




## Applications

# Median-Joining Networks and Bayesian Phylogenies Often Do Not Tell the Same Story

Sungsik Kong<sup>1,2</sup> , Santiago J. Sánchez-Pacheco<sup>3</sup> , Robert Murphy<sup>1,4</sup> 

<sup>1</sup> Department of Ecology and Evolutionary Biology, University of Toronto, <sup>2</sup> Wisconsin Institute for Discovery, University of Wisconsin-Madison, <sup>3</sup> Dirección Académica, Universidad Nacional de Colombia, Sede de La Paz, <sup>4</sup> Reptilia Sanctuary and Education Centre, Vaughan

Keywords: Approximately Unbiased test, Bayesian phylogenetics, phenetics, phylogenetics, Shimodaira-Hasegawa test

<https://doi.org/10.18061/bssb.v2i1.9625>

---

## Bulletin of the Society of Systematic Biologists

---

### Abstract

Inferring phylogenies among intraspecific individuals often yields unresolved relationships (i.e., polytomies). Consequently, methods that compute distance-based abstract networks, like Median-Joining Networks (MJNs), are thought to be more appropriate tools for reconstructing such relationships than traditional trees. Median-Joining Networks visualize all routes of relationships in the form of cycles, if needed, when traditional approaches cannot resolve them. However, the MJN method is a distance-based phenetic approach that does not involve character transformations and makes no reference to ancestor–descendant relationships. Although philosophical and theoretical arguments challenging the implication that MJNs reflect phylogenetic signal in the traditional sense have been presented elsewhere, an empirical comparison with a character-based approach is needed given the increasing popularity of MJN analysis in evolutionary biology. Here, we use the conservative Approximately Unbiased (AU) test to compare 85 cases of branching patterns of cycle-free MJNs and Bayesian Inference (BI) phylogenies using datasets from 55 empirical studies. By rooting the MJN analyses to provide directionality, we report substantial disagreement between computed MJNs and posterior distributions on BI phylogenies. The branching patterns in MJNs and BI phylogenies show significantly different relationships in 37.6% of cases. Among the relationships that do not significantly differ, 96.2% show alternative sets of relationships. Our results indicate that the two methods provide different measures of relatedness in a phylogenetic sense. Finally, our analyses also support previous observations of the statistical hypothesis testing by reconfirming the over-conservativeness of the Shimodaira-Hasegawa test versus the AU test.

## 1 Introduction

Phylogenetic inference involves statistical analysis of the differences among taxa and then represents their evolutionary relationships in a branch-like structure (Felsenstein, 2004). The tree model strictly assumes vertical evolution where a single ancestral taxon splits into two daughter taxa that denotes a speciation event. The resulting diagram consisting of a set of bifurcations often sufficiently describes speciation events but overlook common and important evolutionary histories that also generate novel lineages, like hybridization, due to structural constraints. The ability to infer and illustrate such complex evolutionary phenomena is a critical component for the expansion of evolutionary research (Blair & Ané, 2019; Kong, 2022).

The concept of a phylogenetic network has been proposed as an alternative to the tree model. In brief, a phy-

logenetic network extends the tree model by considering reticulation events (e.g., hybrid speciation, introgression, gene flow, horizontal gene transfer, or recombination) in the form of a directed acyclic graph with at least one vertex with in-degree two and out-degree one. While Huson & Bryant (2005) noted the term network is often described in a way that the authors happen to be interested in practice (e.g., hybridization network (Linder & Rieseberg, 2004) or recombination network (Gusfield & Bansal, 2005)), substantial progress has been made in recent years to establish a solid concept of networks (see Kong (2022) for a review of the various classes of phylogenetic networks and their biological significance).

Networks are largely classified into explicit and abstract (or implicit) networks. It is important to stress that the former model non-treelike evolution due to the complex evolutionary events (hence, they are *true* phylogenetic net-



works), whereas the latter depict non-treelike signal in the data but do not model biological cause of those signals (Degnan, 2018). This is because the abstract networks often focus on whether a distance matrix can be fit onto a topology in a mathematical sense, regardless of the biological mechanism (Bapteste et al., 2013; Debevec & Whitfield, 2012; Huson et al., 2010; Huson & Scornavacca, 2010). In short, no historical ancestor–descendent relationships can be inferred on abstract networks because the non-terminal nodes do not represent ancestral taxa (Solís-Lemus & Ané, 2016). Many efforts to efficiently infer explicit networks have been made (e.g., PhyloNet (Than et al., 2008; Wen & Nakhleh, 2017); SNaQ (Solís-Lemus & Ané, 2016); PhyNEST (Kong et al., 2022)). However, current inference methods require high computational capability, and they become very expensive even to infer the relationships among two dozens of taxa (Hejase & Liu, 2016). In contrast, abstract networks are fast to compute and visualize possible routes of relationships of the nucleotide arrangements in the input sequence alignment (or a set of gene trees) based on overall similarity.

The Median-Joining Network (MJN) is one of the most popular methods to compute abstract networks. The method clusters sequences on the basis of overall similarity (see Kong et al. (2015) for operational details focusing on MJN’s phenetic nature), and the lack of a root precludes identification of the direction of evolution (Sánchez-Pacheco et al., 2020). A rooting option is available when computing MJNs, but this procedure is often bypassed in practice. Moreover, Kong et al. (2015) and Sánchez-Pacheco et al. (2020) argue that the rooting procedure in MJ does not appropriately polarize the character transformation. The cyclic structure in MJNs must be distinguished from the reticulation in explicit networks as they merely illustrate the algorithm’s inability to discern optimal connections (Salzburger et al., 2011). Despite these flaws, the application of MJNs is not waning but rather flourishing with > 12,000 citations as of writing, probably because of the fast computation and freely available, simple to use software NETWORK (available on <http://www.fluxus-engineering.com>; Bandelt et al., 1999).

Bayesian inference (BI) is a widely used method to reconstruct phylogenetic trees (Huelsenbeck & Ronquist, 2001; Yang & Rannala, 1997). Sampling the posterior probability in BI methods is not trivial, and computational efficiency is achieved through application of Markov chain Monte Carlo (MCMC) that approximates the posterior distribution of trees (Li et al., 2000; Mau et al., 1999; Mau & Newton, 1997; Yang & Rannala, 1997). However, BI phylogenies may be unsuitable to infer relationships among closely related intraspecific individuals because the produced tree will have poor resolution as BI samples multiple trees that require merging using some consensus criteria (Posada & Crandall, 2001). Moreover, setting appropriate prior parameters to conduct BI approaches is not always straightforward, and the estimated posterior probability can be excessively liberal when compared to the bootstrap support from a maximum likelihood phylogeny (Suzuki et al., 2002). Nevertheless, BI is considered as one of the most

reliable methods available today for phylogenetic tree reconstruction.

Branching patterns among a set of phylogenetic trees with the same ingroup taxa often differ due to random chance, sampling error, model misidentification, and many others factors. Techniques for measuring and testing the significance of topological incongruence among a set of topologies are used widely. For example, the Shimodaira-Hasegawa (SH) test (Shimodaira & Hasegawa, 1999) is a method that uses non-parametric bootstrapping that compares topologies in a likelihood framework. This test can compare multiple topologies considering the null hypothesis that all the trees tested are equally good explanations of the data, whereas the alternative hypothesis is that one or several trees are better representations of the data. Because the SH test can be very conservative in its rejection of the null hypothesis when the number of candidate trees is large, the Approximately Unbiased (AU) test (Shimodaira, 2002; Shimodaira & Hasegawa, 2001) is proposed in an attempt to ameliorate this bias.

