

# A Likelihood Method for Detecting Trait-Dependent Shifts in the Rate of Molecular Evolution

Itay Mayrose\*<sup>1</sup> and Sarah P. Otto<sup>1</sup>

<sup>1</sup>Department of Zoology, University of British Columbia, Vancouver, BC, Canada

\*Corresponding author: E-mail: mayrose@zoology.ubc.ca.

**Associate editor:** Jeffrey Thorne

## Abstract

Rate heterogeneity within groups of organisms is known to exist even when closely related taxa are examined. A wide variety of phylogenetic and dating methods have been developed that aim either to test for the existence of rate variation or to correct for its bias. However, none of the existing methods track the evolution of features that account for observed rate heterogeneity. Here, we present a likelihood model that assumes that rate variation is caused, in part, by species' intrinsic characteristics, such as a particular life-history trait, morphological feature, or habitat association. The model combines models of sequence and character state evolution such that rates of sequence change depend on the character state of a lineage at each point in time. We test, using simulations, the power and accuracy of the model to determine whether rates of molecular evolution depend on a particular character state and demonstrate its utility using an empirical example with halophilic and freshwater daphniids.

**Key words:** evolutionary models, rate shift, substitution rates, character mapping, molecular clock, dating.

## Introduction

The “molecular clock” hypothesis states that molecular changes accumulate at an approximately constant rate through time in different lineages (Zuckerklund and Pauling 1965). This hypothesis has been a cornerstone in molecular evolutionary research because, if true, it enables us to estimate the time of any divergence event using a single calibration point (e.g., a dated fossil). However, the constant clock hypothesis is often violated, and rate heterogeneity is known to exist even among closely related species (Thomas et al. 2006). Thus, the estimation of divergence times cannot be performed with a single calibration as under the universal clock assumption. Accordingly, a growing number of statistical models have been suggested that allow rates to vary across lineages thereby relaxing the molecular clock assumption (Sanderson 1997; Thorne et al. 1998; Huelsenbeck et al. 2000; Sanderson 2002; Thorne and Kishino 2002; Yang 2004; Drummond et al. 2006; Guindon 2010). It has further become evident that rate variation may have a systematic component (reviewed in Bromham 2009). Various species' characteristics have been shown to influence the rate of molecular evolution, including generation time (Laird et al. 1969; Bromham et al. 1996), metabolic rate (Martin and Palumbi 1993), reproductive mode (Paland and Lynch 2006; Johnson and Howard 2007), sex of carrier (Whittle and Johnston 2002), and habitat (Davies et al. 2004). For example, rates of molecular evolution are consistently higher in herbaceous plants compared with related trees and shrubs (Smith and Donoghue 2008).

In this study, we are interested in the question of whether a particular character state influences the rate of molecular evolution and whether this can be detected.

We limit our discussion to characters with two possible states (i.e., binary). The trait could represent, for example, a particular mating preference, life-history trait, or morphological feature. Two main approaches have been previously used to demonstrate a biological trait effect on the rate of molecular evolution. One approach compares average branch lengths of sister groups under the constraint that all taxa in one clade possess one character state, whereas all taxa in the second clade have the other (e.g., Smith and Donoghue 2008). However, although sister clade comparison is simple and relatively nonparametric, it fails to account for the full information contained within the phylogeny of the species under study. Each sister clade comparison is considered an independent data point without accounting for, for example, the number of species present in each comparison, although higher variance in average branch lengths estimates is expected when fewer taxa are available. Moreover, the method requires accurate knowledge of the phylogeny and excludes all lineages not part of a sister clade contrast. Last, the method lacks an explicit model regarding the different pathways along which the evolution of the character states proceeds. Thus, the character transitions are implicitly assumed to occur simultaneously with the separation of the two lineages (i.e., at their most recent common ancestor). This assumption is reasonable in cases where the character transition itself contributes to speciation (e.g., polyploidy) but will otherwise lead to underestimation of the underlying rate difference because the actual transition may have occurred part way along a branch. Similarly, within each clade, the possibility of multiple or back transitions is ignored.

A second common approach to compare rates of evolution between two predefined groups of sequences uses the relative-rate test (Sarich and Wilson 1973; Wu and

Li 1985; Li and Bousquet 1992). In such a test, one compares the molecular evolutionary distances between one or more species in group A or B and a reference species denoted as the outgroup. Robinson et al. (1998) further presented a weighting scheme in which the tree topology (but not branch lengths) is accounted for. However, as noted by Felsenstein (1988), when more than two taxa are compared the estimated distances become correlated due to phylogenetic structure, thus rendering the test statistically inappropriate. Moreover, relative-rate tests can produce contradictory results depending on the outgroup choice. As the distance between the ingroup and outgroup sequences increases, so does the variance in the distance estimates, thus reducing the accuracy of the test (Robinson et al. 1998). As with other methods that rely on pairwise comparisons, the use of a distant outgroup might also lead to mutational saturation, reducing the ability to detect true rate variation.

An explicit likelihood framework for trait-based rate variation offers a natural solution to these shortcomings. The underlying likelihood model would allow the rate of molecular evolution to depend on the state of a lineage at each point in time. Additionally, the model would allow both the molecular and character states to change anywhere on the tree as long as they are consistent with their realizations at terminal nodes.

Recently, a likelihood-based approach for the detection of a phenotype–genotype association has been presented by O’Connor and Mundy (2009). In their approach, a combined genotype–phenotype rate matrix is constructed and applied under the coevolutionary framework of Pagel (1994). Their method assigns weight parameters that modify the genotypic substitution rate given the state of the phenotype. Importantly, the method assumes that only a fraction of sites are phenotypically associated, whereas the fraction of sites that are unassociated are used to estimate the background substitution rate. Furthermore, their approach does not force the evolution of the trait to have occurred in a particular manner that is the same for all sites influenced by the trait. Rather, the combined phenotype–genotype rate matrix is applied independently to each site considered. Thus, the method of O’Connor and Mundy (2009) cannot be used to detect an association between the phenotype and the background substitution rate. Instead, the method may be used to search for specific sites that are associated with the phenotype, much in the same way as positively selected sites are inferred using codon-based models (e.g., using the PAML package; Yang 2007).

Here, we present a new model framework that combines two well-established sets of models: those that describe sequence evolution and those that describe the evolution of traits into one likelihood framework. Our likelihood calculations rely on the notion that the actual times between present day species and their common ancestor should be equal. Variation in the expected amount of genetic change between these species is due to variation in the species’ characteristics (here assumed to be binary), which in-

fluence the rate of change of their DNA sequences. Ideally, if we knew the exact times at which character transitions occurred, we could scale the branch lengths of the tree by the respective rate associated with each character state. Then, we would calculate the likelihood of the sequence data, given the scaled tree and the model of sequence evolution. In reality, we must infer the timing of the character transitions. Thus, we construct a stochastic model of character evolution and apply the method of Nielsen (2002) to stochastically map characters onto the tree, enabling us to obtain a valid sample of character change histories. Following a detailed description of the model and its likelihood calculations, we apply the method to simulated and real biological data in order to explore the power and accuracy of the method under various evolutionary scenarios.

