The Prevalence and Impact of Model Violations in Phylogenetic Analysis

Suha Naser-Khdour[1],*, Bui Quang Minh[1,2], Wenqi Zhang[1], Eric A. Stone[1], and Robert Lanfear[1]

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

[2]Research School of Computer Science, Australian National University, Canberra, Australian Capital Territory, Australia

*Corresponding author: E-mail: suha.naser@anu.edu.au.

Abstract

In phylogenetic inference, we commonly use models of substitution which assume that sequence evolution is stationary, reversible, and homogeneous (SRH). Although the use of such models is often criticized, the extent of SRH violations and their effects on phylogenetic inference of tree topologies and edge lengths are not well understood. Here, we introduce and apply the maximal matched-pairs tests of homogeneity to assess the scale and impact of SRH model violations on 3,572 partitions from 35 published phylogenetic data sets. We show that roughly one-quarter of all the partitions we analyzed (23.5%) reject the SRH assumptions, and that for 25% of data sets, tree topologies inferred from all partitions differ significantly from topologies inferred using the subset of partitions that do not reject the SRH assumptions. This proportion increases when comparing trees inferred using the subset of partitions that rejects the SRH assumptions, to those inferred from 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. Our results 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).

Key words: 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, 2004; Grenfell et al. 2004; Salipante and Horwitz 2006; Gray et al. 2009; Brady and Salzberg 2011; Dunn et al. 2011).

Most phylogenetic studies use models of sequence 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 (substitution rates between nucleotides or amino acids are equal in both directions), 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 remedied simply by increasing the size of a data set (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 data sets are steadily growing in terms of taxonomic and site sampling, 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 this systematic bias in empirical studies.

One approach to accommodate 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; Jayaswal et al. 2007, 2011, 2014; Knight et al. 2007; Dutheil and Boussau 2008; Sumner et al. 2012; Zou et al. 2012; Groussin et al. 2013; Nguyen et al. 2015; Woodhams et al. 2015). However, such models 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 that assume sequences have evolved under SRH conditions, such as the general time reversible family of models implemented in many of 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 data sets or parts of data sets, and reconstructs 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; Ho et al. 2006), and there are also a posteriori tests for absolute model adequacy which are employed after trees have been estimated (Goldman 1993; Bollback 2002; 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 which sequences and loci are inconsistent with the SRH assumptions will allow us to choose more complex models or to omit some of these sequences and 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 data set increases, because systematic bias has an increasing effect on inferences from larger phylogenetic data sets (Ho and Jermiin 2004; Jermiin et al. 2004; Phillips et al. 2004; Delsuc et al. 2005).

In this article, we evaluate the extent and effect of model violation due to non-SRH evolution using 35 empirical data sets with a total of 3,572 partitions. We determine if the SRH assumptions are violated by extending and applying the matched-pairs tests of homogeneity (Jermiin et al. 2017) to

each partition. We then compare the phylogenetic trees for each data set 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. Our results suggest that violating SRH assumptions can have substantial impacts on phylogenetic inference.

Materials and Methods

Empirical Data Sets

In order to assess the impact of model violation in phylogenetics, we first gathered a representative sample of 35 partitioned empirical data sets 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, that is, such that all loci and all codon positions within protein-coding loci could be identified, we selected the data sets 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 data sets include nucleotide sequences from nuclear, mitochondrial, plastid, and virus genomes, and include protein-coding DNA, introns, intergenic spacers, tRNA, rRNA, and ultraconserved elements. The number of taxa and sites in these data sets range from 27 to 355 and from 699 to 1,079,052, respectively. The clades represented in these data sets include animals, plants, and viruses. We partitioned all data sets to the maximum possible extent based on the biological properties of the data, that is, we divided every locus and every codon position within each protein-coding locus into a separate partition. All partitioning information is available at the github repository (https://github.com/roblanf/SRHtests/tree/master/datasets), and the full details of every data set are provided in table 1 and in supplementary extended table 5, Supplementary Material online.

Workflow Summary

Figure 1 outlines the workflow. For each partition in each data set, we used a new approach based on the three matched-pairs tests of homogeneity to ask whether the evolution of the aligned sequences in the partition rejects the SRH assumptions. The three matched-pairs tests of homogeneity, described in more detail below, test three slightly different assumptions about the historical process that generated each aligned pair of sequences in a given partition. 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). For each test, we classify each partition as *pass* if the result of the test is nonsignificant or *fail* if the result of the test is significant. We then denote the original data set as D_{all},

Table 1

Number of Taxa, Number of Sites, Clade, and Study Reference for Each Data Set That Have Been Used in This Study

Data Set	Study References	Data Set References	Clade	Taxa	Sites
Anderson_2013	Anderson et al. (2014)	Anderson et al. (2013)	Loliginids	145	3,037
Bergsten_2013	Bergsten et al. (2013)	Bergsten et al. (2013)	Dytiscidae	38	2,111
Broughton_2013	Broughton et al. (2013)	Broughton et al. (2013)	Osteichthyes	61	19,997
Brown_2012	Brown et al. (2012)	Brown et al. (2012)	Ptychozoon	41	1,665
Cannon_2016a	Cannon et al. (2016)	Cannon et al. (2016)	Metazoa	78	89,792
Cognato_2001	Cognato and Vogler (2001)	Cognato and Vogler (2001)	Coleoptera: Scolytinae	44	1,897
Day_2013	Day et al. (2013)	Day et al. (2013)	Synodontis	152	3,586
Devitt_2013	Devitt et al. (2013)	Devitt et al. (2013)	Ensatina eschscholtzii klauberi	69	823
Dornburg_2012	Dornburg et al. (2012)	Dornburg et al. (2012)	Teleostei: Beryciformes: Holocentridae	44	5,919
Faircloth_2013	Faircloth et al. (2013)	Faircloth et al. (2013)	Actinopterygii	27	149,366
Fong_2012	Fong et al. (2012)	Fong et al. (2012)	Vertebrata	110	25,919
Horn_2014	Horn et al. (2014)	Horn et al. (2014)	Euphorbia	197	11,587
Kawahara_2013	Kawahara and Rubinoff (2013)	Kawahara and Rubinoff (2013)	Hyposmocoma	70	2,238
Lartillot_2012	Lartillot and Delsuc (2012)	Lartillot and Delsuc (2012)	Eutheria	78	15,117
McCormack_2013	McCormack et al. (2013)	McCormack et al. (2013)	Neoaves	33	1,079,052
Moyle_2016	Moyle et al. (2016)	Moyle et al. (2016)	Oscines	106	375,172
Murray_2013	Murray et al. (2013)	Murray et al. (2013)	Eucharitidae	237	3,111
Oaks_2011	Oaks (2011)	Oaks (2011)	Crocodylia	79	7,282
Rightmyer_2013	Rightmyer et al. (2013)	Rightmyer et al. (2013)	Hymenoptera: Megachilidae	94	3,692
Sauquet_2011	Sauquet et al. (2012)	Sauquet et al. (2011)	Nothofagus	51	5,444
Seago_2011	Seago et al. (2011)	Seago et al. (2011)	Coccinellidae	97	2,253
Sharanowski_2011	Sharanowski et al. (2011)	Sharanowski et al. (2011)	Braconidae	139	3,982
Siler_2013	Siler et al. (2013)	Siler et al. (2013)	Lycodon	61	2,697
Tolley_2013	Tolley et al. (2013)	Tolley et al. (2013)	Chamaeleonidae	203	5,054
Unmack_2013	Unmack et al. (2013)	Unmack et al. (2013)	Melanotaeniidae	139	6,827
Wainwright_2012	Wainwright et al. (2012)	Wainwright et al. (2012)	Acanthomorpha	188	8,439
Wood_2012	Wood et al. (2013)	Wood et al. (2012)	Archaeidae	37	5,185
Worobey_2014a	Worobey et al. (2014)	Worobey et al. (2014)	Influenzavirus A	146	3,432
Worobey_2014b	Worobey et al. (2014)	Worobey et al. (2014)	Influenzavirus A	327	759
Worobey_2014c	Worobey et al. (2014)	Worobey et al. (2014)	Influenzavirus A	92	1,416
Worobey_2014d	Worobey et al. (2014)	Worobey et al. (2014)	Influenzavirus A	355	1,497
Worobey_2014e	Worobey et al. (2014)	Worobey et al. (2014)	Influenzavirus A	340	699
Worobey_2014f	Worobey et al. (2014)	Worobey et al. (2014)	Influenzavirus A	332	2,151
Worobey_2014g	Worobey et al. (2014)	Worobey et al. (2014)	Influenzavirus A	326	2,274
Worobey_2014h	Worobey et al. (2014)	Worobey et al. (2014)	Influenzavirus A	351	2,280

