The Prevalence and Impact of Model Violations in Phylogenetics Analysis

Suha Naser-Khdour*, Bui Quang Minh, Wenqi Zhang, Eric Stone, Robert Lanfear

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

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

Abstract

In phylogenetic inference we commonly use models which assume that sequence evolution is stationary, reversible and homogeneous (SRH). Although such assumptions are often criticized, the extent of SRH violations and their effects on phylogenetic inference are not well understood. Here, we extend the matched-pairs test of symmetry to assess the scale and impact of SRH model violations on 3,572 partitions from 35 published phylogenetic datasets. We show that many partitions (39.5%) reject the SRH assumptions, and that in most datasets, phylogenies inferred from all partitions differ significantly from those inferred using the subset of partitions that do not reject the SRH assumptions. These results suggest that the extent and effects of model violation in phylogenetics may be substantial. They highlight the importance of testing for model violations and possibly excluding partitions that violate models prior to tree reconstruction. They also suggest that further effort in developing models that do not require SRH assumptions could lead to large improvements in the accuracy of phylogenomic inference. The scripts necessary to perform the analysis are available in https://github.com/roblanf/SRHtests, and the new tests we describe are available as a new option in IQ-TREE (http://www.iqtree.org).

Keywords: model violations, phylogenetic inference, test of symmetry, systematic bias

Introduction

Phylogenetics is an essential tool for inferring evolutionary relationships between individuals, species, genes, and genomes. Moreover, phylogenetic trees form the basis of a huge range of other inferences in evolutionary biology, from gene function prediction to drug development and forensics (Eisen 1998; Farrell, et al. 2000; Mäser, et al. 2001; Gardner, et al. 2002; Yao, et al. 2003; Grenfell, et al. 2004; Yao, et al. 2004; Salipante and Horwitz 2006; Gray, et al. 2009; Brady and Salzberg 2011; Dunn, et al. 2011).

Most phylogenetic studies use parametric models of molecular evolution which assume that the evolutionary process follows stationary, reversible and homogeneous (SRH) conditions. Stationarity implies that the marginal frequencies of the nucleotides or amino-acids are constant over time, reversibility implies that the evolutionary process is stationary and undirected, and homogeneity implies that the instantaneous substitution rates are constant along the tree or over an edge (Felsenstein 2004; Yang and Rannala 2012; Jermiin, et al. 2017). However, these simplifying assumptions are often violated by real data (Foster and Hickey 1999; Tarrío, et al. 2001; Paton, et al. 2002; Goremykin and Hellwig 2005; Murray, et al. 2005; Bourlat, et al. 2006; Hyman, et al. 2007; Sheffield, et al. 2009; Nesnidal, et al. 2010; Nabholz, et al. 2011; Martijn, et al. 2018). Such model violation may lead to systematic error that, unlike stochastic error, cannot be solved simply by increasing the size of a dataset (Felsenstein 2004; Ho and Jermiin 2004; Jermiin, et al. 2004; Philippe, et al. 2005; Sullivan and Joyce 2005; Kumar, et al. 2012; Brown and Thomson 2017; Duchene, et al. 2017). As phylogenetic datasets are steadily growing, it is vital that we develop and employ methods to measure and understand the extent to which systematic error affects phylogenetic inference (systematic bias), and explore ways of mitigating systematic bias in empirical studies.

One approach to accounting for data that have evolved under non-SRH conditions is to employ models that relax the SRH assumptions. A number of non-SRH models have been implemented in a variety of software packages (Foster 2004; Lartillot and Philippe 2004; Blanquart and Lartillot 2006; Boussau and Gouy 2006; Knight, et al. 2007; Dutheil and Boussau 2008; Sumner, et al. 2012; Zou, et al. 2012; Groussin, et al. 2013; Jayaswal, et al. 2014; Nguyen, et al. 2015). But despite the availability of these models they remain infrequently used, as searching for optimal phylogenetic trees under these models is computationally demanding (Betancur-r, et al. 2013) and the implementations are often not easy to use. As a result, the vast majority of empirical phylogenetic inferences rely on models which assume that sequences have evolved under SRH conditions, such as the general time reversible (GTR) family of models implemented in the most widely-used phylogenetics software packages (Swofford 2001; Drummond and Rambaut 2007; Guindon, et al. 2010; Ronquist, et al. 2012; Bazinet, et al. 2014; Bouckaert, et al. 2014; Stamatakis 2014; Nguyen, et al. 2015; Höhna, et al. 2016).

Another approach to accounting for data that may have evolved under non-SRH conditions is to test for model violations prior to tree reconstruction. Here, one first screens datasets or parts of datasets, and reconstruct trees exclusively from data that do not reject SRH conditions. A number of methods have been proposed to test for violation of SRH conditions in aligned sequences prior to estimating trees (Bowker 1948; Stuart 1955; Rzhetsky and Nei 1995; Kumar and Gadagkar 2001; Weiss and von Haeseler 2003; Ababneh, et al. 2006; Jermiin, et al. 2009), and there are also *a posteriori* tests for absolute model adequacy which are employed after trees have been estimated (Goldman 1993; Brown and ElDabaje 2009; Brown 2014; Duchene, et al. 2017; Brown and Thomson 2018).

Allowing the data to reject the model when the assumptions of the model are violated is an important approach to reducing systematic bias in phylogenetic inference (Philippe, et al.

2005; Brown 2014). Knowing in advance the sequences and loci that are inconsistent with the SRH assumptions will alert us to choose more complex models for phylogeny reconstruction or potentially to omit these loci from downstream analyses (Kumar and Gadagkar 2001). The need for methods that assess the evolutionary process prior to phylogenetic inference becomes more important as the number of sequences and sites per dataset increases, because systematic bias has an increasing effect on inferences from larger phylogenetic datasets (Ho and Jermiin 2004; Jermiin, et al. 2004; Phillips, et al. 2004; Delsuc, et al. 2005).

In this paper we evaluate the extent and effect of model violation due to non-SRH evolution using 35 empirical datasets with a total of 3,572 partitions. We determine if the SRH assumptions are violated by extending and applying the matched-pairs tests of symmetry (Jermiin, et al. 2017) to each partition. We then compare the phylogenetic trees for each dataset estimated from all of the partitions, the partitions that reject the SRH assumptions, and the partitions that do not reject the SRH assumptions, in order to evaluate the effect violating SRH conditions on phylogenetic inference.

