Taxon sampling and the accuracy of phylogenetic analyses

Tracy A. HEATH[*]   Shannon M. HEDTKE[*]   David M. HILLIS[*, **]

(*Section of Integrative Biology and Center for Computational Biology and Bioinformatics, One University Station C0930,*
The University of Texas, Austin, TX 78712, USA)

Abstract   Appropriate and extensive taxon sampling is one of the most important determinants of accurate phylogenetic estimation. In addition, accuracy of inferences about evolutionary processes obtained from phylogenetic analyses is improved significantly by thorough taxon sampling efforts. Many recent efforts to improve phylogenetic estimates have focused instead on increasing sequence length or the number of overall characters in the analysis, and this often does have a beneficial effect on the accuracy of phylogenetic analyses. However, phylogenetic analyses of few taxa (but each represented by many characters) can be subject to strong systematic biases, which in turn produce high measures of repeatability (such as bootstrap proportions) in support of incorrect or misleading phylogenetic results. Thus, it is important for phylogeneticists to consider both the sampling of taxa, as well as the sampling of characters, in designing phylogenetic studies. Taxon sampling also improves estimates of evolutionary parameters derived from phylogenetic trees, and is thus important for improved applications of phylogenetic analyses. Analysis of sensitivity to taxon inclusion, the possible effects of long-branch attraction, and sensitivity of parameter estimation for model-based methods should be a part of any careful and thorough phylogenetic analysis. Furthermore, recent improvements in phylogenetic algorithms and in computational power have removed many constraints on analyzing large, thoroughly sampled data sets. Thorough taxon sampling is thus one of the most practical ways to improve the accuracy of phylogenetic estimates, as well as the accuracy of biological inferences that are based on these phylogenetic trees.
Key words   consistency, long-branch attraction, phylogenetic accuracy, phylogenomics, systematic error, taxon sampling, Tree of Life.

The past two decades have seen great progress in reconstructing the Tree of Life. The endeavor of inferring the relationships among all living things is not only of intrinsic interest to biologists, but also has many practical applications throughout biology. Phylogenetic trees allow biologists to make predictions about biology, because we can infer when and where various structures, molecules, or behaviors have evolved in living organisms. Trees also provide information about the expected distribution of these features across taxonomic groups. Moreover, phylogenies facilitate the interpretation of comparative observations by accounting for the historical non-independence of organisms when analyzing across various levels of biological organization (e.g., genes, genomes, individuals, populations, species, or clades).

The number of practical applications of phylogenetics continues to grow each year. For example, phylogenetics has become crucial for comparing and interpreting the various genome sequencing projects.

The very reason that many such projects are undertaken is to provide a broad evolutionary spectrum for interpretation of genome function and evolution in the framework of the Tree of Life (Eisen, 1998). Without a phylogenetic framework, every genome would be a new independent mystery, and the detection of gene function would be greatly hindered without comparative analyses. In the absence of phylogenetic comparisons, studying many gene functions in the human genome would require experimentation and manipulation that is not practical or ethical in humans. Therefore, in a very real sense, the Tree of Life helps us to understand how humans function and how we differ from one another and from other species. The same is also true of all the organisms that we eat, of all the organisms that make us sick, of all the organisms that maintain our ecosystems, and of all the organisms that make the biological world interesting, entertaining, and beautiful.

Applications of phylogenetics are by no means limited to the functional and structural study of genomes, however. Phylogenetic applications span much of biology, from human health (Bush et al., 1999) and forensics (Hillis & Huelsenbeck, 1994; Metzker et al., 2002) to conservation biology (Crandall et al., 2000)

and studies of behavior (Martins, 1996). Many of these applications require accurate phylogenetic estimates, not only in terms of tree topologies, but also in branch lengths (for estimation of time and/or the amount of change), ancestral character states (for estimation of evolutionary transitions), and parameters of evolutionary models (for study of evolutionary processes). In general, accurate molecular phylogenetic estimates (estimates that represent true historical relationships among species) are dependent on four primary factors (Swofford et al., 1996): (1) appropriate selection of target genes for analysis; (2) collection of enough sequence data to obtain a robust and repeatable estimate; (3) use of accurate analytical methods; and (4) sufficient taxon sampling for the problem of interest. The first three of these factors often receive the greatest attention from investigators, but increased taxon sampling can be one of the most practical and feasible approaches for improving phylogenetic estimates (Zwickl & Hillis, 2002). Here we explore and review the effects of taxon sampling on phylogenetic analyses and their applications.

1   Dense taxon sampling improves phylogenetic accuracy

Phylogeneticists have long acknowledged that data sets containing a large number of taxa create a more complex computational problem for phylogenetic analysis. As more taxa are added to a phylogenetic data set, the number of possible tree topologies increases very rapidly. In addition, the degree of homoplasy (convergent changes or reversals) increases with the number of taxa (Sanderson & Donoghue, 1989). Regardless, numerous studies on the importance of dense taxon sampling have indicated that introducing additional taxa into a phylogenetic analysis results (on average) in more accurate estimates of evolutionary relationships (Lecointre et al., 1993; Philippe & Douzery, 1994; Hillis, 1996, 1998; Graybeal, 1998; Rannala et al., 1998; Zwickl & Hillis, 2002; Pollock et al., 2002; Poe, 1998a, 2003; DeBry, 2005; Hedtke et al., 2006). These studies represent a broad range of approaches including simulations, examinations of well-studied biological groups, and comparisons to known phylogenies. Each of these approaches has distinct advantages and disadvantages (Hillis, 1995) and together they provide a strong and consistent message about the importance of dense taxon sampling. The benefits of denser taxon sampling are especially evident in conjunction with more thorough searches of solution space (Fig. 1).

Additionally, evaluations of phylogenetic analyses often attribute problematic reconstruction and low resolution to inadequate taxon sampling (e.g., Bremer et al., 1999; Johnson, 2001; Lin et al., 2002; Braun & Kimball, 2002; Chen et al., 2003; Freudenstein et al., 2003; Sorenson et al., 2003; Albrecht et al., 2007).

Although the importance of taxonomic sampling has been intensely investigated, many studies have focused primarily on parsimony and distance methods. Felsenstein (1978) demonstrated that under certain circumstances, parsimony methods are inconsistent, meaning they converge on an incorrect topology as more and more characters are added for a limited sample of taxa. When two non-adjacent taxa share many homoplastic character states along long branches, parsimony methods often interpret such similarity as homology. The resulting tree depicts the two taxa as sister to one another, attributing the shared changes to a branch joining them; this effect is termed long-branch attraction (LBA). Inconsistency is not restricted to parsimony, however, as all phylogenetic reconstruction methods can exhibit this behavior if their assumptions are seriously violated or if there are not enough taxa in the analysis to accurately estimate the parameters of the evolutionary model (Felsenstein, 1978; Hendy & Penny, 1989; DeBry, 1992; Huelsenbeck & Hillis, 1993; Yang, 1994; Huelsenbeck, 1995;
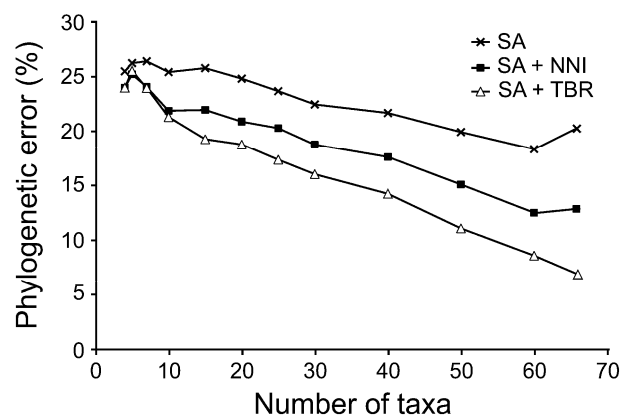

Fig. 1.   Error in phylogenetic reconstruction typically decreases with increased taxon sampling of a given taxonomic group.   The benefits of increased taxon sampling are particularly evident when searches of the solution space are more thorough. In this graph (adapted from Zwickl & Hillis, 2002, fig. 6), phylogenetic error decreases with increased taxon sampling across all analyses. However, the benefits of adding additional taxa are smaller if only the stepwise-addition algorithm (*SA*) is used to find an approximate solution, compared to the more thorough searches provided by stepwise-addition plus nearest-neighbor-interchanges branch-swapping (*SA+NNI*) or tree bisection-reconnection branch-swapping (*SA+TBR*). Analyses of larger data sets generally require more thorough search algorithms (and thus more computational effort), but result in greatly decreased phylogenetic error.

Lockhart et al., 1996; Gascuel et al., 2001; Huelsenbeck & Lander, 2003; Susko et al., 2004; Philippe et al., 2005). For example, maximum likelihood estimation has been shown to be inconsistent in the presence of severe branch-length heterogeneity (heterotachy, a form of non-stationarity) if the substitution process is assumed to be homogeneous across all lineages (Kolaczkowski & Thornton, 2004; Spencer et al., 2005; Philippe et al., 2005). This example emphasizes the need for probabilistic models that incorporate complex evolutionary processes (Yang & Roberts, 1995; Galtier & Gouy; 1998; Foster, 2004; Blanquart & Lartillot, 2006; Gowri-Shankar & Rattray, 2007; Blanquart & Lartillot, 2008; Kolaczkowski & Thornton, 2008).

Including additional taxa in a phylogenetic analysis will increase the accuracy of the inferred topology by dispersing homoplasy across the tree and reducing the effect of long-branch attraction. Hillis (1996) analyzed data simulated on a 228-taxon tree and showed that simple parsimony and distance methods accurately reconstruct the true topology when provided with sequences 5,000 nucleotides in length. At the time, this result was surprising because it seemingly contradicted the common belief that accurate phylogenetic reconstruction from very large data sets was infeasible. Moreover, Hillis et al. (1994b) had previously shown that analyses of much smaller data sets, containing only 4 taxa, required considerably longer sequences to attain the same level of accuracy. The results of Hillis's (1996) large-scale simulation indicated that for phylogenies containing many taxa,

convergent substitutions or reversals (homoplasy) are distributed among the many lineages in the tree and therefore such misleading information is less likely to overwhelm the true phylogenetic signal.

Because inadequate species sampling can result in trees containing relatively long terminal branches, sparsely sampled data sets are more likely to be affected by LBA. Rannala et al. (1998) simulated ultrametric trees under a simple model of cladogenesis to investigate the impact of removing ingroup taxa on the distribution of branch lengths. They demonstrated that decreasing the proportion of sampled taxa leads to an increase in the average length of terminal branches and generates tree shapes that may be susceptible to long-branch attraction (Fig. 2). Huelsenbeck and Lander (2003) simulated sequences using simple evolutionary models and determined that the probability that parsimony is inconsistent becomes greater as the proportion of taxa sampled decreases and substitution rates increase. Even under very simple models of evolution, unweighted parsimony underestimated the number of changes along branches and converged on an incorrect topology (Huelsenbeck & Lander, 2003).

