Syst. Biol. 58(5):501–508, 2009

DOI:10.1093/sysbio/syp045
Advance Access publication on August 20, 2009

Maximum Likelihood Estimates of Species Trees: How Accuracy of Phylogenetic Inference Depends upon the Divergence History and Sampling Design

JOHN E. MCCORMACK, HUATENG HUANG, AND L. LACEY KNOWLES*

Department of Ecology and Evolutionary Biology, and the Museum of Zoology, University of Michigan, Ann Arbor, MI 48109-1079, USA;
**Correspondence to be sent to: Museum of Zoology, 1109 Geddes Avenue, University of Michigan, Ann Arbor, MI 48109-1079, USA; E-mail: knowlesl@umich.edu.*

Abstract.—The understanding that gene trees are often in discord with each other and with the species trees that contain them has led researchers to methods that incorporate the inherent stochasticity of genetic processes in the phylogenetic estimation procedure. Recently developed methods for species-tree estimation that not only consider the retention and sorting of ancestral polymorphism but also quantify the actual probabilities of incomplete lineage sorting are expected to provide an improvement over earlier summary-statistic based approaches that discard much of the information content of gene trees. However, these new methods have yet to be tested on truly challenging evolutionary histories such as those marked by recent rapid speciation where high levels of incomplete lineage sorting and discord among gene trees predominate. Here, we test a new maximum-likelihood method that incorporates stochastic models of both nucleotide substitution and lineage sorting for species-tree estimation. Using a simulation approach, we consider a broad range of species-tree topologies under 2 scenarios representing moderate and severe incomplete lineage sorting. We show that the maximum-likelihood method results in more accurate species trees than a summary-statistic based approach, demonstrating that information contained in discordant gene trees can be effectively extracted using a full probabilistic model. Moreover, we demonstrate that the shape of the original species tree (i.e., the relative lengths of internal branches) has a significant impact on whether the species tree is estimated accurately. In the speciation histories explored here, it is not just the recent origin of species that affects the accuracy of the estimates but the variance in relative species divergence times as well. Additionally, we show that sampling effort (number of individuals and/or loci) and sampling design (ratio of individuals to loci) are both important factors affecting the accuracy of species-tree estimates, which is again affected by the relative timing of divergence among species. The inherent difficulties of estimating relationships when species have undergone a recent radiation are discussed, and in particular, the limitations with maximum-likelihood estimates of species trees that do not consider uncertainty in the estimated gene trees of individual loci. Thus, despite substantial improvements over current summary-statistic based approaches, and the increased sophistication of procedures that incorporate the process of gene lineage coalescence, recent radiations still appear to pose daunting challenges for phylogenetics. [Coalescence; gene tree; lineage sorting; phylogenetics; species tree.]

The species tree is usually the subject of interest for phylogeneticists, yet gene trees comprise the principal data used for species-tree inference. However, we know that gene trees can disagree in topology with the species tree that contains them (Pamilo and Nei 1988; Avise 1989; Takahata 1989; Maddison 1997; Page and Charleston 1997). Incomplete lineage sorting is one of the most common reasons in nature for this discord (Degnan and Rosenberg 2009). It potentially affects all sexually reproducing organisms and predominates in those that have experienced recent divergence (e.g., Sato et al. 1999; Knowles 2000; Dolman and Moritz 2006; Pollard et al. 2006). Thus, there is not only a need for methods for recovering species trees from gene trees, but the development and investigation of these methods have themselves also become exciting areas of research (Degnan and Salter 2005; Buckley et al. 2006; Degnan and Rosenberg 2006; Maddison and Knowles 2006; Pollard et al. 2006; Carstens and Knowles 2007; Liu and Pearl 2007; Mossel and Roch 2007; Brumfield et al. 2008; Knowles and Chan 2008; Liu et al. 2008).

Discord among gene trees poses a serious challenge for phylogenetic estimation. Concatenation of gene sequences (Rokas et al. 2005), once considered standard, is problematic because it does not incorporate the possibility of different evolutionary histories for different genes, and thus overlooks an important and inherent element in the evolution of gene trees (Kubatko and Degnan 2007). Especially for recent divergence, incomplete lineage sorting obscures species relationships (Carstens and Knowles 2007), and stochasticity in the lineage-sorting process can overwhelm phylogenetic signal to the extent that even the most likely gene tree sometimes does not match the species tree (e.g., anomalous gene tree phenomenon; Degnan and Rosenberg 2006; but see Huang and Knowles 2009). Other approaches to species-tree estimation rely on some predefined similarity among gene trees or DNA sequences to identify a species tree. These methods include democratic consensus, which takes the most frequent gene topology (Slowinski and Page 1999; Jennings and Edwards 2005), as well as methods that do not use gene trees at all, such as taxon clustering, which measure similarity in DNA sequence data (Takahata and Nei 1985; Maddison and Knowles 2006). Still others incorporate the idea of gene trees and coalescence, such as methods that estimate species trees by minimizing genealogical discord or the number of deep coalescents, but do not take full advantage of the information content contained in gene trees, such as branch lengths (Page and Charleston 1997; Maddison and Knowles 2006).

