Assessing Confidence in Root Placement on Phylogenies:

An Empirical Study Using Non-Reversible Models for

Mammals

SUHA NASER-KHDOUR*[1], BUI QUANG MINH[1,2], AND ROBERT LANFEAR[1]

[1] DEPARTMENT OF ECOLOGY AND EVOLUTION, RESEARCH SCHOOL OF BIOLOGY,

AUSTRALIAN NATIONAL UNIVERSITY, CANBERRA, AUSTRALIAN CAPITAL TERRITORY,

AUSTRALIA

[2] RESEARCH SCHOOL OF COMPUTER SCIENCE, AUSTRALIAN NATIONAL UNIVERSITY,

CANBERRA, AUSTRALIAN CAPITAL TERRITORY, AUSTRALIA

*Author for Correspondence: E-mail: suha.naser@anu.edu.au

ABSTRACT

Using time-reversible Markov models is a very common practice in phylogenetic analysis, because although we expect many of their assumptions to be violated by empirical data, they provide high computational efficiency. However, these models lack the ability to infer the root placement of the estimated phylogeny. In order to compensate for the inability of these models to root the tree, many researchers use external information such as using outgroup taxa or additional assumptions such as molecular-clocks. In this study, we investigate the utility of non-reversible models to root empirical phylogenies and introduce a new bootstrap measure, the *rootstrap*, which provides information on the statistical support for any given root position.

Availability and implementation: rootstrap support is implemented in IQ-TREE 2 and a tutorial is available at the iqtree webpage http://www.iqtree.org/doc/Rootstrap. In addition, a python script is available at https://github.com/suhanaser/Rootstrap.

23   [phylogenetic inference, root estimation, bootstrap, non-reversible models]

24

25   MAIN TEXT

26   The most widely used method for rooting trees in phylogenetics is the outgroup method.

27   Although the use of an outgroup to root an unrooted phylogeny usually outperforms other

28   rooting methods (Huelsenbeck, et al. 2002), the main challenge with this method is to find an

29   appropriate outgroup (Watrous and Wheeler 1981; Maddison, et al. 1984; Smith 1994;

30   Swofford, et al. 1996; Lyons-Weiler, et al. 1998; Milinkovitch and Lyons-Weiler 1998).

31   Outgroups that are too distantly-related to the ingroup may have substantially different

32   molecular evolution than the ingroup, which can compromise accuracy. And outgroups that are

33   too closely related to the ingroup may not be valid outgroups at all.

34   It is possible to infer the root of a tree without an outgroup using molecular clocks

35   (Huelsenbeck, et al. 2002; Drummond, et al. 2006). A strict molecular-clock assumes that the

36   substitution rate is constant along all lineages, a problematic assumption especially when the

37   ingroup taxa are distantly related such that their rates of molecular evolution may vary.

38   Relaxed molecular-clocks are more robust to deviations from the clock-like behaviour

39   (Drummond, et al. 2006), although previous studies have shown that they can perform poorly

40   in estimating the root of a phylogeny when those deviations are considerable (Tria, et al. 2017).

41   Other rooting methods rely on the distribution of branch lengths, including Midpoint

42   Rooting (MPR) (Farris 1972), Minimal Ancestor Deviation (MAD) (Tria, et al. 2017), and

43   Minimum Variance Rooting (MVR) (Mai, et al. 2017). Such methods also assume a clock-like

44   behaviour; however, they are less dependent on this assumption as the unrooted tree is

45   estimated without it. Similar to inferring a root directly from molecular-clock methods, the

46 accuracy of those rooting methods decreases with higher deviations from the molecular-clock

47 assumption (Mai, et al. 2017).

48   Other less common rooting methods that can be used in the absence of outgroup are:

49 rooting by gene duplication (Dayhoff and Schwartz 1980; Gogarten, et al. 1989; Iwabe, et al.

50 1989), indel-based rooting (Rivera and Lake 1992; Baldauf and Palmer 1993; Lake, et al.

51 2007), rooting the species tree from the distribution of unrooted gene trees (Allman, et al. 2011;

52 Yu, et al. 2011), and probabilistic co-estimation of gene trees and species tree (Boussau, et al.

53 2013).

54   All the methods mentioned above, apart from the molecular-clock, infer the root

55 position independently of the ML tree inference. The only existing approach to include root

56 placement in the ML inference is the application of non-reversible models. Using non-

57 reversible substitution models relaxes the fundamental assumption of time-reversibility that

58 exists in the most widely used models in phylogenetic inference (Jukes and Cantor 1969;

59 Kimura 1980; Hasegawa, et al. 1985; Tavaré 1986; Dayhoff 1987; Jones, et al. 1992; Tamura

60 and Nei 1993; Whelan and Goldman 2001; Le and Gascuel 2008). This in itself is a potentially

61 useful improvement in the fit between models of sequence evolution and empirical data. In

62 addition, since non-reversible models naturally incorporate a notion of time, the position of the

63 root on the tree is a parameter that is estimated as part of the ML tree inference. Since the

64 incorporation of non-reversible models in efficient ML tree inference software is relatively new

65 (Minh, et al. 2020), we still understand relatively little about the ability of non-reversible

66 models to infer the root of a phylogenetic tree, although a recent simulation study has shown

67 some encouraging results (Bettisworth and Stamatakis 2020).

68   Regardless of the rooting method and the underlying assumptions, it is crucial that we

69 are able to estimate the statistical confidence we have in any particular placement of the root

70   on a phylogeny. A number of previous studies have sensibly used ratio likelihood tests such as

71   the Shimodaira-Hasegawa (SH) test (Shimodaira and Hasegawa 1999) and the Approximately

72   Unbiased (AU) test (Shimodaira 2002) to compare a small set of potential root placements,

73   rejecting some alternative root placements in favour of the ML root placement e.g.(Nardi, et

74   al. 2003; Steenkamp, et al. 2006; Jansen, et al. 2007; Moore, et al. 2007; Williams, et al. 2010;

75   Kocot, et al. 2011; Zhou, et al. 2011; Whelan, et al. 2015; Zhang, et al. 2018), these tests are

76   still somewhat limited in that they do not provide the level of support the data have for a certain

77   root position.

78   There is strong empirical evidence that molecular evolutionary processes are rarely

79   reversible (Squartini and Arndt 2008; Naser-Khdour, et al. 2019), but few studies have

80   explored the accuracy of non-reversible substitution models to root phylogenetic trees

81   (Huelsenbeck, et al. 2002; Yap and Speed 2005; Williams, et al. 2015; Cherlin, et al. 2018;

82   Bettisworth and Stamatakis 2020). Most studies that have looked at this question in the past

83   have focused on either simulated datasets (Huelsenbeck, et al. 2002; Jayaswal, et al. 2011;

84   Cherlin, et al. 2018) or relatively small empirical datasets (Yang and Roberts 1995; Yap and

85   Speed 2005; Jayaswal, et al. 2011; Heaps, et al. 2014; Williams, et al. 2015; Cherlin, et al.

86   2018). In both cases, the addressed substitution models were nucleotide models, and to our

87   knowledge, no study has yet investigated the potential of amino acid substitution models in

88   inferring the root placement of phylogenetic trees.

89   In this paper, we focus on evaluating the utility of non-reversible amino acid and

90   nucleotide substitution models to root the trees, and we introduce a new metric, the *rootstrap*

91   *support value*, which estimates the extent to which the data support every possible branch as

92   the placement of a root in a phylogenetic tree. Unlike previous studies that used Bayesian

93   methods with non-reversible substitution models to infer rooted ML trees (Heaps, et al. 2014;

94   Cherlin, et al. 2018), we will conduct our study in a Maximum likelihood framework using IQ-

95 TREE (Minh, et al. 2020). A clear advantage of Maximum likelihood over the Bayesian

96 analysis is that there is no need for a prior on the parameter distributions, which sometimes can

97 affect tree inference (Huelsenbeck, et al. 2002; Cherlin, et al. 2018). Even though estimating

98 the non-reversible model's parameters by maximizing the likelihood function seems more

99 computationally intensive than calculating posterior probabilities (Huelsenbeck, et al. 2002),

100 the IQ-TREE algorithm is sufficiently fast to allow us to estimate root placements, with

101 *rootstrap support* for very large datasets.

102 A recent study investigated the ability of non-reversible nucleotide models to infer the

103 root placement of phylogenetic trees (Bettisworth and Stamatakis 2020). This study showed

104 that IQ-TREE performs competitively with a new rooting tool, RootDigger. In most simulated

105 datasets, IQ-TREE slightly outperformed RootDigger in terms of root placements, but no

106 comparisons were made between RootDigger and IQ-TREE on empirical datasets. Although,

107 RootDigger is significantly faster than IQ-TREE (Bettisworth and Stamatakis 2020), the

108 former is limited to nucleotide substitution models. Since we are interested in both nucleotide

109 and amino acid non-reversible models, we used IQ-TREE for tree and root inference in this

110 study.

