doi:10.1111/evo.1403

Heterogeneity in the rate of molecular sequence evolution substantially impacts the accuracy of detecting shifts in diversification rates

Anat Shafir,[1] Dana Azouri,[1,2] Emma E. Goldberg,[3] and Itay Mayrose[1,4]

[1]School of Plant Sciences and Food security, Tel Aviv University, Ramat Aviv, 69978, Israel

[2]School of Molecular Cell Biology & Biotechnology, Tel Aviv University, Ramat Aviv, 69978, Israel

[3]New Mexico Consortium, Los Alamos, New Mexico 87544

[4]E-mail: itaymay@tauex.tau.ac.il

Received October 17, 2019

Accepted May 17, 2020

As species richness varies along the tree of life, there is a great interest in identifying factors that affect the rates by which lineages speciate or go extinct. To this end, theoretical biologists have developed a suite of phylogenetic comparative methods that aim to identify where shifts in diversification rates had occurred along a phylogeny and whether they are associated with some traits. Using these methods, numerous studies have predicted that speciation and extinction rates vary across the tree of life. In this study, we show that asymmetric rates of sequence evolution lead to systematic biases in the inferred phylogeny, which in turn lead to erroneous inferences regarding lineage diversification patterns. The results demonstrate that as the asymmetry in sequence evolution rates increases, so does the tendency to select more complicated models that include the possibility of diversification rate shifts. These results thus suggest that any inference regarding shifts in diversification pattern should be treated with great caution, at least until any biases regarding the molecular substitution rate have been ruled out.

KEY WORDS: Diversification, extinction, phylogeny reconstruction, sequence evolution rates, speciation.

As species richness varies along the tree of life, there is a great interest in identifying factors that affect the rates by which lineages speciate or go extinct. Numerous studies have revealed that species diversification patterns cannot be explained by constant speciation and extinction rates, but rather by a more complex model, in which diversification rates vary across the tree (Stanley et al. 1981; Strathmann and Slatkin 1983; Ricklefs 2003; Rabosky and Goldberg 2015). Several alternative scenarios may lead to altered diversification rates across a phylogeny. The emergence of a novel trait could spur rapid radiation, which will be manifested by a high diversification rate of a certain subclade compared to the rest of the phylogeny. Other traits, which could be related to morphological, reproductive, or ecological features of a lineage, may transition between several alternative states and these transitions could repeatedly influence speciation and extinction patterns. Alternatively, shifts in diversification rates may be trait independent, for example, due to time-related factors, mass extinctions, or considerations related to species richness. To quantitatively study the heterogeneity of species richness across the tree of life, evolutionary biologists have developed a suit of phylogenetic probabilistic methods that aim to identify where shifts in diversification rates had occurred along a phylogeny and whether they are associated with some traits. Although each of these methods have their strengths and weaknesses, they all assume that the given phylogeny (or posterior set of trees) is correct and is accurately time calibrated. Here, we demonstrate that variation in the rates of molecular sequence evolution could introduce artifacts in the reconstructed phylogenies that lead to mistaken inferences of diversification rate shifts.

A pioneering probabilistic approach for detecting organismal traits that are linked to altered rates of diversification is based on the Binary State Speciation and Extinction (BiSSE) model (Maddison et al. 2007). The incorporation of a character evolution process into the likelihood function of a diversification model allows BiSSE to infer unique speciation and extinction rates for each character state. Originally, BiSSE was developed to examine binary traits only. Further development allowed the examination of multiple states (MuSSE; Fitzjohn 2012), quantitative traits (QuaSSE; Fitzjohn 2010), geographic traits (GeoSSE; Goldberg et al. 2011), and the differentiation between state transitions that coincide with speciation events and those that occur continuously in time along branches of the phylogeny (ClaSSE and BiSSE-ness; Goldberg and Igić 2012; Magnuson-Ford and Otto 2012). These "SSE" methods have been widely used to identify trait-dependent shifts in diversification patterns, and have been shown to have adequate statistical power, particularly when large phylogenies are analyzed (Davis et al. 2013). Yet, it has been demonstrated that this framework tends to incorrectly classify neutral traits as being associated with shifts in diversification when empirical phylogenies are analyzed (Rabosky and Goldberg 2015). This finding was attributed to factors that are not included in the model, such as temporal changes in speciation and extinction rates, or co-distribution with other traits affecting diversification rates (Maddison and Fitzjohn 2015; Rabosky and Goldberg 2015). Beaulieu and O'Meara (2016) have subsequently devised the HiSSE model, which accounts for the existence of a hidden trait that affects diversification patterns and is associated with the observed (focal) trait in some parts of the phylogeny. In addition, these authors introduced a set of more complex neutral models (termed CID for character-independent diversification), which accounts for the possibility of a hidden trait that affects rates of diversification but that evolves independently from the focal trait, thus allowing a fairer likelihood comparison with trait-dependent models.

A parallel suite of methods was developed with the aim to detect diversification rate shifts along a phylogeny regardless of a character trait. Of these trait-independent methods, the most commonly used are MEDUSA (Alfaro et al. 2009) and BAMM (Rabosky 2014). BAMM is a Bayesian approach based on a reversible jump Markov chain Monte Carlo, whereas MEDUSA uses a stepwise model selection approach to infer the diversification model that includes the optimal number of rate shifts. Despite their popularity, each of these methods has been shown to have some shortcomings. Concerns were raised regarding the correctness of the likelihood function of MEDUSA, its usage of the Akaike Information Criterion (AIC) for model selection, biased inferences of diversification rate parameters, and possible elevated rate of false inferences of shifts in diversification rates (May and Moore 2016). Some concerns were also raised regarding BAMM; Moore et al. (2016) pointed to problems with prior sensitivity, and an error in the likelihood function as it does not account for unobserved rate shifts along extinct (or nonsampled) branches (but see Rabosky et al. 2017).

The inference of diversification patterns is tightly linked to the branch lengths distribution of the studied phylogeny. As applied in both trait-dependent and trait-independent diversification analyses, the input phylogeny is assumed to be correct and its branch lengths proportional to time. However, the precise estimation of the tree topology and divergence times from empirical data remains highly challenging. Under the molecular clock hypothesis (Zuckerkandl et al. 1965), time calibration of a phylogeny (i.e., the transformation of a tree that is inferred in units of expected number of substitutions of molecular characters to time) can be performed using a single calibration point. More often than not, however, the constant clock hypothesis is violated and thus various alternative models have been developed to alleviate this assumption, including the widely used autocorrelated and uncorrelated relaxed clock models (Thorne et al. 1998; Sanderson 2002; Drummond et al. 2006; Rannala and Yang 2007). Together with ongoing advances in Bayesian phylogenetics, increasingly complex clock models can be considered, while integrating over uncertainties in the phylogeny and model parameters (Bromham et al. 2018). Nevertheless, dating estimates in Bayesian analyses are sensitive to nonrandom taxon sampling (Beaulieu et al. 2015), the sampling fraction (Hugall and Lee 2007; Heath et al. 2008; Soares and Schrago 2012; Schulte 2013), and the choice of the tree and clock priors (Ho et al. 2005; Lepage et al. 2007; Linder et al. 2011; Crisp et al. 2014; Duchêne et al. 2014; Dos Reis et al. 2015; Magallón et al. 2015). For example, assuming a tree prior that follows the Yule process generally results in older estimates compared to birth-death prior (Crisp et al. 2014; Condamine et al. 2015). The wide array of methodologies and models for the reconstruction of time-calibrated phylogenies also results in wide discrepancies in age estimates, as apparent, for example, in the case of the angiosperms (Magallón et al. 2015). Although the addition of informative calibration points, especially at deeper time points, should result in more consistent estimates of divergence times and reduced sensitivity to model misspecification (Duchêne et al. 2014), the fossil record is in many cases limited, and constructing reliable prior bounds for node ages is highly challenging (Magallón et al. 2015). Thus, despite ongoing developments of relaxed clock methods, divergence time estimation from molecular sequence data still suffers from high degrees of uncertainty (Graur and Martin 2004; Dos Reis et al. 2015; Bromham et al. 2018), especially when variation in the rate of molecular substitution exists among lineages (Dornburg et al. 2012; Wertheim et al. 2012; Beaulieu et al. 2015).

Such impact of heterogeneity in the rate of molecular evolution on divergence time estimation was examined by Beaulieu

et al. (2015), who demonstrated that the age of a crown group can be largely overestimated when variations in the rate of sequence evolution exist near the root of the phylogeny. Similarly, when applying a time-calibration procedure to a phylogeny that exhibits high heterogeneity in sequence evolutionary rates due to a particular character trait (i.e., one state evolves faster than the other), we would expect that the conversion of branch lengths to units of time would be biased in such a way that the more rapidly evolving state will be inferred as being associated with lower diversification rates. Namely, the state with the higher sequence evolutionary rate accumulates more changes per unit of time, and is therefore represented by longer branch lengths relative to the slower evolving state. However, under probabilistic diversification models, long branch lengths could be interpreted as having longer intervals of time between speciation events. This could lead to erroneous association of the faster evolving state with lower diversification rates and a spurious inference of trait-dependent diversification. Therefore, heterogeneity in molecular evolutionary rates could substantially affect not only the inference of the phylogeny, but also the downstream analyses, particularly the analyses of diversification patterns—a concern that we investigate in this study.

Indeed, heterogeneity in sequence evolutionary rates is a widespread phenomenon even among closely related lineages (e.g., Tomoko 1995; Moorjani et al. 2016). Molecular evolutionary rates could be affected by various factors, including the generation time (Rutschmann 2006; Baer et al. 2007; Bromham 2009; Thomas et al. 2010; Ho and Duchêne 2014), DNA repair mechanism (Bromham 2009), metabolic rates (Santos 2012), reproductive mode (Paland and Lynch 2006), body size (Fontanillas et al. 2007; Bromham 2011), plant height (Lanfear et al. 2013), and plant growth form (Smith and Donoghue 2008). The population size is another important factor, as slightly deleterious substitutions accumulate at higher rates in lineages of smaller effective population size due to relaxed efficacy of purifying selection (Ohta 1972; Woolfit and Bromham 2003; Bromham 2009, 2011). Such factors could lead to substantial variation in the molecular evolutionary rates throughout the phylogeny. For example, it was demonstrated that the average substitution rate of herbaceous plants is 2.5 times higher compared to woody plants, reaching 4.75 fold difference between the shrub *Dorstenia* (Moraceae) and its sister herbaceous clade Urticaceae (Smith and Donoghue 2008). Substitution rate heterogeneity in angiosperms was also reported to be associated with the life cycle, with annual lineages exhibiting, on average, 1.6 rate acceleration compared to perennials when internal transcribed spacer sequences were examined, with some annual/perennial sister pairs exhibiting large rate differences of five- to 11-fold (Soria-Hernanz et al. 2008). In invertebrates, surface isopod species were found to evolve up to two times faster compared to subterranean sister species, with up to sixfold difference in the synonymous substitution rate between more distant lineages (Saclier et al. 2018). Johnson and Howard (2007) reported that slightly deleterious mutations accumulate six times faster in asexual snail lineages compared to sexual lineages. Large rate variations are also observed when phylogenetically related clades are examined irrespective of a focal trait. For example, the rate of synonymous substitutions in *Arabidopsis* was found to be six times higher than in *Populus* (Tuskan et al. 2006), and 15-fold difference was observed in the comparison between conifers and angiosperms (Buschiazzo et al. 2012). Within mammals, the substitution rate is highly variable even when closely related species are examined (e.g., sevenfold difference between *Rattus norvegicus* and *Gerbillus nigeriae*) with extreme rate variation, by as much as two order of magnitude, across the clade (Nabholz et al. 2008a). Finally, a wide spectrum of heterogeneity in the molecular evolutionary rate was also revealed in bacteria, with inferred substitution rates ranging by approximately two order of magnitude (Duchêne et al. 2016). For example, in Salmonella, bacterial pathogens that are human specialist evolve 2.3-6.6 times higher than generalist species, whereas livestock associated pathogen species evolved five times faster than its sister lineage (Duchêne et al. 2016).

Considering the extent of this phenomenon, we suspect that heterogeneity in molecular evolutionary rates could lead to systematic biases in the inferred phylogeny and consequently to erroneous inferences regarding lineage diversification patterns in many empirical analyses. We further speculate that asymmetric rates of sequence evolution across the phylogeny due to unaccounted-for factors (e.g., a hidden unexamined trait or time-dependent shifts in the substitution pattern) could also affect the phylogeny in a way that would lead to erroneous inferences of patterns concerning lineage diversification. In the current study, we examine the impact of such variations in the nucleotide substitution rate on the performance of methods that aim to detect shifts in the diversification process. To this end, we simulate phylogenies that are neutral with respect to diversification rates (i.e., identical speciation and extinction rates across all lineages of a phylogeny) but that exhibit asymmetry with respect to the substitution rate—either when the asymmetry is associated with an evolving trait or is elevated in a certain subclade of the phylogeny. We then demonstrate that erroneous detection of shifts in diversification patterns by both trait-dependent and trait-independent methods occurs at a higher frequency as the extent of asymmetry in the substitution rate increases.

Methods

SIMULATION PROCEDURES

Our general simulation procedure is to generate trees with heterogeneity in the rate of sequence evolution, and to test if these

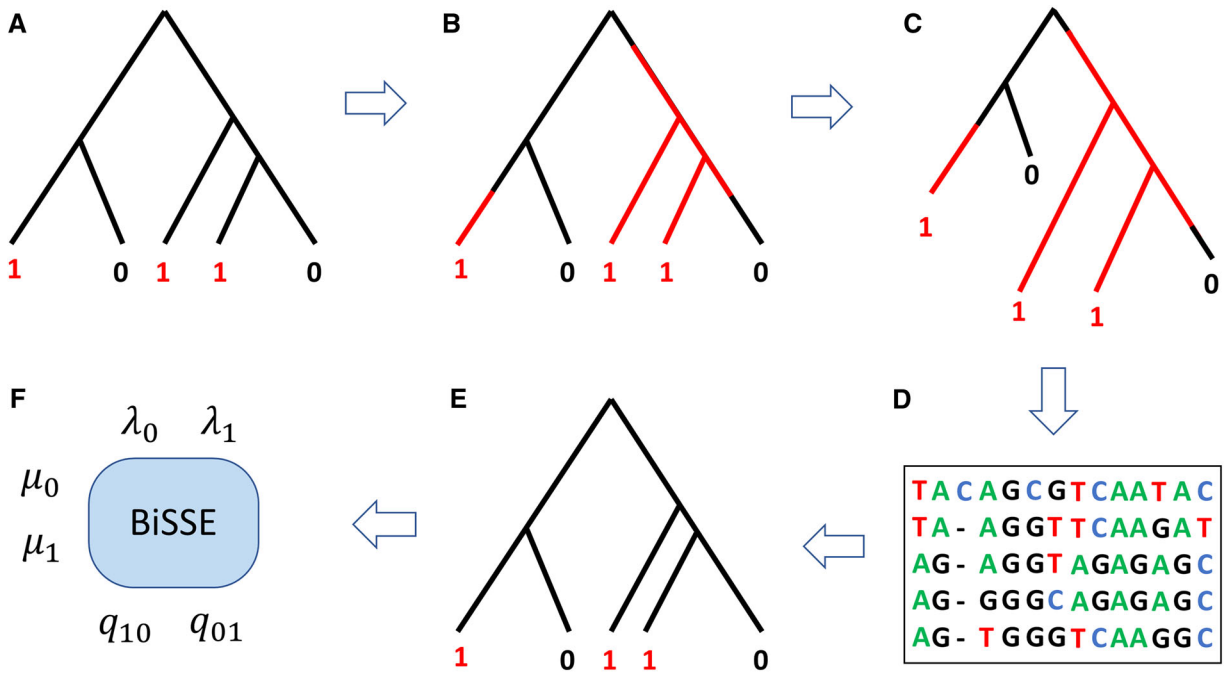


Figure 1. The general simulation outline (exemplified for Simulation Set 1). (A) A simulated ultrametric tree with tip states. (B) The ultrametric tree with the mapped character history; black and red colors represent duration of time spent under state 0 and 1, respectively. (C) The mapped trees are scaled to reflect the effect of the trait on rates of sequence evolution. Here, any red segment of the tree is multiplied by a factor of $r = 2$. (D) A multiple sequence alignment is generated based on the scaled tree. (E) The tree inferred from the multiple sequence alignment. (F) The inferred tree and the tip states are given as input to BiSSE for parameter estimation and hypothesis testing.

yield mistaken inferences regarding shifts in speciation and/or extinction rates. Our simulation procedures consisted of the following six steps (Fig. 1). (1) Generating ultrametric trees. These trees were generated using time homogenous birth-death models (e.g., using traits that are neutral with respect to their effect on lineage diversification). (2) Simulating a process that affects patterns of sequence evolution (e.g., a trait with two states, where state "1" evolves twice as fast as state "0"). (3) Modifying the ultrametric trees according to the simulated process of sequence evolution (e.g., by lengthening segments of the tree evolving under state 1). The branch lengths in the resulting trees are proportional to the expected number of sequence substitutions, and are referred to hereafter as the "scaled trees." (4) Simulating sequence data along the scaled trees, thereby generating multiple sequence alignments. (5) Reconstructing ultrametric (time calibrated) phylogenies given the simulated sequence data. (6) Inferring shifts in the diversification process (using either trait-dependent or trait-independent methods) given the reconstructed phylogenies. Under this simulation scheme, any inference regarding alteration in the speciation or extinction processes is not a correct macroevolutionary conclusion, but an artifact of the asymmetry in sequence evolutionary rates. Below, we detail how we used this general scheme to simulate several different scenarios.