Recently, probabilistic methods for estimating species trees from multiple gene trees have been developed that combine Bayesian or maximum-likelihood models in a coalescent framework to incorporate the inherent stochasticity of gene lineage sorting (Degnan and Salter

2005; Liu and Pearl 2007; Kubatko et al. 2009). Insofar as they have been tested, these methods have been shown to provide accurate estimates of species trees (Edwards et al. 2007; Liu and Pearl 2007). However, they have thus far been applied only to a limited number of divergence scenarios because of computational constraints and have not addressed the truly challenging situation of recent rapid divergence, as occurs during evolutionary radiations.

Here, we test a recently developed method for obtaining maximum-likelihood estimates of species trees (Kubatko et al. 2009) from phylogenies representing the most difficult divergence scenarios to infer-recent radiations with widespread incomplete lineage sorting. Instead of exploring one or a few divergence scenarios, we explore 1000 different species trees that cover a broad spectrum of diversification histories where incomplete lineage sorting predominates. The accuracy of species-tree estimates is examined with respect to the underlying shape of the species tree (specifically, the relative lengths of internal branches) to understand why the accuracy of the estimates for a given sampling effort and divergence time differs among different species trees. To evaluate the extent to which the accuracy of species-tree estimates might be increased by more fully utilizing the information contained in gene trees (i.e., by incorporating the probability of gene trees into the estimation procedure), we also compare the maximum-likelihood method to the summary-statistic based approach of minimizing the number of deep coalescents (Maddison and Knowles 2006), which is also based on coalescent theory, but does not actually incorporate a model of gene lineage coalescence. Although promising, the summary-statistic approach left considerable room for improvement, especially for recent divergences where species trees proved difficult to estimate. More effective use of gene tree information content could have additional consequences of practical interest to many biologists. For example, it might effect the plateau at which the addition of more individuals or loci (i.e., sampling effort) does not improve the accuracy of species-tree estimates or shift the optimal allocation of sequencing to more individuals versus loci or vice versa (i.e., sampling design). We explore these questions by comparing the accuracy of species-tree estimates from the maximum-likelihood based approach (Kubatko et al. 2009) under differing sampling efforts and designs.

Methods

The conceptual design (Fig. 1), which followed Maddison and Knowles (2006), was to 1) generate 500 species trees with random branching patterns to represent the diversity of possible divergence scenarios; 2) simulate the process of gene coalescence along the branches of each species tree with variable numbers of loci and individuals per species; 3) simulate DNA sequence evolution along these coalescent gene trees; 4) estimate gene trees from DNA sequence data; 5) estimate species trees from the estimated gene trees; and 6) assess the accuracy of the estimated species tree from comparison of the estimated species tree with the known species tree from step 1. The degree of discord between a single gene tree and the underlying species tree depends on both the time since divergence and population size; thus, we standardize time in terms of coalescent units (t/N), where $t =$ generations and $N =$ population size. Divergence was simulated at 2 time depths (1N and 10N), corresponding to high and moderate levels of incomplete lineage sorting, respectively. Because we were interested in how sampling effort and sampling design affected the ability to reconstruct the true species tree, we generated sets of data with different overall number and ratios of loci and individuals (see below).

We began by simulating 500 species trees, each with 8 species. Each tree topology was generated at time depth 1N and 10N for a total of 1000 species trees. The topology and branch lengths were randomly generated under a uniform speciation (Yule) model in Mesquite (Maddison and Maddison 2004). Gene coalescence was simulated along the branches of the 500 species trees under a neutral coalescence model implemented with the program MS (Hudson 2002). For each of the 2 time depths, 12 data sets were created with different ratios of individuals to loci (individuals:loci = 1:1, 1:3, 1:9, 1:27, 1:50, 3:1, 3:3, 3:9, 9:1, 9:3, 27:1, 50:1). DNA sequences were generated by simulating molecular evolution along the branches of the coalescent gene trees using Seq-Gen (Rambaut and Grassly 1997). Sequences of 1000 base pairs were generated under an HKY85 model of nucleotide substitution with a transition–transversion ratio of 3.0, a gamma distribution with 4 categories, and a shape parameter of 0.8, assigning base pair states with the probability 0.3 A, 0.2 C, 0.3 T, 0.2 G.