## Materials and Methods

### A Combined Character and Sequence Model

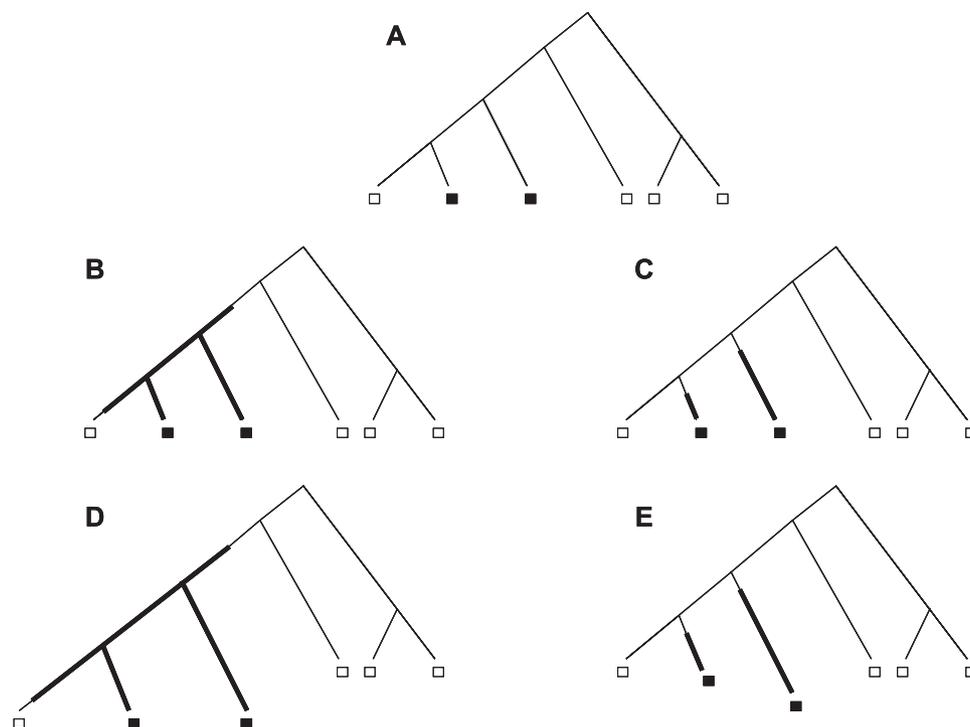
To detect the effect of a binary character trait on the rate of molecular evolution requires data, a phylogenetic tree, and an underlying stochastic model of transitions. The data consist of an assignment of a binary character {0, 1} to the tip taxa (trait data) and sequences for the tip taxa in the form of a multiple alignment (molecular data). Throughout this paper, we assume that the alignment consists of nucleotides, though the model can be applied to amino acid sequences and can be easily extended to model codon evolution. The phylogenetic tree is represented as a rooted ultrametric species tree in which both the topology and branch lengths are assumed to be specified, and the distances from the root to all tips are equal. For simplicity, we first consider the case of a single tree. An extension to multiple plausible trees is detailed below. Finally, the underlying stochastic model consists of two components: a character process, describing the evolution of the binary character along the phylogeny and a molecular evolutionary process, describing the sequence evolution of the given locus (or loci).

The sequence model can be any nucleotide substitution model, such as the general time reversible (GTR; Yang et al. 1994) model or HKY (Hasegawa et al. 1985). In the simulations and analyses presented in this study, the mutation process is based on the HKY model and allows for among-site rate heterogeneity following a gamma distribution (Yang 1994). In this formulation, the sequence model has two free parameters:  $\kappa$ , the transition versus transversion rate bias and  $\alpha$ , the shape parameter of the gamma distribution.

For the character model, we assume a continuous-time Markov process, with transitions between zeros and ones defined by a  $2 \times 2$  rate matrix:

$$Q_c = \mu \cdot \begin{bmatrix} -\pi_1 & \pi_1 \\ \pi_0 & -\pi_0 \end{bmatrix}, \quad (1)$$

where  $\pi_1$  and  $\pi_0$  are the stationary character frequencies.  $\pi_1$  is proportional to the rate of  $0 \rightarrow 1$  transitions and  $\pi_0 = 1 - \pi_1$  is proportional to the rate of  $1 \rightarrow 0$  transitions. The parameter  $\mu$  is introduced to transform the units of branch lengths



**FIG. 1.** Example of scaled trees generated by the two-clock model. (A) The original ultrametric tree. (B + C) Two possible character mappings consistent with the states at the tips of the tree. The mapping in (B) involves one transition from state 0 to state 1 and one transition from state 1 to state 0, whereas the mapping in (C) involves two transitions from 0 to 1. (D + E) The resulting trees after the rescaling step with  $r = 1.5$  are no longer ultrametric. The rescaled trees are then used for likelihood calculations of observing the sequence data. Thick lines represent the time spent in state 1 (black boxes).

to expected number of character transitions per unit of time. The transition probabilities from character state  $i$  to state  $j$  along a branch of length  $t$ , denoted  $P_{ij}(t)$ , can be calculated analytically given the rate matrix (Ross 1996). In order to unite the character and sequence models, we introduce an additional free parameter  $r$ , which denotes the sequence substitution rate while the character is in state 1 relative to state 0. We hereafter refer to this model as the two-clock model. Generally, the model described here is not limited to two character states but can be applied to any number of discrete states (Lewis 2001). In such cases, the model would contain additional free parameters: an additional relative rate parameter for each discrete state and additional free parameters as specified by the extended rate matrix for the characters.

### Likelihood Computations

We are interested in computing the combined likelihood of the character and sequence data:

$$L = P(D_s, D_c | T, r, M_c, M_s), \quad (2)$$

where  $T$  is the input clock-like species tree,  $M_c$  and  $M_s$  are the parameters of the character and sequence models, respectively, and  $D_c$  and  $D_s$  are the character and sequence data, respectively. The steps involved in computing the likelihood under our combined model are as follows (fig. 1). First, a sample of possible character histories is mapped onto the species tree using stochastic mapping (Nielsen 2002), a technique that accounts for the uncertainty in reconstructing discrete state histories by generating a sample of plausible character change histories along the phylogeny, given the model of char-

acter evolution and the character data. Second, the species tree is scaled according to the simulated mappings and  $r$ , the relative substitution rate when in state 1. Specifically, given a simulated history, we adjust each branch according to the amount of time spent in state 1 and in state 0:  $b_i(p_{i1} \times r + p_{i0})$ , where  $b_i$  is the original branch length in the species tree and  $p_{is}$  is the proportion of time the lineage was in state  $s$  along branch  $i$ . We then introduce a parameter,  $B$ , which estimates the total amount of sequence evolution over the tree. This parameter allows the adjusted tree to be stretched or shrunk, as different choices for  $r$  are explored, so that  $r$  is not artificially constrained by its effects on the tree length. Each branch  $i$  is then scaled so that the total tree length is  $B$  given a particular simulated character history:

$$b'_i = B \frac{b_i(p_{i1} \times r + p_{i0})}{\sum_j b_j(p_{j1} \times r + p_{j0})}, \quad (3)$$

where the summation is over all branches in the tree. According to equation (3), for a fixed value of  $B$ , those branches that are predominantly in state 0 will shrink in length (if  $r > 1$ ), whereas branches mostly in state 1 will expand. Consequently, the scaled tree no longer follows a single molecular clock, and the units of time are now measured by the expected number of nucleotide substitutions, given a particular character history.

Following the scaling procedure, the likelihood of the sequence data is calculated for each simulated character history based on the adjusted tree and the nucleotide model using Felsenstein's (1981) pruning algorithm. The final likelihood of the sequence data is the average over

all simulated histories and is multiplied by that of the character data to produce the combined likelihood score:

$$\begin{aligned}
 & P(D_s, D_c | T, r, M_c, M_s) \\
 &= P(D_c | T, r, M_c, M_s) P(D_s | T, r, M_c, M_s, D_c) \\
 &= P(D_c | T, M_c) P(D_s | T, r, M_c, M_s, D_c) \\
 &= P(D_c | T, M_c) \int_{T_\psi} P(D_s | T, r, D_c, T_\psi, M_c, M_s) \\
 &\quad \times P(T_\psi | T, r, M_c, M_s, D_c) dT_\psi \\
 &= P(D_c | T, M_c) \int_{T_\psi} P(D_s | T_\psi, M_s) \\
 &\quad \times P(T_\psi | T, M_c, D_c) dT_\psi \\
 &\cong \frac{P(D_c | T, M_c)}{N} \sum_{j=1}^N P(D_s | T_j', M_s).
 \end{aligned} \tag{4}$$

The integral is over the set of all possible stochastic mappings, with  $T_\psi$  representing one particular mapping and  $T_\psi'$  corresponding to the scaled tree given the particular mapping and the  $r$  parameter. Because the number of possible mappings is infinite, the integral is approximated by a sum where  $T_j$  represents one character mapping randomly sampled from  $P(T_\psi | T, D_c, M_c)$  and  $N$  is the number of sampled mappings.  $P(D_s | T_j', M_s)$  is the likelihood of the sequence data, given the scaled tree from the  $j$ th character mapping,  $T_j'$ , and the model of sequence evolution. We note that because  $T_j'$  is scaled according to the character model and character data, the likelihood of the sequence data depends on the evolution of the characters.  $P(D_c | T, M_c)$  is the likelihood of the character data, given the model of character evolution and the input clock-like species tree,  $T$ , which is here assumed to be specified prior to analysis and does not necessarily rely on the input sequence data (but see Phylogenetic Uncertainty below). All free parameters of the model ( $\pi_1$ ,  $\mu$ , and  $r$  for the character model;  $\kappa$ ,  $\alpha$ , and  $B$  for the sequence model) were estimated by maximum likelihood (ML) using Brent's optimization scheme (Press et al. 2002). In order to avoid getting caught at local maxima, ten random starting points were used during the optimization process.

