

## A Nonstationary Markov Model Detects Directional Evolution in Hymenopteran Morphology

SERAINA KLOPFSTEIN<sup>1,2,3,\*</sup>, LARS VILHELMSSEN<sup>4</sup>, AND FREDRIK RONQUIST<sup>1</sup>

<sup>1</sup>Department of Bioinformatics and Genetics, Swedish Museum of Natural History, SE-104 05 Stockholm, Sweden; <sup>2</sup>The University of Adelaide, ACEBB, Adelaide SA 5005, Australia; <sup>3</sup>Natural History Museum, Department of Invertebrates, CH-3005 Bern, Switzerland; <sup>4</sup>Biosystematics, Natural History Museum of Denmark, DK-2100 Copenhagen Ø, Denmark

\*Correspondence to be sent to: Natural History Museum, Department of Invertebrates, Bernastr. 15, CH-3005 Bern, Switzerland; E-mail: [klopfstein@nmbe.ch](mailto:klopfstein@nmbe.ch).

Received 17 January 2015; reviews returned 22 May 2015; accepted 17 July 2015

Associate Editor: Lars Jermin

**Abstract.**—Directional evolution has played an important role in shaping the morphological, ecological, and molecular diversity of life. However, standard substitution models assume stationarity of the evolutionary process over the time scale examined, thus impeding the study of directionality. Here we explore a simple, nonstationary model of evolution for discrete data, which assumes that the state frequencies at the root differ from the equilibrium frequencies of the homogeneous evolutionary process along the rest of the tree (i.e., the process is nonstationary, nonreversible, but homogeneous). Within this framework, we develop a Bayesian approach for testing directional versus stationary evolution using a reversible-jump algorithm. Simulations show that when only data from extant taxa are available, the success in inferring directionality is strongly dependent on the evolutionary rate, the shape of the tree, the relative branch lengths, and the number of taxa. Given suitable evolutionary rates (0.1–0.5 expected substitutions between root and tips), accounting for directionality improves tree inference and often allows correct rooting of the tree without the use of an outgroup. As an empirical test, we apply our method to study directional evolution in hymenopteran morphology. We focus on three character systems: wing veins, muscles, and sclerites. We find strong support for a trend toward loss of wing veins and muscles, while stationarity cannot be ruled out for sclerites. Adding fossil and time information in a total-evidence dating approach, we show that accounting for directionality results in more precise estimates not only of the ancestral state at the root of the tree, but also of the divergence times. Our model relaxes the assumption of stationarity and reversibility by adding a minimum of additional parameters, and is thus well suited to studying the nature of the evolutionary process in data sets of limited size, such as morphology and ecology. [Bayesian inference; continuous-time Markov model; directional selection; Hymenoptera; morphology; neutral evolution; nonstationary; positive selection, Symphyta.]

Since the dawn of evolutionary biology, the nature of the evolutionary process has been hotly contested. The relative importance of phenomena such as directional evolution, punctuated equilibrium, and stasis is still debated (Eldredge and Gould 1972; Hunt 2007; Pennell et al. 2014) despite numerous attempts to find signatures of the evolutionary process in the morphological, ecological, or molecular characters of extant and extinct organisms (e.g., Carmel et al. 2007; Hunt 2007; Sookias et al. 2012; Finarelli and Goswami 2013; Shoemaker and Clauset 2014). Depending on the groups examined, timescales studied, and methodologies applied, biologists tend to arrive at very different conclusions with respect to the predominant evolutionary mechanisms at play.

The term “directional evolution” is here used to refer to a nonrandom shift in the marginal distribution of traits over evolutionary time. To distinguish such trends or tendencies from stochastic events, it is typically necessary to study change over time in collections of traits or across multiple lineages. Directional change can result from a number of microevolutionary processes, both active and passive. Adaptation or directional selection is one possible explanation. For instance, selection for improved defense against predation has been invoked to explain the increase in body size over evolutionary time observed in various groups (Cope’s Rule; Cope 1885) (note that some authors have argued that the apparent increase in body size is a statistical

artifact; Hone and Benton 2005; Zanno and Makovicky 2012). Directional patterns could also result from biased starting conditions followed by an entirely unbiased process, which would create increasingly extreme traits simply due to an increase in the variance over time. This scenario was put forward by Gould (1996) to explain the apparent increase in complexity during the evolution of life, which he suggested does not reflect an underlying pull toward higher complexity, but is simply a consequence of life beginning at very low complexity levels. In this article, we focus on detecting directional patterns without addressing the specific micro- or macroevolutionary mechanisms at work.

The stochastic framework has proven useful for testing competing hypotheses concerning the nature of the evolutionary process (e.g., Blomberg et al. 2003; Hunt 2006; Pagel and Meade 2006). Besides allowing for a better understanding of the evolution of characters, more realistic stochastic evolutionary models might also increase the accuracy of phylogenetic analyses, especially if the latter are based on morphological or ecological data that match standard models poorly (Felsenstein 1973; Lewis 2001). The use of morphological characters in phylogenetics has experienced a recent resurgence with the introduction of total-evidence dating methods which attempt to accommodate the uncertainty in the phylogenetic placement of fossils using morphological data (Pyron 2011; Ronquist et al. 2012a; Wood et al. 2013). The success of these

methods relies on adequate modeling of morphological evolution.

For continuous morphological characters, various directional models have been considered, for example biased random walks (Hunt 2006), Uhlenbeck and Ornstein processes (Blomberg et al. 2003), or Lévy processes (Landis et al. 2013). However, here we focus on models for discrete morphological data. Virtually all work to date on qualitative character evolution has been based on continuous-time Markov processes running over finite state spaces. In its most general form, a Markov process only has to satisfy the criterion of being memoryless, that is, conditioned on the current state of the process, its future and past are independent (the general Markov model (GMM); Barry and Hartigan 1987). In evolutionary biology, it is commonly assumed that the process is also homogeneous, time-reversible, and at stationarity (reviewed in Jermiin et al. 2008).

Stationarity is achieved when a process has been running for long enough that its current state is no longer dependent on the starting state, but can be considered drawn from the stationary distribution of the process (Fig. 1a). Homogeneity means that the process does not change over time or across lineages (Fig. 1b). Time-reversibility refers to the fact that the stochastic fluctuations of the process at stationarity are indistinguishable regardless of the flow of time (Fig. 1c). These properties are often misunderstood. Rate asymmetries, that is different rates of forward and backward substitutions, have little to do with them; even homogeneous, time-reversible processes at stationarity can have extreme asymmetries in the exchange rates between states. Both heterogeneous and time-irreversible processes may be at stationarity, in which case the system does not display any long-term trends. Directional evolution, as we defined it (see above), is the same as saying that the process is not at stationarity; directional evolution does not imply that the process is necessarily heterogeneous (but it is time-irreversible by definition). Expressed in terms familiar to Bayesian phylogeneticists, directionality simply means that evolution is not yet out of the “burn-in phase” and thus has not yet reached the equilibrium state of the process.

The most commonly used models of molecular evolution assume stationarity, homogeneity, and time-reversibility, but several nonstationary alternatives have been considered, mostly in the context of nucleotide or amino acid sequences that vary in base composition over time or across lineages (Lockhart et al. 1994; Yang and Roberts 1995; Galtier and Gouy 1998; Foster 2004; Jayaswal et al. 2005, 2007, 2011a, 2011b, 2014; Blanquart and Lartillot 2006, 2008; Boussau and Gouy 2006; Boussau et al. 2008; Dutheil et al. 2012; Zou et al. 2012; Groussin et al. 2013). Standard models for discrete morphological data are even simpler than the ones commonly used to model molecular evolution. In addition to stationarity, homogeneity, and time-reversibility, they also assume equal state frequencies and exchangeability rates due to the arbitrary nature of

the state labels of most morphological data sets (Lewis 2001). To our knowledge, nonstationary models have not been considered for discrete morphological data previously.

Here, we explore the simplest possible nonstationary model for morphology: a homogeneous process for binary characters which is not assumed to start at equilibrium. We develop a Bayesian reversible-jump (RJ) approach to test this model against a stationary model and use simulations to identify the conditions under which it is possible to detect directional evolution. Finally, we use our method to study directional evolution in the morphology of the Hymenoptera (ants, wasps, and bees).

The literature on hymenopteran morphology is rich (e.g., Gibson 1986; Rasnitsyn 1988; Vilhelmsen 2001; Krogmann and Vilhelmsen 2006; Beutel and Vilhelmsen 2007; Vilhelmsen et al. 2010) and provides easy access to large sets of qualitative characters. There is also a vivid discussion on directional tendencies among hymenopteran morphologists, for instance toward reduction in wing venation (Brown and Nutting 1949). Xyelids, probably the sister group of all other Hymenoptera (e.g., Ronquist et al. 2012a; Malm and Nyman 2015), have the most complete venation of any hymenopteran, and the pattern is virtually identical to that of the earliest known hymenopteran fossils from the Triassic. The loss of wing veins appears to be partly associated with a reduction in body size, a common trend in several hymenopteran lineages, leading many authors to conclude that wing loss characters are unreliable for reconstructing higher-level hymenopteran phylogeny (Sharkey and Roy 2002). But body size is not the only factor driving the evolution of wing venation since there are both small hymenopterans with rich venation (e.g., xyelids) and large hymenopterans with reduced venation (e.g., orussids, ibaliids, and peleciniids).

Reductions and simplifications from the hymenopteran ground plan have also been observed in muscles, sclerites, and articulations (Vilhelmsen et al. 2010). To our knowledge, however, there has been no attempt to date to quantify these tendencies, nor to analyze them in a probabilistic framework. Here, we use our Bayesian RJ approach to test for directionality in the evolution of hymenopteran wing veins, muscles, and sclerites in a data set including both extant and fossil taxa. Finally, we examine the effect of accounting for directional morphological evolution in the context of total-evidence dating.

