Phylogenomic Analyses to Root the Tree of Life

Nicholas Boffa, u7686660

SCNC2101: Advanced Studies Course

Professor Robert Lanfear

October 25, 2024

# Abstract

An ongoing debate in evolutionary biology revolves around whether sea sponges (Porifera) or comb jellies (Ctenophora) were the first group of organisms to split from all other animals (Metazoa) – that is, which is the root of the Metazoan tree. Generally, biologists investigate this by including non-  
animals in their phylogenetic trees – the closest animal phylum to the non-animals in the tree is then the basal  
lineage. However, it has been suggested that this rooting method is biased towards placing Ctenophora at the root; additionally, which non-animals are included has been shown to change the inferred tree. To avoid these potential issues, this study utilises a rooting method that does not require the use of non-animals in the phylogenetic inference, and has recently been shown to be effective at discerning the root of several taxonomic groups: nonreversible amino acid substitution models. These are models of sequence evolution in which the instantaneous rate of mutation between any two types of amino acid is allowed to be different, and this fact allows the maximum-likelihood root placement to be determined during the tree inference, unlike in standard models. Using four previously published datasets of varying size and composition, maximum likelihood trees were inferred and evaluated. For the three smallest datasets, the inferred tree failed to resolve the 5 Metazoan phyla – Bilateria, Placozoa, Cnidaria, Ctenophora, and Porifera – all as separate groups. For the largest dataset, the nonreversible model resulted in a tree with siliceous sponges (one of the major groups in Porifera) as the sister to all other Metazoans with moderate statistical support, but this finding was not robust to removing small amounts of data, and is not commonly supported by the literature. Thus, though nonreversible models have been shown to have promise in rooting trees over smaller evolutionary distances, they appear limited when applied to the much older Metazoan tree. Since the so-called profile mixture models used by the original authors of the datasets were able to infer better-supported trees, while both the nonreversible and reversible models used in this research could not, future research in this area should focus on developing nonreversible profile mixture models, to both allow for the advantages of rooting the Metazoan tree without non-animals, whilst being more likely to recover a tree consistent with the morphological and phylogenetic consensus of the broader scientific community.

# Introduction

One of the fiercest debates in evolutionary biology concerns the root of the animal tree: whether comb jellies were the earliest-diverging animal lineage (Ctenophora-sister hypothesis), or sponges (Porifera-sister hypothesis). Historically, scientists agreed that Porifera were the first to split due to their apparently simple morphology (Wainright et al., 1993), but ever since Dunn et al. (2008) recovered Ctenophora-sister in their phylogenetic study there has been an extensive back-and-forth. Phylogenetic studies since have typically found that the sister group to all Metazoans is either Ctenophora (e.g. Hejnol et al., 2009; Laumer et al., 2019; Laumer et al., 2018; Li et al., 2021; Ryan et al., 2013; Schultz et al., 2023; Whelan et al., 2015) or Porifera (e.g. Feuda et al., 2017; Philippe et al., 2011; Philippe et al., 2009; Pick et al., 2010; Pisani et al., 2015; Redmond & McLysaght, 2022; Simion et al., 2017), though other topologies are sometimes reported (e.g. Francis & Canfield, 2020; Schierwater et al., 2009). As can be seen from a cursory viewing of the author lists of the papers just cited, which authors helped to write a paper is often an excellent predictor of its conclusions, suggesting this debate may currently be driven more by ideology than evidence. Nevertheless, there are many reasons why this particular phylogeny has proven difficult to resolve, as the choice of model (Li et al., 2021), data (Nosenko et al., 2013; Shen et al., 2017), and taxa used (Pick et al., 2010) in the analysis have all been shown to impact which hypothesis is supported.

One reason for the large amounts of interest in this tree, is that the two competing hypotheses present very different accounts of the diversification of animals. Several morphological features, such as neurons and muscle cells, are possessed by Ctenophora and most Metazoan clades, but not Porifera (Burkhardt, 2022; Dunn et al., 2015). Thus, if the Ctenophora-sister hypothesis is correct, it suggests that either Ctenophora evolved these traits independently of the other groups, or Porifera lost these traits over time due to natural selection (Schultz et al., 2023). If Porifera-sister is correct, then this suggests these traits evolved after Porifera diverged (Schultz et al., 2023). Aside from these important biological implications, the Metazoan tree is also widely studied because of its difficulty, as its resolution may illuminate more accurate and robust methods for phylogeny reconstruction more broadly.

Though many tree-rooting methods exist, the most common uses outgroups. Outgroup rooting involves estimating a tree that includes both species which are inside (the ingroups) and outside (the outgroups) the group of interest, which in this case is Metazoa (Huelsenbeck et al., 2002). The node at which the outgroups join collectively join to the ingroup tree indicates the oldest common ancestor of all the ingroup species, and thus is the root of the ingroup tree (Huelsenbeck et al., 2002). While this method is effective in most cases, in the Metazoan tree it has been shown that including more distantly related outgroups (such as fungi) can make recovering Ctenophora-sister more likely, while using only more closely related outgroups (like choanoflagellates) can make recovering Porifera-sister more likely (Li et al., 2021). The use of outgroups can also exacerbate a systematic bias known as long-branch attraction (LBA). LBA is a phenomenon in phylogenetic inference whereby two lineages with relatively long branch lengths are incorrectly inferred to be more closely related than they are in reality. Since both species have accumulated many more mutations than the others, their genomes are likely to look more similar due to convergent evolution, rather than because they are necessarily closely related (Bergsten, 2005; Li et al., 2021). This is particularly relevant to the Metazoan tree because there is consensus that the branch leading to Ctenophora is relatively long (Li et al., 2021). Hence, a key contention of Porifera-sister studies is that Ctenophora-sister is the result of the long branches of the outgroup and of Ctenophora attracting due to LBA (e.g. Simion et al., 2017). As such, rooting methods which do not use outgroups, such as using nonreversible substitution models, may be more effective at rooting the Metazoan tree, and so will be the focus of this research.

Most models of sequence evolution used by phylogenetic studies are time-reversible – that is, if a site is currently in state A, the probability of the site mutating to state B after some amount of time is assumed to be the same as the probability of the site changing from B to A over the same amount of time (Sumner et al., 2012). Nonreversible models relax this assumption, with the advantage that the likelihood function then depends on the root of the tree (Dang et al., 2022; Felsenstein, 1981). This means that the maximum likelihood (ML) placement of the root can be estimated without the need for outgroups. Thus, an advantage of the nonreversible rooting method becomes clear: because it does not require outgroups, LBA between Ctenophora and the long outgroup branch is eliminated. Though this method has long been known about, it was thought to not be sufficiently powerful to effectively discriminate between all the candidate rooted trees (Huelsenbeck et al., 2002; Yap & Speed, 2005). However, recently, both a study on simulated and empirical nucleotide data (Bettisworth & Stamatakis, 2021), and a study on empirical nucleotide and amino acid data (Naser-Khdour et al., 2022), demonstrated the utility of the nonreversible rooting method. In particular, though Naser-Khdour et al. (2022) were unable to definitively determine the root for two contentious trees in Mammalia, they found that nonreversible amino acid substitution models were particularly effective at identifying the root of resolved trees. This indicates that nonreversible amino acid models may be useful in rooting the Metazoan tree.

The aim of this report is thus to use nonreversible amino acid substitution models to root the Metazoan phylogenetic tree, using previously published Metazoan datasets.

# Methods

## Datasets and Cleaning

Four datasets were chosen from previous studies on the Metazoan tree, with the requirement that they contain 1) an amino acid multiple sequence alignment, with taxa from every Metazoan phylum, and 2) partitioning information describing the locations of loci in the alignment.