A similar likelihood calculation can be carried out across multiple plausible species trees (that are constrained to be ultrametric). These trees may be, for example, a Markov chain Monte Carlo (MCMC) sample of the posterior distribution (Huelsenbeck et al. 2000). In such a case, for each of the  $N$  iterations, we draw uniformly at random a tree from the set of input trees and perform a single stochastic mapping iteration based on that tree. The computation continues identically thereafter according to equation (4).

### Null Model and Model Comparison

A null model, which does not permit a trait-specific shift in the rate of sequence evolution ( $r = 1$ ), allows statistical testing of the null hypothesis that the state of the binary character is not associated with the rate of evolution. In such a case, the difference in the number of free parameters between the two models is one. The likelihood ratio test (LRT) can be used to compare such nested models with the distribution of the maximum log-likelihood difference between the models approaching a  $\chi_1^2$  distribution with increasing amount of data. The validity of this approximation was verified using simulations (e.g., Whelan and Goldman 1999) (results not shown).

The significance test above compared the two-state model to a null model, which assumes that observed rate variation among lineages is due solely to the stochasticity of the substitution process. Alternatively, rate variation may be present but may not be caused by the trait under study. This may lead to rejection of the null hypothesis using the LRT without the trait being associated with the rate of sequence evolution any more than many other traits. A parametric bootstrap approach was thus developed to test for this possibility. This procedure tests whether the observed rate variation is associated with the analyzed trait significantly more often than uncorrelated traits that evolve in a similar manner. Specifically, for each data set analyzed, character data were regenerated by simulating character evolution along the input phylogeny (see below for details regarding the simulation implementation) with the parameters of the character model identical to those inferred by ML for the original data set under the two-clock model but without altering the original sequence data (regenerating  $D_c$ , holding  $D_s$  and  $T$  constant). Given the randomly generated character data at the tips of the tree and the original sequence data, the ML values of the null and the two-clock models were compared. This procedure was repeated a large number of times (200 for the empirical biological data set and 100 for each of the simulated data sets) to produce a distribution for the difference in the maximum log-likelihood obtained under the two models. Thus, in these simulations, substitution rate variation is present and is identical to that in the original data set, but it is decoupled from the observed character data. The log-likelihood difference observed for the original data was then compared with the simulated distribution to produce a  $P$  value. Hereinafter, we refer to this test as the parametric trait bootstrap. We note that for the simulated data sets, using the ML parameter estimates to simulate the evolution of additional traits generated very similar  $P$  values to those generated using the true simulated parameters (not shown). Thus, it seems that the error associated with the ML estimates has little effect on the parametric trait bootstrap procedure.

### Assessing Accuracy via Simulations

Simulations were used to investigate the power and precision of our method and to assess its accuracy in inferring the  $r$  parameter. Simulated data were generated by modeling the evolutionary process given a fixed tree and a given set of model parameters. The simulations were performed by simulating character evolution using the embedded discrete-time Markov chain of the rate matrix  $Q_c$  (i.e., the waiting time in state  $i$  is exponentially distributed with rate  $-Q_{ii}$  and given that a change has occurred, the probability to jump to state  $j$  is  $-Q_{ij}/Q_{ii}$ ). In this way, we recorded, for each tree branch, the time spent in state 0 and in state 1, allowing us to rescale the model tree according to equation (3). We then simulated nucleotide evolution along the scaled tree. Simulations with all tip taxa having the same character state were rerun until both character states were observed. The resulting sequences and character state

**Table 1.** Various Simulation Scenarios Analyzed in this Study.<sup>a</sup>

Simulation Scenario	Number of Taxa	Sequence Length	Tree	$r$	$\mu$	$\pi_1$	$N$
1	20–60	200	Simulated	1–4.2	10	0.5	25–1,000
2	20–60	200	Simulated	1–4.2	10	0.5	100
3	20–60 <sup>b</sup>	100–1,600	Simulated	1–4.2	10	0.5	100
4	20–60 <sup>b</sup>	200	Simulated	1–4.2	1–1,000	0.5	100
5	40	200	Simulated	1–4.2	0.1–200 <sup>c</sup>	0.5	100
6	20–60 <sup>b</sup>	200	Simulated	1–4.2	10	0.1–0.9	100
7	40	200–1,600	Reconstructed <sup>d</sup>	1–4.2	10	0.5	100

<sup>a</sup> Bold italic type indicates the varied parameter.

<sup>b</sup> Results are presented for 40-taxon trees only.

<sup>c</sup> In these simulations, the tree height varied from 0.005 to 10 so that the expected number of character transitions was fixed at 1.0.

<sup>d</sup> See text for description of the different tree reconstruction methods.

assignments at the tips of the tree were used as the data input.

Random trees with different numbers of taxa (20, 40, and 60) were generated according to a birth–death process using the Mesquite program (Maddison WP and Maddison DR 2008) with default parameters (speciation rate 0.3 and extinction rate 0.1) and were scaled so that the distance from the root to the tips, defined as the tree height, is 0.1 nucleotide substitutions per site. Unless otherwise stated, all simulations were conducted with the parameters of the character model set to  $\pi_1 = 0.5$  and  $\mu = 10$  and parameters of the sequence model set to  $\kappa = 2$  and  $\alpha = 1$  to produce character data and sequence alignment of length 200 nucleotides. The different simulation scenarios analyzed in this study are summarized in table 1. For each simulated scenario, 30 independent runs were conducted (each based on an independently generated tree). For each scenario, we varied the  $r$  parameter between 1.0, representing a false positive inference, and 4.2 for which the power of the method was already very high. Simulations with  $r < 1.0$  (representing the case where the rate of sequence evolutionary with character 1 is lower than that with character 0) generated symmetrical results and are thus not presented. The estimated parameter error was defined as  $|r - \hat{r}|/r$ , where  $r$  and  $\hat{r}$  are the true and estimated relative rate parameters.

### Accounting for Phylogenetic Uncertainty

In order to assess the robustness of our method to errors in the phylogeny, the true species tree underlying the simulations was not given as input in the inference step. Instead, ML and Bayesian trees were inferred from the simulated DNA sequences. ML trees were reconstructed using the PhyML software (Guindon and Gascuel 2003) under the HKY + Gamma model with four rate categories. The r8s program (Sanderson 2003) was then applied to ultrametricize the ML tree using the penalized likelihood method (Sanderson 2002). Bayesian trees were reconstructed using MrBayes version 3.2 (Ronquist and Huelsenbeck 2003) under the HKY + Gamma model with four rate categories under a relaxed molecular clock according to a Brownian motion model (Thorne et al. 1998); trees reconstructed according to a strict clock process under the birth–death model (Yang and Rannala 1997) gave highly similar results (not shown). The resulting MrBayes trees are already clock

like, and so an ultrametricization step is unnecessary. For each simulation, two independent MrBayes runs were combined to form one sample. Each analysis consisted of four Markov chains (with heating according to default settings), run for 100,000 generations, and sampled every 200 steps. The first 25% of the sampled trees were considered as burn-in and were discarded from the analyses.