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.
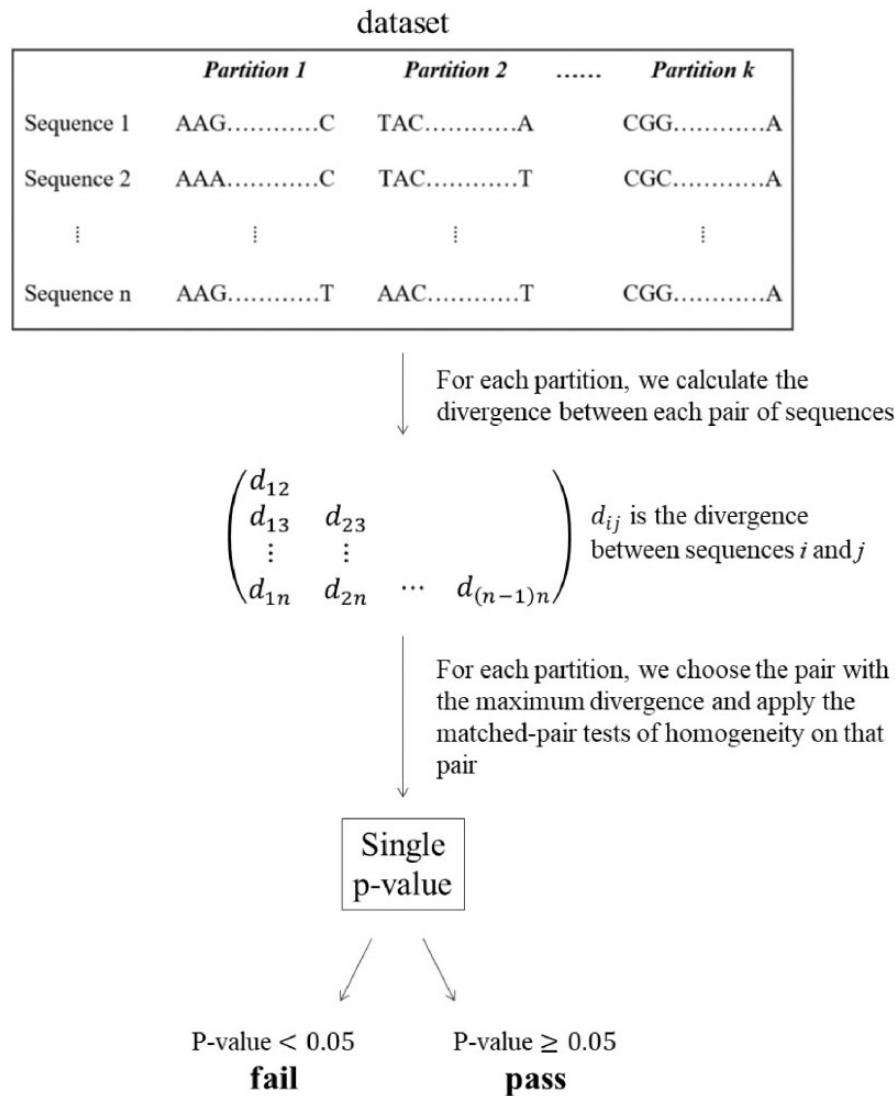
Matched-Pairs Tests of Homogeneity

The three matched-pairs tests of homogeneity that are applied to pairs of sequences are: the MPTS (matched-pairs test of symmetry), MPTMS (matched-pairs test of marginal symmetry), and MPTIS (matched-pairs test of internal symmetry). The statistics are computed on an *m*-by-*m* (*m* is 4 for nucleotides and 20 for amino acids) divergence matrix D with elements d_{ij}, where d_{ij} is the number of alignment sites having

nucleotide (or amino acid) i in the first sequence and nucleotide (or amino acid) j in the second sequence.

The MPTS tests the symmetry of D by computing the Bowker's (1948) test statistic as the χ^2 distance between D and its transpose:

$$S_B^2 = \sum_{1 \le i < j \le m} \frac{(d_{ij} - d_{ji})^2}{(d_{ij} + d_{ji})},$$

where $d_{ij} + d_{ji} > 0$. A P value is then obtained by a χ^2 test with f degrees of freedom, where f is the number of (i, j) pairs for which $d_{ij} + d_{ji} > 0$. A small P value (e.g., <0.05) indicates that the assumption of symmetry is rejected at that significance level, suggesting that evolution is nonstationary, nonhomogeneous, or both (Jermiin et al. 2017).

FIG. 1.—Flow chart of methodology. For each partition in the alignment, we choose the pair of sequences with the maximum divergence and apply the matched-pairs tests of homogeneity on that pair.