The first two datasets, both approximately 10,000 sites long, were obtained from Nosenko et al. (2013). The first contained nonribosomal amino acid sequences, while the second included ribosomal ones, with Nosenko et al. (2013) reporting different inferred topologies for each. Thus, these datasets served as both a small starting point for analysis and an opportunity to test whether the nonreversible model would yield the same topology for both, contrary to the original findings. The third dataset was from Laumer et al. (2018) and was 94,444 amino acids long. To create the dataset, the authors ran the BUSCO algorithm (Simao et al., 2015) on a larger matrix, filtering out loci that were not highly conserved single-copy orthologs in an attempt to minimise both the number of paralogs and the amount of compositional bias in the data, as both are known to bias phylogenetic analyses (Laumer et al., 2018). The fourth and final dataset was 401,632 amino acids long, from Simion et al. (2017). This dataset was one of the largest available, with the authors focusing on broad taxon sampling and automated data filtering to mitigate phylogenetic errors such as misalignments and the inclusion of paralogous groups (Simion et al., 2017). The four datasets will henceforth be referred to as Nosenko2013 nonribosomal, Nosenko2013 ribosomal, Laumer2018, and Simion2017; a brief summary of each is given in Table 1.

Table 1. Summary of the four datasets analysed in this report, including the models used in the original paper, and the hypothesis it supported. Numerical data was manually collected using the log reports of the reversible model IQ-TREE analyses run for each dataset. For more information on the CAT model, which was used by all of the original papers, see Lartillot and Philippe (2004).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Dataset** | **Non-outgroup taxa** | **Loci** | **Sites** | | **Model Used** | **Hypothesis supported** |
| ***Total*** | ***Parsimony-informative*** |
| Nosenko2013 nonribosomal | 61 | 35 | 9187 | 4558 | CAT+Γ | Ctenophora-sister |
| Nosenko2013 ribosomal | 61 | 87 | 14612 | 8236 | CAT+Γ | Porifera-sister |
| Laumer2018 | 52 | 303 | 94444 | 72871 | CAT + GTR + Г4 | Ctenophora-sister |
| Simion2017 | 72 | 1719 | 401632 | 280024 | CAT+ Γ4 | Porifera-sister |

Datasets had previously been obtained and formatted by Cherryh (2024) before being given to the author. The author then manually removed outgroup sequences, and sequence names were standardised to correspond the scientific name of each taxon. This was done using two custom R v4.3.1 (R Core Team, 2023) scripts: nick\_util\_fasta\_processing.R for the Nosenko2013 and Simion2017 data, and laumer\_fasta\_processing.R for the Laumer2018 data. This scripts used ape v5.8 (Paradis & Schliep, 2019), TreeTools v1.12.0 (Smith, 2019), and seqinR v4.2.36 (Charif & Lobry, 2007). The scripts are both a modified version of Cherryh (2023). The R scripts and all data, both as given by Cherryh (2024) and after cleaning, are available on the project GitHub: <https://github.com/nickboffa/metazoan-root/>.

## Phylogenetic Inference

For each dataset, phylogenetic inference was performed using IQ-TREE v2.3.6 (Minh et al., 2020), following the methodology of Naser-Khdour et al. (2022). First, the best-fitting partition scheme (Chernomor et al., 2016) and estimate of the unrooted topology were found using ModelFinder (Kalyaanamoorthy et al., 2017), using only reversible substitution models. This means that for each locus, the best-fitting reversible Q-matrix and rate heterogeneity across sites (RHAS) model was determined by finding the combination which minimises the Bayesian Information Criterion (BIC) score (Kalyaanamoorthy et al., 2017). Briefly, the Q-matrix determines the relative rates of mutation between different amino acids in the substitution model (Felsenstein, 1981), and incorporating a RHAS model allows for the fact that sites may have been evolving at different rates (Kalyaanamoorthy et al., 2017). The command-line input used to perform this analysis was:

iqtree2 --seed 2222 -s ALIGNMENT.fasta -p PARTITION.nex -–prefix REV

The rooted tree was then inferred using an edge-linked nonreversible model. This means that unlike in the previous step, for each locus the same nonreversible Q-matrix is used, though still with the different RHAS models as inferred in the previous step (Naser-Khdour et al., 2022). The nonreversible model is estimated directly from the data using nQmaker method (Dang et al., 2022), which iteratively updates an existing nonreversible Q-matrix to maximise the log-likelihood function until convergence. The ML parameters of the predetermined best-fit RHAS models are also estimated. Additionally, 1000 ultrafast bootstrap (Hoang et al., 2018) replicates are generated to compute the bootstrap and rootstrap support values for each branch, which are an indication of how robust the topology and root placement respectively are to resampling (Naser-Khdour et al., 2022). The command used for this step was thus:

iqtree2 --seed 2222 -s ALIGNMENT.fasta -p REV.best\_scheme.nex -t REV.treefile --model-joint NONREV -B 1000 --prefix NONREV

The approximately unbiased (AU) test (Shimodaira, 2002) was then performed on tree inferred by the nonreversible model. This statistical test compares the likelihood of each potential root placement to that of the ML root, allowing the rejection of placements that are significantly less likely (Shimodaira, 2002). In a slight deviation from the methods of Naser-Khdour et al. (2022), rather than performing the expensive nonreversible model estimation again, the nonreversible model estimated above was manually extracted from the NONREV.iqtree output file and pasted into a manually-created nexus file (called MODEL.nex below). This gave the following command-line input:

iqtree2 -seed 2222 -s ALIGNMENT.fasta -p REV.best\_scheme.nex -mdef MODEL.nex --model-joint MODEL --root-test -zb 100000 -au -te NONREV.treefile --prefix TOP

The exact commands used can be viewed in the log files of each run, available in the GitHub repository of this report.

## Model Fit

To determine how well the reversible and nonreversible models fit each dataset, BIC scores were calculated for every model. This was done automatically during the IQ-TREE analyses described above, and so were manually extracted from the REV.iqtree and NONREV.iqtree summary files of each run, which are again all available on the GitHub repository of this project.

## Analysis and Removal of Influential Sites

A post hoc analysis of the site-wise log-likelihood scores (SLSs) for the Simion2017 dataset was performed. The difference in a site’s SLS for two competing topological hypotheses (∆SLS) is an indication of how influential that site is on the phylogenetic inference (Francis & Canfield, 2020). The SLSs of four different rooted placements were tested: the ML tree in Figure 8 (T1), the Ctenophora-sister tree (T2), the Porifera-sister tree (T3), and the “Ctenophora+Porifera”-sister tree (T4). These topologies, in Newick format, were manually put into a file called simion.trees. To avoid re-estimation of the nonreversible model or RHAS model parameters, the file simion\_full\_model.nex was manually created. It contained both the automatically generated NONREV.best\_model.nex file, and the estimated rate matrix NQ.simion, from the previous analyses.

iqtree2 --seed 2222 -s ALIGNMENT.fasta -p simion\_full\_model.nex -mdef simion\_full\_model.nex -m NQ.simion -z simion.trees -n 0 -wsl --prefix DELTA\_simion

The output DELTA\_simion.sitelh file was then converted to a CSV file using a custom R v4.3.1 (R Core Team, 2023) script called gls\_sls\_graphing.R, using the functionality of tidyverse v2.0.0 (Wickham et al., 2019) and data.table v1.16.0 (Barrett et al., 2024). Converting to a dataframe enabled simple calculation of the differences in site-wise log-likelihood scores (∆SLS) between each pair of topologies listed earlier, which were then visualised.