## MATERIALS AND METHODS

### *Implementation of a Nonstationary Model for Binary Characters*

We implemented a nonstationary model of evolution for binary characters in MrBayes 3.2 (Ronquist et al. 2012b). The model assumes a single instantaneous rate matrix over the whole tree, and allows for the state frequencies at the root to differ from the stationary

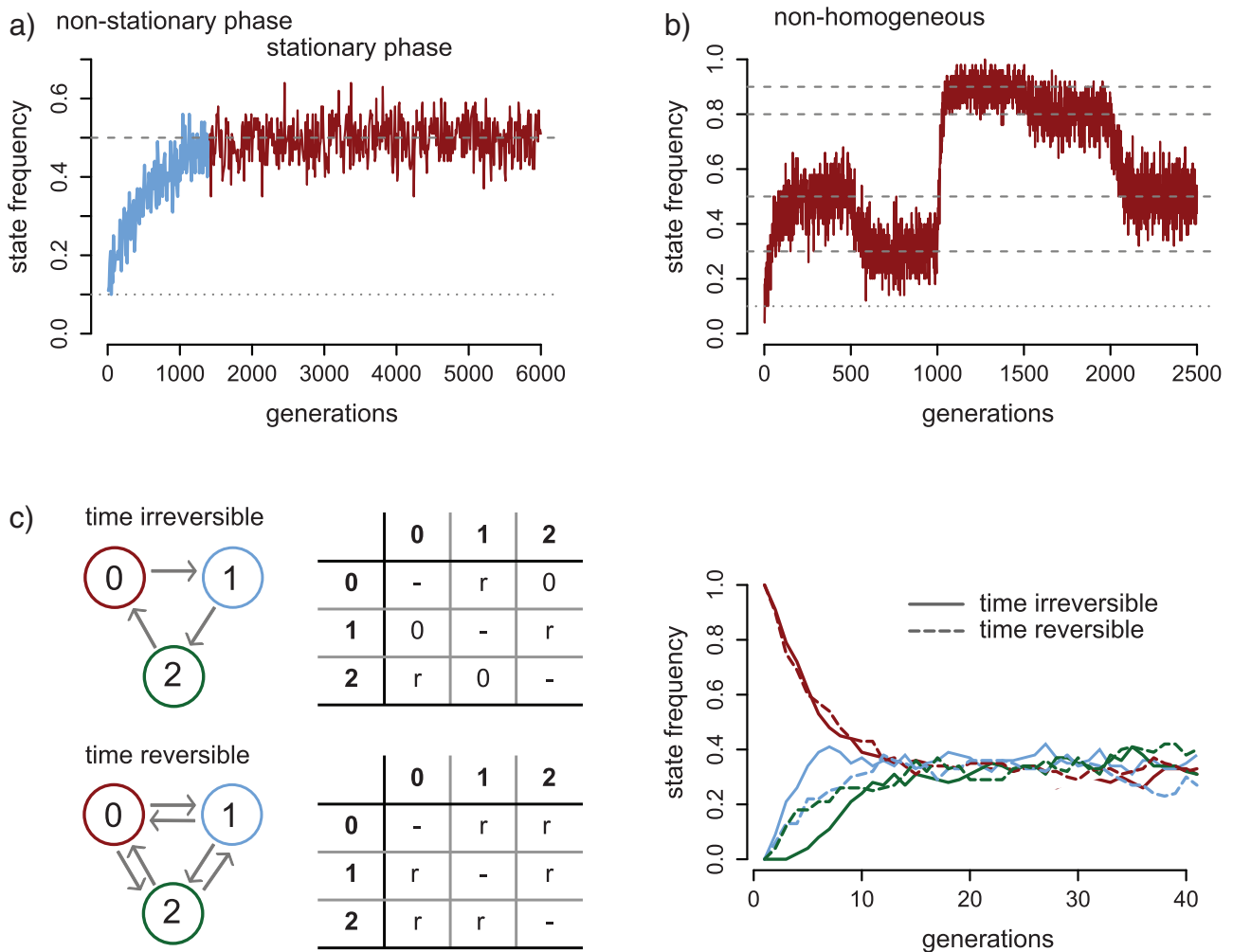


FIGURE 1. Commonly adopted simplifications of the General Markov Model: stationarity, homogeneity, and time-reversibility. a) Nonstationary and stationary phase of a Markov process with differing starting and equilibrium state frequencies (dotted and dashed lines), for a two-state character with symmetric transition probabilities. b) Nonhomogeneous Markov process with changing equilibrium frequencies (dotted and dashed lines) over time. c) Example of time-irreversible and time-reversible processes (left) for a character with three states, with their instantaneous rate matrices (middle); the equilibrium frequencies of both processes are the same but states 1 and 2 follow different trajectories during the nonstationary phase (right). Even at stationarity, there are similar asymmetries in the stochastic fluctuations of the time-irreversible process.

frequencies of the process (Boussau and Gouy 2006). The model thus accounts for nonstationarity, even though the process along the branches is homogeneous. Since the model is directional, probabilities have to be computed on rooted trees, but the position of the root does not have to be specified *a priori*. The additional parameters of the model, the root frequencies, are updated during the Markov chain Monte Carlo (MCMC) procedure in the same way as the stationary frequencies, using standard sliding window and Dirichlet proposal mechanisms. The model, which is currently only available for binary data (data type “restriction”), can be invoked in MrBayes by the command “lset statefreqmodel = directional” (instead of “stationary”, the default), and the prior on the root frequencies is set with the command “prset rootfreq = dirichlet/fixed” (see the “help” function in MrBayes for details).

To allow for direct model testing and to integrate over model uncertainty, we implemented a RJ algorithm, that is, a transdimensional MCMC (Green 1995, 2003) which jumps between the stationary and nonstationary models. We assume equal prior probabilities of the two models, and use a split-merge move as follows. In the split move, we obtain new stationary and root frequencies by two independent draws from a beta distribution centered on the stationary frequencies in the previous generation, with the variance of the distribution used as a tuning parameter. The merge move draws new stationary frequencies from a beta distribution with the mean equal to the average of the previous stationary and root frequencies, under the same tuning parameter as in the split move. This move has a Jacobian of 1, as calculated following the approach detailed in Green (2003) and confirmed by running the model without data, which

returned the prior probabilities for the stationary and directional models. As the directional-evolution model requires a rooted tree, this is assumed throughout, even while the MCMC is sampling under the stationary model and there is thus no information about the placement of the root. When there is a low but significant probability for the directional model, the information about the position of the root is gradually lost while sampling from the stationary model, reducing the acceptance rate of swaps back to the directional model. Under such conditions, a rooting constraint can be introduced to improve mixing. Both the root frequencies and stationary frequencies are updated with sliding-window and Dirichlet proposals. The command in MrBayes to invoke the RJ approach is “lset statefreqmodel = mixed”. The implementations of the directional model and RJ algorithm were tested by running the MCMC without data, which recovered the prior distributions of state frequencies, model probabilities, and topology (results not shown).

### Simulation Study

To study the performance of the model and test for circumstances under which directional evolution could be discerned based on data from only extant taxa, we performed simulations in R using the “ape” package, including the “matexpo” function (Paradis et al. 2004; R Development Core Team 2009), and a number of newly developed functions. All R scripts are available as Supplementary Material on Dryad at <http://dx.doi.org/10.5061/dryad.0sn23>. We use the term “branch” to specify a single edge of the phylogeny (in contrast to “lineage” which can denote a collection of edges) and “tree” to denote a topology with branch lengths.

We tested the influence of the shape of the tree (see below), the number of taxa sampled (8 or 16), the evolutionary rate in expected substitutions from the root to any of the tips (0.01 to 1), and the strength of directionality, that is, the difference between root and stationary frequencies (several combinations of the state frequencies 0.1, 0.3, 0.5, 0.7, and 0.9 were used, either the same at the root and at equilibrium to simulate stationarity, or different to mimic directional evolution). We used three tree shapes that represent extremes in terms of topology and branch lengths: (i) an entirely balanced tree with Grafen branch lengths (where the height of each node is equal to the number of subtended terminals minus one, divided by the root height; Grafen 1989) (tree A); (ii) an entirely unbalanced tree with Grafen branch lengths (tree B); and (iii) an entirely unbalanced tree with all branches of equal lengths (tree C) (Fig. 2). The trees were scaled so that trees A and B, whose branch lengths conform to a molecular clock, have a root-tip distance of one, while tree C was scaled so that the average root-tip distance was 1. On these trees, we simulated 1000 (in the case of trees with eight taxa) or 100 (for 16 taxa) data sets of 100 binary characters each. These data sets

were analyzed in MrBayes under the stationary and directional models and using the RJ approach, running two independent runs with one cold and one heated chain for 100,000 generations, sampling every 100th generation. We assumed that only the variable characters could be observed (using the command “lset coding = variable”), thus effectively excluding the constant characters from the analyses. Such an ascertainment bias is well suited for analyzing morphological or restriction-site data which are the focus here, but is less appropriate for data such as DNA or amino acid sequences where constant characters are also observed. We employed R scripts to study the power to distinguish between directional and stationary evolution, the impact on topology inference and root recovery, and the accuracy of the inferred state frequencies at the root and at equilibrium. Simulations and analyses were run at the National Supercomputer Center at Linköping University in Sweden (NSC).

### Empirical Example: Hymenopteran Morphology

We modified the morphological data set in Ronquist et al. (2012a) by recoding all wing vein, muscle, and sclerite characters so that they followed the same state labeling, with zeros for the state “absent” (or “fused” in muscles and sclerites) and ones for “present” (or “separate” for muscles and sclerites). All other characters were left unchanged. A list of characters is given in Supplementary Table S1, available on Dryad at <http://dx.doi.org/10.5061/dryad.0sn23>.