The MPTMS tests the equality of nucleotide or amino acid composition between two sequences. To do so, MPTMS computes the Stuart's test statistic $S_S^2 = u^T V^{-1} u$ using the difference between nucleotide or amino acid frequencies of two sequences, u, and its variance–covariance matrix, V. In detail, u is given by $u^T = (d_{1\bullet} - d_{\bullet 1}, \ d_{2\bullet} - d_{\bullet 2}, \ldots, d_{k\bullet} - d_{\bullet k})$ where $d_{i\bullet}$ is the sum of d_{ij} over j, $d_{\bullet j}$ is the sum of d_{ij} over i, and, $k = m - 1$. V, the estimated variance–covariance matrix of u under the assumption of marginal symmetry, is defined elementwise by:

$$v_{ij} = \begin{cases} d_{i\bullet} + d_{\bullet i} - 2d_{ii}, & i = j \\ -(d_{ij} + d_{ji}), & i \neq j \end{cases}.$$

A P value is obtained by a χ^2 test with $m-1$ degrees of freedom. A small P value (<0.05) indicates that the stationarity assumption is rejected. Note that when V is not invertible, the Stuart's statistic S_S^2 is ill-defined and the MPTMS is not applicable.

The MPTIS uses the test statistic as the difference between Bowker's and Stuart's statistic:

$S_I^2 = S_B^2 - S_S^2$. S_I^2 is χ^2 distributed with $f - m + 1$ degrees of freedom. A small P value (<0.05) indicates that the homogeneity assumption is rejected.

The MPTS, MPTMS, and MPTIS test different aspects of the symmetry with which differences accumulate between pairs of sequences due to the substitution process. The MPTS is a

comprehensive and sufficient test to determine whether the data comply 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 the other two 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. If the violation of the SRH assumptions stems from changes in the relative 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, that is, there is not a one-to-one correspondence between the three tests and violation of the three SRH conditions (Jermiin et al. 2017).

The three matched-pairs tests of homogeneity are appropriate to test for SRH assumptions as they consider the alignment on a site-by-site basis. The basic intuition that underlies these tests is that two sequences diverging under SRH conditions should accumulate differences symmetrically (e.g., both sequences are equally likely to accumulate at a C to T change at a site in which both originally shared a C). This symmetry of accumulation is reflected by symmetries in the resulting difference matrix, violations of which can be assessed statistically. However, these tests 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 developed an approach to extend the matched-pairs tests of homogeneity to accommodate data sets with more than two sequences.

Maximum Symmetry Test

In order to determine whether a given multiple sequence alignment rejects SRH conditions, we consider only the pair of taxa with the maximum divergence. In order to find the maximum divergent pair, we sum the off-diagonal elements of the divergence matrix and divide by the sum of all elements. We then randomly choose one pair from all the pairs with the maximum divergence score (if there is more than one pair). By using the most divergent sequence pair, we maximize our power to detect model violations without a priori knowledge of the underlying tree topology and the dependencies that it induces in the data. For the maximum divergent pair, we then apply the matched-pair tests of homogeneity and calculate their χ^2 P values. If the obtained P value is <0.05, then we consider that the null hypothesis of SRH evolution is rejected for the corresponding partition and we add it to the D_{fail} data set. Otherwise, we add it to the

D_{pass} data set. We denote our applications of the MPTS, MPTMS, and MPTIS based on the d_{max} *Pair* as MaxSymTest, MaxSymTest$_{mar}$, and MaxSymTest$_{int}$, respectively.

Phylogenetic Inference

We used IQ-TREE (Nguyen et al. 2015) to infer up to seven phylogenetic trees for every data set: T_{all} (all partitions from the original data set; D_{all}); and T_{pass} and T_{fail} based on the D_{pass} and D_{fail} data sets from each of the three tests (MaxSymTest, MaxSymTest$_{mar}$, MaxSymTest$_{int}$), 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), which allows each partition to have its own evolutionary model and edge-linked rate determined by ModelFinder (Kalyaanamoorthy et al. 2017) followed 1,000 ultrafast bootstrap replicates (Hoang et al. 2018).

Distance between Trees

For each of the three tests (MPTS, MPTMS, MPTIS) we calculated the Normalized Path-Difference (NPD) and quartet distance (QD) (Steel and Penny 1993; Sand et al. 2014) 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 nonempty 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 this study, because we are interested only in differences between topologies, we use the variant of the PD metric that ignores branch lengths. In order to compare path distances between trees with different number of taxa, we normalized 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 0 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 >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 nonnegative, the NPD is also guaranteed to be nonnegative.

The QD metric is defined as the fraction of quartets (subsets of four taxa) that induce different subtrees between the two trees being compared. QD ranges between 0 and 1, where 0 means that two trees are identical and 1 means that they do not share any quartet subtrees. Compared with PD, QD has the advantage that its distribution is less sensitive to the underlying distribution of tree topologies (Steel and Penny 1993).

Tree Topology Tests

The NPD and the QD give us measures of the differences between pairs of trees, but they do not tell us whether the

differences are phylogenetically significant in the three data sets (D_{pass}, D_{all}, and D_{fail}) derived from a given test. For example, trees that differ due to stochastic error associated with small data sets may be very different, but such differences may not be statistically significant. To assess the significance of the differences between T_{pass}, T_{all}, and T_{fail}, we used the weighted Shimodaira–Hasegawa (wSH) test (Shimodaira and Hasegawa 1999; Shimodaira 2002) implemented in IQ-TREE with 1,000 RELL replicates (Kishino et al. 1990). Given the alignment (D_{pass}), the wSH test computes a P value for each tree, where a small P value (<0.05) implies that the corresponding tree has a significantly worse likelihood than the best tree in the set of T_{pass}, T_{all}, and T_{fail}. We use D_{pass} for these tests because it is, by definition, the only data set that does not reject the underlying assumptions of the SH test. As such, we only compute sWH P values when D_{pass} is nonempty. Thus, we performed a wSH test for each of the three MaxSymTest variants: each of which asks whether T_{all} and/or T_{fail} can be rejected in favor of T_{pass}.

Correlation between Number of Substitutions and Model Violation

We hypothesized 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 homogeneity. 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 data set 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 associates with whether a partition fails a given test of symmetry, after accounting for differences between the data sets. To calculate the R^2 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 –symtest in IQ-TREE to perform the three MaxSymTest matched-pairs tests of symmetry. In addition, the option –symtest-remove-bad allows users to remove from the final analysis partitions that fail the MaxSymTest. One can change the removal criterion to MaxSymTest$_{mar}$ or MaxSymTest$_{int}$ via the –symtest-type MAR|INT option. In addition, the cutoff P value can be changed using the –symtest-pval NUM option, where the default value is 0.05.

Reproducibility

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 Data Sets

Across all 3,572 partitions analyzed, 573 (16.0%) failed the MaxSymTest, 728 (20.4%) failed the MaxSymTest$_{mar}$, and 312 (2.8%) failed the MaxSymTest$_{int}$. In total, 840 (23.5%) of the partitions failed at least one test.

The proportion of partitions failing each test varied substantially among data sets (fig. 2), but on an average, 21.8% of the partitions in each data set failed the MaxSymTest, 27.5% failed the MaxSymTest$_{mar}$, and 5.1% failed the MaxSymTest$_{int}$.