To differentiate between the effects of tree reconstruction errors stemming from an altered tree topology, branch lengths, or from the ultrametricization step, we compared the power and accuracy of our method under a number of different scenarios with the ML-reconstructed trees: 1) Both topology and branch lengths were estimated using ML; inferred trees were subsequently ultrametricized to be clock like (as above). 2) The tree topologies were forced to equal the true simulated ones, but the branch lengths were inferred and the ultrametricization step carried out. 3) The true tree topologies and branch lengths were scaled according to the rates of substitution expected from the simulated character evolution; this non-clock-like tree was then ultrametricized. 4) Same as (3) but 5–20% of internal nodes served as calibration points and were fixed to their true simulated time points. 5) Species trees were identical to the true simulated ones.

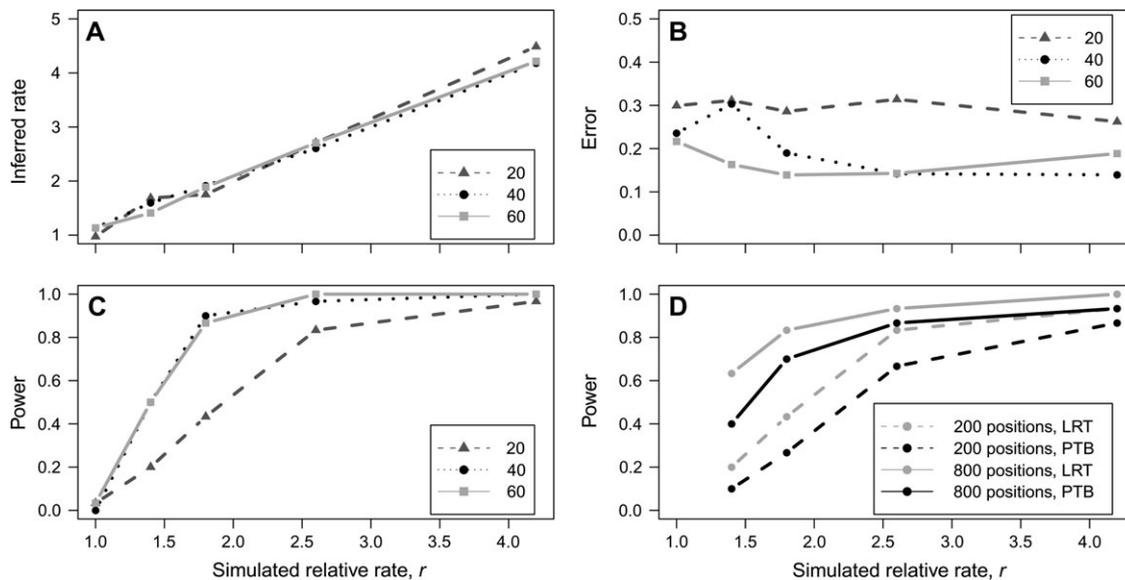
### Implementation and Availability

The models and inference method described here were implemented in C++. The program and source codes are available at <http://www.zoology.ubc.ca/~mayrose/cp/traitRate/>. The inputs to the program are a tree file in Newick format containing one or several alternative trees, a multiple sequence alignment in a number of possible common formats, and a file containing character state assignments for extant taxa in a FASTA format. The program estimates the relative rate parameter and the parameters of the sequence and character models and computes the maximum log-likelihood values of the null and two-clock models.

## Results

### Accuracy as a Function of the Number of Stochastic Mappings

A practical problem associated with methods that rely on mapping mutations onto phylogenies is that the number of



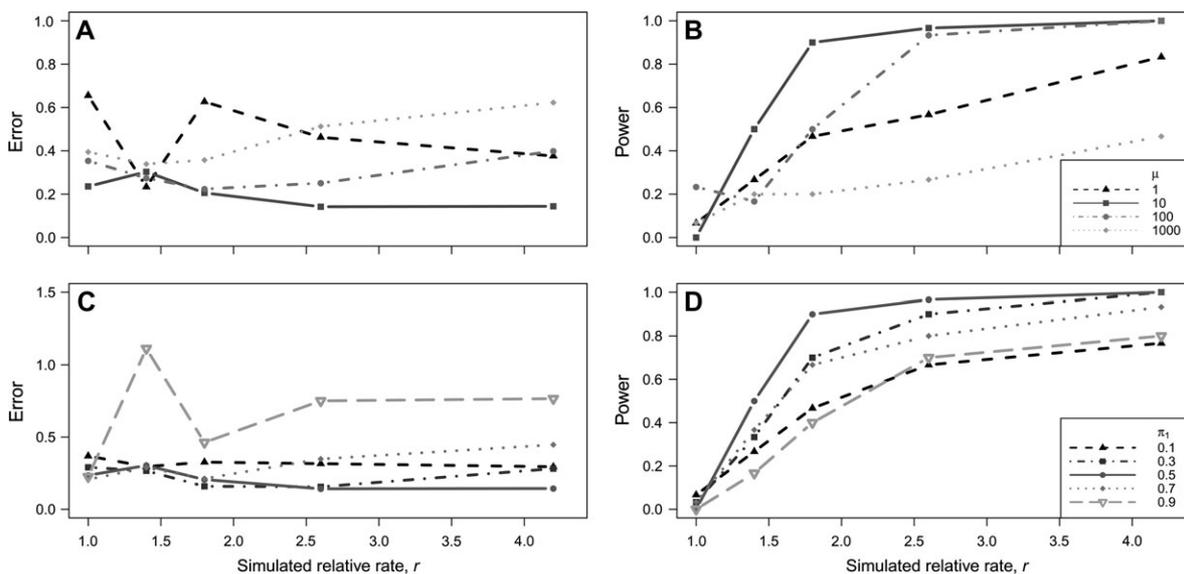
**FIG. 2.** Simulation results for trees with different numbers of taxa. (A) The average value of the estimated  $r$  parameter, (B) its average inference error, defined as  $|r - \hat{r}|/r$ . Panels (C) and (D) show the power of the method to detect trait-dependent rate shifts, using LRT (Panel C and gray curves in Panel D) or using the parametric bootstrap (PTB; black curves in panel D), with a significance threshold of  $\alpha = 0.05$ . In A–C, simulation results with 20, 40, and 60 taxa are plotted in dashed, dotted, and solid lines. In panel D, dashed and solid lines represent simulations with 200 and 800 positions, respectively, using 20 taxa. In each case, the true simulated tree was used to infer rate variation, and simulated  $r$  values were set to 1, 1.4, 1.8, 2.6, and 4.2.

stochastic mappings necessary to obtain a valid approximation for the likelihood is unknown (i.e., when replacing the integral with a sum in eq. 4). Of course, the more mappings,  $N$ , that are used the better the approximation will be. However, the computation time increases linearly with  $N$ . We thus used simulated data, generated with various numbers of taxa and under different values of the  $r$  parameter to study the effect of  $N$  on the approximated likelihood. For each data set, we computed the log-likelihood 100 times, given a fixed set of parameter values, and evaluated its coefficient of variation (CV) given various numbers of mappings. The CV of the log-likelihood scores was below 0.001 for most data sets tested (supplementary fig. S1, Supplementary Material online). In addition, increasing  $N$  above 400 appeared to contribute little additional accuracy. Because the likelihood function is evaluated many times during the search for the ML point, the effects of  $N$  on the estimated ML and on the estimated model parameters are of particular interest. We thus also compared the maximum log-likelihood values and the estimated  $r$  parameter using 25, 50, 100, 200, and 1,000 stochastic mapping iterations (simulation scenario 1; table 1). For all simulations tested, the average difference in maximum log-likelihood values was below 1 with 100 or more mappings (not shown). In addition, the average errors in  $r$  estimates for  $N = 100$  and  $N = 1,000$  were nearly identical (supplementary fig. S1, Supplementary Material online). The difference in the inferred  $r$  parameter was much more noticeable using only 25 or 50 mappings. We thus chose to use 100 stochastic mappings in all subsequent simulations. It should be noted that when individual data sets are an-

alyzed, we recommend using as many mappings as feasible and verifying the replicability of the results with several values of  $N$ .