To determine the effect of removing small numbers of sites on the inferred tree topology, for a proportion of sites , the top and bottom of sites, according to their ∆SLS, were removed from the alignment, and the partition file adjusted accordingly. Again, this was done using a custom R v4.3.1 (R Core Team, 2023) script called site\_remover.R, which relied upon the packages tidyverse v2.0.0 (Wickham et al., 2019), ape v5.8 (Paradis & Schliep, 2019), and seqinR v4.2.36 (Charif & Lobry, 2007). This was done for equal to 1, 0.1, 0.01, and 0.001, based on the ∆SLS values calculated for T1 and T3. The already-estimated nonreversible model, as described in the Simion2017 version of the MODEL.nex file from earlier called NQ.simion.nex, was then used to infer the ML topology of the new alignment, together with bootstrap and rootstrap support values:

iqtree2 --seed 2222 -s ALIGNMENT.fasta -p PARTITION.nex -t NONREV.treefile -mdef NQ.simion.nex -m NQ.simion -B 1000

## Graphing

Results were plotted using R v4.3.1 (R Core Team, 2023), together with the packages cowplot (Wilke, 2024), ggrepel v0.9.6 (Slowikowski, 2024), ggtext v0.1.2 (Wilke & Wiernik, 2022), ggtree v3.10.1 (Yu et al., 2017), patchwork v1.2.0 (Pedersen, 2024), scales v1.3.0 (Wickham et al., 2023), tidyverse v2.0.0 (Wickham et al., 2019), and treeio v1.26.0 (Wang et al., 2020). The scripts used to create each figure are available from the GitHub repository of this report.

# Results

The primary aim of this research was to use nonreversible amino acid substitution models to root the Metazoan tree. After applying this method to four previously published datasets, neither Ctenophora-sister, Porifera-sister, nor Ctenophora+Porifera-sister were recovered in any instance. For each dataset, ignoring root placements, the reversible models produced identical trees to the nonreversible models.

## ML Trees of each Dataset

### Nosenko2013 Nonribosomal Dataset

The ML tree for the Nosenko2013 nonribosomal dataset is shown in Figure 1. It has the root between Bilateria and all other Metazoans, and Ctenophora as a clade within Cnidaria. Additionally, the Cnidarian *Cyanea capillata* is positioned within the Bilaterian clade. Of the 9187 sites in the alignment, only 737 were amino acids for *C. capillata* (8% of the alignment length, versus a mean and median of 64% for all taxa), which is the fewest in the dataset. The ML root had a moderate rootstrap support value of 64.9, and 12/119 (10.1%) branches total received non-zero rootstrap support, while 11/119 (9.24%) were in the confidence set of the AU test.

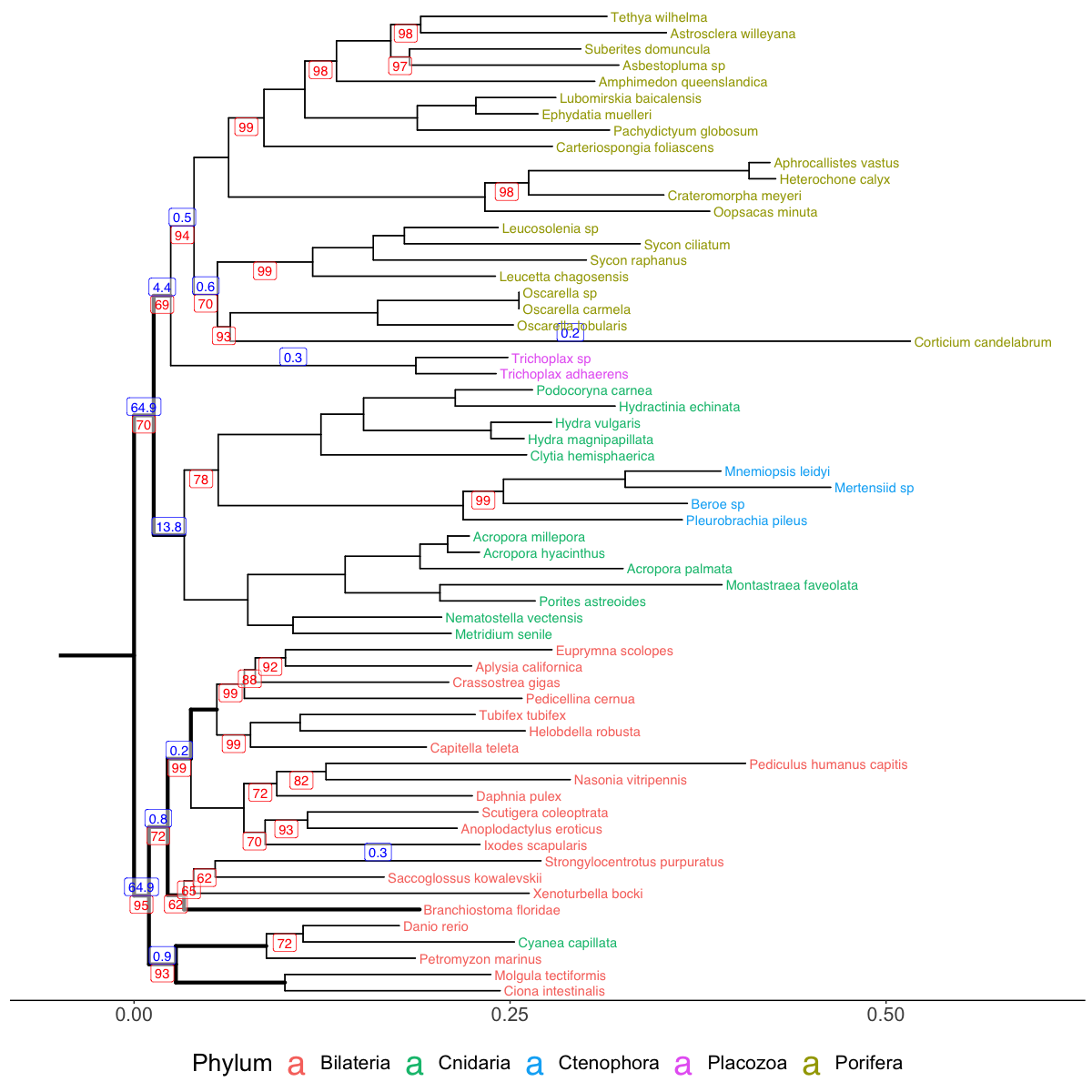


Figure 1. Maximum-likelihood tree inferred from the nonribosomal protein dataset of Nosenko et al. (2013), using a nonreversible substitution model estimated from the data with nQmaker (Dang et al., 2022). Rootstrap support values are shown in blue above the centre of a branch if they are greater than 0, and bootstrap support values are shown in red below the centre of a branch if they are below 100. Branches in the confidence set of the AU test are bold.

### Nosenko2013 Ribosomal Dataset

As shown in Figure 2, the ML root for the tree inferred using the Nosenko2013 ribosomal dataset is located on the branch leading to the Cnidarian *Acropora hyacinthus,* with a rootstrap support of 54.5. The tree also has a clade of Ctenophora and Placozoa as the sister to Porifera. 18/119 (15.1%) branches received non-zero rootstrap support, and 24/119 (20.2%) were in the confidence set of the AU test.