111 MATERIAL AND METHODS

112 *The "Rootstrap" Support, and measurements of error in root placement*

113 To compute rootstrap supports, we conduct a bootstrap analysis, i.e., resampling alignment

114 sites with replacement, to obtain a number of bootstrap trees. We define the *rootstrap* support

115 for each branch in the ML tree, as the proportion of bootstrap trees that have the root on that

116 branch. Since the root can be on any branch in a rooted tree, the rootstrap support values are

117 computed for all the branches including external branches. The sum of the rootstrap support

118 values along the tree are always smaller than or equal to one. A sum that is smaller than one

119   can occur when one or more bootstrap replicates are rooted on a branch that does not occur in

120   the ML tree (Fig. 1).
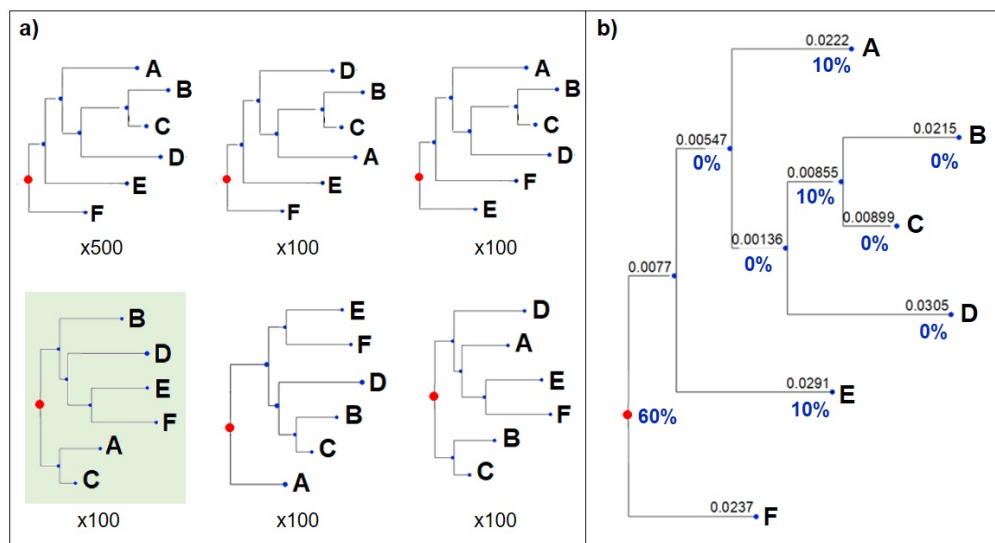

121

122   FIGURE 1.   **Illustration of the rootstrap concept.** (a) The bootstrap replicates trees. (b)
123   The ML tree with the rootstrap support values for each branch. Note that the sum of the
124   rootstrap support values is less than 100% due to 100 bootstrap replicates trees (green) that
125   have their root at a branch that does not exist in the ML tree.

126

127   By definition, the rootstrap support values for internal branches are bounded by the

128   bootstrap support values at those branches. On the other hand, the rootstrap support values for

129   tips (leaf branches) are bounded by 100%, as tips always appear in all the bootstrap trees.

130   If the true position of the root is known (e.g. in simulation studies) or assumed (e.g. in

131   the empirical cases we present below), we can calculate additional measurements of the error

132   of the root placement. We introduce two such measurements here: *root branchlength error*

133   *distance* (rBED) and *root split error distance* (rSED). Since the non-reversible model infers

134   the exact position of the root on a branch, we define the *root branchlength error distance*

135   (rBED) as the range between the minimum and maximum distance between the inferred root

136   position and the "true root" branch. If the true root is on the same branch as the ML tree root,

137   then rBED will be between 0 and the distance between the ML tree root and the farthest point

138   on that branch (Fig. 2). Since rBED is based on branch lengths only, it ignores the absolute

139   number of splits between the ML tree root and the true root; and therefore, the rBED for the

140   true root being on the same ML root branch can be bigger than the rBED for the true root

141   being on a different branch (e.g. Fig. 2). In order to account for the number of splits (nodes)

142   between the ML tree root and the true root, we define *root split error distance* (rSED) as the

143   number of splits between the ML root branch and the branch that is believed to contain the
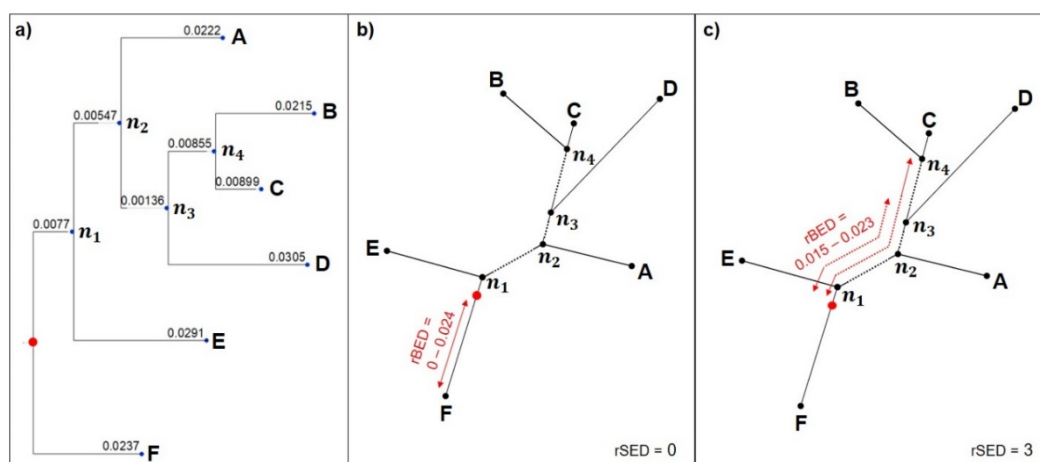
144   true root (Fig. 2).

145

146   FIGURE 2.   An example to illustrate the root error distance. (a) the ML rooted tree, (b) the
147   root branch-length error distance (rBED) if the true root is believed to be on the same ML
148   root branch (rSED = 0), (c) the rBED if the true root is believed to be on the branch between
149   D and the clade of C + B (rSED = 3).

150

151   The rootstrap, rBED, and rSED assess different aspects of the root placement. While the

152   rootstrap offers an indication of the support that the data have for a certain branch to be the

153   root branch, rBED and rSED provide an estimation to the accuracy of the method in

154   estimating the exact root position if the root position is known or assumed in advance. In

155   other words, the rootstrap value is a measure for the robustness of the root placement given

156   the model and the data and can be used on any dataset regardless of whether the true root

157   position is known, while rBED and rSED are measures of the accuracy of the non-reversible

158   model to find the root placement given the data, and require the root position to be known or

159   assumed in advance.

Empirical Datasets

161   Because non-reversible amino acid models require the estimation of a large number of

162   parameters, and because we suspected that the information in any such analysis on the

163   placement of the root branch of a tree might be rather limited, we searched for empirical

164   datasets that met a number of stringent criteria:

165   (1) Existence of both DNA and amino acid multiple sequences alignments (MSA) for the

166   same loci.

167   (2) Genome-scale MSAs to ensure that the MSAs have as much information as possible with

168   which to estimate the non-reversible models' free parameters and the root position. Since

169   we do not know the number of sites required to correctly infer the rooted ML tree, we

170   define 100,000 sites as the minimum number of required sites. This also allows us to

171   subsample the dataset to explore the ability of smaller datasets to infer root positions.

172   (3) Highly-curated alignments: since the quality of the inferred phylogeny is highly

173   dependent on the quality of the MSA (Philippe, et al. 2011), we focussed on datasets that

174   were highly-curated for misalignment, contamination, and paralogy.

175   (4) Existence of several clades for which there is a very strong consensus regarding their root

176   placement. Since we are interested in evaluating the performance of non-reversible

177   models to infer root placements in an empirical rather than a simulation context, we need

178   to identify monophyletic sub-clades for which we can be almost certain about their root

179   position. This enables us to divide the dataset into non-overlapping sub-clades for which

180   we are willing to assume we know the root positions. Furthermore, we define the

181   minimum number of taxa in each sub-dataset as five.