In general, many studies have shown that adding taxa to bisect long branches can mitigate the effect of LBA (Hendy & Penny, 1989; Graybeal, 1998; Poe & Swofford, 1999; Poe, 2003). However, taxon addition should be practiced judiciously to ensure that enough taxa are added to sufficiently partition multiple long branches (Graybeal, 1998; Poe, 2003) and that the new taxa do not introduce additional long branches (Kim, 1998). Prudent taxon addition is particularly
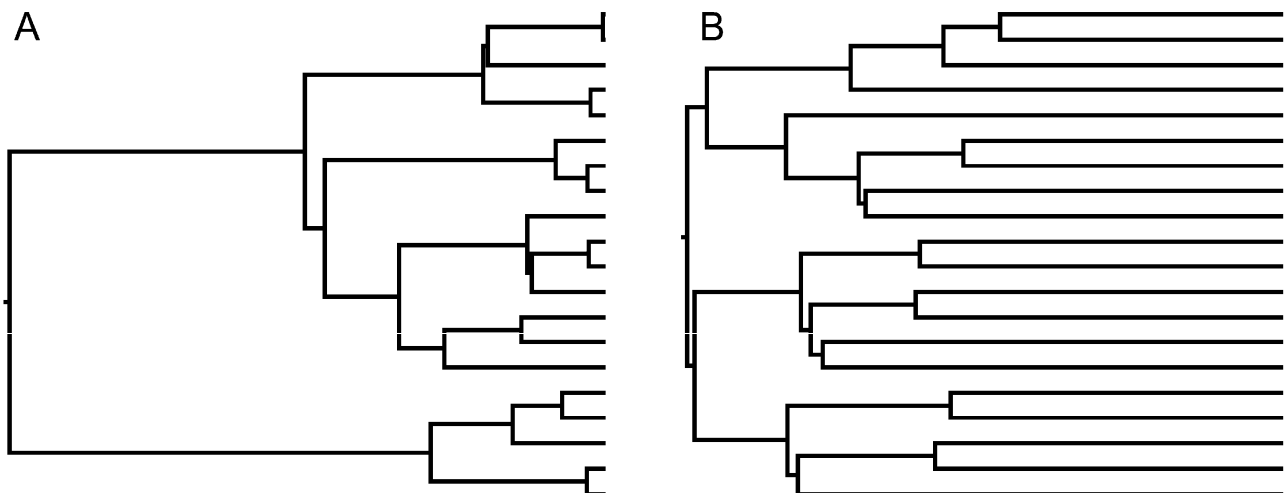

Fig. 2. Two simulations of a birth-death process to model cladogenesis. The speciation rate (λ) and extinction rate (μ) were fixed throughout the simulation and arbitrarily set to $\lambda/\mu=2$. **A,** Phylogenetic tree with complete (100%) taxon sampling (20 taxa total). **B,** Phylogenetic tree with 10% taxon sampling (20 taxa sampled from 200 taxa total). When taxon sampling is low, terminal branch lengths are longer, indicating that sparsely sampled data sets are susceptible to the effects of long-branch attraction. Adapted from Rannala et al. (1998; figs. 1 and 2).

important when conducting parsimony analyses since this method is especially liable to inconsistency due to long-branch attraction. Because parametric methods, such as maximum likelihood, incorporate models that account for unobserved substitutions, these methods are less prone to the effects of long-branch attraction, as long as the models of evolution are adequate. However, enough taxa must be sampled to parameterize these models effectively (Pollock et al., 2002). In addition, longer branches require more accurate models of evolution (because more unobserved changes must be inferred), so increased taxon sampling (which breaks up long branches) greatly benefits parametric methods as well as nonparametric methods. We discuss methods for detecting and minimizing LBA more completely below.

Apart from its effect on topological accuracy, the density of taxon sampling also has an impact on branch-length estimation. Branch lengths provide important information about the amount of change that has occurred over the tree and are critical for applications using phylogenies to make inferences about evolution. Under the parsimony criterion, branch lengths are often underestimated in sparsely sampled regions of the tree because less information is available to infer the history of unobserved substitutions (Fitch & Bruschi, 1987; Fitch & Beintema, 1990). This artifact has been termed the node-density effect (NDE) and may mislead studies that investigate correlations between rates of molecular evolution and biodiversity (Webster et al., 2003; Venditti et al., 2006; Hugall & Lee, 2007). Maximum likelihood, Bayesian, and distance methods are also susceptible to node-density effects, particularly when the assumed model of sequence evolution is overly simple and substitution rates are high (Gojobori et al., 1982; Bruno & Halpern, 1999; Hugall & Lee, 2007). If the density of taxon sampling is increased, additional internal nodes can reveal undetected substitutions and improve estimates of branch lengths.

It has been shown that mis-estimation of branch lengths can, in turn, lead to biased tree topologies (Xia, 2006). Errors in estimates of genetic distance become greater as the amount of divergence between two sequences increases. Pairwise distance methods for phylogenetic reconstruction typically use log-transformed formulae to account for unobserved substitutions (Swofford et al., 1996; Hoyle & Higgs, 2003). When using log-transformed formulae to calculate genetic distances, particularly at high levels of sequence divergence, there is a significant probability that the distance estimates will be undefined even if the "true" model of sequence evolution is assumed (Hoyle & Higgs, 2003). Therefore, when conducting distance-based analyses, it is very important to consider how taxa are sampled and avoid inclusion of highly divergent sequences.

1.1   Increased taxon sampling versus increased sequence length

Increasing the total number of characters in a data set can increase resolution and support for a phylogeny (Hillis et al., 1994a; Graybeal, 1994; Rannala et al., 1998). In particular, increasing character data such as the number of genes (or total number of nucleotides) should reduce stochastic error or character sampling bias (Phillips et al., 2004; Delsuc et al., 2005). The rapidly increasing amount of sequence data available to researchers from whole genomes, expressed sequence tags or cDNA libraries, and individual gene-based studies means that many analyses of these character-rich phylogenetic matrices can greatly reduce stochastic error.

This plethora of sequence data has caused some researchers to argue that large character data sets alone are sufficient to estimate an accurate phylogeny, notwithstanding the conclusions reached by numerous studies showing the importance of taxon sampling. For example, Rosenberg and Kumar (2001) conducted a simulation study indicating that adding taxa to a problematic phylogeny is less effective than adding additional characters. This paper led to a debate in the literature and reanalyses of the Rosenberg and Kumar (2001) data (Zwickl & Hillis, 2002; Pollock et al., 2002; Rosenberg & Kumar, 2003; Hillis et al., 2003). Pollock et al. (2002) reanalyzed the Rosenberg and Kumar (2001) data using a different approach to summarizing results (measurement of error), and Zwickl and Hillis (2002) re-conducted the Rosenberg and Kumar (2001) study with a different approach to study design that examined a fuller spectrum of taxon sampling strategies. Both studies concluded that taxon sampling has a very strong and positive effect on the accuracy of phylogenetic reconstruction, and showed in many cases that increasing the number of taxa had a much greater beneficial effect than increasing the number of characters. However, because of the ever-increasing availability and generation of genomic data and the difficulty of obtaining sequence data for many taxa, the debate about the relative importance of taxon sampling versus character sampling continues in the literature (Hillis et al., 2003; Rosenberg & Kumar, 2003; Rokas et al., 2003; Cummings & Meyer, 2005; Rokas et al., 2005; Hedtke et al., 2006; Gatesy et al., 2007).

More recently, Rokas et al. (2003) argued that, under parsimony, an accurate phylogeny for seven yeast taxa could be obtained using twenty genes randomly selected from a data set of 106, regardless of taxonomic sampling. The authors based this claim on the bootstrap support for each node in their tree: once twenty genes were sampled, the bootstrap support for each node rose above 95%, and the topology was identical to that of the data set using all 106 genes. Therefore, they concluded, accuracy was increased by increasing the number of genes, with the implication that many phylogenetic studies may be using an insufficient number of genes to accurately reconstruct topologies. The most obvious problem with their study design was that the authors equated high bootstrap support with accuracy (rather than more properly with repeatability). Analyzing many characters may result in convergence on a single answer, but small systematic biases can result in convergence on the incorrect answer (a point we explore in the next section).

1.2   Increased character data can reduce stochastic error, but can contribute to systematic error

Increasing the number of nucleotides will not solve inaccurate reconstruction due to method inconsistency, or systematic error. All phylogenetic methods to date have conditions in which they will perform inconsistently. Phylogenomic data sets, which tend to have large numbers of characters with relatively few, widely dispersed taxa, may be particularly prone to problems with long-branch attraction. LBA has been identified as causing spurious relationships in large-scale studies of several taxonomic groups, including mammals (Lin et al., 2002), metazoans (Anderson & Swofford, 2004; Baurain et al., 2007), arthropods (Delsuc et al., 2003) and angiosperms (Stefanovic et al., 2004). When methods are inconsistent, increasing character data can increase statistical support for an inaccurate phylogeny (Huelsenbeck & Hillis, 1993). This occurs because measures of support such as the nonparametric bootstrap proportion or Bayesian posterior probabilities are conditional on the data and the method (Delsuc et al., 2005).

A reanalysis of the Rokas et al. (2003) data set by Hedtke et al. (2006) showed that much of the conflict between genes under parsimony is due to LBA. The taxa sampled are unevenly dispersed across the tree of yeast, and a long branch leading to the outgroup has a tendency to pull particular taxa together across many independent genes. Hedtke et al. (2006) additionally used genes simulated on a more densely sampled yeast phylogeny to demonstrate that for a given bipartition affected by long-branch attraction, bootstrap support for the wrong reconstruction increased as genes were added to the analysis (Fig. 3). However, when taxon sampling was increased, fewer genes were needed to get acceptable support for the correct bipartition. Although Hedtke et al. (2006) did not examine increases in support or accuracy across the entire tree, their results are consistent with the general finding that increased taxon sampling improves accuracy across the tree.

This discussion so far has assumed that all genes share the same evolutionary history, and that by increasing the number of nucleotides a researcher increases the signal for that one history — i.e., that topological conflicts between trees based on single genes are stochastic. However, genes do not always share an identical history, because of horizontal gene transfer, introgression, incomplete lineage sorting, or gene duplication/loss. In some cases, it may be more appropriate to analyze genes separately rather than concatenating all the data (e.g., Ane & Sanderson, 2005), particularly if a researcher is interested in identifying processes that may have affected the evolution of a species.

2   Strategies for effective taxon sampling

We have discussed how increased taxon sampling can generally improve topological estimates. Unfortunately, there is no "magic number" of taxa and genes which ensures the accuracy of a phylogeny. The percentage of taxa sampled within a taxonomic group may be more important than the total number of taxa (Yang & Goldman, 1997; Hillis, 1998). For small clades of closely-related taxa, reduced taxonomic sampling may not be problematic for phylogeny reconstruction (Poe, 1998b). Critically, whether to increase sampling within a taxonomic group of interest should not depend on whether there is statistical support for a topology, as strong support does not indicate a lack of systematic error.