Here, we ask if the branching pattern computed from the MJ algorithm is statistically different from that of estimated through BI analysis. Because the true relationships are unknown, we employ the AU test to statistically test for the significance of the discrepancies. With a strong assumption that the character-based, philosophically defensible BI method always produces a topology that is closer to the true relationship (i.e., actual pattern of ancestor–descendant relationships), the statistical differences between the BI and MJN topologies rejects the null hypothesis that MJN explains the data as good as BI phylogeny, which can be interpreted as the relationship depicted in MJN is unlikely to be a true evolutionary relationships among the taxa. We evaluate the performance of MJN analyses by scoring the percentages where two topologies differ significantly. Through this study, we aim to show that MJN is not an appropriate tool to estimate phylogenetic relationship, thus evolutionary interpretation solely based on MJN can be misleading.

## 2 Methods

### 2.1 Data collection

We collected a set of published studies that computed MJNs in their analyses. We particularly looked for MJNs with no or a few cycles. While star-like MJNs, in which all nodes are individually connected to a central connection node that usually occur when there is a very small variation between sequences (i.e., one to a few nucleotide differences), were most frequent in the literature, we avoided them as they are likely to yield a completely unresolved BI tree (pers. obs.) The star-like MJNs often lack phylogenetic information; however, Bandelt et al. (1995) argue that such topologies characterize demographic expansions around a founder population, which is represented by the central node, in the past. The selected studies were further filtered by the availability of GenBank accession numbers for the DNA sequences used to compute the MJNs.

The outgroup sequences were also retrieved when available in the original article, although they were not available in most cases because MJ does not conduct outgroup rooting for its calculation (Kong et al., 2015; Sánchez-Pacheco et al., 2020). When the outgroup was not specified, we selected outgroup sequences using two strategies: (1) we identified outgroup sequences from other studies on the same organisms in interest or (2) we manually selected them via Basic Local Alignment Search Tool (BLAST) implemented in GENEIOUS R6 version 6.1.8 (available on <https://www.geneious.com>; Kearsse et al., 2012). Because the ingroup sequences were often from a single (or closely related) species that are expected to be very similar with each other, outgroup taxa that were excessively closely or distantly related were not expected to be very useful in reconstructing relationships. We selected multiple taxa within the sister group when possible, which produced more consistent results than any other outgroup selection strategies (Luo et al., 2010), preferably those with low pairwise identical sites (i.e., < 80%) and/or with high identical site (i.e., > 95%) compared to the ingroup sequences.

Multiple sequence alignment was performed with MUSCLE (Edgar, 2004) implemented in GENEIOUS R6. All alignments were performed using default parameters, and they were visually inspected and manually corrected whenever appropriate. Two sets of alignments, one without outgroup (denoted by Alignment 1 hereafter) and another with outgroup (Alignment 2), were saved.

## 2.2 Topology inference

For each Alignment 1, we computed MJN using NETWORK version 4.6.1.2 (Bandelt et al., 1999) with the default setting, where a weight applied to each character state (i.e., nucleotide) and an explicit parameter  $\epsilon$  were set to 10 and 0, respectively. An equal weight was applied to all character differences in the alignment to the computation of MJNs, although Bandelt et al. (1999) suggest to increase the value in case some character states were more informative in identifying the clusterings than the others, assuming mutations in conservative sites were more valuable than the mutations in hypervariable sites. The explicit parameter  $\epsilon$  specified a weighted genetic distance to the sampled sequences in the dataset, within which potential median vectors (unsampled hypothetical vectors that linked either sampled vectors, median vectors, or both) may have been constructed. For example, a greater value of  $\epsilon$  would have resulted in the construction of more median vectors between haplotypes, thus yielding very complex networks (i.e., full median networks) that were often impossible to interpret. Because we aimed to remove cycles in the network, we kept  $\epsilon$  at its lowest setting. The Optional Post-processing/Maximum Parsimony (MP) calculation implemented in NETWORK (Polzin & Daneshmand, 2003), which would have taken all produced median vectors and linkages into account, was not conducted (see Kong et al. (2015) for arguments against the so-called MP calculation in MJN analysis).

Hereafter, we refer to a set of haplotypes that constitutes a cycle in MJN as ‘causative haplotypes.’ Because we were

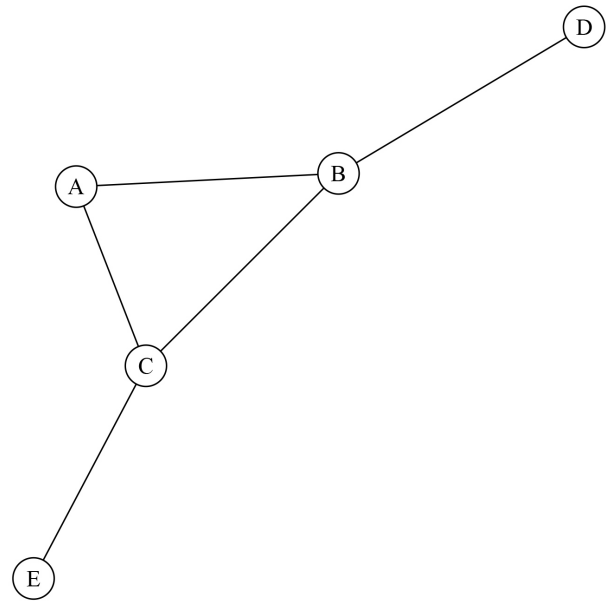


Figure 1. A Median Joining Network (MJN) with five haplotypes (labeled A–E) and a cycle. Eliminating any of the causative haplotypes A, B, or C will result in a network free of cycles. Haplotypes D and E are not causative haplotypes as their deletion will not remove the cycle.

interested in comparing the treelike branching patterns, we manually eliminated any cycle created by the MJ method by arbitrarily selecting and removing one or more causative haplotype(s). For example, in the MJN with a cycle shown in Figure 1 the elimination of any one of the causative haplotypes A, B, or C resulted in a network free of cycles (assuming no median vector replaced the deleted haplotype). In contrast, haplotypes D and E were not causative haplotypes because they are not part of the cycle, and their removal would not eliminate the cycle. More specifically, we removed the sequence(s) that belonged to the selected causative haplotype(s) from Alignment 1. With the truncated Alignment 1 that contained a subset of sequences, we re-executed the MJ algorithm. This procedure was iteratively done until the final network is free of any cycle. We made the same deletion of sequences that belong to the selected causative haplotype(s) for Alignment 2, and this was termed Alignment 2a.

Bayesian inference phylogenies were reconstructed for Alignment 2a and rooted with an outgroup in the alignment. First, we employed MrModelTest version 2.3 (Posada & Crandall, 1998) to select the best DNA substitution model by the Akaike information criterion (AIC). Next, we conducted a single MCMC analysis of  $10^7$  iterations in MrBayes version 3.2.2 (Huelsenbeck & Ronquist, 2001), each of which had four chains (three hot and one cold) with the default priors. One tree was saved every  $10^3$  generations. The analysis was stopped if the standard deviation of split frequencies fell below  $10^{-2}$  after the  $10^7$  generations. If not,  $10^4$  generations were added until the value fell below  $10^{-2}$ . The final 50% majority rule consensus tree was saved. We used the 50% majority rule consensus tree as a final product

of our Bayesian analysis instead of other alternatives (e.g., the maximum clade credibility tree) because the consensus tree is very frequently used in practice to summarize a set of trees collected from the MCMC tree search, and it can be viewed as the optimal tree to report, shown to be more reliable than the maximum clade credibility tree with a lower proportion of incorrect nodes (Holder et al., 2008; O'Reilly & Donoghue, 2017).

Because the BI phylogenies incorporated information about the direction of evolution among the ingroup taxa through outgroup comparison, we used this directionality to manually convert each cycle-free unrooted MJN into a rooted branching topology following Salzburger et al. (2011), with minor modifications. We saved the BI tree and the rooted and transformed MJN topology in Newick format. When the root position that provides the same directionality as in BI phylogeny was not present in MJN, we selected the largest haplotype cluster or the longest edge in MJN as a starting point of the evolutionary direction. When all haplotype clusters were in equal size, we arbitrarily selected a place that produces two clusters that bifurcate from the root where the composition in each cluster is the closest to the two clades formed by the root in BI phylogeny. Note, we distinguished the terms “cluster” and “clades” to refer to a grouping of taxa in MJN and BI tree, respectively, as the latter represents a monophyletic group (i.e., an ancestor and all its descendants) whose information is unavailable in MJN (Kong et al., 2015). In one case where multiple root positions were plausible (shown in case study 1), we produced multiple MJN topologies using different root positions. Finally, the outgroup taxa in the BI phylogenies were removed before statistical comparison to ensure all topologies being tested contained identical sets of leaves.

