher, T. Warnow, K.J. Peterson, P.C.J. Donoghue, D. Pisani. The Interrelationships of Placental Mammals and the Limits of Phylogenetic Inference. Genome Biol. 8(2):330-344, 2016