The fraction of failing partitions also varied with the genome type (e.g., mitochondrial, chloroplast, or nuclear) and context (e.g., protein-coding, UCE, tRNA) from which the partition was sequenced (table 2) although we note that a substantial proportion of the partitions from almost every category failed at least one of the tests (table 2).

There were no clear differences in the substitution models that were selected for the partitions that pass or fail the tests (see supplementary extended tables 1–3, Supplementary Material online). However, we note that the two most-frequently selected substitution models (for 35% of the partitions) were relatively simple: K80 (Kimura 1980) and HKY (Hasegawa et al. 1985).

Model Violation Has a Large Influence on Tree Topologies

Using both MaxSymTest and MaxSymTest$_{mar}$, we compared each tree inferred from each data set (T_{all}) to the corresponding trees estimated from the failed (T_{fail}) and passed (T_{pass}) partitions. Disturbingly, for each of the two tree distance metrics that we considered (NPD and QD), we find that the tree inferred from the original data set tended to be more similar to the tree estimated from the failed partitions (table 3 and supplementary extended table 4, Supplementary Material online). Furthermore, the mean NPD distance between T_{pass} and T_{fail} across all 35 data sets for the MaxSymTest was 0.69, that is, the two trees are 69% as dissimilar as random pairs of trees. This suggests that violations of SRH assumptions drive large changes in tree topologies.

The results of the wSH tests (table 4) confirm that the differences between trees that we observe tend to be statistically significant. For example, when using the MaxSymTest$_{mar}$, T_{pass} is a significantly better description of the D_{pass} data than T_{all} in ~37% of the data sets, and better than T_{fail} in ~89% of the data sets.
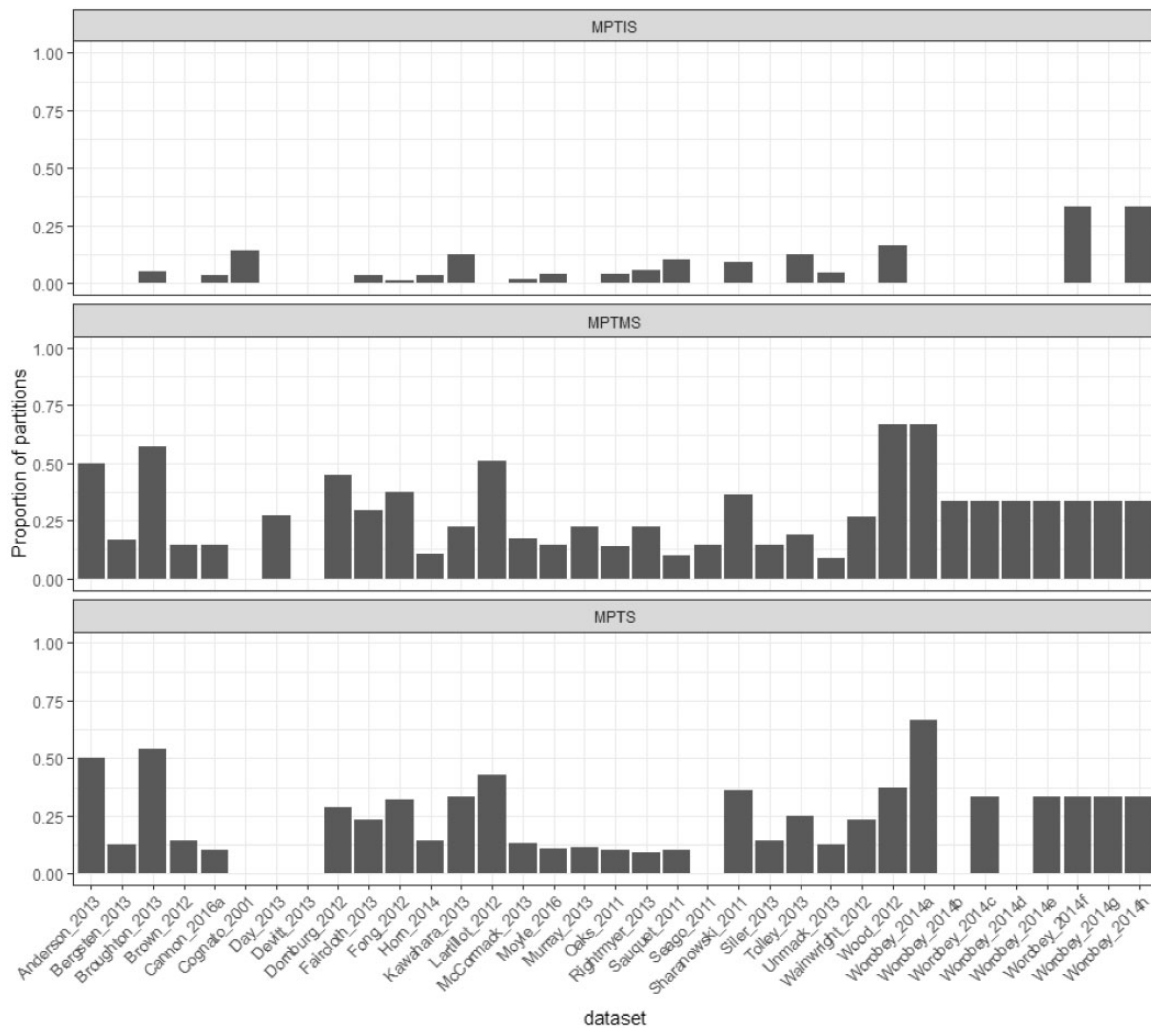
Fig. 2.—The proportion of partitions that reject the null hypothesis of the MaxSymTest, MaxSymTest$_{mar}$, and MaxSymTest$_{int}$ (P value <0.05) in each data set.

Table 2

The Proportion of Partitions That Failed At Least One of the Three Tests—MaxSymTest, MaxSymTest$_{mar}$, and MaxSymTest$_{int}$

Type/Genome	Nuclear	Mitochondrial	Plastid	Virus
First codon positions	20.2%	27.6%	33.3%	25.0%
Second codon positions	21.0%	7.4%	0.0%	25.0%
Third codon positions	76.6%	44.8%	0.0%	75.0%
Other (e.g., intron)	27.8%	100.0%	0.0%	
rRNA	30.0%	25.0%		
UCE	22.5%			
tRNA		0.0%		

The Number of Substitutions Explains Less than One-Third of the Variance in Passing or Failing the Tests of Symmetry

The number of substitutions in a partition explained 27.5% of the variation in whether or not a partition passed or failed the MaxSymTest (supplementary extended fig. 7, Supplementary Material online). This proportion is very similar for

Table 3

The Proportion of Data Sets That Have the Highest NPD Metric (and QD metric) between the Three Comparisons (All-fail, All-pass, Pass–fail) for MaxSymTest, MaxSymTest$_{mar}$, and MaxSymTest$_{int}$