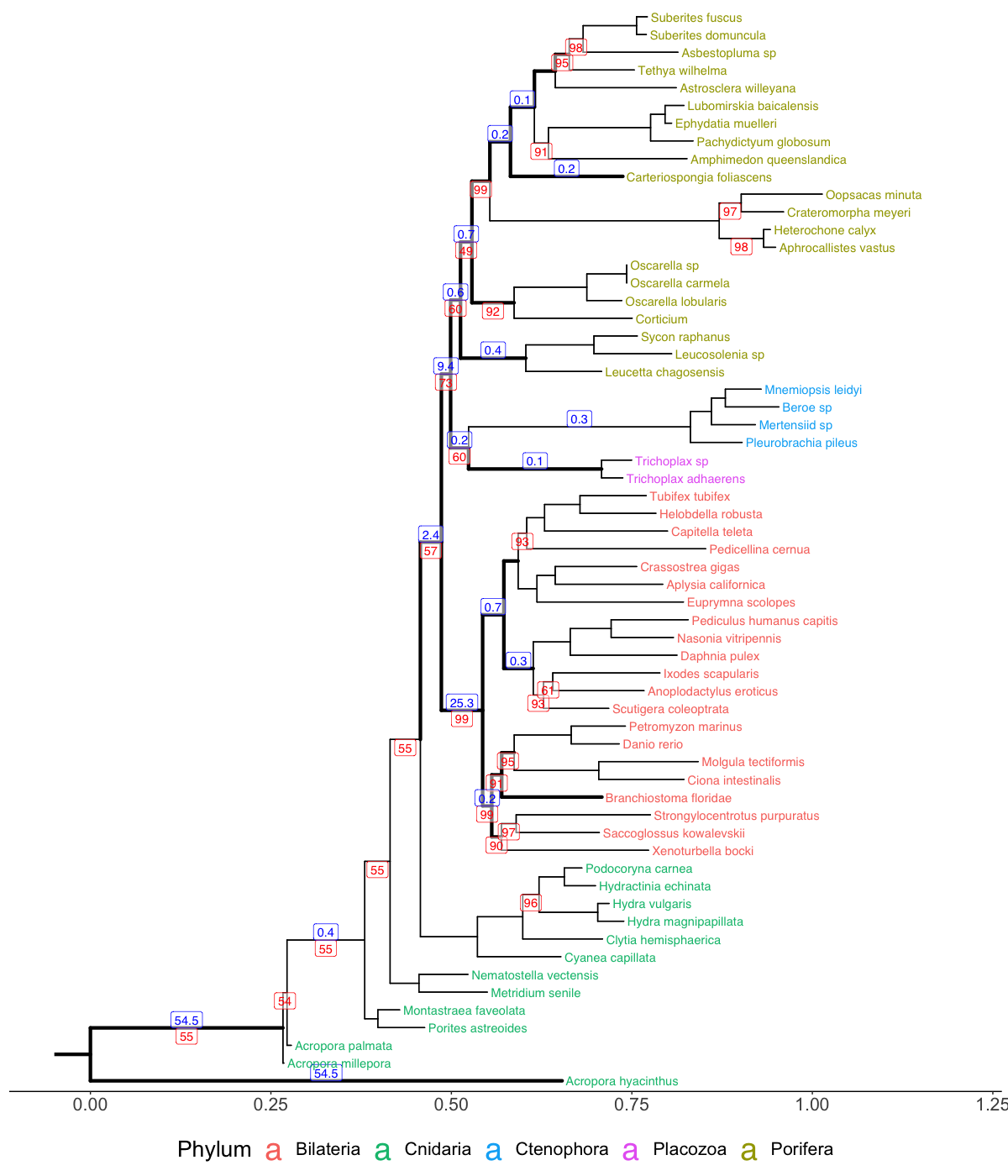


Figure 2. Maximum-likelihood tree inferred from the nonribosomal protein dataset of Nosenko et al. (2013), using a nonreversible substitution model estimated from the data with nQmaker (Dang et al., 2022). Rootstrap support values are shown in blue above the centre of a branch if they are greater than 0, and bootstrap support values are shown in red below the centre of a branch if they are below 100. Branches which were in the confidence set of the AU test are bold.

### Laumer2018 Dataset

The root in the ML tree of the Laumer2018 dataset places a monophyletic clade consisting of Porifera and Ctenophora as the sister group to the rest of Metazoa, though with a very low rootstrap support value of 36.3. Additionally, the tree has Ctenophora being a clade within Porifera, as the sister group to Calcarea. 5/101 (4.95%) branches were in the confidence set of the AU test, and 11/101 (10.9%) had non-zero rootstrap support.

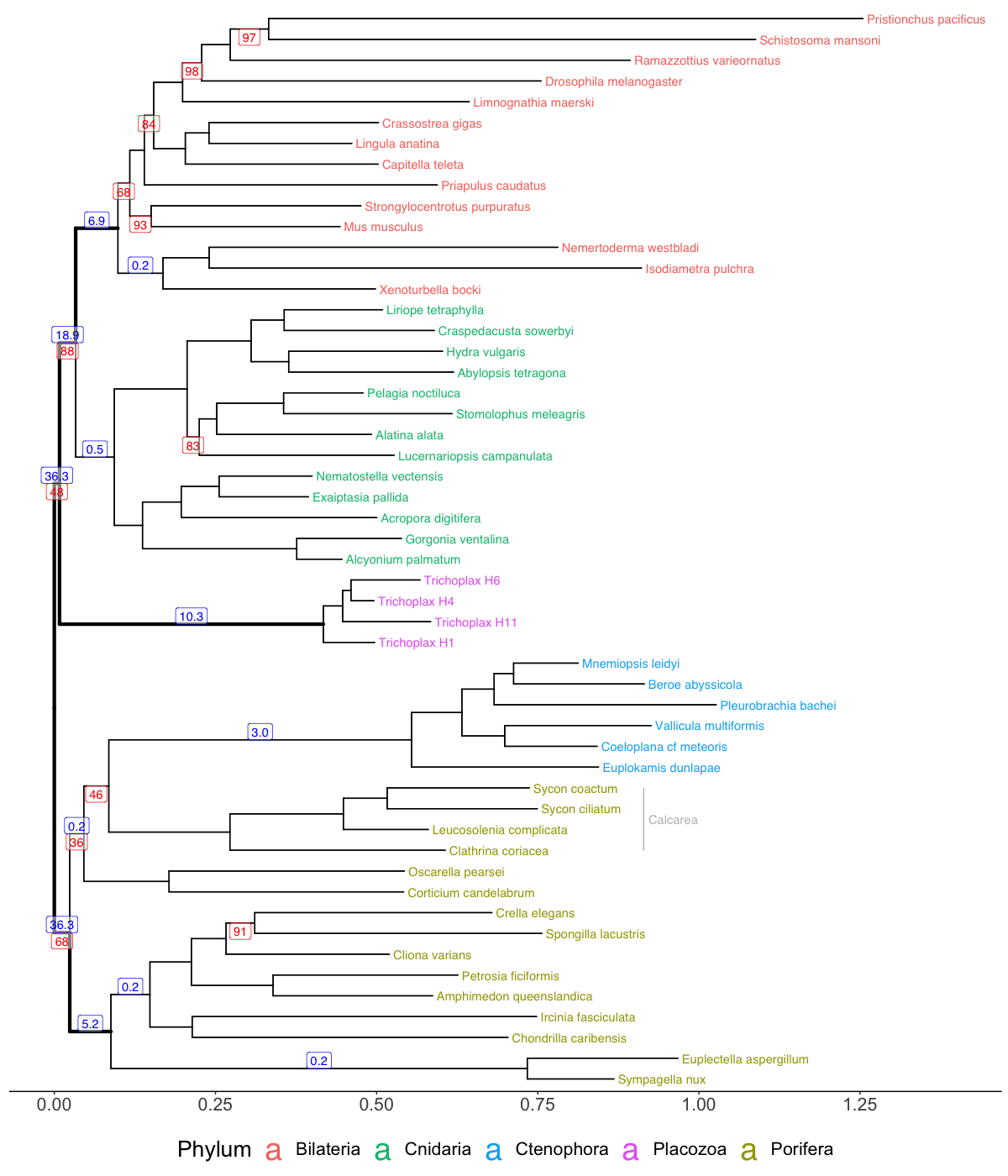


Figure 3. Maximum-likelihood tree inferred from the BUSCO protein dataset in Laumer et al. (2018) using a nonreversible substitution model estimated from the data with nQmaker (Dang et al., 2022). Rootstrap support values are shown in blue above the centre of a branch if they are greater than 0, and bootstrap support values are shown in red below the centre of a branch if they are below 100. Branches in the confidence set of the AU test are bold. The Poriferan class Calcarea is labelled in grey.