Materials and Methods

Empirical datasets

In order to assess the impact of model violation in phylogenetics, we first gathered a representative sample of 35 partitioned empirical datasets that had been used for phylogenetic analysis in recent studies (Table 1). Within the constraints of selecting data that were publicly available and suitably annotated, we selected the datasets to provide as representative a sample as possible of the data types, taxa, and genomic regions most commonly used to infer bifurcating phylogenetic trees from concatenated alignments. These datasets include nucleotide sequences from nuclear, mitochondrial, plastid and virus genomes, and include protein-coding DNA, introns, intergenic spacers, tRNA, rRNA and ultra-conserved elements.

The number of taxa and sites in these datasets range from 27 to 355 and from 699 to 1,079,052 respectively. The clades represented in these datasets include animals, plants and viruses. All the datasets files can be found on Github in nexus format https://github.com/roblanf/SRHtests/tree/master/datasets, and the full details of every dataset are provided in Table 1.

Table 1| Number of taxa, number of sites, clade and study reference for each dataset that has been used in this study

Dataset	Study Reference	Dataset Reference	Clade	Taxa	Sites
Anderson_2013	(Anderson, et al. 2014)	(Anderson, et al. 2013)	loliginids	145	3037
Bergsten_2013	(Bergsten, et al. 2013a)	(Bergsten, et al. 2013b)	Dytiscidae	38	2111
Broughton_2013	(Broughton, et al. 2013b)	(Broughton, et al. 2013a)	Osteichthyes	61	19997
Brown_2012	(Brown, et al. 2012b)	(Brown, et al. 2012a)	Ptychozoon	41	1665
Cannon_2016a	(Cannon, et al. 2016a)	(Cannon, et al. 2016b)	Metazoa	78	89792
Cognato_2001	(Cognato and Vogler 2001b)	(Cognato and Vogler 2001a)	Coleoptera: Scolytinae	44	1897
Day_2013	(Day, Peart, Brown, Friel, et al. 2013)	(Day, Peart, Brown, Bills, et al. 2013)	Synodontis	152	3586
Devitt_2013	(Devitt, Devitt, et al. 2013)	(Devitt, Cameron Devitt, et al. 2013)	Ensatina eschscholtzii klauberi	69	823
Dornburg_2012	(Dornburg, et al. 2012b)	(Dornburg, et al. 2012a)	Teleostei: Beryciformes: Holocentridae	44	5919
Faircloth_2013	(Faircloth, et al. 2013b)	(Faircloth, et al. 2013a)	Actinopterygii	27	149366
Fong_2012	(Fong, et al. 2012b)	(Fong, et al. 2012a)	Vertebrata	110	25919
Horn_2014	(Horn, et al. 2014b)	(Horn, et al. 2014a)	Euphorbia	197	11587
Kawahara_2013	(Kawahara and Rubinoff 2013a)	(Kawahara and Rubinoff 2013b)	Hyposmocoma	70	2238
Lartillot_2012	(Lartillot and Delsuc 2012b)	(Lartillot and Delsuc 2012a)	Eutheria	78	15117
McCormack_2013	(McCormack, et al. 2013b)	(McCormack, et al. 2013a)	Neoaves	33	1079052
Moyle_2016	(Moyle, et al. 2016b)	(Moyle, et al. 2016a)	Oscines	106	375172
Murray_2013	(Murray, et al. 2013a)	(Murray, et al. 2013b)	Eucharitidae	237	3111
Oaks_2011	(Oaks 2011b)	(Oaks 2011a)	Crocodylia	79	7282
Rightmyer_2013	(Rightmyer, et al. 2013b)	(Rightmyer, et al. 2013a)	Hymenoptera: Megachilidae	94	3692
Sauquet_2011	(Sauquet, et al. 2012)	(Sauquet, et al. 2011)	Nothofagus	51	5444
Seago_2011	(Seago, et al. 2011b)	(Seago, et al. 2011a)	Coccinellidae	97	2253
Sharanowski_2011	(Sharanowski, et al. 2011b)	(Sharanowski, et al. 2011a)	Braconidae	139	3982
Siler_2013	(Siler, Oliveros, et al. 2013)	(Siler, Brown, et al. 2013)	Lycodon	61	2697
Tolley_2013	(Tolley, et al. 2013b)	(Tolley, et al. 2013a)	Chamaeleonidae	203	5054
Unmack_2013	(Unmack, et al. 2013b)	(Unmack, et al. 2013a)	Melanotaeniidae	139	6827
Wainwright_2012	Wainwright, Smith, Price, Tang, Sparks,	(Wainwright, Smith, Price, Tang, Sparks,	Acanthomorpha	188	8439

	Ferry, Kuhn, Eytan, et al. (2012)	Ferry, Kuhn and Near 2012)			
Wood_2012	(Wood, et al. 2013)	(Wood, et al. 2012)	Archaeidae	37	5185
Worobey_2014a	(Worobey, et al. 2014b)	(Worobey, et al. 2014a)	Influenzavirus A	146	3432
Worobey_2014b	(Worobey, et al. 2014b)	(Worobey, et al. 2014a)	Influenzavirus A	327	759
Worobey_2014c	(Worobey, et al. 2014b)	(Worobey, et al. 2014a)	Influenzavirus A	92	1416
Worobey_2014d	(Worobey, et al. 2014b)	(Worobey, et al. 2014a)	Influenzavirus A	355	1497
Worobey_2014e	(Worobey, et al. 2014b)	(Worobey, et al. 2014a)	Influenzavirus A	340	699
Worobey_2014f	(Worobey, et al. 2014b)	(Worobey, et al. 2014a)	Influenzavirus A	332	2151
Worobey_2014g	(Worobey, et al. 2014b)	(Worobey, et al. 2014a)	Influenzavirus A	326	2274
Worobey_2014h	(Worobey, et al. 2014b)	(Worobey, et al. 2014a)	Influenzavirus A	351	2280

Workflow summary

Figure 1 outlines the workflow. For each partition in each dataset, we used three matched-pairs tests of symmetry to ask whether the evolution of the aligned sequences rejects the SRH assumptions. The three tests, described in more detail below, comprise: the binomial matched-pairs test of symmetry (BiSymTest); the binomial matched-pairs test of marginal symmetry ($BiSymTest_{mar}$); and the binomial matched-pairs test of internal symmetry ($BiSymTest_{int}$). These approaches test three slightly different assumptions about the historical process that generated the aligned sequences in a given partition, and a significant result from any test suggests that the nature of the evolutionary process required to explain the aligned sequences violates at least one of the three SRH conditions (Jermiin, et al. 2017, and below). For each test, we classify each partition as *pass* if the result of the test is non-significant or *fail* if the result of the test is significant. We then denote the original dataset as D_{all}, while the concatenation of *pass* partitions is denoted D_{pass} and the concatenation of *fail* partitions as D_{fail} (fig. 1).