2.1   Information theory and taxon addition

Several techniques have been developed from information theory to examine where taxa could be added to an analysis to increase the precision of a topological estimate (Goldman, 1998; Massingham & Goldman, 2000; Geuten et al., 2007). The observed information matrix, which is a measure of the sharpness of the curve about the likelihood function, is compared to the expected (or Fisher) information matrix when a branch is added to different parts of the
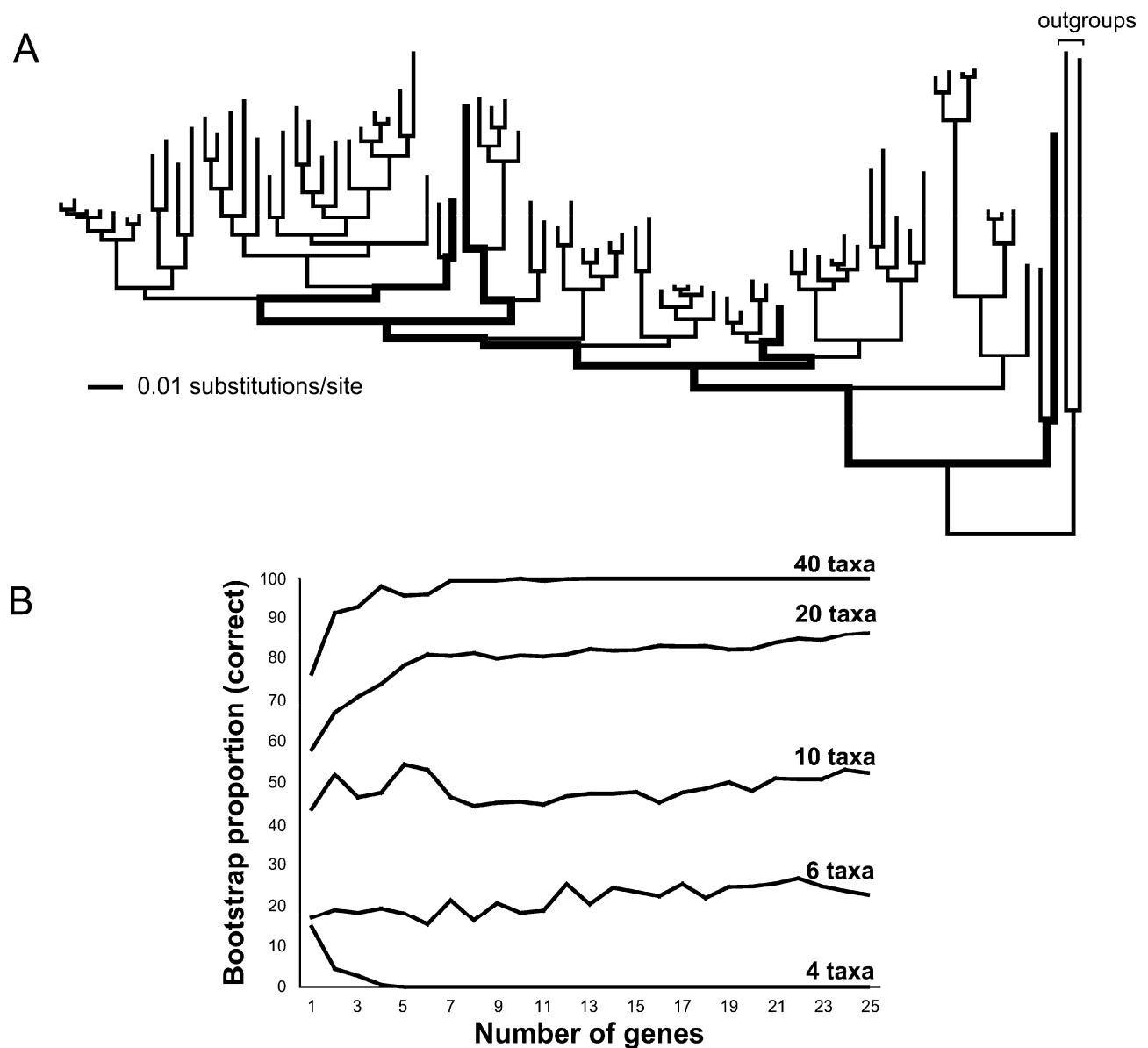
Fig. 3. A, Well-sampled yeast tree used for simulating data. Dark lines represent four taxa known to be susceptible to long-branch attraction under parsimony. B, Results of randomly sampling 1–25 genes for 4–40 taxa. As the number of genes in a parsimony analysis increases, the bootstrap for the correct reconstruction of the four-taxon statement decreases, unless taxon sampling is sufficient to break the long branches. Adapted from Hedtke et al. (2006; figs. 1 and 3).

topology (Geuten et al., 2007). In essence, this comparison indicates the increase in precision of the topology estimate that is gained by adding a taxon along a particular branch of the tree. Based on a data set from angiosperms, Geuten et al. (2007) found that information theory generally supports adding additional taxa close to where the long branch attaches to the rest of the tree, in congruence with other studies based on simulation (Graybeal, 1998; Poe & Swofford, 1999). Unfortunately, these techniques, while promising, have not been rigorously tested, and generally assume that one has strong *a priori* expectations about where additional taxa might fall on the phylogeny.

2.2 Detecting LBA

Most studies that explore specific recommendations for increased taxon sampling focus on detecting and correcting LBA. LBA is often cited when a topology fails to meet *a priori* expectations, but caution must be applied before attributing unexpected

results to method inconsistency (Anderson & Swofford, 2004) or method bias. Many claims of LBA in published studies are *a posteriori* explanations of unexpected results. Often, when taxa are added to an analysis that originally generated an unexpected tree, topological relationships change to match expectations based on traditional taxonomy, and this change is attributed to LBA (e.g. Lin et al., 2002; Stefanovic et al., 2004; Philippe et al., 2005; Baurain et al., 2007). Here we discuss strategies which could provide useful heuristics for exploring whether method inconsistency or bias, particularly LBA, is affecting the analysis.

Disagreement among independent data sets—particularly those based on traditional morphological taxonomy versus molecular analyses—may be the first signal that LBA may be present in the data set (Lin et al., 2002; Chen et al., 2003). If two long branches are unexpectedly drawn together in one or the other analysis, LBA may be the culprit. However, a researcher must also consider the alternative that traditional taxonomy may be wrong (e.g., Ammerman & Hillis, 1992; Van Den Bussche et al., 1998). The two long branches may in fact be sister taxa, or may be brought together by LBA. To distinguish these hypotheses, one technique is to prune each long-branched taxon successively from the analysis, and observe whether the topology changes (Bergsten, 2005). Presumably, if the position of a long-branched taxon changes dependent upon the inclusion of other long-branched taxa, LBA may be implicated. Unfortunately, this test is not definitive, because other characteristics of the excluded taxa could be affecting the topology as a whole.

Simulated data generated under different hypotheses could be analyzed to compare possible topologies and get a sense of potential error rates in the analysis (Van Den Bussche et al., 1998; Sanderson et al., 2000). In this parametric bootstrapping approach, data are simulated under each hypothesis using model parameters estimated from the biological sequence data. For example, one could identify the alternative hypothesis by running a phylogenetic analysis constrained to find the best tree which does *not* place two suspect taxa as sister groups. If analyses of replicate data sets simulated on this tree tend to bring the two suspect taxa together more often than expected by chance, then one cannot reject the possibility that the two taxa are grouped due to systematic error or bias. However, as with all parametric approaches, investigators should consider the adequacy of the simulation model. Under-parameterized models may underestimate the potential for LBA.

2.3   Outgroup sampling

Outgroup taxa tend to be on long branches, either because of the processes of cladogenesis, because of extinction events between the outgroup and the ingroup, or because of inappropriate selection of outgroup taxa. When the outgroup is distantly related to the ingroup, long branches in the ingroup can be drawn toward the base of the tree by LBA, affecting ingroup relationships (Hillis, 1998; Rannala et al., 1998) or misplacing the root of the tree (Holland et al., 2003). Holland et al. (2003) used simulated data to demonstrate this effect not just for parsimony, but for maximum likelihood and distance analyses as well. Graybeal (1998) suggested that error could be reduced through the use of multiple outgroup taxa separated by short internal branches.

One method to examine whether outgroup choice is affecting the topology of the ingroup is to run the analysis using the ingroup taxa only (Holland et al., 2003; Bergsten, 2005). If the ingroup topology is influenced by the long branch leading to the outgroup, the unrooted topology with and without the outgroup will change. This technique will not allow the researcher to distinguish whether the rooting of the tree (placement of the outgroup) is correct; only whether the outgroup is influencing the ingroup topology. A second method to evaluate outgroup choice is to simulate a number of random sequences which are each used to root the tree (Sullivan & Swofford, 1997). This can indicate the relative probabilities for rooting positions when no historical signal is present in the data for the outgroup. In both cases, if use of a particular outgroup leads to LBA problems, sampling more outgroup taxa may assist in detecting homoplasy and reducing the effects of the long outgroup branch.

2.4   Ingroup sampling

If LBA is detected within the ingroup, long-branch subdivision by addition of taxa could mitigate this effect (Fig. 4; Hendy & Penny, 1989; Hillis, 1996). The strategy of long-branch subdivision has been examined using both simulated (e.g., Graybeal, 1998; Poe & Swofford, 1999; Poe, 2003) and biological (e.g., Poe, 1998a; Baurain et al., 2007) data. In parsimony analyses, there are cases in which the addition of one taxon can actually cause LBA; this has been demonstrated using simulated data in the four-taxon case when the new taxon introduces a new long branch (Rannala et al., 1998; Poe & Swofford, 1999), or when there are three long branches, and breaking one causes the remaining two to become drawn together (Fig. 4; Poe & Swofford, 1999; Poe, 2003). In this case the new taxon introduces an
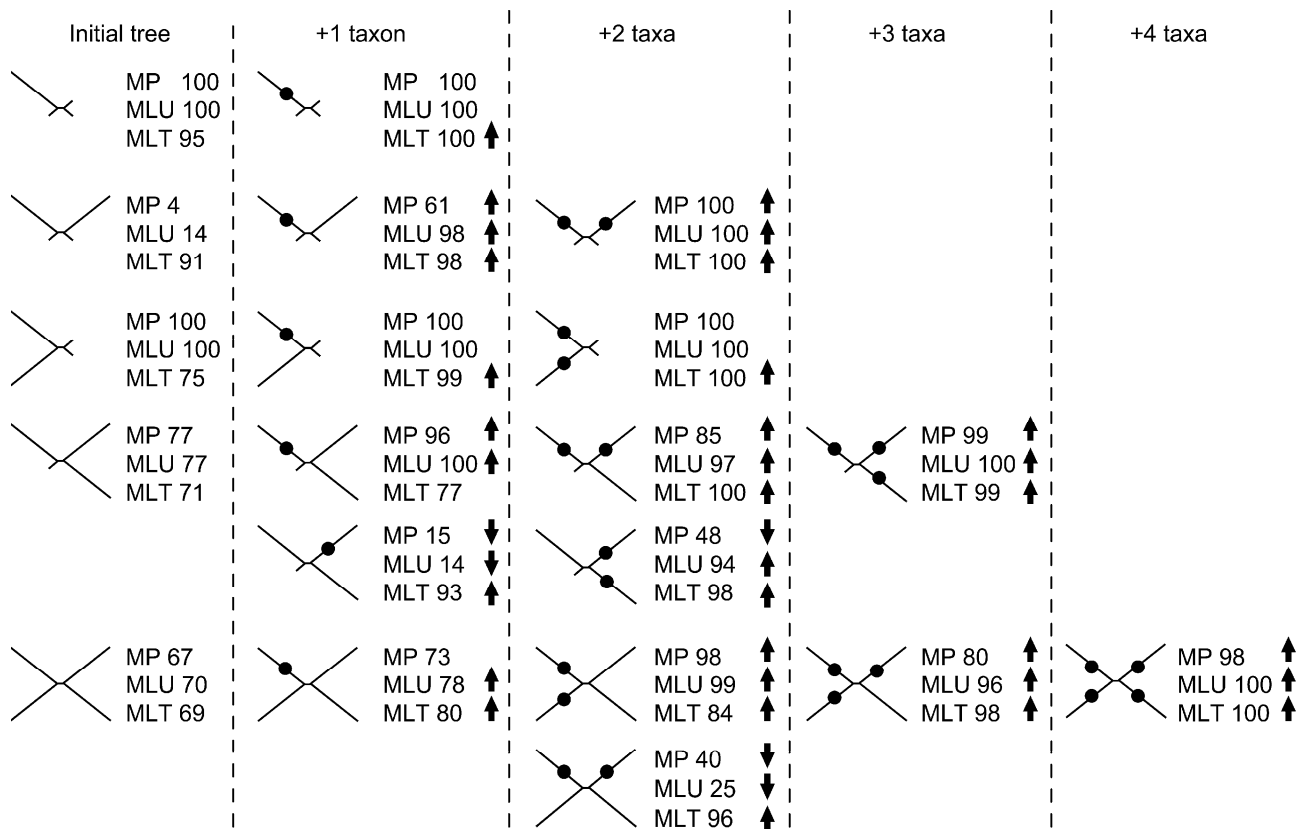
Fig. 4. Effects of long-branch subdivision on accuracy in four-taxon trees, based on simulated data. For each tree, branch lengths of long branches are 0.5, of short terminal branches 0.1, and of internal branches 0.05. Dots indicate when a long branch was broken at length 0.1 from the internal node. Poe (2003) evaluated 6 different reconstruction methods, 3 are summarized in this figure: unweighted maximum parsimony (*MP*), maximum likelihood with an under-parameterized model of substitution (*MLU*), and maximum likelihood with the true model of substitution (*MLT*). The numbers are the percentage of time the true, four-taxon tree was recovered in 100 replicate simulations. The arrows indicate increased (↟) or decreased (↡) accuracy as a result of the added taxa, and no arrow indicates that the accuracy was unaffected. Adapted from Poe (2003; fig. 3).

