The Influence of Model Violation on Phylogenetic Inference: A Simulation Study

SUHA NASER-KHDOUR*1, BUI QUANG MINH2, AND ROBERT LANFEAR1

1 Department of Ecology and Evolution, Research School of Biology, Australian National University, Canberra, Australian Capital Territory, Australia
2 School of Computing, Australian National University, Canberra, Australian Capital Territory, Australia

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

ABSTRACT

Phylogenetic inference typically assumes that the data has evolved under Stationary, Reversible and Homogeneous (SRH) conditions. Many empirical and simulation studies have shown that assuming SRH conditions can lead to significant errors in phylogenetic inference when the data violates these assumptions. Yet, many simulation studies focused on extreme non-SRH conditions that represent worst-case scenarios and not the average empirical dataset. In this study, we simulate datasets under various degrees of non-SRH conditions using empirically derived parameters to mimic real data and examine the effects of incorrectly assuming SRH conditions on inferring phylogenies. Our results show that maximum likelihood inference is generally quite robust to a wide range of SRH model violations but is inaccurate under extreme convergent evolution.

[Phylogenetic inference, model violations, systematic bias, simulations, evolution under non-SRH conditions, test of symmetry]
Markov processes are commonly used in model-based phylogenetic analyses such as maximum likelihood (ML) and Bayesian inference (Felsenstein 2004; Yang 2006). A Markov model is represented by an instantaneous rate matrix $Q$ of size $4 \times 4$ for DNA or $20 \times 20$ for protein sequences, that describes the substitution rates between nucleotides or amino-acids (henceforth denoted as states), respectively. The Markovian propriety is convenient because the probabilities of the next states only depend on the current states, independently of how the current states had evolved (Felsenstein 1981; Felsenstein 1983; Yang 1994; Swofford, et al. 1996; Yang 2006). For mathematical simplicity and computational tractability, most studies assume that the Markov model is stationary, reversible, and homogeneous (SRH) (Kimura 1980; Felsenstein 1981; Hasegawa, et al. 1985; Tavaré 1986; Tamura and Nei 1993; Yang 1994). Homogeneity means that a single $Q$ matrix operates along all edges of the tree, i.e., all substitution rates stay constant through time. Stationarity means that the state frequencies also remain constant along all edges of the tree. Reversibility means that the rate of change from state A to another state B is the same as the backward substitution rate from B to A.

The assumptions of homogeneity, stationarity, and reversibility come at the cost of complying with biological reality (Roberts and Yang 1995; Foster and Hickey 1999; Foster 2004; Ababneh, et al. 2006). For example, the reversibility assumption implies that the likelihood of a tree topology will be the same regardless of the placement of the root (Felsenstein 1981). Moreover, a reversible substitution model has up to 8 free rate parameters for nucleotides and 208 for amino acids, while a non-reversible substitution model has up to 11 free rate parameters for nucleotides and 379 for amino acids, provided that the model is still stationary and homogeneous (Yang 1994). These degrees of freedom can increase dramatically if the model is non-stationary or/and non-homogeneous (Barry and Hartigan...
Using stationary, reversible, and homogeneous substitution models to infer a phylogeny from data that has evolved under more complex conditions compromises the consistency of the ML estimation (Felsenstein 2004). Ideally, we would like to use data that comply with the assumptions of the models we apply, or alternatively, use models that are not violated by the data in hand. However, the use of non-SRH models is computationally demanding and is often not practical in large datasets. On the other hand, removing data that do not comply with the SRH assumption will come at a cost of losing phylogenetic information. Both simulation (Huelsenbeck and Hillis 1993; Hillis, et al. 1994; Galtier and Gouy 1998; Ho and Jermiin 2004; Jermiin, et al. 2004; Boussau and Gouy 2006) and empirical (Phillips, et al. 2004; Collins, et al. 2005; Nguyen, et al. 2012; Betancur, et al. 2013; Naser-Khdour, et al. 2019) studies have shown that applying SRH models to data that have evolved under more complex conditions can lead to significant errors in phylogenetic inference. However, most of these simulation studies have used parameters that do not reflect most empirical datasets, and sometimes represent extreme conditions such as the independent convergence of distantly-related taxa to a GC content that differs substantially from the rest of the taxa in the tree. While these simulations are based on biological observations such as the evolution of extreme GC content differences among closely related bacteria (Mooers and Holmes 2000), they do not represent the degree of violation of SRH conditions typical of most datasets. Indeed, apart from extreme cases it remains relatively poorly understood to what extent different types and degrees of violations of the SRH conditions affect phylogenetic inference.
In this study, we examine the influence of violating the SRH assumptions on phylogenetic inference with SRH models using parameters that are derived from thousands of empirical datasets. We simulate nucleotide alignments under various non-stationary (and thus non-reversible) or/and non-homogeneous conditions and examine the effects of incorrectly assuming SRH conditions on inferring phylogenies from these data. Moreover, we examine the ability of different methods to detect non-SRH evolution across multiple sequence alignments. Several tests for detecting non-SRH evolution in nucleotide and amino acid alignments have been introduced (Lanave, et al. 1984; Lanave, et al. 1986; von Haeseler, et al. 1993; Lockhart, et al. 1994; Kumar and Gadagkar 2001; Phillips and Penny 2003; Weiss and von Haeseler 2003; Foster 2004; Ababneh, et al. 2006; Ho, et al. 2006; Jermiin, et al. 2019; Naser-Khdour, et al. 2019). However, these tests are rarely used in phylogenetic analysis (Jermiin, et al. 2004; Jermiin, et al. 2009), likely because many of them are difficult to apply in practice. In this study we focussed on three tests for detecting non-SRH evolution that are implemented in the widely-used IQ-TREE software (Minh, et al. 2020): the MaxSymTests (Naser-Khdour, et al. 2019), the compositional chi-square test (Preparata and Saccone 1987) as implemented in IQ-TREE (Nguyen, et al. 2015), and the test of non-stationarity proposed by Weiss and von Haeseler (Weiss and von Haeseler 2003). The MaxSymTests ask whether there is evidence in a single alignment that evolutionary symmetry imposed by SRH evolution is violated, and is a relatively new extension of similar tests designed for pairs of sequences (Jermiin, et al. 2019). The Weiss and von Haeseler (WH) test checks the homogeneity of the substitution model across the tree based on the pairwise sequence comparisons and performs a parametric bootstrap to assess the statistical significance (Weiss and von Haeseler 2003). The compositional chi-square test checks if the state composition of each sequence in the alignment is similar to the average state composition of the whole alignment, and is commonly-used to detect and sometimes remove
sequences that clearly violate the SRH conditions (e.g. Aouad, et al. 2018; Liu, et al. 2018; Martijn, et al. 2018; Puttick, et al. 2018; Song, et al. 2018; Fan, et al. 2020). The Chi-square test gives researchers a way of understanding whether each sequence in an alignment has state frequencies that are plausible given the overall state frequencies of the alignment. We know of no existing test which combines individual chi-square tests to assess whether the state frequencies across all sequences of an alignment is plausible under an SRH model. It is possible to do this with model adequacy tests, but this requires one to first fit a full model and a tree (Foster 2004; Brown and ElDabaje 2009; Duchene, et al. 2017), while our current work focusses on tests that can be performed quickly and efficiently on very large datasets prior to tree inference. We therefore use two different approaches in this study to leverage the information in from individual chi-square tests.

The two approaches we take to using information from chi-square tests reflect different ways of balancing false-positive and false-negative outcomes, and so may be thought of as appropriate for different situations. Our first approach to using the chi-square tests is to take the most conservative possible approach and score an alignment as violating SRH assumptions if at least one sequence fails the test. Using the Chi-square frequencies in this way is very conservative, and liable to have a high false positive rate that increases with the number of sequences in an alignment. However, in some practical cases when many loci are available but only a small number can be used for analyses, e.g. selecting ~50 loci for a Bayesian analysis out of many thousands available from whole genomes, a conservative approach such as this with a high false positive rate may be warranted. Our second approach is less conservative. In this approach we record the proportion of sequences in an alignment that fail the Chi-square test, and ask whether this proportion is correlated with the degree of non-stationarity in the simulations. This approach may be more useful in practical cases where researchers wish to rank a set of loci with respect to the severity of model violations.
MATERIALS AND METHODS

Simulations

In order to investigate the ability of SRH models to correctly infer topologies and branch lengths from non-SRH data, we devised a new approach that allows us to simulate alignments gradually ranging from true SRH conditions (with identical base frequencies and identical reversible substitution processes on every branch of the topology) to the most extreme violation with completely unrelated base frequencies and non-reversible substitution processes on every branch of the topology. For an alignment of $m$ taxa and $n$ sites, we will denote the set of all branches in the rooted tree $\tau$ as $\Phi = \{1, \ldots, l\}$.

We simulate data under two different simulation schemes as follows:

1. An inheritance scheme designed to reflect the evolutionary process, in which each node in the tree inherits its substitution processes from its parent with a constant strength of inheritance modified by the branch length connecting the two nodes. The scheme reflects the continuity of evolutionary processes that are changing through time along a phylogenetic tree.

2. A two-matrix scheme designed to reflect previous approaches to simulating non-SRH evolution, where two independent subtrees (that are not sisters nor descendants of each other) have an identical substitution process and that is distinct from the substitution process that operates on the rest of the tree. This scheme resembles convergent evolution.

Applying these two schemes allows us to ask how evolutionarily-inspired non-SRH simulations are affected by SRH assumptions (scheme 1) and then to directly compare these to the more extreme forms of non-SRH evolution that are more often simulated (scheme 2).
THE INFLUENCE OF MODEL VIOLATION ON PHYLOGENETIC INFERENCE

We will describe both simulation approaches in more detail below. But we start by describing how we choose model parameters for our simulations.

