

1 **Coalescent-based phylogenetic inference from genes with unequivocal historical signal**
2 **suggests a polytomy at the root of the placental mammal tree of life**

3
4 Filipe R. R. Moreira and Carlos G. Schrago

5
6 *Department of Genetics, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil*

7
8
9
10
11
12 *Address for correspondence:

13 Carlos G. Schrago

14 Universidade Federal do Rio de Janeiro

15 Instituto de Biologia, Departamento de Genética, CCS, A2-092

16 Rua Prof. Rodolpho Paulo Rocco, SN

17 Cidade Universitária

18 Rio de Janeiro, RJ

19 CEP: 21941-617

20 BRAZIL

21 Phone: +55 21 3938-6397

22 Email: carlos.schrago@gmail.com

23
24 Keywords: Phylogenomics; Placental root; Polytomy; Phylogenetic information

38 **Abstract**

39 Finding the correct place of the root of the Placentalia tree of life is an unsolved problem in
40 mammalian systematics. Three major competing hypotheses were proposed, alternating the position
41 of the supraordinal taxa Afrotheria, Xenarthra, or Atlantogenata as sister lineages to the remaining
42 placentals. While all three phylogenetic associations were inferred by studies conducted so far,
43 recent assessments applying heterogeneous models and coalescent-based phylogenetic methods
44 found robust support for the Atlantogenata hypothesis. Current developments in theoretical
45 phylogenetics, such as the demonstration that concatenating genes may bias the likelihood function,
46 and that summary coalescent-based phylogenetic methods are sensitive to errors in gene tree
47 estimation, calls for a reevaluation of the early placental split problem. We investigated the
48 phylogenetic relationships between placental superorders by restricting the analysis to subsets of
49 genes with unequivocal phylogenetic signal. In contrast to previous works, we show that the
50 hypothesis of a polytomy at the placental root could not be rejected under the multispecies
51 coalescent model. This result endorses conclusions from analyses of retroposon insertion patterns.
52 We provide an analytical framework to assess empirical polytomies employing summary coalescent
53 methods and topological tests, helping the investigation of poorly resolved nodes in the tree of life.

54

55 **Introduction**

56 Rooting the placental tree of life is a major challenge in mammalian systematics (Teeling
57 and Hedges, 2013). Three hypotheses are generally considered regarding the phylogenetic
58 association of the major mammalian lineages, positioning Afrotheria (e.g. elephants and hyraxes;
59 (Asher, 2007; McCormack et al., 2012; Murphy et al., 2001; Nikolaev et al., 2007; Nishihara et al.,
60 2007; Romiguier et al., 2013), Xenarthra (e.g., armadillos and sloths; (Kriegs et al., 2006; O’Leary
61 et al., 2013; Waddell et al., 2001), or Atlantogenata, i.e., Afrotheria + Xenarthra (Hallström et al.,
62 2007; Hallström and Janke, 2008; Kjer and Honeycutt, 2007; Meredith et al., 2011; Morgan et al.,
63 2013; Murphy et al., 2007; Prasad et al., 2008; Song et al., 2012; Tarver et al., 2016; Wildman et al.,
64 2007), as the sister group to the remaining placentals (**Figure 1**). So far, studies employing different
65 data sources and methods, including genomic data, were unable to unambiguously resolve the early
66 placental split, motivating the proposition that the first divergences of the placental tree were not
67 strictly bifurcating (Churakov et al., 2009; Hallström and Janke, 2010; Nishihara et al., 2009).
68 Alternatively, other works suggested that this phylogenetic incongruence may be attributed to the
69 effects of incomplete lineage sorting (ILS) (Song et al., 2012), long branch attraction (LBA)
70 (Romiguier et al., 2013), and model misspecification (Morgan et al., 2013; Tarver et al., 2016).
71 Notwithstanding, assessments of this topological conflict adopting heterogeneous substitution
72 models (Morgan et al., 2013; Tarver et al., 2016) and summary coalescent methods (Song et al.,
73 2012; Tarver et al., 2016) found robust support for the Atlantogenata hypothesis, which was also
74 recovered when previously published datasets were re-examined with best-fitting models (Tarver et
75 al., 2016).