asymmetrical pattern of homoplasy and long-branch attraction results in an incorrect reconstruction. In addition, adding only one taxon may not be sufficient to alleviate LBA if the branch is sufficiently long (Poe, 2003; Hedtke et al., 2006) or if the added taxon is placed towards the tip (Graybeal, 1998). However, both these effects are diminished if enough taxa are added along long branches (Poe, 1998a; Anderson & Swofford, 2004; Hedtke et al., 2006). For example, in the Hedtke et al. (2006) simulation study, adding taxa increased accuracy for a particular bipartition only when the added taxa divided a long branch that was causing long-branch attraction.

2.5 Adding taxa with missing data

Adding taxa with incomplete character information to a supermatrix has primarily been evaluated in parsimony-based morphological analyses in reference to fossil data. Incomplete fossil data are not always beneficial, as they may reduce support for some nodes

(Wiens, 2003; Wiens, 2005; Cobbett et al., 2007). However, it appears that well-chosen fossil data can be helpful in breaking long branches, even if incomplete, as these morphological characters can be informative about character reconstruction at branching nodes (Donoghue et al., 1989; Huelsenbeck, 1991). The effect of missing character data on sequence analysis is still being debated, with some researchers arguing that adding taxonomic data is beneficial even if the resulting supermatrix has a large proportion of missing data (e.g., 25%: Philippe et al., 2004; 75%: Wiens & Reeder, 1995; 95%: McMahon & Sanderson, 2006). However, Lemmon et al. (in press) used simulated data to demonstrate that missing sequence data can positively mislead model-based methods. This depends in part on the relative rates of evolution for sites with and without missing data, and the topological position of those taxa with missing data. This is an area of active research, and caution should be used

when combining taxa with missing data until this issue has been more completely explored using both simulated and biological data sets.

2.6   Alternatives to increased taxon sampling

Several other techniques of combating method inconsistency have been suggested. Because inconsistency or bias results from violation of model assumptions (e.g., not accurately modeling multiple substitutions), finding a better-fitting model could solve the problem (Olsen, 1987; Whelan & Goldman, 2001; Lartillot & Philippe, 2004; Anderson & Swofford, 2004; Delsuc et al., 2005; Baurain et al., 2007; Lartillot et al., 2007). To reduce inconsistency due to LBA, it has been suggested that either eliminating fast-evolving sites from the analysis (Delsuc et al., 2002; Delsuc et al., 2005; Rodriguez-Ezpeleta et al., 2007) or coding all the data to represent only less frequent transversions ("RY" coding: Phillips et al., 2004; Delsuc et al., 2005) would reduce saturation and compositional bias in the data set, and thus reduce LBA. Using amino acid data (with models which take into account site heterogeneity, e.g., Lartillot & Philippe, 2004) may be another alternative to using raw sequence data. However, the actual effect of saturated sequences on phylogenetic analyses has been incompletely explored. Hillis (1998) simulated sequence data on a tree with such long branch lengths that the sequences would not be recognizable as homologous, but phylogenetic methods were still able to reconstruct the correct tree when taxon sampling was sufficient. Exclusion of data (whether by excluding sites or reducing their information) may be useful in eliminating the problem of a particular set of long branches, but this may be at the expense of resolution in other regions of the tree.

3   Taxon sampling affects parameter estimation

Many advances in phylogenetic analysis over the past two decades have involved model-based approaches, such as maximum likelihood and Bayesian analyses (Swofford et al., 1996; Ronquist & Huelsenbeck, 2003; Felsenstein, 2004). In general, these parametric methods outperform nonparametric methods in both simulations and experimental studies (Hillis et al., 1994a; Huelsenbeck, 1995; Cunningham et al., 1997). However, accurate phylogenetic results from model-based studies depend, at least in part, on reasonably accurate parameter estimates for the models of evolution (Goldman, 1993; Hillis et al.,

1994b; Cunningham et al., 1998; Lemmon & Moriarty, 2004; Brown & Lemmon, 2007). One of the reasons that increased taxon sampling results in more accurate phylogenetic estimation for these model-based methods is that sampling additional taxa also improves parameter estimation (Pollock et al., 1999; Sullivan et al., 1999; Pollock & Bruno, 2000; Pollock et al., 2002). In addition, as branch lengths are shortened, there are fewer unobserved changes that need to be inferred, so the accuracy of the inference becomes less dependent on the model of evolution.

In addition to their effect on phylogenetic analyses, the parameters of evolutionary models are themselves of interest to biologists. These parameters are often gene-specific, so collecting genomic-scale data from many genes across only a few taxa does little to improve our estimates of the details of evolutionary models. Instead, a thorough taxon-sampling approach is needed for each gene. Of course, the evolutionary processes may not be static across the Tree of Life for any given gene, so models that account for non-stationarity in these processes can provide better descriptions of evolutionary history (Yang & Roberts, 1995; Galtier & Gouy, 1998; Foster, 2004; Blanquart & Lartillot, 2006; Boussau & Gouy, 2006; Gowri-Shankar & Rattray, 2007; Blanquart & Lartillot, 2008; Kolaczkowski & Thornton, 2008). These models relax the assumption of time-homogeneity and can be used to detect signatures of complex evolutionary processes, such as compositional heterogeneity or heterotachy (branch-length heterogeneity), that are known to exist in biological data (Lockhart et al., 1992; Foster et al., 1997; Mooers & Holmes, 2000; Lopez et al., 2002; Jermiin et al., 2004; Ane et al., 2005). Non-stationary models can greatly increase the need for even more thorough taxon sampling, because the model parameters may need to be estimated multiple times across the tree, rather than once for all taxa. It is important to note, however, that under non-stationary models, the number of parameters can increase as more sequences are added, thus increasing the computational difficulty of phylogenetic reconstruction from large data sets. Nonetheless, this obstacle may be mitigated by the use of carefully constructed priors in a Bayesian MCMC framework (Yang, 2006) and with the development of computational methods for calculating likelihoods from non-reversible models (Boussau & Gouy, 2006).

Some of the parameters that have been shown to be important for phylogenetic estimation include site-specific rates of evolutionary change; rates of change across first, second, and third positions of

codons; rates of change relative to changes in functional groups of amino-acid residues; relative rates of the various classes of transitions and transversions between nucleotide states; branch-specific rates of evolutionary change; and taxon-specific differences in base composition (Olsen, 1987; Steel et al., 1993; Hasegawa & Hashimoto, 1993; Hillis et al., 1993; Leipe et al., 1993; Goldman & Yang, 1994; Steel, 1994; Swofford et al., 1996). The number of taxa that are needed to effectively estimate these parameters differ greatly across the parameters, but all of the estimates are improved by more thorough taxon sampling. For instance, Pollock and Bruno (2000) noted significant improvement in parameter estimation (and in turn, phylogenetic estimation) as their taxon samples increased from 4 to 8 to 16 to 24 taxa. They concluded that both phylogenetic reconstruction and estimation of unknown evolutionary processes show greater improvement through increasing taxon sampling than by increasing sequence length. In some cases, reasonable parameter estimates may be obtained from external data sources, such as the HIV database, and then applied to a more limited set of taxa in the phylogenetic analysis (Hillis, 1999). However, for most taxa, the appropriate comparative data must be obtained by the investigator for a specific group of species under study.

4 Dense taxon sampling improves inferences of evolutionary processes

Beyond simply broadening our understanding of species relationships, phylogenetic trees are essential tools used in many areas of biology. Phylogenies are often used to explain broad evolutionary patterns and processes such as the evolution of adaptive traits, ancestral character states, the timing of species divergences, and variation in evolutionary rates. Many of the applications developed for these types of analyses require robust and accurate estimates of phylogeny (topology, branch lengths, and root position). This is an important consideration in and of itself; however, post-tree reconstruction applications are sensitive to reduced levels of data sampling, even when provided with an accurate phylogenetic tree.

4.1 Comparative methods

Comparative analyses are a fundamental component in the fields of evolutionary biology, behavior, and ecology. The development of statistical methods that incorporate phylogenetic trees (Felsenstein, 1985) have allowed for robust and reliable tests of the

evolution of adaptive traits and the processes that might drive diversification. For example, these methods have been used to reveal patterns in the biodiversity of marine teleost fishes (Alfaro et al., 2007) and to show that independent origins of dietary specialization have been a major factor in the evolution of defensive mechanisms in neotropical poison frogs (Darst et al., 2005). Comparative analyses of character evolution using phylogenetic comparative methods require attention to adequate sampling at many levels. At the intraspecific level, poor sampling of organismal attributes can lead to measurement error, which may result in an underestimation of the variance of contrasts between sister taxa (Ricklefs & Starck, 1996). Generation of a robust phylogeny is extremely important since different comparative methods have different ways of dealing with topological uncertainty (Purvis et al., 1994). In addition, fewer taxa (and thus fewer internal nodes for calculating contrasts) can lead to increased variance and uncertainty in the results. Ackerly (2000) used simulated data to show that the statistical power of several comparative tests decreased as the sample size of taxa decreased, and that careful attention should be paid to how species are sampled for these analyses. Biased taxon sampling, particularly with respect to the characters of interest, can lead to systematic biases in the calculation of statistical correlations between characters. The results presented by Ackerly (2000) indicate that uniform, random sampling of taxa does not introduce error in phylogenetic comparative methods.