**Estimating Empirical Parameter Distributions and Tree Topologies for Simulations**

Both of our simulation approaches require us to choose base frequency vectors and rate matrices with which to simulate alignments. Generating these at random could limit the applicability of our results because it is unlikely that randomly-generated base frequency vectors or rate matrices would reflect reality. To address this, we instead estimated base frequency vectors and rate matrices from a large collection of empirical alignments, and then used these parameters for our simulations.

In order to estimate the distributions of the empirical base frequencies ($\Pi$) and the substitution rates ($X$) we used 32,666 partitions from 49 nucleotide datasets (Appendix Table A.1). Consisting of different types of loci (codon positions, rRNA, tRNA, introns, intergenic spacers and UCEs) and genomes (nuclear, mitochondria, virus, plastid). Since different partitions of the genome evolve differently, for each partition, we ran IQ-TREE with a GTR model and free rate heterogeneity across sites (Yang 1995) with 4 categories + invariant sites. This gave us the distributions of 32,666 estimates of each parameter in the GTR matrix $(A\leftrightarrow C, A\leftrightarrow G, A\leftrightarrow T, C\leftrightarrow G, C\leftrightarrow T, G\leftrightarrow T)$ and the distribution of each base frequency ($\pi_A, \pi_C, \pi_G, \pi_T$).

We use a similar approach to estimate the distribution of branch lengths. Estimating branch lengths from each partition separately could be misleading because there tends to be a high stochastic error in branch lengths estimated from short single-partition alignments (Kumar, et al. 2012). Therefore, in order to estimate the empirical distribution of the branch lengths, we instead estimated a single set of branch lengths from each of our 49 nucleotide datasets and complemented these with an additional 18 amino-acid datasets. For each dataset,
we ran IQ-TREE with the best-fit fully-partitioned model (Chernomor, et al. 2016), which allows each partition to have its own evolutionary model and edge-linked rates determined by ModelFinder (Kalyaanamoorthy, et al. 2017). We then rooted the tree with the outgroup taxa (if provided) and extracted the empirical branch lengths of the ingroup ($T$) for each of the 33,178 partitions from 67 nucleotide and amino acid datasets.

Finally, for each parameter in $X$ (5 parameters - G↔T equals to 1) and $\Pi$ (4 parameters), and for each distribution in $T$ (67 distributions - each dataset is an independent distribution) we find the best-fit distribution from 36 common probability distributions using the Kolmogorov-Smirnov test using SciPy (Virtanen, et al. 2020). We then sampled parameters for our simulations from these best-fit distributions. Since the parameters of $\Pi$ are not independent, to sample a base-frequency vector we randomly sampled a parameter from each of the four base-frequency’s best-fit distribution and then normalized these parameters to sum to 1.

The tree topology $\tau$ is derived from birth-death simulations with speciation rate $\lambda$, extinction rate $\mu$ and the fraction of sampled taxa $f$ using TreeSim package with a fixed number of extant species (Stadler 2011). In principle, it is possible to estimate the speciation and extinction rates from empirical data (Nee, et al. 1994; Rannala and Yang 1996; Magallon and Sanderson 2001). However, not knowing the fraction of sampled taxa a priori will tend to bias such estimates (Stadler 2013; Hua and Lanfear 2018). Because of the challenges of reliably estimating empirical speciation and extinction rates, we instead randomly sampled the speciation rate, the extinction rate and the fraction of sampled taxa from uniform distributions, to attempt to cover all the realistic regions of the parameter space.

$$\lambda \sim U(0,1), \quad \mu \sim U(0, \lambda), \quad f \sim U(0,1)$$

Note that under these conditions $\lambda$ is always greater than $\mu$. 
THE INFLUENCE OF MODEL VIOLATION ON PHYLOGENETIC INFERENCE

We simulated datasets with 20, 40, 60, 80 and 100 taxa. For each number of taxa ($m$), we simulated 3960 topologies with random speciation rate ($\lambda$), random extinction rate ($\mu$) and random fraction of sampled taxa ($f$). For each of these topologies, we then randomly choose a distribution from set $T$ and sampled the branch lengths from this distribution ($2m–2$ branch lengths in total).

Other Python libraries that we used for the simulations are NumPy (Walt, et al. 2011), pandas (McKinney 2010) and ETE3 (Huerta-Cepas, et al. 2016). The python scripts for all simulations can be found on Github (https://github.com/suhanaser/empiricalGTRdist).

Inheritance Evolution: Inheritance Scheme Simulations

An evolutionary scenario would, ideally, have each lineage inheriting the parameters of its molecular evolutionary process from its parent lineage. At one extreme — where inheritance is perfect and the original evolutionary process is SRH, such a process would define a molecular evolutionary process that is SRH across the entire topology by simply defining a single SRH model at the root node. At the other extreme, where the association between parent and offspring lineages is no better than random and the original process is not SRH, there is no association between parent and offspring lineages and the process is maximally non-SRH. To mimic this situation, we designed a simulation approach that allows us to vary the homogeneity and stationarity assumptions both independently and together.

Our inheritance scheme allows us to vary the degree to which a single alignment has evolved under SRH conditions by simply adjusting the strength of inheritance of the substitution process and the base frequencies either jointly via a parameter we call $\rho$, or independently via parameters $\nu$ and $\omega$ respectively. When the inheritance parameters are set to 1 and the model at the root of the tree is reversible, the model will conform to SRH conditions. We can simulate increasing violation of SRH conditions simply by decreasing the
inheritance parameters towards zero. When the relevant inheritance parameter is less than one, each branch inherits some proportion of its substitution model from the parent branch, while the remaining proportion of the model is selected at random from the empirical parameter distributions. In practice, the parameter in a descendant branch is calculated as the weighted sum of the parameter in the parent branch (where the weight is the inheritance parameter) and a randomly-generated parameter from the appropriate empirical distribution (where the weight is one minus the inheritance parameter).

We simulated data under five different categories of conditions using this scheme, in order to examine independently and together the effects of relaxing the stationarity and homogeneity assumptions.

1. **SRH conditions (Fig. 1a).**—In the simplest case for a model that conforms to the SRH assumptions, where model parameters are generated from the empirical distributions. This describes a model in which all branches inherit this reversible model from their parent branch without variation, such that all branches on the tree have the same reversible substitution model, conforming to the SRH assumptions.

2. **Relaxing the stationarity assumption (Fig. 1b).**—In order to hold the homogeneity assumption but relax the stationarity assumption, we introduce a parameter called $\nu$ ($0 \leq \nu \leq 1$) that allows to vary the state frequency at the root while still keeping the same rate matrix for all branches of the tree. Mathematically, this can be described as:

$$
\begin{align*}
Q_i &= Q_0 = \pi_0 S_0 \\
\pi_{\text{root}} &= \nu^{d_{\text{root}}} \pi_0 + (1 - \nu^{d_{\text{root}}}) \pi \\
\end{align*}
$$

Where $Q_i$ is the substitution rate matrix operating on branch $i$, and $d_{\text{root}}$ is the branch length of the root branch.
THE INFLUENCE OF MODEL VIOLATION ON PHYLOGENETIC INFERENCE

When $\nu = 1$, $\pi_{\text{root}}$ is equal to $\pi_0$ and this scheme boils down to the first SRH condition. When $\nu = 0$, $\pi_{\text{root}}$ is equal to $\pi$, meaning that the root frequency is generated separately from $\pi_0$. $\pi_{\text{root}}$ will vary between these two extremes when $\nu$ is between 0 and 1, with lower $\nu$ reflecting a larger deviation from stationary conditions.

3. Relaxing the homogeneity assumption (Fig. 1c).—In order to hold the stationarity assumption but relax the homogeneity assumption we need to simulate data in which $\nu$ is set to 1 (such that all branches have the same base frequencies as the root node), but we introduce a parameter $\omega$ that varies between zero and one (such that the inheritance of the parameters of the $Q$ matrix ranges from completely random to near-perfect). We can describe this mathematically as follows:

\[
\begin{align*}
Q_i &= \omega^d_i \pi_0 S_j + (1 - \omega^d_i) \pi_0 S \\
\pi_{\text{root}} &= \pi_0 \\
\end{align*}
\]

Where $Q_i$ is the process operating on branch $i$, $S_j$ are the substitution rates on the parent branch of branch $i$, and $d_i$ is the branch length of the branch $i$.

4. Relaxing the stationarity and homogeneity assumptions simultaneously but independently (Fig. 1d).—We can simulate non-stationary and non-homogeneous data by setting both $\nu$ and $\omega$ to values less than one. When we relax both assumptions, we will allow $Q_i$ and $\pi_{\text{root}}$ to vary simultaneously but independently:

\[
\begin{align*}
Q_i &= \omega^d_i \pi_0 S_j + (1 - \omega^d_i) \pi_0 S \\
\pi_{\text{root}} &= \nu^d_{\text{root}} \pi_0 + (1 - \nu^d_{\text{root}}) \pi \\
\end{align*}
\]

\[
\begin{align*}
\omega &\in \mathbb{R}: 0 \leq \omega < 1 \\
\nu &\in \mathbb{R}: 0 \leq \nu < 1 \\
\end{align*}
\]
Relaxing the stationarity and homogeneity assumptions jointly (Fig. 1e).—While the 4th set of simulation conditions, above, allows us to vary homogeneity and stationarity jointly but independently, it suffers from the limitation that we have a maximum of two base frequency vectors in the tree ($\pi_{\text{root}}$ and $\pi_0$). To relax this assumption further, we will allow $Q_i$ to vary while $\pi_{\text{root}}$ stays fixed. In those settings, both homogeneity and stationarity will increase with $\rho$.

$$
\begin{align*}
    \{Q_i &= \rho \pi_0 S_0 + (1 - \rho)\pi S \quad i \in \{\Phi\}, \quad \{\rho \in \mathbb{R}: 0 \leq \rho \leq 1\} \\
    \pi_{\text{root}} &= \pi_0
\end{align*}
$$

**Figure 1.** An example of 5 taxon tree with different degrees for homogeneity and stationarity. (a) stationary and homogeneous, (b) stationarity but not homogeneous, (c) non-stationary but homogeneous, (d) non-stationary and non-homogeneous where the stationarity and homogeneity assumptions are relaxed simultaneously but independently, (e) non-stationary and non-homogeneous where the stationarity and homogeneity assumptions are relaxed jointly.