GENERATION OF SIMULATED TREES

Simulation Set 1: Sequence evolutionary rates affected by a known trait (trait-dependent methods)

This simulation set examined whether a focal trait that affects rate of sequence evolution could be mistakenly associated with differences in speciation or extinction rate by trait-dependent diversification methods, such as BiSSE and HiSSE. Specifically, four sets of 100 ultrametric trees were generated, such that each set contained a fixed number of taxa: 50, 100, 200, and 400. All trees were generated under a symmetric BiSSE model, as implemented in the R package diversitree, with the following rates: $0 = \lambda_1 = 0.3$, $\mu_0 = \mu_1 = 0.01$, and $q_{01} = q_{10} = 0.1$. The root state was set to be "0," such that "0" is the ancestral state and "1" is the derived state. The trees were rescaled to a fixed depth (defined as the distance from the root to tips) of 1. The duration of time each segment of the tree was simulated with character state 0 or 1 was obtained using the function history.from.sim.discrete in diversitree. Then, the branch lengths of each ultrametric tree T were scaled according to the character history and the r parameter, representing the sequence substitution rate while the character state is 1 relative to state 0. Four different values of r were examined: 1 (namely, the rate of substitution under state 1 is equal to that of state 0), 2, 4, and 8. This set of values represents different degrees of rate heterogeneities, from low rate to very high

rate variation, thus capturing a range of rate heterogeneity observed in nature. Each branch b_i in T was scaled according to $b_i^{'} = b_i(p_1 \times r + p_0 \times 1)$, where $b_i^{'}$ is the new rescaled branch, and p_1 and p_0 are the proportions of time spent in state 1 and state 0 along that branch, respectively. To keep all trees at the same scale, each tree was then rescaled, such that its total sum of branch lengths was equal before and after the scaling procedure. These tree manipulations on ultrametric trees resulted in nonultrametric trees (termed hereafter "scaled trees"), such that the branch lengths represent the expected number of nucleotide substitutions. Finally, an outgroup species was added to each tree, with depth of $1.2d'$, where d' is the maximal depth of the rescaled tree. The branch length from the new root to the descendant internal node (which was previously the root) was set to be $0.2d'$.

Simulation Set 2: Sequence evolutionary rates affected by a hidden trait (trait dependent)

This simulation set examined whether asymmetries in the substitution rate of a hidden, unexamined, trait could cause the focal trait to be erroneously inferred as affecting diversification rates. The procedures for this simulation scenario were identical to those detailed for Simulation Set 1, but here the history of the trait that was used to generate the tree (and to produce the trait data that is used in the follow-up diversification analysis) was ignored. Instead, a second (hidden) trait that is neutral with respect to the diversification process, but affects the rate of sequence evolution, was simulated along the generated ultrametric trees and the history of this trait was used to scale the trees. This trait was simulated using the MK2 model (Lewis 2001) and $q = 0.1$, constraining the transition rates to be equal in both directions (from state 0 to state 1, and from state 1 to state 0). A representative history of the hidden trait was obtained using the make.simmap() function in the phytools R package (Revell 2012). This history was used to scale the tree according to the parameter $r \in \{1,2,4,8\}$.

Simulation Set 3: A single shift in sequence evolutionary rates (trait independent)

This simulation set aimed to test whether asymmetric rates of sequence evolution in a certain subtree of the phylogeny could lead to mistaken inference of shifts in diversification rates by trait-independent methods such as MEDUSA and BAMM. We followed the same procedure as in Simulation Set 1, but here, the r parameter represented the ratio between the substitution rate in a descendant subtree following a rate shift and the ancestral substitution rate. The location of the shift was set at a node that was the ancestor of 40% of the taxa in the phylogeny. If such a node did not exist, we chose the node that was the ancestor of the nearest number of taxa, and in any case did not exceed 45% of the total number of taxa. Once the shifted node was identified, the lengths of all branches descending from this node were multiplied by $r \in \{1,2,4,8\}$, and the procedures for generating the scaled tree were repeated as detailed above.

Simulation Set 4: Sequence evolutionary rates affected by a hidden trait (trait independent)

Unlike Simulation Set 3, which examined the scenario of a single transition in sequence evolutionary rates, in this simulation set we tested whether asymmetries in the substitution rate of a hidden, unexamined, trait, which could cause multiple transitions in sequence evolutionary rates, could lead to erroneous identification of shifts in diversification rates by trait-independent methods. Hence, we followed the same procedure as in Simulation Set 1, so that the trees were simulated simultaneously with a trait, and were rescaled according to r based on the character history. However, the trait data were ignored when shifts in diversification were inferred by the trait-independent methods MEDUSA and BAMM.

Simulation Set 5: Known trait affecting both speciation and substitution rates (trait dependent)

This simulation set examined potential biases in trait-dependent diversification methods when a focal trait affects both rates of sequence evolution and rates of speciation. Thus, this simulation set does not examine erroneous rejection of the null hypothesis, but rather statistical power. The general procedure was similar to Simulation Set 1, except that the speciation rate was not symmetrical under the two states. In Simulation Set 5a, we examined the scenario where the state with the higher rate of sequence evolution (state 1) also had a higher speciation rate ($\lambda_0 = 0.1, \quad \lambda_1 = 0.3$), whereas in Simulation Set 5b, the state with the lower rate of sequence evolution speciated more rapidly ($\lambda_0 = 0.3, \quad \lambda_1 = 0.1$). All other parameters remained unchanged. In both cases, the substitution rate of state 1 was higher than that of state 0 according to $r \in \{1,2,4,8\}$.

MULTIPLE SEQUENCE ALIGNMENT SIMULATIONS

The scaled trees were used to simulate multiple sequence alignments using INDELible (Fletcher and Yang 2009). The multiple sequence alignments were simulated under HKY+Gamma model with four rate categories and gamma shape parameter $\alpha = 0.5$. The base frequencies parameters were set to be the following: $f(\mathrm{A}) = 0.202$, $f(\mathrm{C}) = 0.202$, $f(\mathrm{G}) = 0.298$, and $f(\mathrm{T}) = 0.298$ (representing average nucleotide frequencies in mammals; Swindell et al. 2012). The transition versus transversion ratio, κ, was set to 2, and root sequence length was set to 500. Indels (insertion and deletion events) were drawn from a Zipfian distribution with maximum indel length (M) of 25 and the Zipfian parameter (a) set to 1.8. The indel rate parameter was set to 0.001 for 50 taxa

trees, and 0.0005 otherwise, such that the alignment length is at most twice as the root length. To consider the potential effect of sequence length on diversification analyses, we repeated the simulation procedure of Simulation Set 1 and 4 with the root sequence length set to 250 and 2000.

PHYLOGENY RECONSTRUCTION

The multiple sequence alignments created by Indelible were used to infer maximum likelihood (ML) and Bayesian trees. ML trees were reconstructed with PhyML (Guindon et al. 2010) using the GTR+I+G model (with four categories of the Gamma distribution). The outgroup species that was included in the simulation was used to root the tree and was then removed. Bayesian trees were inferred with MrBayes (Huelsenbeck and Ronquist 2001; Ronquist et al. 2012) using the strict clock model, as well as the autocorrelated and uncorrelated relaxed clock models. For each clock model, the trees were inferred using the GTR+I+G nucleotide substitution model with four rate categories for the discrete gamma distribution. Due to the intensive computing time, only the uncorrelated relaxed clock model was applied in Simulation Sets 2 and 5. The IGR relaxed clock model (Lepage et al. 2007) was used to account for clock variation across lineages in the relaxed uncorrelated model, and the TK02 autocorrelated model was used for the relaxed autocorrelated model (Thorne and Kishino 2002). The prior on the clock rate was set to have a lognormal distribution. Additionally, two fixed calibration points were set: the divergence time between the outgroup and all other taxa was set to 1.2, whereas the most recent common ancestor of all ingroup taxa was calibrated to 1. All ingroup taxa were constrained to form a monophyletic group. The Markov Chain Monte Carlo (MCMC) analysis was run for 2,500,000 generations with a sampling frequency of 5000. The burn-in fraction was set to 0.25, and the tree with the maximum log likelihood values was chosen as a point estimate for downstream analyses. We note that the Bayesian inferred trees are ultrametric, whereas the ML inferred trees are not, and thus require a time calibration procedure before being given as input to downstream diversification analyses.

The ML-inferred trees were time calibrated by two methods: PATHd8 (Britton et al. 2007) and penalized likelihood (Sanderson 2002). For the execution of PATHd8, we specified the sequence length as 500 and the root as a calibration point fixed to age 1. Penalized likelihood (PL) inferences were obtained using r8s version 1.81 (Sanderson 2003) with the Truncated Newton optimization. We fixed the age of the root to 1, the ftol parameter to 10^{-9}, the minRateFactor parameter to 0.005, and the minDurFactor parameter to 10^{-8}. r8s was executed twice: once to identify the optimal smoothing parameter over a range of possible values (from 0.0001 to 10,000 increasing in log scale of 10) using a cross validation procedure, and again with the optimal smoothing parameter that was chosen as the one with the minimal chi-square error. In both runs, the number of sites was set to the alignment length, with the artificially added outgroup species excluded from the sequence data.

TRAIT-DEPENDENT DIVERSIFICATION INFERENCE

Two BiSSE models were fitted to the time-calibrated trees using the original simulated tip states as the input trait data. As a null model, we used a four parameter BiSSE model, where the extinction and speciation rates under both states of the trait were constrained to be equal, whereas the transition rates were free to vary ($\lambda_1 = \lambda_0,\ \mu_1 = \mu_0,\ q_{10},\ q_{01}$). The alternative model contained six free parameters, where speciation, extinction, and transition rates were free to vary for each trait state ($\lambda_1, \lambda_0,\ \mu_1, \mu_0,\ q_{10},\ q_{01}$). Model selection was performed according to the AIC, such that the null hypothesis was rejected if the difference in AIC scores between the alternative and null models was at least 2. Results obtained using the likelihood ratio test and an α value of 0.05 were highly similar.

We also examined the impact of models that consider the possibility that hidden traits may affect diversification patterns. To this end, the CID-2 and HiSSE models (Beaulieu and O'Meara 2016) were fitted to each simulated dataset using the R package "hisse." In this case, four candidate models were considered: two neutral models and two state-dependent diversification models. The four-parameter BiSSE model ($\lambda_1 = \lambda_0,\ \mu_1 = \mu_0,\ q_{10},\ q_{01}$) and the CID-2 model (a CID model for character-independent diversification with a binary hidden trait) with three transition rates (equal transition rates between the two hidden traits and two asymmetrical transitions rates between the observed states) were considered as null models, whereas the six-parameter BiSSE model and the full HiSSE model (where the hidden states are allowed to alter diversification rates within each observed state) were regarded as alternative state-dependent diversification models. Model selection was performed according to the AIC criterion, such that the null hypothesis of no association between the trait and diversification rate was rejected if the difference in AIC scores between the best alternative model and best null model was at least 2 (i.e., $\Delta\text{AIC} > 2$ in support of the alternative models).

TRAIT-INDEPENDENT DIVERSIFICATION INFERENCE

For Simulation Sets 3 and 4, we used MEDUSA and BAMM to detect shifts in diversification rates across the phylogeny. MEDUSA analysis was performed using the MEDUSA package in R. As parameters for the MEDUSA function, we used the birth-death model, and allowed for unresolved trees, because there could be time-calibrated trees with zero branch lengths. The calibrated trees were given to MEDUSA as input, and if the best model inferred by MEDUSA contained one or more rate shifts, it was regarded as a false inference of shifts in diversification rates.

For BAMM analyses, we specified an MCMC run of 10,000,000 generations with sampling frequency of 5000. Model selection was performed according to the Bayes Factor (BF), such that the null model of zero shifts was rejected when the BF of a more complex model relative to the null model was greater than 1.0 (Mitchell and Rabosky 2017). Two different priors, with the expected number of shifts set to 1 or to 10, were initially examined. The selected models in both cases were nearly identical (99 out of 100 runs), as was previously reported (Mitchell and Rabosky 2017). Similarly, the estimates of speciation and extinction rates were highly correlated using the two different priors ($r^2 = 0.97$ and 0.92, respectively). We chose to use the broader prior of 10 expected shifts because the chain converged faster (reached effective sample size >200) in nearly all runs. However, in few cases (four trees out of 11,200 in Simulation Set 3, and nine trees out of 11,200 in Simulation Set 4) BAMM analysis with the broader prior did not complete within a reasonable amount of time (i.e., more than two weeks), and the narrower prior was used instead. Because trees with zero-branch lengths cannot be analyzed in BAMM, such branches were assigned to length 10^{-8}. Following this adjustment, the distances from the root to all tips were maintained equal by extending the lengths of terminal branches to match that of the tip with longest distance. As a control to this procedure, we applied MEDUSA on this set of adjusted phylogenies, before and after the modifications, and results were identical in all cases.

Results

KNOWN TRAIT AND MODEL COMPARISON (SIMULATION SET 1)

We used Simulation Set 1 to assess the impact of asymmetric rates of sequence evolution on the tendency of the BiSSE likelihood framework to incorrectly select models that include distinct speciation and extinction rates for each character state. To this end, we counted the number of simulations for which the alternative hypothesis (in which speciation and extinction rates depend on the state of the binary state) fitted significantly better than the null model, in which the speciation and extinction rates are constrained to be equal for the two states. First, we examined the results when $r = 1$ (i.e., the simulated trait does not affect the rate of sequence evolution) in five different types of phylogenies that were given as input to BiSSE: ML trees that were inferred with PhyML and then time calibrated using either one of the widely used methods PL and PATHd8 (hereafter referred to as ML-PL and ML-PATHd8, respectively); and the Bayesian trees inferred by MrBayes using the autocorrelated, uncorrelated, and strict clock models (hereafter referred to as Bayes-TK02, Bayes-IGR, and Bayes-strict, respectively). As a reference, we also examined the error rate when the time calibration step was performed directly on the true simulated trees (i.e., the trees that were used to simulate the sequence data). The proportion of simulations with incorrect inference of state-dependent diversification that was obtained with the Bayesian inferred trees as well as with two ML-ultrametrization methods was around the expected value of 0.05 (between 0.01 and 0.06), and similar to those obtained using the true trees (Table S1).

Next, we examined the error rate while increasing the effect of the character trait on the nucleotide substitution rate (i.e., higher values of the r parameter). Notably, the proportion of simulations in which the selected model incorrectly included trait-dependent diversification rates markedly increased with the simulated r parameter and when more taxa were simulated (Fig. 2). This pattern was especially evident for the ML-PATHd8 and the Bayesian trees inferred using the strict clock model and the IGR uncorrelated relaxed clock model, for which the error rate reached nearly 100% when 400 taxa were simulated. The high error rate for the strict-clock model was anticipated, because the assumption of molecular clock is violated in our simulation procedures with $r > 1$. For the Bayesian autocorrelated-TK02 trees, although the error rate generally increased with higher values of r, it was lower compared to the other time-calibration methods. In accordance with these results, the log likelihood difference between the null and alternative models steadily increased with higher values of the simulated r parameter (Fig. S1).

We further examined whether the spurious inferences of trait-dependent diversification are affected by the inclusion of models that account for the possibility that hidden traits influence diversification patterns. To this end, the CID-2 and full-HiSSE models (Beaulieu and O'Meara 2016) were considered as additional null and alternative models, respectively. Our results revealed that also in this case the error rate increased with higher values of r and with larger number of taxa, reaching high levels and, with some minor variations, similar to that obtained using BiSSE-only models (Fig. 2). These results were in accordance with the relative support obtained for each of the four models, as computed using Akaike weights (Fig. S2). In small trees and low values of r, the support of the full HiSSE model was lower compared to that obtained by CID-2 (although both were lower compared to their respective BiSSE models), such that the error rate was moderate. In most tree types, the average Akaike weight of full-HiSSE model increased with larger trees and with the value of the simulated r parameter, reaching nearly maximal support (and very high error rate) in trees with 200 or 400 taxa and $r \geq 4$. For the Bayes-IGR trees, however, the support for the full-HiSSE and CID-2 models was very small compared to the two BiSSE models, in accordance with the observed error rate that did not differ substantially upon inclusion of the HiSSE and CID-2 models.