To investigate the impact of model violation on phylogenetic inference, we infer and compare three phylogenetic trees, T_{all}, T_{pass} and T_{fail}, estimated from D_{all}, D_{pass} and D_{fail}, respectively.

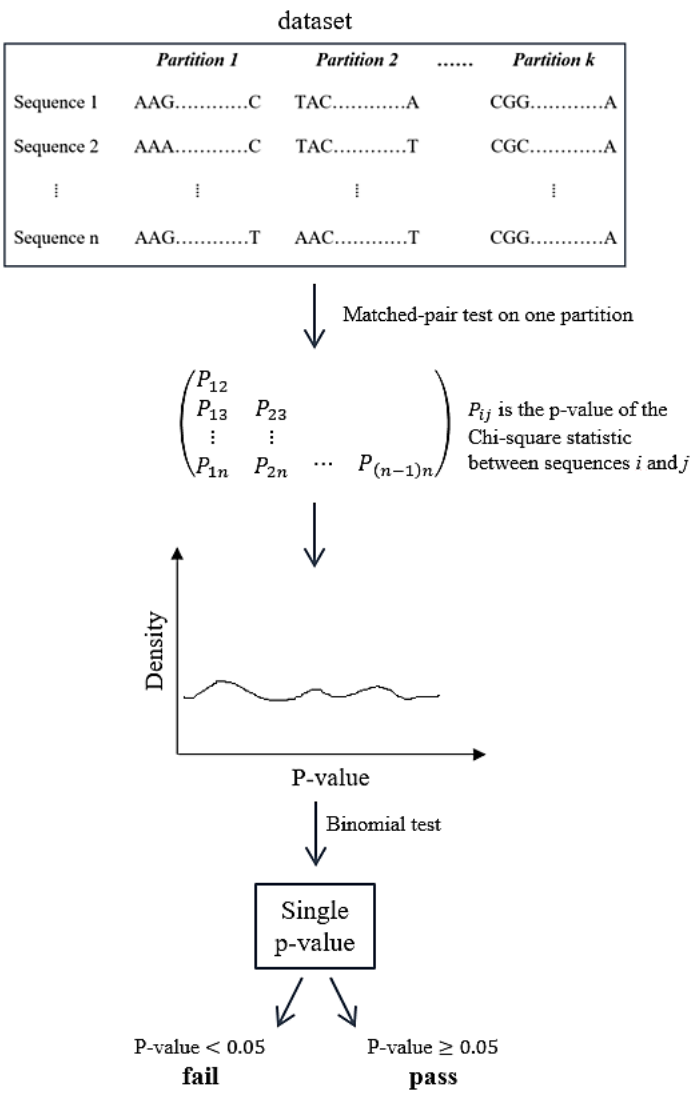

Fig. 1| Flow chart of methodology. After application of the matched-pairs tests of symmetry on each possible pair of sequences for each partition, we apply a binomial test on the p-values of all possible pairs of sequences of each partition to derive a P-value for that partition.

Matched-pairs tests of symmetry

The three binomial matched-pairs of symmetry that we describe below extend three tests of symmetry that are applied to pairs of sequences: the MPTS (matched-pairs test of symmetry), MPTMS (matched-pairs test of marginal symmetry), and MPTIS (matched-pairs test of internal symmetry). The MPTS, MPTMS and MPTIS test different aspects of the symmetry with which substitutions accumulate between pairs of sequences. The three test statistics are

asymptotically distributed as chi-squared under the assumption that the Markov processes operating on different branches are independent (Ababneh, et al. 2006). The statistics are computed on divergence matrices derived from a site-by-site comparison of two homologous sequences. Accordingly, using the chi-square test provides an indication of whether or not it is reasonable to assume that the pair of sequences has evolved under SRH conditions (Jermiin, et al. 2017).

The MPTS is a comprehensive and sufficient test to determine whether the data complies with the SRH assumptions (Jermiin, et al. 2017), but it cannot provide any information about the source of this violation. Some information on the underlying source of model violation may be obtained by performing two further tests of symmetry, the MPTMS and the MPTIS. If the violation of the SRH assumptions stems from differences in base composition between the sequences, this should affect the marginal symmetry of the sequence pair, which can in principle be detected by the MPTMS. While if the violation of the SRH assumptions stems from differences in substitution rates over time, this should affect the internal symmetry of the sequence pair, which can in principle be detected by the MPTIS. However, even after performing all three tests, it is difficult to ascertain which of the three SRH assumptions is violated during the evolutionary process because the relationships between the SRH conditions and the three matched-pair tests is neither bijective nor injective, i.e. there is no one-to-one correspondence between the three tests and violation of the three SRH conditions (Jermiin, et al. 2017).