**Convergent Evolution: The Two-Matrix Scheme Simulations**

Previous studies for simulating non-SRH evolution on phylogenies have used an approach in which two distantly related branches undergo severe but correlated changes in the molecular evolutionary process. To compare this approach to the more evolutionarily-motivated approach described above, we randomly chose two nodes that are not sisters and not descendants of each other and assigned a different rate matrix (denoted by $S_1 \pi_1$) from the rest of the tree to all their descendant branches (Fig. 2).
Figure 2. Non-stationary and non-homogeneous process with two different rate matrices $Q_0$ and $Q_1$.

Simulation Parameters

The simulation parameters that we use in this study are the strength of inheritance of the substitution process ($\omega$), strength of inheritance of the base frequencies ($\nu$), strength of inheritance of the substitution process and base frequencies ($\rho$), number of sites ($n$), and number of taxa ($m$) where the parameter space is:

$\omega, \nu, \rho \in \{0, 10^{-25}, 10^{-20}, 10^{-15}, 10^{-10}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 0.1, 1.0\}$

$n \in \{100, 1000, 10000\}$

$m \in \{20, 40, 60, 80, 100\}$

The inheritance weight parameters ($\omega, \nu, \rho$) were chosen to represent an even spread of corrected inheritance weights (i.e., the inheritance weights raised to the power of $d$, where $d$ is the branch length) between zero and one. The number of taxa and number of sites are chosen to reflect typical sizes of empirical datasets. For simulation under the inheritance scheme, we simulated 10 alignments of each combination of $n, m, \nu$, and $\omega$ or $n, m$, and $\rho$ for a total of 19,800 simulations. For simulation under the two-matrix scheme, we simulated 1000 alignments of each combination of $n$ and $m$ for a total of 15,000 simulations.
Our first goal is to understand how the incorrect use of SRH models on data that have evolved under non-SRH processes can affect phylogenetic inference. To do this, we compare the tree topologies and branch lengths estimated with SRH models in IQ-TREE to the topologies and branch lengths used to simulate each dataset. For each simulated alignment, we ran IQ-TREE with ModelFinder (Kalyaanamoorthy, et al. 2017) and 1000 ultrafast bootstrap replicates (Hoang, et al. 2018). In order to assess the ability of SRH models to infer the correct tree topology we then compared the simulated tree topology to the estimated tree topology using three different metrics – normalized Robinson-Foulds distance (Robinson and Foulds 1981), Quartet distance (Estabrook, et al. 1985), and the Path-difference distance (Steel and Penny 1993). The normalized Robinson-Foulds distance between two trees is the fraction of internal branches that appear in one tree but not the other. It ranges from 0 to 1, where 0 means that the two trees are topologically identical and 1 means that the two trees have no branches in common. In order to assess the accuracy of branch length estimates, we tested whether the estimated branch lengths and the original branch lengths are drawn from the same distribution using the two-sample Kolmogorov-Smirnov test.

Detecting non-SRH Processes

We, therefore, tested the ability of three tests implemented in IQ-TREE to detect violation of the SRH assumptions: the MaxSymTests (Naser-Khdour, et al. 2019), and the compositional Chi-square test, and the WvH test (Weiss and von Haeseler 2003). These three tests only need the composition of the alignment and therefore can be used with any analysis in IQ-TREE by adding the appropriate options to the command line, except for the Chi-square test that runs automatically for each alignment (Table 1).

| TABLE 1. IQ-TREE option for each test |
| test | IQ-TREE option |
|      |                |
Since the Chi-square test tells us whether each sequence in the alignment fails the compositional homogeneity assumption, we use two different approaches that leverage the results of the Chi-square test (see also the Introduction):

1) A very conservative approach that we denote as $\text{Chi}^2_{\text{cons}}$. In this approach, we consider the alignment to fail the Chi-square test if one or more of the sequences in the alignment fails the test.

2) A less conservative ranking approach that we denote as $\text{Chi}^2_{\text{rank}}$. We record for each alignment what proportion of sequences that fail Chi-square test.

In the first case, we ask whether the proportion of replicate simulated alignments with one or more sequences failing the Chi-square test increases with the degree of violation of SRH conditions in the simulations. In the second case, we ask whether the proportion of sequences that fail the Chi-square test increases with the degree of violation of the SRH conditions in the simulations.

RESULTS

Empirical Distributions

We derived the empirical distributions of the substitution model parameters, the nucleotide frequencies, and the proportion of invariant sites from 32,666 nucleotide alignments (Appendix Table A.2). The empirical distribution of branch lengths we derived from 67 nucleotide and amino acid alignments consist of 33,178 partitions (Appendix Table A.1).
Using Kolmogorov-Smirnov test, we found the best-fit probability distribution for each one of these empirical distributions (Table 2, Appendix Table A.2, Appendix Figs. A.1-3).

**TABLE 2. The best-fit probability distribution by Kolmogorov-Smirnov test**

| Parameter | Best-fit distribution       | Shape (α) | Scale (β) | Location (x₀) |
|-----------|----------------------------|-----------|-----------|---------------|
| A→C       | Log-Laplace                | 1.695     | 1.636     | -0.152        |
| A→G       | Log-Laplace                | 1.465     | 4.930     | -0.140        |
| A→T       | Inverse-Weibull            | 3.015     | 1.841     | -1.154        |
| C→G       | Inverse-Weibull            | 1.793     | 1.651     | -0.741        |
| C→T       | Log-Laplace                | 1.551     | 5.182     | -0.175        |
| \( \pi_A \) | Generalized-logistic     | 0.557     | 0.026     | 0.313         |
| \( \pi_T \) | Exponential-Weibull      | 0.843, 4.872 | 0.294 | -0.001        |
| \( \pi_C \) | Exponential-normal       | 1.769     | 0.027     | 0.173         |
| \( \pi_G \) | Power-log-normal          | 0.090, 0.039 | 0.614 | -0.471        |
| %I        | Beta                       | 0.577, 4.707 | 2.162 | -5.437        |
| Branch length | Power-log-normal    | 1.208, 1.443 | 0.017 | -7.1 e-05     |

**Phylogenetic Inference is Unaffected by Violation of SRH Conditions in an inheritance Framework**

Surprisingly, our results for the inheritance simulation scheme show that there is no detectable relationship between the severity with which SRH conditions were violated during the simulations and the accuracy of the tree topology or the tree length inferred from the simulated data. Specifically, we saw no relationship between the inheritance weight and the normalized RF (Robinson-Foulds), QD (Quartet Distance), or NPD (Normalized Path Difference) metrics in any of our inheritance simulations (Fig. 3, Appendix Figs. A.4-7). These metrics measure the difference between the inferred tree and the tree from which the alignment was simulated. If stronger violation of the SRH conditions affects phylogenetic inference we should expect to see that the distances are higher when the inheritance weight is lower, because a lower inheritance weight implies stronger model violation through less homogeneity (for the rate matrix) and less stationarity (for the base
THE INFLUENCE OF MODEL VIOLATION ON PHYLOGENETIC INFERENCE

frequencies). In addition, our results show that the proportion of simulated datasets for which
the simulated tree is recovered from the simulated alignment is constant at around 0.25 in the
inheritance scheme simulations regardless of the inheritance weight (Fig. 3, Appendix Fig.
A.4). Finally, we see no correlation between the inheritance weights and the proportion of
datasets that fail a Kolmogorov-Smirnov test comparing the true and estimated branch
lengths, suggesting that violation of SRH assumptions in our evolutionary framework has no
detectable effect on the estimation of branch lengths (Fig. 4, Appendix Fig. A.19).

Tree Topologies, but not Branch Lengths, are Affected by Severe and Convergent Violation of
SRH Conditions

Our results show that convergent violation of SRH assumptions by allowing two
distantly related branches to have identical substitution models has increasingly severe effects
on phylogenetic inference as the severity of the changes in the substitution models increases.
Under the two-matrix scheme, we expect to see higher distances between the true tree and the
estimated tree when there are larger Euclidian distances between the original matrix and the
matrix under which the divergent clades evolve. In two out of the three metrics (Robison-
Foulds and Path-Difference) we found a weak but significant correlation between the distance
between the matrices and the distance between the topologies (Fig. 3, Appendix Fig. A.4,
Appendix Figs. A.8-10). However, in the third metric (Quartet Distance) we found no
correlation. Notably, the distance between the true tree and the estimated tree increases only
when the Euclidean distance between the two matrices is very high. Nevertheless, the
proportion of simulated datasets for which the simulated tree is recovered from the simulated
alignment declines exponentially as the difference between the matrices in the two-matrix
scheme increases (Fig. 3).
Figure 3. Normalized Robinson-Foulds distance between the estimated tree topology and the original tree topology as a function of the inheritance weight ($\nu$, $\omega$, $\rho$) in the first simulation scheme, and the distance between the two matrices (delta(Q)) in the second simulation scheme. The small plots show the proportion of datasets in which the distance between the estimated topology and the original topology equals zero as a function of the inheritance weight and the distance between the two matrices. If violation of SRH model assumptions increases topological error, we expect the nRF distance to increase towards the right of each plot. The figure shows that for the first simulation scheme, which mimics a stochastic evolutionary process, there is no detectable association between violation of SRH conditions and topological error. For the second simulation scheme, which mimics an extreme convergent situation, topological error increases with increasing violation of SRH conditions.
The proportion of analyses in which the simulated tree is recovered positively declines from around 0.20 when there is no model violation to zero when the Euclidean distance between the matrices is around 2000, confirming that even the lowest levels of SRH violation have detectable negative effects on phylogenetic inference under the two-matrix scheme.

Finally, we see no correlation between the Euclidean distance between the two matrices and the proportion of datasets that fail a Kolmogorov-Smirnov test comparing the distributions of the true and estimated branch lengths, suggesting that violation of SRH assumptions in the convergent framework has limited effects on the estimation of branch lengths (Fig. 4, Appendix Fig. A.22).