### Simion2017 Dataset

The ML tree for the Simion2017 dataset, shown in Figure 4, has the root in the middle of Porifera, with a clade comprised of Hexactinellida and Demospongiae (together known as Silicea) as the sister to all other Metazoans. This root has a moderate rootstrap support of 79.7. Overall, 5/141 (3.55%) of branches received non-zero rootstrap support, including the Ctenophora-sister root, the Ctenophora+Porifera-sister root, the Porifera-sister root, and the branch leading to “Calcarea+Homoscleromorpha” clade. Only 2/141 (1.42%) of branches were in the confidence set of the AU test – the ML root, and the Porifera-sister root. The Porifera-sister root also received the second highest rootstrap support in the tree, with a rootstrap support value of 10.2.

# 

Figure 4. Maximum-likelihood tree inferred from the protein dataset of Simion et al. (2017) using a nonreversible substitution model estimated from the data with nQmaker (Dang et al., 2022). Rootstrap support values are shown in blue above the centre of a branch if they are greater than 0, and bootstrap support values are shown in red below the centre of a branch if they are below 100. Branches in the confidence set of the AU test are bold. Poriferan clades, in grey, are labelled as per Simion et al. (2017).

## Model Evaluation

The BIC scores of both the reversible and nonreversible model were calculated for all datasets, and are shown in Table 2. Only the nonreversible model estimated using the Nosenko2013 nonribosomal dataset had a worse fit than its reversible counterpart, with the other three all having a lower BIC score for the nonreversible model.

Table 2. BIC scores for each dataset of both the nonreversible and reversible model, as reported in the summary ‘.iqtree’ file of each analysis generated automatically by IQ-TREE. The difference in score between the models is also shown: . A positive thus indicates that the nonreversible model fit the data better than the reversible model.

|  |  |  |  |
| --- | --- | --- | --- |
| **Dataset** | **BIC** | | **ΔBIC** |
| ***Reversible model*** | ***Nonreversible model*** |
| Nosenko2013 nonribosomal | 471496 | 472652 | -1156 |
| Nosenko2013 ribosomal | 822158 | 819768 | 2390 |
| Laumer2018 | 8308998 | 8304071 | 4926 |
| Simion2017 | 27432102 | 27348360 | 83742 |

## Site Removal

Figure 5 shows a pairwise analysis of ∆SLS values for the four different rooted topologies. It shows that generally, Ctenophora-sister (T2) was favoured by more sites – 58% compared to 33% against T1, 58% to 32% against T3, and 59% to 29% against T4 – but has a lower overall log-likelihood than T1 and T3. Additionally, the majority of sites in Figure 5 have a negligible ∆SLS – only a few show strong preference for each hypothesis. Differences in log-likelihood between all topologies are also all less than or equal to 21.

A screenshot of a computer

Description automatically generated

Figure 5. The distribution of site-wise log-likelihood scores (∆SLS) for each site in the Simion2017 dataset, for each choice of two out of four rooted topologies: the ML root (T1), Ctenophora-sister (T2), Porifera-sister (T3), and Ctenophora+Porifera-sister (T4). Green indicates sites that support the first listed topology, and red indicates support for the second. The overall difference in log-likelihood between the two topologies, ∆Log-L, is the sum of all ∆SLS values, or equivalently is the log-likelihood of the first topology listed (green), minus the log-likelihood of the second (red).

As the two roots in the confidence set of the AU test were the Porifera-sister root and the ML root, the distribution of ∆SLS for these two topologies – T3 and T1 respectively – was further investigated. Upon removing progressively more of the most informative sites, the rootstrap support value of the ML root went from 79.7 in the original tree (Figure 4), to 70.0 after removing 0.001% of sites, and to 56.2 after removing 0.01% of sites. The inferred topology then changed after removing 0.1% of sites, putting Ctenophora instead as the basal lineage, followed by paraphyletic sponges. Upon removing 1% of sites from the alignment, the topology changed again to place a monophyletic clade of Ctenophora and Hexactinellida as the sister to other Metazoans, followed by Demospongiae, then the rest of Porifera. These ML topologies, inferred for each amount of data removed, is shown in Figure 6.

# 

Figure 6. The ML trees inferred from the Simion2017 dataset after removing the top and bottom 0.0005%, 0.005%, 0.05%, and 0.5% of most influential sites, in accordance with their site-likelihood score difference between the ML paraphyletic-Porifera topology, and the Porifera-sister topology. The nonreversible model estimated by nQmaker for the full Simion2017 dataset was used in all analyses. Subplot titles indicate the proportion of the alignment which was removed. Tree leaves (i.e. terminal branches) are coloured in accordance with the phylum of the species they lead to. Rootstrap support values are shown in blue above the centre of a branch if they are greater than 0, and bootstrap support values are shown in red below the centre of a branch if they are below 100.

# Discussion

The aim of this research was to use nonreversible amino acid substitution models to root the Metazoan tree, using several previously published datasets. This was successfully done for each dataset, though the biological plausibility of the results varied. No analyses recovering phylogenies that are well-supported by any recent literature.

The Nosenko2013 nonribosomal dataset placed the root between Bilateria and the other phyla. Although this hypothesis lacks support in the literature, it has some biological plausibility. Both Placozoa and Porifera, unlike other Metazoan phyla, lack neurons, muscle cells, and internal digestive systems (Laumer et al., 2018). Their grouping in a clade would be consistent with a scenario in which the ancestral Metazoan possessed these traits, before they were lost by the ancestor of Porifera and Placozoa but retained in other phyla. Thus, despite the lack of support in the literature, these facets of the topology should not be dismissed immediately. However, the placement of the Cnidarian *C. capillata* as a sister to the Bilaterian *Danio rerio*, and the recovery of Ctenophora within Cnidaria, raise concerns. While the former could be attributed to poor sequence quality, the scientific consensus places Ctenophora as a separate phylum to Cnidaria (King & Rokas, 2017). In the original paper, Nosenko et al. (2013) used the nonribosomal dataset with the Bayesian CAT model (Lartillot & Philippe, 2004), recovering support for Ctenophora as the earliest-branching lineage, followed by Porifera, Placozoa, and then a clade of Bilateria and Cnidaria; this is extremely different to topology inferred in this report. The nonreversible model also had a higher BIC score than the reversible model, suggesting there was insufficient phylogenetic signal present to justify the large increase in parameters associated with nonreversible models (Naser-Khdour et al., 2022). This, together with the lack of supporting evidence in the literature, suggests little confidence should be placed in the accuracy of the Nosenko2013 nonribosomal tree.