### 2.3 Analysis

The topologies obtained from MJN and BI analyses were statistically compared using an AU test (Shimodaira, 2002) implemented in the software CONSEL (Shimodaira & Hasegawa, 2001). We also conducted the SH test implemented in PAUP\* version 4.0b10 (Swofford, 2003). Site-wise log-likelihood matrix for each tree using Alignment 1 was obtained using PAUP\* followed by execution of SH test with  $10^3$  bootstrap replicates using the Resampling of Estimated Log Likelihoods (RELL) method (Kishino et al., 1990). Because CONSEL did not read the matrix from PAUP\*, the output matrix was transformed into CONSEL-readable format using a custom python script written by Mark Holder. The transformed matrix file was treated using `makermt` function in CONSEL to read the matrix and generate  $10^5$  bootstrap replicates using the RELL method. The resulting file with extension `.rmt` was processed with CONSEL, which read the bootstrap-replicates of log-likelihood from `makermt` and calculated  $p$ -values for the AU test. We used `catpv` function to visualize the final  $p$ -values with the significance level  $\alpha$  set as 0.05.

The AU test, which was originally intended to assess the confidence for a tree among a set of optimal trees, computes the  $p$ -value to the difference in likelihood among the input topologies based on the same dataset that represents

the possibility that the tree is the true tree (Shimodaira, 2002). This property of  $p$ -value may be informative if the tree was inferred under a single methodology or in different methods that are equally reliable. We have prior knowledge that MJ is a philosophically unsuitable method for phylogenetic inference due to its phenetic-based algorithm. Therefore, we assumed that the  $p$ -value only reflected significant differences in topologies and not the probability of being correct.

## 3 Results

[Figure 2](#) summarizes the number of sequences, the scale (i.e., intra- or interspecific), and type (i.e., mitochondrial or nuclear DNA) of data of the 85 empirical datasets extracted from 55 studies analyzed in this study. Forty-nine (57.6%) out of 85 datasets centered on one species, and 36 (42.4%) contained two or more species. Fifty out of 85 datasets (58.8%) were mitochondrial loci, and 35 (41.2%) involved nuclear genes. Additional information on the sequence length, nucleotide model selected,  $p$ -values from AU and SH tests, and the source where the sequences were extracted from for each case is summarized in the Supplementary Material 1. Focusing on a diversity of organisms, a total of 4,149 ingroup DNA sequences were analyzed, and each dataset contained from seven to 212 sequences. A total of 265 outgroup sequences were included, ranging from one to six per dataset. Seventy-five out of 85 datasets contained less than 100 sequences, and, among these, 57 datasets contained less than 50 sequences. The shortest length of sequence alignment was 235 base pairs (bp), and the longest was 6490 bp. Further, 68 out of 85 datasets had the sequence length of less than 1000 bp after alignment, and only one dataset had sequence length of more than 2000 bp. The computed MJNs for each dataset are available in the Supplementary Material 2. Mesquite executable NEXUS files, each of which contains sequence alignment, 50% majority rule consensus BI phylogeny, and the transformed MJN topology for each case, are provided in the Supplementary Material 3.

Topologies differed significantly (i.e., AU  $p \leq 0.05$ ) in 32 out of 85 cases (37.6%). Significant topological discrepancies were not dependent on whether the dataset was intra- or interspecific, mitochondrial or nuclear DNA, or the sequence length of the dataset. While the difference of likelihood between the two topologies was generally small (i.e.,  $< 100$ ) in most of the cases, four cases showed relatively large difference (191.7–802.2), and significance was found for all four. Among the 53 cases where the significance was not observed, MJN and BI topologies showed some discrepancies in branching patterns in 51 cases (96.3%), and the topologies were identical in two cases only. The discrepancies included positions of clades (or clusters in MJN), constituents of a clade (or cluster), or the branching pattern of the topology (see case studies below).

In several cases where significant discrepancies were observed, the MJN topology had greater likelihood than the BI phylogeny. This was not surprising because the 50% majority rule consensus topology represents a summarization of

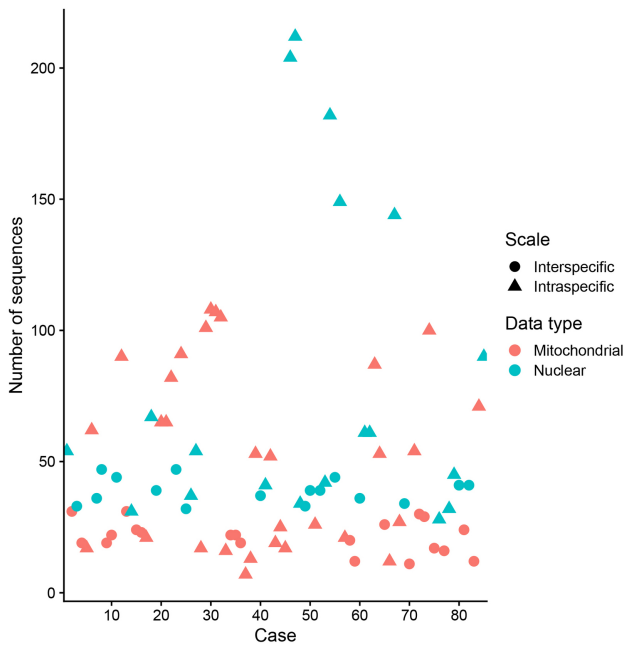


Figure 2. Summary of the 85 analyses (X-axis represents case number) conducted in this study. Y-axis represents the number of sequences. Each point represents a dataset, where circles and triangles represent whether the dataset contains inter- or intraspecific terminals, respectively. Red and green color represents whether the loci in the datasets are mitochondrial or nuclear DNA.

a set of sampled trees from the posterior distribution and often contains unresolved nodes, which consequently leads to reduced likelihood. Nevertheless, we did not make any interpretation based on the likelihood score because the MJ method does not perform philosophically suitable phylogenetic inference, as mentioned in the Methods section. Instead, we only focused on if two topologies are incongruent. In the following section, we show two cases where BI topologies had greater likelihood than MJNs'.

While SH and AU tests generally exhibited similar results, an overly conservative pattern in the SH test was observed as stressed by Strimmer & Rambaut (2002) and Shi et al. (2005). When two competing topologies were not significantly different, both tests agreed in all cases. The two tests agreed in 16 out of 32 cases when the significance was observed between the topologies. However, the two tests did not agree in 16 out of 32 cases when discrepancies between the two topologies were observed, and, in all of these cases, significance found using the AU test was veiled in the SH test.

### 3.1 Case study 1: Significantly different topologies

Yu et al. (2012) analyzed mitochondrial DNA sequences from *Ochotona curzoniae* (the plateau pika) of the Qinghai-Tibetan Plateau, China. We selected 27 D-Loop sequences (GenBank accession numbers JN165313–17; JN165320; JN165322; JN165324–29; JN165332, JN1653323; JN165335–45; JN165350) that were included in the MJN

Table 1. The Approximately Unbiased (AU) and Shimodaira-Hasegawa (SH) test results for four topologies from BI and MJNs using three different roots of *Ochotona curzoniae* D-Loop sequences from Yu et al. (2012). For each tests for the topology listed under the column labeled as Item, Obs. represents the observed loglikelihood differences, and AU and SH represent their computed  $p$ -values.