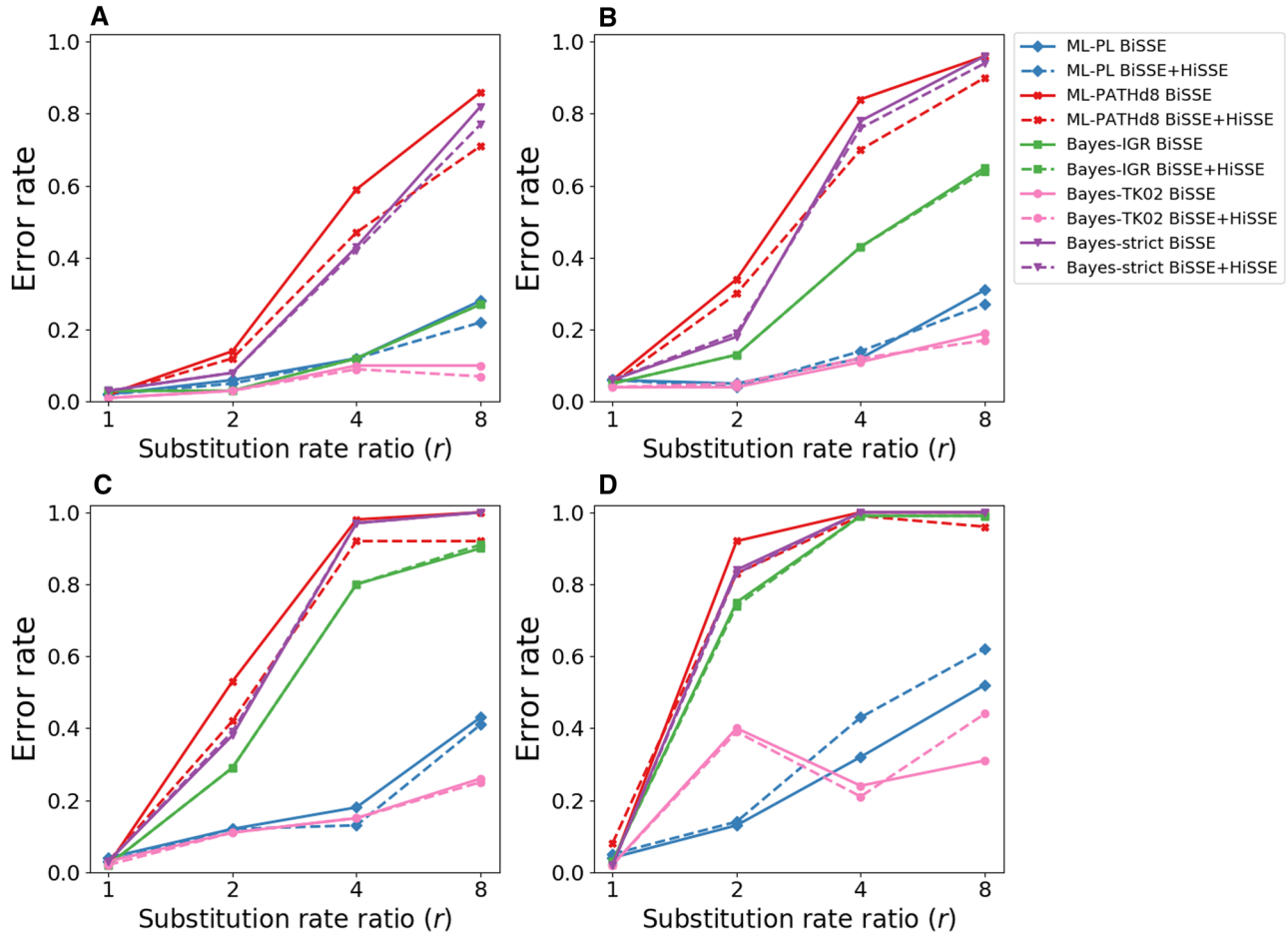


Figure 2. **False inference of trait-dependent diversification of BiSSE and HiSSE models when the focal trait of diversification analysis is associated with altered rates of sequence evolution (Simulation Set 1). Proportion of simulations where the neutral trait was erroneously associated with altered rates of diversification as a function of the substitution rate ratio (*r*) in simulations that include (A) 50, (B) 100, (C) 200, and (D) 400 taxa. The solid lines represent the results obtained for the two BiSSE models (null and alternative with four and six parameters, respectively), whereas the dashed lines when additionally including the HiSSE and CID-2 models. Results are presented for input trees inferred using: ML-PL (blue), ML-PATHd8 (red), Bayes-IGR (green), Bayes-TK02 (pink), and Bayes-strict (purple). In all cases, the null model was rejected using a ΔAIC threshold of 2.**

DISTINGUISHING BETWEEN ERRORS IN PHYLOGENY RECONSTRUCTION AND TIME-CALIBRATION ON FALSE POSITIVE INFERENCE OF DIVERSIFICATION SHIFTS (SIMULATION SET 1)

The high rates of the false inference of trait-dependent diversification reported in the analyses above were based on trees that were inferred from simulated sequence data. We next examined whether these results are driven by errors in tree reconstruction (altered tree topology or branch lengths) or by errors in the time-calibration procedure. To differentiate between these two possibilities, the time-calibration step was performed directly on the simulated trees, which were scaled according to the *r* parameter and then used to simulate the sequence data rather than on the ML-inferred trees. We refer to these trees as scaled-PL and scaled-PATHd8, corresponding to the two different calibration methods used. We observed very similar error rates when using the scaled trees compared to the ML inferred trees, both when comparing the null and alternative BiSSE models or when additionally including the CID-2 and full-HiSSE models in an expanded set of four models (Fig. S3). These results thus indicate that the main cause for the elevated error rate when inferring trait-dependent shifts in diversification rates is due to errors in the time-calibration procedure, rather than errors in the inference of the topology or in the estimation of branch lengths when these are inferred in units that are proportional to genetic distances.

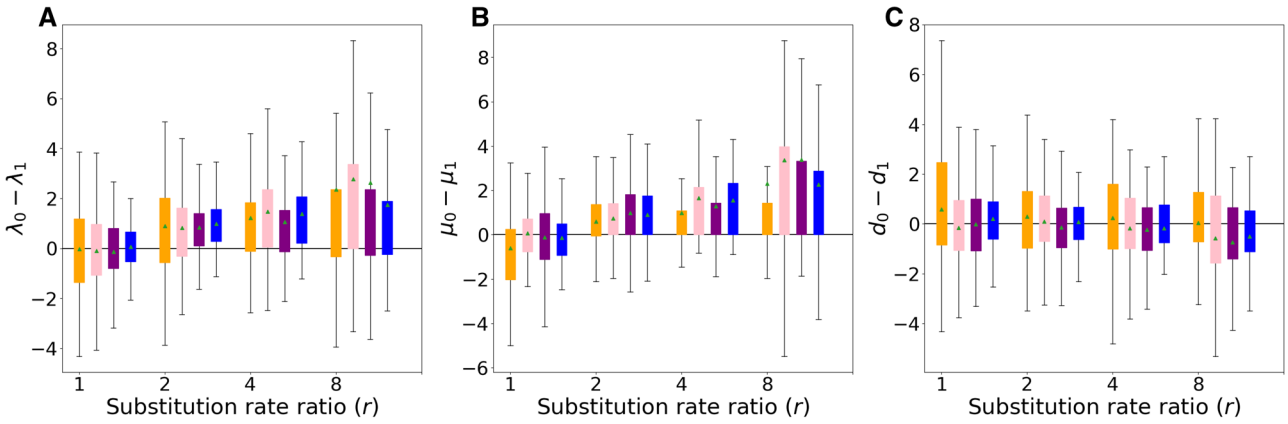


Figure 3. Bias in parameter estimation of BiSSE as a function of asymmetry in sequence evolution rate (Simulation Set 1). Box plots of the inferred (A) $\lambda_0 - \lambda_1$, (B) $\mu_0 - \mu_1$, and (C) $d_0 - d_1$ as a function of the simulated substitution rate ratio (r) parameter. Within each panel, results obtained using trees of 50, 100, 200, and 400 taxa are ordered from left to right, represented by orange, pink, purple, and blue, respectively. The triangle within each box denotes the average of 100 replicates. The horizontal line represents the true value. The results shown here correspond to those obtained using the ML-PL trees (see Figure S4 for results obtained using other methods).

KNOWN TRAIT AND DIVERSIFICATION RATES ESTIMATES (SIMULATION SET 1)

To asses which parameters are most affected by asymmetry in sequence evolutionary rates and whether there is a systematic bias, we evaluated the difference between the inferred speciation, extinction, and net diversification (speciation minus extinction) rates of the two binary states. Specifically, we tested the values of $\lambda_0 - \lambda_1$; $\mu_0 - \mu_1$; and $d_0 - d_1$, where d_0 and d_1 denote the net diversification rate under state 0 and state 1, respectively. For $r = 1$, the estimated λ_0 and λ_1 were very similar, such that the average $(\lambda_0 - \lambda_1)$ was close to 0, indicating no bias. For larger values of r, however, the difference between λ_0 and λ_1 became more substantial, where, on average, λ_0 was inferred to be larger than λ_1, indicating a bias of lower inferred speciation rate for the character state with higher rate of sequence evolution. This trend became more noticeable as the number of taxa increased (Fig. 3A). Inferences of the extinction rates showed very similar patterns, where the average estimates of $(\mu_0 - \mu_1)$ were close to zero for $r = 1$ and became larger with higher values of the simulated r parameter (Fig. 3B), again indicating a consistent bias for lower inferred extinction rate for the character state with the higher rate of sequence evolution. The abovementioned biases in the inferences of trait-dependent speciation and extinction rates seem to counterbalance each other when net diversification rates are concerned, as the average inference of $(d_0 - d_1)$ was around zero for nearly all taxa sets and simulated r values (Fig. 3C). The results reported above were based on the ML-PL tree reconstruction strategy, which were similar to those obtained using ML-PATHd8 and Bayes-strict trees, although the biases in the latter two were generally larger (Figs. S4A and S4B, respectively). The results obtained using the Bayes-IGR trees also showed elevated values of $(\lambda_0 - \lambda_1)$ with increasing values of r, but there was no clear pattern in the estimations of $(\mu_0 - \mu_1)$, so the bias in the inferred net diversification rates increased with higher values of r (Fig. S4C). In the Bayes-TK02 trees, the estimated difference in speciation, extinction, and net diversification rates did not demonstrate any particular trend for most values of r (Fig. S4D).

HIDDEN TRAIT, MODEL COMPARISONS, AND DIVERSIFICATION RATE ESTIMATES (SIMULATION SET 2)

We next examined whether the existence of a hidden trait (i.e., a trait that is not the focus of the diversification analysis) that is associated with altered rates of sequence evolution also lead to erroneous inference regarding the diversification patterns of a focal trait (Simulation Set 2). Our results revealed that as the bias in the sequence evolutionary rates of the hidden trait increased, so did the error rate of BiSSE. As in the previous simulation set, this pattern was apparent under all tree reconstruction strategies (ML-PL, ML-PATHd8, and Bayes-IGR) and was more noticeable when larger trees were simulated (Fig. 4). Yet, the proportion of false positive inferences was substantially lower compared to Simulation Set 1 where the focal trait itself was associated with altered rates of sequence evolution (compare Fig. 4 to Fig. 2). The results were qualitatively similar when models that account for a hidden trait (CID-2 and full-HiSSE models) were included in the set of examined models. For the smaller set of trees (with 50 and 100 taxa), the inclusion of these two models resulted in a shallow reduction in the spurious inference of trait-dependent diversification. However, in the case of trees with 200 and 400 taxa, the error rate was similar, and in some cases became even higher for the expanded set of models compared to using the two BiSSE models only. In all examined cases, the error rates were very similar whether the input trees were time calibrated given

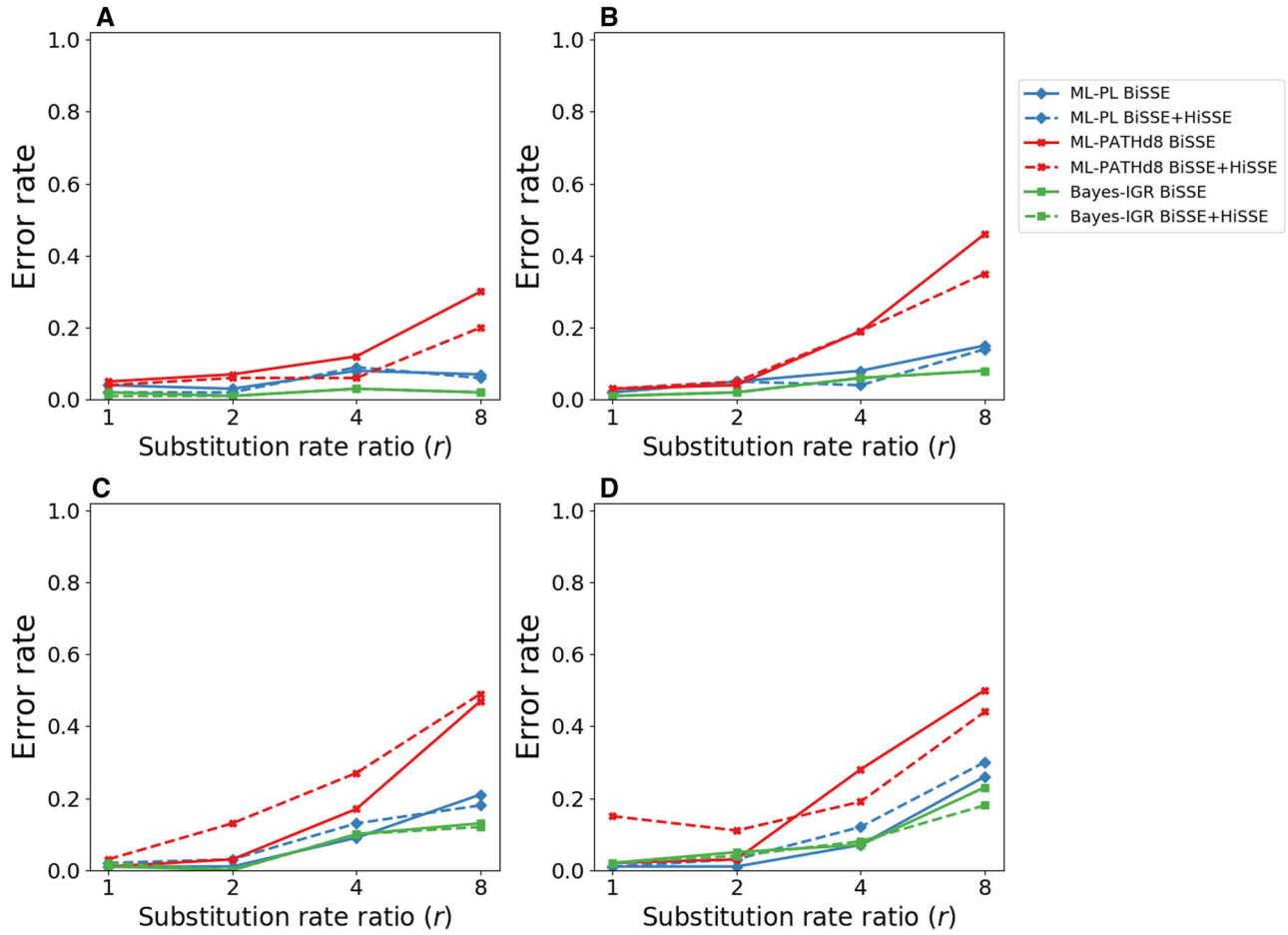


Figure 4. **The rate of false inference of trait-dependent diversification of BiSSE and HiSSE models when a hidden trait is associated with altered rates of sequence evolution (Simulation Set 2). Proportion of simulations in which the null model was rejected as a function of the substitution rate ratio (*r*) in (A) 50, (B) 100, (C) 200, and (D) 400 taxa sets. Solid lines represent the results obtained for the two BiSSE models (null and alternative with four and six parameters, respectively), and the dashed lines additionally include the HiSSE and CID-2 models. Results are presented for input trees inferred using: ML-PL (blue), ML-PATHd8 (red), and Bayes-IGR (green). In all cases, the null model was rejected using a Δ*AIC* threshold of 2.**

the simulated (scaled) trees, or those reconstructed based on simulated sequence data (Fig. S5), indicating again that the major source of the error is in the time-calibration procedure. Examination of parameter estimates revealed no bias toward a particular character state. This is expected because the hidden trait that affects rates of sequence evolution is disconnected from the focal trait (see Fig. S6 for ML-PL trees).

THE INFERENCE OF SHIFTS IN DIVERSIFICATION RATES WITH NO TRAIT DATA (SIMULATION SET 3 AND 4)

The performance of trait-independent diversification analyses, represented by the popular methods MEDUSA and BAMM, was first evaluated when shifts in the substitution rate occur independently of a trait (Simulation Set 3). To this end, we evaluated a simple scenario in which a single subtree of a phylogeny exhibits a shift in the nucleotide substitution rate, which is r times higher compared to the rest of the phylogeny. In the case of $r = 1$ (i.e., representing no shift in the substitution rate in any clade), the proportion of simulations with falsely inferred shifts in diversification rates obtained using MEDUSA was around the expected value of 0.05 and was always below 0.05 in the case of BAMM. For both methods, the error rate increased with the simulated r parameter and with the number of taxa (Fig. 5), reaching high values (0.3-1.0) when large trees were simulated. Additionally, across all simulation scenarios the error rate of BAMM was lower than that of MEDUSA.