	T_{fail}	T_{pass}
MaxSymTest		
T_{all}	14.3% (4.8%)	4.8% (4.8%)
T_{pass}	80.9% (90.4%)	
MaxSymTest$_{mar}$		
T_{all}	8.3% (0.0%)	8.3% (4.2%)
T_{pass}	83.4% (95.8%)	
MaxSymTest$_{int}$		
T_{all}	28.6% (28.6%)	0.0% (0.0%)
T_{pass}	71.4% (71.4%)	

MaxSymTest$_{mar}$ (24.4%) (supplementary extended fig. 8, Supplementary Material online), but is dramatically lower for the MaxSymTest$_{int}$ (1.8%) (supplementary extended fig. 9, Supplementary Material online). Thus, although the number

Table 4

The Proportion of Data Sets That Have a Significant P Value in the Weighted SH Test When Using D_{pass} As the Input Alignment for the Test

	T_{all}	T_{fail}
MaxSymTest	25%	79%
MaxSymTest$_{mar}$	37%	89%
MaxSymTest$_{int}$	4%	28%

of substitutions in a partition is a highly significant ($P < 2e\text{-}16$) predictor of passing or failing any of the tests, that it explains only about a quarter of the variation suggests that other factors, such as underlying differences in the extent to which partitions violate the SRH assumptions, are driving the remaining ~75% of the variation.

Model Violation Due to Non-SRH Evolution Affects the Inferred Relationship between Even-Toed and Odd-Toed Ungulates in the Tree of Mammals

To examine the effects of model violation in more detail, we selected two data sets for more detailed consideration. Conflicting support for the placement of Xenacoelomorpha, the clade that contains Xenoturbella and Acoelomorpha, in the tree of life across different analyses has led to various hypotheses about the evolution of Bilateria (Cannon et al. 2016). In addition, the interordinal relationships in Laurasiatheria, especially the relationships between Fereuungulata (Perissodactyla, Cetartiodactyla, Carnivora, and Pholidota), in the tree of placental mammals is controversial (Cao et al. 1998; Zhou et al. 2012). It has been suggested that such inferences might be strongly affected by model violation and systematic error (Cao et al. 1998; Delsuc et al. 2005; Philippe et al. 2011; Tsagkogeorga et al. 2013). To assess whether data that pass or fail the MaxSymTest$_{mar}$ show different signals regarding the evolution of the Bilateria and the superorder Laurasiatheria, we examined in more detail the T_{all}, T_{pass}, and T_{fail} trees from recent studies that explored the tree of placental mammals (Lartillot and Delsuc 2012) and the tree of all animals (Cannon et al. 2016). The mammals' data set comprises 78 mammalian taxa, including 73 placental mammals with 51 partitions representing the first, second, and third codon positions of the 17 genes (Lartillot and Delsuc 2012). The tree reconstructed from all of the partitions (T_{all}) and the tree reconstructed from the partitions that pass the MaxSymTest (T_{pass}, 29 partitions) both show Perissodactyla (odd-toed ungulates) as a sister group to Cetartiodactyla (even-toed ungulates) (fig. 3a and supplementary extended figs. 4 and 5, Supplementary Material online). Even so, the bootstrap support for this branch is not high: 73% for T_{all} and 34% for T_{pass}. On the other hand, the tree reconstructed from the data that fail the MaxSymTest (T_{fail}, 22 partitions) shows Perissodactyla as the sister group to the clade that

contains Carnivora + Pholidota with 49% bootstrap support (fig. 3b and supplementary extended fig. 6, Supplementary Material online).

The animal data set comprises 76 metazoan taxa, 2 choanoflagellate outgroups, 212 genes, and 424 partitions representing first and second codon positions (Cannon et al. 2016). The tree reconstructed from all of the partitions (T_{all}) is identical to the trees reconstructed from the 381 partitions that pass the MaxSymTest (T_{pass}), the partitions that fail the MaxSymTest (T_{pass}, 43 partitions), and the tree shown in the original paper from both DNA and amino acid data (Cannon et al. 2016), which places Xenacoelomorpha as the sister group of Nephrozoa (Deuterostomia and Protostomia) with 100% bootstrap support (supplementary extended figs. 1–3, Supplementary Material online).

Discussion

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

Our results show great heterogeneity in the extent of model violation among different data sets and partitions. This is demonstrated by the varying proportion of partitions that failed the matched-pairs tests of homogeneity in each data set 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 far from constrained to these genomic regions. For example, in a data set of placental mammals, of the 22 partitions that failed the MaxSymTest, only 11 are third codon positions. The tree inferred from the partitions that show significant violation of the SRH conditions (T_{fail}) differs in its topology from the tree inferred from the partitions that do not show significant violation of the SRH conditions (T_{pass}) with respect to the interordinal relationships in Laurasiatheria (fig. 3). The tree inferred from partitions that violate the SRH conditions (T_{fail}) is consistent with the results from the original paper in that it places Perissodactyla as a sister group to Carnivora + Pholidota (Lartillot and Delsuc 2012). However, other studies using ML analysis show Perissodactyla to be a sister group to Cetartiodactyla (Graur et al. 1997; Murphy et al. 2001; Tsagkogeorga et al. 2013; Liu et al. 2017), which is also the relationship we find in this study with the tree inferred from partitions that do not show significant violation of the SRH assumptions.

Examining the results of the two other tests (MaxSymTest$_{mar}$ and MaxSymTest$_{int}$) we noticed that all the
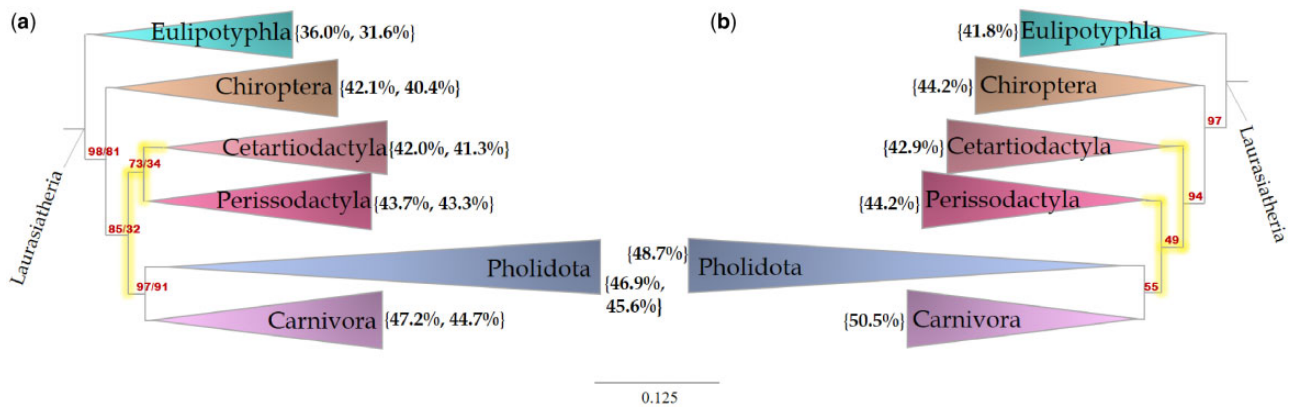
Fig. 3.—Maximum-likelihood trees of mammalian relationships based on analysis of Lartillot 2012 data set. (a) The tree inferred from all 51 partitions and from the 29 partitions that passed the MaxSymTest. (b) The tree inferred from 22 partitions that failed the MaxSymTest. Red numbers at the internal branches indicate the bootstrap support values that are <100% under the best fitting model. Numbers in curly brackets show the GC content (in panel a, %GC and bootstrap support values are for T_{all} and T_{pass}, respectively).