Topology	Item	Obs.	AU	SH
1	BI	-13.2	0.960	0.954
4	MJN, root C	13.2	0.046	0.055
3	MJN, root B	13.2	0.045	0.055
2	MJN, root A	15.1	0.031	0.042

presented in Figure 3B of Yu et al. (2012). Note that some haplotypes used to compute the original MJN were excluded in our analysis to eliminate unwanted cycles. We selected three outgroup sequences of *O. collaris* (AF348080), *O. princeps* (AJ537415), and *Lepus sinensis* (KM362831), where the former two were suggested in the original study, and the latter was selected using BLAST.

The final alignment was 672 bp long. We set  $\epsilon = 0$  and weight = 10 for computing the MJN. No pre- or post-processing option was taken. Each sequence represented an active, unique haplotype (i.e., 27 terminal vertices exist in the MJN), and a total of 142 mutations were counted for the shortest network. For BI analysis, the GTR+I+ $\Gamma$  substitution model was selected for the sequence evolution, and a single MCMC analysis of  $10^7$  iterations was performed.

Figure 3 shows the BI 50% majority-rule consensus phylogeny (Fig. 3a), MJN (Fig. 3e), and three transformed MJNs using the root positions A, B, and C (Fig. 3b–d, respectively) that are indicated in the MJN. The BI phylogeny had four primary clades, where the members of Clades 1, 2, 3, and 4 are colored in blue, green, purple, and red, respectively, in Figure 3a. The posterior probability for each node that represented the most recent common ancestor (MRCA) for each clade was high. These values ranged from 0.86 to 0.99, except for Clade 4, which had 0.62. Four clusters in the MJN corresponded to the BI phylogeny (the sequences that belong to the Clades 1–4 in the BI phylogeny are colored using the same scheme described above in Figure 3e), but no root position that would provide the same evolutionary direction as in the BI phylogeny was available. The three most probable root positions were A, B, and C as shown in yellow stars in Figure 3e. Roots A and B were placed on the edges with the longest lengths. Although the edge between Clusters 1 (blue) and 2 (green) was long, it was not selected as a possible root placement because the two clusters were sister taxa in the BI tree with MRCA that is not the root. Because a root may not necessarily position on long edges, we also identified root C as it grouped Clusters 1 and 2 despite the short length that separated some members of cluster 4. None of the topologies with roots A, B, or C resolved mono-

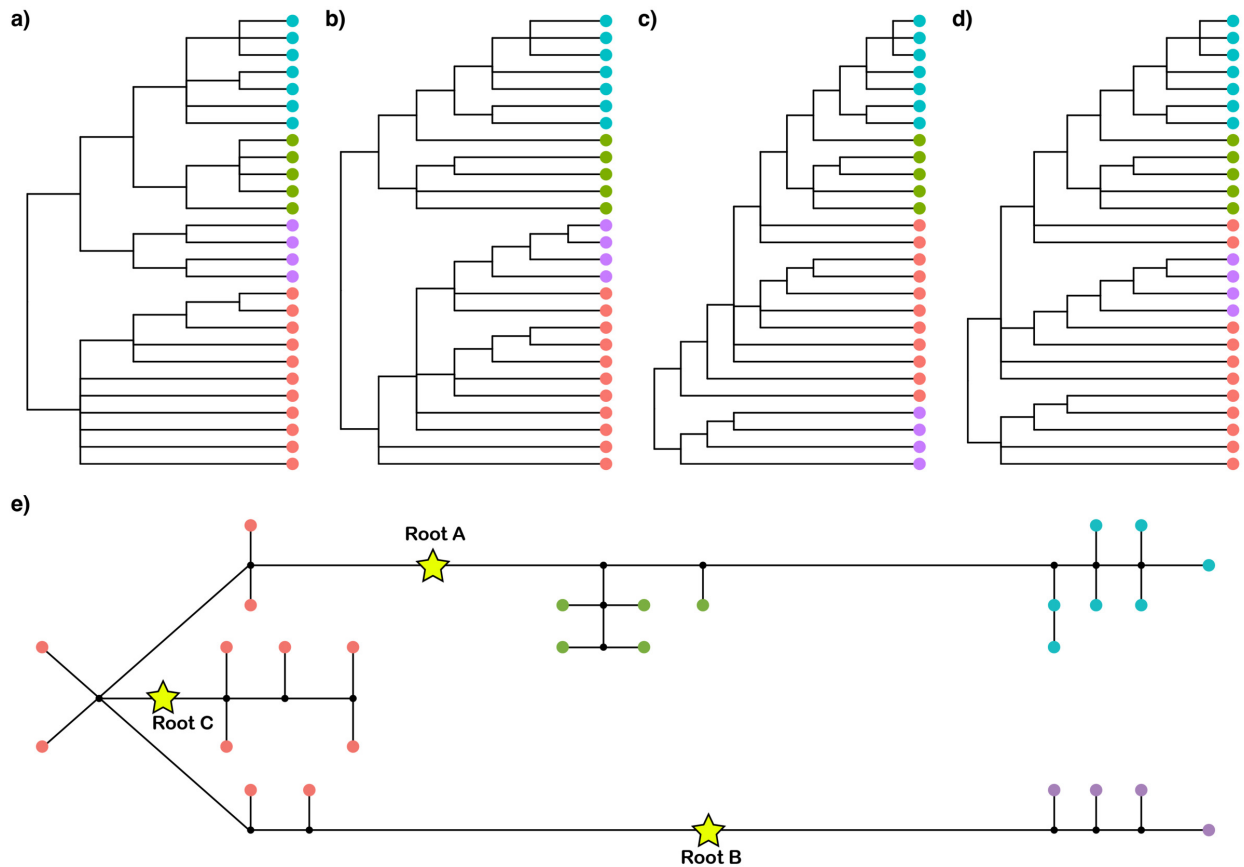


Figure 3. (a) Reconstructed 50% majority-rule consensus Bayesian inference (BI) phylogeny using 27 *Ochotona curzoniae* D-Loop sequences from Yu et al. (2012), rooted with outgroup sequences (posterior probabilities and outgroup sequence are not shown here). Four clades were identified and color coded with blue, green, purple, and red for Clades 1, 2, 3, and 4, respectively. (b) MJN topology using the Root A, (c) Root B, and (d) Root C that are selected and labeled in the computed MJN in (e) with yellow stars. We used the same color scheme mentioned above to mark the members of each clade for all topologies and MJN. Small black dots in (e) represent computed median vectors.

Table 2. AU and SH test results for topologies from BI and MJN of *Coreoperca whiteheadi* cytochrome b sequences from Cao et al. (2013). Obs. represents the observed log-likelihood differences. AU and SH represent  $p$ -values for each tests for the topology listed under the column labeled as Item.

Topology	Item	Obs.	AU	SH
1	BI	-4.3	0.857	0.825
2	MJN	4.3	0.143	0.175

phyly of Cluster 4 (Figs. 3b–d). PAUP\* was used to calculate site-wise log-likelihood scores for all four rooted topologies using the GTR+I+ $\Gamma$  nucleotide substitution model.

The AU test (Table 1) showed that the BI topology differed significantly in branching pattern from all three MJN topologies ( $p \leq 0.05$ ). In comparison, the SH test showed that the BI topology differed significantly from the MJN using root A only, although root positions B and C had small  $p$ -values.

### 3.2 Case study 2: Insignificantly different topologies

We used the mitochondrial cytochrome b sequence dataset of *Coreoperca whiteheadi* (East Asian perch) from Cao et al. (2013). We selected 53 sequences of *C. whiteheadi* (JN315557–JN315609) along with outgroup sequences from *Lateolabrax japonicus* (DQ3515-31) and *Siniperca chuatsi* (KC888071), as suggested by the authors, plus *Trachinotus rhodopus* (AY050739) selected via BLAST. The length of final alignment was 1141 bp.

Analyses replicated the MJN in Figure 3a of Cao et al. (2013). The 11 haplotypes contained from one to 19 sequences, and the final network identified 203 mutations among the haplotypes. We used the GTR+I+ $\Gamma$  nucleotide substitution model, as selected by MrModelTest using AIC, for the construction of BI phylogeny. The identified position of the root on the MJN (see Fig. 4c) provided the same directionality as in the BI phylogeny (Fig. 4a). The root occurred on the longest edge of the MJN (Fig. 4c), which represented 106 nucleotide differences. We converted the MJN into a tree using the selected root position (Fig. 4b).