To evaluate which reconstruction steps (tree inference or time calibration) underlie the high rate of false inference of shifts in diversification rates, we also evaluated the performance of MEDUSA and BAMM when the scaled trees were time calibrated using PL and PATHd8 methods (termed "scaled-PL" and

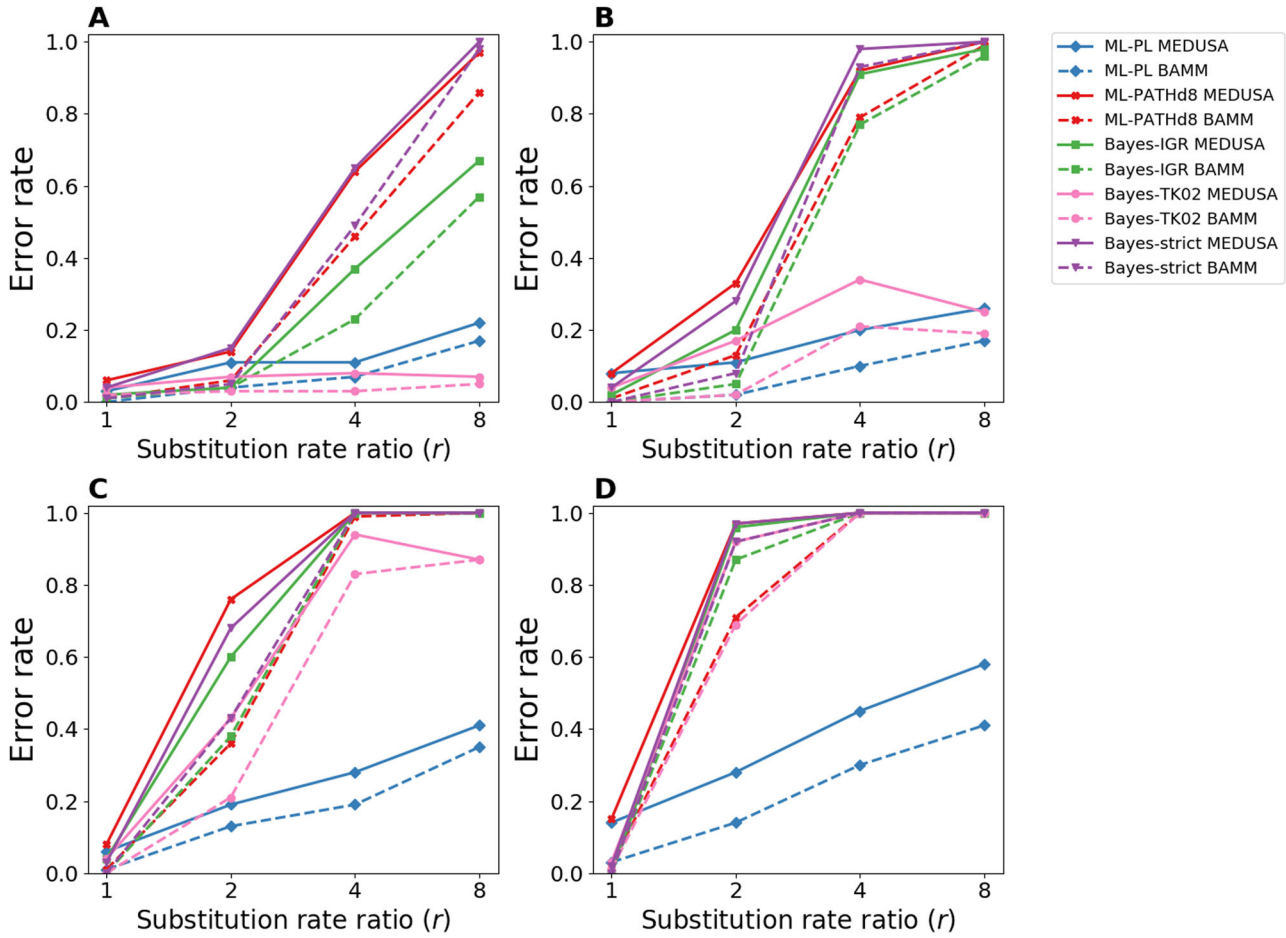


Figure 5. False inference of shifts in diversification rates using MEDUSA and BAMM when a shift in sequence evolution occurs in a certain subclade of a phylogeny (Simulation Set 3). Proportion of simulations where the diversification analysis erroneously inferred shifts in diversification rates as a function of the substitution rate ratio (*r*) in (A) 50, (B) 100, (C) 200, and (D) 400 taxa sets, using different time-calibration methods. The solid lines represent the results obtained using MEDUSA, whereas the dashed lines represent the results obtained by BAMM. The different colors represent different time-calibration methods: ML-PL (blue), ML-PATHd8 (red), Bayes-IGR (green), Bayes-TK02 (pink), and Bayes-strict (purple).

"scaled-PATHd8," respectively) and given as input, rather than the trees that were reconstructed from the simulated sequence data. A similar trend of higher incorrect identification of shifts in diversification with more data and with higher values of the simulated *r* parameter was obtained (Figs. S7 and S8). However, the error rate of the scaled trees was substantially lower compared to the reconstructed trees when time calibration was performed using PL, whereas for PATHd8 it was similar across the two sets of trees.

We next evaluated the performance of MEDUSA and BAMM on datasets in which a shift in the rate of sequence evolution is associated with a transition between organismal traits (Simulation Set 4). Here again, the rate of falsely inferred shifts in diversification rates increased with increasing values of *r* and in larger trees (Fig. 6), although the error rate was generally lower compared to that of Simulation Set 3, particularly for Bayes-IGR trees whose error rate was only slightly higher than the expected 0.05.

KNOWN TRAIT AND TRUE SPECIATION RATE DIFFERENCES (SIMULATION SETS 5a AND 5b)

All simulations described above concentrated on the effect of shifts in the substitution rate on methods that detect shifts in the diversification process in phylogenies that evolve under a time-homogenous diversification process. In such cases, rejection of the null hypothesis leads to an incorrect conclusion of trait-dependent diversification. However, based on the results obtained above, it is expected that in phylogenies that truly evolve under a trait-dependent heterogeneous diversification pattern, alterations in the rate of sequence evolution can either strengthen or weaken any potential diversification signal. To test this prediction, two sets of simulations were examined. In Simulation Set

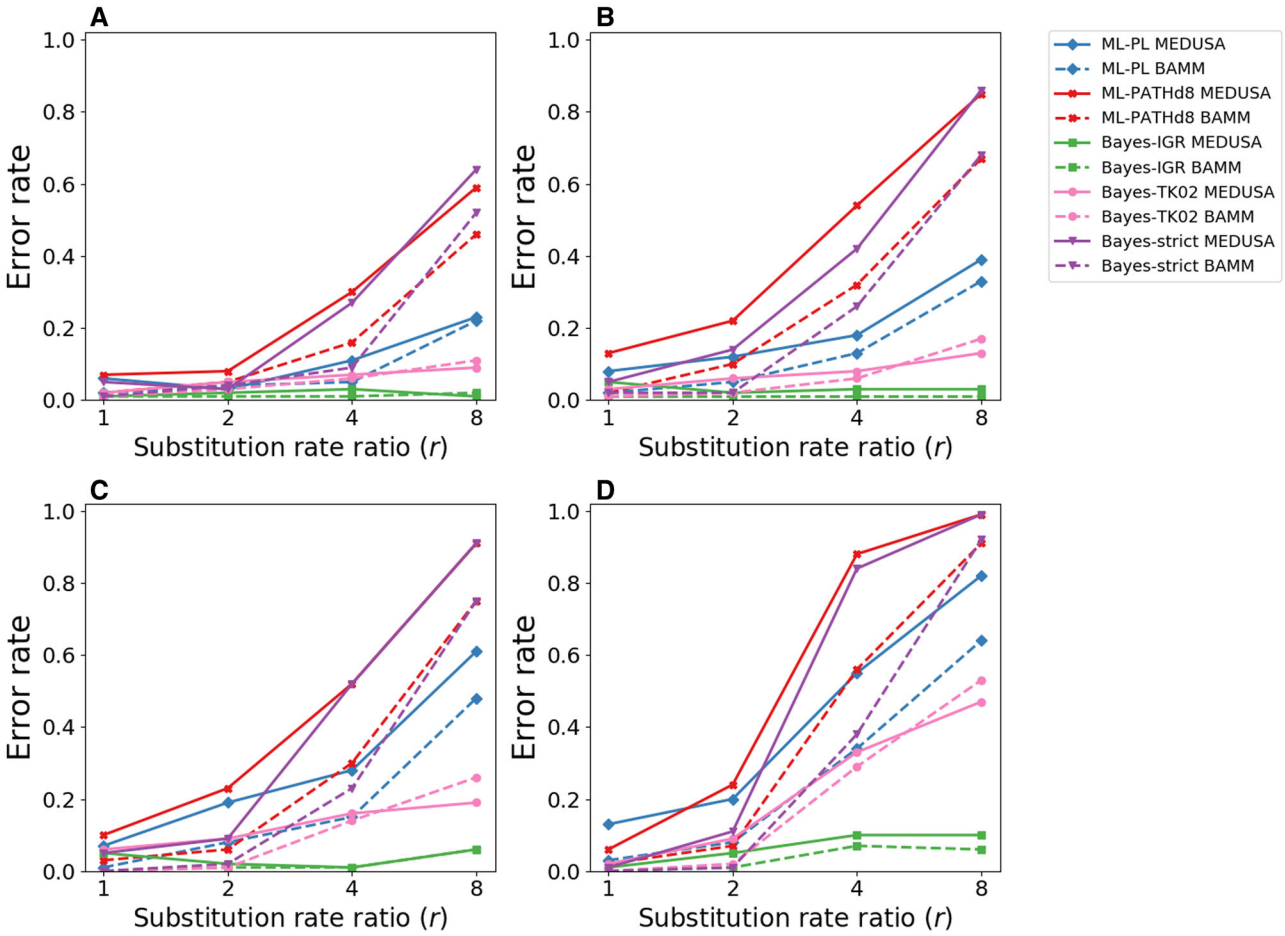


Figure 6. False inference of shifts in diversification rates using MEDUSA and BAMM when the asymmetry in sequence evolution rates is trait dependent (Simulation Set 4). Proportion of simulations where the diversification analyses erroneously inferred shifts in diversification rates as a function of the substitution rate ratio (*r*) in (A) 50, (B) 100, (C) 200, and (D) 400 taxa sets, using different time-calibration methods. The solid lines represent the results obtained using MEDUSA, whereas the dashed lines represent the results obtained by BAMM. The different colors represent different time-calibration methods: ML-PL (blue), ML-PATHd8 (red), Bayes-IGR (green), Bayes-TK02 (pink), and Bayes-strict (purple).

5a, the state with the higher speciation rate was also associated with higher rates of sequence evolution, whereas Simulation Set 5b examined the opposite scenario (i.e., $\lambda_1 > \lambda_0$ in Simulation Set 5a, and $\lambda_0 > \lambda_1$ in Simulation Set 5b; in both cases the substitution rate of state 1 was r times higher than that of state 0, $r \in \{1,2,4,8\}$). When no asymmetry in sequence evolution was simulated ($r = 1$), the power of BiSSE to correctly reject the null hypothesis was relatively similar for both simulation scenarios. However, the effect of increasing r values was markedly different for the two simulation types (Fig. 7). Namely, in Simulation Set 5a the power of BiSSE to correctly detect the faster speciation state steadily decreased with increasing r values, whereas in Simulation Set 5b the power sharply increased. Moreover, in Simulation Set 5a and when the bias in sequence evolution rate was high ($r = 8$), not only did we observe low statistical power to reject the null hypothesis, but in certain cases the inferred speciation rates pointed at the opposite direction than the true values (i.e., the null model was rejected but $\lambda_0$was erroneously inferred to be larger than λ_1).

THE EFFECT OF SEQUENCE LENGTH ON THE FALSE POSITIVE RATE IN DIVERSIFICATION RATES ANALYSES (SIMULATION SETS 1 AND 4)

We examined whether longer sequences would result in more accurate tree inference and consequently lower tendency of inferring shifts in diversification rates. We thus simulated datasets with sequence lengths of 250, 500, and 2000 positions and examined the error rate for both trait-dependent and trait-independent diversification rates. We also computed the branch score (Bs) tree distance (Kuhner and Felsenstein 1994, as implemented in the phangorn R package; Schliep 2011), between the inferred and true trees (i.e., those that were generated based on birth-death

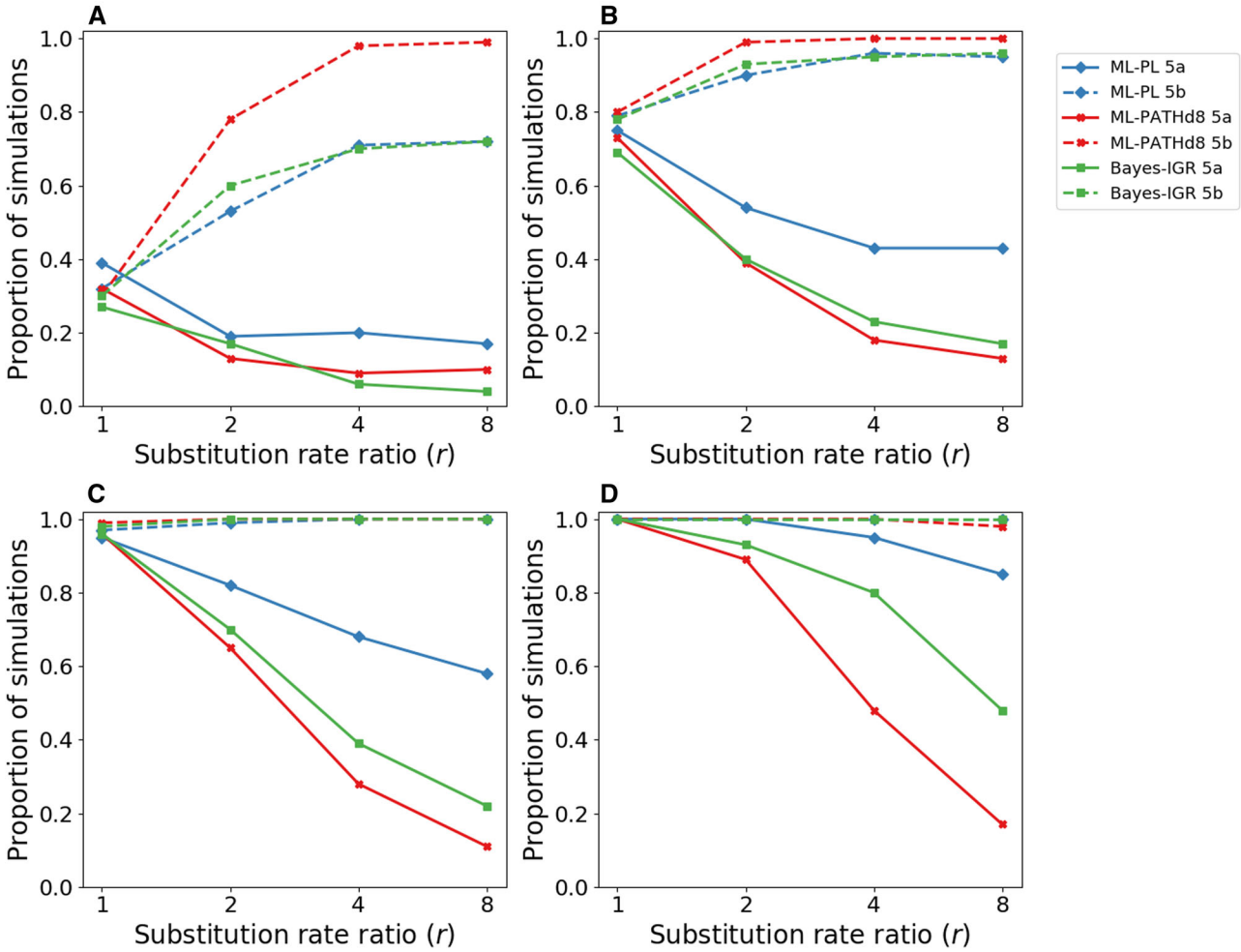


Figure 7. Power of BiSSE when the faster evolving state has higher (Simulation Set 5a) or lower (Simulation Set 5b) speciation rates. Proportion of simulations in which the null model was correctly rejected as a function of the rate substitution ratio (*r*) in (A) 50, (B) 100, (C) 200, and (D) 400 taxa sets. The solid and dashed lines represent the results obtained when the state with the higher rate of sequence evolution had a higher (Simulation Set 5a) and lower (Simulation Set 5b) speciation rate, respectively. Different colors correspond to different tree reconstruction strategies: ML-PL (blue), ML-PATHd8 (red), and Bayes-IGR (green). In all cases, the null model was rejected using a ΔAIC threshold of 2 and the inference of $(\lambda_0 - \lambda_1)$ was in the correct direction (i.e., in Simulation Set 5a, datasets in which the null model was correctly rejected but $\lambda_1$was erroneously inferred to be smaller than $\lambda_0$were not counted as true positive).

models in the initial step of the simulation). As might be expected, the distance between the true and inferred trees increased for trees inferred from shorter sequences (Fig. S9). In accordance, for both types of diversification inference methods, shorter alignments consistently led to higher false positive rate. We further compared the results of the ML-PL trees to those of the scaled-PL, where the time-calibration is applied directly on the simulated trees, thus representing the case of perfectly inferred trees from sequence data. The results demonstrated that the error rates for trees inferred from longer sequences were closer to the error rate observed in the scaled-PL trees, reaching almost identical results for sequence length of 2000, whose distances from the true trees were the lowest.

Discussion

Detecting shifts in the diversification process has significant implications for the study of adaptive radiation and species selection, and may contribute to the efforts of conserving the biodiversity on Earth. To this end, probabilistic trait-dependent and trait-independent phylogenetic methods are commonly applied to study shifts in the diversification pattern. In this study, we speculated that asymmetric rates of sequence evolution lead to systematic biases in the reconstruction of time-calibrated trees, which in turn could lead to erroneous detection of trait-dependent or trait-independent shift in the diversification rate.

To assess the influence of asymmetric rates of sequence evolution on diversification analyses, we simulated phylogenies that are neutral with respect to the diversification pattern but that exhibit shifts in the rate of nucleotide substitutions. Our results demonstrate that as the heterogeneity in the rate of sequence evolution along a phylogeny increases, so does the tendency of various types of diversification analyses to reject the neutral model of homogenous diversification pattern and to incorrectly select more complex models that include shifts in the diversification process. This was demonstrated in the following scenarios. First, for trait-dependent methods, in cases where the focal trait of the diversification analysis is associated with altered rates of sequence evolution. Second, for trait-dependent methods, whenever there exists a trait (unexamined or hidden) that affects the rate of sequence evolution in the examined phylogeny. Third, for trait-independent inference methods, whenever a subclade in the examined phylogeny exhibits a shift in the rate of sequence evolution. Fourth, for trait-independent inference methods, whenever there exists a trait that affects the rate of sequence evolution in the examined phylogeny. Obviously, the examined scenarios are a small subset of all possible ways in which shifts in the molecular substitution rates can be modeled. However, the fact that all examined scenarios revealed major biases in the inference of shifts in the diversification pattern points to a more general conclusion: robust inferences of rates of lineage diversification should consider more closely possible biases in the rate of molecular evolution within the clade under study. This conclusion is further strengthened by an additional set of simulations that demonstrated that the statistical power to correctly infer state-dependent diversification shifts is reduced (or elevated) when the state with the higher speciation rate is also the one with higher (or lower) rate of nucleotide substitution.

Additionally, we demonstrated that the rate of false inferences of shifts in diversification rates increases with the size of the phylogeny (i.e., number of taxa). This result raises concern for two main reasons. First, as the size of the phylogeny increases, so does the phenotypic diversity exhibited within the analyzed clade, increasing the probability for genuine differences in the molecular substitution rate. Second, the statistical power and accuracy of BiSSE were shown to be rather low for small clades (containing less than 200 species), whereas results obtained using larger clades are considered more trustworthy (Davis et al. 2013). The effect of the data size observed here suggests that results obtained using such large trees should also be treated cautiously, unless heterogeneity in the substitution rate is ruled out.