### Assessing Accuracy via Simulations

Simulations were used to investigate the power and accuracy of our method in estimating  $r$ , the relative rate of molecular evolution in species with character state 1 compared with state 0, as a function of the number of taxa available (simulation scenario 2; table 1). For all cases examined (20, 40, or 60 taxa), the average inferred  $r$  was close to the simulated value (fig. 2A). With few taxa, the average  $r$  estimates were slightly higher than the true  $r$  value that was simulated due to a few simulation runs with high inferred  $r$  values. As expected, accuracy increased when more taxa were available. For 60 taxa, the average inferred  $r$  was very close to the simulated value (fig. 2A), and the inference error was  $<25\%$  for all simulations considered (fig. 2B). The power of the method was high even with relatively small  $r$  values, provided that enough data (in terms of the number of taxa and the sequence length) were available. As can be seen in figure 2C, with 40 or more taxa, the percentage of simulated runs in which the null model would be rejected based on the LRT was above 90% for  $r$  values above 1.8 and above 50% for  $r \geq 1.4$ . With 20 taxa, the power was higher than 80% for  $r$  values above 2.6. In addition, the standard error around the average inferred values generally decreased with the number of taxa simulated (supplementary figs. S6–S7, Supplementary Material online). Simulations with sequences ranging from 100 to 1,600 nucleotides in length (simulation scenario 3; table 1) indicated that power



**FIG. 3.** Simulation results for different parameter values of the character model. Average error of the estimated  $r$  parameter and power are plotted as a function of the simulated  $r$  parameter for different  $\mu$  (A–B) and  $\pi_1$  (C–D) values. Power (panels B and D) describes the % of runs where the null hypothesis was rejected based on the LRT ( $\alpha = 0.05$ ). In panels A–B the simulated  $\pi_1$  parameter was fixed at 0.5, and in panels C–D the simulated  $\mu$  parameter was fixed at 10. The legend inside panel B specifies different values for the simulated  $\mu$  parameter and applies also to panel A. The legend inside panel D specifies different values for the simulated  $\pi_1$  parameter and applies also to panel C. All simulations presented are with 40-taxon trees.

and accuracy significantly increased with the amount of sequence data (supplementary fig. S2, Supplementary Material online). To shorten computational time, all simulation results presented in the text were conducted with a sequence length of 200 nucleotides, which is shorter than the sequence length of most biological data sets. In this respect, the presented power and accuracy of the method can be considered underestimates.

The significance test detailed above compared the two-clock model to a null model that assumes a single clock. However, it is possible that the two-state model better fit the data because there was rate variation present that was unrelated to the analyzed trait. A second statistical test, the parametric trait bootstrap (see Materials and Methods), compared the log-likelihood difference between the two models for the true character data versus character data that were randomly generated. In the vast majority of cases, the significance threshold derived from the parametric trait bootstrap was higher compared with that for the LRT. This trend was especially pronounced when longer sequences and higher  $r$  values were simulated (not shown). As with the LRT, the power of the parametric trait bootstrap test increased substantially with the amount of data (i.e., sequence length) and with higher values of  $r$  (fig. 2D). We thus interpret the LRT as initial evidence for trait-dependent rate variation. The parametric trait bootstrap test, which is computationally more intensive, should then be used to determine if the observed rate variation is more likely associated with the analyzed trait compared with other traits that evolve in a similar manner. We note, however, that a significant result using the parametric trait

bootstrap test does not preclude the possibility that other traits also contribute to the underlying rate variation. In the following, we explore a very large number of simulated data sets and so we restrict our power analyses to the LRT to avoid excessive computation time. Accordingly, we refer to the power of the method achieved under the LRT criterion only as a proxy to investigate the influence of varying parameters on the performance of the method. When we analyze a specific empirical data set, however, we apply the parametric trait bootstrap test as well.

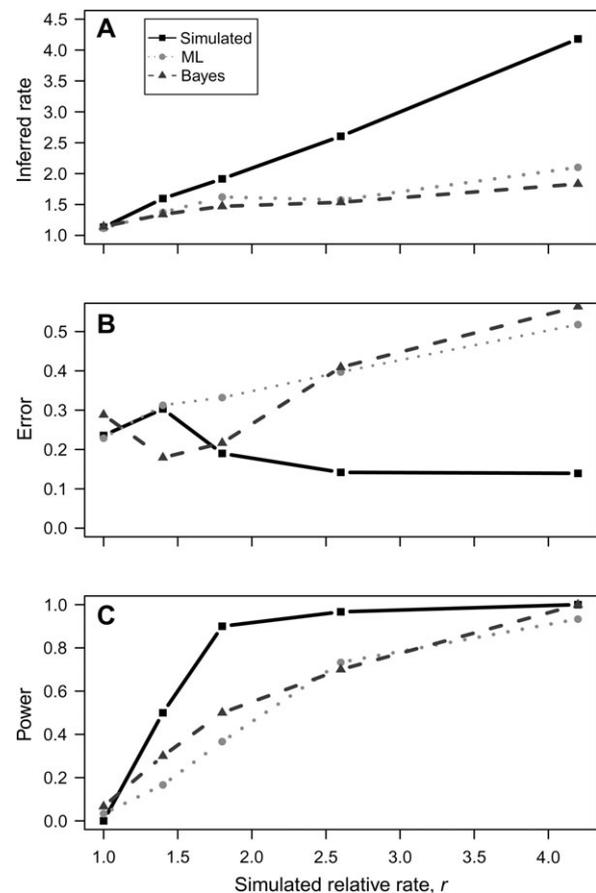
We next tested the accuracy of our approach given different values for the parameters of the character model (simulation scenarios 4–6; table 1). As can be seen in figure 3, the power and accuracy of inferring the  $r$  parameter dropped considerably when the rates of character change were very low ( $\mu = 1$ ) or very high ( $\mu = 1,000$ ). With low  $\mu$  values, there were merely one or two character state changes across the whole tree, leading to inaccurate inference of the character model parameters and little information about the relative rate of molecular evolution in one of the character states. This resulted in errors inferring the  $r$  parameter and low power to detect rate shifts. With very high  $\mu$  values, both character states are expected to be present along most branches of the phylogeny, and the rate of sequence evolution along each branch tends toward the average rate. Importantly,  $\mu$  specifies the scaling factor that multiplies the branch lengths of the input tree (in which branch lengths are measured as average number of nucleotide substitutions) to best fit the character data and thus its exact values depend on the input tree. The above simulations were performed with trees that were scaled so that

the tree height is 0.1. We also performed simulations with different tree heights while keeping the expected number of character transitions fixed at 1.0 (e.g., tree height of 0.01 and  $\mu = 100$ ; simulation scenario 5 in [table 1](#)). In these simulations, the average inference errors of the  $r$  parameter were quite similar across tree heights in the range 0.05–1.0, whereas inferior results were obtained for extremely short trees (tree height of 0.005) because very few nucleotide substitutions occurred, or for very long trees (tree height of 10) where site saturation became a problem ([supplementary fig. S3, Supplementary Material online](#)). Our simulation results also indicated that the most accurate results are achieved when the transition rates between the two character states are similar (i.e.,  $\pi_0 = \pi_1 = 0.5$ ). With very asymmetric transition rates ( $\pi_1 = 0.1$  or 0.9), either the rare state is very transient (when  $\mu$  is high) and so has little effect on the rate of molecular evolution across a branch or the rare state is hardly encountered (low  $\mu$ ), thus reducing the power of the test ([fig. 3C–D](#)). Interestingly, the errors associated with high  $\pi_1$  values were greater compared with low  $\pi_1$  values (compare  $\pi_1 = 0.9$  to 0.1 in [fig. 3C](#)), implying that less accurate  $r$  estimates are obtained when the rare phenotype is slow evolving and leaves little signal on the sequence data. Despite giving less accurate  $r$  estimates, simulations with extreme  $\pi_1$  values tended to result in more accurate  $\pi_1$  and  $\mu$  estimates ([supplementary fig. S4, Supplementary Material online](#)).