**Figure 4.** The percentage of datasets that pass the KS test (a dataset passes when there is no evidence to suggest that the inferred branch length distributions differ) as a function of the inheritance weights ($\nu, \rho, \omega$) (left-hand panel) and maximum Euclidian distance between the two matrices used to simulate the data (right-hand panel).

Tests for Detecting non-SRH Processes are Successful but have High False-Negative Rates

As expected, the ability of all three MaxSym tests to reject the null hypothesis of stationarity and homogeneity improves as the inheritance weight in the evolutionary
simulations decreases (i.e. as the violation of SRH conditions increases), the distance between the two matrices in the convergent simulations increases, and the number of sites in the alignment increases (Fig. 5, Appendix Figs. A.11-13). Moreover, the three MaxSym tests have a reasonable false positive rate of approximately 4.5% (Appendix Table A.4). However, they also have very high false-negative rates of 50-90%, depending on the test and the particular simulation conditions (Fig. 5, Appendix Table A.3, Appendix Figs. A.14-16). In the two-matrix scheme simulations, the false negative rates of MaxSym, MaxSym$_{int}$, and MaxSym$_{int}$ tests are 67%, 66%, and 87%, respectively. Thus, across all simulation conditions, a significant result from a MaxSymTest can be reliably interpreted as indicating that an alignment violates the SRH conditions, but the test will fail to identify many such alignments.

Similarly to the MaxSym tests results, the $Chi^2_{cons}$ and $Chi^2_{rank}$ tests show an increase in the proportion of alignments and/or sequences that fail the test in each dataset as the inheritance weight decreases, and the number of sites increases (Fig. 5a, Appendix Fig. A.17). The false-positive rates of the $Chi^2_{cons}$ test is 6% (Appendix Table A.6). The false-negative rate of the $Chi^2_{cons}$ test in the inheritance-scheme simulations is 57% (Appendix Table A.5). Moreover, similar to the MaxSym tests, in the two-matrix scheme simulation, the percentage of datasets that pass the $Chi^2_{cons}$ decreases logarithmically the higher the distance between the two matrices (Fig. 5b, Appendix Fig. A.18). The false negative rate of the $Chi^2_{cons}$ test under extreme convergent evolution is the smallest of all the tests considered here under these conditions, and it is around 44% (Appendix Table A.5).

In the inheritance-scheme simulations, similarly to the MaxSym tests, the $Chi^2_{cons}$ and $Chi^2_{rank}$, the WvH test shows an increase in the proportion of alignments that fail the test as the inheritance weight ($\omega$ and $\rho$) decreases. However, $\nu$ has no effect on the proportion of alignments that fail the WvH test. The false-positive rates of the WvH test is 3.5% (Appendix
THE INFLUENCE OF MODEL VIOLATION ON PHYLOGENETIC INFERENCE

Table A.8), which is lower than any of the MaxSym tests or the \( \chi^2_{\text{cons}} \) test. In addition, the false negative rate of the WvH test (Appendix Table A.7) in the inheritance-scheme simulations is lower than all the other tests (~30%) but it is still high under the two-matrix scheme simulations (~67%). Yet, due to numerical instability, the WvH test could be only applied to half of the datasets in the two simulation schemes.

![Graphs showing the influence of model violation on phylogenetic inference](image)

(a)
Figure 5. The mean percentage of datasets that pass each of the MaxSym tests, the WvH test, and the Chi-square test as a function of (a) the inheritance weights \((v, p, o)\) and (b) maximum Euclidian distance between the two matrices. We define datasets that pass the Chi2Cons test as datasets where all the sequences pass the Chi-square test. On the other hand, Chi2Rank shows the proportion of sequences that pass the Chi-square test in each dataset.

MaxSymTest\textsubscript{int} is a good predictor of correct tree inference

A key question for empiricists is whether tests of model adequacy are likely to improve phylogenetic inference. To explore this in our simulation framework, we asked whether datasets that are rejected by the tests we evaluated tended to be associated with more phylogenetic tree error than those that were not rejected. To do this, we used three different metrics of tree distance (the normalized Robison-Foulds (RF), Path-Difference, and Quartet distance) and asked whether datasets that fail the test (i.e. have detectable non-SRH processes) tended to result in trees that were further from the true tree (i.e. had higher nRF distances) when analysed using SRH models. All three showed very similar results, so we
show the normalized Robinson-Foulds results here (Fig. 6) and the other metrics in the supplementary information (Appendix Fig. A.23a, Appendix Fig. A.24a).

For the inheritance scheme simulations we found as expected that datasets which failed the MaxSym tests were associated with trees much further from the true tree than those that passed the tests, although there was substantial variation within each category (Fig. 6a, Appendix Fig. A.23a, Appendix Fig. A.24a). Surprisingly, this pattern was reversed for the \( \text{Chi}_\text{cons}^2 \) test, and there was a very small difference in tree distances with the WvH test (Fig. 6a). Welch’s t-test results suggest all of the differences are statistically significant (p<<0.05, Fig. 6a).

For the two-matrix simulations the only test for which datasets which failed were associated with trees further from the true tree was the MaxSym\(_{\text{int}}\) test (Fig. 6b, Appendix Fig. A.23b, Appendix Fig. A.24b). For all other tests, datasets which failed the test were associated with trees that were markedly closer to the true tree than datasets which passed the tests (Fig. 6b, Appendix Fig. A.23b, Appendix Fig. A.24b). Again, Welch’s t-test results suggest all of the differences are statistically significant (p<<0.05, Fig. 6b).
Figure 6. The normalized Robison-Foulds distance as a function of passing/failing the tests in (a) the inheritance scheme simulations and (b) the two-matrix scheme simulations. The p-value in each sub-figure is calculated from the Welch’s t-test statistics.

Discussion

Using two different simulation schemes, we explored the impact of violating the assumption of evolution under stationary, reversible, and homogeneous (SRH) conditions on ML phylogenetic tree inference. Our study extends the simulations in many previous studies by simulating data under an evolutionary scenario in which molecular evolutionary models evolve along a phylogeny. Our results show that the inference of phylogenetic tree topologies and branch lengths are surprisingly robust to violations of SRH assumptions under an evolutionary scheme. But similarly to previous studies, we show that in extreme cases of convergent molecular evolution the incorrect assumption of SRH conditions can severely mislead phylogenetic inference.

The first simulation scheme we introduced in this paper, which we called the inheritance scheme, allows tree branches to inherit their substitution process from their ancestor. The second simulation scheme, which we called the two-matrix scheme, is similar to previous studies and allows two distantly related monophyletic sub-trees to evolve with a
different evolutionary process from the rest of the tree (Galtier and Gouy 1995; Jermiin, et al. 2004; Jayaswal, et al. 2011; Duchene, et al. 2017).

Surprisingly, our results show no correlation between errors in the topology or branch length inference and any of the inheritance scheme parameters, even in extreme cases where the evolutionary process is completely heterogeneous and non-stationary. These results indicate that ML tree inference with SRH models is surprisingly robust to even quite extreme violations of the SRH conditions.

Under the two-matrix simulation scheme, we found a small but significant increase in topological inference error and the extent of the violation of the SRH assumptions. Specifically, the more extreme the evolutionary convergence, the larger the errors in topological inference that assumes SRH conditions. Despite this, we found no correlation between branch length inference and the distance between the two matrices. These results emphasize the limitations of ML inference to operate under certain model violations, especially when these violations are highly imbalanced along the tree, as in the case of the two-matrix scheme simulations. These results indicate that the inference of the substitution model is more influenced by the imbalance of the model violation distribution along the tree than by the model violation itself. This conclusion agrees well with all previous simulation studies of similar simulation conditions (e.g. Jermiin, et al. 2004; Duchene, et al. 2017; Jermiin, et al. 2019).

In this study, we also tested the power of the MaxSym tests, WvH test and two variations of the Chi-squared test to detect model violation due to non-SRH evolution. Our results show that those tests were able to detect some model violation in both simulation schemes. As expected, the power of all tests to detect model violation due to non-SRH evolution improves dramatically as alignment length increases, reflecting simply the larger
amount of information available in longer alignments. However, the power of most of the
tests we looked at was somewhat limited— even in the best-case scenario when violation of
the SRH conditions was severe, most tests were able to detect this violation in less than 50% of the simulated datasets (Fig. 5). The two exceptions were the WvH test, which was able to
detect the vast majority of datasets simulated with model violation under the inheritance
scheme simulations (Fig. 6a) and the conservative Chi-Square test, which was able to detect
the majority of datasets simulated with model violation under the convergent evolution
scheme. However, the WvH test could not be applied to half of the datasets in our simulations
due to numerical instability, suggesting that it may be less useful for detecting violations of
SRH conditions in practice than the other tests.

The utility of any test of model adequacy in practice is likely to be tied to the amount
of phylogenetic error that a test helps empiricists avoid. All models used in phylogenetic
analyses are gross oversimplifications of highly complex molecular evolutionary processes,
and so merely detecting violations of models is necessary but not sufficient for a model
adequacy test to be useful. Because of this, we asked for each test whether the datasets that
fail the test were associated with more or less topological inference error than the datasets
that passed the test. Surprisingly, the only test that performed consistently well in this regard
was the MaxSymTest_int. Under the inheritance scheme simulations, all three MaxSym tests
are good predictors of phylogenetic accuracy; trees that pass any of those tests are closer to
the true tree than trees that fail. The WvH and $\text{Chi}^2_{\text{cons}}$ tests on the other hand are bad
predictors of phylogenetic accuracy; trees that fail the $\text{Chi}^2_{\text{cons}}$ test are usually closer to the
tree, while there is only a small difference between trees that fail and trees that pass the
WvH test. Surprisingly, under the convergent simulation scheme, the MaxSymTest_int is the
only test for which datasets that pass the test are closer to the true tree than datasets that fail.
the test (Fig. 6). For all other tests, the datasets that pass the test were substantially further from the true tree than those that fail the test.