Interestingly, our results regarding the accuracy of the inferred diversification rates parameters demonstrated that, in most cases, the state with the higher substitution rate tends to be falsely associated with lower speciation rate, as well as with lower extinction rate (Fig. 3). Indeed, and as previously hypothesized (Smith and Donoghue 2008; Beaulieu and O'Meara 2016), the longer branch lengths emitted by the state with the higher rate of substitution cause these lineages to appear older than they actually are, biasing estimates of diversification rates. However, because the inference of both speciation and extinction rates were biased in the same direction, they tended to cancel out as far as the inference of net diversification rates is concerned. These results indicate that the comparison of net diversification rates provides more robust inferences than teasing apart the individual effect of speciation and extinction as well as a more robust alternative for hypothesis testing. Still, under some tree reconstruction strategies (e.g., the Bayesian uncorrelated IGR clock model), this pattern was not observed as both the difference in speciation rates ($\lambda_0 - \lambda_1$) and net diversification rates ($d_0 - d_1$) between the two states tended to increase with the extent of substitution rate heterogeneity.

Several studies have previously demonstrated that errors in divergence time estimation could deteriorate the inference of diversification rates. For instance, Marin and Hedges (2018) demonstrated that undersampling of sequence sites can lead to artificially increased speciation rate toward the present due to underestimation of more recent branching times. Duchêne et al. (2017) examined the possible effect of substitution rate heterogeneity on the inference of diversification rates, and demonstrated that positive association between the two processes reduces the statistical power to detect speedup or slowdown in diversification. On the other hand, Sarver et al. (2019) found that although the accuracy of estimating the global diversification rate decreases with increasing substitution rate heterogeneity, the resulting estimates are still reasonable unless the tree and clock models deviate substantially from the generating process. Notably, in these studies the examined rate variation was simulated stochastically with no association to a biological trait, which better fit the assumptions of the clock model used for molecular dating. As noted by Bromham et al. (2018), however, the currently implemented relaxed clock models might not align well with biological reality because most of them assume random fluctuations in molecular evolutionary rates, although in many cases rate heterogeneity is expected to change directionally according to a gradual change of some biological trait. In the present study, we simulated scenarios where rate heterogeneity was linked to trait evolution, and our results demonstrated a significant impact of heterogeneity in sequence evolutionary rates on diversification rate estimates, thus suggesting that the existing methods for molecular dating are not able to capture such kinds of rate variation.

Our results here suggest that errors in the inferred divergence times impact the accuracy of diversification rate estimates. In most simulation scenarios, the tendencies to select models with multiple diversification regimes were similar when the time

calibration step was performed on trees that were inferred from the simulated sequence data using ML techniques or directly given the simulated trees (termed here, "scaled trees"). In fact, the latter can be viewed as the phylogenies that accurately represent the genetic distances among the simulated taxa (and thus the best possible ML inferred phylogeny). This suggests that the main cause for the elevated error rate for both trait-dependent and trait-independent diversification methods is due to errors in the time-calibration procedure. These findings were consistent through all the examined simulation scenarios, except for Simulation Set 3, where the error rate was substantially lower for the scaled trees compared to the ML inferred trees, indicating that in this case inaccuracies in the reconstructed phylogeny, prior to time calibration, also contribute to spurious inferences of altered diversification rates. It is possible that such a scenario, where rate variation occurs along a single branch, could be accommodated fairly accurately by the penalized likelihood model (Sanderson 2002), because the assumption that substitution rates are correlated across adjacent branches holds across nearly all branches of the phylogeny.

To further examine whether errors in divergence time estimation are the main drivers of erroneous detection of shifts in diversification patterns, we examined the impact of sequence evolutionary rate heterogeneity on the accuracy of phylogeny inference, as measured by the Bs distance (Kuhner and Felsenstein 1994). We observed that increased rate heterogeneity led to larger errors in phylogeny reconstruction, which were reflected in increasing Bs distance as a function of *r*. This trend was common to all the examined time-calibration methods and all simulation scenarios (Fig. S10). Nevertheless, there was not a consistent correspondence between the elevated Bs distance and the error rate of the diversification analyses. For example, in Simulation Sets 1 and 4, the Bs distance for the Bayes-strict trees was among the lowest compared to the other examined methods clock but the error rate was among the highest in both trait-dependent and trait-independent analyses. Similarly, although the Bs distance for the Bayes-IGR trees was generally the largest, in MEDUSA/BAMM analyses (Simulation Set 4) they demonstrated the lowest error rate of incorrectly inferred shifts in diversification rates, whereas the same trees showed relatively high rates of error in the trait-dependent diversification analyses (Simulation Set 1). These complexities indicate that further examination is required to uncover the impact of errors in phylogeny inference, and divergence time estimation in particular, on diversification analyses.

It is important to note that although errors in tree reconstruction may impact all phylogenetic analyses, they are particularly important for methods that infer diversification rates. As noted by Beaulieu and O'Meara (2016), trait-dependent diversification methods maximize the joint probability of the observed tip states and the observed tree, given the model. Thus, SSE models are not only models for trait evolution, but are rather joint models for trait evolution and the phylogeny. Consequently, even if the tip states are consistent with a given transition model, the model still may not fit well if the given tree violates some of the underlying model assumptions. Under such modeling settings, the accuracy of the given time-calibrated phylogeny is of great importance to the performance of trait-dependent diversification analyses. The same rational holds for trait-independent methods, such as BAMM and MEDUSA, which tightly rely on the input tree in their likelihood calculations. Unfortunately, accurate estimation of dated phylogenies is a highly challenging task, fraught with uncertainties (Graur and Martin 2004; Brown and Smith 2018), and the assumption that the time-calibrated tree is absolutely correct is usually violated. Therefore, considering the abundance of evolutionary studies that focus on diversification rate estimates and the ease by which dated phylogenies can be obtained, a big step forward could be the development of methods that would inform us on the reliability of these inferences.

In our simulations, we applied several methods for diversification inference that are at present among the most widely used. Recently, a new method was proposed for the inference of trait-independent diversification shifts (Höhna et al. 2019), implemented in the RevBayes environment (Hohna et al. 2016). This method is based on a numerical integration approach using stochastic mapping and, in contrast to MEDUSA and BAMM, incorporates diversification rate heterogeneity of extinct lineages into the likelihood computations. Because this method is computationally highly demanding, we were not able to apply it on our extensive simulation testing. However, in an initial exploration of a subset of trees consisting of 50 taxa we obtained erroneous inference of diversification rate shifts in all examined cases. This incorrect inference could stem from prior sensitivity, which was already reported by Höhna et al. (2019). A more accurate inference would apply the stepping stone procedure to estimate the BFs for the number of rate shifts, but this was not attempted here due to the exceedingly long running times.

The results of this study enforce the challenge in studying the linkage between rates of genome evolution and lineage diversification. For example, it has been hypothesized that bursts of genomic change ("punctuated evolution") are correlated temporarily with speciation events, with elevated rates of genomic change either driving speciation or concomitant with other changes that occur during the speciation process. For example, cladogenesis may increase sequence evolutionary rates due to increased rate of adaptive evolution in genomic regions associated with local adaptation (Losos et al. 1997), or by elevating the genome wide substitution rate due to the more rapid accumulation of slightly deleterious mutations in small isolated populations (Ohta 1973; Venditti and Pagel 2010). On the other hand, higher substitution rate may lead to faster reproductive isolation (Orr and Turelli

2001), or to higher standing variation that increases local adaptations (Nabholz et al. 2008b; Nosil et al. 2009). Indeed, several studies have used phylogenetic and genomic data to examine the link between molecular evolutionary rates and the formation rate of new species (Barraclough and Savolainen 2001; Jobson and Albert 2002; Pagel et al. 2006; Lancaster 2010; Lanfear et al. 2010; Venditti and Pagel 2010). However, despite the variety of studies concluding that there is a correlation between speciation and molecular evolutionary rates, there is no evidence that this correlation exists in all groups of species and that it is not confounded with other factors (Pennell et al. 2014). According to our study, however, when such correlation exists, a genuine association between these two factors is tricky to identify because elevation in the substitution rates could quickly lead to misidentification of the trait as associated with diversification rates (as demonstrated in Simulation Set 5a), which was also found by Duchêne et al. (2017). As a case in point, Rolland et al. (2014) applied the GeoSSE model (Goldberg et al. 2011) and demonstrated that most mammalian orders exhibit higher diversification rates in the tropics compared to temperate regions but this pattern was not observed in other mammalian orders, such as Carnivora, where both the speciation and extinction rates were higher in temperate regions. However, elevated substitution rates of tropical carnivorous lineages, as reported by Gillman et al. (2009, 2010), could weaken a potential signal of high diversification of these lineages, and thus future investigation should interpret results in light of possible associations between these two processes.

As shown here, trait-dependent heterogeneity in the substitution rates tends to bias inferred node times in such a way that the slowly evolving state is erroneously inferred to speciate faster and to get extinct at higher rates compared to the fast-evolving state. However, it is plausible that this effect is not unidirectional. Namely, when the dated phylogeny is inferred from a multiple sequence alignment, asymmetry in diversification rates of a trait under study may bias inferences of trait-dependent rates of molecular evolution (e.g., interpreting the state with the higher diversification rate as being associated with lower substitution rate). To investigate such a possibility, we simulated trees and sequence data under varying levels of speciation rate asymmetry dictated by a binary trait and inferred trait-dependent rates of substitution using the traitRate program (Mayrose and Otto 2011). Our preliminary analysis did not reveal elevated type 1 error rates for increasing levels of speciation rate asymmetry (results not shown) and thus we believe such an effect is not substantial.

Recently, Louca and Pennell (2019) argued that speciation and extinction rates per se are meaningless to estimate, especially if time heterogeneity is assumed, because there are numerous parameter combinations that can give rise to the same branch lengths distribution. These authors suggested the use of the "pulled speciation rate" and the "pulled diversification rate" (λ_p and r_p, respectively) as more meaningful statistics. Although these findings raise serious concerns regarding the validity of most currently existing methods for diversification analyses, a promising direction would be to estimate λ_p and r_p with relation to an evolving trait (i.e., in an SSE setting) or in a clade-specific manner. Our study, however, suggests that such attempts could suffer from the same shortcomings as those revealed here. Certainly, this is an important direction for future research.

One promising direction to alleviate the artifacts of asymmetry in molecular evolutionary rates would be to account directly for trait-specific rates of sequence evolution during phylogenetic inference, for example, within a general relaxed clock model. In this case, the substitution rate of each branch should be drawn from a distribution that depends on the state of the trait along that branch. An expanded probabilistic model could further account for the possible effect of the trait on diversification rates, thus simultaneously reconstructing the phylogeny and providing estimates of diversification rates within a single likelihood framework.

Until this is accomplished, it seems that as a prerequisite for the analysis of diversification rate shifts, one should test for heterogeneity in the rate of molecular substitutions along the examined phylogeny, whether associated with some trait (e.g., Lartillot and Poujol 2011; Mayrose and Otto 2011) or not (e.g., Whelan et al. 2011). A parallel concern is related to the analysis of large phylogenies, in which the assumption of a constant molecular clock is less likely to hold (Sanderson 2002; Welch and Bromham 2005; Britton et al. 2007; Quental and Marshall 2010; Tamura et al. 2012; Simpson et al. 2018). In this case, it could be beneficial to split the phylogeny into several (nonoverlapping) subclades, which are more likely to follow the clock hypothesis. When several clades are examined in parallel, it is less likely that an unaccounted-for trait would bias the inferred phylogenies in a similar manner (which would cause trait-dependent methods to be biased toward higher diversification rates of the same character state). Importantly, when multiple clades are examined, it is also important to construct the null distribution of the relative diversification rates expected for such a meta-analysis, because this might deviate from a symmetrical Gaussian distribution (e.g., as was found by Sabath et al. [2016] when examining the association between diversification rates and the plant sexual system) and to additionally consider the possible effect of alterations in the rate of sequence evolution. Obviously, such an approach represents a heuristic bypass to the problem discussed throughout this study and loses information on the relationships among the subclades. We suggest that directly incorporating trait information when modeling among-lineage heterogeneity in the sequence evolutionary rate has great potential to increase the accuracy of phylogeny reconstruction, particularly in the estimation of divergence times, and should improve the performance of many analy-

ses that rely on the reconstructed phylogeny, for which the inference of shifts in diversification rates is one prominent example.

AUTHOR CONTRIBUTIONS

All authors designed the study, helped in interpreting the results, and provided inputs on the draft. AS performed the analyses. AS and IM drafted the manuscript. IM supervised the study.

ACKNOWLEDGMENTS

We thank the associated editor, M. W. Pennell, and an anonymous reviewer for providing many thoughtful and helpful reviews. This study was supported in part by the Binational Science Foundation (BSF-2013286 to IM and EEG), and by National Science Foundation DEB (1655478/1940868 to EEG) and National Science Foundation-BSF (1655478 to EEG and IM) and by PhD fellowships from The Council for Higher Education program for excellent PhD students and from the Fast & Direct Ph.D Program by the Argentinean Friends of TAU to DA, and from the Edmond J. Safra Center for Bioinformatics at Tel-Aviv University to AS.

DATA ARCHIVING

All simulated data and scripts to reproduce results will be deposited in the Dryad repository: https://doi.org/10.5061/dryad.37pvmcvgp.

LITERATURE CITED

Alfaro, M. E., F. Santini, C. Brock, H. Alamillo, A. Dornburg, D. L. Rabosky, G. Carnevale, and L. J. Harmon. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. Proc. Natl. Acad. Sci. 106:13410–13414.

Baer, C. F., M. M. Miyamoto, and D. R. Denver. 2007. Mutation rate variation in multicellular eukaryotes: causes and consequences. Nat. Rev. Genet. 8:619–631.

Barraclough, T. G., and V. Savolainen. 2001. Evolutionary rates and species diversity in flowering plants. Evolution 55:677–683.

Beaulieu, J. M., and B. C. O'Meara. 2016. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. Syst. Biol. 65:583–601.

Beaulieu, J. M., B. C. O'Meara, P. Crane, and M. J. Donoghue. 2015. Heterogeneous rates of molecular evolution and diversification could explain the Triassic age estimate for angiosperms. Syst. Biol. 64:869–878.

Britton, T., C. L. Anderson, D. Jacquet, S. Lundqvist, and K. Bremer. 2007. Estimating divergence times in large phylogenetic trees. Syst. Biol. 56:741–752.

Bromham, L. 2009. Why do species vary in their rate of molecular evolution? Biol. Lett. 5:401–404.

———. 2011. The genome as a life-history character: why rate of molecular evolution varies between mammal species. Philos. Trans. R. Soc. B Biol. Sci. 366:2503–2513.

Bromham, L., S. Duchêne, X. Hua, A. M. Ritchie, D. A. Duchêne, and S. Y. W. Ho. 2018. Bayesian molecular dating: opening up the black box. Biol. Rev. 93:1165–1191.

Brown, J. W., and S. A. Smith. 2018. The past sure is tense: on interpreting phylogenetic divergence time estimates. Syst. Biol. 67:340–353.

Buschiazzo, E., C. Ritland, J. Bohlmann, and K. Ritland. 2012. Slow but not low: genomic comparisons reveal slower evolutionary rate and higher dN/dS in conifers compared to angiosperms. BMC Evol. Biol. 12:8.

Condamine, F. L., N. S. Nagalingum, C. R. Marshall, and H. Morlon. 2015. Origin and diversification of living cycads: a cautionary tale on the impact of the branching process prior in Bayesian molecular dating. BMC Evol. Biol. 15:65.

Crisp, M. D., N. B. Hardy, and L. G. Cook. 2014. Clock model makes a large difference to age estimates of long-stemmed clades with no internal calibration: a test using Australian grasstrees. BMC Evol. Biol. 14:263–279.

Davis, M. P., P. E. Midford, and W. Maddison. 2013. Exploring power and parameter estimation of the BiSSE method for analyzing species diversification. BMC Evol. Biol. 13:38.

Dornburg, A., M. C. Brandley, M. R. McGowen, and T. J. Near. 2012. Relaxed clocks and inferences of heterogeneous patterns of nucleotide substitution and divergence time estimates across whales and dolphins (Mammalia: Cetacea). Mol. Biol. Evol. 29:721–736.

Dos Reis, M., Y. Thawornwattana, K. Angelis, M. J. Telford, P. C. J. Donoghue, and Z. Yang. 2015. Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. Curr. Biol. 25:2939–2950.

Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4:e88.

Duchêne, D. A., X. Hua, and L. Bromham. 2017. Phylogenetic estimates of diversification rate are affected by molecular rate variation. J. Evol. Biol. 30:1884–1897.

Duchêne, S., R. Lanfear, and S. Y. W. Ho. 2014. The impact of calibration and clock-model choice on molecular estimates of divergence times. Mol. Phylogenet. Evol. 78:277–289.

Duchêne, S., K. E. Holt, F. X. Weill, S. Le Hello, J. Hawkey, D. J. Edwards, M. Fourment, and E. C. Holmes. 2016. Genome-scale rates of evolutionary change in bacteria. Microb. Genom. 2:1–12.

Fitzjohn, R. G. 2010. Quantitative traits and diversification. Syst. Biol. 59:619–633.

———. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. Methods Ecol. Evol. 3:1084–1092.

Fletcher, W., and Z. Yang. 2009. INDELible: a flexible simulator of biological sequence evolution. Mol. Biol. Evol. 26:1879–1888.

Fontanillas, E., J. J. Welch, J. A. Thomas, and L. Bromham. 2007. The influence of body size and net diversification rate on molecular evolution during the radiation of animal phyla. BMC Evol. Biol. 7:1–12.