76 Despite these results, controversy remains, as methods employed in previous analyses are
77 not free of biases from different sources. For instance, it has been shown that, when sequences are
78 concatenated for carrying out supermatrix analysis, only a small fraction of sites may be responsible
79 for choosing between alternative topologies (Shen et al., 2017). Moreover, it has been demonstrated

80 that the likelihood function used in phylogenetics is inappropriate when genes are concatenated,
81 resulting in positively misleading estimates under high levels of gene trees/species tree
82 incongruence (Degnan and Rosenberg, 2006; Kubatko and Degnan, 2007; Mendes and Hahn, 2018;
83 Roch and Steel, 2015), which might be the case of the early placental split (Song et al., 2012; Tarver
84 et al., 2016). On the other hand, methods that avoid concatenation by accounting for the long-term
85 effects of ILS were criticized in the context of placental phylogeny (Springer and Gatesy, 2016).
86 The criticism lies on the fact that summary coalescent-based phylogenetic methods treat gene trees
87 as error-free observations (Song et al., 2012). As maximum likelihood trees are point estimates,
88 treating them as observations imply overconfidence, which may bias phylogeny estimation by
89 approaches derived from the multispecies coalescent (MSC) model (Xi et al., 2015). This issue has
90 been largely ignored by analyses of the placental tree relying on coalescent-based methods
91 (McCormack et al., 2012; Song et al., 2012; Tarver et al., 2016). It is thus conceivable that previous
92 studies on placental phylogeny were impacted by genes with ambiguous signal for resolving the
93 early divergences.

94 Following this reasoning, in this work, we were prompted to investigate, under the MSC
95 model, which phylogenetic hypothesis regarding the early placental split was significantly favored
96 when only subsets of phylogenetically informative genes were employed. We examined the
97 mammalian phylogeny from two subsets of genes trees that either presented unequivocal
98 phylogenetic resolution for the early placental split or consisted of topologies that did not overfit
99 data.

100

101 **Material and Methods**

102 Our analysis was carried out using 14,526 orthologous coding sequences retrieved from the
103 OrthoMam v9 data base (Douzery et al., 2014), which were filtered to consist solely of genes whose
104 maximum likelihood (ML) tree contained an outgroup and at least one representative of each
105 placental superorder, namely, Xenarthra, Afrotheria, Laurasiatheria and Euarchontoglires.
106 Furthermore, to reduce the phylogenetic problem to a 4-terminals tree, which makes the topological
107 space computationally accessible, only ML trees wherein superorders were recovered as
108 monophyletic were kept. We thus focused on the phylogenetic resolution of the early placental
109 divergence, allowing us to evaluate all the three possible topological hypotheses (**Figure 1**). After
110 this filtering step, our dataset was composed of 5,862 alignments, hereafter referred to as the *full*
111 dataset.

112 Gene tree estimation and choice of substitution models were performed with the IQ-TREE
113 software (Kalyaanamoorthy et al., 2017; Nguyen et al., 2014). The statistical support for the three
114 topological configurations at the root was assessed by the approximately-unbiased (AU) test
115 (Shimoidara, 2002), also implemented in IQ-TREE. Genes with unequivocal phylogenetic signal,
116 i.e., those with exclusive support for a single topology, were grouped, constituting the *AU* dataset.
117 Finally, we also built a dataset encompassing genes that, under the corrected Akaike information
118 criterion (AICc) (Hurvich et al., 1998), rejected the hypothesis of a polytomy between the
119 mammalian superorders, suggesting a strictly bifurcating tree as the best fit model. Species tree
120 inference was conducted independently for each dataset (*full*, *AU*, and *AICc*) employing the
121 coalescent-based algorithm implemented in the Astral-III software (Zhang et al., 2017), which also
122 performs hypothesis testing of polytomies according to the MSC model expectation (Sayyari and
123 Mirarab, 2018). A significance level of 5% was used throughout this study. Astral was also used to
124 estimate branch lengths in coalescent units, which is relevant for interpreting gene tree distribution
125 in light of the MSC, in particular to investigate whether the branching pattern that defines the deep

126 splits between the major lineages of placentals lies in the anomaly zone (Degnan and Rosenberg,
127 2006).