The BI phylogeny had four primary clades (the members in Clades 1, 2, 3, and 4 are colored in red, green, blue,

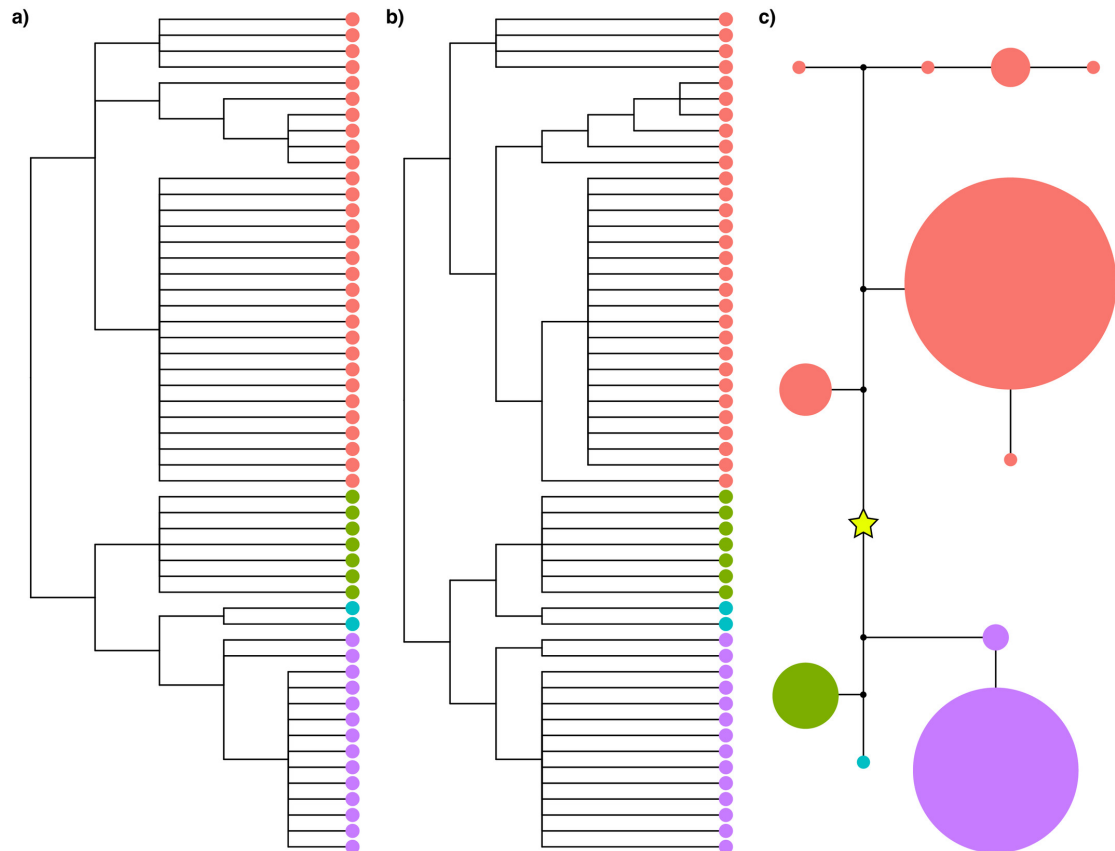


Figure 4. (a) Reconstructed 50% majority-rule consensus BI phylogenetic tree using 53 *Coreoperca whiteheadi* cytochrome b sequences from Cao et al. (2013), rooted with outgroup sequences (posterior probabilities and outgroup sequences are not shown). (b) MJN topology using the identified root position in MJN shown in (c) (i.e., yellow star). Four clades were identified in (a), and we color coded clade 1, 2, 3, and 4 using red, green, blue, and purple, respectively. Same color scheme was applied for (b) and (c). Small black dots in (c) represent computed median vectors.

and purple, respectively, in [Figure 4a](#)). The relationships between Clades 2, 3, and 4 were inconsistent between the BI phylogeny and MJN topology ([Figs. 4a and 4b](#), respectively). In the BI phylogeny, Clades 3 and 4 shared the MRCA with posterior probability of 1.0. However, the MJN grouped Clusters 2 with 3. Despite the differences in branching pattern, the relationships depicted in BI phylogeny and MJN were not shown to be significantly different based on both the AU and SH test ([Table 2](#)). Although no significant difference occurred in the AU and SH tests, the biological interpretations based on the MJN would be problematic due to differences in the resolved branching order.

## 4 Discussion

### 4.1 Problems of Median-Joining Network analysis

Our analyses find statistically significant discrepancies between topologies from BI and MJ in 37.6% of the cases evaluated based on AU test output. If BI produces a more reliable evolutionary history than MJ, then MJ method too often fails to produce a correct suite of relationships for confident evolutionary inference. Even in the 62.4% of cases where statistical significance does not occur, differ-

ences in branching patterns often represent different evolutionary histories among the sequences (see case study 2). Therefore, MJ largely produces misleading hypotheses of relationships and character evolution. Phenetic-based MJ has many unrealistic assumptions, such as ignoring homoplastic evolution. This has led many authors to criticize phenetic methods for failing to produce ‘true’ phylogenies (Cheema & Dicks, 2009; Lawrence et al., 2002). While MJ and BI produce marginally different sets of relationships in terms of the AU test output in more than half of the cases, sometimes the differences are critical, for example, when MJ fails to correctly identify monophyletic groups.

One of the biggest problems with MJNs, and probably all un-rooted networks, is the absence of evolutionary direction (Kong et al., 2015). Identification of monophyletic groups is only possible given direction, i.e., ancestor–descendent relationships. An outgroup serves to infer hypothetical ancestral states and, in doing so, roots an un-rooted branching structure (Lyons-Weiler et al., 1998; Maddison et al., 1984; Sanderson & Shaffer, 2002; Smith, 1994; Watrous & Wheeler, 1981). Although the software NETWORK allows the user to root the network, this option merely links the outgroup sequence to the most similar haplotype of the already produced ingroup network (Kong et al., 2015; Sánchez-Pacheco et al., 2020). For example,

Sakaguchi et al. (2012) employed the rooting option to root the MJN for 26 chloroplast haplotypes, yet the branching pattern differed from the inferred MP phylogeny using the same outgroup. Additionally, the option only allows one sequence to be entered, yet more than one outgroup taxon may be necessary to obtain reproducible results (Maddison et al., 1984; Smith, 1994).

Given its phenetic nature, overall similarity influences the MJ algorithm very strongly, especially in computations involving gaps and ambiguous states. When the amount of gaps is large due to excessive amounts of indels, MJ can create complicated cyclic structures that are almost impossible to interpret. Consequently, it has been recommended to minimize gap sites prior to analysis (Bandelt et al., 1999), but this can be a very subjective process (DeSalle et al., 1994). Moreover, gaps can be phylogenetically informative, improve branch support, and even change a topology (Nagy et al., 2012; Simmons & Norton, 2014). Trimming by the simple removal of the gaps can result in the loss of otherwise informative sites, and, thus, it does not necessarily lead to better trees (Dessimoz & Gil, 2010; Löytynoja & Goldman, 2008; Wong et al., 2008; Wu et al., 2012).

The treatment of ambiguous states in the construction of MJNs is also problematic. Bandelt et al. (1999) suggested using ambiguous states in the alignment infrequently. However, MJ tolerates ambiguity codes such as  $X = R$  (purine) = {A, G} or  $X = N$  = {A, T, C, G} and others, where  $X$  represents the set of nucleotide states that specify the ambiguous state. Here, the MJ algorithm replaces the ambiguous state by assigning the most common state of the other minimally distant sequences in prior to the computation. This procedure is arbitrary and can result in the loss of potentially informative variations in the data. To illustrate the problem, we use three hypothetical sequences with five bp each—SQ1=CAACG, SQ2=GCCAC, and SQ3=AGCGA (Fig. 5a)—used in the construction of Figure 3 of Bandelt et al. (1999). Each sequence differs at three sites from each other: sites 1 and 2, plus 3 for SQ1, 4 for SQ2, and 5 for SQ3. Figure 5h shows the MJN constructed using the alignment in Fig. 5a.