4.2 Ancestral character states

An integral component of phylogenetic comparative analyses and other evolutionary applications is the reconstruction of ancestral character states. These methods use phylogenetic trees and branch lengths to infer the states of discrete or continuous characters at ancestral nodes, and have been used to reconstruct such diverse ancestral characters as the advertisement calls of frogs in the genus *Physalaemus* (Ryan & Rand, 1995, 1998), the fruiting-body forms of homobasidiomycetes (Hibbett, 2004), and ancient bacterial protein sequences (Gaucher et al., 2003). Dense taxon sampling is also an important consideration for ancestral-state reconstruction methods. Salisbury and Kim's (2001) analyses of simulated data and trees indicated that the accuracy of parsimony ancestral-state estimation decreases with reduced taxon sampling and increased rates of character evolution (Fig. 5). Because parsimony methods do not account for unobserved changes, they usually underestimate the number of changes along a branch (Fitch & Bruschi, 1987;
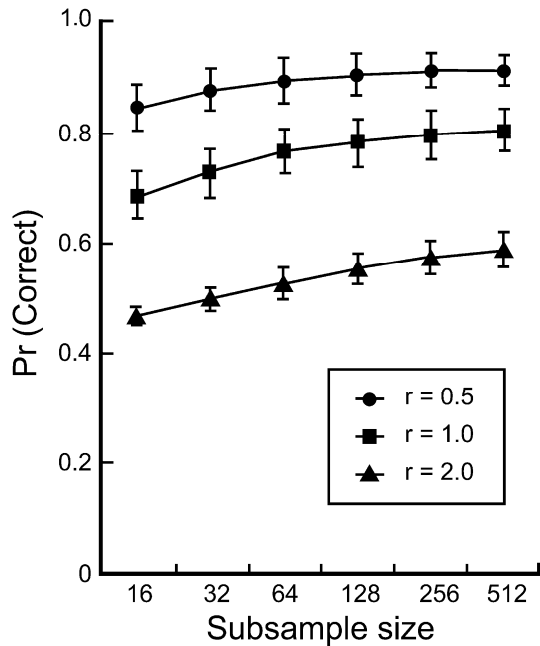
Fig. 5. The mean probabilities, *Pr (Correct),* of correctly estimating the root state of a binary character evolving at 3 different rates (*r*) on subsamples of 512-taxon, pure-birth model tree topologies. Each point is the mean for a sample of 100 trees and the error bars represent the ± 1 standard deviation. Adapted from Salisbury and Kim (2001; fig. 1).

Fitch & Beintema, 1990; Huelsenbeck & Lander, 2003). Dense taxon sampling can reduce this effect and improve the accuracy of parsimony ancestral-state estimates. Maximum likelihood and Bayesian methods for reconstructing ancestral states have also been developed (Pagel, 1994; Schluter et al., 1997; Pagel, 1999; Huelsenbeck & Bollback, 2001; Pagel et al., 2004). These parametric ancestral-state reconstruction methods are also sensitive to high rates of character evolution. However, Schluter et al. (1997) showed that parsimony ancestral-state reconstruction methods often fail to identify ambiguous-node state estimates. Conversely, maximum likelihood and Bayesian methods are less likely to provide misleading results because these methods incorporate branch-length information and explicit models of character evolution and quantify uncertainty in ancestral-state estimates (provided that the model assumptions are adequate). Bayesian approaches, in particular, use Markov chain Monte Carlo sampling to accommodate and quantify uncertainty in the tree topology, branch lengths, ancestral states, and model parameters (Huelsenbeck & Bollback, 2001; Pagel et al., 2004). Denser taxon sampling reduces the number of unobserved evolu-

tionary events, and so is also expected to simplify and improve the reconstruction of ancestral states in model-based analyses.

4.3 Divergence time estimation

A primary field of research in evolutionary biology involves estimation of the timing and rate of evolutionary processes. In these applications, phylogenetic trees are used to date speciation events and infer lineage-specific substitution rates. Reliable estimates of species divergence times are fundamental components for understanding historical biogeography, testing hypotheses of adaptive character evolution, and estimating speciation and extinction rates. However, divergence time estimation is hindered by the fact that the rate of evolution and time are intrinsically linked when inferring genetic distances between lineages. Several methods test for variation in the rates of molecular evolution or tease apart the rate of substitution and time by applying models for estimating lineage-specific substitution rates. These methods include strict molecular clock models (Zuckerkandl & Pauling, 1962; Langley & Fitch, 1974), local molecular clocks (Kishino & Hasegawa, 1990; Rambaut & Bromham, 1998; Yoder & Yang, 2000; Yang & Yoder, 2003), non-parametric and semi-parametric methods for estimating autocorrelated substitution rates (Sanderson, 1997, 2002), and Bayesian methods for estimating autocorrelated and uncorrelated rates (Thorne et al., 1998; Huelsenbeck et al., 2000; Kishino et al., 2001; Thorne & Kishino, 2002; Drummond et al., 2006; Lepage et al., 2006). These various approaches have been applied to a number of biological data sets (e.g., Yang & Yoder, 2003; Smith et al., 2006; Bell, 2007; Hugall et al., 2007; Roelants et al., 2007; Zhou & Holmes, 2007). Current implementations of most of these methods require a fixed tree topology and sometimes fixed branch lengths (Thorne & Kishino, 2002; Sanderson, 2003; Lepage et al., 2007; for exceptions see Drummond et al., 2006). Because of their reliance on phylogenetic data, these methods can be sensitive to taxon sampling density. Robinson et al. (1998) evaluated the effect of reduced taxon sampling on the performance of the relative-rates test. The relative-rates test (Sarich & Wilson, 1973; Wu & Li, 1985) is used to compare the substitution rates between two species and has been extended for analyzing larger phylogenetic trees to detect rate variation (Li & Bousquet, 1992; Takezaki et al., 1995). The simulation study of Robinson et al. (1998) showed that increased proportions of taxon sampling improved the accuracy of the relative-rates test.

Most of the work exploring the accuracy of molecular dating methods has revealed that these methods are very sensitive to the fossil calibrations used and little is known about the impact of taxon sampling on divergence time estimates (Yang & Rannala, 2006; Rutschmann et al., 2007; Hugall et al., 2007). A recent study by Hug and Roger (2007) used two biological data sets with low levels of taxon sampling (30 metazoan taxa with two outgroup species, and a 36 taxon data set that spanned all eukaryotes) and concluded that, for these data sets, reduced taxon sampling was not an important factor in the estimation of node times. However, their analyses showed that the choice and application of fossil calibration points resulted in a significant impact on the estimates of node ages. From their results, Hug and Roger (2007) recommended that biologists should focus on improving the number and quality of their fossil calibrations and not on increasing taxon sampling, provided there are enough taxa to obtain a reliable estimate of phylogeny. However, because of the sparsely sampled data sets used in this study and the demonstrated extreme sensitivity of these data to fossil constraints, Hug and Roger's (2007) results may not apply to a more general set of conditions and the importance of dense taxon sampling for estimating species divergence times is still an open question.

Node-density effects, as a result of uneven taxon sampling, may adversely affect molecular dating analyses (Hugall & Lee, 2007). Based on the studies demonstrating the sensitivity of divergence time estimation methods to fossil calibration choice (Near & Sanderson, 2004; Near et al., 2005; Roger & Hug, 2006; Yang & Rannala, 2006; Ho, 2007; Hugall et al., 2007; Rutschmann et al., 2007), together with studies emphasizing the importance of increased taxon sampling on phylogenetic reconstruction methods and the estimation of evolutionary parameters (Lecointre et al., 1993; Hillis, 1996, 1998; Graybeal, 1998; Rannala et al., 1998; Pollock & Bruno, 2000; Zwickl & Hillis, 2002; Pollock et al., 2002; Poe, 2003; DeBry, 2005; Hedtke et al., 2006), we recommend increased collection of fossils and improved taxon sampling density for these types of analyses, whenever possible. Maximizing the number of fossil calibration points goes hand-in-hand with increasing taxon sampling because densely sampled trees provide a greater number of internal nodes on which an investigator can place a fossil calibration. Moreover, investigators are far more restricted by the availability of fossils and other types of information for calibrating divergences than by the availability of extant taxa. Further investigation using simulations and well-sampled data sets of living and fossil taxa should help shed light on this issue. Because extensive taxon sampling (especially of fossil taxa) is sometimes impractical, Bayesian methods for divergence time estimation present promising opportunities to account for uncertainty in phylogenies by simultaneously estimating the tree topology and branching times (Drummond et al., 2006). These methods can also incorporate information on taxon sampling density in the form of priors on the distribution of divergence times (Yang & Rannala, 1997, 2006).

4.4 Evaluating diversification rates

Phylogenetic trees are fundamental for understanding variation in species diversity. Methods for elucidating patterns of speciation and extinction measure the shape of phylogenies to detect shifts in diversification rates or to estimate global net diversification rates. Phylogenetic tree shape can be measured by quantifying how node ages are distributed over time or by calculating the degree of asymmetry among lineages in the tree. Measures of tree shape can be compared to a null model that assumes all lineages have experienced the same rate of diversification (Shao & Sokal, 1990; Kirkpatrick & Slatkin, 1993; Nee et al., 1994b; Pybus & Harvey, 2000; Agapow & Purvis, 2002). Analyses of the temporal distribution of diversification events use branch lengths obtained from time-adjusted phylogenies to estimate and detect large shifts in speciation and extinction rates (Nee et al., 1994b; Pybus & Harvey, 2000). For example, Becerra (2005) applied these methods to investigate temporal and biogeographic processes that may have shaped the diversity of the plant genus *Bursera*. The results of this study indicate that the radiation of this group is associated with the establishment of tropical dry forest habitat in Mexico. However, inadequate taxon sampling has a significant impact on these methods (Nee et al., 1994a). Nee et al. (1994a) used lineages-by-time plots to show that incomplete taxon sampling can result in an apparent reduction in the rate of diversification over time, even when the tree evolved under constant rates of speciation and extinction.

Analyses based on topology measure asymmetry in the distribution of lineages over a tree to test for changes in diversification rates. These methods evaluate the balance either at a single node or over the entire tree (Shao & Sokal, 1990; Kirkpatrick & Slatkin, 1993; Agapow & Purvis, 2002) and are often used to detect patterns characteristic of rapid radiations in phylogenetic trees (Guyer & Slowinski, 1993; Chan &

Moore, 1999, 2002). The degree of taxon sampling is an important consideration when conducting these analyses. Several studies have shown that published phylogenies are (on average) much more imbalanced than expected under a model assuming constant diversification rates (Guyer & Slowinski, 1991; Heard, 1992; Mooers, 1995; Purvis & Agapow, 2002; Holman, 2005; Blum & François, 2006; Heath et al., 2008). Mooers (1995) compared the level of tree imbalance in a collection of published phylogenies and found that incomplete trees are more imbalanced than completely sampled phylogenies. Another study by Heath et al. (2008) examined the effect of random taxon sampling on empirical trees and on phylogenies simulated under different models of cladogenesis. They found that reduced taxon sampling of empirical trees and trees simulated under variable and autocorrelated speciation and extinction rates causes an increase in node imbalance. These results suggest that poor taxon sampling leads to an increase in the apparent rate variation because of the overrepresentation of older nodes. The bias caused by incomplete species sampling must be considered when using phylogenies to test hypotheses about species diversity.