ROOTSTRAP SUPPORT

182   We initially identified a number of genome-scale datasets that contained large numbers of

183   nucleotide and amino acid MSAs. In many cases, it was difficult to determine whether these

184   alignments had been rigorously curated, and even more challenging to find datasets for which

185   the root position of a number of subclades could be assumed with confidence. The only dataset

186   that met all of our criteria was a dataset of placental mammals with 78 ingroup taxa and

187   3,050,199 amino acids (Wu, et al. 2019). This dataset was originally published as an MSA

188   (Liu, et al. 2017) based on very high-quality sequences from Ensembl, NCBI, and GenBank

189   databases. After receiving detailed critiques for potential alignment errors (Gatesy and Springer

190   2017), the dataset was further processed to remove potential sources of bias and error, and an

191   updated version of the dataset was recently published (Wu, et al. 2018). The fact that this

192   alignment comes from one of the most well-studied clades on the planet, has been

193   independently curated and critiqued by multiple groups of researchers and includes truly

194   genome-scale data, makes it ideally suited for our study. The curated alignments can be found

195   on figshare (https://figshare.com/s/622e9e0a156e5233944b) under the name "Wu_2018_aa"

196   and "Wu_2018_dna" for the amino-acid and nucleotide alignments respectively.

197

198   *Selecting Clades with a Well-Defined Root*

199   Since our main objective in this study is to evaluate the effectiveness of non-reversible

200   models and the rootstrap value in estimating and measuring the support for a given root

201   placement on empirical datasets, we must identify a collection of sub-clades of the larger

202   mammal dataset for which it is reasonable to assume a root position. We acknowledge, of

203   course, that outside a simulation framework it is not possible to be certain of the position of the

204   root position of a clade. Nevertheless, it is possible to identify clades for which the position of

205   the root is well supported and non-controversial, thus minimising the chances that the

206   assumption of a particular root position is incorrect. To achieve this, we analysed the root

9

207   position of each order and superorder in the dataset, and defined "*well-defined clades*" that

208   fulfilled **all** of the following criteria:

209   (1) It contains at least five taxa. This ensures that the probability of obtaining a random ML

210   rooted tree to be at most 0.95%. For clades with four taxa, there are 15 different rooted

211   topologies, and therefore a 6.7% probability to get any particular root position by chance.

212   On the other hand, for clades with at least five taxa, there are at least 105 different rooted

213   topologies and maximum probability of 0.95% to randomly get a particular root position

214   by chance.

215   (2) The bootstrap support for the branch leading to that clade in the phylogenetic tree

216   calculated from the whole dataset is 100%: since the bootstrap value indicates the

217   support the data have for a certain branch, we also require 100% support for the first

218   direct descendants in the clade (Appendix Fig. A.1). This requirement ensures that there

219   is strong support in the dataset for the root position of the clade when the entire dataset is

220   rooted with an outgroup.

221   (3) The site concordance factor (sCF) for the first direct descendants in the clade is

222   significantly greater than 33%. The site Concordance Factor (sCF) is calculated by

223   comparing the support of each site in the alignment for the different arrangements of

224   quartet around a certain branch. In other words, an sCF of 33% means equal support for

225   any of the possible arrangements. Therefore, we require that the sCF of the deepest two

226   levels of branches leading to that clade is significantly greater than 33%. Moreover, we

227   require that the gene Concordance Factor (gCF) for the first direct descendants in the

228   clade to be significantly greater than 33% of the sum of the gene concordance factor and

229   the two Discordance Factors (gDF1 and gDF2). The gCF of a branch is calculated as the

230   proportion of gene trees containing that branch, and gDFs are calculated as the

231   proportion of gene trees containing one of the two other resolutions of that branch. Since

232   for each branch in a bifurcating tree there are three possible arrangements of clades

233   around that branch, we ignore all gene trees that do not contain one of these

234   arrangements (e.g. gene trees that contribute to neither the gCF nor the gDFs). Although

235   there is no threshold regarding the required proportion of genes concordant with a certain

236   branch, for convenience, we define branches with gCF significantly greater than 33% of

237   the sum gCF+gDF1+gCF2 as branches that are concordant with enough genes in the

238   alignment (Minh, et al. 2020). To test whether the sCF and the gCF are significantly

239   greater than 33%, we use a simple binomial test with a success probability of 0.33. The

240   gCF,gDF1,gCF2 and sCF values are based on the tree estimated from the amino acid

241   dataset.

242   (4) At least 95% of the studies that have been published in the last decade support this clade:

243   we searched google scholar for all published papers since 2009 that determine the root of

244   the addressed clade. We then checked if at least 95% of those papers agree that the root

245   position of the clade matches that in the ML tree we estimate from the whole dataset (see

246   supplementary material).

247   *Estimating unrooted Phylogenies*

248   For the whole nucleotide and amino-acid datasets with ingroup and outgroup taxa, we

249   inferred the unrooted phylogeny using IQ-TREE2 (Minh, et al. 2020) with the best-fit fully

250   partitioned model (Chernomor, et al. 2016) and edge-linked substitution rates (Duchene, et al.

251   2020). We then determined the best-fit reversible model for each partition using ModelFinder

252   (Kalyaanamoorthy, et al. 2017). See the algorithm for finding well-defined clades in

253   Appendix Algorithm A.1.

11

254   *Estimating Rooted phylogenies*

255   For each well-defined clade, we first removed all other taxa from the tree and then

256   sought to infer the root of the well-defined clade using non-reversible models without

257   outgroups. Using the best partitioning scheme from the reversible analysis, we inferred the

258   rooted tree for each well-defined clade with the non-reversible models for amino acid (NR-

259   AA) and nucleotide (NR-DNA) sequences (Minh, et al. 2020). This approach fits a 12-

260   parameter non-reversible model for DNA sequences, and a 380-parameter non-reversible

261   model for amino acids. Details of the command lines used are provided in the supplementary

262   material section "Algorithm A.2". Each analysis returns a rooted tree. We performed 1000

263   non-parametric bootstraps of every analysis to measure the rootstrap support.

264   To assess the performance of the rootstrap and the ability of non-reversible models to

265   estimate the root of the trees on smaller datasets, we also repeated every analysis on

266   subsamples of the complete dataset. For each well-defined clade, we performed analysis on

267   the complete dataset (100%) as well as datasets with 10%, 1% and 0.1% of randomly-

268   selected loci from the original alignment.

269   *The confidence set of root branches using the Approximately Unbiased test*

270   In addition to the rootstrap support, we calculate the confidence set of all the branches

271   that may contain the root of the ML tree using the Approximately Unbiased (AU) test

272   (Shimodaira 2002). To do this, we re-root the ML tree with all possible placements of the root

273   (one placement for each branch) and calculate the likelihood of each tree. Using the AU test,

274   we then ask which root placements can be rejected in favour of the ML root, using an alpha

275   value of 5%. We define the *root branches confidence set* as the set of root branches that are not

276   rejected in favour of the ML root placement. An important difference between the AU test and

277   the rootstrap support is that the AU test is conditioned on a single ML tree topology, but the

278   rootstrap support is not. Because of this, they provide quite different information about the

279   position of the root. The AU test assumes that the ML tree topology is true, and then seeks to

280   determine the confidence set of root placements conditioned on that topology. The confidence

281   set for the AU test will always therefore contain at least the ML root branch. The rootstrap does

282   not assume any particular topology, and instead asks how many times a particular root position

283   appears across a set of bootstrap replicates. Because of this, it is possible for every branch in

284   the ML topology to receive 0% rootstrap support. This can occur if none of the branches in the

285   ML topology appear as the root branch in any of the bootstrap topologies.

286   *Reducing systematic bias by removing third codon positions and loci that fail the MaxSym*

287   *test*

288   As it is common in many phylogenetic analyses to remove third codon positions from

289   the alignment (Swofford, et al. 1996), we wanted to assess the effect of removing third codon

290   positions on the root inference and the rootstrap values in nucleotide datasets. For that

291   purpose, we remove all the third codon positions from the nucleotide alignments and re-ran

292   the analysis using the NR-DNA model.

293   Moreover, although the NR-AA and NR-DNA models relax the reversibility assumption,

294   they still assume stationarity and homogeneity. To reduce the systematic bias produced by

295   violating these assumptions, we used the MaxSym test (Naser-Khdour, et al. 2019) to remove

296   loci that violate those assumptions in the nucleotide and amino acid datasets, and then re-ran

297   all analyses as above.