Gillman, L. N., D. J. Keeling, H. A. Ross, and S. D. Wright. 2009. Latitude, elevation and the tempo of molecular evolution in mammals. Proc. R. Soc. B Biol. Sci. 276:3353–3359.

Gillman, L. N., D. J. Keeling, R. C. Gardner, and S. D. Wright. 2010. Faster evolution of highly conserved DNA in tropical plants. J. Evol. Biol. 23:1327–1330.

Goldberg, E. E., and B. Igić. 2012. Tempo and mode in plant breeding system evolution. Evolution 66:3701–3709.

Goldberg, E. E., L. T. Lancaster, and R. H. Ree. 2011. Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. Syst. Biol. 60:451–465.

Graur, D., and W. Martin. 2004. Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. Trends Genet. 20:80–86.

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

Heath, T. A., S. M. Hedtke, and D. M. Hillis. 2008. Taxon sampling and the accuracy of phylogenetic analyses. J. Syst. Evol. 46:239–257.

Ho, S. Y. W., and S. Duchêne. 2014. Molecular-clock methods for estimating evolutionary rates and timescales. Mol. Ecol. 23:5947–5965.

Ho, S. Y. W., M. J. Phillips, A. J. Drummond, and A. Cooper. 2005. Accuracy of rate estimation using relaxed-clock models with a critical focus on the early metazoan radiation. Mol. Biol. Evol. 22:1355–1363.

Höhna, S., W. A. Freyman, Z. Nolen, J. Huelsenbeck, M. R. May, and B. R. Moore. 2019. A Bayesian approach for estimating branch-specific speciation and extinction rates. bioRxiv 49:555805.

Hohna, S., M. J. Landis, T. A. Heath, B. Boussau, N. Lartillot, B. R. Moore, J. P. Huelsenbeck, and F. Ronquist. 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. Syst. Biol 65:726–736.

Huelsenbeck, John P., and F., Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17: 754–755.

Hugall, A. F., and M. S. Y. Lee. 2007. The likelihood node density effect and consequences for evolutionary studies of molecular rates. Evolution 61:2293–2307.

Jobson, R. W., and V. A. Albert. 2002. Molecular rates parallel diversification contrasts between carnivorous plant sister lineages. Cladistics 18:127–136.

Johnson, S. G., and R. S. Howard. 2007. Contrasting patterns of synonymous and nonsynonymous sequence evolution in asexual and sexual freshwater snail lineages. Evolution 61:2728–2735.

Kuhner, M. K., and J. Felsenstein. 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. Mol. Biol. Evol. 11:459–468.

Lancaster, L. T. 2010. Molecular evolutionary rates predict both extinction and speciation in temperate angiosperm lineages. BMC Evol. Biol. 10:162.

Lanfear, R., S. Y. W. Ho, D. Love, and L. Bromham. 2010. Mutation rate is linked to diversification in birds. Proc. Natl. Acad. Sci. 107:20423–20428.

Lanfear, R., S. Y. W. Ho, T. Jonathan Davies, A. T. Moles, L. Aarssen, N. G. Swenson, L. Warman, A. E. Zanne, and A. P. Allen. 2013. Taller plants have lower rates of molecular evolution. Nat. Commun. 4:1879.

Lartillot, N., and R. Poujol. 2011. A phylogenetic model for investigating correlated evolution of substitution rates and continuous phenotypic characters. Mol. Biol. Evol. 28:729–744.

Lepage, T., D. Bryant, H. Philippe, and N. Lartillot. 2007. A general comparison of relaxed molecular clock models. Mol. Biol. Evol. 24:2669–2680.

Lewis, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst. Biol. 50:913–925.

Linder, M., T. Britton, and B. Sennblad. 2011. Evaluation of bayesian models of substitution rate evolution-parental guidance versus mutual independence. Syst. Biol. 60:329–342.

Losos, J. B., K. I. Warheit, and T. W. Schoener. 1997. Adaptive differentiation following experimental island colonization in Anolis lizards. Nature 387:70–73.

Louca, S., and M. W. Pennell. 2019. Phylogenies of extant species are consistent with an infinite array of diversification histories. bioRxiv: 719435.

Maddison, W. P., P. E. Midford, and S. P. Otto. 2007. Estimating a binary character's effect on speciation and extinction. Syst. Biol. 56:701–710.

Maddison, W. P. M., and R. G. Fitzjohn. 2015. The unsolved challenge to phylogenetic correlation tests for categorical characters. Syst. Biol. 64:127–136.

Magallón, S., S. Gómez-Acevedo, L. L. Sánchez-Reyes, and T. Hernández-Hernández. 2015. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. New Phytol. 207:437–453.

Magnuson-Ford, K., and S. P. Otto. 2012. Linking the investigations of character evolution and species diversification. Am. Nat. 180:225–245.

Marin, J., and S. B. Hedges. 2018. Undersampling genomes has biased time and rate estimates throughout the tree of life. Mol. Biol. Evol. 35:2077–2084.

May, M. R., and B. R. Moore. 2016. How well can we detect lineage-specific diversification-rate shifts? A simulation study of sequential AIC methods. Syst. Biol. 65:1076–1084.

Mayrose, I., and S. P. Otto. 2011. A likelihood method for detecting trait-dependent shifts in the rate of molecular evolution. Mol. Biol. Evol. 28:759–770.

Mitchell, J. S., and D. L. Rabosky. 2017. Bayesian model selection with BAMM: effects of the model prior on the inferred number of diversification shifts. Methods Ecol. Evol. 8:37–46.

Moore, B. R., S. Höhna, M. R. May, B. Rannala, and J. P. Huelsenbeck. 2016. Critically evaluating the theory and performance of Bayesian analysis of macroevolutionary mixtures. Proc. Natl. Acad. Sci. 113: 9569–9574.

Moorjani, P., C. E. G. Amorim, P. F. Arndt, and M. Przeworski. 2016. Variation in the molecular clock of primates. Proc. Natl. Acad. Sci. 113:10607–10612.

Nabholz, B., S. Glémin, and N. Galtier. 2008a. Strong variations of mitochondrial mutation rate across mammals - the longevity hypothesis. Mol. Biol. Evol. 25:120–130.

Nabholz, B., J. F. Mauffrey, E. Bazin, N. Galtier, and S. Glemin. 2008b. Determination of mitochondrial genetic diversity in mammals. Genetics 178:351–361.

Nosil, P., D. J. Funk, and D. Ortiz-Barrientos. 2009. Divergent selection and heterogeneous genomic divergence. Mol. Ecol. 18:375–402.

Ohta, T. 1972. Population size and rate of evolution. J. Mol. Evol. 1:305–314.

———. 1973. Slightly deleterious mutant substitutions in evolution. Nature 246:96–98.

Orr, H. A., and M. Turelli. 2001. The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities. Evolution 55:1085–1094.

Pagel, M., C. Venditti, and A. Meade. 2006. Large punctuational contribution of speciation to evolutionary divergence at the molecular level. Science 314:119–121.

Paland, S., and M. Lynch. 2006. Transitions to asexuality result in excess amino acid substitutions. Science 311:990–992.

Pennell, M. W., L. J. Harmon, and J. C. Uyeda. 2014. Speciation is unlikely to drive divergence rates. Trends Ecol. Evol. 29:72–73.

Quental, T. B., and C. R. Marshall. 2010. Diversity dynamics: molecular phylogenies need the fossil record. Trends Ecol. Evol. 25:434–441.

Rabosky, D. L. 2014. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. PLoS One 9:e89543.

Rabosky, D. L., and E. E. Goldberg. 2015. Model inadequacy and mistaken inferences of trait-dependent speciation. Syst. Biol. 64:340–355.

Rabosky, D. L., J. S. Mitchell, and J. Chang. 2017. Is BAMM flawed? Theoretical and practical concerns in the analysis of multi-rate diversification models. Syst. Biol. 66:477–498.

Rannala, B., and Z. Yang. 2007. Inferring speciation tunes under an episodic molecular clock. Syst. Biol. 56:453–466.

Revell, L. J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3:217–223.

Ricklefs, R. R. E. 2003. Global diversification rates of passerine birds. Proc. R. Soc. Lond. Ser. B Biol. Sci. 270:2285–2291.

Rolland, J., F. L. Condamine, F. Jiguet, and H. Morlon. 2014. Faster speciation and reduced extinction in the tropics contribute to the mammalian latitudinal diversity gradient. PLoS Biol. 12:e1001775.

Ronquist, F., M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. C. Suchard, J. P. Huelsenbeck. 2012. MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61:539–542. http://doi.org/10.1093/sysbio/sys029.

Rutschmann, F. 2006. Molecular dating of phylogenetic trees: a brief review of current methods that estimate divergence times. Divers. Distrib. 12:35–48.

Sabath, N., E. E. Goldberg, L. Glick, M. Einhorn, T. L. Ashman, R. Ming, S. P. Otto, J. C. Vamosi, and I. Mayrose. 2016. Dioecy does not consistently accelerate or slow lineage diversification across multiple genera of angiosperms. New Phytol. 209:1290–1300.

Saclier, N., C. M. François, L. Konecny-Dupre, N. Lartillot, L. Guéguen, L. Duret, F. Malard, C. J. Douady, and T. Lefébure. 2018. Life history traits impact the nuclear rate of substitution but not the mitochondrial rate in isopods. Mol. Biol. Evol. 35:2900–2912.

Sanderson, M. J. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol. Biol. Evol. 19:101–109.

———. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics 19:301–302.

Santos, J. C. 2012. Fast molecular evolution associated with high active metabolic rates in poison frogs. Mol. Biol. Evol. 29:2001–2018.

Sarver, B. A. J., M. W. Pennell, J. W. Brown, S. Keeble, K. M. Hardwick, J. Sullivan, and L. J. Harmon. 2019. The choice of tree prior and molecular clock does not substantially affect phylogenetic inferences of diversification rates. PeerJ 7:e6334.

Schliep, K. P. 2011. phangorn: phylogenetic analysis in R. Bioinformatics 27:592–593.

Schulte, J. A. 2013. Undersampling taxa will underestimate molecular divergence dates: an example from the South American lizard clade Liolaemini. Int. J. Evol. Biol. 2013:1–12.

Simpson, A. G., P. J. Wagner, S. L. Wing, and C. B. Fenster. 2018. Binary-state speciation and extinction method is conditionally robust to realistic violations of its assumptions. BMC Evol. Biol. 18:69.

Smith, S. A., and M. J. Donoghue. 2008. Rates of molecular evolution are linked to life history in flowering plants. Science 322:86–89.

Soares, A. E. R., and C. G. Schrago. 2012. The influence of taxon sampling and tree shape on molecular dating: an empirical example from mammalian mitochondrial genomes. Bioinform. Biol. Insights 6:129–143.

Soria-Hernanz, D. F., O. Fiz-Palacios, J. M. Braverman, and M. B. Hamilton. 2008. Reconsidering the generation time hypothesis based on nuclear ribosomal ITS sequence comparisons in annual and perennial angiosperms. BMC Evol. Biol. 8:1–17.

Stanley, S. M., III, P. W. Signor, S. Lidgard, and A. F. Karr 1981. Natural clades differ from "random" clades: simulations and analyses. Paleobiology 7:115–127.

Strathmann, R. R., and M. Slatkin. 1983. The improbability of animal phyla with few species. Paleobiology 9:97–106.

Swindell, W. R., A. Johnston, L. Sun, X. Xing, G. J. Fisher, M. L. Bulyk, J. T. Elder, and J. E. Gudjonsson. 2012. Meta-profiles of gene expression during aging: limited similarities between mouse and human and an unexpectedly decreased inflammatory signature. PLoS One 7:e33204.

Tamura, K., F. U. Battistuzzi, P. Billing-Ross, O. Murillo, A. Filipski, and S. Kumar. 2012. Estimating divergence times in large molecular phylogenies. Proc. Natl. Acad. Sci. 109:19333–19338.

Thomas, J. A., J. J. Welch, R. Lanfear, and L. Bromham. 2010. A generation time effect on the rate of molecular evolution in invertebrates. Mol. Biol. Evol. 27:1173–1180.

Thorne, J. L., and H. Kishino. 2002. Divergence time and evolutionary rate estimation with multilocus data. Syst. Biol. 51:689–702.

Thorne, J. L., H. Kishino, and I. S. Painter. 1998. Estimating the rate of evolution of the rate of molecular evolution. Mol. Biol. Evol. 15:1647–1657.

Tomoko, O. 1995. Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. J. Mol. Evol. 40:56–63.

Tuskan, G. A., S. DiFazio, S. Jansson, J. Bohlmann, I. Grigoriev, U. Hellsten, M. Putnam, S. Ralph, S. Rombauts, A. Salamov, et al. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313:1596–1604.

Venditti, C., and M. Pagel. 2010. Speciation as an active force in promoting genetic evolution. Trends Ecol. Evol. 25:14–20.

Welch, J. J., and L. Bromham. 2005. Molecular dating when rates vary. Trends Ecol. Evol. 20:320–327.

Wertheim, J. O., M. Fourment, and S. L. Kosakovsky Pond. 2012. Inconsistencies in estimating the age of HIV-1 subtypes due to heterotachy. Mol. Biol. Evol. 29:451–456.

Whelan, S., B. P. Blackburne, and M. Spencer. 2011. Phylogenetic substitution models for detecting heterotachy during plastid evolution. Mol. Biol. Evol. 28:449–458.

Woolfit, M., and L. Bromham. 2003. Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes. Mol. Biol. Evol. 20:1545–1555.

Zuckerkandl, E., L. Pauling, and C. Basu. 1965. Evolutionary divergence and convergence in proteins. Pp. 97–166 *in* V. Bryson and H. Vogel, eds. Evolving genes and proteins. Academic Press, New York.

Associate Editor: Y. Brandvain
Handling Editor: T. Chapman

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1: The proportion of simulations where the simulated neutral trait was erroneously associated with shifts in diversification ($r = 1$; Simulation Set 1).
Figure S1: Log likelihood difference between BiSSE null and alternative models as a function of asymmetry in sequence evolution (Simulation Set 1).
Figure S2: The average Akaike weights of each of the examined BiSSE and HiSSE models (Simulation Set 1).
Figure S3: False inference of trait-dependent diversification of BiSSE+HiSSE in scaled versus inferred trees (Simulation Set 1).
Figure S4: Bias in parameter estimation of BiSSE as a function of asymmetry in sequence evolution rate (Simulation Set 1).
Figure S5: False inference of trait-dependent diversification of BiSSE+HiSSE in scaled versus inferred trees (hidden trait, Simulation Set 2).
Figure S6: Bias in parameter estimation of the BiSSE model as a function of asymmetry in sequence evolution rate of a hidden trait (Simulation Set 2).
Figure S7: False inference of shifts in diversification rates of MEDUSA in scaled versus inferred trees (Simulation Set 3).
Figure S8: False inference of shifts in diversification rates of BAMM in scaled versus inferred trees (Simulation Set 3).
Figure S9: The effect of sequence length on the false positive rate of diversification analyses (simulation sets 1 and 4)
Figure S10: Branch-score distances between the time-calibrated trees and the true trees.

Supporting information

Tables

Table S1: The proportion of simulations where the simulated neutral trait was erroneously associated with shifts in diversification (*r* = 1; Simulation Set 1). An incorrect inference of state-dependent diversification was considered when the null BiSSE model was rejected using a ΔAIC threshold of 2. As expected, values are all around the expected value of 0.05.

Number of taxa	True trees	ML-PATHd8	ML-PL	Bayes-IGR	Bayes-TK02	Bayes-strict
50	0.04	0.02	0.02	0.03	0.01	0.03
100	0.06	0.06	0.06	0.05	0.04	0.06
200	0.02	0.02	0.04	0.02	0.03	0.03
400	0.03	0.02	0.04	0.03	0.02	0.02

Figures

Figure S1: Log likelihood difference between BiSSE null and alternative models as a function of asymmetry in sequence evolution (Simulation Set 1). The log likelihood difference between the null and alternative BiSSE models when given trees that were inferred using different tree reconstruction strategies: (A) ML-PL (B) ML-PATHd8, (C) Bayes-IGR, (D) Bayes-TK02, and (E) Bayes-strict, using trees of 50, 100, 200, and 400 taxa. The red horizontal line inside each box indicates the median. From left to right, the orange, pink, purple and blue colors correspond to simulations with r = 1, 2, 4, 8, respectively.