partitions that failed the MaxSymTest also failed the MaxSymTest$_{mar}$, suggesting that those partitions are violating the models mainly due to nonstationarity. Based on this observation, GC content may drive the differences between the trees inferred from all partitions and those inferred from partitions that failed neither MaxSymTest nor MaxSymTest$_{mar}$. Trees with partitions that violate the models tend to group together clades with similar GC content (e.g., as in Betancur-r et al. 2013). However, it is hard to discern any clear evidence for this from examining the GC content of the clades (fig. 3). Yet, our results show that all the clades in the partitions that failed the MaxSymTest have on an average a higher GC content (fig. 3).

The results of our study also provide some insight into the likely cause of model violation in the data sets we examined. Figure 2 shows that violation of marginal symmetry (assessed with MaxSymTest$_{mar}$) was much more common than violation of internal symmetry (assessed with MaxSymTest$_{int}$). This suggests that nonstationarity, which is associated with marginal symmetry, is likely a more common cause of systematic bias than nonhomogeneity in the data sets that we examined (see also Jayaswal et al. 2005; Ababneh et al. 2006; Song et al. 2010). Yet, the difference between the proportion of partitions that failed the MaxSymTest$_{mar}$ and the proportion of partitions that failed the MaxSymTest$_{int}$ could also be due to the higher power of the MaxSymTest$_{mar}$. Either way, this result hints that the development and application of nonstationary models (Yang 1994; Roberts and Yang 1995; Yap and Speed 2005) may be an important avenue toward reducing systematic bias in future analyses. Moreover, our results show a clear preference for simple substitution models with a single transition/transversion ratio over more complex models such as general time reversible. This suggests that developing nonstationary models with a single parameter for the transition/transversion ratio might be sufficient to reduce systematic bias in phylogenetic analysis.

One limitation of using the tests that we propose in this article 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 >3,500 partitions from published data sets, roughly ~25% 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 implies that the remaining ~75% of the variance in whether a partition passes or fails a test could be attributable to other processes, such as variation in the extent of model violation among partitions. 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), the tests we propose still have significant power to identify partitions that show the evidence of model violation. It is possible that removing such partitions from phylogenetic analyses may improve the accuracy of results by reducing the overall burden of model violation on the inference of the tree topology. We hope that our implementation of these tests in the user-friendly software IQ-TREE will allow empirical phylogeneticists to continue to explore whether this is the case.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

Downloaded from https://academic.oup.com/gbe/article-abstract/11/12/3341/5571717 by R.G. Menzies Library, Building #2, Australian National University user on 05 February 2020

Literature Cited

Ababneh F, Jermiin LS, Ma C, Robinson J. 2006. Matched-pairs tests of homogeneity with applications to homologous nucleotide sequences. Bioinformatics 22(10):1225–1231.

Anderson FE, Bergman A, Cheng SH, Pankey MS, Valinassab T. 2014. Lights out: the evolution of bacterial bioluminescence in Loliginidae. Hydrobiologia 725(1):189–203. In: Dryad Data Repository. doi:10.5061/dryad.93s3n.

Barton K. 2009. MuMIn: multi-model inference, R package version 0.12.0. Available from: http://r-forge. r-project. org/projects/mumin/, last accessed May 28, 2019.

Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4. J Stat Softw 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(5):812–818.

Bergsten J, Nilsson AN, Ronquist F. 2013. Bayesian tests of topology hypotheses with an example from diving beetles. Syst Biol. 62(5):660–673. In: Dryad Data Repository. doi:10.5061/dryad.s631d.

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(5):763–785.

Blanquart S, Lartillot N. 2006. A Bayesian compound stochastic process for modeling nonstationary and nonhomogeneous sequence evolution. Mol Biol Evol. 23(11):2058–2071.

Bogdanowicz D, Giaro K, Wrobel B. 2012. TreeCmp: comparison of trees in polynomial time. Evol Bioinformatics. 8:475–487.

Bollback JP. 2002. Bayesian model adequacy and choice in phylogenetics. Mol Biol Evol. 19(7):1171–1180.

Bouckaert R, et al. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 10(4):e1003537.

Bourlat SJ, et al. 2006. Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. Nature 444(7115):85.

Boussau B, Gouy M. 2006. Efficient likelihood computations with non-reversible models of evolution. Syst Biol. 55(5):756–768.

Bowker AH. 1948. A test for symmetry in contingency tables. J Am Stat Assoc. 43(244):572–574.

Brady A, Salzberg S. 2011. PhymmBL expanded: confidence scores, custom databases, parallelization and more. Nat Methods. 8(5):367.

Broughton RE, Betancur RR, Li C, Arratia G, Orti G. 2013. Multi-locus phylogenetic analysis reveals the pattern and tempo of bony fish evolution. PLoS Curr. 5. In: Dryad Data Repository. doi:10.1371/currents.tol.2ca8041495ffafd0c92756e75247483e.

Brown JM. 2014. Detection of implausible phylogenetic inferences using posterior predictive assessment of model fit. Syst Biol. 63(3):334–348.

Brown JM, ElDabaje R. 2009. PuMA: Bayesian analysis of partitioned (and unpartitioned) model adequacy. Bioinformatics 25(4):537–538.

Brown JM, Thomson RC. 2017. Bayes factors unmask highly variable information content, bias, and extreme influence in phylogenomic analyses. Syst Biol. 66(4):517–530.

Brown JM, Thomson RC. 2018. Evaluating model performance in evolutionary biology. Annu Rev Ecol Evol Syst. 49:null.

Brown RM, Siler CD, Das I, Min Y. 2012. 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(3):915–921. In: Dryad Data Repository. doi:10.5061/dryad.7bn0fr99.

Cannon JT, et al. 2016. Xenacoelomorpha is the sister group to Nephrozoa. Nature 530(7588):89–93. In: Dryad Data Repository. doi:10.5061/dryad.493b7.

Cao Y, et al. 1998. Conflict among individual mitochondrial proteins in resolving the phylogeny of eutherian orders. J. Mol. Evol. 47:307–322.

Chernomor O, von Haeseler A, Minh BQ. 2016. Terrace aware data structure for phylogenomic inference from supermatrices. Syst Biol. 65(6):997–1008.