The tree inferred from the Nosenko2013 ribosomal dataset placed the root on the terminal branch of the Cnidarian *Acropora hyacinthus*. Not only is this placement unprecedented in the literature, but the tree itself is biologically implausible, as it suggests that the distance between two closely related species, *A. hyacinthus* and *Acropora palmata*, is comparable to the length of the entire Metazoan tree (see Figure 2). In the original paper, Nosenko et al. (2013) applied the CAT model to the ribosomal dataset and recovered support for the Porifera-sister hypothesis—a topology that is markedly different from the one obtained in this report. Given the highly unusual and unlikely tree topology inferred here, the results of the Nosenko2013 ribosomal dataset are likely incorrect.

Despite using a much larger alignment than the two Nosenko2013 datasets, the tree inferred with the Laumer2018 dataset also had a topology unsupported by the literature. The basal lineage was found to be a monophyletic clade of Porifera and Ctenophora. While this has been found before (e.g. Francis & Canfield, 2020), this has only ever been with Ctenophora and Porifera as distinct phyla. In this analysis, however, Ctenophora was placed not as the sister group to all of Porifera, but specifically to Calcarea. While such extensive evolutionary changes could theoretically occur over the long branch separating Ctenophora from Porifera, there is no supporting evidence for this scenario in the literature. Thus, the topology inferred from the nonreversible model estimated using the Laumer2018 dataset is also unlikely to be correct. For comparison, the original paper used a CAT model to recover the branching lineages, in order, as Ctenophora, Porifera, then Placozoa, leaving a clade of Bilateria and Cnidaria remaining (Laumer et al., 2018).

The tree inferred using the Simion2017 dataset was the only one to have an identical unrooted topology as its original paper. Correspondingly, all species were placed in their commonly agreed upon phyla. The tree has a clade of Demospongeia and Hexactinellida – referred to as Silicea by Botting and Muir (2018) – as the sister to all other Metazoans. This would suggest that the ancestral Metazoan was likely a sponge (Botting & Muir, 2018). Support for this root was relatively robust to resampling of the data, with a rootstrap support value of 79.7 – the highest rootstrap support for the ML root of any of the four trees in this report. While this root position has been recovered before (e.g. Sperling et al., 2009; Sperling et al., 2007), more recent phylogenetic studies invariably recover Porifera as a monophyletic group (Botting & Muir, 2018). This includes the paper the dataset is from, which supported a Porifera-sister topology, again using a CAT model (Simion et al., 2017). Hence, despite the greater statistical support for the ML root, the literature still suggests that the results of the Simion2017 dataset are unlikely to be correct.

Since the Simion2017 tree was the only one to recover a topology previously observed in the literature, it was the only tree subjected to the post hoc analysis of ∆SLS values. This analysis revealed that when comparing the SLS scores of Ctenophora-sister to both the Porifera-sister topology and the ML topology, Ctenophora-sister consistently had support from a greater proportion of sites, despite its lower overall log-likelihood. This is consistent with Shen et al. (2017), who found that Ctenophora-sister was supported by a greater percentage of sites in their alignment than both Porifera-sister and Ctenophora+Porifera-sister. This could explain why the removal of sites with the top and bottom 0.05% of ∆SLS values for the T1 and T3 topologies caused the tree to shift to Ctenophora-sister. However, after removing the top and bottom 0.5% of sites, the sister group of Metazoa shifted once again to be a “Ctenophora+Hexactinillida” clade. The fact the original ML topology was not robust to the removal of a small fraction of sites, and the small differences in log-likelihood between the ML tree and other topologies in generally, suggests again that limited confidence should be placed in its accuracy. As the majority of the phylogenetic signal was located in a few sites, which hypothesis was recovered was highly dependent on which sites were included in the analysis, a finding shared by Francis and Canfield (2020). This may also explain why the rootstrap support value of the ML root was only moderate, since subsamples of the data which randomly did not contain these few, highly influential sites, may have led to other root placements. Thus, the Metazoan tree was also unable to be conclusively rooted by the nonreversible model when using the Simion2017 dataset.

Considering the results of all four datasets, the nonreversible models used in this research do not appear to be particularly effective at rooting the Metazoan tree, as they were unable to recover a topology which has been supported by any recent phylogenetic study. One reason for this may be how distantly related species in the Metazoan tree are. Bettisworth and Stamatakis (2021) suggested that nonreversible rooting may be less effective when there are few, distantly related species. Additionally, though Naser-Khdour et al. (2022) were able to repeatedly root trees correctly, the clades they tested in were all within Mammalia, which diverged much more recently than Metazoa (King & Rokas, 2017). The analysis of the entirety of Metazoa presented here correspondingly involves much greater evolutionary distances; thus, the relative lack of power of the nonreversible rooting method, together with the diluted phylogenetic signal involved with such large timescales, may have contributed to difficulties in the nonreversible model being able to effectively discern the true root. This theory is supported by the fact that as the size of the dataset increased, so too did the statistical confidence in the ML root. While the nonribosomal and ribosomal Nosenko2013 datasets had 9.24% and 20.2% of branches in the confidence set of the AU test respectively, the Laumer2018 dataset had 4.95%, and the Simion2017 dataset only had 1.42%. This suggests that the larger datasets contained more phylogenetic signal, that then enabled the AU test to reject more potential root positions. However, even if even larger alignments are used, the results of this analysis do not suggest that accuracy of the inferred root would necessarily improve.

Nonreversible models generally fit the datasets used better than the reversible models. This is demonstrated by Table 2, in which the BIC scores of the reversible model was better for the Nosenko2013 nonribosomal dataset only, while the BIC scores indicated better fit of the nonreversible model for the three larger datasets. This has been found by other recent studies into nonreversible models, and highlights that the reversibility assumption common to most phylogenetic models is often violated by the data they are applied to (Dang et al., 2022; Naser-Khdour et al., 2022). However, as the reversible model used here allows each partition to have a different rate matrix, while only a single rate matrix is used across the whole alignment in the nonreversible model, it is possible that this could have caused the reversible models’ BIC scores to be better than they would be otherwise (Naser-Khdour et al., 2022).

Regardless of their relative model fit, the results of the nonreversible and reversible models were identical for each dataset (ignoring the root placement, as this was not inferred in the reversible analyses). Thus, only the reversible analysis of the Simion2017 dataset produced an unrooted topology that agrees with the literature – depending on the root placement, the ML tree (shown rooted in Figure 4) could be Ctenophora-sister or Porifera-sister. Conversely, the CAT models used by all of the original papers for each dataset were able to produce well-supported phylogenies (Laumer et al., 2018; Nosenko et al., 2013; Simion et al., 2017). CAT models were developed in a Bayesian framework, and they allow sites to evolve under different substitution models; in particular, the frequency vector, or stationary distribution, of the model for a particular site is allowed to vary (Lartillot & Philippe, 2004). Models which do this are known as profile mixture models, and were borne out of the observation that for some sites, only a specific subset of residues can be favourably selected for by evolution (Baños et al., 2024). Profile mixture models have been implemented in a ML framework for some time for reversible substitution models (e.g. Quang et al., 2008), but not for nonreversible ones. Given they have generally been found to provide better fit to data (Baños et al., 2024), and as the CAT model proved more reliably accurate than both of the non-profile mixture models for all datasets in this report, a logical direction for future research is to develop nonreversible profile mixture models. This would still allow trees to be rooted without using outgroups, whilst potentially recovering more accurate phylogenies.

For the Metazoan tree in general, as traditional molecular phylogenetics methods involving DNA and protein sequences have failed to provide a resolution, using information about organisms which changes at a slower rate than these sequences, could mean more phylogenetic signal remains with which to root the tree. For example, Schultz et al. (2023) analysed historical chromosomal rearrangements of sections of the genome, as these are very unlikely to be reversed, showing support for Ctenophora-sister. Similarly analysing slower-evolving traits, such as the structure of encoded proteins rather than the protein sequence themselves, could similarly prove insightful (Malik et al., 2020). If enough such methods are developed, and are shown to be accurate for resolved phylogenies, whilst consistently supporting the same hypothesis, there may still be some hope for a definitive answer to which is the basal Metazoan lineage.