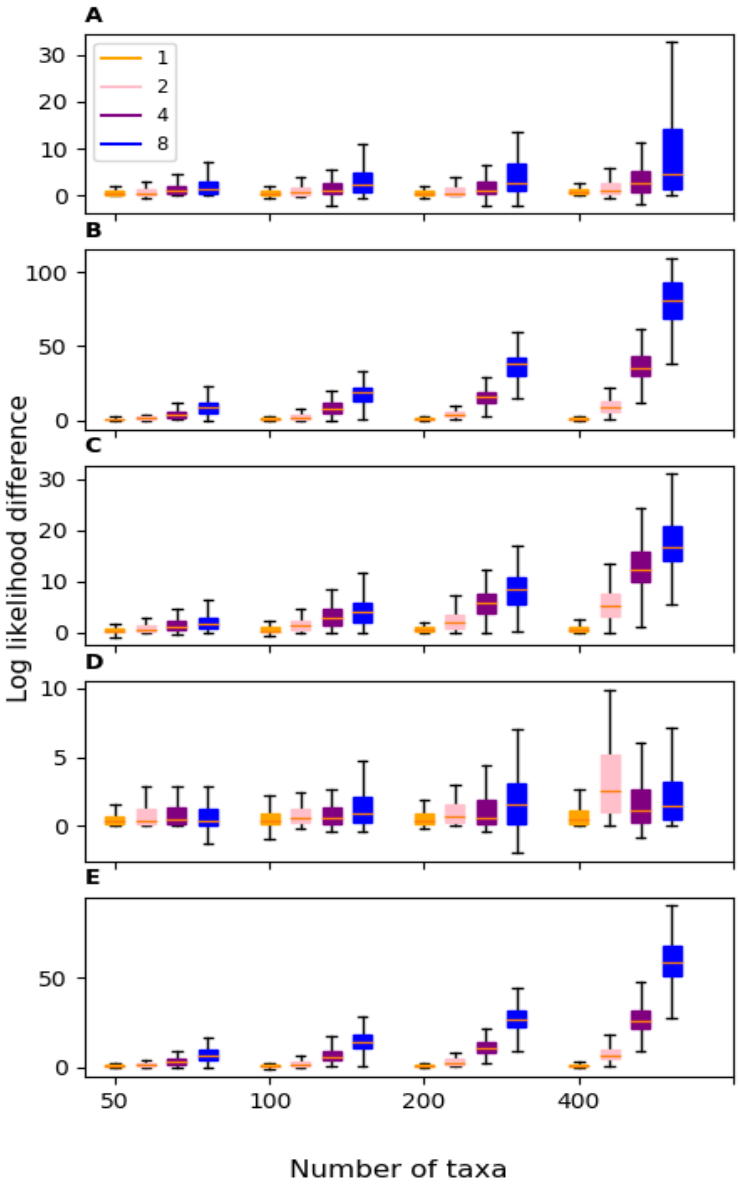

A
B
C
D
E
1
2
4
8
Log likelihood difference
Number of taxa
50
100
200
400

Figure S2: The average Akaike weights of each of the examined BiSSE and HiSSE models (Simulation Set 1). Akaike weights were computed given different tree reconstruction strategies: (A) ML-PL (first row), (B) ML-PATHd8 (second row), (C) Bayes-IGR (third row), (D) Bayes-TK02 (fourth row), and (E) Bayes-strict (fifth row). Within each row, each panel represents results for different number of taxa (from left to right): 50 taxa, 100 taxa, 200 taxa, and 400 taxa. The different colors represent the Akaike weights obtained for the different models: BiSSE-null (orange), CID-2 (purple), BiSSE-SDD (pink), and full-HiSSE (blue). For each substitution rate ratio *r*, the left bar represents the Akaike weights of the candidate models when analyzing only the two BiSSE models (BiSSE-null and BiSSE-SDD), while the right bar represents the Akaike weights of the candidate models when considering additionally the CID-2 and full-HiSSE models.

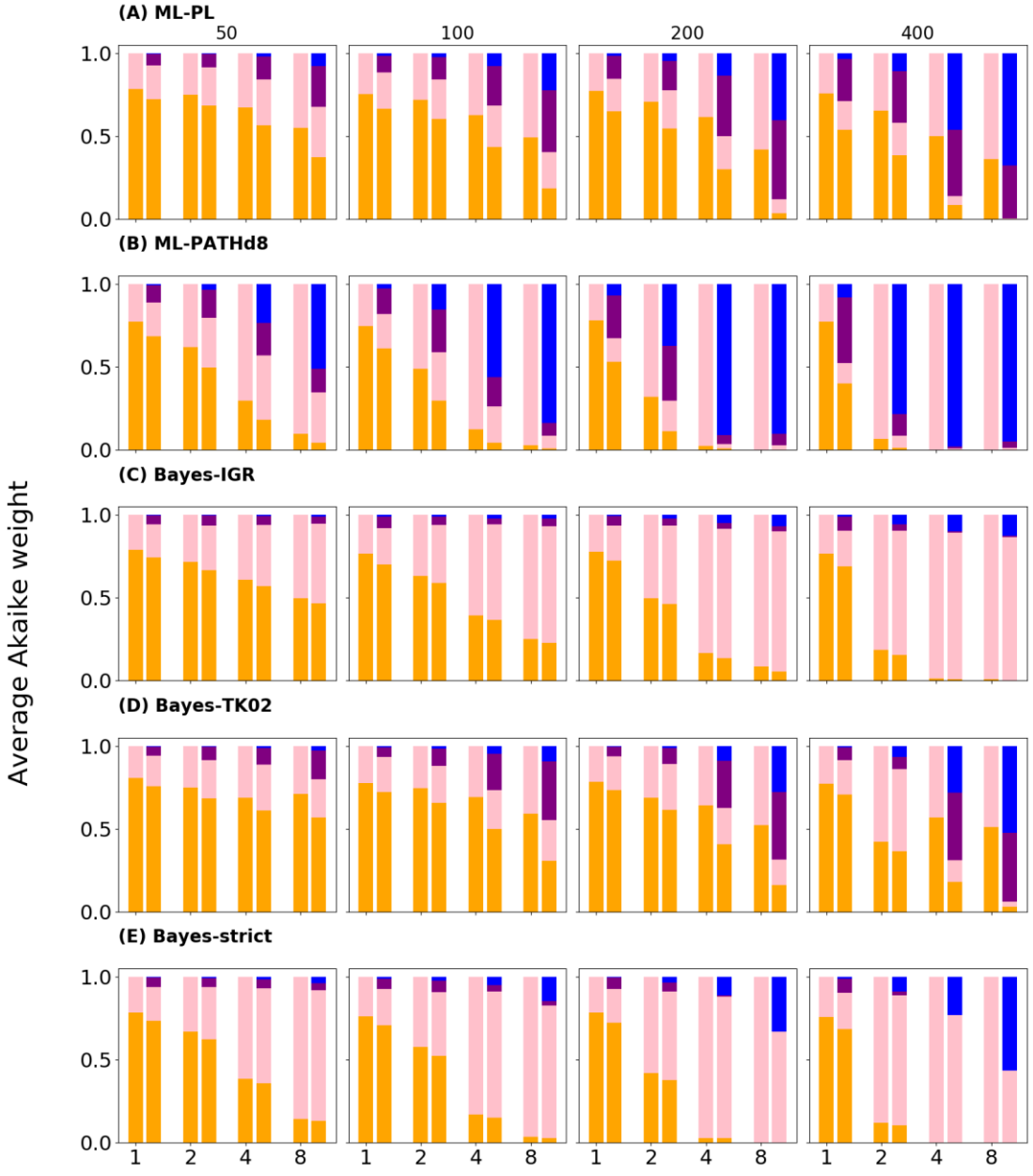

(A) ML-PL
50
100
200
400
(B) ML-PATHd8
(C) Bayes-IGR
(D) Bayes-TK02
(E) Bayes-strict
1.0
0.5
0.0
1
2
4
8
Average Akaike weight
Substitution rate ratio (r)

Figure S3: **False inference of trait-dependent diversification of BiSSE+HiSSE in scaled versus inferred trees (Simulation Set 1)**. Proportion of simulations in which the trait was erroneously associated with diversification rates as a function of the substitution rate ratio (*r*) in simulations that include: (A) 50, (B) 100, (C) 200, and (D) 400 taxa. The solid lines represent the results obtained for the inferred trees, while the dashed lines refer to the results for the scaled trees. Results are presented for input trees inferred using different time-calibration methods: PL (blue) and PATHd8 (red). In all cases the null model was rejected using a ΔAIC threshold of 2.

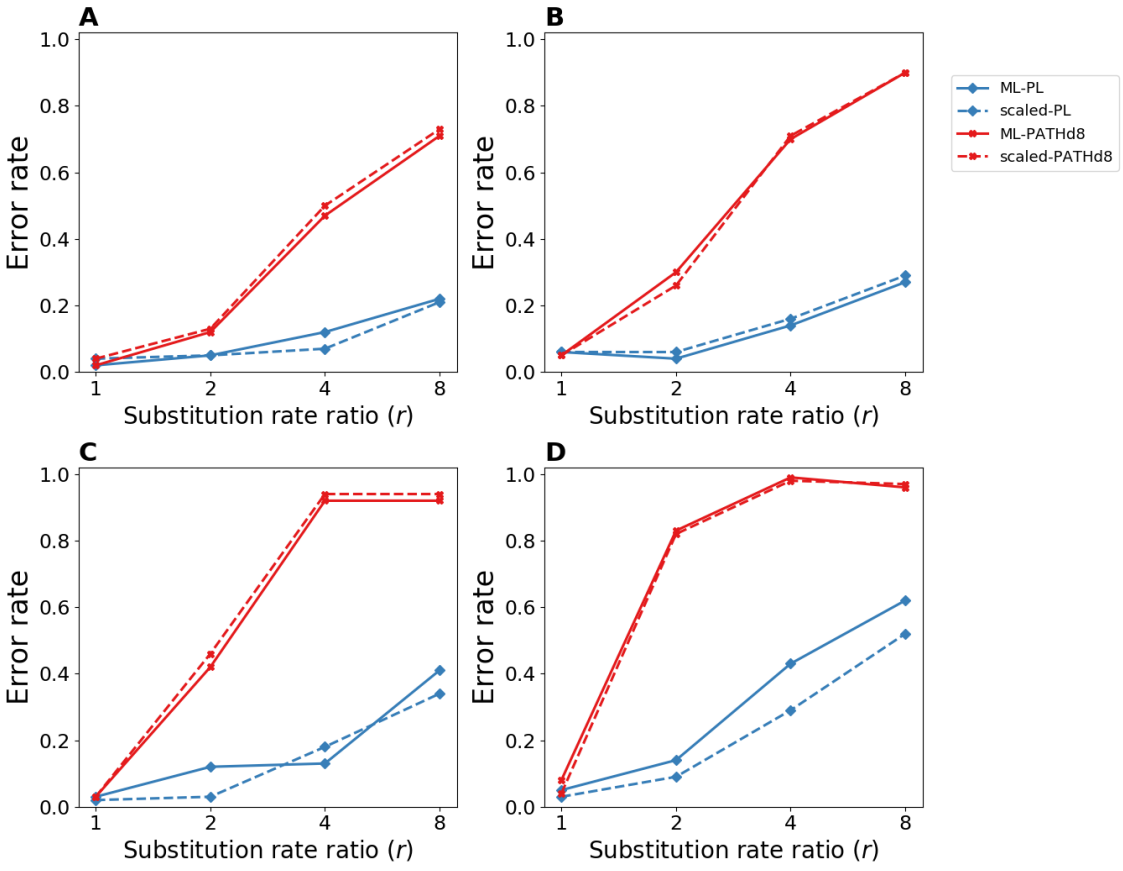


Figure S4: Bias in parameter estimation of BiSSE as a function of asymmetry in sequence evolution rate (Simulation Set 1). Box plots of the inferred diversification parameter estimates $(\lambda_0 - \lambda_1)$ (first column), $(\mu_0 - \mu_1)$ (second column), and $(d_0 - d_1)$ (third column), as a function of the simulated substitution rate ratio (r) parameter. Each row represents different tree reconstruction method: (A) ML-PATHd8, (B) Bayes-strict, (C) Bayes-IGR, and (D) Bayes-TK02. Within each panel, results obtained using trees of 50, 100, 200, and 400 taxa are ordered from left to right and are colored orange, pink, purple, and blue, respectively. The green triangle within each box plot denotes the average of 100 replicates. The horizontal brown line represents the true value.

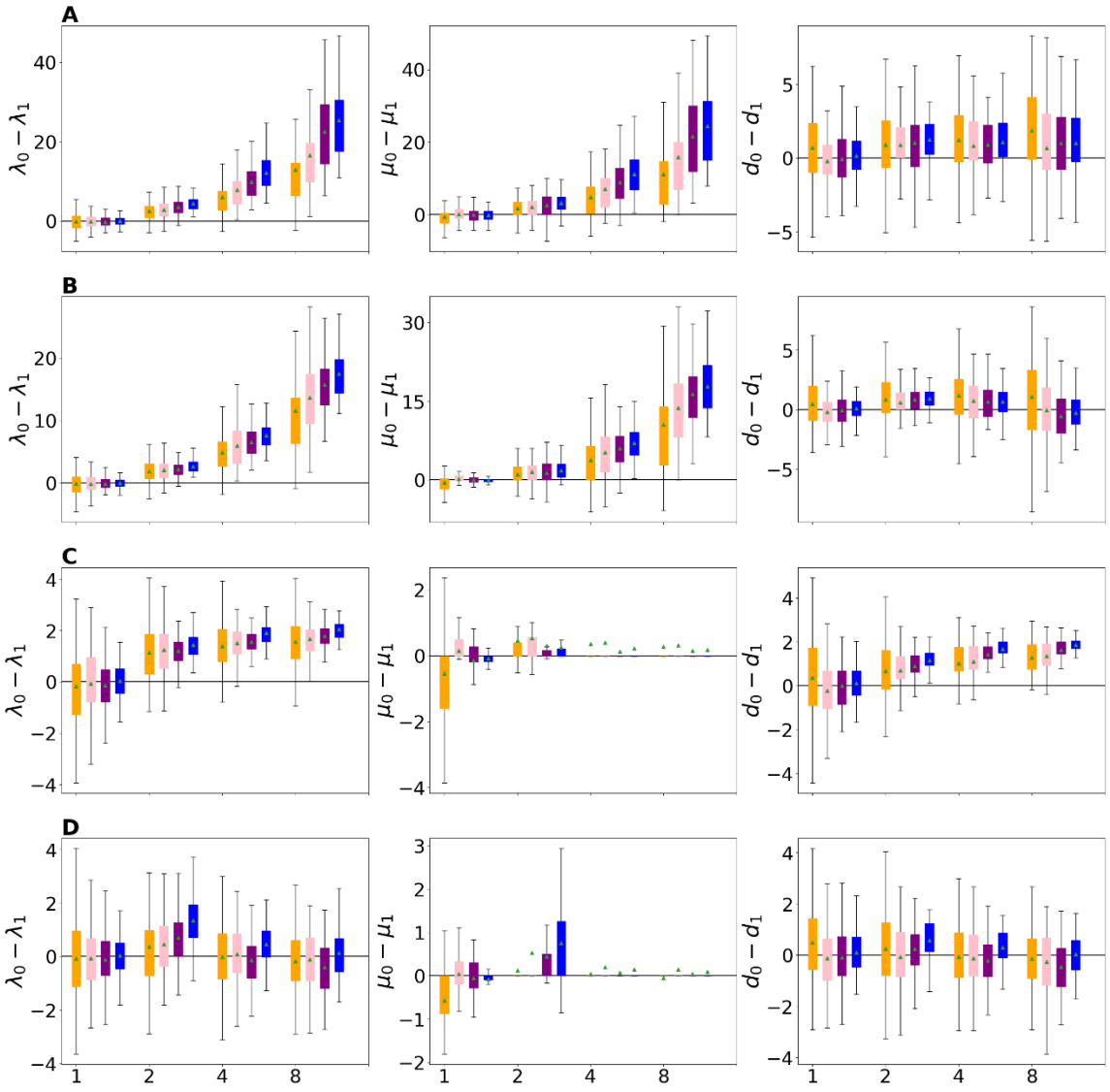

A
B
C
D
$\lambda_0 - \lambda_1$
$\mu_0 - \mu_1$
$d_0 - d_1$
1
2
4
8
Substitution rate ratio (r)

Figure S5: **False inference of trait-dependent diversification of BiSSE+HiSSE in scaled versus inferred trees (hidden trait, Simulation Set 2)**. Proportion of simulations in which the neutral trait was erroneously inferred as associated with diversification rates as a function of the substitution rate ratio (*r*) in simulations with (A) 50, (B) 100, (C) 200, and (D) 400 taxa. The solid lines represent the results obtained for the inferred trees and the dashed lines for the scaled trees. Results are presented for input trees inferred using different time-calibration methods: PL (blue) and PATHd8 (red). In all cases the null model was rejected using a ΔAIC threshold of 2.

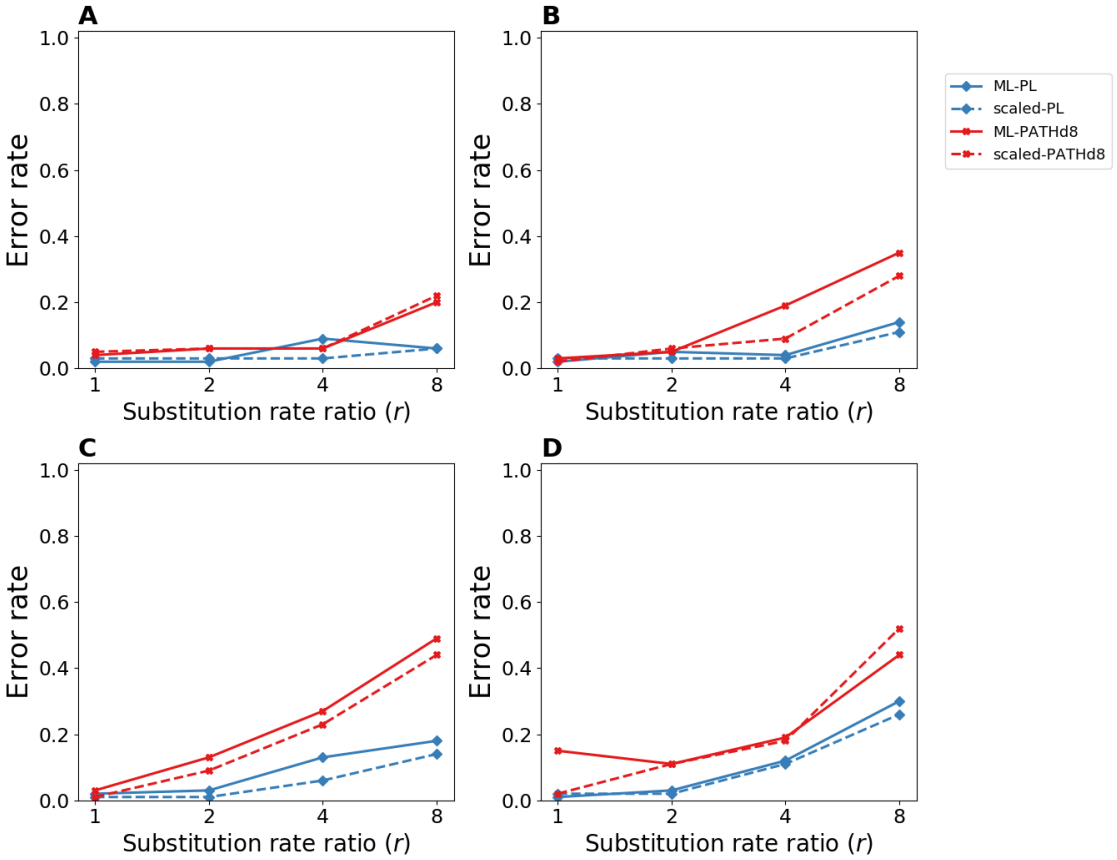


Figure S6: Bias in parameter estimation of the BiSSE model as a function of asymmetry in sequence evolution rate of a hidden trait (Simulation Set 2). Box plots of the inferred (A) (λ_0 -λ_1), (B) ($\mu_0 - \mu_1$), and (C) ($d_0 - d_1$) as a function of the simulated substitution rate ratio (*r*) parameter. Within each panel, results obtained using trees of 50, 100, 200, and 400 taxa are ordered from left to right and are colored orange, pink, purple, and blue, respectively. The green triangle within each box plot denotes the average of 100 replicates. The horizontal brown line represents the true value. The results shown here correspond to ML-PL trees.