It is challenging to disentangle why some tests of the SRH assumptions tend to detect datasets that are associated with more topological error, while others show the opposite tendency (Fig. 6), although we suspect this is often driven by the interplay of the power of the tests, phylogenetic signal, and stochastic error in tree estimation. Across all simulation conditions, the only test which consistently showed the desirable behaviour from an empirical standpoint (i.e. where datasets that fail the test are associated with more topological error) was the MaxSymTest\textsubscript{int}. All other tests showed evidence of having the opposite tendency (Fig. 6) in at least some simulation conditions. In the case of the WvH test, for which alignments that fail the test were consistently associated with less topological inference error when analysed under SRH models, we suspect that the underlying reason may be the reliance of the test on a parametric bootstrap. The WvH test depends fundamentally on a tree estimated with an SRH model to estimate the null distribution of the test statistic. If this tree is wrong, as we show occurs under model violation, then the null distribution may be incorrect and the test misled. For the other tests we suspect that the tendency is driven largely by the fact that datasets with few informative sites will tend to both pass the tests and be associated with high topological error, with both caused by the limited information in the data, although further work is needed to understand these relationships in more detail.

Nevertheless, the observation that across all simulation conditions, datasets that fail the MaxSymTest\textsubscript{int} are associated with higher topological error do suggest that violations of the homogeneity assumption might be the most important when it comes to phylogenetic inference with SRH models, since the MaxSymTest\textsubscript{int} tests primarily for violations of homogeneity.
These results combined with the results from the inheritance scheme simulations, emphasize the need to use different methods and tests for model violation in phylogenetic analyses since each test can capture a different aspect of model violation. A new phylogenetic protocol (Jermiin, et al. 2020) stresses the need to validate the assumptions of the models in advance. If the data in hand violates the model’s assumptions then different models or methods should be considered. A surprising result from this work is that the MaxSymTest is a good predictor for phylogenetic accuracy. Yet, one should bear in mind that this test has the highest false-negative rate among all of the tests examined in this study.

It is noteworthy that our results from the different simulation schemes agree with the results from empirical data (Naser-Khdour, et al. 2019). They emphasize the impact of model violation due to non-SRH evolution on phylogenetic inference and suggest that reducing model violation in phylogenetic analysis by using the new protocol of phylogenetic inference (Jermiin, et al. 2020) or using more complex substitution models e.g. (Galtier and Gouy 1998; Tamura and Kumar 2002; Blanquart and Lartillot 2008; Dutheil, et al. 2012; Zou, et al. 2012; Groussin, et al. 2013; Jayaswal, et al. 2014) has the potential to improve phylogenetic accuracy.

For the purpose of this study, in order to simulate data that mimic as closely as possible empirical alignments, we extracted the empirical distributions of base frequencies, substitution rates, proportion of invariable sites, and branch lengths from tens of thousands of empirical datasets. In addition to their use in this paper, these empirical distributions, along with their best-fit distributions may be useful for a wide variety of simulation studies, or for specifying prior distributions for Bayesian phylogenetic methods.
THE INFLUENCE OF MODEL VIOLATION ON PHYLOGENETIC INFERENCE

FUNDING

This work was supported by an Australian Research Council (Grant No. DP200103151 to R.L., B.Q.M.) and by a Chan-Zuckerberg Initiative grant to B.Q.M and R.L.

REFERENCES

Ababneh F, Jermiin LS, Ma C, Robinson J. 2006. Matched-pairs tests of homogeneity with applications to homologous nucleotide sequences. Bioinformatics 22:1225-1231.

Anderson FE, Bergman A, Cheng SH, Pankey MS, Valinassab T. 2013. Data from: Lights out: the evolution of bacterial bioluminescence in Loliginidae. In: Dryad Data Repository.

Anderson FE, Bergman A, Cheng SH, Pankey MS, Valinassab T. 2014. Lights out: the evolution of bacterial bioluminescence in Loliginidae. Hydrobiologia 725:189-203.

Aouad M, Taib N, Oudart A, Lecocq M, Gouy M, Brochier-Armanet C. 2018. Extreme halophilic archaea derive from two distinct methanogen Class II lineages. Mol Phylogenet Evol 127:46-54.

Ballesteros JA, Sharma PP. 2019a. A Critical Appraisal of the Placement of Xiphosura (Chelicerata) with Account of Known Sources of Phylogenetic Error. Syst. Biol.

Ballesteros JA, Sharma PP. 2019b. Data from: A critical appraisal of the placement of Xiphosura (Chelicerata) with account of known sources of phylogenetic error. In: Dryad.

Barry D, Hartigan JA. 1987. Statistical Analysis of Hominoid Molecular Evolution. Statistical Science 2:191-207.

Becker EA, Yao AI, Seitzer PM, Kind T, Wang T, Eigenheer R, Shao KS, Yarov-Yarovoy V, Facciotti MT. 2016. A Large and Phylogenetically Diverse Class of Type 1 Opsins Lacking a Canonical Retinal Binding Site. PLoS One 11:e0156543.

Becker EA, Yao AI, Seitzer PM, Kind T, Wang T, Eigenheer R, Shao KSY, Yarov-Yarovoy V, Facciotti MT. 2017. Data from: A large and phylogenetically diverse class of type 1 opsins lacking a canonical retinal binding site. In: Dryad.
Bergsten J, Nilsson AN, Ronquist F. 2013a. Bayesian tests of topology hypotheses with an example from diving beetles. Syst. Biol. 62:660-673.

Bergsten J, Nilsson AN, Ronquist F. 2013b. Data from: Bayesian tests of topology hypotheses with an example from diving beetles. In: Dryad Data Repository.

Betancur RR, Li C, Munroe TA, Ballesteros JA, Orti G. 2013. Addressing gene tree discordance and non-stationarity to resolve a multi-locus phylogeny of the flatfishes (Teleostei: Pleuronectiformes). Syst. Biol. 62:763-785.

Blanquart S, Lartillot N. 2008. A site- and time-heterogeneous model of amino acid replacement. Mol. Biol. Evol. 25:842-858.

Borowiec ML, Lee EK, Chiu JC, Plachetzki DC. 2016. Data from: Extracting phylogenetic signal and accounting for bias in whole-genome data sets supports the Ctenophora as sister to remaining Metazoa. In: Dryad Digital Repository.

Borowiec ML, Lee EK, Chiu JC, Plachetzki DC. 2015. Extracting phylogenetic signal and accounting for bias in whole-genome data sets supports the Ctenophora as sister to remaining Metazoa. BMC Genomics 16:987.

Boussau B, Gouy M. 2006. Efficient likelihood computations with nonreversible models of evolution. Syst. Biol. 55:756-768.

Branstetter MG, Danforth BN, Pitts JP, Faircloth BC, Ward PS, Buffington ML, Gates MW, Kula RR, Brady SG. 2017a. Data from: Phylogenomic insights into the evolution of stinging wasps and the origins of ants and bees. In: Dryad Digital Repository.

Branstetter MG, Danforth BN, Pitts JP, Faircloth BC, Ward PS, Buffington ML, Gates MW, Kula RR, Brady SG. 2017b. Phylogenomic Insights into the Evolution of Stinging Wasps and the Origins of Ants and Bees. Curr. Biol. 27:1019-1025.

Broughton RE, Betancur RR, Li C, Arratia G, Orti G. 2013a. Data from: Multi-locus phylogenetic analysis reveals the pattern and tempo of bony fish evolution. In: Dryad Data Repository.

Broughton RE, Betancur RR, Li C, Arratia G, Orti G. 2013b. Multi-locus phylogenetic analysis reveals the pattern and tempo of bony fish evolution. PLoS Curr 5.

Brown JM, ElDabaje R. 2009. PuMA: Bayesian analysis of partitioned (and unpartitioned) model adequacy. Bioinformatics 25:537-538.

Brown RM, Siler CD, Das I, Min PY. 2012a. Data from: Testing the phylogenetic affinities of Southeast Asia’s rarest geckos: Flap-legged geckos (Luperosaurus), Flying geckos (Ptychozoon) and their relationship to the pan-Asian genus Gekko. In: Dryad Data Repository.

Brown RM, Siler CD, Das I, Min Y. 2012b. Testing the phylogenetic affinities of Southeast Asia's rarest geckos: Flap-legged geckos (Luperosaurus), Flying
THE INFLUENCE OF MODEL VIOLATION ON PHYLOGENETIC INFERENCE

geckos (Ptychozoon) and their relationship to the pan-Asian genus Gekko. Mol Phylogenet Evol 63:915-921.

Cannon JT, Vellutini BC, Smith J, 3rd, Ronquist F, Jondelius U, Hejnol A. 2016a. Xenacoelomorpha is the sister group to Nephrozoa. Nature 530:89-93.

Cannon JT, Vellutini BC, Smith J, Ronquist F, Jondelius U, Hejnol A. 2016b. Data from: Xenacoelomorpha is the sister group to Nephrozoa. In: Dryad Data Repository.

Chen M-Y, Liang D, Zhang P. 2015a. Data from: Selecting question-specific genes to reduce incongruence in phylogenomics: a case study of jawed vertebrate backbone phylogeny. In: Dryad.

Chen MY, Liang D, Zhang P. 2015b. Selecting Question-Specific Genes to Reduce Incongruence in Phylogenomics: A Case Study of Jawed Vertebrate Backbone Phylogeny. Syst. Biol. 64:1104-1120.

Chernomor O, von Haeseler A, Minh BQ. 2016. Terrace Aware Data Structure for Phylogenomic Inference from Supermatrices. Syst. Biol. 65:997-1008.

Cognato AI, Vogler AP. 2001a. Data from: Exploring data interaction and nucleotide alignment in a multiple gene analysis of Ips (Coleoptera: Scolytinae). In: Dryad Data Repository.