Binomial matched-pairs tests of symmetry

The three matched-pairs tests of symmetry were designed to ask whether any single pair of sequences rejects the SRH conditions (Jermiin, et al. 2017). To ask whether a given partition rejects SRH conditions, we extend the matched-pairs tests of symmetry to accommodate

datasets with more than two sequences. To do so, we apply the matched-pair tests of symmetry to every pair of sequences in in a partition, resulting in $\binom{n}{2}$ chi-squared p-values for each partition, where n is the number of sequences in the partition (fig. 1). Under the null hypothesis of SRH evolution, the marginal distribution of each p-value should be uniform on the interval [0,1], suggesting a 5% chance that any one of them falls below 0.05. We use this logic to assess when too many p-values were too small, in which case a partition is deemed to fail the SRH conditions. Specifically, we count how many chi-square p-values are less than 0.05 and compare the result to a Binomial distribution with $\binom{n}{2}$ trials and success probability 0.05. If the corresponding Binomial p-value is smaller than 0.05, we classify the partition as *fail*; otherwise, the partition is labelled as *pass*. While this approach ignores the dependencies among p-values as obtained from pairwise sequence comparisons, a full accounting of the dependencies among p-values would require knowledge of the underlying phylogeny. Given that these tests aim to determine *a priori* whether it is feasible to build a reliable phylogeny, it would be inappropriate to use an estimated phylogeny as part of the test. Furthermore, even ignoring dependencies among p-values, if the evolutionary process underlying an alignment was generated by an evolutionary process that violates SRH conditions on many branches, we would still expect that alignment to fail the binomial test. To avoid confusion between the binomial and pairwise tests, we denote the binomial extensions of the MPTS, MPTMS, and the MTPIS as BiSymTest, $\text{BiSymTest}_{\text{mar}}$, and $\text{BiSymTest}_{\text{int}}$ respectively.

Phylogenetic inference

We used IQ-TREE (Nguyen, et al. 2015) to infer up to seven phylogenetic trees for every dataset: T_{all} (all partitions from the original dataset; D_{all}); and T_{pass} and T_{fail} based on the D_{pass} and D_{fail} datasets from each of the three tests (BiSymTest, $\text{BiSymTest}_{\text{mar}}$, $\text{BiSymTest}_{\text{int}}$), provided that there was at least one partition in each category. We ran IQ-TREE using the

default settings with the best-fit fully-partitioned model (Chernomor, et al. 2016) with edge-linked rates determined by ModelFinder (Kalyaanamoorthy, et al. 2017) followed 1000 ultrafast bootstrap replicates (Hoang, et al. 2018).

Distance between trees

For each of the three tests (MPTS, MPTMS, MPTIS) we calculated the Normalised Path-Difference (NPD) between all three possible pairs of trees (T_{all} vs. T_{pass}; T_{all} vs. T_{fail}; and T_{pass} vs. T_{fail}), as long as D_{pass} and D_{fail} were non-empty and so T_{pass} and T_{fail} had been estimated. The path-difference metric (PD) is defined as the Euclidean distance between pairs of taxa (Steel and Penny 1993; Mir and Russello 2010). In order to compare path distances between trees with different number of taxa, we normalised PD (to obtain NPD) by the mean of a null distribution of PDs generated from 10K random pairs of trees with the same number of taxa (Bogdanowicz, et al. 2012). Thus, an NPD of zero indicates an identical pair of trees, an NPD of 1 indicates that a pair of trees is as similar as a pair of randomly-selected trees with the same number of taxa; and an NPD greater than 1 indicates a pair of trees that are less similar than a randomly-selected pair of trees with the same number of taxa. Since path differences are always non-negative, the NPD is also guaranteed to be non-negative.

Tree topology tests

The NPD gives us a measure of the differences between pairs of trees, but it does not tell us whether the differences are phylogenetically significant differences underpinned by strong differences in signal in the three datatsets (D_{pass}, D_{all}, and D_{fail}) derived from a given test. For example, trees that differ due to stochastic error associated with small datasets may be very different, but such differences may not be statistically significant. To assess the significance of the differences between trees, we used the weighted Shimodaira-Hasegawa (wSH) test (Shimodaira and Hasegawa 1999; Shimodaira 2002) implemented in IQ-TREE with 1000

RELL replicates (Kishino, et al. 1990), between all three pairs of trees for each test (T_{all} vs. T_{pass}; T_{all} vs. T_{fail}; and T_{pass} vs. T_{fail}), as long as D_{pass} and D_{fail} were non-empty. The wSH test returns a p-value for a given set of trees and an alignment, and assesses whether the more likely tree given a particular alignment is a significantly better description of the data than the less likely trees. Thus, for each set of trees (e.g. T_{all}, T_{pass} and T_{fail}) we performed three wSH tests, one for each underlying alignment (D_{all}, D_{pass} and D_{fail}). In total, this resulted in up to 9 wSH tests for every partition and each of the three tests (MPTS, MPTMS, and MPTIS).

Comparing symmetry tests to random subsets of partitions

It is conceivable that simply dividing the partitions of any dataset into two groups would produce two different trees. In order to assess whether the division of the datasets into D_{pass} and D_{fail} based on the matched-pairs tests of symmetry led to differences between trees that are more extreme than those we might expect by chance, we generated a null distribution of wSH p-values for the comparison of D_{pass} and D_{fail} for each dataset as follows. For each of the 24 datasets with more than 20 possible combinations of partitions, we created 20 new divisions of the dataset into two groups of partitions, where the two groups contained the same number of partitions as the observed D_{pass} and D_{fail} subsets for that dataset. For each of these 20 replicates, we inferred a tree for both subsets of the data and compared these two trees using the wSH test as above. We then calculate a single wSH test using the replicate dataset with the same number of partitions as D_{fail} as the focal dataset. This gives us a null distribution of 20 wSH p-values for each of the 24 datasets with sufficient partitions to employ this approach. For each dataset, we then compared the observed wSH p-value from comparing T_{pass} and T_{fail} using to this null distribution. This allows us to ask whether dividing datasets into two groups based on the results of matched-pairs tests of symmetry leads to more extreme differences between phylogenetic trees than might be expected by chance.

Correlation between number of substitutions and model violation

We hypothesised that partitions with more substitutions may be more likely to violate the SRH assumptions, since substitutions form the raw data for the matched-pairs tests of symmetry. To assess this, we fitted a linear mixed-effects model for each of the three tests using the glmer function from the lme4 package in R (Bates, et al. 2015). In this model, we treat each partition as a datapoint, the number of substitutions measured for that partition as a fixed effect, and the dataset from which that partition was taken as a random effect. This allows us to estimate the extent to which the number of substitutions in a partition correlates with whether a partition fails a given test of symmetry, after accounting for differences between the datasets. To calculate the R squared value we use the r.squaredGLMM function from the MuMIn package in R (Barton 2009; Nakagawa and Schielzeth 2013).