5 Innovations in reconstruction algorithms and the analysis of large data sets

Until recently, computational constraints on phylogenetic analyses made inclusion of large numbers of taxa impractical for many biologists. However, developments in computational power, parallel computation, and phylogenetic algorithms have greatly decreased computational constraints for phylogenetic analyses of many taxa, even for the most computationally demanding parametric approaches (e.g., Brauer at al., 2002; Lemmon & Milinkovitch, 2002; Guindon & Gascuel, 2003; Ronquist & Huelsenbeck, 2003; Stamatakis, 2006; Minh et al., 2005; Zwickl, 2006). Quick but imprecise clustering techniques, such as the widely used neighbor-joining algorithm (Saitou & Nei, 1987), are rapidly being replaced by methods that more thoroughly explore solution space using a clearly defined model of evolution (implemented in such programs as RAXML, GARLI, PhyML, IQPNNI, MRBAYES, and PAUP*; see http://evolution.genetics.washington.edu/phylip/software.html for more information). Analyses of hundreds to thousands of taxa have become routine for parsimony, maximum likelihood, and Bayesian approaches, and analyses of tens-of-thousands of taxa are now feasible, even for parametric methods. Therefore, computa-

tional constraints can no longer be viewed as a serious impediment to thorough taxon sampling. Instead, the limitations of taxon sampling have now shifted to problems of taxon availability and the constraints of specimen and data collection. Given the many benefits of thorough taxon sampling summarized in this paper, we advise biologists to carefully consider taxon sampling design in planning, conducting, and interpreting phylogenetic analyses. In many cases, increasing taxon sampling is one of the most practical and beneficial approaches to increasing the accuracy of phylogenetic estimates and the biological inferences that are derived from phylogenetic trees.

References

Ackerly DD. 2000. Taxon sampling, correlated evolution, and independent contrasts. Evolution 54: 1480–1492.

Agapow PM, Purvis A. 2002. Power of eight tree shape statistics to detect nonrandom diversification: a comparison by simulation of two models of cladogenesis. Systematic Biology 51: 866–872.

Albrecht C, Kuhn K, Streit B. 2007. A molecular phylogeny of Planorboidea (Gastropoda, Pulmonata): insights from enhanced taxon sampling. Zoologica Scripta 36: 27–39.

Alfaro ME, Santini F, Brock CD. 2007. Do reefs drive diversification in marine teleosts? Evidence from the pufferfish and their allies (order Tetraodontiformes). Evolution 61: 2104–2126.

Ammerman LK, Hillis DM. 1992. A molecular test of bat relationships: Monophyly or diphyly? Systematic Biology 41: 222–232.

Anderson FE, Swofford DL. 2004. Should we be worried about long-branch attraction in real data sets? Investigations using metazoan 18S rDNA. Molecular Phylogenetics and

Evolution 33: 440–451.

Ane C, Burleigh JG, McMahon MM, Sanderson MJ. 2005. Covarion structure in plastid genome evolution: A new statistical test. Molecular Biology and Evolution 22: 914–924.

Ane C, Sanderson MJ. 2005. Missing the forest for the trees: Phylogenetic compression and its implications for inferring complex evolutionary histories. Systematic Biology 54: 146–157.

Baurain D, Brinkmann H, Philippe H. 2007. Lack of resolution in the animal phylogeny: Closely spaced cladogeneses or undetected systematic errors? Molecular Biology and Evolution 24: 6–9.

Becerra JX. 2005. Timing the origin and expansion of the Mexican tropical dry forest. Proceedings of the National Academy of Sciences USA 102: 10919–10923.

Bell CD. 2007. Phylogenetic placement and biogeography of the North American species of Valerianella (Valerianaceae: Dipsacales) based on chloroplast and nuclear DNA. Molecular Phylogenetics and Evolution 44: 929–941.

Bergsten J. 2005. A review of long-branch attraction. Cladistics 21: 163–193.

Blanquart S, Lartillot N. 2006. A Bayesian compound stochastic process for modeling nonstationary and nonhomogeneous sequence evolution. Molecular Biology and Evolution 23: 2058–2071.

Blanquart S, Lartillot N. 2008. A site- and time-heterogeneous model of amino acid replacement. Molecular Biology and Evolution 25: 842–858.

Blum MGB, Françios O. 2006. Which random processes describe the Tree of Life? A large-scale study of phylogenetic tree imbalance. Systematic Biology 55: 685–691.

Boussau B, Gouy M. 2006. Efficient likelihood computations with nonreversible models of evolution. Systematic Biology 55: 756–768.

Brauer MJ, Holder MT, Dries LA, Zwickl DJ, Lewis PO, Hillis DM. 2002. Genetic algorithms and parallel processing in maximum-likelihood phylogeny inference. Molecular Biology and Evolution 19: 1717–1726.

Braun EL, Kimball RT. 2002. Examining basal avian divergences with mitochondrial sequences: Model complexity, taxon sampling, and sequence length. Systematic Biology 51: 614–625.

Bremer B, Jansen RK, Oxelman B, Backlund M, Lantz H, Kim KJ. 1999. More characters or more taxa for a robust phylogeny—case study from the coffee family (Rubiaceae). Systematic Biology 48: 413–435.

Brown JM, Lemmon AR. 2007. The importance of data partitioning and the utility of bayes factors in Bayesian phylogenetics. Systematic Biology 56: 643–655.

Bruno WJ, Halpern AL. 1999. Topological bias and inconsistency of maximum likelihood using wrong models. Molecular Biology and Evolution 16: 564–566.

Bush RM, Fitch WM, Bender CA, Cox NJ. 1999. Positive selection on the H3 hemagglutinin gene of human influenza virus A. Molecular Biology and Evolution 16: 1457–1465.

Chan KMA, Moore BR. 1999. Accounting for mode of speciation increases power and realism of tests of phylogenetic asymmetry. American Naturalist 153: 332–346.

Chan KMA, Moore BR. 2002. Whole-tree methods for detecting differential diversification rates. Systematic Biology 51: 855–865.

Chen WJ, Bonillo C, Lecointre G. 2003. Repeatability of clades as a criterion of reliability: A case study for molecular phylogeny of Acanthomorpha (Teleostei) with larger number of taxa. Molecular Phylogenetics and Evolution 26: 262–288.

Cobbett A, Wilkinson M, Wills MA. 2007. Fossils impact as hard as living taxa in parsimony analyses of morphology. Systematic Biology 56: 753–766.

Crandall KA, Bininda-Emonds ORP, Mace GM, Wayne RK. 2000. Considering evolutionary processes in conservation biology. Trends in Ecology and Evolution 15: 290–295.

Cummings MP, Meyer A. 2005. Magic bullets and golden rules: Data sampling in molecular phylogenetics. Zoology 108: 329–336.

Cunningham CW, Jeng K, Husti J, Badgett M, Molineux IJ, Hillis DM, Bull JJ. 1997. Parallel molecular evolution of deletions and nonsense mutations in bacteriophage T7. Molecular Biology and Evolution 14: 113–116.

Cunningham CW, Zhu H, Hillis DM. 1998. Best-fit maximum-likelihood models for phylogenetic inference: Empirical tests with known phylogenies. Evolution 52: 978–987.

Darst CR, Menendez-Guerrero PA, Coloma LA, Cannatella DC. 2005. Evolution of dietary specialization and chemical defense in poison frogs (Dendrobatidae): A comparative analysis. American Naturalist 165: 56–69.

DeBry RW. 1992. The consistency of several phylogeny-inference methods under varying evolutionary rates. Molecular Biology and Evolution 9: 537–551.

DeBry RW. 2005. The systematic component of phylogenetic error as a function of taxonomic sampling under parsimony. Systematic Biology 54: 432–440.

Delsuc F, Brinkmann H, Philippe H. 2005. Phylogenomics and the reconstruction of the Tree of Life. Nature Reviews Genetics 6: 361–375.

Delsuc F, Phillips MJ, Penny D. 2003. Comment on "Hexapod origins: Monophyletic or paraphyletic?" Science 301: 1482.

Delsuc F, Scally M, Madsen O, Stanhope MJ, de Jong WW, Catzeflis FM, Springer MS, Douzery EJP. 2002. Molecular phylogeny of living xenarthrans and the impact of character and taxon sampling on the placental tree rooting. Molecular Biology and Evolution 19: 1656–1671.

Donoghue MJ, Doyle JA, Gauthier J, Kluge AG, Rowe T. 1989. The importance of fossils in phylogeny reconstruction. Annual Review of Ecology and Systematics 20: 431–460.

Drummond AJ, Ho SY, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biology 4: e88.

Eisen JA. 1998. Phylogenomics: Improving functional predictions for uncharacterized genes by evolutionary analysis. Genome Research 8: 163–167.

Felsenstein J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Systematic Zoology 27: 401–410.

Felsenstein J. 1985. Phylogenies and the comparative method. American Naturalist 125: 1–15.

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

Fitch WM, Beintema JJ. 1990. Correcting parsimonious trees for unseen nucleotide substitutions—the effect of dense branching as exemplified by ribonuclease. Molecular Biology and Evolution 7: 438–443.

Fitch WM, Bruschi M. 1987. The evolution of prokaryotic ferredoxins—with a general method correcting for unobserved substitutions in less branched lineages. Molecular Biology and Evolution 4: 381–394.

Foster PG. 2004. Modeling compositional heterogeneity. Systematic Biology 53: 485–495.

Foster PG, Jermiin LS, Hickey DA. 1997. Nucleotide composition bias affects amino acid content in proteins coded by animal mitochondria. Journal of Molecular Evolution 44: 282–288.

Freudenstein JV, Pickett KM, Simmons MP, Wenzel JW. 2003. From basepairs to birdsongs: phylogenetic data in the age of genomics. Cladistics 19: 333–347.

Galtier N, Gouy M. 1998. Inferring pattern and process: maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. Molecular Biology and Evolution 15: 871–879.

Gascuel O, Bryant D, Denis F. 2001. Strengths and limitations of the minimum evolution principle. Systematic Biology 50: 621–627.

Gatesy J, DeSalle R, Wahlberg N. 2007. How many genes should a systematist sample? Conflicting insights from a phylogenomic matrix characterized by replicated incongruence. Systematic Biology 56: 355–363.

Gaucher EA, Thomson JM, Burgan MF, Benner SA. 2003. Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins. Nature 425: 285–288.

Geuten K, Massingham T, Darius P, Smets E, Goldman N. 2007. Experimental design criteria in phylogenetics: Where to add taxa. Systematic Biology 56: 609–622.

Gojobori T, Ishii K, Nei M. 1982. Estimation of average number of nucleotide substitutions when the rate of substitution varies with nucleotide. Journal of Molecular Evolution 18: 414–423.

Goldman N. 1993. Simple diagnostic statistical tests of models for DNA substitution. Journal of Molecular Evolution 37: 650–661.

Goldman N. 1998. Phylogenetic information and experimental design in molecular systematics. Proceedings of the Royal Society of London Series B-Biological Sciences 265: 1779–1786.

Goldman N, Yang Z. 1994. Codon-based model of nucleotide substitution for protein-coding DNA-sequences. Molecular Biology and Evolution 11: 725–736.

Gowri-Shankar V, Rattray M. 2007. A reversible jump method for Bayesian phylogenetic inference with a nonhomogeneous substitution model. Molecular Biology and Evolution 24: 1286–1299.