Cognato AI, Vogler AP. 2001b. Exploring data interaction and nucleotide alignment in a multiple gene analysis of Ips (Coleoptera: Scolytinae). Syst. Biol. 50:758-780.

Collins TM, Fedrigo O, Naylor GJ. 2005. Choosing the best genes for the job: the case for stationary genes in genome-scale phylogenetics. Syst. Biol. 54:493-500.

Crawford NG, Faircloth BC, McCormack JE, Brumfield RT, Winker K, Glenn TC. 2012a. Data from: More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. In: Dryad Digital Repository.

Crawford NG, Faircloth BC, McCormack JE, Brumfield RT, Winker K, Glenn TC. 2012b. More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. Biol. Lett. 8:783-786.

Day JJ, Peart CR, Brown KJ, Friel JP, Bills R, Moritz T. 2013a. Continental diversification of an African catfish radiation (Mochokidae: Synodontis). Syst. Biol. 62:351-365.

Day JJ, Peart CR, Brown KJ, Friel JP, Bills R, Moritz T. 2013b. Data from: Continental diversification of an African catfish radiation (Mochokidae: Synodontis). In: Dryad Data Repository.

Devitt TJ, Cameron Devitt SE, Hollingsworth BD, McGuire JA, Moritz C. 2013. Data from: Montane refugia predict population genetic structure in the Large-blotched Ensatina salamander. In: Dryad Data Repository.
Devitt TJ, Devitt SE, Hollingsworth BD, McGuire JA, Moritz C. 2013. Montane refugia predict population genetic structure in the Large-blotched Ensatina salamander. Mol. Ecol. 22:1650-1665.

Dornburg A, Moore JA, Webster R, Warren DL, Brandley MC, Iglesias TL, Wainwright PC, Near TJ. 2012a. Data from: Molecular phylogenetics of squirrelfishes and soldierfishes (Teleostei: Beryciformes: Holocentridae): reconciling more than 100 years of taxonomic confusion. In: Dryad Data Repository.

Dornburg A, Moore JA, Webster R, Warren DL, Brandley MC, Iglesias TL, Wainwright PC, Near TJ. 2012b. Molecular phylogenetics of squirrelfishes and soldierfishes (Teleostei: Beryciformes: Holocentridae): reconciling more than 100 years of taxonomic confusion. Mol Phylogenet Evol 65:727-738.

Duchene DA, Duchene S, Ho SYW. 2017. New Statistical Criteria Detect Phylogenetic Bias Caused by Compositional Heterogeneity. Mol. Biol. Evol. 34:1529-1534.

Dutheil JY, Galtier N, Romiguier J, Douzery EJ, Ranwez V, Boussau B. 2012. Efficient selection of branch-specific models of sequence evolution. Mol. Biol. Evol. 29:1861-1874.

Estabrook GF, McMorris F, Meacham CA. 1985. Comparison of undirected phylogenetic trees based on subtrees of four evolutionary units. Systematic Zoology 34:193-200.

Faircloth BC, Sorenson L, Santini F, Alfaro ME. 2013a. Data from: A phylogenomic perspective on the radiation of ray-finned fishes based upon targeted sequencing of ultraconserved elements (UCEs). In: Dryad Data Repository.

Faircloth BC, Sorenson L, Santini F, Alfaro ME. 2013b. A Phylogenomic Perspective on the Radiation of Ray-Finned Fishes Based upon Targeted Sequencing of Ultraconserved Elements (UCEs). PLoS One 8:e65923.

Fan L, Wu D, Goremykin V, Xiao J, Xu Y, Garg S, Zhang C, Martin WF, Zhu R. 2020. Phylogenetic analyses with systematic taxon sampling show that mitochondria branch within Alphaproteobacteria. Nat Ecol Evol 4:1213-1219.

Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17:368-376.

Felsenstein J. 2004. Inferring Phylogenies. Sunderland, Massachusetts: Sinauer Associates, Inc.

Felsenstein J. 1983. Statistical inference of phylogenies. Journal of the Royal Statistical Society: Series A (General) 146:246-262.

Fong JJ, Brown JM, Fujita MK, Boussau B. 2012a. Data from: A phylogenomic approach to vertebrate phylogeny supports a turtle-archosaur affinity and a possible paraphyletic Lissamphibia. In: Dryad Data Repository.
THE INFLUENCE OF MODEL VIOLATION ON PHYLOGENETIC INFERENCE

Fong JJ, Brown JM, Fujita MK, Boussau B. 2012b. A phylogenomic approach to vertebrate phylogeny supports a turtle-archosaur affinity and a possible paraplyletic lissamphibia. PLoS One 7:e48990.

Foster PG. 2004. Modeling compositional heterogeneity. Syst. Biol. 53:485-495.

Foster PG, Hickey DA. 1999. Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. J. Mol. Evol. 48:284-290.

Galtier N, Gouy M. 1998. Inferring pattern and process: maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. Mol. Biol. Evol. 15:871-879.

Galtier N, Gouy M. 1995. Inferring phylogenies from DNA sequences of unequal base compositions. Proc Natl Acad Sci U S A 92:11317-11321.

Groussin M, Boussau B, Gouy M. 2013. A branch-heterogeneous model of protein evolution for efficient inference of ancestral sequences. Syst. Biol. 62:523-538.

Hasegawa M, Kishino H, Yano T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22:160-174.

Hillis DM, Huelsenbeck JP, Cunningham CW. 1994. Application and accuracy of molecular phylogenies. Science 264:671-677.

Ho JW, Adams CE, Lew JB, Matthews TJ, Ng CC, Shahabi-Sirjani A, Tan LH, Zhao Y, Easteal S, Wilson SR, et al. 2006. SeqVis: visualization of compositional heterogeneity in large alignments of nucleotides. Bioinformatics 22:2162-2163.

Ho SY, Jermiin L. 2004. Tracing the decay of the historical signal in biological sequence data. Syst. Biol. 53:623-637.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol. Biol. Evol. 35:518-522.

Horn JW, Xi Z, Riina R, Peirson JA, Yang Y, Dorsey BL, Berry PE, Davis CC, Wurdack KJ. 2014a. Data from: Evolutionary bursts in Euphorbia (Euphorbiaceae) are linked with photosynthetic pathway. In: Dryad Data Repository.

Horn JW, Xi Z, Riina R, Peirson JA, Yang Y, Dorsey BL, Berry PE, Davis CC, Wurdack KJ. 2014b. Evolutionary bursts in Euphorbia (Euphorbiaceae) are linked with photosynthetic pathway. Evolution 68:3485-3504.

Hua X, Lanfear R. 2018. The influence of non-random species sampling on macroevolutionary and macroecological inference from phylogenies. Methods in Ecology and Evolution 9:1353-1362.

Huelsenbeck JP, Hillis DM. 1993. Success of Phylogenetic Methods in the Four-Taxon Case. Syst. Biol. 42:247-264.
Huerta-Cepas J, Serra F, Bork P. 2016. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. Mol. Biol. Evol. 33:1635-1638.

Irisarri I, Baurain D, Brinkmann H, Delsuc F, Sire J, Kupfer A, Petersen J, Jarek M, Meyer A, Vences M, et al. 2017a. Data from: Phylotranscriptomic consolidation of the jawed vertebrate timetree. In: Dryad Digital Repository.

Irisarri I, Baurain D, Brinkmann H, Delsuc F, Sire JY, Kupfer A, Petersen J, Jarek M, Meyer A, Vences M, et al. 2017b. Phylotranscriptomic consolidation of the jawed vertebrate timetree. Nat Ecol Evol 1:1370-1378.

Jarvis ED, Mirarab S, Aberer A, Houde P, Li C, Ho S, Faircloth BC, Nabholz B, Howard JT, Suh A, et al. 2014. Data from: Phylogenomic analyses data of the avian phylogenomics project. In: GigaScience Database.

Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SY, Faircloth BC, Nabholz B, Howard JT, et al. 2015. Phylogenomic analyses data of the avian phylogenomics project. Gigascience 4:4.

Jayaswal V, Ababneh F, Jermiin LS, Robinson J. 2011. Reducing model complexity of the general Markov model of evolution. Mol. Biol. Evol. 28:3045-3059.

Jayaswal V, Wong TK, Robinson J, Poladian L, Jermiin LS. 2014. Mixture models of nucleotide sequence evolution that account for heterogeneity in the substitution process across sites and across lineages. Syst. Biol. 63:726-742.

Jermiin L, Ho SY, Ababneh F, Robinson J, Larkum AW. 2004. The biasing effect of compositional heterogeneity on phylogenetic estimates may be underestimated. Syst. Biol. 53:638-643.

Jermiin LS, Catullo RA, Holland BR. 2020. A new phylogenetic protocol: dealing with model misspecification and confirmation bias in molecular phylogenetics. NAR Genom Bioinform 2:1qaa041.

Jermiin LS, Ho JWK, Lau KW, Jayaswal V. 2009. SeqVis: a tool for detecting compositional heterogeneity among aligned nucleotide sequences. In. Bioinformatics for DNA sequence analysis: Springer. p. 65-91.

Jermiin LS, Lovell DR, Misof B, Foster PG, Robinson J. 2019. Software for Detecting Heterogeneous Evolutionary Processes across Aligned Sequence Data. bioRxiv:828996.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14:587-589.

Kawahara AY, Rubinoff D. 2013a. Convergent evolution of morphology and habitat use in the explosive Hawaiian fancy case caterpillar radiation. J. Evol. Biol. 26:1763-1773.
Kawahara AY, Rubinoff D. 2013b. Data from: Convergent evolution in the explosive Hawaiian Fancy Cased caterpillar radiation. In: Dryad Data Repository.

Kimura M. 1980. A Simple Method for Estimating Evolutionary Rates of Base Substitutions through Comparative Studies of Nucleotide-Sequences. J. Mol. Evol. 16:111-120.

Kumar S, Filipski AJ, Battistuzzi FU, Kosakovsky Pond SL, Tamura K. 2012. Statistics and truth in phylogenomics. Mol. Biol. Evol. 29:457-472.