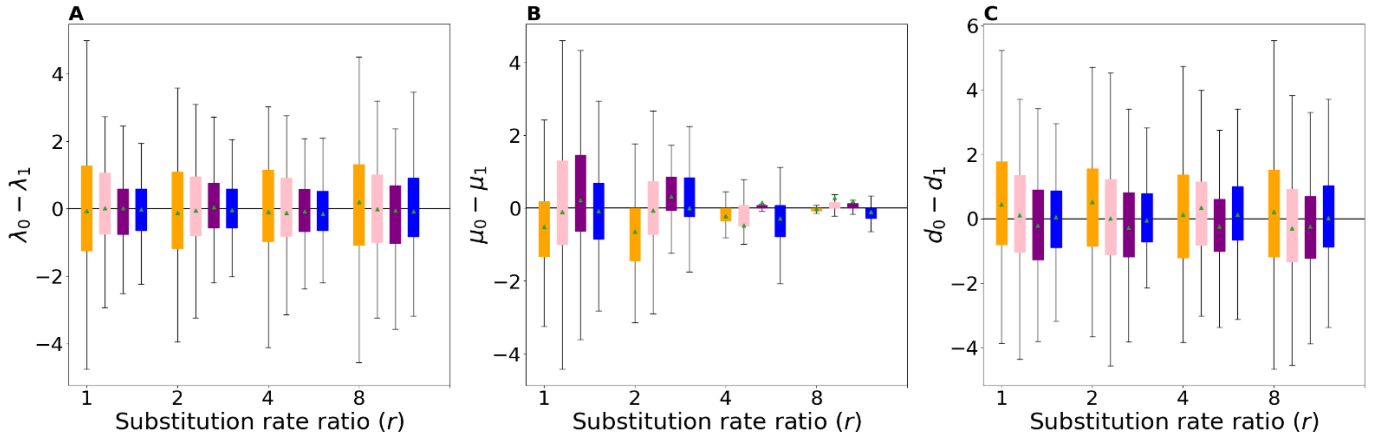


Figure S7: **False inference of shifts in diversification rates of MEDUSA in scaled versus inferred trees (Simulation Set 3)**. Proportion of simulations in which MEDUSA erroneously inferred shifts in diversification as a function of the substitution rate ratio (*r*) in simulations with: (A) 50, (B) 100, (C) 200, and (D) 400 taxa, using different time-calibration methods. The solid lines represent the results obtained using trees inferred from simulated sequence data: ML-PL (blue) and ML-PATHd8 (red), while the dashed lines represent the results obtained using trees that were time calibrated based on the scaled trees: scaled-PL (blue) and scaled-PATHd8 (red).

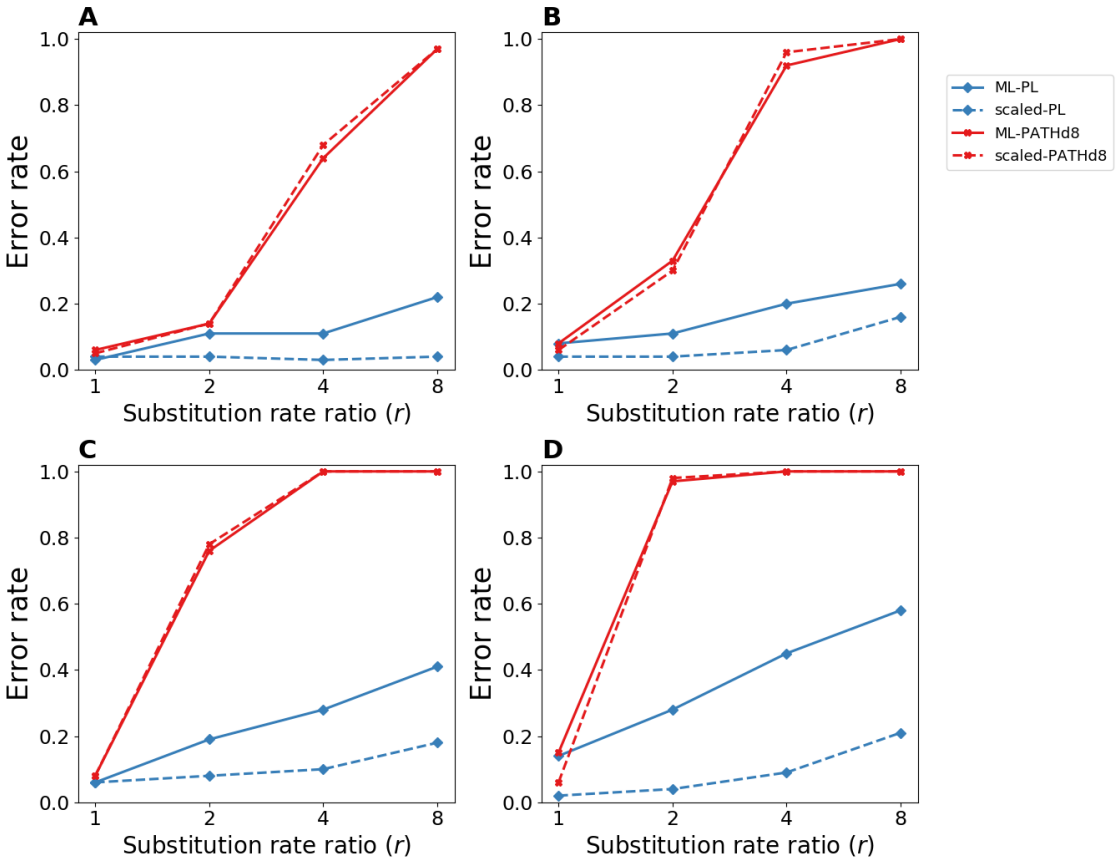


Figure S8: False inference of shifts in diversification rates of BAMM in scaled versus inferred trees (Simulation Set 3). Proportion of simulations in which BAMM erroneously inferred shifts in diversification rates as a function of the substitution rate ratio (*r*) in simulations with: (A) 50, (B) 100, (C) 200 and (D) 400 taxa, using different time-calibration methods. The solid lines represent the results obtained using trees inferred from simulated sequence data: ML-PL (blue) and ML-PATHd8 (red), while the dashed lines represent the results obtained using trees that were time calibrated based on the scaled trees: scaled-PL (blue) and scaled-PATHd8 (red).

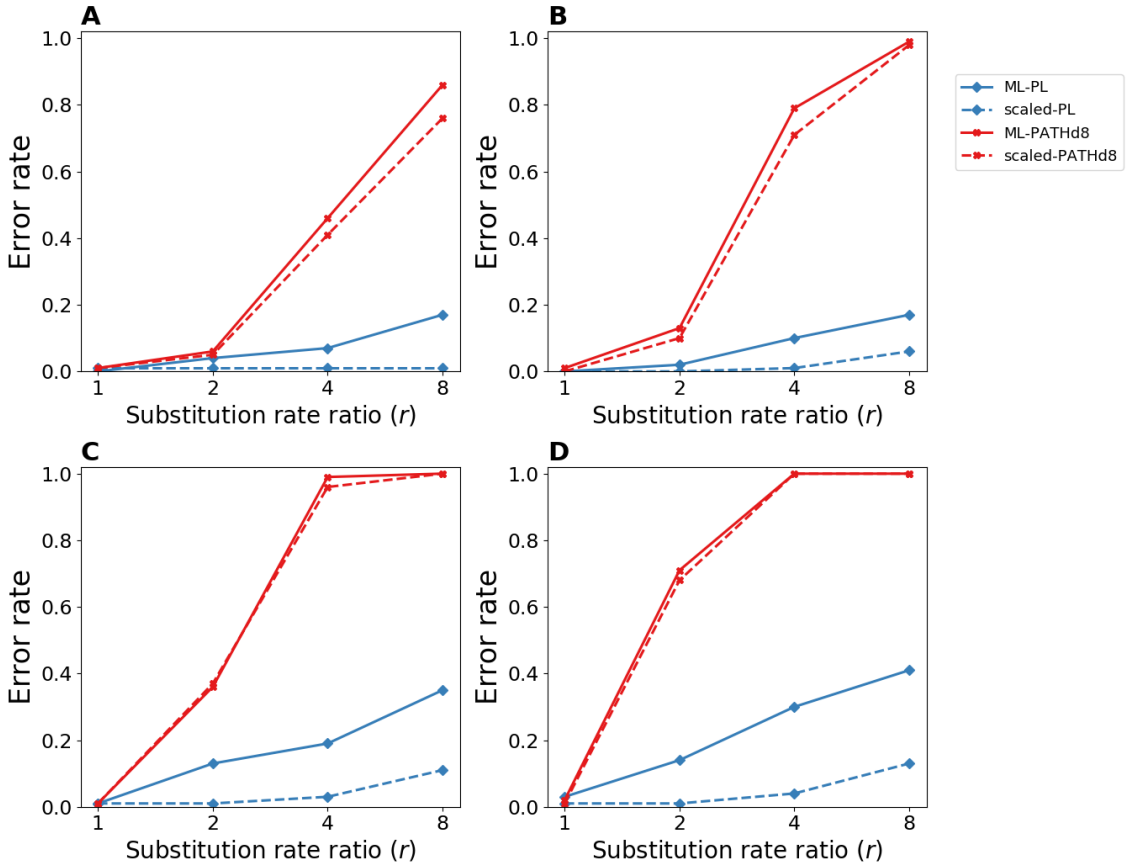


Figure S9: The effect of sequence length on the false positive rate of diversification analyses (simulation sets 1 and 4)

(A-B) Proportion of simulations in which the null model was erroneously rejected as a function of *r* in simulations with different sequence lengths for (A) trait-dependent diversification using BiSSE+HiSSE and (B) trait-independent diversification using MEDUSA. (C) The tree distance between the reconstructed and true trees as a function of the substitution rate ratio (*r*) in simulations with different sequence lengths. Distances were computed using the branch score distance (Kuhner and Felsenstein 1994). Simulations with different sequence lengths are represented by different colors: 250 (blue), 500 (red), and 2000 (green). The solid and dashed lines represent results obtained using trees inferred using ML-PL and Bayes-TK02, respectively. The dotted black line represents the scaled-PL trees. All simulations were conducted using trees of 200 taxa.

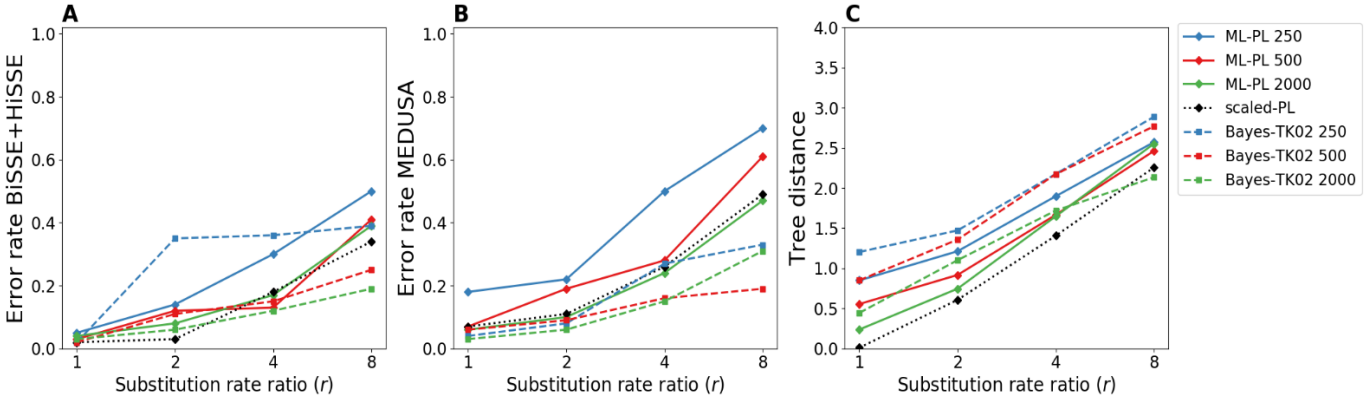
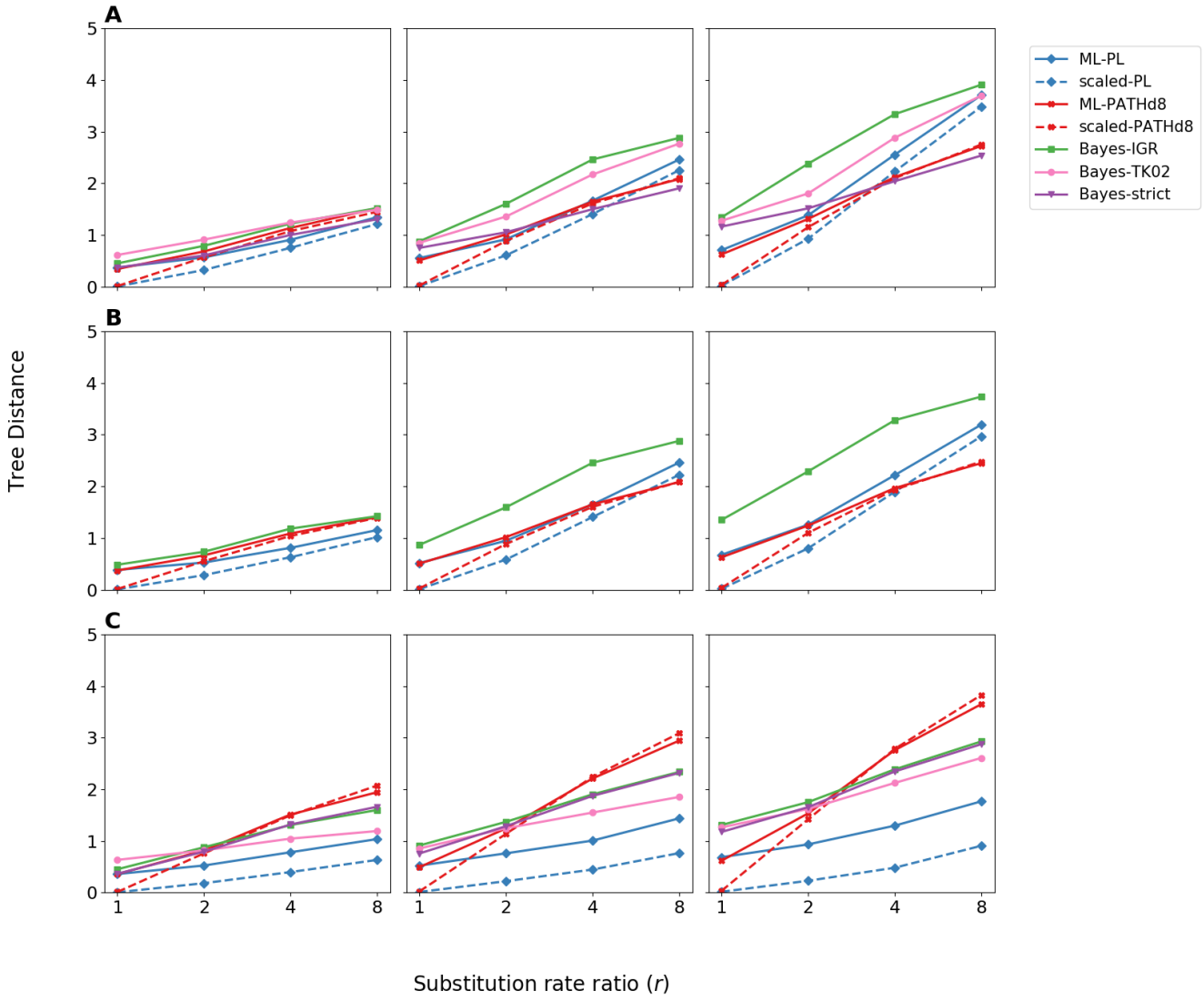


Figure S10: **Branch-score distances between the time-calibrated trees and the true trees**. Branch score tree distance between the reconstructed and true trees as a function of the substitution rate ratio (*r*) obtained in (A) Simulation Set 1+4, (B) Simulation Set 2, (C) Simulation Set 3 using different time-calibration methods. The solid lines represent results obtained using trees inferred from simulated sequence data: ML-PL (blue), ML-PATHd8 (red), Bayes-IGR (green), Bayes-TK02 (pink), and Bayes-strict (purple), while the dashed lines represent the results obtained using trees that were time calibrated based on the scaled trees: scaled-PL (blue) and scaled-PATHd8 (red). For each Simulation Set, each panel depicts results obtained using trees with different number of taxa (from left to right: 50, 200 and 400 taxa).


A
B
C
Tree Distance
Substitution rate ratio (r)
ML-PL
scaled-PL
ML-PATHd8
scaled-PATHd8
Bayes-IGR
Bayes-TK02
Bayes-strict