Graybeal A. 1994. Evaluating the phylogenetic utility of genes—a search for genes informative about deep divergences among vertebrates. Systematic Biology 43: 174–193.

Graybeal A. 1998. Is it better to add taxa or characters to a difficult phylogenetic problem? Systematic Biology 47: 9–17.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology 52: 696–704.

Guyer C, Slowinski JB. 1991. Comparisons of observed phylogenetic topologies with null expectations among 3 monophyletic lineages. Evolution 45: 340–350.

Guyer C, Slowinski JB. 1993. Adaptive radiation and the topology of large phylogenies. Evolution 47: 253–263.

Hasegawa M, Hashimoto T. 1993. Ribosomal-RNA trees misleading? Nature 361: 23.

Heard SB. 1992. Patterns in tree balance among cladistic, phenetic, and randomly generated phylogenetic trees. Evolution 46: 1818–1826.

Heath TA, Zwickl DJ, Kim J, Hillis DM. 2008. Taxon sampling affects inferences of macroevolutionary processes from phylogenetic trees. Systematic Biology 57: 160–166.

Hedtke SM, Townsend TM, Hillis DM. 2006. Resolution of phylogenetic conflict in large data sets by increased taxon sampling. Systematic Biology 55: 522–529.

Hendy MD, Penny D. 1989. A framework for the quantitative study of evolutionary trees. Systematic Zoology 38: 297–309.

Hibbett D. 2004. Trends in morphological evolution in homobasidiomycetes inferred using maximum likelihood: a comparison of binary and multistate approaches. Systematic Biology 53: 889–903.

Hillis DM. 1995. Approaches for assessing phylogenetic accuracy. Systematic Biology 44: 3–16.

Hillis DM. 1996. Inferring complex phylogenies. Nature 383: 130–131.

Hillis DM. 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. Systematic Biology 47: 3–8.

Hillis DM. 1999. Phylogenetics and the study of HIV. In: Crandall KA ed. The Evolution of HIV. Baltimore: The Johns Hopkins University Press. 105–121.

Hillis DM, Allard MW, Miyamoto MM. 1993. Analysis of DNA sequence data: phylogenetic inference. Methods in Enzymology 224: 456–487.

Hillis DM, Huelsenbeck JP. 1994. Support for dental HIV transmission. Nature 369: 24–25.

Hillis DM, Huelsenbeck JP, Cunningham CW. 1994a. Application and accuracy of molecular phylogenies. Science 264: 671–677.

Hillis DM, Huelsenbeck JP, Swofford DL. 1994b. Hobgoblin of phylogenetics. Nature 369: 363–364.

Hillis DM, Pollock DD, McGuire JA, Zwickl DJ. 2003. Is sparse taxon sampling a problem for phylogenetic inference? Systematic Biology 52: 124–126.

Ho SYW. 2007. Calibrating molecular estimates of substitution rates and divergence times in birds. Journal of Avian Biology 38: 409–414.

Holland BR, Penny D, Hendy MD. 2003. Outgroup misplacement and phylogenetic inaccuracy under a molecular clock—A simulation study. Systematic Biology 52: 229–238.

Holman EW. 2005. Nodes in phylogenetic trees: The relation between imbalance and number of descendent species. Systematic Biology 54: 895–899.

Hoyle DC, Higgs PG. 2003. Factors affecting the errors in the estimation of evolutionary distances between sequences.

Molecular Biology and Evolution 20: 1–9.

Huelsenbeck JP. 1991. When are fossils better than extant taxa in phylogenetic analysis. Systematic Zoology 40: 458–469.

Huelsenbeck JP. 1995. Performance of phylogenetic methods in simulation. Systematic Biology 44: 17–48.

Huelsenbeck JP, Bollback JP. 2001. Empirical and hierarchical Bayesian estimation of ancestral states. Systematic Biology 50: 351–366.

Huelsenbeck JP, Hillis DM. 1993. Success of phylogenetic methods in the 4-taxon case. Systematic Biology 42: 247–264.

Huelsenbeck JP, Lander KM. 2003. Frequent inconsistency of parsimony under a simple model of cladogenesis. Systematic Biology 52: 641–648.

Huelsenbeck JP, Larget B, Swofford D. 2000. A compound Poisson process for relaxing the molecular clock. Genetics 154: 1879–1892.

Hug LA, Roger AJ. 2007. The impact of fossils and taxon sampling on ancient molecular dating analyses. Molecular Biology and Evolution 24: 1889–1897.

Hugall AF, Foster R, Lee MS. 2007. Calibration choice, rate smoothing, and the pattern of tetrapod diversification according to the long nuclear gene RAG-1. Systematic Biology 56: 543–563.

Hugall AF, Lee MSY. 2007. The likelihood node density effect and consequences for evolutionary studies of molecular rates. Evolution 61: 2293–2307.

Jermiin LS, Ho SYW, Ababneh F, Robinson J, Larkum AWD. 2004. The biasing effect of compositional heterogeneity on phylogenetic estimates may be underestimated. Systematic Biology 53: 638–643.

Johnson KP. 2001. Taxon sampling and the phylogenetic position of passeriformes: Evidence from 916 avian cytochrome b sequences. Systematic Biology 50: 128–136.

Kim J. 1998. Large-scale phylogenies and measuring the performance of phylogenetic estimators. Systematic Biology 47: 43–60.

Kirkpatrick M, Slatkin M. 1993. Searching for evolutionary patterns in the shape of a phylogenetic tree. Evolution 47: 1171–1181.

Kishino H, Hasegawa M. 1990. Converting distance to time: application to human evolution. Methods in Enzymology 183: 550–570.

Kishino H, Thorne JL, Bruno WJ. 2001. Performance of a divergence time estimation method under a probabilistic model of rate evolution. Molecular Biology and Evolution 18: 352–361.

Kolaczkowski B, Thornton JW. 2004. Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. Nature 431: 980–984.

Kolaczkowski B, Thornton JW. 2008. A mixed branch length model of heterotachy improves phylogenetic accuracy [online]. Molecular Biology and Evolution. doi:10.1093/molbev/msn042. Available from mbe.oxfordjournals.org [accessed 3 March 2008].

Langley CH, Fitch WM. 1974. An examination of the constancy of the rate of molecular evolution. Journal of Molecular Evolution 3: 161–177.

Lartillot N, Brinkmann H, Philippe H. 2007. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. BMC Evolutionary Biology 7 (Suppl. 1): S4.

Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Molecular Biology and Evolution 21: 1095–1109.

Lecointre G, Philippe H, Van Le HL, Le Guyader H. 1993. Species sampling has a major impact on phylogenetic inference. Molecular Phylogenetics and Evolution 2: 205–224.

Leipe DD, Gunderson JH, Nerad TA, Sogin ML. 1993. Small subunit ribosomal RNA+ of *Hexamita inflata* and the quest for the first branch in the eukaryotic tree. Molecular and Biochemical Parasitology 59: 41–48.

Lemmon AR, Brown JM, Stanger-Hall K, Lemmon EM. The effect of missing data in likelihood-based phylogenetics. Systematic Biology. (in press)

Lemmon AR, Milinkovitch MC. 2002. The metapopulation genetic algorithm: An efficient solution for the problem of phylogeny estimation. Proceedings of the National Academy of Sciences USA 99: 10516–10521.

Lemmon AR, Moriarty EC. 2004. The importance of proper model assumption in Bayesian phylogenetics. Systematic Biology 53: 265–277.

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

Lepage T, Lawi S, Tupper P, Bryant D. 2006. Continuous and tractable models for the variation of evolutionary rates. Mathematical Biosciences 199: 216–233.

Li P, Bousquet J. 1992. Relative-rate test for nucleotide substitutions between two lineages. Molecular Biology and Evolution 9: 1185–1189.

Lin YH, McLenachan PA, Gore AR, Phillips MJ, Ota R, Hendy MD, Penny D. 2002. Four new mitochondrial genomes and the increased stability of evolutionary trees of mammals from improved taxon sampling. Molecular Biology and Evolution 19: 2060–2070.

Lockhart PJ, Howe CJ, Bryant DA, Beanland TJ, Larkum AWD. 1992. Substitutional bias confounds inference of cyanelle origins from sequence data. Journal of Molecular Evolution 34: 153–162.

Lockhart PJ, Larkum AWD, Steel MA, Waddell PJ, Penny D. 1996. Evolution of chlorophyll and bacteriochlorophyll: The problem of invariant sites in sequence analysis. Proceedings of the National Academy of Sciences USA 93: 1930–1934.

Lopez P, Casane D, Philippe H. 2002. Heterotachy, an important process of protein evolution. Molecular Biology and Evolution 19: 1–7.

Martins EP. 1996. Phylogenies and the comparative method in animal behavior. Oxford: Oxford University Press.

Massingham T, Goldman N. 2000. EDIBLE: experimental design and information calculations in phylogenetics. Bioinformatics 16: 294–295.

McMahon MM, Sanderson MJ. 2006. Phylogenetic supermatrix analysis of GenBank sequences from 2228 papilionoid legumes. Systematic Biology 55: 818–836.

Metzker ML, Mindell DP, Liu XM, Ptak RG, Gibbs RA, Hillis DM. 2002. Molecular evidence of HIV-1 transmission in a criminal case. Proceedings of the National Academy of

Sciences USA 99: 14292–14297.

Minh BQ, Vinh LS, von Haeseler A, Schmidt HA. 2005. pIQPNNI: parallel reconstruction of large maximum likelihood phylogenies. Bioinformatics 21: 3794–3796.

Mooers AO. 1995. Tree balance and tree completeness. Evolution 49: 379–384.

Mooers AO, Holmes EC. 2000. The evolution of base composition and phylogenetic inference. Trends in Ecology and Evolution 15: 365–369.

Near TJ, Meylan PA, Shaffer HB. 2005. Assessing concordance of fossil calibration points in molecular clock studies: An example using turtles. American Naturalist 165: 137–146.

Near TJ, Sanderson MJ. 2004. Assessing the quality of molecular divergence time estimates by fossil calibrations and fossil-based model selection. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 359: 1477–1483.

Nee S, Holmes EC, May RM, Harvey PH. 1994a. Extinction rates can be estimated from molecular phylogenies. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 344: 77–82.

Nee S, May RM, Harvey PH. 1994b. The reconstructed evolutionary process. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 344: 305–411.

Olsen GJ. 1987. Earliest phylogenetic branchings: comparing rRNA-based evolutionary trees inferred with various techniques. Cold Spring Harbor Symposia on Quantitative Biology 52: 825–837.

Pagel M. 1994. Detecting correlated evolution on phylogenies—a general-method for the comparative-analysis of discrete characters. Proceedings of the Royal Society of London Series B-Biological Sciences 255: 37–45.

Pagel M. 1999. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. Systematic Biology 48: 612–622.

Pagel M, Meade A, Barker D. 2004. Bayesian estimation of ancestral character states on phylogenies. Systematic Biology 53: 673–684.

Philippe H, Douzery E. 1994. The pitfalls of molecular phylogeny based on four species, as illustrated by the cetacea/artiodactyla relationships. Journal of Mammalian Evolution 2: 133–152.

Philippe H, Lartillot N, Brinkmann H. 2005. Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. Molecular Biology and Evolution 22: 1246–1253.