First, we analyzed the wing vein (23 characters), muscle (56 characters), and sclerite data (67 characters) in combination with the remaining morphological characters (208 characters), invoking a separate directional or stationary model for each of the binary partitions, under a “variable” coding bias and gamma-distributed among-character rate variation. We used a more restrictive and a more diffuse branch length prior, that is, an exponential distribution of mean 0.1 and of mean 1, respectively. The morphology-only analyses employed four independent runs of one cold and three heated chains each, sampling every 1000th of 10 million generations. Convergence was assessed through the potential scale reduction factor (PSRF, smaller than 1.01 unless mentioned otherwise) for scalar parameters and the average standard deviation of split frequencies (ASDSF, below 0.02) for the topology. To calculate Bayes factors for model comparison, we used the stepping-stone algorithm (Xie et al. 2011) with 49 steps of 1,000,000 generations each plus 1,000,000 generations as initial burn-in, and 250,000 generations burn-in within each step. The alpha parameter was set to 0.4, and sampling went from the posterior to the prior. All of the steps reached an ASDSF below 0.03, except for the last 10 steps (sampling close to the prior) in which no split reached the minimum frequency of 0.1.

We then added the molecular data from Ronquist et al. (2012a) to improve the estimates of topology and relative branch lengths, once more running the

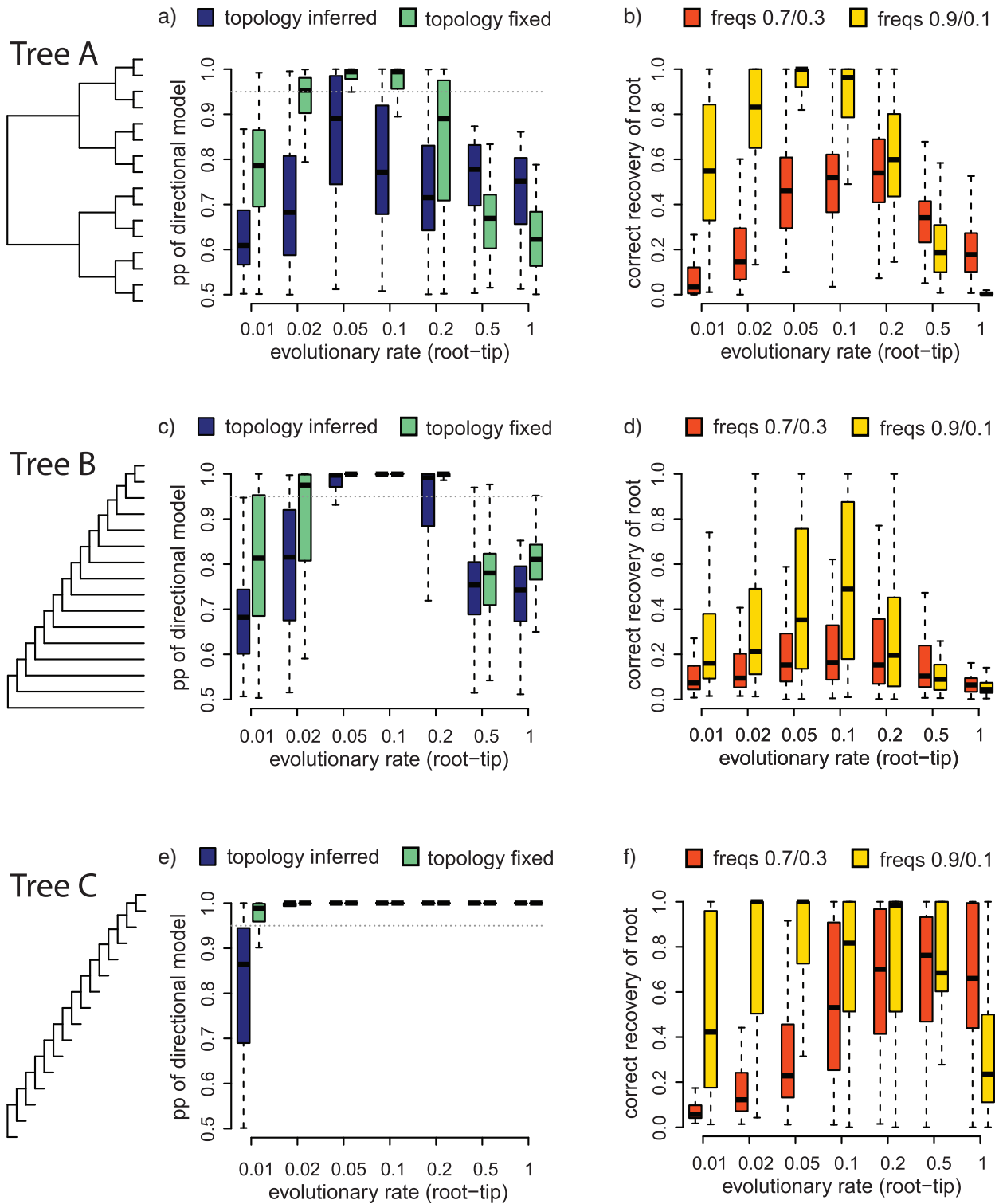


FIGURE 2. Impact of tree shape on the recovery of the directional model and the correct rooting of the tree. One hundred data sets of 100 characters were simulated for each setting. (a, c, e) Recovered posterior probability (pp) of the directional model under a fixed or inferred topology. Data sets of 100 characters each were simulated with a frequency of state 0 at the root of 0.1 and under a process with a stationary frequency of state 0 of 0.9. The *x*-axis shows the evolutionary rate in expected changes between root and tips. The horizontal line indicates the significance threshold at 0.95. (b, d, f) Proportion of correctly rooted trees inferred under the directional model for two different levels of asymmetry; root and stationary frequencies of 0.7 and 0.3, and of 0.9 and 0.1, respectively.

binary morphological partitions under the stationary and directional models using the RJ approach. The molecular data were partitioned into genes and first and second versus third codon positions in the protein-coding genes, and a separate RJ algorithm over the GTR-model subspace (Huelsenbeck et al. 2004) was used for each partition, combined with discrete gamma-distributed among-site rate variation and a proportion of invariant sites. We used an exponential distribution with rate 10 (mean 0.1) as the branch-length prior. Settings for the MCMC search and convergence diagnostics were the same as in the case of the morphology-only analyses. Model testing was performed using the RJ algorithm, but no stepping-stone approach, given the slow convergence close to the prior observed for the morphology-only analyses and the large computational effort associated with this algorithm.

Finally, we repeated the total-evidence dating analysis as detailed in Ronquist et al. (2012a), including fossil morphology and fossil ages in the data matrix, with the updated morphological data set and under the stationary and directional models as well as with the RJ approach. Topology convergence was lower than in the nonclock analyses (ASDSF of 0.047, 0.055, and 0.035, respectively, for the directional, stationary, and RJ analyses), and the clock-rate estimates showed some deviations between runs as well (PSRF of 1.07, 1.04, and 1.02, respectively). All analyses were run at the National Supercomputer Center at Linköping University in Sweden (NSC). Alignments and trees are deposited on TreeBase ([www.treebase.org](http://www.treebase.org), study number S16060).

## RESULTS

### *Recovery of the Directional Model in Simulated Data Sets*

Correct recovery of the directional model depends on several factors. First of all, the evolutionary rate has a major impact on correct inference (Figs. 2 and 3, see also Supplementary Files S3, available on Dryad at <http://dx.doi.org/10.5061/dryad.0sn23>). While very low rates (0.01–0.05 expected substitutions between root and tip) result in too little information for distinguishing between the models, rates of 0.5 and especially 1 seem already too saturated in many cases. An evolutionary rate of 0.1 or 0.2, sometimes 0.5, retains most information about the process and the tree.

The shape of the tree also plays a major role. Detecting directional evolution is most difficult on a balanced tree (tree A; Fig. 2a,b), followed by the unbalanced tree with ultrametric branch lengths (tree B; Fig. 2c,d). The unbalanced tree with branches of equal length (tree C; Fig. 2e,f) is the most favorable for inferring directional evolution. In all cases, fixing the topology to the true tree for the analysis has a positive effect. Increasing the size of the tree from 8 to 16 taxa also increases the success in correctly recovering the directional model (Fig. 3a,b), especially if directionality is strong. The proportion of analyses that recovers the correct topology is lowest for tree B, medium for tree A, and highest for tree C (Fig. 3e,f,

see also Supplementary Files S3, available on Dryad at <http://dx.doi.org/10.5061/dryad.0sn23>). In general, the chance of inferring the entire topology correctly is much higher for trees of 8 than for those of 16 taxa, as one might expect for a limited number of characters. Compared with a stationary model, topology inference improves noticeably under the directional model when the data have been simulated with a minimum level of directionality. Rooting of the tree under the directional model is easiest for trees A and C and hardest for the unbalanced tree B with ultrametric branch lengths (Fig. 2b,d,f). In all cases, stronger directionality (i.e., larger differences between the root and stationary frequencies) improves the correct placement of the root.

The success of inferring the root frequencies and stationary frequencies of the directional model varies with the evolutionary rate measured from root to tip (Fig. 3c,d). At low rates, the process is still far from reaching its equilibrium state, and the estimated stationary state frequencies are strongly biased toward the starting frequencies. At rates above 0.5, saturation diminishes the signal about the true state frequencies at the root. The most accurate inference of both root and stationary state frequencies occurs at intermediate rates (0.05 to 0.2), but even in those cases, the difference between root and stationary frequencies is often underestimated.