298   *Applying the methods to two clades whose root position is uncertain*

299   In addition to the well-defined clades, we used the methods we propose here to infer

300   the root of two clades of mammals whose root position is controversial; Chiroptera and the

301   Cetartiodactyla.

13

302   There is a controversy around the root of the Chiroptera (bats) in literature. The two

303   most popular hypotheses are: 1) the Microchiroptera-Megachiroptera hypothesis; where the

304   root is placed between the Megachiroptera, which contains the family Pteropodidae, and the

305   Microchiroptera, which contains all the remaining Chiroptera families. This hypothesis is

306   well supported in the literature (Agnarsson, et al. 2011; Meredith, et al. 2011). However,

307   more recent studies seem to provide less support for this hypothesis; 2) the

308   Yinpterochiroptera-Yangochiroptera hypothesis, in which the Yangochiroptera clade includes

309   most of Microchiroptera and the Yinpterochiroptera clade includes the rest of

310   Microchiroptera and all of Megachiroptera. There is growing support for this hypothesis in

311   the literature (Meganathan, et al. 2012; Tsagkogeorga, et al. 2013; Ren, et al. 2018; Reyes-

312   Amaya and Flores 2019).

313   Similar to Chiroptera, the root of Cetartiodactyla remains contentious in the literature.

314   The three main hypotheses regarding the root of Cetartiodactyla are: 1) Tylopoda as the sister

315   group for all other cetartiodactylans; 2) Suina as the sister group for all other

316   cetartiodactylans; 3) the monophyletic clade containing Tylopoda and Suina as the sister

317   group for all other cetartiodactylans.

318   To ascertain whether certain sites or loci had very strong effects on the placement of

319   the root we follow the approach of Shen et. al. (Shen, et al. 2017) and calculate the difference

320   in site-wise log-likelihood scores (ΔSLS) and gene-wise log-likelihood scores (ΔGLS) between

321   the supported root positions for each clade. Moreover, we analysed subsamples of each dataset

322   to test the limits of using non-reversible models to root trees with smaller datasets.

323

324   RESULTS

325   *Inference of the mammal tree and selection of well-defined clades*

326   The trees inferred from the whole datasets with the nucleotide-reversible model and

327   the amino-acid-reversible model (Appendix Fig. A.2, Appendix Fig. A.3, Appendix Table

328   A.2) are consistent with the published tree (Liu, et al. 2017). Five clades met all the criteria of

329   well-defined clades, namely, Afrotheria, Bovidae, Carnivora, Myomorpha, and Primates in

330   both amino acid and nucleotide datasets (see Appendix Table A.1 and Appendix Table A.2).

331   Trees in Newick format can be found on github:

332   https://github.com/suhanaser/Rootstrap/tree/master/trees

333   *High accuracy of the AA non-reversible model in inferring the root*

334   Using NR-AA, we inferred the correct root with very high rootstrap support for all

335   five well-defined clades when all loci were used (Appendix Table A.3). Moreover, for all the

336   five clades, the true root was the only root placement in the confidence set of the AU test.

337   The average running time of the NR-AA model (model estimation + tree search + bootstrap +

338   root inference) is 929 hrs on one core 2.6GHz CPU. However, using the optimal number of

339   cores for each dataset reduced the average running time to 43.5 hrs per dataset.

340   Our results show that using only 10% of the sites in the amino acid alignments

341   (around 300,000 alignment columns) still gave very high rootstrap support values (> 98%)

342   for four of the five well-defined clades (Fig. 3) with no correlation between rSED and rBED

343   and the size of the dataset (Table A.3). Moreover, in three of five well-defined clades, 1% of

344   the sites (around 30,000 alignment columns) was enough to give a very high rootstrap

345   support value for the assumed correct root placement. Using only 0.1% of the sites (around

346   3000 alignment columns) decreased the rootstrap support value noticeably in all datasets

347   (Appendix Table A.3). These values are shown for each dataset in Figure 3, where the X-axis

15

348   is plotted in terms of parsimony-informative sites to allow for a more direct comparison

349   between datasets, and to assist those applying these methods in deciding whether to use them

350   on their own data. Although the rootstrap support for the true root improves as the number of

351   parsimony-informative sites increase, in some datasets (e.g. Afrotheria nucleotide dataset)
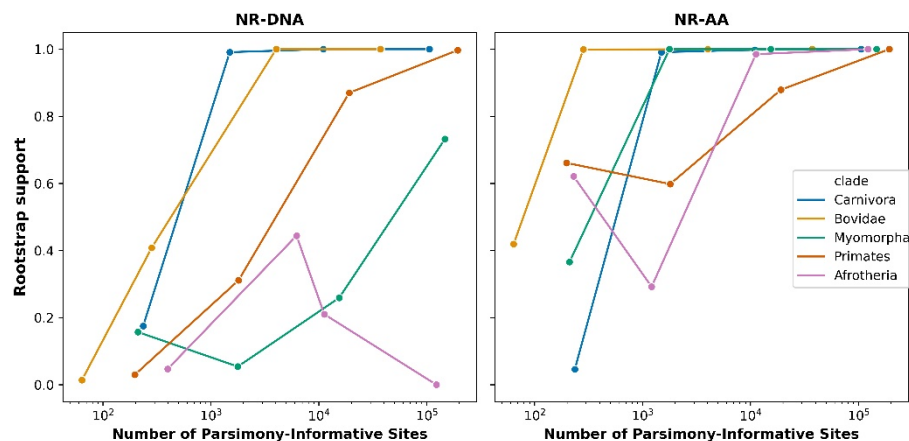
352   this is not the case (Fig. 3).

353
354   FIGURE 3.   The rootstrap support value for each clade as a function of the number of
355   parsimony-informative sites.

356

357   The non-reversible amino acid models were strongly preferred to the reversible models on the

358   complete datasets (BIC values were 93943 to 235958 units better for the non-reversible

359   models), and for the datasets with 10% of loci subsampled (BIC values were 3577 to 15082

360   units better for the non-reversible models), but the opposite was true for the datasets 1% and

361   0.1% of the loci subsampled (e.g. BIC values were between 2102 and 2712 units worse for

362   the non-reversible models for the 0.1% subsampled datasets; see Table A.7 for full results).

363
364   *Poor performance of the DNA non-reversible model in inferring the root*

365   We correctly inferred the root for four out of the five nucleotide datasets with the NR-

366   DNA model, when all loci were used. However, the rootstrap support was generally lower

367   than in the amino-acid datasets (Fig. 3, Appendix Tables A.3 and A.4). Similar to amino-acid

368   datasets, there is no correlation between rSED and rBED and the size of the dataset (Table

369   A.4). The average running time of the NR-DNA model (model estimation + tree search +

370   bootstrap + root inference) is 35.7 hrs on one core 2.6GHz CPU and 4 hours when the

371   optimal number of cores for each dataset were used.

372   In contrast to the NR-AA model, there is no conclusive preference for the NR-DNA

373   model over the reversible DNA model for the datasets we analysed (Table A.8). In fact, The

374   BIC values of the NR-DNA models are always worse than reversible models regardless to the

375   size of the nucleotide dataset except for three clades when all loci were included (Table A.8).

376   In two of the datasets (Myomorpha and Primates) were the NR-DNA model was better than

377   the reversible model the root placement was inferred correctly with high rootstrap support

378   (>95%). In fact, the Afrotheria nucleotide dataset is the only dataset in which the non-

379   reversible model was better than the reversible model but the root placement was inferred

380   incorrectly.

381   Our results show that removing the third codon positions does not improve the

382   rootstrap support value. In contrast, in some datasets removing third codon positions

383   decreased the rootstrap support value and increased the rSED (Table 1).

384   TABLE 1.   Rootstrap support and rSED values in whole nucleotide datasets and
385   nucleotide datasets without third codon positions.

Clades	All loci		Without 3rd	
	rootstrap	rSED	rootstrap	rSED
Afrotheria	0.0%	2	0.0%	2
Primates	99.7%	0	90.1%	0
Myomorpha	73.2%	0	15.8%	1
Carnivora	100.0%	0	100.0%	0
Bovidae	100.0%	0	82.5%	0

386

17

387   *Removing loci that violate the stationarity and homogeneity assumptions improves the*

388   *rootstrap support*

389   As expected, our results show that removing loci that fail the MaxSym test improves

390   the rootstrap support values when the rootstrap support value was less than 100% and/or the

391   root placement was inferred incorrectly, as the case in some nucleotide datasets (Table 2).

392   TABLE 2.   Rootstrap support values in whole datasets and datasets with loci that passed
393   the MaxSym test only.