Philippe H, Snell EA, Bapteste E, Lopez P, Holland PWH, Casane D. 2004. Phylogenomics of eukaryotes: Impact of missing data on large alignments. Molecular Biology and Evolution 21: 1740–1752.

Philippe H, Zhou Y, Brinkmann H, Rodrigue N, Delsuc F. 2005. Heterotachy and long-branch attraction in phylogenetics. BMC Evolutionary Biology 5: 50.

Phillips MJ, Delsuc F, Penny D. 2004. Genome-scale phylogeny and the detection of systematic biases. Molecular Biology and Evolution 21: 1455–1458.

Poe S. 1998a. The effect of taxonomic sampling on accuracy of phylogeny estimation: Test case of a known phylogeny.

Molecular Biology and Evolution 15: 1086–1090.

Poe S. 1998b. Sensitivity of phylogeny estimation to taxonomic sampling. Systematic Biology 47: 18–31.

Poe S. 2003. Evaluation of the strategy of long-branch subdivision to improve the accuracy of phylogenetic methods. Systematic Biology 52: 423–428.

Poe S, Swofford DL. 1999. Taxon sampling revisited. Nature 398: 299–300.

Pollock DD, Bruno WJ. 2000. Assessing an unknown evolutionary process: Effect of increasing site-specific knowledge through taxon addition. Molecular Biology and Evolution 17: 1854–1858.

Pollock DD, Taylor WR, Goldman N. 1999. Coevolving protein residues: Maximum likelihood identification and relationship to structure. Journal of Molecular Biology 287: 187–198.

Pollock DD, Zwickl DJ, McGuire JA, Hillis DM. 2002. Increased taxon sampling is advantageous for phylogenetic inference. Systematic Biology 51: 664–671.

Purvis A, Agapow PM. 2002. Phylogeny imbalance: taxonomic level matters. Systematic Biology 51: 844–854.

Purvis A, Gittleman JL, Luh HK. 1994. Truth or consequences—effects of phylogenetic accuracy on 2 comparative methods. Journal of Theoretical Biology 167: 293–300.

Pybus OG, Harvey PH. 2000. Testing macro-evolutionary models using incomplete molecular phylogenies. Proceedings of the Royal Society of London Series B-Biological Sciences 267: 2267–2272.

Rambaut A, Bromham L. 1998. Estimating divergence dates from molecular sequences. Molecular Biology and Evolution 15: 442–448.

Rannala B, Huelsenbeck JP, Yang Z, Nielsen R. 1998. Taxon sampling and the accuracy of large phylogenies. Systematic Biology 47: 702–710.

Ricklefs RE, Starck JM. 1996. Applications of phylogenetically independent contrasts: A mixed progress report. Oikos 77: 167–172.

Robinson M, Gouy M, Gautier C, Mouchiroud D. 1998. Sensitivity of the relative-rate test to taxonomic sampling. Molecular Biology and Evolution 15: 1091–1098.

Rodriguez-Ezpeleta N, Brinkmann H, Roure B, Lartillot N, Lang BF, Philippe H. 2007. Detecting and overcoming systematic errors in genome-scale phylogenies. Systematic Biology 56: 389–399.

Roelants K, Gower DJ, Wilkinson M, Loader SP, Biju SD, Guillaume K, Moriau L, Bossuyt F. 2007. Global patterns of diversification in the history of modern amphibians. Proceedings of the National Academy of Sciences USA 104: 887–892.

Roger AJ, Hug LA. 2006. The origin and diversification of eukaryotes: problems with molecular phylogenetics and molecular clock estimation. Philosophical Transactions of the Royal Society B-Biological Sciences 361: 1039–1054.

Rokas A, Kruger D, Carroll SB. 2005. Animal evolution and the molecular signature of radiations compressed in time. Science 310: 1933–1938.

Rokas A, Williams BL, King N, Carroll SB. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature 425: 798–804.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models.

Bioinformatics 19: 1572–1574.

Rosenberg MS, Kumar S. 2001. Incomplete taxon sampling is not a problem for phylogenetic inference. Proceedings of the National Academy of Sciences USA 98: 10751–10756.

Rosenberg MS, Kumar S. 2003. Taxon sampling, bioinformatics, and phylogenomics. Systematic Biology 52: 119–124.

Rutschmann F, Eriksson T, Abu Salim K, Conti E. 2007. Assessing calibration uncertainty in molecular dating: The assignment of fossils to alternative calibration points. Systematic Biology 56: 591–608.

Ryan MJ, Rand AS. 1995. Female responses to ancestral advertisement calls in túngara frogs. Science 269: 390–392.

Ryan MJ, Rand AS. 1998. Evoked vocal response in male túngara frogs: pre-existing biases in male responses? Animal Behavior 56: 1509–1516.

Saitou N, Nei M. 1987. The neighbor-joining method—a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4: 406–425.

Salisbury BA, Kim JH. 2001. Ancestral state estimation and taxon sampling density. Systematic Biology 50: 557–564.

Sanderson MJ. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Molecular Biology and Evolution 14: 1218–1231.

Sanderson MJ. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Molecular Biology and Evolution 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.

Sanderson MJ, Donoghue MJ. 1989. Patterns of variation in levels of homoplasy. Evolution 43: 1781–1795.

Sanderson MJ, Wojciechowski MF, Hu JM, Khan TS, Brady SG. 2000. Error, bias, and long-branch attraction in data for two chloroplast photosystem genes in seed plants. Molecular Biology and Evolution 17: 782–797.

Sarich VM, Wilson AC. 1973. Generation time and genomic evolution in primates. Science 179: 1144–1147.

Schluter D, Price T, Mooers AO, Ludwig D. 1997. Likelihood of ancestor states in adaptive radiation. Evolution 51: 1699–1711.

Shao KT, Sokal RR. 1990. Tree balance. Systematic Zoology 39: 266–276.

Smith AB, Pisani D, Mackenzie-Dodds JA, Stockley B, Webster BL, Littlewood DT. 2006. Testing the molecular clock: molecular and paleontological estimates of divergence times in the Echinoidea (Echinodermata). Molecular Biology and Evolution 23: 1832–1851.

Sorenson MD, Oneal E, Garcia-Moreno J, Mindell DP. 2003. More taxa, more characters: the hoatzin problem is still unresolved. Molecular Biology and Evolution 20: 1484–1498.

Spencer M, Susko E, Roger AJ. 2005. Likelihood, parsimony, and heterogeneous evolution. Molecular Biology and Evolution 22: 1161–1164.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22: 2688–2690.

Steel M. 1994. Recovering a tree from the leaf colourations it generates under a Markov model. Applied Mathematics Letters 7: 19–23.

Steel M, Szekely L, Erdos PL, Waddell P. 1993. A complete family of phylogenetic invariants for any number of taxa under Kimura 3ST model. New Zealand Journal of Botany 31: 289–296.

Stefanovic S, Rice DW, Palmer JD. 2004. Long branch attraction, taxon sampling, and the earliest angiosperms: Amborella or monocots? BMC Evolutionary Biology 4: 35.

Sullivan J, Swofford DL. 1997. Are guinea pigs rodents? The utility of models in molecular phylogenetics. Journal of Mammalian Evolution 4: 77–86.

Sullivan J, Swofford DL, Naylor GJP. 1999. The effect of taxon sampling on estimating rate heterogeneity parameters of maximum-likelihood models. Molecular Biology and Evolution 16: 1347–1356.

Susko E, Inagaki Y, Roger AJ. 2004. On inconsistency of the neighbor-joining, least squares, and minimum evolution estimation when substitution processes are incorrectly modeled. Molecular Biology and Evolution 21: 1629–1642.

Swofford DL, Olsen JL, Waddell PJ, Hillis DM. 1996. Phylogenetic inference. In: Hillis DM, Moritz C, Mable BK eds. Molecular systematics. Sunderland, MA: Sinauer Associates. 407–514.

Takezaki N, Rzhetsky A, Nei M. 1995. Phylogenetic test of the molecular clock and linearized trees. Molecular Biology and Evolution 12: 823–833.

Thorne JL, Kishino H. 2002. Divergence time and evolutionary rate estimation with multilocus data. Systematic Biology 51: 689–702.

Thorne JL, Kishino H, Painter IS. 1998. Estimating the rate of evolution of the rate of molecular evolution. Molecular Biology and Evolution 15: 1647–1657.

Van Den Bussche RA, Baker RJ, Huelsenbeck JP, Hillis DM. 1998. Base compositional bias and phylogenetic analyses: A test of the "flying DNA" hypothesis. Molecular Phylogenetics and Evolution 10: 408–416.

Venditti C, Meade A, Pagel M. 2006. Detecting the node-density artifact in phylogeny reconstruction. Systematic Biology 55: 637–643.

Webster AJ, Payne RJH, Pagel M. 2003. Molecular phylogenies link rates of evolution and speciation. Science 301: 478–478.

Whelan S, Goldman N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Molecular Biology and Evolution 18: 691–699.

Wiens JJ. 2003. Missing data, incomplete taxa, and phylogenetic accuracy. Systematic Biology 52: 528–538.

Wiens JJ. 2005. Can incomplete taxa rescue phylogenetic analyses from long-branch attraction? Systematic Biology 54: 731–742.

Wiens JJ, Reeder TW. 1995. Combining data sets with different numbers of taxa for phylogenetic analysis. Systematic Biology 44: 548–558.

Wu CI, Li WH. 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. Proceedings of the National Academy of Sciences USA 82: 1741–1745.

Xia X. 2006. Topological bias in distance-based phylogenetic methods: problems with over- and under-estimated genetic

distances. Evolutionary Bioinformatics 2: 375–387.

Yang Z. 1994. Statistical properties of the maximum-likelihood method of phylogenetic estimation and comparison with distance matrix-methods. Systematic Biology 43: 329–342.

Yang Z. 2006. Computational Molecular Evolution. New York: Oxford University Press.

Yang Z, Goldman N. 1997. Are big trees indeed easy? Trends in Ecology and Evolution 12: 357–357.

Yang Z, Rannala B. 1997. Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo method. Molecular Biology and Evolution 14: 717–724.

Yang Z, Rannala B. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Molecular Biology and Evolution 23: 212–226.

Yang Z, Roberts D. 1995. On the use of nucleic-acid sequences to infer early branchings in the Tree of Life. Molecular Biology and Evolution 12: 451–458.

Yang Z, Yoder AD. 2003. Comparison of likelihood and Bayesian methods for estimating divergence times using multiple gene loci and calibration points, with application to a radiation of cute-looking mouse lemur species. Systematic Biology 52: 705–716.

Yoder AD, Yang Z. 2000. Estimation of primate speciation dates using local molecular clocks. Molecular Biology and Evolution 17: 1081–1090.

Zhou Y, Holmes EC. 2007. Bayesian estimates of the evolutionary rate and age of hepatitis B virus. Journal of Molecular Evolution 65: 197–205.

Zuckerkandl E, Pauling L. 1962. Molecular disease, evolution, and genetic heterogeneity. In: Kasha M, Pullman B eds. Horizons in Biochemistry. New York: Academic Press. 189–225.

Zwickl DJ. 2006. Genetic algorithm approaches for the phylogenetic analyses of large biological sequence datasets under the maximum likelihood criterion. Ph.D. Dissertation. Austin, TX: The University of Texas at Austin.

Zwickl DJ, Hillis DM. 2002. Increased taxon sampling greatly reduces phylogenetic error. Systematic Biology 51: 588–598.