When the trees are not fixed to the true topology and very high rates are assumed (two or more expected substitutions between root and tips), we observe false positives in the recovery of the directional model, and the same is the case with entirely random data (Supplementary File S4, available on Dryad at <http://dx.doi.org/10.5061/dryad.0sn23>). Further analyses show that under such circumstances, that is without reliable information about topology and relative branch lengths, and in combination with an inadequate branch length prior (we used an exponential distribution with rate 10 throughout), the analyses recover a balanced tree with extreme directionality because this parameter combination increases the probability of observing a large number of substitutions on short branches.

### *Nonclock Analyses of the Hymenopteran Data Set*

The analysis of the morphological data alone results in highest posterior probabilities for the directional model in the case of wing veins and muscles, and just below significant support for the sclerites (Table 1 and Fig. 4a–c). The root and stationary state frequencies differ most strongly in the wing veins (Fig. 4a), followed by the muscles (Fig. 4b), while they are rather similar in the case of the sclerites (Fig. 4c). The Bayes factor estimated via stepping-stone sampling also strongly favors the directional model (71.6 on the log scale). While MCMC sampling produces very similar topology samples across runs under all models, convergence on tree height and partition-specific rate multipliers is slower under the

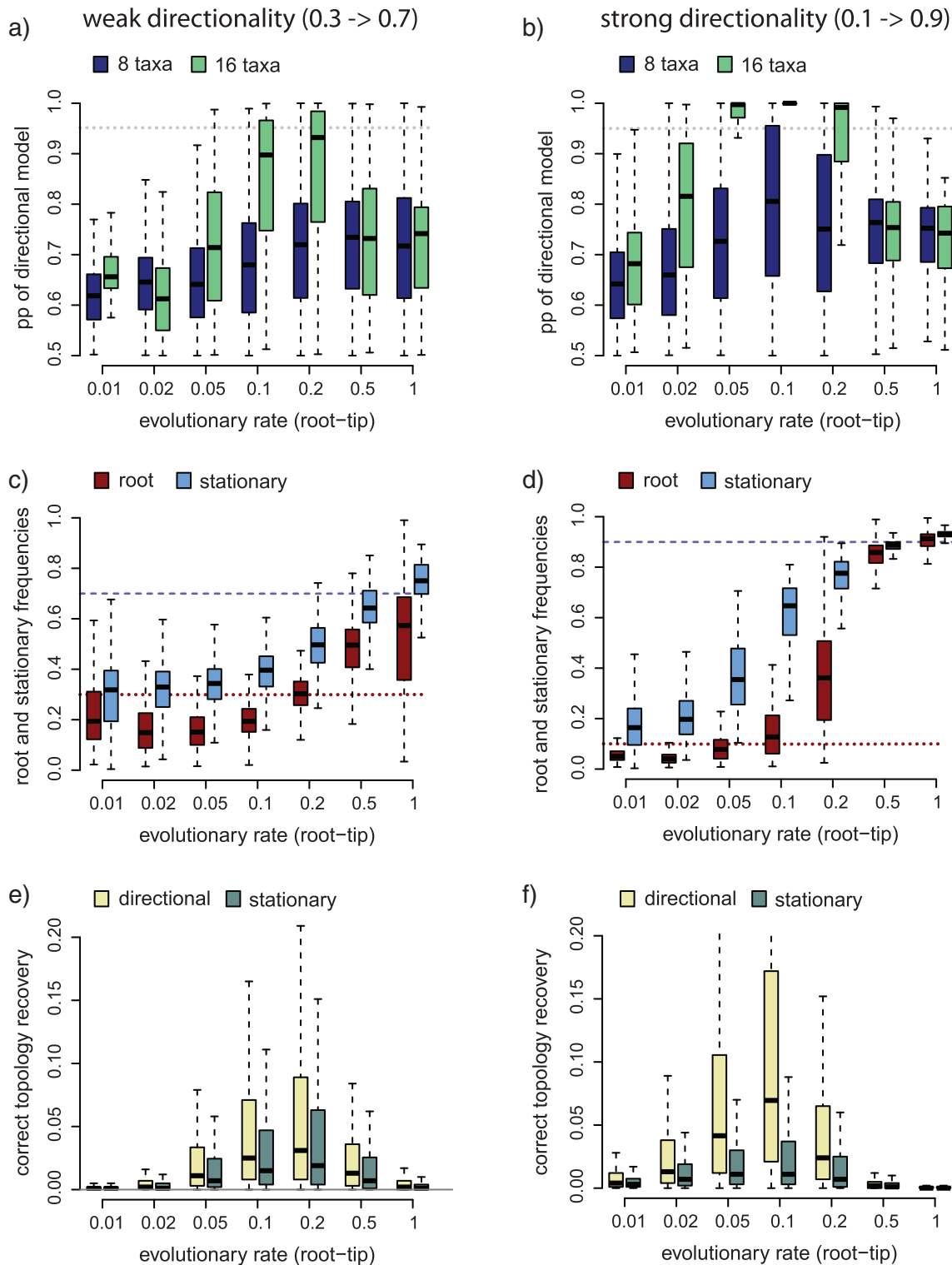


FIGURE 3. Simulation results from data sets with medium directionality (root and stationary frequencies of state 0 of 0.3 and 0.7, respectively; a, c, and e) and strong directionality (0.1 and 0.9, respectively; b, d, and f). One hundred data sets of 100 characters were simulated for each setting, using a completely unbalanced tree of eight taxa (unless stated otherwise) with clock-like branch lengths (Tree B). (a, b) Recovered posterior probability of the directional model for trees of 8 versus 16 taxa. (c, d) Estimated frequencies of state 0 at the root and at equilibrium. The horizontal lines indicate the simulation conditions for root (dotted line) and equilibrium frequencies (dashed line). (e, f) Proportion of times the correct, unbalanced 8-taxon topology was recovered when analyzing the data under the correct (directional) or under the incorrect (stationary) model.

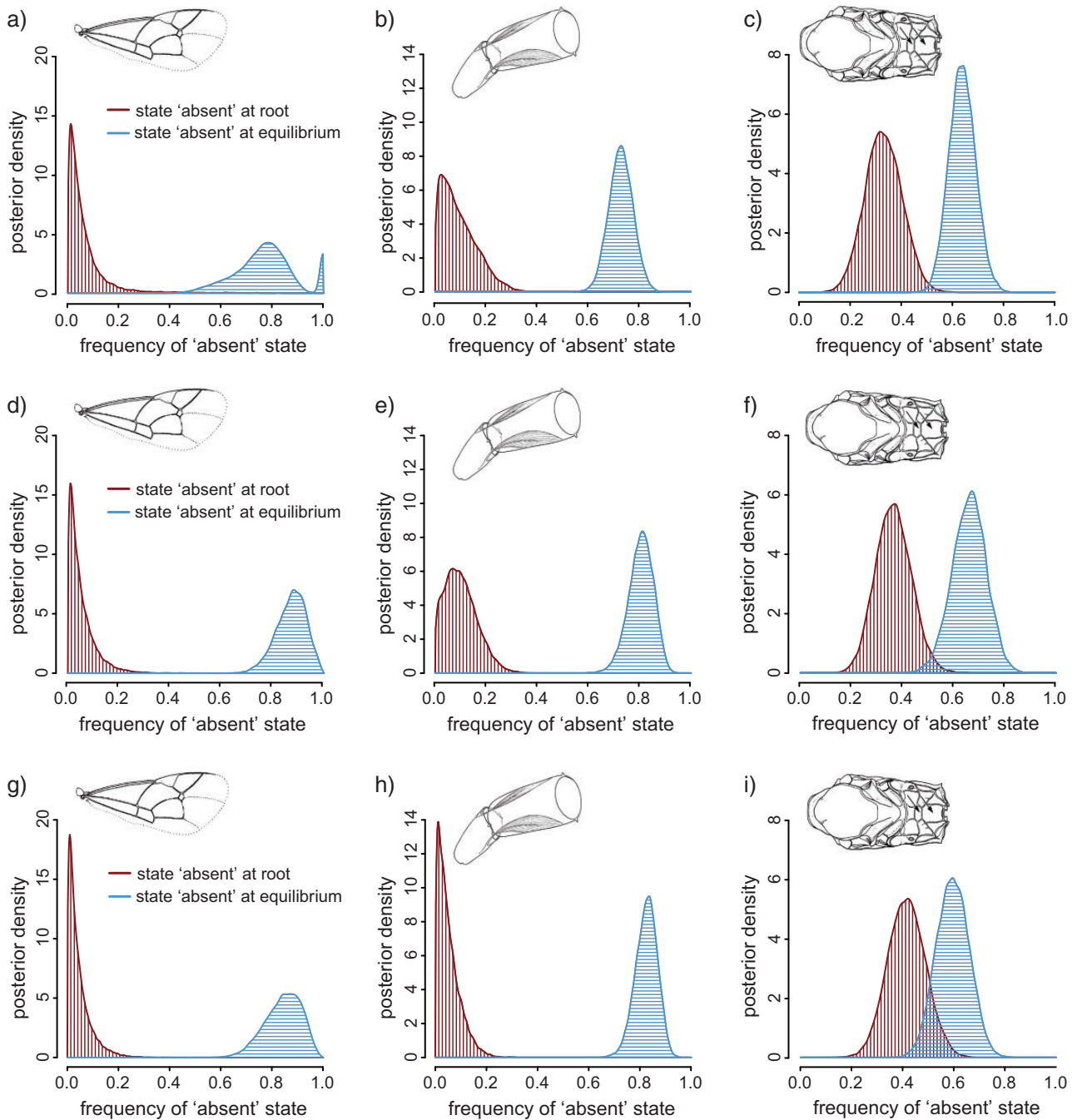


FIGURE 4. Posterior densities of the state frequency for the state “absent” at the root and at equilibrium as estimated from the analysis of morphology alone (a, b, c), of the combined morphology-molecules data set (d, e, f), and with added fossil data analyzed in a total-evidence dating approach (g, h, i). The different morphological partitions are wing veins (a, d, g), muscles (b, e, h), and sclerites (c, f, i).

directional model (PSRF values around 1.04), which indicates a problem with estimating relative branch lengths under this model when the amount of data is limited. The estimated tree height and rate multipliers are also quite sensitive to the branch-length prior. The tree height estimate increases from a median of 6.59 under an exponential branch-length prior with mean 0.1 to 117 under a prior with mean 1. Convergence, especially of the partition-specific rate multipliers, is much worse

under the more diffuse branch-length prior, with PSRF values of up to 1.86.