	Amino Acid		Nucleotide	
Clade	all loci	Passed MaxSym	all loci	Passed MaxSym
Afrotheria	100.0%	100.0%	0.0%	8.4%
Primates	100.0%	100.0%	99.7%	99.9%
Myomorpha	100.0%	100.0%	73.2%	88.3%
Carnivora	100.0%	100.0%	100.0%	100.0%
Bovidae	100.0%	100.0%	100.0%	100.0%

394

395   *Microchiroptera-Megachiroptera or Yinpterochiroptera-Yangochiroptera?*

396   Using the whole amino acid dataset, our results show 65.5% rootstrap support for the

397   Yinpterochiroptera-Yangochiroptera hypothesis and 23.2% for the Microchiroptera -

398   Megachiroptera hypothesis. The remaining11.3% of the rootstrap support goes to supporting

399   the branch leading to Rhinolophoidea as root branch of the bats (Fig. 4). Removing amino

400   acid loci that fail the MaxSym test (110 loci) gives similar results, with 65.9% rootstrap

401   support for the Yinptero-Yango hypothesis and 25.6% rootstrap support for the Micro-Mega

402   hypothesis. In both cases, the AU test could not reject any of the three root positions that

403   received non-zero rootstrap support (Appendix Table A.5).
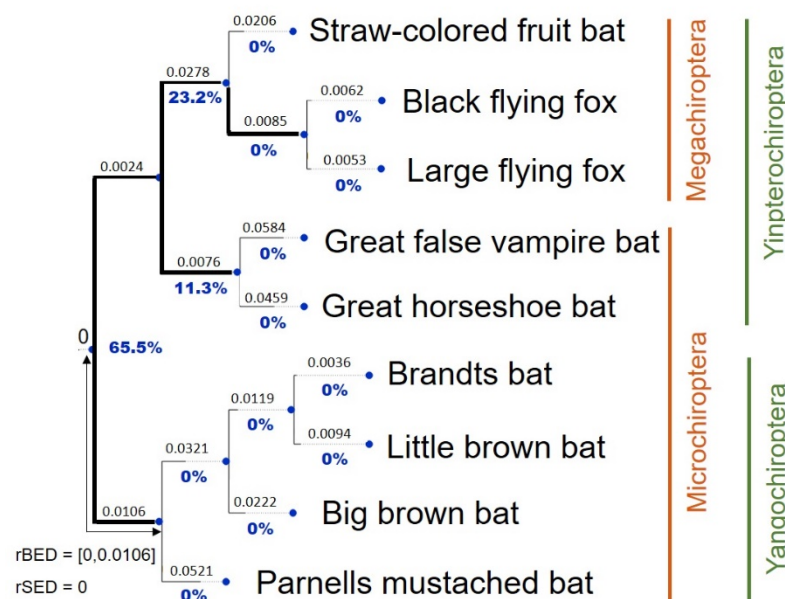
ROOTSTRAP SUPPORT

404

405　FIGURE 4.　　　The ML rooted tree as inferred from the whole Chiroptera amino acid dataset.
406　Bold branches are branches in the AU confidence set. Blue values under each branch are the
407　rootstrap support values.

408　　　　Using the NR-DNA model gives 100% rootstrap support for the Yinptero-Yango

409　hypothesis, and we can confidently reject the Micro-Mega hypothesis in favour of the

410　Yinptero-Yango hypothesis using the AU test (Appendix Fig. A.4). Yet, removing nucleotide

411　loci that fail the MaxSym test (~25% of the loci) decreases the support for the Yinptero-

412　Yango hypothesis to 90.1%, although we can still confidently reject the Micro-Mega

413　hypothesis using the AU test (Appendix Table A.5).

414　　　　Interestingly, when we randomly subsample 10%, 1%, and 0.1% of the loci in the

415　nucleotide dataset, we consistently get the Yinptero-Yango hypothesis as the ML tree and the

416　solely rooted topology in the AU confidence set (Appendix Table A.5). Moreover, the

417　rootstrap support value for the Yinptero-Yango hypothesis increases and the rootstrap support

418　value for the Micro-Mega hypothesis decreases as more parsimony-informative sites are

419　added to the alignment, for both nucleotide and amino acid datasets (Fig. 5, Appendix Table

420　A.5). These results are consistent with previous studies that used smaller datasets (Appendix
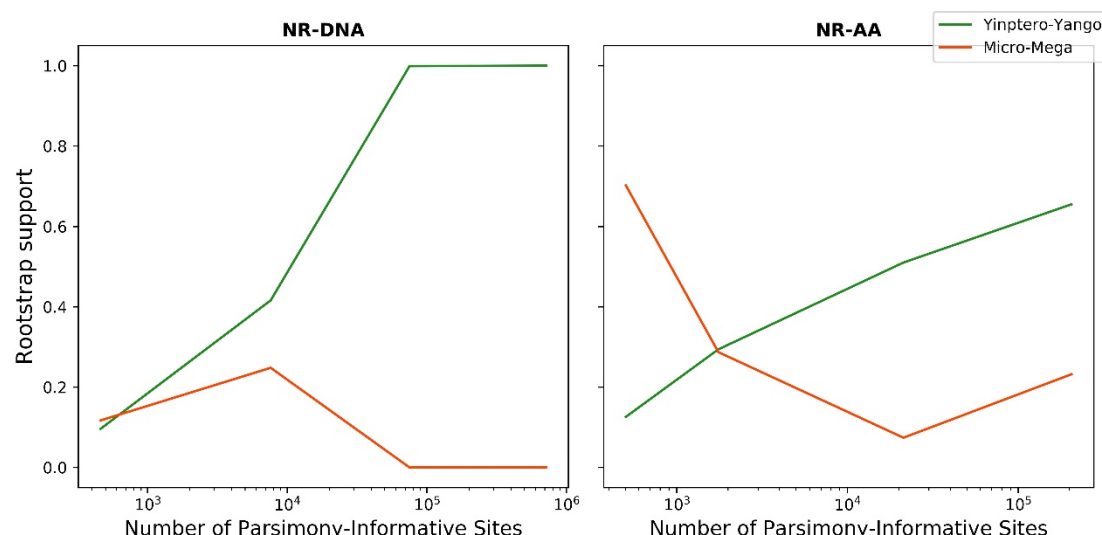
421　Figure A.10)

19

FIGURE 5. Rootstrap support value as a function of the number of parsimony-informative characters in the Chiroptera nucleotide and amino acid datasets using the NR-DNA model (to the left) and the NR-AA model (to the right).

The ΔGLS and ΔSLS values (Shen, et al. 2017) reveal that approximately half of the nucleotide and amino acid loci prefer the Yinptero-Yango hypothesis while the other half prefers Micro-Mega hypothesis. Furthermore, slightly less than half of the nucleotide sites prefer the Yinptero-Yango hypothesis. However, more than two-thirds of the amino acid sites prefer the Yinptero-Yango hypothesis (Appendix Fig. A.5). The distributions of ΔGLS and ΔSLS (Appendix Fig. A.6) show that a small proportion of the amino acid loci (~1%) have very strong support for the Micro-Mega hypothesis, and removing those loci from the alignment increased the rootstrap support for the Yinptero-Yango hypothesis to 76.6%. Nonetheless, both root placements are still in the confidence set of the AU test (Appendix Table A.5) with the amino acid dataset. On the other hand, removing nucleotide loci with the highest absolute ΔGLS value still gives the Yinptero-Yango hypothesis as the ML tree and the sole topology in the AU confidence set. Although the nucleotide data show a clear preference to the Yinptero-Yango hypothesis, in terms of BIC scores, the NR-DNA model preforms worse than reversible models in all datasets except for the dataset where we removed loci that failed the MaxSym test (Table A.5). On the other hand, the NR-AA

ROOTSTRAP SUPPORT

441   performs better than reversible models in big datasets (Table A.5). Yet, the amino acid data

442   do not allow us to distinguish between the two leading hypotheses for the placement of the

443   root of the Chiroptera based on rooting with non-reversible models (Table A.5).