### Phylogenetic Uncertainty

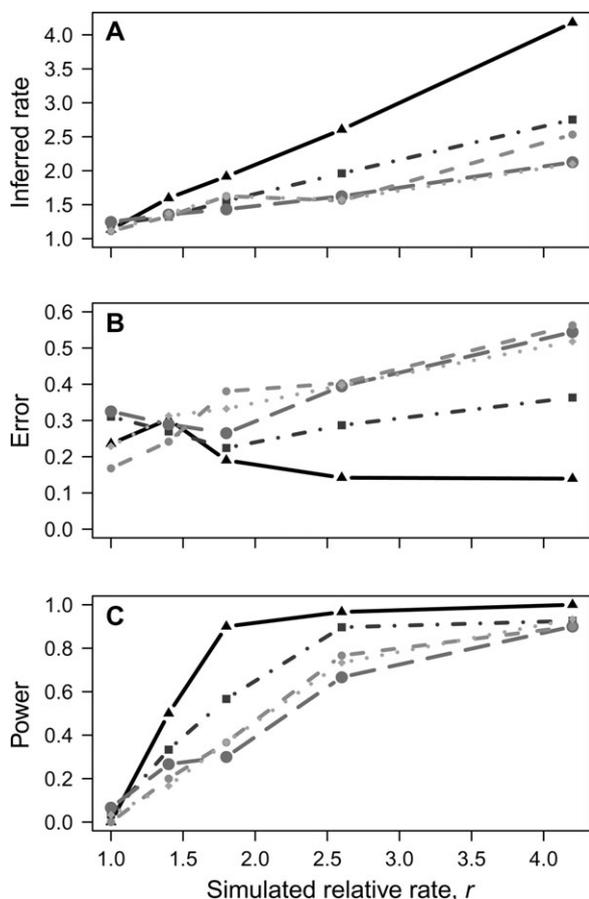
The simulations detailed above assumed that the given phylogeny is the correct one. This assumption certainly cannot be achieved in most biological studies. We thus investigated the robustness of our method to phylogenetic errors using two approaches. First, a single ML-based phylogeny was reconstructed from the simulated sequence data (see [Material and Methods](#)) and was given as input to our method. Second, a large number of possible trees were sampled from the posterior distribution using Bayesian MCMC techniques. Both these options assume that the sequence data available to reconstruct the species tree are identical to the sequence data that are used to study rate variation, although in practice, the reconstruction of the species tree may rely on additional data other than those analyzed for rate variation. As may be expected, the accuracy of the method decreased when the input tree was inferred from the simulated data rather than being identical to the tree used to simulate the data. As shown in [figure 4](#), the error in the estimation of the relative rate parameter,  $r$ , and the power were very similar for both phylogenetic inference methods. Although the power of both approaches increased substantially when longer sequences were simulated (simulation scenario 7; [table 1](#)), the errors in estimating the  $r$  parameter depended only weakly on sequence length ([supplementary fig. S5, Supplementary Material online](#)).

Noticeably, the  $r$  parameter was always underestimated when the tree was inferred ([fig. 4A](#)). This bias in estimating  $r$  could be caused by higher rates of change being inter-



**FIG. 4.** Simulation results using different tree reconstruction methods. (A) The average value of the estimated  $r$  parameter, (B) its average inference error, and (C) power based on the LRT ( $\alpha = 0.05$ ) are plotted for different methods of inferring the species tree: when the simulated model tree was given as input (solid), reconstructed using ML (dotted), or averaged over an MCMC sample (dashed) as indicated in the legend inside panel A that applies also to panels B and C. All simulations presented are with 40-taxon trees.

preted as longer periods of divergence by tree inference methods, which could alter the inferred tree topology and branch lengths. Alternatively, the bias might originate from the ultrametrization step. To differentiate between these possibilities, we compared the power and accuracy of our method under a number of different scenarios with the ML-reconstructed trees (see [Material and Methods](#)). Our simulation results indicated that errors in reconstructing the gene trees had negligible effect on accuracy. As shown in [figure 5](#), power and accuracy were very similar when gene trees were reconstructed or assumed to be identical to the true simulated ones (in these cases, the non-clock-like trees were then ultrametricized). Thus, the major source for the reduced power of the ML-reconstructed trees stemmed from inaccuracies when ultrametricizing gene trees. We conjecture that this occurs because ultrametricization is done without regard to the character states and thus the procedure changes every branch to some extent rather than preferentially adjusting the branches with



**Fig. 5.** Simulation results using different tree input scenarios. (A) The average value of the estimated  $r$  parameter, (B) its average inference error, and (C) power based on the LRT ( $\alpha = 0.05$ ) are plotted for different methods of inferring the species tree: 1) both topology and branch lengths were reconstructed using ML and then ultrametricized (dotted), 2) only the branch lengths were reconstructed and then ultrametricized (dashed), 3) the simulated tree was first rescaled according to the rates of substitution expected from the simulated character evolution and then ultrametricized (long-dash), 4) like case (3), but 20% of the internal nodes were time calibrated (dotted dashed), or 5) the true simulated tree was given as input (solid). See Material and Methods for descriptions of the different tree reconstruction scenarios. All simulations presented are with 40-taxon trees.

high substitution rates (see Discussion). Indeed, the accuracy of the inferred  $r$  parameter increased substantially as a function of the percentage of internal nodes that were correctly dated. Note, however, that relaxed-clock trees generated with MrBayes also led to  $r$  being underestimated (fig. 4) even though there is no ultrametricization step involved. This is presumably because the relaxed-clock model also altered branch lengths in a manner that did not match the character-dependent transitions used in the simulations.

### Biological Example: Halophilic and Freshwater Crustaceans

We exemplify the use of our probabilistic method in detecting habitat-dependent shifts in evolutionary rates by

analyzing saline and freshwater species of the crustacean genus *Daphnia*. Colbourne et al. (2006) presented a molecular phylogeny of the genus based on mitochondrial genes with emphasis on species from Australia, where saline lakes are common. Their results indicated that at least three habitat shifts have occurred from freshwater to saline environments. Using the RRTree program (Robinson-Rechavi and Huchon 2000), the authors detected higher substitution rates in selected lineages of *Daphnia* occupying saline habitats, where both ionic and ultraviolet (UV) exposure are extreme. Here, we use our method to reanalyze the data of Colbourne et al. (2006). A combined sequence alignment of the mitochondrial 12S and 16S rDNA genes was obtained from Colbourne JK (personal communication) and was pruned to contain sequences of the 28 species with a clearly defined habitat assignment as either saline or freshwater. A phylogenetic tree was reconstructed using the phyML program (Guindon and Gascuel 2003) with the GTR substitution model (Yang et al. 1994) and four rate categories. The tree was then converted to be clock like using the penalized likelihood approach implemented in r8s (Sanderson 2002). Our two-clock method indicated a significant rate acceleration in halophiles compared with freshwater crustaceans across the whole phylogeny ( $r = 1.4$ ; log-likelihood difference = 14; LRT  $P < 0.01$ ; parametric trait bootstrap  $P < 0.01$ ; identical results were obtained with either 100 or 500 stochastic mappings). Similar results were obtained whether or not outgroup taxa were included in the analysis. The inferred character model parameters were  $\pi_1 = 0.63$  and  $\mu = 2.3$  (corresponding to four expected character transitions across the given tree). As indicated by our simulations, both these values are in the range where our method achieves high power and accuracy (fig. 3). We note that the inferred  $r$  value is likely an underestimate of the true rate difference between saltwater and freshwater species because of the consistent underestimation of  $r$  when ML-reconstructed phylogenies were used instead of the true simulated ones (fig. 4). The higher substitution rates in halophiles has been ascribed to the mutagenic effect of high salt concentration and/or to higher levels of damaging UV radiation in saline waters (Hebert et al. 2002). In contrast, using the model-based method of O'Connor and Mundy (2009), we did not detect an association between habitat and rate of substitution (log-likelihood difference of 0.23 between alternative and null models;  $P = 0.5$ ). This is presumably because saline habitats cause a genome-wide increase in the rate of substitution, whereas the method of O'Connor and Mundy (2009) aims to detect phenotype-genotype associations that are limited to a fraction of sites only. Using the relative-rate test with the RRTree program, results were highly dependent on the outgroup choice. A significant rate increase in saline habitats was observed when *Daphnia dubia* and *D. occidentalis* were placed as the outgroup species ( $P = 0.0001$  and  $0.0007$ , respectively), whereas no difference was observed with the more distant outgroup *Scapholebris* ( $P = 0.78$ ). A nonsignificant difference was also observed when *D. ephemeralis*, a basal ingroup taxon, was designated as