Adding molecular data resolves the problem of inferring branch lengths and rate multipliers. The posterior probability of the directional model remains maximal for wing veins and muscles, but drops to 0.8 for the sclerite partition (Table 1), with partly overlapping posterior densities recovered for the root and stationary state frequencies (Fig. 4f).



TABLE 1. Inferred root and equilibrium frequencies and posterior probabilities of the directional model as estimated through an RJ approach for three morphological character sets

|   | Frequency at root | Frequency at equilibrium | pp of directional model |
|---|-------------------|--------------------------|-------------------------|
| Morphology only                             |                   |                          |                         |
| Wing veins                                  | 0.04              | 0.78                     | 1.00                    |
| Muscles                                     | 0.08              | 0.73                     | 1.00                    |
| Sclerites                                   | 0.33              | 0.64                     | 0.948                   |
| Morphology and molecular data               |                   |                          |                         |
| Wing veins                                  | 0.03              | 0.88                     | 1.00                    |
| Muscles                                     | 0.09              | 0.81                     | 1.00                    |
| Sclerites                                   | 0.36              | 0.66                     | 0.80                    |
| Dated total-evidence analysis, with fossils |                   |                          |                         |
| Wing veins                                  | 0.03              | 0.83                     | 1.00                    |
| Muscles                                     | 0.04              | 0.81                     | 1.00                    |
| Sclerites                                   | 0.41              | 0.59                     | 0.38                    |

#### *Adding Fossil and Time Information*

Fossils and the corresponding time information promise a much better chance for inferring the actual evolutionary process than when relying on data from extant taxa alone. The fossil data added here are very unevenly distributed among the character partitions, with 99% of the wing veins coded for the fossils, 7.2% of the sclerites, but none of the muscles. Adding the fossil information in a total-evidence dating approach does not change the main conclusions, with wing veins and muscles still achieving highest support for the directional model (Table 1). The difference between estimated root and stationary frequencies increases even further and estimates are more precise than those from the analyses with extant taxa alone (Fig. 4g,h). The posterior densities of the state frequencies of the sclerite partition, however, now overlap broadly (Fig. 4i), indicating no support for the directional model for this partition. The extent of directionality in certain character systems becomes apparent when plotting the wing veins of selected fossil and extant taxa on a random tree drawn from the posterior distribution (the topology of which is concordant with the majority-rule consensus tree from the RJ analysis, Fig. 5).

The stationary and directional models, as well as the RJ approach sampling under both models, all retrieve similar median ages for Hymenoptera of approximately 305 Ma (Table 2). However, the precision of the estimates increases from the stationary to the directional model, and even further in the RJ approach, with the 95% credibility intervals decreasing from 63 to 38 million years (Table 2).

## DISCUSSION

### *A Nonstationary Markov Model*

While the standard model for analyzing discrete morphological characters assumes stationarity of the evolutionary process (Lewis 2001), the nonstationary Markov model for binary characters suggested here

allows the modeling of biological processes that have not yet reached their equilibrium state. The only relaxation of the Lewis model is that the state frequencies at the root of the phylogenetic tree are allowed to differ from the equilibrium frequencies of the homogeneous process running along the rest of the tree. The rate matrix is potentially asymmetric, allowing different forward and backward rates between the two states. The model can also accommodate heterogeneous distributions of state frequencies among terminals, if these are caused by differences in evolutionary rates across lineages.

In terms of macroevolutionary inference, the model proposed here might be useful to distinguish between stasis and directional change. While the state frequencies at the root coincide with the stationary frequencies in the former, the difference between them drives directionality in the latter. This model is especially useful in contexts where the evolutionary process within a group is believed to differ from the conditions in its sister groups and thus at the root of the tree, for instance after the appearance of a key innovation, the colonization of a new geographical area, or the occupation of a new ecological niche. While we apply the model to morphological data coded with nonarbitrary (“presence–absence”) state labels (Lewis 2001), it is also suited for other binary data types, for example presence or absence of introns, restriction sites, gene duplications, or lateral gene transfers, or even behavioral or ecological traits (e.g., Nylander et al. 2004; Alekseyenko et al. 2008; Klopstein et al. 2010; Stern et al. 2010; Klopstein and Ronquist 2013). While currently only implemented for binary data, an extension of the model to more states is possible, for instance to model the evolution of nucleotide or amino acid sequences.

Other variations of the finite-state Markov model that also relax the assumptions of stationarity, time-reversibility, and homogeneity have been proposed (for an overview, see Jermin et al. 2008; Jayaswal et al. 2014). Barry and Hartigan (1987) detailed what is now referred to as the GMM, which assumes separate, arbitrary substitution matrices along each branch of the phylogeny and is thus both very flexible and very parameter rich. It has been demonstrated to suffer from identifiability problems (Zou et al. 2011). For amino acids, Blanquart and Lartillot (2008) suggested a model that is both site and time heterogeneous; even though heterogeneous, it can be seen as stationary in the broad sense because they assume that the state frequencies at the root of the tree correspond to the stationary frequencies of the mixture of their heterogeneous processes. The same applies to a three-state Markov model employed by Stern et al. (2010) to study lateral gene transfer, and to two more general models proposed by Jayaswal et al. (2011b).

Truly nonstationary models were suggested repeatedly (Yang and Roberts 1995; Galtier and Gouy 1998; Pagel and Meade 2004; Jayaswal et al. 2005, 2007; Blanquart and Lartillot 2006; Zou et al. 2012) to address phylogenetic inference from nucleotide or amino acid sequences that vary in base composition over time. Most of them introduced restricted versions of the GMM

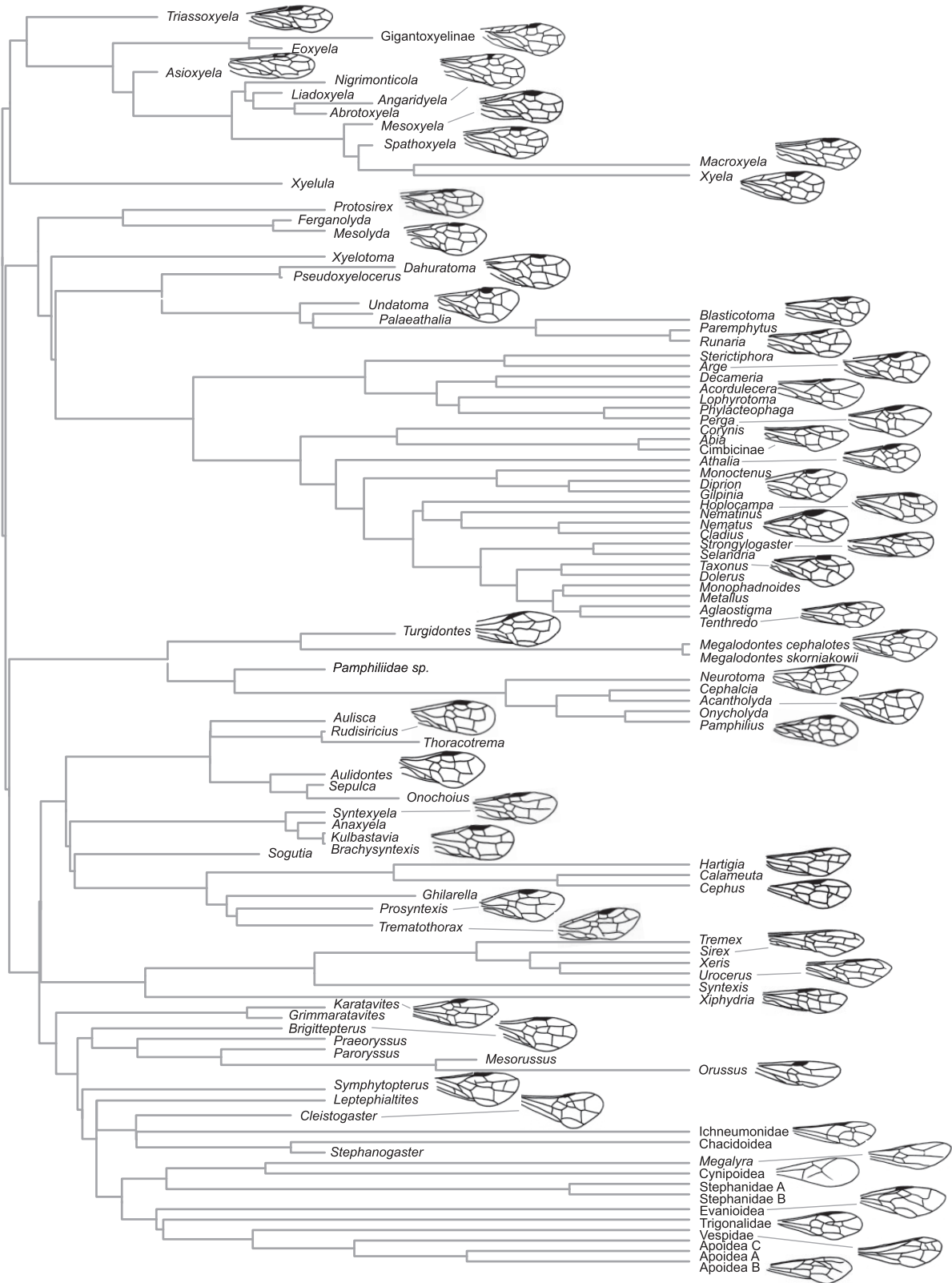


FIGURE 5. Wing veins of selected fossil and extant taxa shown on a dated total-evidence tree from the posterior tree distribution of the RJ analysis.