Kumar S, Gadagkar SR. 2001. Disparity index: a simple statistic to measure and test the homogeneity of substitution patterns between molecular sequences. Genetics 158:1321-1327.

Lanave C, Preparata G, Saccone C, Serio G. 1984. A new method for calculating evolutionary substitution rates. J. Mol. Evol. 20:86-93.

Lanave C, Tommasi S, Preparata G, Saccone C. 1986. Transition and transversion rate in the evolution of animal mitochondrial DNA. BioSyst. 19:273-283.

Lartillot N, Delsuc F. 2012a. Data from: Joint reconstruction of divergence times and life-history evolution in placental mammals using a phylogenetic covariance model. In: Dryad Data Repository.

Lartillot N, Delsuc F. 2012b. Joint reconstruction of divergence times and life-history evolution in placental mammals using a phylogenetic covariance model. Evolution 66:1773-1787.

Leaché AD, Chavez AS, Jones LN, Grummer JA, Gottscho AD, Linkem CW. 2015. Phylogenomics of phrynosomatid lizards: conflicting signals from sequence capture versus restriction site associated DNA sequencing. Genome Biol Evol 7:706-719.

Liu ZQ, Liu YF, Kuermanali N, Wang DF, Chen SJ, Guo HL, Zhao L, Wang JW, Han T, Wang YZ, et al. 2018. Sequencing of complete mitochondrial genomes confirms synonymization of Hyalomma asiaticum asiaticum and kozlovi, and advances phylogenetic hypotheses for the Ixodidae. PLoS One 13:e0197524.

Lockhart PJ, Steel MA, Hendy MD, Penny D. 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. Mol. Biol. Evol. 11:605-612.

Looney BP, Ryberg M, Hampe F, Sanchez-Garcia M, Matheny PB. 2016. Into and out of the tropics: global diversification patterns in a hyperdiverse clade of ectomycorrhizal fungi. Mol. Ecol. 25:630-647.
Looney BP, Ryberg M, Hampe F, Sánchez-García M, Matheny PB. 2015. Data from: Into and out of the tropics: global diversification patterns in a hyper-diverse clade of ectomycorrhizal fungi. In: Dryad.

Magallon S, Sanderson MJ. 2001. Absolute diversification rates in angiosperm clades. Evolution 55:1762-1780.

Martijn J, Vosseberg J, Guy L, Offre P, Ettema TJ. 2018. Deep mitochondrial origin outside the sampled alphaproteobacteria. Nature.

McCormack JE, Harvey MG, Faircloth BC, Crawford NG, Glenn TC, Brumfield RT. 2013a. Data from: A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. In: Dryad Data Repository.

McCormack JE, Harvey MG, Faircloth BC, Crawford NG, Glenn TC, Brumfield RT. 2013b. A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. PLoS One 8:e54848.

McKinney W. 2010. Data Structures for Statistical Computing in Python.

Meiklejohn KA, Faircloth BC, Glenn TC, Kimball RT, Braun EL. 2016a. Analysis of a Rapid Evolutionary Radiation Using Ultraconserved Elements: Evidence for a Bias in Some Multispecies Coalescent Methods. Syst. Biol. 65:612-627.

Meiklejohn KA, Faircloth BC, Glenn TC, Kimball RT, Braun EL. 2016b. Data from: Analysis of a rapid evolutionary radiation using ultraconserved elements (UCEs): Evidence for a bias in some multi-species coalescent methods. In: Dryad.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol. Biol. Evol. 37:1530-1534.

Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, Frandsen PB, Ware J, Flouri T, Beutel RG, et al. 2014a. Data from: Phylogenomics resolves the timing and pattern of insect evolution. In: Dryad Digital Repository.

Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, Frandsen PB, Ware J, Flouri T, Beutel RG, et al. 2014b. Phylogenomics resolves the timing and pattern of insect evolution. Science 346:763-767.

Mooers AO, Holmes EC. 2000. The evolution of base composition and phylogenetic inference. Trends Ecol. Evol. 15:365-369.

Moyle RG, Oliveros CH, Andersen MJ, Hosner PA, Benz BW, Manthey JD, Travers SL, Brown RM, Faircloth BC. 2016a. Data from: Tectonic collision and uplift of Wallacea triggered the global songbird radiation. In: Dryad Data Repository.
THE INFLUENCE OF MODEL VIOLATION ON PHYLOGENETIC INFERENCE

Moyle RG, Oliveros CH, Andersen MJ, Hosner PA, Benz BW, Manthey JD, Travers SL, Brown RM, Faircloth BC. 2016b. Tectonic collision and uplift of Wallacea triggered the global songbird radiation. Nat Commun 7:12709.

Murray EA, Carmichael AE, Heraty JM. 2013a. Ancient host shifts followed by host conservatism in a group of ant parasitoids. Proc Biol Sci 280:20130495.

Murray EA, Carmichael AE, Heraty JM. 2013b. Data from: Ancient host shifts followed by host conservatism in a group of ant parasitoids. In: Dryad Data Repository.

Naser-Khdour S, Minh BQ, Zhang W, Stone EA, Lanfear R. 2019. The Prevalence and Impact of Model Violations in Phylogenetic Analysis. Genome Biol Evol.

Near TJ, Dornburg A, Eytan RI, Keck BP, Smith WL, Kuhn KL, Moore JA, Price SA, Burbirk FT, Friedman M, et al. 2013a. Data from: Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. In: Dryad.

Near TJ, Dornburg A, Eytan RI, Keck BP, Smith WL, Kuhn KL, Moore JA, Price SA, Burbirk FT, Friedman M, et al. 2013b. Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. Proc Natl Acad Sci U S A 110:12738-12743.

Nee S, May RM, Harvey PH. 1994. The reconstructed evolutionary process. Philos Trans R Soc Lond B Biol Sci 344:305-311.

Nguyen AD, Gotelli NJ, Cahan SH. 2016a. Data from: The evolution of heat shock protein sequences, cis-regulatory elements, and expression profiles in the eusocial Hymenoptera. In: Dryad.

Nguyen AD, Gotelli NJ, Cahan SH. 2016b. The evolution of heat shock protein sequences, cis-regulatory elements, and expression profiles in the eusocial Hymenoptera. BMC Evol. Biol. 16:15.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32:268-274.

Oaks JR. 2011a. Data from: A time-calibrated species tree of Crocodylia reveals a recent radiation of the true crocodiles. In: Dryad Data Repository.

Oaks JR. 2011b. A time-calibrated species tree of Crocodylia reveals a recent radiation of the true crocodiles. Evolution 65:3285-3297.

Phillips MJ, Delsuc F, Penny D. 2004. Genome-scale phylogeny and the detection of systematic biases. Mol. Biol. Evol. 21:1455-1458.

Phillips MJ, Penny D. 2003. The root of the mammalian tree inferred from whole mitochondrial genomes. Mol Phylogenet Evol 28:171-185.
Prebus M. 2017a. Data from: Insights into the evolution, biogeography and natural
history of the acorn ants, genus Temnothorax Mayr (Hymenoptera: Formicidae).
In: Dryad.
Prebus M. 2017b. Insights into the evolution, biogeography and natural history of the
acorn ants, genus Temnothorax Mayr (hymenoptera: Formicidae). BMC Evol.
Biol. 17:250.
Preparata G, Saccone C. 1987. A simple quantitative model of the molecular clock. J.
Mol. Evol. 26:7-15.
Puttick MN, Morris JL, Williams TA, Cox CJ, Edwards D, Kenrick P, Pressel S,
Wellman CH, Schneider H, Pisani D, et al. 2018. The Interrelationships of Land
Plants and the Nature of the Ancestral Embryophyte. Curr. Biol. 28:733-745
Pyron RA, Wiens JJ. 2011. A large-scale phylogeny of Amphibia including over 2800
species, and a revised classification of extant frogs, salamanders, and caecilians.
Mol Phylogenetic Evol 61:543-583.
Pyron RA, Wiens JJ, Alexander Pyron R. 2011. Data from: A large-scale phylogeny of
Amphibia including over 2800 species, and a revised classification of extant
frogs, salamanders, and caecilians. In: Dryad.
Ran JH, Shen TT, Wang MM, Wang XQ. 2018a. Data from: Phylogenomics resolves the
deep phylogeny of seed plants and indicates partial convergent or homoplasmic
evolution between Gnetales and angiosperms. In: Dryad Digital Repository.
Ran JH, Shen TT, Wang MM, Wang XQ. 2018b. Phylogenomics resolves the deep
phylogeny of seed plants and indicates partial convergent or homoplasmic
evolution between Gnetales and angiosperms. Proc Biol Sci 285:20181012.
Rannala B, Yang Z. 1996. Probability distribution of molecular evolutionary trees: A
new method of phylogenetic inference. J. Mol. Evol. 43:304-311.
Reddy S, Kimball RT, Pandey A, Hosner PA, Braun MJ, Hackett SJ, Han K, Harshman
J, Huddleston CJ, Kingston S, et al. 2017a. Data from: Why do phylogenomic
data sets yield conflicting trees? Data type influences the avian tree of life more
than taxon sampling. In: Dryad Digital Repository.
Reddy S, Kimball RT, Pandey A, Hosner PA, Braun MJ, Hackett SJ, Han KL, Harshman
J, Huddleston CJ, Kingston S, et al. 2017b. Why Do Phylogenomic Data Sets
Yield Conflicting Trees? Data Type Influences the Avian Tree of Life more than
Taxon Sampling. Syst. Biol. 66:857-879.
Richart CH, Hayashi CY, Hedin M. 2016a. Data from: Phylogenomic analyses resolve
an ancient trichotomy at the base of Ischyropsalidoidea (Arachnida, Opiliones)
THE INFLUENCE OF MODEL VIOLATION ON PHYLOGENETIC INFERENCE

despite high levels of gene tree conflict and unequal minority resolution frequencies. In: Dryad.