444   *The ambiguous root of Cetartiodactyla*

445   The ML tree inferred with the whole amino acid dataset places the clade containing

446   Tylopoda (represented by its only extant family; Camelidae) and Suina as the sister group to

447   all other cetartiodactylans with 71.8% rootstrap support (Fig. 6). Yet, The AU test did not

448   reject Tylopoda alone as the sister group to all other cetartiodactylans. On the other hand, the

449   ML tree inferred with the whole nucleotide dataset places Tylopoda as the only sister group

450   to all other cetartiodactylans with 71.0% rootstrap support, and we can confidently reject the

451   Tylopoda + Suina hypothesis using the AU test (Appendix Fig. A.7).

452   Removing the amino acid loci that failed the MaxSym test (~1%) still places Tylopoda +

453   Suina as the sister group to all other cetartiodactylans, yet, it decreases the rootstrap support

454   for the Tylopoda + Suina hypothesis to 63.3% and increases the rootstrap support for the

455   Tylopoda hypothesis to 28.5%. However, we still cannot reject either of the hypotheses using
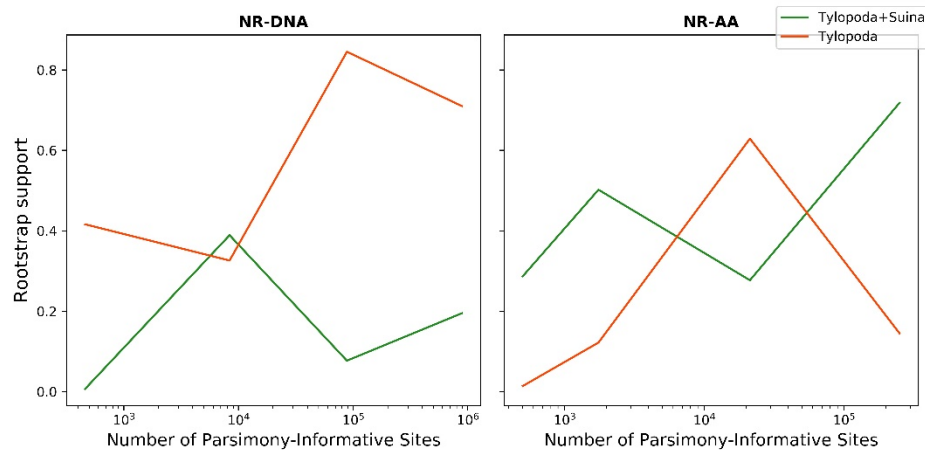
456   the AU test (Appendix Table A.6).

21

FIGURE 6.   The ML rooted tree of as inferred from the whole Cetartiodactyla amino acid dataset. Bold branches are branches in the AU confidence set. Blue values under each branch are the rootstrap support values.

Removing the nucleotide loci that failed the MaxSym test (~1%) still places Tylopoda as the only sister group to all other cetartiodactylans and the only rooted topology in the AU confidence set. However, it decreases the rootstrap support for the Tylopoda hypothesis to 68.7% and increases the rootstrap support for the Tylopoda + Suina hypothesis to 20.1% (Appendix Table A.6).

The results from the subsample datasets are mixed (Fig. 7). Analyses on smaller datasets show no clear pattern in the placement of the root (Appendix Table A.6), leading us to conclude only that the analyses of the whole dataset is likely to provide the most accurate result, but that it is plausible that adding more data may lead to different conclusions in the future.

ROOTSTRAP SUPPORT

FIGURE 7.   rootstrap support value as a function of the number of parsimony-informative characters in the Cetartiodactyla nucleotide and amino acid datasets using the NR-DNA model (to the left) and the NR-AA model (to the right).

ΔGLS analyses reveal that approximately, half of the amino acid and nucleotide loci favour the Tylopoda+Suina hypothesis, while the other half of loci favour the Tylopoda hypothesis (Appendix Figs. A.8-9). On the other hand, two-thirds of the amino acid sites and more than 80% of the nucleotide sites favour the Tylopoda+Suina hypothesis. Removing 1% of the amino acid loci with the highest absolute ΔGLS values still places Tylopoda + Suina as the sister group to all other cetartiodactylans. However, the rootstrap support of the Tylopoda + Suina decreased to 63.2% and the rootstrap support for the Tylopoda hypothesis remains approximately the same (~14.5%), while the rootstrap support for the Suina hypothesis increases from 13.7% to 22.4%. Yet, both the Tylopoda + Suina hypothesis and the Tylopoda hypothesis are in the confidence set of the AU test, while the Suina hypothesis is rejected by the AU test (Appendix Table A.6).

Removing 1% of the nucleotide loci with the highest absolute ΔGLS values gives the Tylopoda+Suina as the sister group to all other cetartiodactylans with 39.7% rootstrap support. However, the solely rooted topology in the AU confidence set is the topology in which the root is placed on the branch leading to Suina (Appendix Table A.6). Similar to Chiroptera and the well-defined clades, NR-AA model preforms, in terms of the BIC score,

491   better than reversible models in big amino-acid datasets, while the NR-DNA performs worse

492   than reversible models in all datasets (Table A.6). We conclude that neither the nucleotide

493   nor the amino acid data are adequate to infer the root placement of Cetartiodactyla with non-

494   reversible models.

495   DISCUSSION

496   In this paper, we introduced a new measure of support for the placement of the root in

497   a phylogenetic tree, the rootstrap support value, and applied it to empirical amino acid and

498   nucleotide datasets inferred using non-reversible models implemented in IQ-TREE (Minh, et

499   al. 2020). The rootstrap is a useful measure because it can be used to assess the statistical

500   support for the placement of the root in any rooted tree, regardless of the rooting method. In a

501   Maximum Likelihood setting, interpretation of the rootstrap support is similar to the

502   interpretation of the classic nonparametric bootstrap. In a Bayesian setting, the same

503   procedure could be used to calculate the posterior probability of the root placement given a

504   posterior distribution of trees. It is noteworthy that the rootstrap support value is not a

505   measure of the accuracy of the root placement and therefore should not be interpreted as

506   such. However, it provides information about the robustness of the root inference with regard

507   to resampling the data. This interpretation is consistent with the interpretation of the

508   nonparametric bootstrap (Holmes 2003) but with regard to the root placement instead of the

509   whole tree topology.

510   In addition to the rootstrap support value, we introduced another two metrics; the root

511   branch-length error distance (rBED), and the root split error distance rSED. Similar to the

512   rootstrap metric, these additional metrics can be used in with any approach that generates

513   rooted phylogenetic trees. We note that both metrics require the true position of the root to be

539   datasets we analysed: we show that removing loci that violate the stationarity and

540   homogeneity assumptions improves the accuracy and statistical support for the placement of

541   the root. This is not surprising since the robustness of the rootstrap, similar to the bootstrap,

542   relies on the consistency of the inference method, so removing systematic bias should

543   improve its performance.

544   We used the non-reversible approach to rooting trees along with the rootstrap support

545   to assess the evidence for different root placements in the Chiroptera and Cetartiodactyla.

546   Using the amino acid datasets we found that in both cases, although there tended to be higher

547   rootstrap support for one hypothesis, neither of the current hypotheses for either dataset could

548   be rejected. These results emphasize the importance of the rootstrap support value as a

549   measure of the robustness of the root estimate given the data. In both the Chiroptera and

550   Cetartiodactyla datasets the root placement varied among subsamples of the dataset, and the

551   rootstrap support reflects this uncertainty. However, checking the stability of root placement

552   estimate by randomly subsampling from the whole Chiroptera dataset show an obvious trend

553   towards the Yinpterochiroptera-Yangochiroptera hypothesis as the dataset increases in size.

554   This trend is consistent with a small number of influential sites or loci having their signal

555   progressively drowned out in favour of the Yinpterochiroptera-Yangochiroptera hypothesis

556   as more data are added to the alignment. In both the Chiroptera and Cetartiodactyla cases, the

557   amino acid data is inadequate to distinguish between certain root placements. On the other

558   hand, in both the Chiroptera and Cetartiodactyla, the nucleotide datasets appear to show

559   stronger support for a single root placement.

560   Comparing BIC scores of reversible and non-reversible models show that in most of

561   the nucleotide datasets the reversible model was a much better fit to the data than the NR-

562   DNA model. This is likely due to the limitations of the method we used to infer the NR-DNA

563   model. Specifically, when inferring the trees with reversible DNA models, we used a

564   partitioned model such that each partition was able to have an independent DNA substitution

565   model. On the other hand, when we inferred the NR-DNA model we estimated a single

566   model for the entire alignment. Thus, the NR-DNA model we inferred was unable to account

567   for heterogeneity in the evolutionary process among partitions, possibly leading to its worse

568   fit to the data when assessed using BIC scores. This suggests that using either mixture models

569   or partitioned models may improve the fit of non-reversible DNA models to the data. The

570   DNA results are consistent with results from previous study using the NR-DNA model and

571   RootDigger (Bettisworth and Stamatakis 2020), although that study did not compare the

572   performance of IQ-TREE and RootDigger on empirical datasets. Its results indicate that the

573   NR-DNA model in IQ-TREE could not infer the correct root placement for any of the three

574   tested datasets.

575   Our results demonstrate that the amino-acid non-reversible model can often be

576   surprisingly accurate for inferring the root placement of phylogenies in the absence of

577   additional information (such as outgroups) or assumptions (such as molecular clocks). In all

578   of the well-defined clades that we examined, the non-reversible amino-acid model

579   successfully identified the root that we identified a-priori as correct, and with very high

580   rootstrap support. Importantly, the non-reversible amino-acid models also tended to fit the

581   data far better than their reversible counterparts did. Indeed, we show that root placements

582   appear to be accurate even with datasets as small as 50 well-curated loci between fairly

583   closely-related taxa such as orders of mammals. Nevertheless, the application of the non-

584   reversible amino acid models to two clades where the root position has previously been

585   contentious failed to shed much additional light on the true root placement. Thus, while we

586   show that the use of non-reversible models certainly has promise, we also show that it is no

587   silver bullet.