TABLE 2. Age estimates for the ancestor of extant Hymenoptera from total-evidence dating analyses under the stationary and directional models and under an RJ approach sampling under both models

|                   | Median age <sup>a</sup> | 95% credibility interval <sup>1</sup> | Interval width |
|-------------------|-------------------------|---------------------------------------|----------------|
| Stationary model  | 305.5                   | 287–350                               | 63             |
| Directional model | 305.0                   | 289–333                               | 44             |
| RJ approach       | 304.2                   | 289–327                               | 38             |

<sup>a</sup>The realized prior on the age of Hymenoptera, as obtained from an MCMC run without data, had a median of 544.5 and a credibility interval of 376–749.

which allow for some heterogeneity of the process, but use far less parameters. In their models, only the state frequencies are allowed to vary among branches, while a single, simplified vector of exchangeability rates is assumed for the entire tree. To circumvent the problem of overparameterization that comes with branch-wise models, alternatives were suggested that either use a fixed number of different compositions (Foster 2004), use clustering methods and/or likelihood-based model testing to determine optimal numbers of parameters (Jayaswal et al. 2011a, 2014; Dutheil et al. 2012; Groussin et al. 2013), or employ a Poisson process to model changes in the stationary state frequencies on the tree (Blanquart and Lartillot 2006).

All these models relax the assumption of stationarity while at the same time allowing for heterogeneity of the process along the tree. However, assuming time-heterogeneity beyond the origin of the studied group might not always be necessary to achieve an adequate fit of the model to the data at hand. Our model can be seen as the simplest possible nonstationary model; it will thus be particularly useful for situations where there is a limited amount of data and in contexts where the largest deviations from stationarity are expected at the root of the tree.

#### *Performance in Simulations*

Our simulations suggest that the best conditions for detecting directional evolution are unbalanced trees with nonclock branch lengths, a fixed topology, many taxa, and a strong directional drive (the latter being determined by the difference between the state frequencies at the root and at equilibrium). But even under such favorable conditions, the signal left by the directional process can only be detected for a range of evolutionary rates that typically span 0.1 to 0.5 expected substitutions between the root and the tips of the tree. Interestingly, similar rates have been shown to also be optimal for phylogenetic inference and branch length estimation (Goldman 1998; Felsenstein 2004; Susko and Roger 2012).

The shape of the tree also has a profound effect on the detection of directional evolution, with unbalanced trees being beneficial, especially if they have short terminal branches attaching close to the root. Such trees might result in contexts where a key innovation

speeds up speciation and evolutionary rates at the same time (Mooers and Heard 1997). The shape of the hymenopteran tree, with the very slowly evolving family Xyelidae as the sister group to the remaining taxa (Ronquist et al. 2012a), is certainly reminiscent of such a pattern.

Even under favorable conditions and at suitable evolutionary rates, the differences between stationary and root frequencies are often underestimated in our simulations, as stationary frequencies were not yet reached in the terminals before the signal of the root frequencies began to erode. This relates to the difficulties associated with inferring evolutionary processes in the distant past from extant taxa only. Even when using adequate models that allow for directionality, inference based on extant taxa only can be misleading, both positively and negatively, while the incorporation of fossil information leads to a dramatic increase in accuracy and precision both in simulated and empirical data (Slater et al. 2012; Finarelli and Goswami 2013). Nevertheless, especially under certain tree shapes, the signal left by directional evolution might still be detectable in extant taxa using parametric statistical inference under appropriate models.

Allowing the state frequencies at the root to differ from the stationary state frequencies requires the computation of rooted instead of unrooted trees. Our simulations show that adequately modeled, directional evolution can even be used to infer the root of a tree, at least for a certain period after the onset of the process (Fig. 3e,f, Huelsenbeck et al. 2002). Moreover, not accounting for directionality cannot only mislead estimation of the ancestral states in a phylogeny (Susko and Roger 2013), but also of the phylogeny itself, as is evident from our simulation results and has been shown previously for model misspecifications in general (e.g., Felsenstein 1978; Huelsenbeck and Rannala 2004).

Our simulations were intended to mimic morphological data sets and were thus performed assuming an ascertainment bias which excludes constant characters. Especially at low evolutionary rates, which result in numerous constant characters, this will probably have an impact on the inference of directionality, as constant characters can retain a lot of information about the conditions at the root. For studies based on different data types, for instance nucleotide or amino acid sequences which contain constant characters as well, recovery rates of the directional model might thus be higher. Further studies on simulated and empirical data sets are needed to evaluate the interplay between ascertainment biases and the inference of directional evolution.

At high evolutionary rates and also with random data, we found erroneous preference for the directional model under inappropriate branch-length priors. Support for the directional model was often very high under such conditions, but largely disappeared when additional information was provided about the topology and relative branch lengths, or if a more adequate branch-length prior was used. To improve confidence in the

inference of directionality, it is thus advisable to examine the sensitivity to and adequacy of the branch-length prior, and to add complementary information about the topology and relative branch lengths, for instance by adding molecular data. Including information from fossil taxa, if available, can also help address the problem of false positives.

### *The Evolution of Hymenopteran Morphology*

Our results strongly support directional evolution in hymenopteran wing venation and musculature, while the picture is more complex in the case of sclerites. Even though the posterior probability of directionality in the evolution of sclerites was just shy of the 95% mark in the pure-morphology analyses, adding molecular or fossil data weakened the support considerably. Therefore, it seems likely that the signal in the pure-morphology analyses is spurious. Why is there no directionality in the evolution of sclerites? One possibility is that the hymenopteran exoskeleton evolves under conflicting selection pressures. On the one hand, there may be advantages to the structural simplifications that can result from the loss of sclerites. On the other hand, the exoskeleton might compensate for the loss of muscles by increasing in complexity, using more parts and more complex interactions between parts to allow better control of movement by fewer muscles. Several examples could be mentioned to support this idea. However, until we achieve a more profound understanding of the functional morphology of the hymenopteran body, this largely remains the realm of speculation.

With respect to wing venation and musculature, our results suggest that hymenopteran evolution has essentially been a story of losses. In the case of wing venation, the systematic trend toward simplification across multiple lineages is beautifully illustrated by the fossil record (Fig. 5). In fact, the frequencies of presence and absence of wing veins among extant terminals included in our analysis are quite different from the estimated stationary frequencies, suggesting that this character system is still far from reaching equilibrium. The same holds true for muscle characters. It is interesting to note that we were able to document such strong directional trends despite the fact that our analysis included so few representatives of the Apocrita, the explosive terminal radiation of originally insect-parasitic hymenopterans. It is especially evident within this group that reduction of morphological features is correlated with reduction in body size, with the most simplified morphologies found among the smallest hymenopteran taxa, for example Ceraphronoidea (Miko et al. 2013), Chalcidoidea (Krogmann and Vilhelmsen 2006), Mymaromatidae (Vilhelmsen and Krogmann 2006), and Platygasteridae (Austin et al. 2005). Whether the trend toward simplification in hymenopteran morphology is due to directional selection, for instance through lower costs of producing fewer wing veins or muscles, or due to developmental constraints favoring losses over gains, is still unclear. However, the fact that

the trends are so strong even among basal hymenopteran lineages, mostly including insects of moderate to large body size, suggests that reduction in body size is at least not the only causal factor. Dollo-like evolution (Gould 1970; Dollo 1893), that is, the irreversibility of losses of complex traits, probably also plays an important role. As evolution toward smaller body sizes has happened several times independently and was probably reversed multiple times during hymenopteran evolution, and as the loss of many muscles and wing veins seems to be a consequence of small body size (Vilhelmsen et al. 2010), one could even envision an interaction of these two causes, combining in a ratchet-like mechanism leading to an accumulation of losses in different body-size-dependent character systems (c.f. Lukes et al. 2011).

Few other studies have found empirical support for directional evolution. Hunt (2007) compiled a comprehensive set of over 250 fossil sequences monitoring trait evolution in single lineages over mostly 0.5 to 6 million years. He found that only 5% of these sequences showed statistical support for directional change as opposed to stasis or unbiased random walks, even though these fossil sequences had often been recorded with such a change in mind. Hunt's study differs strongly from our analysis, both in temporal or taxonomical scope and in methodology. Using data from several lineages in a phylogenetic framework, as we exemplify here, certainly increases the power to detect directional patterns dramatically. Furthermore, instead of examining single characters, our data set included 23–67 characters in each of the morphological partitions, and our simulations even assumed 100 characters, which certainly increased the statistical power considerably. However, a researcher might in many cases only have a few characters or even a single character for which she would like to test directionality; it is reasonable to assume that our model is useful also in such a context. The reduced power can be compensated for by increasing the number of terminals sampled; additional simulations are needed to evaluate the required number. In the context of character-dependent diversification (the BiSSE model, Maddison et al. 2007), the authors of a recent simulation study suggest using at least 300 terminals (Davis et al. 2013), but our model is less complex than theirs and might thus be sufficiently powerful even with a smaller sample. Sanderson (1993) studied rate asymmetry and found that given a sufficient evolutionary rate, as few as 32 taxa can be enough to distinguish between evolutionary models.