128

129 **Results**

130 The distribution of ML gene trees supporting each of the three possible phylogenetic
131 hypotheses of the early placental split was approximately equivalent in the *full* dataset. Out of 5,862
132 gene trees, 1,910 (32.6%) supported the Atlantogenata hypothesis, whereas 2,045 (34.9%)
133 supported Afrotheria, and 1,907 (32.5%) favored the Xenarthra tree. Therefore, the majority of gene
134 trees, as well as MSC-topology from Astral analysis, supported the Afrotheria tree. The coalescent-
135 based polytomy test also rejected the null hypothesis of a polytomic topology for this distribution of
136 gene trees ($p = 0.04$). However, when each gene from the *full* dataset was subjected to the AU test,
137 we found that only 101 genes (0.02%) presented unequivocal phylogenetic signal for a single
138 topology concerning the early placental split (the *AU* dataset). Each of the three possible topological
139 associations were represented with the following frequencies among these gene trees: 37.6%
140 supporting Atlantogenata, 22.8% Afrotheria and 39.6% Xenarthra. When Astral analysis was run
141 using the *AU* dataset, the Xenarthra topology was inferred as the species tree, although the null
142 hypothesis of a polytomy at the root of placental mammals was not rejected ($p = 0.08$).

143 According to the AICc test, 2,079 gene trees, nearly 35% of the *full* data set, overfit data.
144 The distribution of topologies inferred from the remaining 3,782 genes that favored a strictly
145 bifurcating tree as the best fit model (the *AICc* dataset) presented a similar frequency of each early
146 placental phylogenetic hypothesis, namely, Atlantogenata: 1,259 (33.3%), Afrotheria: 1,308
147 (34.6%), and Xenarthra: 1,216 (32.2%). Astral inferred the Afrotheria as the placental species tree
148 using the *AICc* dataset, but the hypothesis of a polytomic association among superorders was also
149 not rejected ($p = 0.29$). Branch lengths establishing the pattern of the early placental split in the *full*,
150 *AU*, and *AICc* datasets measured 0.0235, 0.0924, and 0.0167 coalescent units respectively (**Table**
151 **1**).

152

153

154 **Discussion**

155 Our results indicated that when genes with significant phylogenetic signal were used,
156 Afrotheria or Xenarthra were the likely sister lineages of the remaining placentals. However, in both
157 datasets, the hypothesis of a polytomy at the root of the placental tree was not rejected. When it
158 comes to phylogenomics, all three phylogenetic hypotheses depicted in **Figure 1** were obtained by
159 combining different methods and datasets. For instance, support for the Afrotheria tree was found
160 when datasets less prone to LBA artifacts were employed (Romiguier et al., 2013), and when
161 ultraconserved elements were analyzed (McCormack et al., 2012). With a single exception (Kriegs
162 et al., 2006), the Xenarthra tree has been rarely recovered by molecular data, although it stands as
163 the preferred hypothesis of morphologists (O’Leary et al., 2013). In our analyses, we found no
164 support for the Atlantogenata tree. This hypothesis was recently endorsed by Tarver *et al.* 2016,
165 who implemented MSC methods and argued that correcting for heterogeneous substitution
166 processes supports Atlantogenata as sister lineage to the remaining placentals. On the other hand,
167 most previous works have relied on concatenated supermatrices (Hallström et al., 2007; Hallström
168 and Janke, 2008; McCormack et al., 2012; Morgan et al., 2013; Nikolaev et al., 2007; Nishihara et
169 al., 2007; Romiguier et al., 2013; Wildman et al., 2007), which may be affected by the few-
170 sites/large-effect issue (Shen et al., 2017). In face of these conflicting results, our study indicates
171 that this lack of resolution for the early placental split might be due to the failure of the distribution
172 of topologies from genes with significant phylogenetic signal to reject a polytomic association
173 among mammalian superorders.

174 In this sense, it is meaningful to investigate whether the branching pattern of the early
175 placental split lies in the anomaly zone, the region of parameter space where the most common gene
176 tree is incongruent with the species tree (Degnan and Rosenberg, 2006). Even though it has been
177 argued that the anomaly zone is unlikely to be a problem when studying empirical datasets (Huang
178 and Knowles, 2009), the branch lengths supporting resolutions for the early placental split in all

179 species trees inferred here were shorter than ~ 0.156 coalescent units (**Table 1**). Therefore, the
180 theoretical possibility of obtaining anomalous gene trees distribution exists when investigating
181 supraordinal relationship of placental mammals. Under this scenario, phylogenetic inference from
182 concatenated sequences will be biased (Degnan and Rosenberg, 2006; Kubatko and Degnan, 2007;
183 Mendes and Hahn, 2018; Roch and Steel, 2015).