Because MJ simply identifies relationships based on overall similarity, the resulting network is the same when the character states that are unique for each haplotype (i.e., site 3 for SQ1, 4 for SQ2, and 5 for SQ3) differ from another character state, as long as it still distinguishes them. For example, when the sites 3, 4, and 5 in SQ1, SQ2, and SQ3, respectively, have nucleotide state A and then changed to T (Fig. 5b) or a gap (Fig. 5c), the network structure remains as in Figure 5h. Similarly, when the sites 3, 4, and 5 in SQ1, SQ2, and SQ3, respectively, are replaced by ambiguous sites (Fig. 5d) or a mixture of nucleotide, gap, and an ambiguous site (Fig. 5e), the network structure remains unaltered.

Figures 5d and 5e use ambiguous nucleotides that are not the same as in other sequences. For example, ambiguous site R exists in site 3 of SQ1; it can be either A or G, but not C (i.e., the state SQ2 and SQ3 possess at site 3). Similar ambiguous replacements for the other cases illustrate the problem. When we replace SQ1 with N (= A,T,C,G) or M (= A,C), the MJ algorithm arbitrarily switches the site to C

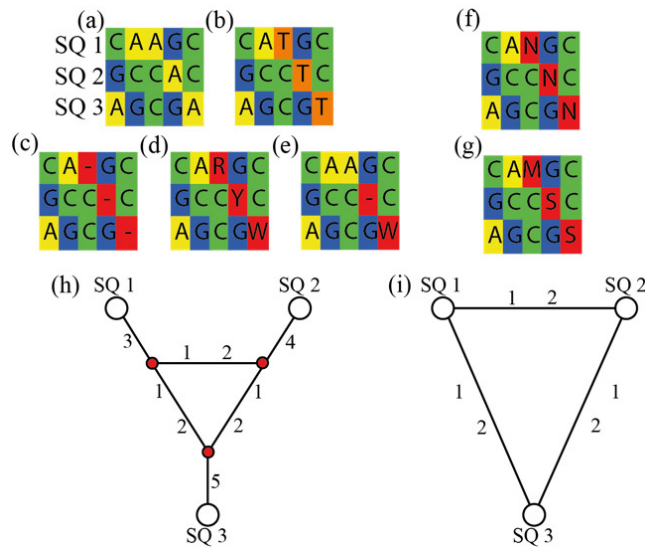


Figure 5. Example of the treatment of ambiguous states in the construction of MJNs. Sequence alignments (a–e) result in the MJN shown in (h), whereas the alignments (f) and (g) result in the MJN (i). Note the two MJNs have different biological interpretations as the algorithm arbitrarily removes variabilities in the presence of the ambiguous sites.

by comparing it with the minimally distant sequence; this preference eliminates the median vector on the edges (Fig. 5f, 5g, and 5i) based solely on algorithm-assigned overall similarity. Networks in Figures 5h and 5i do not have identical biological interpretations, and more complex problems can arise when evaluating longer sequences and more taxa.

## 4.2 Excessive conservatism of the SH test

Researchers often compare the significance of one or a set of optimal trees from different phylogenetic methods (e.g., Kazlauskas et al., 2019; Naser-Khdour et al., 2019; Spence et al., 2021) or how strongly competing topologies can be rejected relative to the preferred one (e.g., Coleman et al., 2021; Espeland et al., 2018; Hime et al., 2020; Miller et al., 2020; Yan et al., 2021) using SH and AU tests. They employ similar rationale, yet selection-bias in the SH test can lead to overconfidence in incorrect trees when more than two trees are compared (Goldman et al., 2000; Shimodaira & Hasegawa, 1999). Moreover, Strimmer & Rambaut (2002) pointed out that the SH test can also be very conservative, and the number of trees in confidence set can increase with the number of trees being compared. Thus, Shimodaira (2002) proposed the AU test, which is less conservative than the SH test.

By comparing the outputs of AU and SH tests, we reconfirm the over-conservativeness of the SH test. For cases in which significant discrepancies exist between the two topologies in at least one of the tests, the two tests arrive at different confidence values in 16 out of 32 cases. Here, the SH test fails to detect significant discrepancies found by the AU test. In other words, the AU test confidently selects one topology out of the two topologies tested, but the SH test



selects both trees as being equivalent (or the difference is insignificant) due to its conservative nature. This observation is congruent with the reanalysis of mammalian mitochondrial protein-coding sequences in Shimodaira (2002). In the reanalysis, the SH test resulted in eight trees where their  $p$ -values were not significant at  $\alpha = 0.05$  out of 15 competing hypotheses, whereas AU resulted in only six.

## 5 Conclusions

In conclusion, MJ analysis is appealing because it is computationally simple and fast and yields attractive graphical representation of the data. Nevertheless, MJ should not be used in evolutionary studies given its distance-based phenetics, the absence of direction (Kong et al., 2015; Sánchez-Pacheco et al., 2020), the inherent assumption that all lineages evolve at equal rates (i.e., overall similarity reflects phylogeny), and the extent of statistically significant mismatches between the MJN and BI topologies. The frequency that MJ shows problematic branching patterns, and thus incorrect inferences of evolutionary relationships among taxa, can be higher than 95% (as only two out of 85 cases show identical relationships between MJN and BI topologies). Indeed, the branching patterns produced by MJN analyses differ statistically significantly in major ways from BI trees in 37.6% of our case studies. This is not a trivial problem in an era when researchers are using phylogenetic trees to help solve a wide range of problems, from identifying the source of human pathogens to making conserva-

tion decisions. Such enterprises require the use of the most robust phylogenies possible, which, in turn, requires the use of defensible and rigorous methods for reconstructing trees. Our study reinforces the decades-old recognition that phenetic-based algorithms are not a defensible way to hypothesize evolutionary relationships.

## Funding

NSERC Discovery Grant 3148 supported the research. Funding for S.J.S-P was provided by an Ontario Graduate Scholarship at the University of Toronto.

## Acknowledgements

We thank two anonymous reviewers for the constructive comments. We also thank M. Holder for advice conducting the AU test.

## Supporting Information

Supplementary Materials available from the Dryad Digital Repository: <https://doi.org/10.5061/dryad.f4tqrfj708>