Even though our model was clearly better than the stationary model in explaining the evolution of hymenopteran morphology, it might be possible to improve the fit to this particular data set even further. In particular, it would be interesting to explore the possibility of allowing a limited number of changes between evolutionary regimes across the tree, instead of assuming a single regime. For instance, the poorly sampled Apocrita represent a group that has undergone rather dramatic changes at its base, with the shift to a parasitic lifestyle, the appearance of a wasp waist,

and strongly increased rates of diversification (Heraty et al. 2011; Aguiar et al. 2013; Klopstein et al. 2013). One might suspect that these changes are associated with detectable modifications in the overall process of morphological evolution. A first step toward modeling process heterogeneity would thus be to allow for a separate model in this part of the tree. However, similar shifts might of course have occurred in other parts of the tree as well. Ideally, one would like to infer both the number of different evolutionary regimes and the points of change between them. In the extreme case, such a heterogeneous model of morphological evolution might estimate different regimes on each branch of the tree (cf. Yang and Roberts 1995; Galtier and Gouy 1998; Jayaswal et al. 2005), which might lead to overparameterization. Some model component that puts a fairly strong constraint on the number of regimes would be advisable for typically rather scarce morphological characters, such as a compound Poisson process or a Dirichlet process prior (cf. Blanquart and Lartillot 2006). Unfortunately, models like these are often difficult to sample from efficiently using standard Bayesian MCMC algorithms, and the statistical power is likely to decrease rapidly with the number of postulated evolutionary regimes. How to model the process that changes evolutionary regimes is a critical issue in this context.

Another interesting possibility would be to consider dependence among discrete morphological characters. We treated the characters as independent of each other, which could potentially be problematic if their evolution was in fact tightly linked, even though the effect of such linkage among characters on the inference of directionality is not entirely clear. If character change is fundamentally directional, and the linkage is a simple matter of the state of some characters affecting the rate of change of others, then the net effect might simply be overconfidence in the directional nature of evolution. One could also imagine a scenario in which atomic changes are essentially nondirectional, but the dependence among characters creates the impression of directionality. However, to be able to distinguish such character-dependence scenarios from independent directional and stationary models, we will probably need detailed models of the nature of the potential linkage among morphological characters. In formulating such models, both developmental and functional links need to be considered.

In hymenopteran wing veins, we generally observe a reduction from the posterior to the anterior margin of the wing and from its apex toward the base, which could form the basis of character-dependence models. In muscles and sclerites, there are a few cases of potential developmental or functional constraints, for instance in serially homologous structures such as the meso- and metafurca, meso- and metatibial spurs, and certain sets of muscles in the meso- and metathorax, or in functionally linked muscle groups. Even though it would be interesting to try and model such character-linkage phenomena, we probably need larger data sets

(more terminals and more characters) to convincingly demonstrate their presence. For instance, it seems likely that the underlying developmental constraints in wing venation (e.g., the presence of more basal veins being necessary for the formation of a more apical vein) will only start playing a major role once a certain number of wing veins have been lost, which is not the case in the relative modest range of venational complexity seen in the taxa sampled for this study. Similarly, the observed distributions of skeletal and musculature features among the taxa examined here suggest that it would be difficult to demonstrate any potential character linkages in these systems (Vilhelmsen et al. 2010). Nevertheless, modeling character dependence certainly has a lot of potential, as exemplified for molecular evolution (e.g., accounting for secondary or tertiary structure in rRNA and proteins; Schoeniger and von Haeseler 1994; Robinson et al. 2003). While not yet feasible for most morphological character systems, the formulation of more complex evolutionary models will in the future greatly increase the rigor of phylogenetic analyses involving character sets that violate the standard models; further research in this area is sorely needed. Another possible extension to our approach would be to model the correlation between the evolution of morphological characters and the evolutionary process itself. Using this approach, one might for instance be able to account for the influence of body size on the extent and speed of evolutionary reductions in morphological character systems. Such a framework has been presented for quantitative characters based on multivariate Brownian motion models (Lartillot and Poujol 2011). This type of model remains to be extended to discrete characters, but the Brownian framework would actually be adequate for the study of the correlation between body size and evolutionary transition rates, both of which are quantitative traits. For Hymenoptera, the problem here lies more with the data than the analysis framework: It appears unlikely that a sample of less than 1% of the described taxa would be sufficient to infer body size evolution with any degree of precision.

These are just a few examples of possible extensions to our model. Clearly, there is still much to learn from more sophisticated probabilistic modeling and statistical inference of morphological evolution.

#### SUPPLEMENTARY MATERIAL

Data available from the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.0sn23>.

#### FUNDING

This work was supported by the Swiss National Science Foundation [fellowship grant PA00P3\_145377 to S.K.] and by the Swedish Research Council [grant VR-2011-5622 to F.R.]. Computations were performed using the Swedish National Infrastructure for Computing under project SNIC 2013/1-194.

## ACKNOWLEDGMENTS

The authors thank associate editor Lars Jermiin, editor-in-chief Frank Anderson, Jeffrey Thorne, and an anonymous reviewer for useful comments that helped improve an earlier version of this article.

## REFERENCES

- Aguilar A.P., Deans A.R., Engel M.S., Forshage M., Huber J.T., Jennings J.T., Johnson N.F., Lelej A.S., Longino J.T., Lohrmann V., Miko I., Ohl M., Rasmussen C., Taeger A., and Yu D.S. 2013. Order Hymenoptera. *Zootaxa* 3703:1–82.
- Alekseyenko A.V., Lee C.J., Suchard M.A. 2008. Wagner and Dollo: A stochastic duet by composing two parsimonious solos. *Syst. Biol.* 57:772–784.
- Austin A.D., Johnson N.F., Dowton M. 2005. Systematics, evolution and biology of scelionid and platygastriid wasps. *Annu. Rev. Entomol.* 50:553–582.
- Barry, D., Hartigan J.A. 1987. Statistical analysis of hominoid molecular evolution. *Stat. Sci.* 2:191–210.
- Beutel R.G., Vilhelmsen L. 2007. Head anatomy of Xyelidae (Hexapoda: Hymenoptera) and phylogenetic implications. *Organ. Divers. Evol.* 7:207–230.
- Blanquart, S., Lartillot N. 2006. A Bayesian compound stochastic process for modeling nonstationary and nonhomogeneous sequence evolution. *Mol. Biol. Evol.* 23:2058–2071.
- Blanquart, S., Lartillot, N. 2008. A site- and time-heterogeneous model of amino acid replacement. *Mol. Biol. Evol.* 25:842–858.
- Blomberg S.P., Garland T.J., Ives A.R. 2003. Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution* 57:717–745.
- Boussau B., Blanquart S., Necsulea A., Lartillot N., Gouy M. 2008. Parallel adaptations to high temperatures in the Archean eon. *Nature* 456:942–945.
- Boussau B., Gouy M. 2006. Efficient likelihood computations with nonreversible models of evolution. *Syst. Biol.* 55:756–768.
- Brown W.L.J., Nutting W.L. 1949. Wing venation and the phylogeny of the Formicidae (Hymenoptera). *Trans. Am. Entomol. Soc.* 75:113–132.
- Carmel L., Wolf Y.I., Rogozin I.B., Koonin E.V. 2007. Three distinct modes of intron dynamics in the evolution of eukaryotes. *Genome Res.* 17:1034–1044.
- Cope E.D. 1885. On the evolution of the Vertebrata, progressive and retrogressive. *Am. Nat.* 19:140–148.
- Davis M.P., Midford P.E., Maddison W.P. 2013. Exploring power and parameter estimation of the BiSSE method for analyzing species diversification. *BMC Evol. Biol.* 13:38.
- Dollo L. 1893. Les lois de l'évolution. *Bull. Soc. Belge Géol. Paléont. Hydrol.* 7:164–6.
- Dutheil J.Y., Galtier N., Romiguier J., Douzery E.J.P., Ranwez V., Boussau B. 2012. Efficient selection of branch-specific models of sequence evolution. *Mol. Biol. Evol.* 29:1861–1874.
- Eldredge N., Gould S.J. 1972. Punctuated equilibria: An alternative to phyletic gradualism. In: T. J. Schopf M., editor. *Models in paleobiology*. San Francisco: Freeman Cooper. p. 82–115.
- Felsenstein J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. *Am. J. Human Genet.* 25:471–492.
- Felsenstein J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* 27:401–410.
- Felsenstein J. 2004. *Inferring phylogenies*. Sunderland (MA): Sinauer Associates.
- Finarelli J.A., Goswami A. 2013. Potential pitfalls of reconstructing deep time evolutionary history with only extant data, a case study using the Canidae (Mammalia, Carnivora). *Evolution* 67:3678–3685.
- Foster P.G. 2004. Modeling compositional heterogeneity. *Syst. Biol.* 53:485–495.
- Galtier N., Gouy M. 1998. Inferring pattern and process: Maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. *Mol. Biol. Evol.* 15:871–879.
- Gibson G.A.P. 1986. Evidence for monophyly and relationships of Chalcidoidea, Mymaridae, and Mymarommatidae (Hymenoptera: Terebrantes). *Can. Entomol.* 118:205–240.
- Goldman N. 1998. Phylogenetic information and experimental design in molecular systematics. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* 265:1779–1786.
- Gould S.J. 1970. Dollo on Dollo's law: Irreversibility and the status of evolutionary laws. *J. Hist. Biol.* 3:189–212.
- Gould S.J. 1996. *Full house: The spread of excellence from Plato to Darwin*. New York: Harmony Books.
- Grafen A. 1989. The phylogenetic regression. *Phil. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 326:119–156.
- Green P.J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82:711–732.
- Green P.J. 2003. Trans-dimensional Markov chain Monte Carlo. In: Green P.J., Hjort N.L., Richardson S., editors. *Highly structured stochastic systems*. Oxford: Oxford University Press. p. 179–198.
- Groussin M., Boussau B., Gouy M. 2013. A branch-heterogeneous model of protein evolution for efficient inference of ancestral sequences. *Syst. Biol.* 62:523–538.
- Heraty J.M., Ronquist F., Carpenter J.M., Hawks D., Schulmeister S., Dowling A.P.G., Murray D.L., Munro J., Wheeler W.C., Schiff N., Sharkey M.J. 2011. Evolution of the hymenopteran megaradiation. *Mol. Phylogenet. Evol.* 60:73–88.
- Hone D.W.E., Benton M.J. 2005. The evolution of large size: How does Cope's Rule work? *Trends Ecol. Evol.* 20:4–6.
- Huelsenbeck J.P., Bollback J.P., Levine A.M. 2002. Inferring the root of a phylogenetic tree. *Syst. Biol.* 51:32–43.
- Huelsenbeck J.P., Larget B., Alfaro M.E. 2004. Bayesian phylogenetic model selection using reversible-jump Markov chain Monte Carlo. *Mol. Biol. Evol.* 21:1123–1133.
- Huelsenbeck J.P., Rannala B. 2004. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Syst. Biol.* 53:904–913.
- Hunt G. 2006. Fitting and comparing models of phyletic evolution: Random walks and beyond. *Paleobiology* 32:578–601.
- Hunt G. 2007. The relative importance of directional change, random walks, and stasis in the evolution of fossil lineages. *Proc. Natl. Acad. Sci. USA* 104:18404–18408.
- Jayaswal V., Ababneh F., Jermiin L.S., Robinson J. 2011a. Reducing model complexity of the general Markov model of evolution. *Mol. Biol. Evol.* 24:1–21.
- Jayaswal V., Jermiin L., Robinson J. 2005. Estimation of phylogeny using a general Markov model. *Evol. Bioinform. Online* 1:62–80.
- Jayaswal V., Jermiin L.S., Poladian L., Robinson J. 2011b. Two stationary nonhomogeneous Markov models of nucleotide sequence evolution. *Syst. Biol.* 60:74–86.
- Jayaswal V., Robinson J., Jermiin L. 2007. Estimation of phylogeny and invariant sites under the general Markov model of nucleotide sequence evolution. *Syst. Biol.* 56:155–162.
- Jayaswal V., Wong T.K.F., Robinson J., Poladian L., Jermiin L.S. 2014. Mixture models of nucleotide sequence evolution that account for heterogeneity in the substitution process across sites and across lineages. *Syst. Biol.* 63:726–742.
- Jermiin L.S., Jayaswal V., Ababneh F., Robinson J. 2008. Phylogenetic model evaluation. In: Keith J.M., editor. *Bioinformatics: Data, sequence analysis, and evolution*. Totawa: Humana Press. p. 331–364.
- Klopfstein S., Quicke D.L.J., Kropf C. 2010. The evolution of antennal courtship in diplazontine parasitoid wasps (Hymenoptera, Ichneumonidae, Diplazontinae). *BMC Evol. Biol.* 10:218.
- Klopfstein S., Ronquist F. 2013. Convergent intron gains in hymenopteran elongation factor-1a. *Mol. Phylogenet. Evol.* 67:266–276.
- Klopfstein S., Vilhelmsen L., Heraty J.M., Sharkey M.J., Ronquist F. 2013. The hymenopteran tree of life: Evidence from protein-coding genes and objectively aligned ribosomal data. *PLoS One* 8:e69344.
- Krogmann L., Vilhelmsen L. 2006. Phylogenetic implications of the mesosomal skeleton in Chalcidoidea (Hymenoptera: Apocrita)—tree searches in a jungle of homoplasy. *Invertebr. Syst.* 20:615–674.