184 By restricting our analysis to subsets of genes with significant phylogenetic signal, our study
185 also avoided setbacks of summary coalescent methods (e.g., Astral) when gene tree uncertainty is
186 not accounted for. It is worth noting that, although Tarver *et al.* (2016) implied that few genes have
187 unequivocal signal for resolving the placental root by likelihood comparisons, which we
188 corroborated via proper topological tests, the authors did not restrict their MSC phylogeny inference
189 to this subset of unequivocal genes. Instead, they controlled for errors in gene trees by filtering
190 topologies with non-parametric bootstrap thresholds (Felsenstein, 1985). The bootstrap, however, is
191 not a formal statistical test between alternative topologies, and such procedure may have resulted in
192 topological overconfidence.

193 The hypothesis of a polytomy at the root of placentals has been previously advocated by
194 studies relying on retroposon insertions as phylogenetic markers (Churakov *et al.*, 2009; Nishihara
195 *et al.*, 2009). Alternatively, using supermatrices, Hallström and Janke (2010) argued that the
196 phylogenetic relationship among placental superorders is best represented by a phylogenetic
197 network. While the proposition of a placental root polytomy is not entirely novel, this is the first
198 coalescent-based analysis supporting this view, contradicting previous phylogenomic studies that
199 were possibly affected by the very issues that motivated our reexamination of this problem
200 (Hallström *et al.*, 2007; Hallström and Janke, 2008; McCormack *et al.*, 2012; Morgan *et al.*, 2013;
201 Nikolaev *et al.*, 2007; Nishihara *et al.*, 2007; Romiguier *et al.*, 2013; Song *et al.*, 2012; Tarver *et al.*,
202 2016; Wildman *et al.*, 2007). Our results were not able to devise whether the placental root is the
203 case of a soft (dataset limited) or hard (real) polytomy (Whitfield and Lockhart, 2007), but we

204 showed how the prior examination of the historical signal of genes, associated with a control for
205 systematic biases, impacts the resolution of a hard phylogenetic problem.

206 As conservative as our approach was, it illustrated how the MSC model may be used to test
207 for polytomies in empirical datasets (Edwards, 2009). Future application of this framework is
208 promising in unveiling rapid radiations in the deeper nodes of the tree of life (Whitfield and
209 Lockhart, 2007). As has been argued by Shen et al. (Shen et al., 2017), branches that fail to reject a
210 polytomy should be considered unresolved, implying a more accurate view of the signal available in
211 data. We here provided evidence that this is the case of the early split between superorders of
212 Placentalia.

213

214

215 **Acknowledgements**

216 This research was supported by grants 310974/2015-1 and 440954/2016-9 from the Brazilian
217 Research Council (CNPq) to CGS.

218

219

220

221 **References**

222 Asher, R.J., 2007. A web-database of mammalian morphology and a reanalysis of placental
223 phylogeny. *BMC Evol. Biol.* 7, 1–10. <https://doi.org/10.1186/1471-2148-7-108>

224 Churakov, G., Kriegs, J.O., Baertsch, R., Zemann, A., Brosius, J., Schmitz, J., 2009. Mosaic
225 retroposon insertion patterns in placental mammals. *Genome Res.* 19, 868–875.
226 <https://doi.org/10.1101/gr.090647.108>

227 Degnan, J.H., Rosenberg, N.A., 2006. Discordance of Species Trees with Their Most Likely Gene
228 Trees. *Plos Genet.* 2. <https://doi.org/10.1371/journal.pgen.0020068>

229 Douzery, E.J.P., Scornavacca, C., Romiguier, J., Belkhir, K., Galtier, N., Delsuc, F., Ranwez, V.,
230 2014. OrthoMaM v8 : A Database of Orthologous Exons and Coding Sequences for
231 Comparative Genomics in Mammals. *Mol Biol Evol* 31, 1923–1928.
232 <https://doi.org/10.1093/molbev/msu132>