Submitted: July 17, 2023 EDT, Accepted: August 01, 2023 EDT

## References

- Bandelt, H. J., Forster, P., & Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, *16*(1), 37–48. <https://doi.org/10.1093/oxfordjournals.molbev.a026036>
- Bandelt, H. J., Forster, P., Sykes, B. C., & Richards, M. B. (1995). Mitochondrial portraits of human populations using median networks. *Genetics*, *141*(2), 743–753. <https://doi.org/10.1093/genetics/141.2.743>
- Bapteste, E., van Iersel, L., Janke, A., Kelchner, S., Kelk, S., McInerney, J. O., Morrison, D. A., Nakhleh, L., Steel, M., Stougie, L., & Whitfield, J. (2013). Networks: Expanding evolutionary thinking. *Trends in Genetics*, *29*(8), 439–441. <https://doi.org/10.1016/j.tig.2013.05.007>
- Blair, C., & Ané, C. (2019). Phylogenetic trees and networks can serve as powerful and complementary approaches for analysis of genomic data. *Systematic Biology*, *69*(3), 593–601. <https://doi.org/10.1093/sysbio/syz056>
- Cao, L., Liang, X.-F., Tang, W., & Zhao, J. (2013). Phylogeography of *Coreoperca whiteheadi* (Perciformes: Coreoperca) in China based on mitochondrial and nuclear gene sequences. *Biochemical Systematics and Ecology*, *50*, 223–231. <https://doi.org/10.1016/j.bse.2013.04.007>
- Cheema, J., & Dicks, J. (2009). Computational approaches and software tools for genetic linkage map estimation in plants. *Briefings in Bioinformatics*, *10*(6), 595–608. <https://doi.org/10.1093/bib/bbp045>
- Coleman, G. A., Davín, A. A., Mahendrarajah, T. A., Szánthó, L. L., Spang, A., Hugenholtz, P., Szöllösi, G. J., & Williams, T. A. (2021). A rooted phylogeny resolves early bacterial evolution. *Science*, *372*(6542), eabe0511. <https://doi.org/10.1126/science.abe0511>
- Debevec, A. H., & Whitfield, J. B. (2012). Introduction to phylogenetic networks.—David A. Morrison. *Systematic Biology*, *62*(1), 177–178. <https://doi.org/10.1093/sysbio/sys056>
- Degnan, J. H. (2018). Modeling hybridization under the network multispecies coalescent. *Systematic Biology*, *67*(5), 786–799. <https://doi.org/10.1093/sysbio/syy040>
- DeSalle, R., Wray, C., & Absher, R. (1994). Computational problems in molecular systematics. *Experientia Supplementum*, *69*, 353–370. [https://doi.org/10.1007/978-3-0348-7527-1\\_21](https://doi.org/10.1007/978-3-0348-7527-1_21)
- Dessimoz, C., & Gil, M. (2010). Phylogenetic assessment of alignments reveals neglected tree signal in gaps. *Genome Biology*, *11*(4), R37. <https://doi.org/10.1186/gb-2010-11-4-r37>
- Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. <https://doi.org/10.1093/nar/gkh340>
- Espeland, M., Breinholt, J., Willmott, K. R., Warren, A. D., Vila, R., Toussaint, E. F. A., Maunsell, S. C., Aduse-Poku, K., Talavera, G., Eastwood, R., Jarzyna, M. A., Guralnick, R., Lohman, D. J., Pierce, N. E., & Kawahara, A. Y. (2018). A comprehensive and dated phylogenomic analysis of butterflies. *Current Biology*, *28*(5), 770–778.e5. <https://doi.org/10.1016/j.cub.2018.01.061>
- Felsenstein, J. (2004). *Inferring phylogenies*. Sinauer Associates.
- Goldman, N., Anderson, J. P., & Rodrigo, A. G. (2000). Likelihood-based tests of topologies in phylogenetics. *Systematic Biology*, *49*(4), 652–670. <https://doi.org/10.1080/106351500750049752>
- Gusfield, D., & Bansal, V. (2005). A fundamental decomposition theory for phylogenetic networks and incompatible characters. *Lecture Notes in Computer Science*, *3500*, 217–232. [https://doi.org/10.1007/11415770\\_17](https://doi.org/10.1007/11415770_17)
- Hejase, H. A., & Liu, K. J. (2016). A scalability study of phylogenetic network inference methods using empirical datasets and simulations involving a single reticulation. *BMC Bioinformatics*, *17*(1), 422. <https://doi.org/10.1186/s12859-016-1277-1>
- Hime, P. M., Lemmon, A. R., Lemmon, E. C. M., Prendini, E., Brown, J. M., Thomson, R. C., Kratovil, J. D., Noonan, B. P., Pyron, R. A., Peloso, P. L. V., Kortyna, M. L., Keogh, J. S., Donnellan, S. C., Mueller, R. L., Raxworthy, C. J., Kunte, K., Ron, S. R., Das, S., Gaitonde, N., ... Weisrock, D. W. (2020). Phylogenomics reveals ancient gene tree discordance in the Amphibian Tree of Life. *Systematic Biology*, *70*(1), 49–66. <https://doi.org/10.1093/sysbio/syaa034>
- Holder, M. T., Sukumaran, J., & Lewis, P. O. (2008). Justification for reporting the majority-rule consensus tree in Bayesian phylogenetics. *Systematic Biology*, *57*(5), 814–821. <https://doi.org/10.1080/10635150802422308>

- Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17(8), 754–755. <https://doi.org/10.1093/bioinformatics/17.8.754>
- Huson, D. H., & Bryant, D. (2005). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, 23(2), 254–267. <https://doi.org/10.1093/molbev/msj030>
- Huson, D. H., Rupp, R., & Scornavacca, C. (2010). *Phylogenetic networks: Concepts, algorithms and applications*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511974076>
- Huson, D. H., & Scornavacca, C. (2010). A survey of combinatorial methods for phylogenetic networks. *Genome Biology and Evolution*, 3, 23–35. <https://doi.org/10.1093/gbe/evq077>
- Kazlauskas, D., Varsani, A., Koonin, E. V., & Krupovic, M. (2019). Multiple origins of prokaryotic and eukaryotic single-stranded DNA viruses from bacterial and archaeal plasmids. *Nature Communications*, 10(1), 3425. <https://doi.org/10.1038/s41467-019-11433-0>
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), 1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>
- Kishino, H., Miyata, T., & Hasegawa, M. (1990). Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *Journal of Molecular Evolution*, 31(2), 151–160. <https://doi.org/10.1007/bf02109483>
- Kong, S. (2022). Digest: Frequent hybridization in *Darevskia* rarely leads to the evolution of asexuality. *Evolution*, 76(9), 2216–2217. <https://doi.org/10.1111/evo.14587>
- Kong, S., Sánchez-Pacheco, S. J., & Murphy, R. W. (2015). On the use of median-joining networks in evolutionary biology. *Cladistics*, 32(6), 691–699. <https://doi.org/10.1111/cla.12147>
- Kong, S., Swofford, D. L., & Kubatko, L. S. (2022). Inference of phylogenetic networks from sequence data using composite likelihood. *bioRxiv*. <https://doi.org/10.1101/2022.11.14.516468>
- Lawrence, J. G., Hatfull, G. F., & Hendrix, R. W. (2002). Imbroglions of viral taxonomy: Genetic exchange and failings of phenetic approaches. *Journal of Bacteriology*, 184(17), 4891–4905. <https://doi.org/10.1128/jb.184.17.4891-4905.2002>
- Li, S., Pearl, D. K., & Doss, H. (2000). Phylogenetic tree construction using Markov Chain Monte Carlo. *Journal of the American Statistical Association*, 95(450), 493–508. <https://doi.org/10.1080/01621459.2000.10474227>
- Linder, C. R., & Rieseberg, L. H. (2004). Reconstructing patterns of reticulate evolution in plants. *American Journal of Botany*, 91(10), 1700–1708. <https://doi.org/10.3732/ajb.91.10.1700>
- Löytynoja, A., & Goldman, N. (2008). Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science*, 320(5883), 1632–1635. <https://doi.org/10.1126/science.1158395>
- Luo, A., Zhang, Y.-Z., Qiao, H., Murphy, R., Shi, W., & Zhu, C.-D. (2010). Outgroup selection in tree construction: A case study of the family Halictidae (Hymenoptera: Apoidea). *Acta Entomologica Sinica*, 53, 192–201.
- Lyons-Weiler, J., Hoelzer, G. A., & Tausch, R. J. (1998). Optimal outgroup analysis. *Biological Journal of the Linnean Society*, 64(4), 493–511. <https://doi.org/10.1111/j.1095-8312.1998.tb00346.x>
- Maddison, W. P., Donoghue, M. J., & Maddison, D. R. (1984). Outgroup analysis and parsimony. *Systematic Zoology*, 33(1), 83. <https://doi.org/10.2307/2413134>
- Mau, B., & Newton, M. A. (1997). Phylogenetic inference for binary data on dendrograms using Markov Chain Monte Carlo. *Journal of Computational and Graphical Statistics*, 6(1), 122–131. <https://doi.org/10.1080/10618600.1997.10474731>
- Mau, B., Newton, M. A., & Larget, B. (1999). Bayesian phylogenetic inference via Markov Chain Monte Carlo Methods. *Biometrics*, 55(1), 1–12. <https://doi.org/10.1111/j.0006-341x.1999.00001.x>
- Miller, J. M., Cullingham, C. I., & Peery, R. M. (2020). The influence of a priori grouping on inference of genetic clusters: Simulation study and literature review of the DAPC method. *Heredity*, 125(5), 269–280. <https://doi.org/10.1038/s41437-020-0348-2>