588   Where a reliable outgroup taxon can be found, without the issues that can confound

589   the inference of root placements using outgroups (Dalevi, et al. 2001; Braun and Kimball

590   2002; Graham, et al. 2002; Brady, et al. 2006), we suggest relying on the use of outgroups.

591   Nevertheless, where no reliable outgroups can be found, or where there is some reason to

592   question the position of a root inferred using an outgroup (e.g. references about questionable

593   outgroup rooting), our study suggests that using non-reversible models can provide a useful

594   additional line of evidence for the position of the root of a phylogeny. We note also that the

595   rootstrap value and the AU test could be used to provide estimates of the uncertainty of root

596   placement using an outgroup taxon

597   Our work suggests a practical approach to inferring the root of a phylogenetic tree

598   using non-reversible models. First, estimate an unrooted tree topology using the best

599   reversible models available, excluding outgroup sequences. Next, fix the tree topology and

600   use the best non-reversible models available to infer the Maximum Likelihood (ML) root

601   position of that tree. Finally, determine to what extent the ML root position should be trusted.

602   The degree of trust that researchers should put in an inferred ML root position should be

603   influenced by three factors (noting of course that all phylogenetic inferences are susceptible

604   to be misled by model misspecification). First, the fit of the non-reversible model to the data

605   should be better than the fit of the reversible model. This can be assessed using common

606   criteria like AICc or BIC scores. A better fit of the non-reversible model provides some

607   assurance that the data contain sufficient signal that using a non-reversible model is advisable

608   in the first place. Our results show that the root placement was inferred correctly with high

609   rootstrap support in 12 out of the 13 datasets in which the non-reversible model was

610   preferable.

611   In the absence of a better fit for a non-reversible model, we do not think any inferred

612   ML root position should be trusted. Second, root positions with higher rootstrap support

613   should be trusted more, because a higher rootstrap support indicates less variance among sites

614   in the signal for the placement of the root. Third, ML root positions should be trusted more

615   when the number of root placements included in the confidence set of an AU test is small,

616   because a smaller confidence set indicates that there is less uncertainty in the root placement

617   when the analysis is conditioned on the full alignment and the unrooted ML tree topology. A

618   conservative approach to inferring root placements with non-reversible models would be to

619   consider any root placement that has a substantial fraction of the rootstrap support and/or is

620   included in the set of possible root placements identified by the AU test as a possible root

621   placement given the assumptions of the model.

622   We hope that the combination of non-reversible, rootstrap support, and AU tests will

623   add another tool to the phylogeneticist's arsenal when it comes to inferring rooted

624   phylogenies.

625   FUNDING

628   REFERENCES

629   Agnarsson I, Zambrana-Torrelio CM, Flores-Saldana NP, May-Collado LJ. 2011. A

630   time-calibrated species-level phylogeny of bats (Chiroptera, Mammalia). PLoS

631   Curr 3:RRN1212.

632   Allman ES, Degnan JH, Rhodes JA. 2011. Identifying the rooted species tree from the

633   distribution of unrooted gene trees under the coalescent. J. Math. Biol. 62:833-

634   862.

635   Baldauf SL, Palmer JD. 1993. Animals and fungi are each other's closest relatives:

636   congruent evidence from multiple proteins. Proceedings of the National

637   Academy of Sciences 90:11558-11562.

638   Bettisworth B, Stamatakis A. 2020. RootDigger: a root placement program for

639   phylogenetic trees. bioRxiv:2020.2002.2013.935304.

640   Boussau B, Szollosi GJ, Duret L, Gouy M, Tannier E, Daubin V. 2013. Genome-scale

641   coestimation of species and gene trees. Genome Res. 23:323-330.

642   Brady SG, Schultz TR, Fisher BL, Ward PS. 2006. Evaluating alternative hypotheses

643   for the early evolution and diversification of ants. Proc Natl Acad Sci U S A

644   103:18172-18177.

645   Braun EL, Kimball RT. 2002. Examining Basal avian divergences with mitochondrial

646   sequences: model complexity, taxon sampling, and sequence length. Syst. Biol.

647   51:614-625.

648   Cherlin S, Heaps SE, Nye TMW, Boys RJ, Williams TA, Embley TM. 2018. The Effect

649   of Nonreversibility on Inferring Rooted Phylogenies. Mol. Biol. Evol. 35:984-

650   1002.

651   Chernomor O, von Haeseler A, Minh BQ. 2016. Terrace Aware Data Structure for

652   Phylogenomic Inference from Supermatrices. Syst. Biol. 65:997-1008.

653   Dalevi D, Hugenholtz P, Blackall LL. 2001. A multiple-outgroup approach to resolving

654   division-level phylogenetic relationships using 16S rDNA data. Int. J. Syst.

655   Evol. Microbiol. 51:385-391.

656   Dayhoff M. 1987. A model of evolutionary change in proteins. Atlas of protein

657   sequence and structure 5:suppl. 3.

658   Dayhoff M, Schwartz R. 1980. Prokaryote evolution and the symbiotic origin of

659   eukaryotes. Endocytobiology: endosymbiosis and cell biology: a synthesis of

660   recent research 1:63-84.

661   Dayhoff M, Schwartz R, Orcutt B. 1978. A model of evolutionary change in proteins.

662   Atlas of protein sequence and structure 5:345-352.

663   Drummond AJ, Ho SY, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and

664   dating with confidence. PLoS Biol. 4:e88.

665   Duchene DA, Tong KJ, Foster CSP, Duchene S, Lanfear R, Ho SYW. 2020. Linking

666   Branch Lengths across Sets of Loci Provides the Highest Statistical Support for

667   Phylogenetic Inference. Mol. Biol. Evol. 37:1202-1210.

668   Farris JS. 1972. Estimating Phylogenetic Trees from Distance Matrices. Am. Nat.

669   106:645-&.

670   Gatesy J, Springer MS. 2017. Phylogenomic red flags: Homology errors and zombie

671   lineages in the evolutionary diversification of placental mammals. Proc Natl

672   Acad Sci U S A 114:E9431-E9432.

673   Gogarten JP, Kibak H, Dittrich P, Taiz L, Bowman EJ, Bowman BJ, Manolson MF,

674   Poole RJ, Date T, Oshima T, et al. 1989. Evolution of the vacuolar H+-ATPase:

675   implications for the origin of eukaryotes. Proc Natl Acad Sci U S A 86:6661-

676   6665.

677   Graham SW, Olmstead RG, Barrett SC. 2002. Rooting phylogenetic trees with distant

678   outgroups: a case study from the commelinoid monocots. Mol. Biol. Evol.

679   19:1769-1781.

680   Hasegawa M, Kishino H, Yano T. 1985. Dating of the human-ape splitting by a

681   molecular clock of mitochondrial DNA. J. Mol. Evol. 22:160-174.

682   Heaps SE, Nye TM, Boys RJ, Williams TA, Embley TM. 2014. Bayesian modelling of

683   compositional heterogeneity in molecular phylogenetics. Stat Appl Genet Mol

684   Biol 13:589-609.

685   Holmes S. 2003. Bootstrapping phylogenetic trees: Theory and methods. Statistical

686   Science 18:241-255.

687   Huelsenbeck JP, Bollback JP, Levine AM. 2002. Inferring the Root of a Phylogenetic

688   Tree. Syst. Biol. 51:32-43.

689   Iwabe N, Kuma K, Hasegawa M, Osawa S, Miyata T. 1989. Evolutionary relationship of
690   archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of
691   duplicated genes. Proc Natl Acad Sci U S A 86:9355-9359.
692   Jansen RK, Cai Z, Raubeson LA, Daniell H, Depamphilis CW, Leebens-Mack J, Muller
693   KF, Guisinger-Bellian M, Haberle RC, Hansen AK, et al. 2007. Analysis of 81
694   genes from 64 plastid genomes resolves relationships in angiosperms and
695   identifies genome-scale evolutionary patterns. Proc Natl Acad Sci U S A
696   104:19369-19374.
697   Jayaswal V, Ababneh F, Jermiin LS, Robinson J. 2011. Reducing model complexity of
698   the general Markov model of evolution. Mol. Biol. Evol. 28:3045-3059.
699   Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data
700   matrices from protein sequences. Comput. Appl. Biosci. 8:275-282.
701   Jukes TH, Cantor C. 1969. Evolution of protein molecules. In: Munro HN, editor. In
702   Mammalian Protein Metabolism.
703   Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017.
704   ModelFinder: fast model selection for accurate phylogenetic estimates. Nat.
705   Methods 14:587-589.
706   Kimura M. 1980. A Simple Method for Estimating Evolutionary Rates of Base
707   Substitutions through Comparative Studies of Nucleotide-Sequences. J. Mol.
708   Evol. 16:111-120.
709   Kocot KM, Cannon JT, Todt C, Citarella MR, Kohn AB, Meyer A, Santos SR, Schander
710   C, Moroz LL, Lieb B, et al. 2011. Phylogenomics reveals deep molluscan
711   relationships. Nature 477:452-456.
712   Lake JA, Herbold CW, Rivera MC, Servin JA, Skophammer RG. 2007. Rooting the tree
713   of life using nonubiquitous genes. Mol. Biol. Evol. 24:130-136.
714   Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. Mol.
715   Biol. Evol. 25:1307-1320.