233 Edwards, S. V, 2009. IS A NEW AND GENERAL THEORY OF MOLECULAR SYSTEMATICS
234 EMERGING ? *Evolution* (N. Y). 1–19. <https://doi.org/10.1111/j.1558-5646.2008.00549.x>

- 235 Felsenstein, J.S., 1985. Confidence limits on phylogenies: An approach using the bootstrap.
236 *Evolution* (N. Y). 39, 783–791.
- 237 Hallström, B.M., Janke, A., 2010. Mammalian evolution may not be strictly bifurcating. *Mol. Biol.*
238 *Evol.* 27, 2804–2816. <https://doi.org/10.1093/molbev/msq166>
- 239 Hallström, B.M., Janke, A., 2008. Resolution among major placental mammal interordinal
240 relationships with genome data imply that speciation influenced their earliest radiations. *BMC*
241 *Evol. Biol.* 13, 1–13. <https://doi.org/10.1186/1471-2148-8-162>
- 242 Hallström, B.M., Kullberg, M., Nilsson, M.A., Janke, A., 2007. Phylogenomic data analyses
243 provide evidence that Xenarthra and Afrotheria are sister groups. *Mol. Biol. Evol.* 24, 2059–
244 2068. <https://doi.org/10.1093/molbev/msm136>
- 245 Huang, H., Knowles, L.L., 2009. What Is the Danger of the Anomaly Zone for Empirical
246 Phylogenetics? *Syst. Biol.* 58, 527–536. <https://doi.org/10.1093/sysbio/syp047>
- 247 Hurvich, C.M., Simonoff, J.S., Tsai, C., 1998. Information Criterion Smoothing parameter selection
248 in nonparametric regression using an improved Akaike information criterion. *J. R. Stat. Soc.*
249 *Ser. B (Statistical Methodol.* 60, 271–293. <https://doi.org/10.1109/ISSPA.2005.1581081>
- 250 Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., Haeseler, A. Von, Jermini, L.S., 2017.
251 ModelFinder : fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14.
252 <https://doi.org/10.1038/nmeth.4285>
- 253 Kjer, K.M., Honeycutt, R.L., 2007. Site specific rates of mitochondrial genomes and the phylogeny
254 of eutheria. *BMC Evol. Biol.* 7, 1–9. <https://doi.org/10.1186/1471-2148-7-8>
- 255 Kriegs, J.O., Churakov, G., Kiefmann, M., Jordan, U., 2006. Retroposed Elements as Archives for
256 the Evolutionary History of Placental Mammals. *PLoS Biol.* 4.
257 <https://doi.org/10.1371/journal.pbio.0040091>
- 258 Kubatko, L.S., Degnan, J.H., 2007. Inconsistency of Phylogenetic Estimates from Concatenated
259 Data under Coalescence. *Syst. Biol.* 56, 17–24. <https://doi.org/10.1080/10635150601146041>
- 260 McCormack, J.E., Faircloth, B.C., Crawford, N.G., McCormack, J.E., Faircloth, B.C., Crawford,
261 N.G., Gowaty, P.A., Brumfield, R.T., Glenn, T.C., 2012. Ultraconserved elements are novel
262 phylogenomic markers that resolve placental mammal phylogeny when combined with
263 species-tree analysis. *Genome Res.* <https://doi.org/10.1101/gr.125864.111>
- 264 Mendes, F.K., Hahn, M.W., 2018. Why Concatenation Fails Near the Anomaly Zone. *Syst. Biol.* 67,
265 158–169. <https://doi.org/10.1093/sysbio/syx063>
- 266 Meredith, R.W., Janecka, J.E., Gatesy, J., Ryder, O.A., Fisher, C.A., Teeling, E.C., Goodbla, A.,
267 Eizirik, E., Simao, T.L.L., Stadler, T., Rabosky, D.L., Honeycutt, R.L., Flynn, J.J., Ingram,
268 C.M., Steiner, C., Williams, T.L., Robinson, T.J., Burk-Herrick, A., Westerman, M., Ayoub,
269 N.A., Springer, M.S., Murphy, W.J., 2011. Impacts of the Cretaceous Terrestrial Revolution
270 and KPg Extinction on Mammal Diversification. *Science* (80-.). 334, 521–524.
271 <https://doi.org/10.1126/science.1211028>