Richart CH, Hayashi CY, Hedin M. 2016b. Phylogenomic analyses resolve an ancient trichotomy at the base of Ischyropsalidoidea (Arachnida, Opiliones) despite high levels of gene tree conflict and unequal minority resolution frequencies. Mol Phylogenet Evol 95:171-182.

Richart MG, Griswold T, Brady SG. 2013a. Data from: Phylogeny and systematics of the bee genus Osmia (Hymenoptera: Megachilidae) with emphasis on North American Melanosmia: subgenera, synonymies, and nesting biology revisited. In: Dryad Data Repository.

Richart MG, Griswold T, Brady SG. 2013b. Phylogeny and systematics of the bee genus Osmia (Hymenoptera: Megachilidae) with emphasis on North American Melanosmia: subgenera, synonymies and nesting biology revisited. Syst. Entomol. 38:561-576.

Roberts D, Yang Z. 1995. On the use of nucleic acid sequences to infer early branchings in the tree of life. Mol. Biol. Evol. 12:451-458.

Robinson DF, Foulds LR. 1981. Comparison of Phylogenetic Trees. Math. Biosci. 53:131-147.

Sauquet H, Ho SY, Gandolfo MA, Jordan GJ, Wilf P, Cantrill DJ, Bayly MJ, Bromham L, Brown GK, Carpenter RJ, et al. 2012. Testing the impact of calibration on molecular divergence times using a fossil-rich group: the case of Nothofagus (Fagales). Syst. Biol. 61:289-313.

Sauquet H, Ho SYW, Gandolfo MA, Jordan GJ, Wilf P, Cantrill DJ, Bayly MJ, Bromham L, Brown GK, Carpenter RJ, et al. 2011. Data from: Testing the impact of calibration on molecular divergence times using a fossil-rich group: the case of Nothofagus (Fagales). In: Dryad Data Repository.

Seago AE, Giorgi JA, Li J, Ślipiński A. 2011a. Data from: Phylogeny, classification and evolution of ladybird beetles (Coleoptera: Coccinellidae) based on simultaneous analysis of molecular and morphological data. In: Dryad Data Repository.

Seago AE, Giorgi JA, Li J, Ślipiński A. 2011b. Phylogeny, classification and evolution of ladybird beetles (Coleoptera: Coccinellidae) based on simultaneous analysis of molecular and morphological data. Mol. Phylogen. Evol. 60:137-151.

Sharanskiowski BJ, Dowling APG, Sharkey MJ. 2011a. Data from: Molecular phylogenetics of Braconidae (Hymenoptera: Ichneumonoidea) based on multiple nuclear genes and implications for classification. In: Dryad Data Repository.
Sharanowski BJ, Dowling APG, Sharkey MJ. 2011b. Molecular phylogenetics of Braconidae (Hymenoptera: Ichneumonoidea), based on multiple nuclear genes, and implications for classification. Syst. Entomol. 36:549-572.

Shen X. 2018. Data from: Tempo and mode of genome evolution in the budding yeast subphylum. In: Figshare.

Shen XX, Opulente DA, Kominek J, Zhou X, Steenwyk JL, Buh KV, Haase MAB, Wisecaver JH, Wang M, Doering DT, et al. 2018. Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. Cell 175:1533-1545 e1520.

Siler C, Brown RM, Oliveros CH, Santanen A. 2013. Data from: Multilocus phylogeny reveals unexpected diversification patterns in Asian Wolf Snakes (genus Lycodon). In: Dryad Data Repository.

Siler CD, Oliveros CH, Santanen A, Brown RM. 2013. Multilocus phylogeny reveals unexpected diversification patterns in Asian wolf snakes (genus Lycodon). Zool. Scr. 42:262-277.

Smith BT, Harvey MG, Faircloth BC, Glenn TC, Brumfield RT. 2014a. Data from: Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. In: Dryad Digital Repository.

Smith BT, Harvey MG, Faircloth BC, Glenn TC, Brumfield RT. 2014b. Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. Syst. Biol. 63:83-95.

Song N, Lin A, Zhao X. 2018. Insight into higher-level phylogeny of Neuropterida: Evidence from secondary structures of mitochondrial rRNA genes and mitogenomic data. PLoS One 13:e0191826.

Stadler T. 2013. How can we improve accuracy of macroevolutionary rate estimates? Syst. Biol. 62:321-329.

Stadler T. 2011. Simulating Trees with a Fixed Number of Extant Species. Syst. Biol. 60:676-684.

Steel MA, Penny D. 1993. Distributions of Tree Comparison Metrics - Some New Results. Syst. Biol. 42:126-141.

Swoford DL, Olsen GJ, Waddell PJ, Hillis DM. 1996. Phylogenetic Inference. In. Molecular systematics: Sunderland, Mass.: Sinauer Associates. p. 407-514.

Tamura K, Kumar S. 2002. Evolutionary distance estimation under heterogeneous substitution pattern among lineages. Mol. Biol. Evol. 19:1727-1736.

Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10:512-526.
THE INFLUENCE OF MODEL VIOLATION ON PHYLOGENETIC INFERENCE

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

Nguyen MAT, Gesell T, von Haeseler A. 2012. ImOSM: intermittent evolution and robustness of phylogenetic methods. Mol. Biol. Evol. 29:663-673.

Tolley KA, Townsend TM, Vences M. 2013a. Data from: Large-scale phylogeny of chameleons suggests African origins and Eocene diversification. In: Dryad Data Repository.

Tolley KA, Townsend TM, Vences M. 2013b. Large-scale phylogeny of chameleons suggests African origins and Eocene diversification. Proc Biol Sci 280:20130184.

Unmack PJ, Allen GR, Johnson JB. 2013a. Data from: Phylogeny and biogeography of rainbowfishes (Melanotaeniidae) from Australia and New Guinea. In: Dryad Data Repository.

Unmack PJ, Allen GR, Johnson JB. 2013b. Phylogeny and biogeography of rainbowfishes (Melanotaeniidae) from Australia and New Guinea. Mol Phylogenet Evol 67:15-27.

Varga T, Krizsán K, Földi C, Dima B, Sánchez-García M, Sánchez-Ramírez S, Szöllősi GJ, Szarkándi JG, Papp V, Albert L, et al. 2019a. Data from: Megaphylogeny resolves global patterns of mushroom evolution. In: Dryad.

Varga T, Krizsán K, Földi C, Dima B, Sánchez-García M, Sánchez-Ramírez S, Szöllősi GJ, Szarkándi JG, Papp V, Albert L, et al. 2019b. Megaphylogeny resolves global patterns of mushroom evolution. Nat Ecol Evol 3:668-678.

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, et al. 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods 17:261-272.

von Haeseler A, Janke A, Pääbo S. 1993. Molecular phylogenetics. Verhandlungen der Deutschen Zoologischen Gesellschaft= Proceedings of the German Zoological Society 86:119-129.

Wainwright PC, Smith WL, Price SA, Tang KL, Sparks JS, Ferry LA, Kuhn KL, Near TJ. 2012. Data from: The evolution of pharyngognathy: a phylogenetic and functional appraisal of the pharyngeal jaw key innovation in labroid fishes and beyond. In: Dryad Data Repository.

Wainwright PC, Smith WL, Price SA, Tang KL, Sparks JS, Ferry LA, Kuhn KL, Near TJ, Eytan RI. 2012. The evolution of pharyngognathy: a phylogenetic and functional appraisal of the pharyngeal jaw key innovation in labroid fishes and beyond. Syst. Biol. 61:1001-1027.
Walt Svd, Colbert SC, Varoquaux G. 2011. The NumPy Array: A Structure for Efficient Numerical Computation. Computing in Science & Engineering 13:22-30.

Weiss G, von Haeseler A. 2003. Testing Substitution Models Within a Phylogenetic Tree. Mol. Biol. Evol. 20:572-578.

Welch BL. 1947. The generalization of ‘STUDENT'S’ problem when several different population variances are involved. Biometrika 34:28-35.

Whelan NV, Kocot KM, Moroz TP, Mukherjee K, Williams P, Paulay G, Moroz LL, Halanych KM. 2017a. Author Correction: Ctenophore relationships and their placement as the sister group to all other animals. Nat Ecol Evol 1:1783.

Whelan NV, Kocot KM, Moroz TP, Mukherjee K, Williams P, Paulay G, Moroz LL, Halanych KM. 2017b. Data from: Ctenophora Phylogeny Datasets and Core Orthologs. In: Figshare.

Wood HM, Matzke NJ, Gillespie RG, Griswold CE. 2012. Data from: Treating fossils as terminal taxa in divergence time estimation reveals ancient vicariance patterns in the palpimanoid spiders. In: Dryad Data Repository.

Wood HM, Matzke NJ, Gillespie RG, Griswold CE. 2013. Treating fossils as terminal taxa in divergence time estimation reveals ancient vicariance patterns in the palpimanoid spiders. Syst. Biol. 62:264-284.

Worobey M, Han G, Rambaut A. 2014a. Data from: A synchronized global sweep of the internal genes of modern avian influenza virus. In: Dryad Data Repository.

Worobey M, Han GZ, Rambaut A. 2014b. A synchronized global sweep of the internal genes of modern avian influenza virus. Nature 508:254-257.

Wu S, Edwards S, Liu L. 2019. Data from: Genome-scale DNA sequence data and the evolutionary history of placental mammals. In: Figshare.

Wu S, Edwards S, Liu L. 2018. Genome-scale DNA sequence data and the evolutionary history of placental mammals. Data Brief 18:1972-1975.

Yang Z. 2006. Computational Molecular Evolution. Oxford, UNITED KINGDOM: Oxford University Press USA - OSO.

Yang Z. 1994. Estimating the pattern of nucleotide substitution. J. Mol. Evol. 39:105-111.

Yang Z. 1995. A space-time process model for the evolution of DNA sequences. Genetics 139:993.

Zou L, Susko E, Field C, Roger AJ. 2012. Fitting nonstationary general-time-reversible models to obtain edge-lengths and frequencies for the Barry–Hartigan model. Syst. Biol. 61:927-940.