Software implementation

We implemented a new option --bisymtest in IQ-TREE to perform the three matched pairs tests of symmetry and their associated binomial extensions. In addition, the option --bisymtest-remove-bad allows users to remove from the final analysis partitions that fail the BiSymTest. One can change the removal criterion to $\text{BySymTest}_{\text{mar}}$ or $\text{BySemTest}_{\text{int}}$ via the --bisymtest-type MAR|INT option. In addition, the cut-off p-value can be changed using the --bisymtest-pval NUM option, where the default value is 0.05.

The GitHub repository https://github.com/roblanf/SRHtests contains the raw data and Python and R scripts necessary to perform all analyses reported in this study.

Results

Violation of SRH conditions is common across 35 empirical datasets

Across all 3,572 partitions analysed, 724 (20.3%) failed the BiSymTest, 1309 (36.7%) failed the BiSymTest_{mar}, and 143 (4.0%) failed the BiSymTest_{int}. 1383 (39.4%) of partitions failed at least one test. The proportion of partitions failing each test varied substantially among datasets (fig. 2), but on average 39.5% of the partitions in each dataset failed the BiSymTest, 49.1% failed the BiSymTest_{mar}, and 8.5% failed the BiSymTest_{int}.

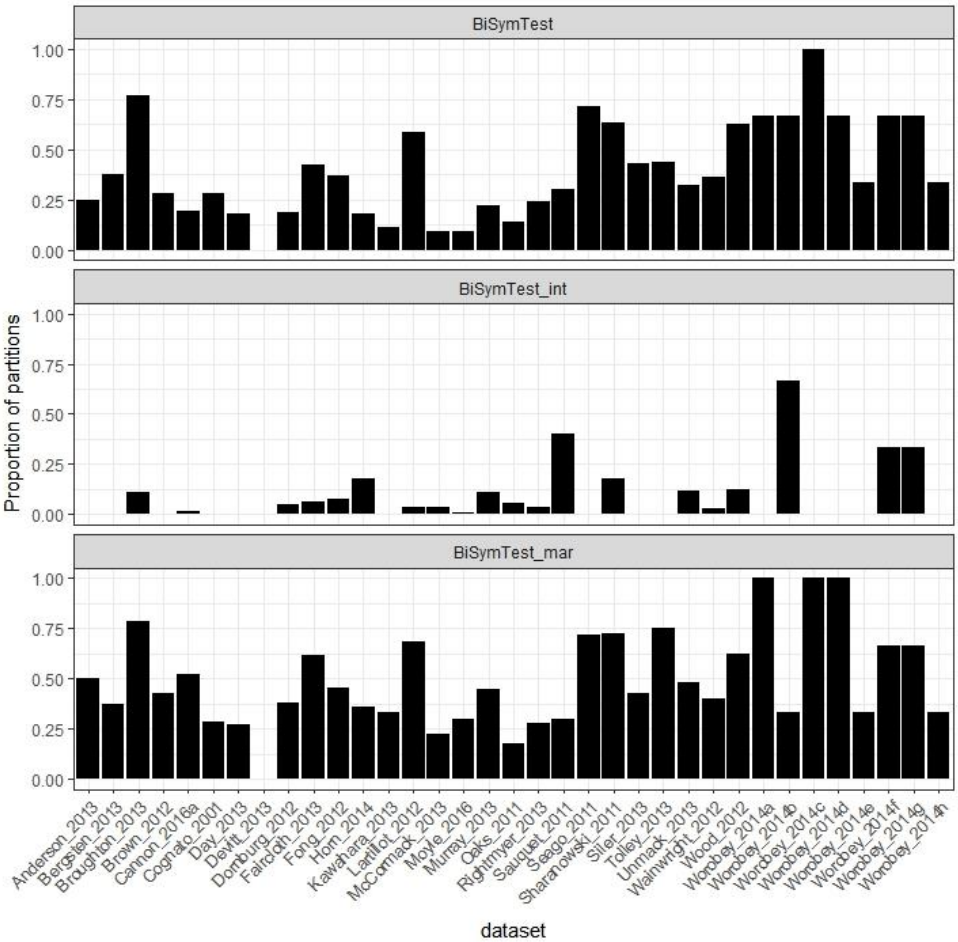

Fig. 2| The percentage of partitions that reject the null hypothesis of the BiSymTest, BiSymTest_{mar} and BiSymTest_{int} in each dataset. The number of partitions that fail the test in each dataset is the number of partitions that have a binomial p-value of less than 0.05.

The propensity of partitions to fail the matched-pairs tests of symmetry also varied with the type of genome (e.g. mitochondrial, chloroplast, or nuclear) and the type of locus (e.g. protein-coding, UCE, tRNA) from which the partition was sequenced (Table 2, Extended Data fig. 7), although we note that a substantial proportion of the partitions from almost every category failed at least one of the tests (Table 2).

Table 2| The proportion of partitions that failed at least one of the three binomial matched-pairs tests of symmetry (BiSymTest, BiSymTest$_{mar}$, BiSymTest$_{int}$)

Type / genome	nuclear	mitochondrial	plastid	virus
1st codon positions	49.1%	63.3%	33.3%	87.5%
2nd codon positions	42.6%	30.0%	33.3%	50.0%
3rd codon positions	85.0%	80.0%	33.3%	87.5%
Other (e.g. intron)	33.3%	100.0%	83.3%	
rRNA	60.0%	50.0%		
UCE	34.0%			
tRNA		0.0%		

Model violation has a strong influence on tree topologies

For the BiSymTest and the BiSymTest$_{mar}$, the tree estimated from the whole dataset, T_{all}, tended to be more similar to the tree estimated from the partitions that failed each test, T_{fail}, than to the tree estimated from the partitions that passed each test, T_{pass} (Table 3). Whereas T_{fail} and T_{pass} tended to be the most different pair of trees. The results of the wSH tests (Table 4) confirm that this pattern is driven by statistically significant differences between the three trees. For example, when using the BiSymTest, T_{fail} rejects T_{all} in ~25% of the datasets, and T_{pass} rejects T_{all} in >50% of the datasets, confirming that T_{all} is more similar to T_{fail} than to T_{pass} in terms of tree topology. Similarly, T_{fail} and T_{pass} reciprocally reject each other in almost all datasets, confirming that these two trees are significantly different. The results from the