- 272 Morgan, C.C., Foster, P.G., Webb, A.E., Pisani, D., Mcinerney, J.O., Connell, M.J.O., 2013.
273 Heterogeneous Models Place the Root of the Placental Mammal Phylogeny. *Mol Biol Evol* 30,
274 2145–2156. <https://doi.org/10.1093/molbev/mst117>
- 275 Murphy, W.J., Eizirik, E., Johnson, W.E., 2001. Molecular phylogenetics and the origins of
276 placental mammals. *Nature* 409.
- 277 Murphy, W.J., Pringle, T.H., Crider, T.A., Springer, M.S., Miller, W., 2007. Using genomic data to
278 unravel the root of the placental mammal phylogeny. *Genome Res.* 413–421.
279 <https://doi.org/10.1101/gr.5918807.17>
- 280 Nguyen, L., Schmidt, H.A., Haeseler, A. Von, Minh, B.Q., 2014. IQ-TREE : A Fast and Effective
281 Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol* 32,
282 268–274. <https://doi.org/10.1093/molbev/msu300>
- 283 Nikolaev, S., Montoya-burgos, J.I., Margulies, E.H., Comparative, N., Program, S., 2007. Early
284 History of Mammals Is Elucidated with the ENCODE Multiple Species Sequencing Data. *Plos*
285 *Genet.* 3. <https://doi.org/10.1371/journal.pgen.0030002>
- 286 Nishihara, H., Maruyama, S., Okada, N., 2009. Retroposon analysis and recent geological data
287 suggest near-simultaneous divergence of the three superorders of mammals. *Proc. Natl. Acad.*
288 *Sci.* 106, 5235–5240. <https://doi.org/10.1073/pnas.0809297106>
- 289 Nishihara, H., Okada, N., Hasegawa, M., 2007. Rooting the eutherian tree : the power and pitfalls of
290 phylogenomics. *Genome Biol.* 8. <https://doi.org/10.1186/gb-2007-8-9-r199>
- 291 O’Leary, M.A., Leary, M.A.O., Bloch, J.I., Flynn, J.J., Gaudin, T.J., Giallombardo, A., Giannini,
292 N.P., Goldberg, S.L., Kraatz, B.P., Luo, Z., Meng, J., Ni, X., Novacek, M.J., Perini, F.A.,
293 Randall, Z.S., Rougier, G.W., Sargis, E.J., Silcox, M.T., Simmons, N.B., Spaulding, M.,
294 Velazco, P.M., Weksler, M., Wible, J.R., Cirranello, A.L., 2013. The Placental Mammal
295 Ancestor and the Post-K-Pg Radiation of Placentals. *Science* (80-.). 662.
296 <https://doi.org/10.1126/science.1229237>
- 297 Prasad, A.B., Allard, M.W., Green, E.D., 2008. Confirming the phylogeny of mammals by use of
298 large comparative sequence data sets. *Mol. Biol. Evol.* 25, 1795–1808.
299 <https://doi.org/10.1093/molbev/msn104>
- 300 Roch, S., Steel, M., 2015. Likelihood-based tree reconstruction on a concatenation of aligned
301 sequence data sets can be statistically inconsistent. *Theor. Popul. Biol.* 100, 56–62.
302 <https://doi.org/10.1016/j.tpb.2014.12.005>
- 303 Romiguier, J., Ranwez, V., Delsuc, F., Galtier, N., J.P.Douzery, E., 2013. Less Is More in
304 Mammalian Phylogenomics : AT-Rich Genes Minimize Tree Conflicts and Unravel the Root of
305 Placental Mammals. *Mol Biol Evol* 30, 2134–2144. <https://doi.org/10.1093/molbev/mst116>
- 306 Sayyari, E., Mirarab, S., 2018. Testing for polytomies in phylogenetic species trees using quartet
307 frequencies. *arXiv:1708.08491v4*. <https://doi.org/10.3390/GENES9030132>