Cognato AI, Vogler AP. 2001. Exploring data interaction and nucleotide alignment in a multiple gene analysis of Ips (Coleoptera: Scolytinae). Syst Biol. 50(6):758–780. In: Dryad Data Repository. doi:10.5061/dryad.678.

Day JJ, et al. 2013. Continental diversification of an African catfish radiation (Mochokidae: Synodontis). Syst Biol. 62(3):351–365. In: Dryad Data Repository. doi:10.5061/dryad.b6225.2.

Delsuc F, Brinkmann H, Philippe H. 2005. Phylogenomics and the reconstruction of the tree of life. Nat Rev Genet. 6(5):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. doi:10.5061/dryad.k9g50.

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(6):1650–1665. In: Dryad Data Repository. doi:10.5061/dryad.k9g50.

Dornburg A, et al. 2012. Molecular phylogenetics of squirrelfishes and soldierfishes (Teleostei: Beryciformes: Holocentridae): reconciling more than 100 years of taxonomic confusion. Mol Phylogenet Evol. 65(2):727–738. In: Dryad Data Repository. doi:10.5061/dryad.3t19n.

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

Duchene DA, Duchene S, Ho S. 2017. New statistical criteria detect phylogenetic bias caused by compositional heterogeneity. Mol Biol Evol. 34(6):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(7345):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(3):163–167.

Faircloth BC, Sorenson L, Santini F, Alfaro ME. 2013. A phylogenomic perspective on the radiation of ray-finned fishes based upon targeted sequencing of ultraconserved elements (UCEs). PLoS One 8(6):e65923. In: Dryad Data Repository. doi:10.5061/dryad.j015n.

Farrell LE, Roman J, Sunquist ME. 2000. Dietary separation of sympatric carnivores identified by molecular analysis of scats. Mol Ecol. 9(10):1583–1590.

Felsenstein J. 2004. Inferring phylogenies. Sunderland (MA): Sinauer Associates.

Fong JJ, Brown JM, Fujita MK, Boussau B. 2012. A phylogenomic approach to vertebrate phylogeny supports a turtle-archosaur affinity and a possible paraphyletic lissamphibia. PLoS One 7(11):e48990. In: Dryad Data Repository. doi:10.5061/dryad.25j6h.

Foster PG. 2004. Modeling compositional heterogeneity. Syst Biol. 53(3):485–495.

Foster PG, Hickey DA. 1999. Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. J Mol Evol. 48(3):284–290.

Gardner MJ, et al. 2002. Genome sequence of the human malaria parasite Plasmodium falciparum. Nature 419(6906):498.

Goldman N. 1993. Statistical tests of models of DNA substitution. J Mol Evol. 36(2):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(1–2):93–103.

3350 *Genome Biol. Evol.* 11(12):3341–3352 doi:10.1093/gbe/evz193 Advance Access publication September 19, 2019

Graur D, Gouy M, Duret L. 1997. Evolutionary affinities of the order Perissodactyla and the phylogenetic status of the superordinal taxa Ungulata and Altungulata. Mol Phylogenet Evol. 7(2):195–200.

Gray RD, Drummond AJ, Greenhill SJ. 2009. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. Science 323(5913):479.

Grenfell BT, et al. 2004. Unifying the epidemiological and evolutionary dynamics of pathogens. Science 303(5656):327.

Groussin M, Boussau B, Gouy M. 2013. A branch-heterogeneous model of protein evolution for efficient inference of ancestral sequences. Syst Biol. 62(4):523–538.

Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 59(3):307–321.

Hasegawa M, Kishino H, Yano T-A. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol. 22(2):160–174.

Ho JW, et al. 2006. SeqVis: visualization of compositional heterogeneity in large alignments of nucleotides. Bioinformatics 22(17):2162–2163.

Ho SY, Jermiin L. 2004. Tracing the decay of the historical signal in biological sequence data. Syst Biol. 53(4):623–637.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. 35(2):518–522.

Höhna S, et al. 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. Syst Biol. 65(4):726–736.

Horn JW, et al. 2014. Evolutionary bursts in Euphorbia (Euphorbiaceae) are linked with photosynthetic pathway. Evolution 68(12):3485–3504. In: Dryad Data Repository. doi:10.5061/dryad.sb1j1.

Hyman IT, Ho SY, Jermiin LS. 2007. Molecular phylogeny of Australian Helicarionidae, Euconulidae and related groups (Gastropoda: Pulmonata: Stylommatophora) based on mitochondrial DNA. Mol Phylogenet Evol. 45(3):792–812.

Jayaswal V, Ababneh F, Jermiin LS, Robinson J. 2011. Reducing model complexity of the general Markov model of evolution. Mol Biol Evol. 28(11):3045–3059.

Jayaswal V, Jermiin LS, Robinson J. 2005. Estimation of Phylogeny Using a General Markov Model. Evol Bioinform 1:62–80.

Jayaswal V, Robinson J, Jermiin L. 2007. Estimation of phylogeny and invariant sites under the general Markov model of nucleotide sequence evolution. Syst Biol. 56(2):155–162.

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(5):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(4):638–643.

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(6):587–589.

Kawahara AY, Rubinoff D. 2013. Convergent evolution of morphology and habitat use in the explosive Hawaiian fancy case caterpillar radiation. J Evol Biol. 26(8):1763–1773. In: Dryad Data Repository. doi:10.5061/dryad.gh895.

Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. 16(2):111–120.

Kishino H, Miyata T, Hasegawa M. 1990. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. J Mol Evol. 31(2):151–160.

Knight R, et al. 2007. PyCogent: a toolkit for making sense from sequence. Genome Biol. 8(8):R171.

Kumar S, Filipski AJ, Battistuzzi FU, Kosakovsky Pond SL, Tamura K. 2012. Statistics and truth in phylogenomics. Mol Biol Evol. 29(2):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(3):1321–1327.

Lartillot N, Delsuc F. 2012. Joint reconstruction of divergence times and life-history evolution in placental mammals using a phylogenetic covariance model. Evolution 66(6):1773–1787. In: Dryad Data Repository. doi:10.5061/dryad.tt28qk6f.

Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol Biol Evol. 21(6):1095–1109.

Liu L, et al. 2017. Genomic evidence reveals a radiation of placental mammals uninterrupted by the KPg boundary. Proc Natl Acad Sci U S A. 114(35):E7282–E7290.

Martijn J, Vosseberg J, Guy L, Offre P, Ettema TJ. 2018. Deep mitochondrial origin outside the sampled alphaproteobacteria. Nature 557(7703):101.

Mäser P, et al. 2001. Phylogenetic relationships within cation transporter families of Arabidopsis. Plant Physiol. 126(4):1646.

McCormack JE, et al. 2013. A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. PLoS One 8(1):e54848. In: Dryad Data Repository. doi:10.5061/dryad.sd080.