716   Liu L, Zhang J, Rheindt FE, Lei F, Qu Y, Wang Y, Zhang Y, Sullivan C, Nie W, Wang

717   J, et al. 2017. Genomic evidence reveals a radiation of placental mammals

718   uninterrupted by the KPg boundary. Proc Natl Acad Sci U S A 114:E7282-

719   E7290.

720   Lyons-Weiler J, Hoelzer GA, Tausch RJ. 1998. Optimal outgroup analysis. Biol. J.

721   Linn. Soc. 64:493-511.

722   Maddison WP, Donoghue MJ, Maddison DR. 1984. Outgroup Analysis and Parsimony.

723   Systematic Zoology 33:83-103.

724   Mai U, Sayyari E, Mirarab S. 2017. Minimum variance rooting of phylogenetic trees

725   and implications for species tree reconstruction. PLoS One 12:e0182238.

726   Meganathan PR, Pagan HJ, McCulloch ES, Stevens RD, Ray DA. 2012. Complete

727   mitochondrial genome sequences of three bats species and whole genome

728   mitochondrial analyses reveal patterns of codon bias and lend support to a basal

729   split in Chiroptera. Gene 492:121-129.

730   Meredith RW, Janecka JE, Gatesy J, Ryder OA, Fisher CA, Teeling EC, Goodbla A,

731   Eizirik E, Simao TL, Stadler T, et al. 2011. Impacts of the Cretaceous Terrestrial

732   Revolution and KPg extinction on mammal diversification. Science 334:521-524.

733   Milinkovitch MC, Lyons-Weiler J. 1998. Finding optimal ingroup topologies and

734   convexities when the choice of outgroups is not obvious. Mol. Phylogen. Evol.

735   9:348-357.

736   Minh BQ, Hahn MW, Lanfear R. 2020. New methods to calculate concordance factors

737   for phylogenomic datasets. Mol. Biol. Evol.

738   Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A,

739   Lanfear R. 2020. IQ-TREE 2: New Models and Efficient Methods for

740   Phylogenetic Inference in the Genomic Era. Mol. Biol. Evol. 37:1530-1534.

741   Moore MJ, Bell CD, Soltis PS, Soltis DE. 2007. Using plastid genome-scale data to

742   resolve enigmatic relationships among basal angiosperms. Proc Natl Acad Sci U

743   S A 104:19363-19368.

744 Nardi F, Spinsanti G, Boore JL, Carapelli A, Dallai R, Frati F. 2003. Hexapod origins:

745   monophyletic or paraphyletic? Science 299:1887-1889.

746 Naser-Khdour S, Minh BQ, Zhang W, Stone EA, Lanfear R. 2019. The Prevalence and

747   Impact of Model Violations in Phylogenetic Analysis. Genome Biol Evol.

748 Philippe H, Brinkmann H, Lavrov DV, Littlewood DT, Manuel M, Worheide G, Baurain

749   D. 2011. Resolving difficult phylogenetic questions: why more sequences are not

750   enough. PLoS Biol. 9:e1000602.

751 Ren M, Sun HJ, Bo SQ, Zhang SY, Hua PY. 2018. Parallel amino acid deletions of

752   prestin protein in two dramatically divergent bat lineages suggest the complexity

753   of the evolution of echolocation in bats. Acta Chiropterologica 20:311-317.

754 Reyes-Amaya N, Flores D. 2019. Hypophysis size evolution in Chiroptera. Acta

755   Chiropterologica 21:65-74.

756 Rivera MC, Lake JA. 1992. Evidence That Eukaryotes and Eocyte Prokaryotes Are

757   Immediate Relatives. Science 257:74-76.

758 Shen XX, Hittinger CT, Rokas A. 2017. Contentious relationships in phylogenomic

759   studies can be driven by a handful of genes. Nat Ecol Evol 1:126.

760 Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection.

761   Syst. Biol. 51:492-508.

762 Shimodaira H, Hasegawa M. 1999. Multiple comparisons of log-likelihoods with

763   applications to phylogenetic inference. Mol. Biol. Evol. 16:1114-1116.

764 Smith AB. 1994. Rooting Molecular Trees - Problems and Strategies. Biol. J. Linn. Soc.

765   51:279-292.

766 Squartini F, Arndt PF. 2008. Quantifying the Stationarity and Time Reversibility of the

767   Nucleotide Substitution Process. Molecular Biology and Evolution 25:2525-

768   2535.

769 Steenkamp ET, Wright J, Baldauf SL. 2006. The protistan origins of animals and fungi.

770   Mol. Biol. Evol. 23:93-106.

ROOTSTRAP SUPPORT

771 Swofford D, Olsen G, Waddell P. 1996. Phylogenetic inference. In: David M. Hillis
772   CM, Barbara K. Mable, editor. Molecular Systematics, 2nd edn: Sunderland,
773   Mass. : Sinauer Associates. p. 407-513.

774 Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the
775   control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol.
776   Evol. 10:512-526.

777 Tavaré S. 1986. Some probabilistic and statistical probles in the analysis of DNA
778   sequences. Lectures on Mathematics in the Life Sciences 17.

779 Tria FDK, Landan G, Dagan T. 2017. Phylogenetic rooting using minimal ancestor
780   deviation. Nat Ecol Evol 1:193.

781 Tsagkogeorga G, Parker J, Stupka E, Cotton JA, Rossiter SJ. 2013. Phylogenomic
782   analyses elucidate the evolutionary relationships of bats. Curr. Biol. 23:2262-
783   2267.

784 Watrous LE, Wheeler QD. 1981. The out-Group Comparison Method of Character
785   Analysis. Systematic Zoology 30:1-11.

786 Whelan NV, Kocot KM, Moroz LL, Halanych KM. 2015. Error, signal, and the
787   placement of Ctenophora sister to all other animals. Proc Natl Acad Sci U S A
788   112:5773-5778.

789 Whelan S, Goldman N. 2001. A general empirical model of protein evolution derived
790   from multiple protein families using a maximum-likelihood approach. Mol. Biol.
791   Evol. 18:691-699.

792 Williams KP, Gillespie JJ, Sobral BW, Nordberg EK, Snyder EE, Shallom JM,
793   Dickerman AW. 2010. Phylogeny of gammaproteobacteria. J. Bacteriol.
794   192:2305-2314.

795 Williams TA, Heaps SE, Cherlin S, Nye TM, Boys RJ, Embley TM. 2015. New
796   substitution models for rooting phylogenetic trees. Philos Trans R Soc Lond B
797   Biol Sci 370:20140336.

35

798   Wu S, Edwards S, Liu L. 2019. Data from: Genome-scale DNA sequence data and the

799   evolutionary history of placental mammals. In: Figshare.

800   Wu S, Edwards S, Liu L. 2018. Genome-scale DNA sequence data and the evolutionary

801   history of placental mammals. Data Brief 18:1972-1975.

802   Yang ZH, Roberts D. 1995. On the Use of Nucleic-Acid Sequences to Infer Early

803   Branchings in the Tree of Life. Mol. Biol. Evol. 12:451-458.

804   Yap VB, Speed T. 2005. Rooting a phylogenetic tree with nonreversible substitution

805   models. BMC Evol. Biol. 5:2.

806   Yu Y, Warnow T, Nakhleh L. 2011. Algorithms for MDC-based multi-locus phylogeny

807   inference: beyond rooted binary gene trees on single alleles. J. Comput. Biol.

808   18:1543-1559.

809   Zhang SQ, Che LH, Li Y, Dan L, Pang H, Slipinski A, Zhang P. 2018. Evolutionary

810   history of Coleoptera revealed by extensive sampling of genes and species. Nat

811   Commun 9:205.

812   Zhou X, Xu S, Yang Y, Zhou K, Yang G. 2011. Phylogenomic analyses and improved

813   resolution of Cetartiodactyla. Mol Phylogenet Evol 61:255-264.

814