# References

Baños, H., Susko, E., & Roger, A. J. (2024). Is over-parameterization a problem for profile mixture models? *Systematic Biology*, *73*(1), 53-75. https://doi.org/10.1093/sysbio/syad063

Barrett, T., Dowle, M., Srinivasan, A., Gorecki, J., Chirico, M., Hocking, T., & Schwendinger, B. (2024). *data.table: Extension of `data.frame`*.In https://CRAN.R-project.org/package=data.table

Bergsten, J. (2005). A review of long-branch attraction. *Cladistics*, *21*(2), 163-193. https://doi.org/10.1111/j.1096-0031.2005.00059.x

Bettisworth, B., & Stamatakis, A. (2021). Root Digger: A root placement program for phylogenetic trees. *BMC Bioinformatics*, *22*(1), 225. https://doi.org/10.1186/s12859-021-03956-5

Botting, J. P., & Muir, L. A. (2018). Early sponge evolution: A review and phylogenetic framework. *Palaeoworld*, *27*(1), 1-29. https://doi.org/10.1016/j.palwor.2017.07.001

Burkhardt, P. (2022). Ctenophores and the evolutionary origin(s) of neurons. *Trends in Neurosciences*, *45*(12), 878-880. https://doi.org/10.1016/j.tins.2022.09.001

Charif, D., & Lobry, J. R. (2007). SeqinR 1.0-2: A contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In U. Bastolla, M. Porto, H. E. Roman, & M. Vendruscolo (Eds.), *Structural approaches to sequence evolution: Molecules, networks, populations* (pp. 207-232). Springer Verlag.

Chernomor, O., von Haeseler, A., & Minh, B. Q. (2016). Terrace aware data structure for phylogenomic inference from supermatrices. *Systematic Biology*, *65*(6), 997-1008. <https://doi.org/10.1093/sysbio/syw037>

Cherryh, C. (2024). Unpublished PhD thesis. Australian National University.

Cherryh, C. (2023). *util\_tree\_processing.R*.In <https://github.com/caitlinch/metazoan-mixtures/blob/main/code/util_tree_processing.R>

Dang, C. C., Minh, B. Q., McShea, H., Masel, J., James, J. E., Vinh, L. S., & Lanfear, R. (2022). nQMaker: Estimating time nonreversible amino acid substitution models. *Systematic Biology*. https://doi.org/10.1093/sysbio/syac007

Dunn, C. W., Hejnol, A., Matus, D. Q., Pang, K., Browne, W. E., Smith, S. A., Seaver, E., Rouse, G. W., Obst, M., Edgecombe, G. D., Sorensen, M. V., Haddock, S. H. D., Schmidt-Rhaesa, A., Okusu, A., Kristensen, R. M., Wheeler, W. C., Martindale, M. Q., & Giribet, G. (2008). Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature*, *452*(7188), 745-749. https://doi.org/10.1038/nature06614

Dunn, C. W., Leys, S. P., & Haddock, S. H. D. (2015). The hidden biology of sponges and ctenophores. *Trends in Ecology & Evolution*, *30*(5), 282-291. https://doi.org/10.1016/j.tree.2015.03.003

Felsenstein, J. (1981). Evolutionary trees from DNA-sequences: A maximum-likelihood approach. *Journal of Molecular Evolution*, *17*(6), 368-376. https://doi.org/10.1007/Bf01734359

Feuda, R., Dohrmann, M., Pett, W., Philippe, H., Rota-Stabelli, O., Lartillot, N., Wörheide, G., & Pisani, D. (2017). Improved modeling of compositional heterogeneity supports sponges as sister to all other animals. *Current Biology*, *27*(24), 3864-3870. https://doi.org/10.1016/j.cub.2017.11.008

Francis, W. R., & Canfield, D. E. (2020). Very few sites can reshape the inferred phylogenetic tree. *PeerJ*, *8*. https://doi.org/10.7717/peerj.8865

Hejnol, A., Obst, M., Stamatakis, A., Ott, M., Rouse, G. W., Edgecombe, G. D., Martinez, P., Baguñá, J., Bailly, X., Jondelius, U., Wiens, M., Müller, W. E. G., Seaver, E., Wheeler, W. C., Martindale, M. Q., Giribet, G., & Dunn, C. W. (2009). Assessing the root of bilaterian animals with scalable phylogenomic methods. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1677), 4261-4270. https://doi.org/10.1098/rspb.2009.0896

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, *35*(2), 518-522. https://doi.org/10.1093/molbev/msx281

Huelsenbeck, J. P., Bollback, J. P., & Levine, A. M. (2002). Inferring the root of a phylogenetic tree. *Systematic Biology*, *51*(1), 32-43. https://doi.org/10.1080/106351502753475862

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587-+. https://doi.org/10.1038/Nmeth.4285

King, N., & Rokas, A. (2017). Embracing Uncertainty in Reconstructing Early Animal Evolution. *Current Biology*, *27*(19), R1081-R1088. https://doi.org/10.1016/j.cub.2017.08.054

Lartillot, N., & Philippe, H. (2004). A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Molecular Biology and Evolution*, *21*(6), 1095-1109. https://doi.org/10.1093/molbev/msh112

Laumer, C. E., Fernández, R., Lemer, S., Combosch, D., Kocots, K. M., Riesgo, A., Andrade, S. C. S., Sterrer, W., Sorensen, M. V., & Giribet, G. (2019). Revisiting metazoan phylogeny with genomic sampling of all phyla. *Proceedings of the Royal Society B-Biological Sciences*, *286*(1906). https://doi.org/10.1098/rspb.2019.0831

Laumer, C. E., Gruber-Vodicka, H., Hadfield, M. G., Pearse, V. B., Riesgo, A., Marioni, J. C., & Giribet, G. (2018). Support for a clade of Placozoa and Cnidaria in genes with minimal compositional bias. *Elife*, *7*. https://doi.org/10.7554/eLife.36278

Li, Y. N., Shen, X. X., Evans, B., Dunn, C. W., & Rokas, A. (2021). Rooting the animal tree of life. *Molecular Biology and Evolution*, *38*(10), 4322-4333. https://doi.org/10.1093/molbev/msab170

Malik, A. J., Poole, A. M., & Allison, J. R. (2020). Structural phylogenetics with confidence. *Molecular Biology and Evolution*, *37*(9), 2711-2726. https://doi.org/10.1093/molbev/msaa100

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, *37*(5), 1530-1534. https://doi.org/10.1093/molbev/msaa015

Naser-Khdour, S., Minh, B. Q., & Lanfear, R. (2022). Assessing confidence in root placement on phylogenies: An empirical study using nonreversible models for mammals. *Systematic Biology*, *71*(4), 959-972. https://doi.org/10.1093/sysbio/syab067

Nosenko, T., Schreiber, F., Adamska, M., Adamski, M., Eitel, M., Hammel, J., Maldonado, M., Müller, W. E. G., Nickel, M., Schierwater, B., Vacelet, J., Wiens, M., & Wörheide, G. (2013). Deep metazoan phylogeny: When different genes tell different stories. *Molecular Phylogenetics and Evolution*, *67*(1), 223-233. https://doi.org/10.1016/j.ympev.2013.01.010

Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*, 526-528. https://doi.org/10.1093/bioinformatics/bty633