Mir A, Russello F. 2010. The mean value of the squared path-difference distance for rooted phylogenetic trees. J Math Anal Appl. 371(1):168–176.

Moyle RG, et al. 2016. Tectonic collision and uplift of Wallacea triggered the global songbird radiation. Nat Commun. 7(1):12709. In: Dryad Data Repository. doi:10.5061/dryad.nf01p.

Murphy WJ, et al. 2001. Molecular phylogenetics and the origins of placental mammals. Nature 409(6820):614–618.

Murray EA, Carmichael AE, Heraty JM. 2013. Ancient host shifts followed by host conservatism in a group of ant parasitoids. Proc Biol Sci. 280(1759):20130495. In: Dryad Data Repository. doi:10.5061/dryad.qn57t.

Murray S, Jørgensen MF, Ho SY, Patterson DJ, Jermiin LS. 2005. Improving the analysis of dinoflagellate phylogeny based on rDNA. Protist 156(3):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(8):2197–2210.

Nakagawa S, Schielzeth H. 2013. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. Methods Ecol Evol. 4(2):133–142.

Nesnidal MP, Helmkampf M, Bruchhaus I, Hausdorf B. 2010. Compositional heterogeneity and phylogenomic inference of metazoan relationships. Mol Biol Evol. 27(9):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(1):268–274.

Oaks JR. 2011. A time-calibrated species tree of Crocodylia reveals a recent radiation of the true crocodiles. Evolution 65(11):3285–3297. In: Dryad Data Repository. doi:10.5061/dryad.5k9s0.

Paton T, Haddrath O, Baker AJ. 2002. Complete mitochondrial DNA genome sequences show that modern birds are not descended from transitional shorebirds. Proc R Soc Lond B. 269(1493):839–846.

Philippe H, Delsuc F, Brinkmann H, Lartillot N. 2005. Phylogenomics. Annu Rev Ecol Evol Syst. 36(1):541–562.

Philippe H, et al. 2011. Acoelomorph flatworms are deuterostomes related to Xenoturbella. Nature 470(7333):255.

Phillips MJ, Delsuc F, Penny D. 2004. Genome-scale phylogeny and the detection of systematic biases. Mol Biol Evol. 21(7):1455–1458.

Rightmyer MG, Griswold T, Brady SG. 2013. 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(3):561–576. In: Dryad Data Repository. doi:10.5061/dryad.jd5ff.

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(3):451–458.

Ronquist F, et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 61(3):539–542.

Rzhetsky A, Nei M. 1995. Tests of applicability of several substitution models for DNA sequence data. Mol Biol Evol. 12(1):131–151.

Salipante SJ, Horwitz MS. 2006. Phylogenetic fate mapping. Proc Natl Acad Sci U S A. 103(14):5448.

Sand A, et al. 2014. tqDist: a library for computing the quartet and triplet distances between binary or general trees. Bioinformatics 30(14):2079–2080.

Sauquet H, 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(2):289–313. In: Dryad Data Repository. doi:10.5061/dryad.qq106tm4.

Seago AE, Giorgi JA, Li J, Ślipiński A. 2011. Phylogeny, classification and evolution of ladybird beetles (Coleoptera: Coccinellidae) based on simultaneous analysis of molecular and morphological data. Mol Phylogenet Evol. 60(1):137–151. In: Dryad Data Repository. doi:10.5061/dryad.dc1r2.

Sharanowski BJ, Dowling APG, Sharkey MJ. 2011. Molecular phylogenetics of Braconidae (Hymenoptera: Ichneumonoidea), based on multiple nuclear genes, and implications for classification. Syst Entomol. 36(3):549–572. In: Dryad Data Repository. doi:10.5061/dryad.1688p.

Sheffield NC, Song H, Cameron SL, Whiting MF. 2009. Nonstationary evolution and compositional heterogeneity in beetle mitochondrial phylogenomics. Syst Biol. 58(4):381–394.

Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. Syst Biol. 51(3):492–508.

Shimodaira H, Hasegawa M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol Biol Evol. 16(8):1114–1116.

Siler CD, Oliveros CH, Santanen A, Brown RM. 2013. Multilocus phylogeny reveals unexpected diversification patterns in Asian wolf snakes (genus *Lycodon*). Zool Scr. 42(3):262–277. In: Dryad Data Repository. doi:10.5061/dryad.cp6gg.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30(9):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(3–4):412–416.

Sullivan J, Joyce P. 2005. Model selection in phylogenetics. Annu Rev Ecol Evol Syst. 36(1):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. Sunderland (MA): Sinauer Associates.

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(8):1464–1473.

Tolley KA, Townsend TM, Vences M. 2013. Large-scale phylogeny of chameleons suggests African origins and Eocene diversification. Proc R Soc B. 280(1759):20130184. In: Dryad Data Repository. doi:10.5061/dryad.11350.

Tsagkogeorga G, Parker J, Stupka E, Cotton JA, Rossiter SJ. 2013. Phylogenomic analyses elucidate the evolutionary relationships of bats. Curr Biol. 23(22):2262–2267.

Unmack PJ, Allen GR, Johnson JB. 2013. Phylogeny and biogeography of rainbowfishes (Melanotaeniidae) from Australia and New Guinea. Mol Phylogenet Evol. 67(1):15–27. In: Dryad Data Repository. doi:10.5061/dryad.qq846.

Wainwright PC, et al. 2012. The evolution of pharyngognathy: a phylogenetic and functional appraisal of the pharyngeal jaw key innovation in labroid fishes and beyond. Syst Biol. 61(6):1001–1027.

Weiss G, von Haeseler A. 2003. Testing substitution models within a phylogenetic tree. Mol Biol Evol. 20(4):572–578.

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(2):264–284. In: Dryad Data Repository. doi:10.5061/dryad.7231d.2.

Woodhams MD, Fernandez-Sanchez J, Sumner JG. 2015. A new hierarchy of phylogenetic models consistent with heterogeneous substitution rates. Syst Biol. 64(4):638–650.

Worobey M, Han GZ, Rambaut A. 2014. A synchronized global sweep of the internal genes of modern avian influenza virus. Nature 508(7495):254–257. In: Dryad Data Repository. doi:10.5061/dryad.m04j9.

Yang Z. 1994. Estimating the pattern of nucleotide substitution. J Mol Evol. 39(1):105–111.

Yang Z, Rannala B. 2012. Molecular phylogenetics: principles and practice. Nat Rev Genet. 13(5):303–314.

Yao H, et al. 2003. An accurate, sensitive, and scalable method to identify functional sites in protein structures. J Mol Biol. 326(1):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):1–6.

Yap VB, Speed T. 2005. Rooting a phylogenetic tree with nonreversible substitution models. BMC Evol Biol. 5:2.

Zhou X, et al. 2012. Phylogenomic analysis resolves the interordinal relationships and rapid diversification of the laurasiatherian mammals. Syst Biol. 61(1):150–164.

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(6):927–940.

Associate editor: David Bryant