the outgroup ( $P = 0.16$ ). Thus, our two-clock likelihood approach showed less sensitivity to the inclusion of outgroups than the previous method used to analyze the data.

## Discussion

The molecular clock assumption may always be formally incorrect but it may sometimes be almost correct. An unsolved question is: when do the weaknesses of the molecular clock assumption outweigh its convenience? Thorne and Kishino (2006)

According to the molecular clock hypothesis, the same amount of genetic change would accumulate between a common ancestor and all its descendants. Variation among the observed distances is therefore attributed solely to the stochastic appearance and fixation of mutations within a lineage. More often than not, however, the molecular clock hypothesis is rejected when biological sequence data are analyzed. Although methodologies have been devised to test for deviations from a universal clock and to correct for rate heterogeneity (reviewed in Rutschmann 2006; Thorne and Kishino 2006; Ho 2009), the present method may offer some biological basis to the correction performed. Our method acknowledges that species may evolve under different molecular rates due to their intrinsic characteristics, for example, distinct morphological traits or habitats. Thus, even when one clock does not exist, trait-specific clocks may be present. Here, we presented a probabilistic model that accounts for trait-dependent acceleration in the rates of molecular evolution within a phylogenetic context. A key strength of our approach is that it does not rely on an outgroup taxon nor does it rely solely on the data contained in sister group relationships. Beside direct implications to systematic and phylogenetic studies, this model might prove useful in the detection of diverse physiochemical agents, phenotypes, environmental factors, or life-history traits underlying the substitution process.

Our main emphasis here was to devise a method for testing evolutionary rate differences between two types of taxa grouped according to a particular character of interest. Along lineages of each type, the molecular clock is still assumed, thus allowing for a model with relatively few free parameters. Integrating character state evolution into phylogenetic inference and dating methodologies is a promising future direction, which can better account for and explain the observed rate variation. Currently, rate variation is modeled with parametric distributions, which usually enhance the fit to the data, but from which little insight can be gained regarding why rates vary. Current dating techniques model rate variation either as correlated among adjacent branches (e.g., Thorne et al. 1998; Sanderson 2002) or drawn randomly from a specified type of rate distribution (Drummond et al. 2006; see Welch and Bromham 2005 for additional relaxed-clock alternatives). Under the autocorrelated model, substitution rates change gradually presumably due to similarities in the characteristics of descendent species, which are not explicitly mod-

eled. According to the random rates model, the pattern of rate change follows punctuated evolution at each node. The model investigated here combines elements of these two modeling extremes. Specifically, the type of change is punctuated (whenever the trait changes), but rates are still determined by species' intrinsic characteristics. Qualitatively, our model may be more similar to the compound Poisson process suggested by Huelsenbeck et al. (2000), which allows rates to change in a stepwise manner at any point on the tree (but not back and forth between two rates). However, our use of character data as an additional source of information may add some biological tractability to the results. Two extensions to our model are possible that may prove useful. First, instead of assuming a single rate for each character state, rate distributions may be assumed allowing for rate variation also within a character state. A second promising future direction would consider multiple traits, which would allow users to assess the relative impact of different characters on substitution rates. Accordingly, a rate shift between adjacent branches would be gradual when a small fraction of the analyzed traits changes at any particular branch, whereas a punctuated change is expected when several character traits change simultaneously.

It is important to note that the power of our model to detect trait-dependent shifts in the rate of sequence evolution depends on the amount of variability in the sequence data and, even more importantly, in the character data at hand. The power of the model will decrease if the rare character type is represented merely once or twice and is associated with a slowdown in evolutionary rate. To obtain reliable estimates, it is necessary that there is a substantial amount of evolutionary time spent in each character state along a tree. This is more likely to be met if a few character transitions occur deep in the tree or if several transitions occur toward the tips. The reliance on sequence data variability seems to be less crucial, as long as some variability is present (in our 40-taxon simulations with 200 nucleotides, this corresponds to tree heights larger than 0.05) but not so much variability that site saturation becomes a problem. Consequently, longer sequences should be used if the taxa being compared have recently diverged and denser taxon sampling should be attempted if saturation is an issue.

In this study, we concentrated on the comparison of substitution rates using nucleotide sequences. However, the same model can be applied to protein sequences in order to analyze differences in either the selection or the mutation pressure. Sequence evolution would then be represented with one of the commonly used amino acid models (Dayhoff et al. 1978). A more parameter-rich model would utilize sequences at the codon level to differentiate between the various scenarios. For example, researchers could allow one rate parameter to describe the shift in the rate of synonymous substitutions and another for the shift in rate of nonsynonymous substitutions associated with a change in character state. By contrasting these two types of relative rates, we would be able to infer the

relative roles of selection and mutation in causing substitution rate variation of a particular gene.

Currently, our model applies to data sets that are composed of a single locus. When multiple genes are combined, the set of divergence times may be different for each genealogy due to recombination. Commonly, different genes are assumed to share a common set of divergence times (e.g., Thorne and Kishino 2002). In this regard, Rasmussen and Kellis (2007) found that a large fraction of gene- and species-tree incongruencies are attributed to algorithmic inaccuracies rather than biological factors. These authors used a large number of orthologous sets of genes obtained from wholly sequenced genomes to construct lineage-specific substitution rates. By doing so, more accurate gene trees were obtained because longer branches were expected for faster-evolving lineages. Using our model, lineage-specific substitution rates can similarly be obtained, although here, the information external to the analyzed molecular locus is derived from phenotypic data rather than whole-genome sequences.

As shown by our simulations, the suggested model tends to yield conservative values of the relative rate parameter,  $r$ , whenever the sequence data are used also to reconstruct the species phylogeny. We conjecture that directly accounting for trait-specific rates of evolution during the phylogenetic inference step may have a profound impact on estimating phylogenies and divergence times. If not explicitly modeled, trait-specific rates of evolution may mislead phylogenetic reconstruction methods by erroneously grouping taxa having a similar high rate of evolution, thus causing the inferred phylogenies to be particularly susceptible to long branch attraction (Felsenstein 1978; Huelsenbeck and Hillis 1993). Likewise, dating methods are expected to date slowly evolving groups as more recent and rapidly evolving groups as older than their true ages (Smith and Donoghue 2008). The Bayesian approach described by Huelsenbeck et al. (2003) could be applied to explore alternative character histories within a MCMC framework and could be integrated as one component of a broader Bayesian phylogenetic methodology (e.g., Ronquist and Huelsenbeck 2003; Drummond and Rambaut 2007). This promising next step would allow coestimation of the relative rate parameter and the phylogeny of the species under study, potentially enhancing the accuracy of the phylogenetic tree, dates of character transitions, and rate variation during evolution.