$BiSymTest_{int}$ show a different pattern from the other two tests, such that T_{all} tends to be more similar to T_{pass} than to T_{fail} (Tables 3, 4)

Table 3| The proportion of datasets that have a significant p-value in the weighted SH test. Rows present the dataset used as an input and therefore the ML tree for the topology test

		T_{all}	T_{pass}	T_{fail}
BiSymTest	D_{all}		75%	58%
	D_{pass}	50%		87%
	D_{fail}	25%	100%	
$BiSymTest_{mar}$	D_{all}		83%	54%
	D_{pass}	54%		96%
	D_{fail}	21%	96%	
$BiSymTest_{int}$	D_{all}		30%	80%
	D_{pass}	20%		90%
	D_{fail}	60%	70%	

Table 4| The NPD metric mean and confidence interval for each test

		T_{fail}	T_{pass}
BiSymTest	T_{all}	0.40 (0.29,0.52)	0.67 (0.41,0.94)
	T_{pass}	0.82 (0.57,1.06)	
$BiSym\ Test_{mar}$	T_{all}	0.31 (0.22,0.40)	0.65 (0.40,0.90)
	T_{pass}	0.72 (0.48,0.96)	
$BiSym\ Test_{int}$	T_{all}	0.65 (0.12,1.18)	0.28 (0.11,0.45)
	T_{pass}	0.76 (0.26,1.26)	

Differences in tree topologies are frequently more severe than expected by chance

In total we were able to perform the randomisation test on 15 datasets. In 10 of these 15 datasets, the observed p-value was smaller than any of the replicate p-values suggesting that the difference between T_{pass} and T_{fail} measured by the wSH test was larger than would be expected by randomly dividing a dataset into two subsets of partitions.

The number of substitutions explains up to half of the variance in passing or failing the tests of symmetry

The number of substitutions in a partition explained 52% of the variation in whether or not a partition passed or failed the BiSymTest. This proportion is very similar for the $BiSymTest_{mar}$ (53%), but is dramatically lower for the $BiSymTest_{int}$ (0.5%). Accordingly, the number of substitutions in a partition is a highlight significant ($p<2e-16$) predictor of passing or failing the BiSymTest and the $BiSymTest_{mar}$, but is not a significant predictor of passing or failing the $BiSymTest_{int}$.

Model violation affects the internal relationships of Spiralia and the position of Xenacoelomorpha

To examine the effects of model violation in more detail, we selected a single dataset for more detailed consideration. Conflicting support for the placement of Xenacoelomorpha in the tree of life across different analyses led to various hypotheses regarding to the evolution of Bilateria (Cannon, et al. 2016a). It has been suggested that such inferences might be strongly affected by model violation and systematic error (Philippe, et al. 2011). To assess whether data that pass or fail the BiSymTest show different signals regarding the evolution of the Bilateria, we examined in more detail the T_{all}, T_{pass}, and T_{fail} trees from the dataset of Cannon et al. (2016a). This dataset comprises 76 metazoan taxa, 2 choanoflagellate outgroups, 212 genes and 424 partitions representing the first and second codon positions of the 212 genes (Cannon, et al. 2016b). The tree reconstructed from all of the partitions (T_{all}) is identical to the tree reconstructed from the partitions that pass the BiSymTest (T_{pass}), and both are identical to the tree shown in the original paper from both DNA and amino acid data (Canon, et al. 2016a), which places Xenacoelomorpha as the sister group of Nephrozoa (Deuterostomia and Protostomia) with 100% bootstrap support (fig. 3, Extended Data figs. 1-

2). The tree reconstructed from the data that fail the BiSymTest (T_{fail}) on the other hand (inferred from the 82 partitions) shows Xenoturbella as the sister group to Nephrozoa with 62% bootstrap support and Platyzoa as a sister of Acoelomorpha with 90% bootstrap support. Although the whole dataset – including the failed partitions – supports Xenacoelomorpha as a sister of Nephrozoa, it is clear from the results that the partitions that comprise D_{fail} support a very different set of relationships to those that comprise D_{pass}.

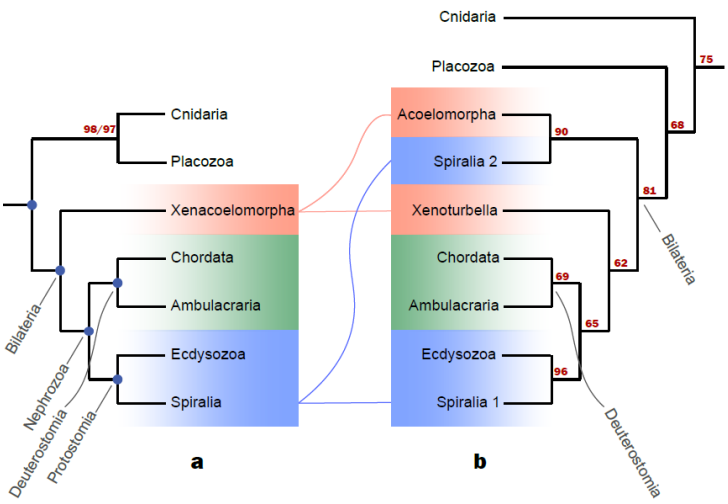

Fig. 3| Maximum likelihood trees of Metazoan relationships based on analysis of Cannon 2016 dataset. a) the T_{all} inferred from all 424 partitions and T_{pass} inferred from partitions that passed the test of symmetry. **b**) the tree inferred from partitions that failed the test of symmetry. Red numbers at the internal nodes indicate the bootstrap support values and filled blue circles represent 100% bootstrap support under the best fitting model. The 98/97 bootstrap support values at the Cindaria-Placozoa node in (a) are for T_{all} and T_{pass} respectively.

Discussion

In this paper, we show that model violation is prevalent and has a strong impact on tree reconstruction in many phylogenetic datasets. This impact varies a lot between different datasets and different types of partitions. The trees inferred from different groups of partitions from the same dataset often have topologies that are biologically and statistically significantly different.