- Nagy, L. G., Kocsubé, S., Csanádi, Z., Kovács, G. M., Petkovits, T., Vágvölgyi, C., & Papp, T. (2012). Re-Mind the gap! Insertion – deletion data reveal neglected phylogenetic potential of the nuclear ribosomal internal transcribed spacer (ITS) of fungi. *PLoS ONE*, 7(11), e49794. <https://doi.org/10.1371/journal.pone.0049794>
- Naser-Khdour, S., Minh, B. Q., Zhang, W., Stone, E. A., & Lanfear, R. (2019). The prevalence and impact of model violations in phylogenetic analysis. *Genome Biology and Evolution*, 11(12), 3341–3352. <https://doi.org/10.1093/gbe/evz193>
- O'Reilly, J. E., & Donoghue, P. C. J. (2017). The efficacy of consensus tree methods for summarizing phylogenetic relationships from a posterior sample of trees estimated from morphological data. *Systematic Biology*, 67(2), 354–362. <https://doi.org/10.1093/sysbio/syx086>
- Polzin, T., & Daneshmand, S. V. (2003). On Steiner trees and minimum spanning trees in hypergraphs. *Operations Research Letters*, 31(1), 12–20. [https://doi.org/10.1016/s0167-6377\(02\)00185-2](https://doi.org/10.1016/s0167-6377(02)00185-2)
- Posada, D., & Crandall, K. A. (1998). MODELTEST: Testing the model of DNA substitution. *Bioinformatics*, 14(9), 817–818. <https://doi.org/10.1093/bioinformatics/14.9.817>
- Posada, D., & Crandall, K. A. (2001). Intraspecific gene genealogies: Trees grafting into networks. *Trends in Ecology & Evolution*, 16(1), 37–45. [https://doi.org/10.1016/s0169-5347\(00\)02026-7](https://doi.org/10.1016/s0169-5347(00)02026-7)
- Sakaguchi, S., Qiu, Y.-X., Liu, Y.-H., Qi, X.-S., Kim, S.-H., Han, J., Takeuchi, Y., Worth, J. R. P., Yamasaki, M., Sakurai, S., & Isagi, Y. (2012). Climate oscillation during the Quaternary associated with landscape heterogeneity promoted allopatric lineage divergence of a temperate tree *Kalopanax septemlobus* (Araliaceae) in East Asia. *Molecular Ecology*, 21(15), 3823–3838. <https://doi.org/10.1111/j.1365-294x.2012.05652.x>
- Salzburger, W., Ewing, G. B., & Von Haeseler, A. (2011). The performance of phylogenetic algorithms in estimating haplotype genealogies with migration: Estimating haplotype genealogies. *Molecular Ecology*, 20(9), 1952–1963. <https://doi.org/10.1111/j.1365-294x.2011.05066.x>
- Sánchez-Pacheco, S. J., Kong, S., Pulido-Santacruz, P., Murphy, R. W., & Kubatko, L. (2020). Median-joining network analysis of SARS-CoV-2 genomes is neither phylogenetic nor evolutionary. *Proceedings of the National Academy of Sciences*, 117(23), 12518–12519. <https://doi.org/10.1073/pnas.2007062117>
- Sanderson, M. J., & Shaffer, H. B. (2002). Troubleshooting molecular phylogenetic analyses. *Annual Review of Ecology and Systematics*, 33(1), 49–72. <https://doi.org/10.1146/annurev.ecolsys.33.010802.150509>
- Shi, X., Gu, H., Susko, E., & Field, C. (2005). The comparison of the confidence regions in phylogeny. *Molecular Biology and Evolution*, 22(11), 2285–2296. <https://doi.org/10.1093/molbev/msi226>
- Shimodaira, H. (2002). An Approximately Unbiased test of phylogenetic tree selection. *Systematic Biology*, 51(3), 492–508. <https://doi.org/10.1080/10635150290069913>
- Shimodaira, H., & Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution*, 16(8), 1114–1116. <https://doi.org/10.1093/oxfordjournals.molbev.a026201>
- Shimodaira, H., & Hasegawa, M. (2001). CONSEL: For assessing the confidence of phylogenetic tree selection. *Bioinformatics*, 17(12), 1246–1247. <https://doi.org/10.1093/bioinformatics/17.12.1246>
- Simmons, M. P., & Norton, A. P. (2014). Divergent maximum-likelihood-branch-support values for polytomies. *Molecular Phylogenetics and Evolution*, 73, 87–96. <https://doi.org/10.1016/j.ympev.2014.01.018>
- Smith, A. B. (1994). Rooting molecular trees: Problems and strategies. *Biological Journal of the Linnean Society*, 51(3), 279–292. <https://doi.org/10.1111/j.1095-8312.1994.tb00962.x>
- Solís-Lemus, C., & Ané, C. (2016). Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLOS Genetics*, 12(3), e1005896. <https://doi.org/10.1371/journal.pgen.1005896>
- Spence, M. A., Mortimer, M. D., Buckle, A. M., Minh, B. Q., & Jackson, C. J. (2021). A comprehensive phylogenetic analysis of the serpin superfamily. *Molecular Biology and Evolution*, 38(7), 2915–2929. <https://doi.org/10.1093/molbev/msab081>
- Strimmer, K., & Rambaut, A. (2002). Inferring confidence sets of possibly misspecified gene trees. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 269(1487), 137–142. <https://doi.org/10.1098/rspb.2001.1862>
- Suzuki, Y., Glazko, G. V., & Nei, M. (2002). Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. *Proceedings of the National Academy of Sciences*, 99(25), 16138–16143. <https://doi.org/10.1073/pnas.212646199>

- Swofford, D. L. (2003). *Phylogenetic Analysis Using Parsimony (\*and Other Methods)*. Version 4. Sinauer Associates.
- Than, C., Ruths, D., & Nakhleh, L. (2008). PhyloNet: A software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics*, 9(1), 322. <https://doi.org/10.1186/1471-2105-9-322>
- Watrout, L. E., & Wheeler, Q. D. (1981). The outgroup comparison method of character analysis. *Systematic Zoology*, 30(1), 1. <https://doi.org/10.2307/2992297>
- Wen, D., & Nakhleh, L. (2017). Coestimating reticulate phylogenies and gene trees from multilocus sequence data. *Systematic Biology*, 67(3), 439–457. <https://doi.org/10.1093/sysbio/syx085>
- Wong, K. M., Suchard, M. A., & Huelsenbeck, J. P. (2008). Alignment uncertainty and genomic analysis. *Science*, 319(5862), 473–476. <https://doi.org/10.1126/science.1151532>
- Wu, M., Chatterji, S., & Eisen, J. A. (2012). Accounting for alignment uncertainty in phylogenomics. *PLoS ONE*, 7(1), e30288. <https://doi.org/10.1371/journal.pone.0030288>
- Yan, L., Pape, T., Meusemann, K., Kutty, S. N., Meier, R., Bayless, K. M., & Zhang, D. (2021). Monophyletic blowflies revealed by phylogenomics. *BMC Biology*, 19(1), 230. <https://doi.org/10.1186/s12915-021-01156-4>
- Yang, Z., & Rannala, B. (1997). Bayesian phylogenetic inference using DNA sequences: A Markov Chain Monte Carlo Method. *Molecular Biology and Evolution*, 14(7), 717–724. <https://doi.org/10.1093/oxfordjournals.molbev.a025811>
- Yu, F., Li, S., Kilpatrick, W. C., McGuire, P. M., He, K., & Wei, W. (2012). Biogeographical study of Plateau Pikas *Ochotona curzoniae* (Lagomorpha, Ochotonidae). *Zoological Science*, 29(8), 518–526. <https://doi.org/10.2108/zsj.29.518>