## Supplementary Material

Supplementary figures S1–S8 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

## Acknowledgments

We are grateful to Arne Mooers and Wayne Maddison for insightful remarks, to John Colbourne for providing sequence alignment for the *Daphnia* data set, to Rich FitzJohn and Mark Rendel for critically reading the manuscript, and to the anonymous referees and the associate

editor for many constructive comments. This work was supported by a Killam postdoctoral fellowship (to I.M.) and the Discovery Grant from the National Science and Engineering Research Council, Canada (to S.P.O.).

## References

- Bromham L. 2009. Why do species vary in their rate of molecular evolution? *Biol Lett.* 5:401–404.
- Bromham L, Rambaut A, Harvey PH. 1996. Determinants of rate variation in mammalian DNA sequence evolution. *J Mol Evol.* 43:610–621.
- Colbourne JK, Wilson CC, Hebert PDN. 2006. The systematics of Australian *Daphnia* and *Daphniopsis* (Crustacea: Cladocera): a shared phylogenetic history transformed by habitat-specific rates of evolution. *Biol J Linn Soc.* 89:469–488.
- Davies TJ, Savolainen V, Chase MW, Moat J, Barraclough TG. 2004. Environmental energy and evolutionary rates in flowering plants. *Proc Biol Sci.* 271:2195–2200.
- Dayhoff MO, Schwartz RM, Orcutt BC. 1978. A model of evolutionary change in proteins. In: Dayhoff M, editor. Atlas of protein sequence and structure. Washington (DC): National Biomedical Research Foundation. p. 345–352.
- Drummond AJ, Ho SY, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* 4:e88.
- Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol.* 7:214.
- Felsenstein J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Syst Zool.* 27:401–410.
- Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol.* 17:368–376.
- Felsenstein J. 1988. Phylogenies from molecular sequences—inference and reliability. *Annu Rev Genet.* 22:521–565.
- Guindon S. 2010. Bayesian estimation of divergence times from large sequence alignments. *Mol Biol Evol.* 27:1768–1781.
- Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* 52:696–704.
- Hasegawa M, Kishino H, Yano T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol.* 22:160–174.
- Hebert PDN, Remigio EA, Colbourne JK, Taylor DJ, Wilson CC. 2002. Accelerated molecular evolution in halophilic crustaceans. *Evolution* 56:909–926.
- Ho SY. 2009. An examination of phylogenetic models of substitution rate variation among lineages. *Biol Lett.* 5:421–424.
- Huelsenbeck JP, Hillis DM. 1993. Success of phylogenetic methods in the 4-taxon case. *Syst Biol.* 42:247–264.
- Huelsenbeck JP, Larget B, Swofford D. 2000. A compound poisson process for relaxing the molecular clock. *Genetics* 154:1879–1892.
- Huelsenbeck JP, Nielsen R, Bollback JP. 2003. Stochastic mapping of morphological characters. *Syst Biol.* 52:131–158.
- Huelsenbeck JP, Rannala B, Masly JP. 2000. Accommodating phylogenetic uncertainty in evolutionary studies. *Science* 288:2349–2350.
- Johnson SG, Howard RS. 2007. Contrasting patterns of synonymous and nonsynonymous sequence evolution in asexual and sexual freshwater snail lineages. *Evolution* 61:2728–2735.
- Laird CD, McConaughy BL, McCarthy BJ. 1969. Rate of fixation of nucleotide substitutions in evolution. *Nature* 224:149–154.
- Lewis PO. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst Biol.* 50:913–925.

- Li P, Bousquet J. 1992. Relative-rate test for nucleotide substitutions between 2 lineages. *Mol Biol Evol.* 9:1185–1189.
- Maddison WP, Maddison DR. 2008. Mesquite: a modular system for evolutionary analysis. Version 2.5. Available from: <http://mesquiteproject.org>.
- Martin AP, Palumbi SR. 1993. Body size, metabolic rate, generation time, and the molecular clock. *Proc Natl Acad Sci U S A* 90:4087–4091.
- Nielsen R. 2002. Mapping mutations on phylogenies. *Syst Biol.* 51:729–739.
- O'Connor TD, Mundy NI. 2009. Genotype-phenotype associations: substitution models to detect evolutionary associations between phenotypic variables and genotypic evolutionary rate. *Bioinformatics* 25:i94–i100.
- Pagel M. 1994. Detecting correlated evolution on phylogenies—a general method for the comparative analysis of discrete characters. *Proc R Soc Lond Ser B Biol Sci.* 255:37–45.
- Paland S, Lynch M. 2006. Transitions to asexuality result in excess amino acid substitutions. *Science.* 311:990–992.
- Press WH, Teukolsky SA, Vetterling WT, Flannery BP. 2002. Numerical recipes in C++: the art of scientific computing. New York: Cambridge University Press.
- Rasmussen MD, Kellis M. 2007. Accurate gene-tree reconstruction by learning gene- and species-specific substitution rates across multiple complete genomes. *Genome Res.* 17:1932–1942.
- Robinson M, Gouy M, Gautier C, Mouchiroud D. 1998. Sensitivity of the relative-rate test to taxonomic sampling. *Mol Biol Evol.* 15:1091–1098.
- Robinson-Rechavi M, Huchon D. 2000. RRTree: relative-rate tests between groups of sequences on a phylogenetic tree. *Bioinformatics.* 16:296–297.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:572–1574.
- Ross SM. 1996. Stochastic processes. New York: John Wiley & Sons.
- Rutschmann F. 2006. Molecular dating of phylogenetic trees: a brief review of current methods that estimate divergence times. *Divers Distrib.* 12:35–48.
- Sanderson MJ. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol Biol Evol.* 14:1218–1231.
- Sanderson MJ. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol Biol Evol.* 19:101–109.
- Sanderson MJ. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* 19:301–302.
- Sarich VM, Wilson AC. 1973. Generation time and genomic evolution in primates. *Science* 179:1144–1147.
- Smith SA, Donoghue MJ. 2008. Rates of molecular evolution are linked to life history in flowering plants. *Science* 322:86–89.
- Thomas JA, Welch JJ, Woolfit M, Bromham L. 2006. There is no universal molecular clock for invertebrates, but rate variation does not scale with body size. *Proc Natl Acad Sci U S A.* 103:7366–7371.
- Thorne JL, Kishino H. 2002. Divergence time and evolutionary rate estimation with multilocus data. *Syst Biol.* 51:689–702.
- Thorne JL, Kishino H. 2006. Estimation of divergence times from molecular sequence data. In: Nielsen R, editor. Statistical methods in molecular evolution. New York: Springer. p. 233–256.
- Thorne JL, Kishino H, Painter IS. 1998. Estimating the rate of evolution of the rate of molecular evolution. *Mol Biol Evol.* 15:1647–1657.
- Welch JJ, Bromham L. 2005. Molecular dating when rates vary. *Trends Ecol Evol.* 20:320–327.
- Whelan S, Goldman N. 1999. Distributions of statistics used for the comparison of models of sequence evolution in phylogenetics. *Mol Biol Evol.* 16:1292–1299.
- Whittle CA, Johnston MO. 2002. Male-driven evolution of mitochondrial and chloroplastial DNA sequences in plants. *Mol Biol Evol.* 19:938–949.
- Wu CI, Li WH. 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. *Proc Natl Acad Sci U S A.* 82:1741–1745.
- Yang Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J Mol Evol.* 39:306–314.
- Yang Z. 2004. A heuristic rate smoothing procedure for maximum likelihood estimation of species divergence times. *Acta Zool Sin.* 50:645–656.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591.
- Yang Z, Goldman N, Friday A. 1994. Comparison of models for nucleotide substitution used in maximum-likelihood phylogenetic estimation. *Mol Biol Evol.* 11:316–324.
- Yang Z, Rannala B. 1997. Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo method. *Mol Biol Evol.* 14:717–724.
- Zuckerkandl E, Pauling L. 1965. Evolutionary divergence and convergence in proteins. In: Bryson V, Vogel HJ, editors. *Evolving genes and proteins*. New York: Academic Press.