Our results show great variation in the extent of model violation among different datasets and partitions. This is demonstrated by the different proportion of partitions that failed the matched-pairs tests of symmetry in each dataset and in each genomic context (codon position, rRNA, tRNA, UCE or other) and type of genome (nuclear, mitochondrial, plastid and virus). Model violations are most frequently observed in the third codon positions for viral, mitochondrial and nuclear genomes and intergenic spacers in plastid sequences. Yet, our results affirm that non-SRH evolution is not constrained to these genomic regions. For example, in a dataset of first and second codon positions from 212 genes (Cannon, et al. 2016b), 21% of partitions showed significant evidence of violating the SRH assumptions. The tree inferred from the partitions that show significant model violation differs a great deal in its topology from the tree inferred from the partitions that do not show significant model violation, particularly with respect to the placement of the focal taxon Xenoturbella (fig. 3).

The results of our study also provide some insights into the likely cause of model violation in the datasets we examined. Figure 2 shows that violation of marginal symmetry (assessed with $\text{BiSymTest}_{\text{mar}}$) was much more common that violation of internal symmetry (assessed with $\text{BiSymTest}_{\text{int}}$). This suggests that non-stationarity, which is associated with marginal symmetry, is a more common cause of systematic bias than non-homogeneity in the datasets that we examined (see also Jayaswal, et al. 2005; Ababneh, et al. 2006; Song, et al. 2010). This suggests that the development and application of non-stationary models (e.g. Yang and Roberts 1995; Yap and Speed 2005) may be an important avenue towards reducing systematic bias in future analyses.

One limitation of the binomial tests that we propose here is that their power will be limited if there are few differences between the sequences being examined. Indeed, our analyses show that in our representative sample of more than 3500 partitions from published datasets, roughly ~50% of the variance in whether a partition passes or fails a given test can be

attributed to the number of observed differences between the sequences. Nevertheless, this suggests that the remaining ~50% of the variance in whether a partition passes or fails a test could be attributable to variation in the extent of model violation among partitions (the number is much higher for $BiSymTest_{int}$). This suggests that we should be cautiously optimistic: although a lack of power on small or slowly-evolving partitions may induce some false negatives (i.e. failures to identify partitions that have evolved under non-SRH conditions), removing the partitions that show the strongest evidence of model violation may still improve phylogenetic analyses by reducing the overall burden of model violation on the inference of the tree topology. We hope that our implementation of these tests in user-friendly software IQ-TREE will allow empirical phylogeneticists to continue to explore whether this is the case.

Acknowledgments

The authors would like to thank Lars Jermiin for discussions.

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.

Barton K. 2009. MuMIn: multi-model inference, R package version 0.12. 0. http://r-forge. r-project. org/projects/mumin/.

Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting Linear Mixed-Effects Models Using lme4. 2015 67:48.

Bazinet AL, Zwickl DJ, Cummings MP. 2014. A gateway for phylogenetic analysis powered by grid computing featuring GARLI 2.0. Syst. Biol. 63:812-818.

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-r R, Li C, Munroe TA, Ballesteros JA, Ortí 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. 2006. A Bayesian compound stochastic process for modeling nonstationary and nonhomogeneous sequence evolution. Mol. Biol. Evol. 23:2058-2071.

Bogdanowicz D, Giaro K, Wrobel B. 2012. TreeCmp: Comparison of Trees in Polynomial Time. Evol Bioinform 8:475-487.

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comp. Biol. 10:e1003537.

Bourlat SJ, Juliusdottir T, Lowe CJ, Freeman R, Aronowicz J, Kirschner M, Lander ES, Thorndyke M, Nakano H, Kohn AB. 2006. Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. Nature 444:85.

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

Bowker AH. 1948. A test for symmetry in contingency tables. J Am Stat Assoc 43:572-574.

Brady A, Salzberg S. 2011. PhymmBL expanded: confidence scores, custom databases, parallelization and more. Nat. Methods 8:367.

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. 2014. Detection of Implausible Phylogenetic Inferences Using Posterior Predictive Assessment of Model Fit. Syst. Biol. 63:334-348.

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

Brown JM, Thomson RC. 2017. Bayes Factors Unmask Highly Variable Information Content, Bias, and Extreme Influence in Phylogenomic Analyses. Syst. Biol. 66:517-530.

Brown JM, Thomson RC. 2018. Evaluating Model Performance in Evolutionary Biology. Annu Rev Ecol Evol S 49:null.

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 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.

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.

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

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

Delsuc F, Brinkmann H, Philippe H. 2005. Phylogenomics and the reconstruction of the tree of life. Nature Reviews Genetics 6:361.

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.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7:214.

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

Dunn M, Greenhill SJ, Levinson SC, Gray RD. 2011. Evolved structure of language shows lineage-specific trends in word-order universals. Nature 473:79.

Dutheil J, Boussau B. 2008. Non-homogeneous models of sequence evolution in the Bio++ suite of libraries and programs. BMC Evol. Biol. 8:255.

Eisen JA. 1998. Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. Genome Res. 8:163-167.

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.

Farrell LE, Roman J, Sunquist ME. 2000. Dietary separation of sympatric carnivores identified by molecular analysis of scats. Mol. Ecol. 9:1583-1590.

Felsenstein J. 2004. Inferring phylogenies: Sinauer associates Sunderland, MA.

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.

Fong JJ, Brown JM, Fujita MK, Boussau B. 2012b. A phylogenomic approach to vertebrate phylogeny supports a turtle-archosaur affinity and a possible paraphyletic 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.

Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S. 2002. Genome sequence of the human malaria parasite Plasmodium falciparum. Nature 419:498.

Goldman N. 1993. Statistical tests of models of DNA substitution. J. Mol. Evol. 36:182-198.

Goremykin V, Hellwig F. 2005. Evidence for the most basal split in land plants dividing bryophyte and tracheophyte lineages. Plant Syst. Evol. 254:93-103.

Gray RD, Drummond AJ, Greenhill SJ. 2009. Language Phylogenies Reveal Expansion Pulses and Pauses in Pacific Settlement. Science 323:479.

Grenfell BT, Pybus OG, Gog JR, Wood JLN, Daly JM, Mumford JA, Holmes EC. 2004. Unifying the Epidemiological and Evolutionary Dynamics of Pathogens. Science 303:327.

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.

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59:307-321.

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. Molecular Biology and Evolution 35:518-522.