- 308 Shen, X., Hittinger, C.T., Rokas, A., 2017. Contentious relationships in phylogenomic studies can
309 be driven by a handful of genes. *Nat. Publ. Gr.* 1, 1–10. [https://doi.org/10.1038/s41559-017-](https://doi.org/10.1038/s41559-017-0126)
310 0126
- 311 Shimoidara, H., 2002. An Approximately Unbiased Test of Phylogenetic Tree Selection. *Syst. Biol.*
312 51, 492–508. <https://doi.org/10.1080/10635150290069913>
- 313 Song, S., Liu, L., Edwards, S. V, Wu, S., 2012. Resolving conflict in eutherian mammal phylogeny
314 using phylogenomics and the multispecies coalescent model. *Proc. Natl. Acad. Sci.* 112.
315 <https://doi.org/10.1073/pnas.1518753112>
- 316 Springer, M.S., Gatesy, J., 2016. The gene tree delusion. *Mol. Phylogenet. Evol.* 94, 1–33.
317 <https://doi.org/10.1016/j.ympev.2015.07.018>
- 318 Tarver, J.E., Reis, M., Mirarab, S., Moran, R.J., Parker, S., Reilly, J.E.O., King, B.L., Connell,
319 M.J.O., Asher, R.J., Warnow, T., Peterson, K.J., Donoghue, P.C.J., Pisani, D., 2016. The
320 Interrelationships of Placental Mammals and the Limits of Phylogenetic Inference. *Genome*
321 *Biol. Evol.* 8, 330–344. <https://doi.org/10.1093/gbe/evv261>
- 322 Teeling, E.C., Hedges, S.B., 2013. Making the Impossible Possible : Rooting the Tree of Placental
323 Mammals. *Mol Biol Evol* 30, 1999–2000. <https://doi.org/10.1093/molbev/mst118>
- 324 Waddell, P.J., Kishino, H., Ota, R., 2001. A Phylogenetic Foundation for Comparative Mammalian
325 Genomics. *Genome Informatics* 154, 141–154.
- 326 Whitfield, J.B., Lockhart, P.J., 2007. Deciphering ancient rapid radiations. *Trends Ecol. Evol.* 22.
327 <https://doi.org/10.1016/j.tree.2007.01.012>
- 328 Wildman, D.E., Uddin, M., Opazo, J.C., Liu, G., Lefort, V., Guindon, S., Gascuel, O., Grossman,
329 L.I., Romero, R., Goodman, M., 2007. Genomics , biogeography , and the diversification of
330 placental mammals. *Proc. Natl. Acad. Sci.* 104.
- 331 Xi, Z., Liu, L., Davis, C.C., 2015. Genes with minimal phylogenetic information are problematic
332 for coalescent analyses when gene tree estimation is biased. *Mol. Phylogenet. Evol.* 92, 63–71.
333 <https://doi.org/10.1016/j.ympev.2015.06.009>
- 334 Zhang, C., Sayyari, E., Mirarab, S., 2017. ASTRAL-III: Increased Scalability and Impacts of
335 Contracting Low Support Branches. *Comp. Genomics. RECOMB-CG 2017* 10562, 53–75.
336 <https://doi.org/10.1007/978-3-319-67979-2>
337
338
339

340 **Table 1.** Summary of results from analyses employing three different filters on the alignments from
341 OrthoMam.

Dataset	Number of genes	Gene tree distribution (Atlg : Afro : Xnrt)	Astral species tree	Null hypothesis of polytomy	Branch length in coalescent units
Full	5,862	32.6 : 34.9 : 32.5	Afrotheria	Rejected ($p = 0.04$)	0.0235
AU	101	37.6 : 22.8 : 39.6	Xenarthra	Not rejected ($p = 0.08$)	0.0924
AICc	3,782	33.3 : 34.6 : 32.2	Afrotheria	Not rejected ($p = 0.3$)	0.0167

342

343 **Figure legend.**

344 **Figure 1.** The three phylogenetic hypotheses for the position of the root of placental mammals (a).
345 Distribution of gene tree topologies in the *full*, *AICc*, and *AU* datasets, with their respective species
346 tree resolution inferred with Astral-III. The *p*-value of the polytomy test (above) and lengths in
347 coalescent units (bellow) were given for branches responsible for resolving the polytomy at the root
348 (b).