- Landis M.J., Schraiber J.G., Liang M. 2013. Phylogenetic analysis using Lévy processes: Finding jumps in the evolution of continuous traits. *Syst. Biol.* 62:193–204.
- Lartillot N., Poujol R. 2011. A phylogenetic model for investigating correlated evolution of substitution rates and continuous phenotypic characters. *Mol. Biol. Evol.* 28:729–744.
- Lewis P.O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst. Biol.* 50:913–925.
- Lockhart P.J., Steel M.A., Hendy M.D., Penny D. 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. *Mol. Biol. Evol.* 11:605–612.
- Lukes J., Archibald J.M., Keeling P.J., Doolittle W.F., Gray M.W. 2011. How a neutral evolutionary ratchet can build cellular complexity. *IUBMB Life* 63:528–537.
- Maddison W.P., Midford P.E., Otto S.P. 2007. Estimating a binary character's effect on speciation and extinction. *Syst. Biol.* 56:701–710.
- Malm T., Nyman T. 2015. Phylogeny of the symphytan grade of Hymenoptera: New pieces into the old jigsaw (fly) puzzle. *Cladistics* 2014:1–17.
- Miko I., Masner L., Johannes E., Yoder M.J., Deans A.R. 2013. Male terminalia of Ceraphronoidea: Morphological diversity in an otherwise monotonous taxon. *Insect Syst. Evol.* 44:261–347.
- Mooers A.O., Heard S.B. 1997. Inferring evolutionary process from phylogenetic tree shape. *Quart. Rev. Biol.* 72:31–54.
- Nylander J.A.A., Ronquist F., Huelsenbeck J.P., Nieves-Aldrey J.L. 2004. Bayesian phylogenetic analysis of combined data. *Syst. Biol.* 53:47–67.
- Pagel M., Meade A. 2004. A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. *Syst. Biol.* 53:571–581.
- Pagel M., Meade A. 2006. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. *Am. Nat.* 167:808–825.
- Paradis E., Claude J., Strimmer K. 2004. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290.
- Pennell M.W., Harmon L.J., Uyeda J.C. 2014. Is there room for punctuated equilibrium in macroevolution? *Trends Ecol. Evol.* 29:23–32.
- Pyron R.A. 2011. Divergence time estimation using fossils as terminal taxa and the origins of Lissamphibia. *Syst. Biol.* 60:466–481.
- R Development Core Team. 2009. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
- Rasnitsyn A.P. 1988. An outline of evolution of the hymenopterous insects (Order Vespida). *Orient. Insects* 22:115–145.
- Robinson D.M., Jones D.T., Kishino H., Goldman N., Thorne J.L. 2003. Protein evolution with dependence among codons due to tertiary structure. *Mol. Biol. Evol.* 20:1692–1704.
- Ronquist F., Klopfstein S., Vilhelmsen L., Schulmeister S., Murray D.L., Rasnitsyn A.P. 2012a. A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. *Syst. Biol.* 61:973–999.
- Ronquist F., Teslenko M., Van der Mark P., Ayres D.L., Darling A., Höhna S., Larget B., Liu L., Suchard M.A., Huelsenbeck J.P. 2012b. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61:539–542.
- Sanderson M.J. 1993. Reversibility in evolution: A maximum likelihood approach to character gain/loss bias in phylogenies. *Evolution* 47:236–252.
- Schoeniger M., von Haeseler A. 1994. A stochastic model and the evolution of autocorrelated DNA sequences. *Mol. Phylogenet. Evol.* 3:240–247.
- Sharkey M.J., Roy A. 2002. Phylogeny of the Hymenoptera: A reanalysis of the Ronquist et al. (1999) reanalysis, emphasizing wing venation and apocritan relationships. *Zool. Scr.* 31:57–66.
- Shoemaker L., Clausen A. 2014. Body mass evolution and diversification within horses (family Equidae). *Ecol. Lett.* 17:211–220.
- Slater G.J., Harmon L.J., Alfaro M.E. 2012. Integrating fossils with molecular phylogenies improves inference of trait evolution. *Evolution* 66:3931–3944.
- Sookias R.B., Butler R.J., Benson R.B.J. 2012. Rise of dinosaurs reveals major body-size transitions are driven by passive processes of trait evolution. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* 279:2180–2187.
- Stern A., Mayrose I., Penn O., Shaul S., Gophna U., Pupko T. 2010. An evolutionary analysis of lateral gene transfer in thymidylate synthase enzymes. *Syst. Biol.* 59:212–225.
- Susko E., Roger A.J. 2012. The probability of correctly resolving a split as an experimental design criterion in phylogenetics. *Syst. Biol.* 61:811–821.
- Susko E., Roger A.J. 2013. Problems with estimation of ancestral frequencies under stationary models. *Syst. Biol.* 62:330–338.
- Vilhelmsen L. 2001. Phylogeny and classification of the extant basal lineages of the Hymenoptera (Insecta). *Zool. J. Linn. Soc.* 131:393–442.
- Vilhelmsen L., Krogmann L. 2006. Skeletal anatomy of the mesosoma of *Palaeomyia anomala* (Blood & Kryger, 1922) (Hymenoptera: Mymaromatidae). *J. Hym. Res.* 15:290–306.
- Vilhelmsen L., Mikó I., Krogmann L. 2010. Beyond the wasp-waist: Structural diversity and phylogenetic significance of the mesosoma in apocritan wasps (Insecta: Hymenoptera). *Zool. J. Linn. Soc.* 159:22–194.
- Wood H.M., Matzke N.J., Gillespie R.G., Griswold C.E. 2013. Treating fossils as terminal taxa in divergence time estimation reveals ancient vicariance patterns in the palpimanoid spiders. *Syst. Biol.* 62:264–284.
- Xie W., Lewis P.O., Fan Y., Kuo L., Chen M.-H. 2011. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Syst. Biol.* 60:150–160.
- Yang Z., Roberts D. 1995. On the use of nucleic acid sequences to infer early branchings in the tree of life. *Mol. Biol. Evol.* 12:451–458.
- Zanno L.E., Makovicky P.J. 2012. No evidence for directional evolution of body mass in herbivorous theropod dinosaurs. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* 280:20122526.
- Zou L., Susko E., Field C., Roger A.J. 2011. The parameters of the Barry and Hartigan general Markov model are statistically nonidentifiable. *Syst. Biol.* 60:872–875.
- Zou L., Susko E., Field C., Roger A.J. 2012. Fitting nonstationary general-time-reversible models to obtain edge-lengths and frequencies for the Barry–Hartigan model. *Syst. Biol.* 61:927–940.