Pedersen, T. L. (2024). *patchwork: The Composer of Plots*.In https://CRAN.R-project.org/package=patchwork

Philippe, H., Brinkmann, H., Lavrov, D. V., Littlewood, D. T. J., Manuel, M., Wörheide, G., & Baurain, D. (2011). Resolving difficult phylogenetic questions: Why more sequences are not enough. *Plos Biology*, *9*(3). https://doi.org/10.1371/journal.pbio.1000602

Philippe, H., Derelle, R., Lopez, P., Pick, K., Borchiellini, C., Boury-Esnault, N., Vacelet, J., Renard, E., Houliston, E., Quéinnec, E., Da Silva, C., Wincker, P., Le Guyader, H., Leys, S., Jackson, D. J., Schreiber, F., Erpenbeck, D., Morgenstern, B., Wörheide, G., & Manuel, M. (2009). Phylogenomics revives traditional views on deep animal relationships. *Current Biology*, *19*(8), 706-712. https://doi.org/10.1016/j.cub.2009.02.052

Pick, K. S., Philippe, H., Schreiber, F., Erpenbeck, D., Jackson, D. J., Wrede, P., Wiens, M., Alié, A., Morgenstern, B., Manuel, M., & Wörheide, G. (2010). Improved phylogenomic taxon sampling noticeably affects nonbilaterian relationships. *Molecular Biology and Evolution*, *27*(9), 1983-1987. https://doi.org/10.1093/molbev/msq089

Pisani, D., Pett, W., Dohrmann, M., Feuda, R., Rota-Stabelli, O., Philippe, H., Lartillot, N., & Wörheide, G. (2015). Genomic data do not support comb jellies as the sister group to all other animals. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(50), 15402-15407. https://doi.org/10.1073/pnas.1518127112

Quang, L. S., Gascuel, O., & Lartillot, N. (2008). Empirical profile mixture models for phylogenetic reconstruction. *Bioinformatics*, *24*(20), 2317-2323. https://doi.org/10.1093/bioinformatics/btn445

R Core Team. (2023). *R: A language and environment for statistical computing*.In R Foundation for Statistical Computing. https://www.R-project.org/

Redmond, A. K., & McLysaght, A. (2022). Evidence for sponges as sister to all other animals from partitioned phylogenomics with mixture models and recoding. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-33707-w

Ryan, J. F., Pang, K., Schnitzler, C. E., Nguyen, A. D., Moreland, R. T., Simmons, D. K., Koch, B. J., Francis, W. R., Havlak, P., Smith, S. A., Putnam, N. H., Haddock, S. H. D., Dunn, C. W., Wolfsberg, T. G., Mullikin, J. C., Martindale, M. Q., Baxevanis, A. D., & Progra, N. C. S. (2013). The genome of the ctenophore Mnemiopsis leidyi and its implications for cell type evolution. *Science*, *342*(6164), Article 1242592. https://doi.org/10.1126/science.1242592

Schierwater, B., Eitel, M., Jakob, W., Osigus, H. J., Hadrys, H., Dellaporta, S. L., Kolokotronis, S. O., & DeSalle, R. (2009). Concatenated analysis sheds light on early metazoan evolution and fuels a modern "Urmetazoon" hypothesis. *Plos Biology*, *7*(1), 36-44. https://doi.org/10.1371/journal.pbio.1000020

Schultz, D. T., Haddock, S. H. D., Bredeson, J. V., Green, R. E., Simakov, O., & Rokhsar, D. S. (2023). Ancient gene linkages support ctenophores as sister to other animals. *Nature*, *618*(7963), 110-117. https://doi.org/10.1038/s41586-023-05936-6

Shen, X. X., Hittinger, C. T., & Rokas, A. (2017). Contentious relationships in phylogenomic studies can be driven by a handful of genes. *Nature Ecology & Evolution*, *1*(5). https://doi.org/10.1038/s41559-017-0126

Shimodaira, H. (2002). An approximately unbiased test of phylogenetic tree selection. *Systematic Biology*, *51*(3), 492-508. https://doi.org/10.1080/10635150290069913

Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210-3212. https://doi.org/10.1093/bioinformatics/btv351

Simion, P., Philippe, H., Baurain, D., Jager, M., Richter, D. J., Di Franco, A., Roure, B., Satoh, N., Quéinnec, E., Ereskovsky, A., Lapébie, P., Corre, E., Delsuc, F., King, N., Wörheide, G., & Manuel, M. (2017). A large and consistent phylogenomic dataset supports sponges as the sister group to all other animals. *Current Biology*, *27*(7), 958-967. https://doi.org/10.1016/j.cub.2017.02.031

Slowikowski, K. (2024). *ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'*.In https://CRAN.R-project.org/package=ggrepel

Smith, M. R. (2019). *TreeTools: create, modify and analyse phylogenetic trees*.In Comprehensive R Archive Network.

Sperling, E. A., Peterson, K. J., & Pisani, D. (2009). Phylogenetic-signal dissection of nuclear housekeeping genes supports the paraphyly of sponges and the monophyly of eumetazoa. *Molecular Biology and Evolution*, *26*(10), 2261-2274. https://doi.org/10.1093/molbev/msp148

Sperling, E. A., Pisani, D., & Peterson, K. J. (2007). Poriferan paraphyly and its implications for Precambrian palaeobiology. *Geological Society, London, Special Publications*, *286*(1), 355-368. https://doi.org/doi:10.1144/SP286.25

Sumner, J. G., Jarvis, P. D., Fernández-Sánchez, J., Kaine, B. T., Woodhams, M. D., & Holland, B. R. (2012). Is the General Time-Reversible Model Bad for Molecular Phylogenetics? *Systematic Biology*, *61*(6), 1069-1074. https://doi.org/10.1093/sysbio/sys042

Wainright, P. O., Hinkle, G., Sogin, M. L., & Stickel, S. K. (1993). Monophyletic origins of the metazoa: An evolutionary link with fungi. *Science*, *260*(5106), 340-342. https://doi.org/10.1126/science.8469985

Wang, L. G., Lam, T. T. Y., Xu, S., Dai, Z., Zhou, L., Feng, T., Guo, P., Dunn, C. W., Jones, B. R., Bradley, T., Zhu, H., Guan, Y., Jiang, Y., & Yu, G. (2020). treeio: an R package for phylogenetic tree input and output with richly annotated and associated data. *Molecular Biology and Evolution*, *37*(2), 599-603. https://doi.org/10.1093/molbev/msz240

Whelan, N. V., Kocot, K. M., Moroz, L. L., & Halanych, K. M. (2015). Error, signal, and the placement of Ctenophora sister to all other animals. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(18), 5773-5778. https://doi.org/10.1073/pnas.1503453112

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., . . . Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., Pedersen, T. L., & Seidel, D. (2023). *scales: Scale Functions for Visualization*.In https://CRAN.R-project.org/package=scales

Wilke, C. O. (2024). *cowplot: Streamlined plot theme and plot annotations for 'ggplot2'*.In https://CRAN.R-project.org/package=cowplot

Wilke, C. O., & Wiernik, B. M. (2022). *ggtext: Improved Text Rendering Support for 'ggplot2'*.In https://CRAN.R-project.org/package=ggtext

Yap, V. B., & Speed, T. (2005). Rooting a phylogenetic tree with nonreversible substitution models. *BMC Evolutionary Biology*, *5*, 2. https://doi.org/10.1186/1471-2148-5-2

Yu, G. C., Smith, D. K., Zhu, H. C., Guan, Y., & Lam, T. T. Y. (2017). ggtree: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, *8*(1), 28-36. https://doi.org/10.1111/2041-210x.12628