Höhna S, Landis MJ, Heath TA, Boussau B, Lartillot N, Moore BR, Huelsenbeck JP, Ronquist F. 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. Syst. Biol. 65:726-736.

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.

Hyman IT, Ho SY, Jermiin LS. 2007. Molecular phylogeny of Australian Helicarionidae, Euconulidae and related groups (Gastropoda: Pulmonata:

Stylommatophora) based on mitochondrial DNA. Mol. Phylogen. Evol. 45:792-812.

Jayaswal V, Jermiin LS, Robinson J. 2005. Estimation of phylogeny using a general Markov model. Evol Bioinform 1:117693430500100005.

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, 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, Jayaswal V, Ababneh FM, Robinson J. 2017. Identifying Optimal Models of Evolution. In: Keith JM, editor. Bioinformatics. Melbourne: Humana Press, New York, NY. p. 379-420.

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.

Kishino H, Miyata T, Hasegawa M. (Kishino1990 co-authors). 1990. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. J. Mol. Evol. 31:151-160.

Knight R, Maxwell P, Birmingham A, Carnes J, Caporaso JG, Easton BC, Eaton M, Hamady M, Lindsay H, Liu Z. 2007. PyCogent: a toolkit for making sense from sequence. Genome biology 8:R171.

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.

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.

Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol. Biol. Evol. 21:1095-1109.

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

Mäser P, Thomine S, Schroeder JI, Ward JM, Hirschi K, Sze H, Talke IN, Amtmann A, Maathuis FJM, Sanders D, et al. 2001. Phylogenetic Relationships within Cation Transporter Families of Arabidopsis. Plant Physiol. 126:1646.

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.

Mir A, Russello F. 2010. The mean value of the squared path-difference distance for rooted phylogenetic trees. J Math Anal Appl 371:168-176.

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.

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.

Murray S, Jørgensen MF, Ho SY, Patterson DJ, Jermiin LS. 2005. Improving the analysis of dinoflagellate phylogeny based on rDNA. Protist 156:269-286.

Nabholz B, Künstner A, Wang R, Jarvis ED, Ellegren H. 2011. Dynamic Evolution of Base Composition: Causes and Consequences in Avian Phylogenomics. Mol. Biol. Evol. 28:2197-2210.

Nakagawa S, Schielzeth H. 2013. A general and simple method for obtaining R2 from generalized linear mixed-effects models. Methods in Ecology and Evolution 4:133-142.

Nesnidal MP, Helmkampf M, Bruchhaus I, Hausdorf B. 2010. Compositional Heterogeneity and Phylogenomic Inference of Metazoan Relationships. Mol. Biol. Evol. 27:2095-2104.

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.

Paton T, Haddrath O, Baker AJ. 2002. Complete mitochondrial DNA genome sequences show that modern birds are not descended from transitional shorebirds. Proceedings of the Royal Society of London B: Biological Sciences 269:839-846.

Philippe H, Brinkmann H, Copley RR, Moroz LL, Nakano H, Poustka AJ, Wallberg A, Peterson KJ, Telford MJ. 2011. Acoelomorph flatworms are deuterostomes related to Xenoturbella. Nature 470:255.

Philippe H, Delsuc F, Brinkmann H, Lartillot N. 2005. Phylogenomics. Annu Rev Ecol Evol S 36:541-562.

Phillips MJ, Delsuc Fdr, Penny D. 2004. Genome-Scale Phylogeny and the Detection of Systematic Biases. Mol. Biol. Evol. 21:1455-1458.

Rightmyer 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.

Rightmyer 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.

Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61:539-542.

Rzhetsky A, Nei M. 1995. Tests of applicability of several substitution models for DNA sequence data. Mol. Biol. Evol. 12:131-151.

Salipante SJ, Horwitz MS. 2006. Phylogenetic fate mapping. Proceedings of the National Academy of Sciences 103:5448.

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.

Sharanowski 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.

Sheffield NC, Song H, Cameron SL, Whiting MF. 2009. Nonstationary Evolution and Compositional Heterogeneity in Beetle Mitochondrial Phylogenomics. Syst. Biol. 58:381-394.

Shimodaira H, Hasegawa M. 1999. Multiple Comparisons of Log-Likelihoods with Applications to Phylogenetic Inference. Mol. Biol. Evol. 16:1114.

Shimodaira HH. 2002. An Approximately Unbiased Test of Phylogenetic Trees. Systematic biology 51:492-508.

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.

Song H, Sheffield NC, Cameron SL, Miller KB, Whiting MF. 2010. When phylogenetic assumptions are violated: base compositional heterogeneity and among-site rate variation in beetle mitochondrial phylogenomics. Syst. Entomol. 35:429-448.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312-1313.

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

Stuart A. 1955. A Test for Homogeneity of the Marginal Distributions in a Two-Way Classification. Biometrika 42:412-416.

Sullivan J, Joyce P. 2005. Model selection in phylogenetics. Annual Review of Ecology Evolution and Systematics 36:445-466.

Sumner JG, Fernandez-Sanchez J, Jarvis PD. 2012. Lie Markov models. J. Theor. Biol. 298:16-31.

Swofford DL. 2001. Paup*: Phylogenetic analysis using parsimony (and other methods) 4.0. B5.

Tarrío R, Rodríguez-Trelles F, Ayala FJ. 2001. Shared nucleotide composition biases among species and their impact on phylogenetic reconstructions of the Drosophilidae. Mol. Biol. Evol. 18:1464-1473.

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.

Wainwright PC, Smith WL, Price SA, Tang KL, Sparks JS, Ferry LA, Kuhn KL, Eytan RI, Near TJ. 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.

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.

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

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.

Yang Z, Rannala B. 2012. Molecular phylogenetics: principles and practice. Nat. Rev. Genet. 13:303-314.

Yao H, Kristensen DM, Mihalek I, Sowa ME, Shaw C, Kimmel M, Kavraki L, Lichtarge O. 2003. An accurate, sensitive, and scalable method to identify functional sites in protein structures. J. Mol. Biol. 326:255-261.

Yao Y-G, Bravi CM, Bandelt H-J. 2004. A call for mtDNA data quality control in forensic science. Forensic Sci. Int. 141:1-6.

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.