Maximum likelihood gene trees were estimated from the simulated DNA sequences using genetic algorithm for rapid likelihood inference (GARLI) version 0.951 (Zwickl 2006) available at http://www.bio.utexas.edu/faculty/antisense/garli/Garli.html. GARLI analyses were conducted specifying the HKY85 model family (6 substitution rate categories) with gamma rate categories set to 4, and base pair state frequencies and the number of invariant sites estimated by the program. Maximum-likelihood trees were selected after 10,000 generations if no significant improvement in likelihood was observed, with the significant topological improvement level set at 0.01 (first condition for termination); the final solution was selected when the total improvement in the likelihood score was <0.05 compared with the last solution obtained (second condition for termination). All other GARLI settings involved in the genetic algorithm were default values, per recommendations of the developer (Zwickl 2006). Estimated gene trees were midpoint rooted and converted to ultrametric trees in PAUP* (Swofford 2000) for further species tree analysis.

The most likely species tree was calculated analytically with the program STEM, version 1.01 (Kubatko et al. 2009); θ was set to 0.01 to match the θ under which

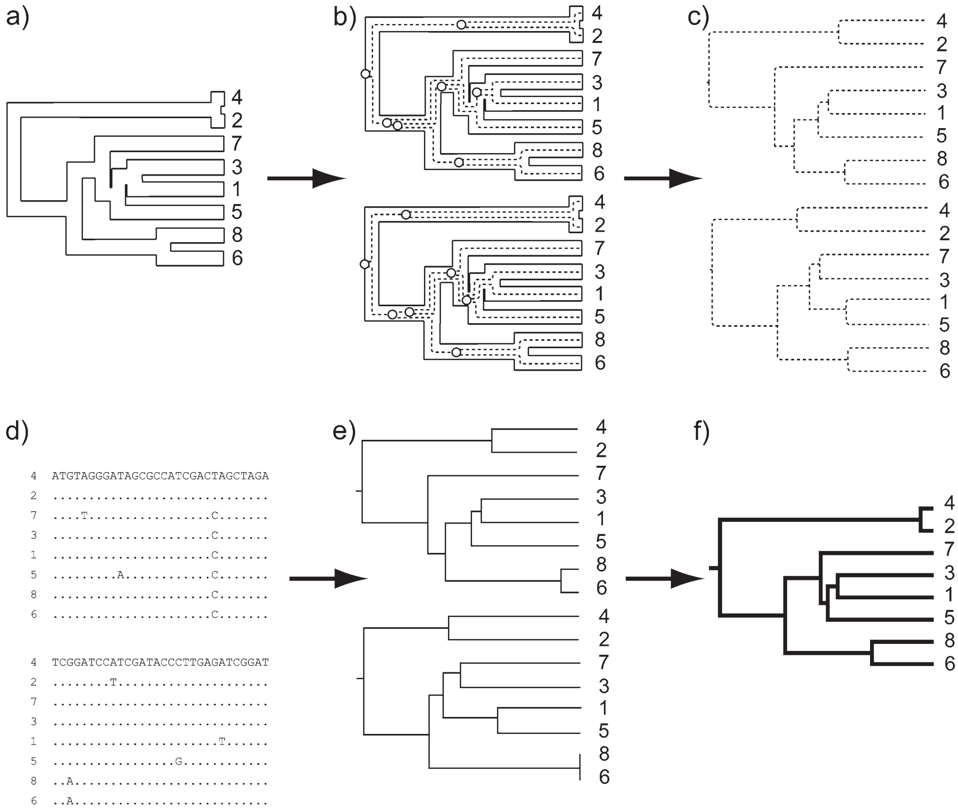
a)
4
2
7
3
1
5
8
6
b)
4
2
7
3
1
5
8
6
4
2
7
3
1
5
8
6
c)
4
2
7
3
1
5
8
6
4
2
7
3
1
5
8
6
d)
4 ATGTAGGGATAGCGCCATCGACTAGCTAGA
2 ....
7 ....T.................C.......
3 ....C.......
1 ....C.......
5 .........A............C.......
8 ....C.......
6 ....C.......
4 TCGGATCCATCGATACCCTTGAGATCGGAT
2 ....T.....................
7 ....
3 ....
1 ....T......
5 ....G............
8 ..A...........................
6 ..A...........................
e)
4
2
7
3
1
5
8
6
4
2
7
3
1
5
8
6
f)
4
2
7
3
1
5
8
6

FIGURE 1. Conceptual design for evaluating the accuracy of estimated species trees. a) Species trees were simulated through a random speciation process. b) Gene coalescence (dotted lines) was simulated within the branches of each species tree for variable numbers of individuals and loci. Two loci are illustrated with 1 individual per species showing some coalescence events (circles), which results in c) coalescent gene trees whose topologies do not match the species tree. d) DNA sequences were simulated along the branches of the coalescent gene trees. e) Maximum likelihood gene trees were estimated for each locus from the nucleotide data. f) Species trees were estimated from the estimated gene trees by minimizing deep coalescence or calculating a maximum-likelihood species trees; results were compared with the known species tree from (a).

the data were simulated. STEM is a coalescent-based program that can analytically derive the maximum-likelihood estimate for a species tree (both branch lengths and topology) given a set of gene trees with branch length information (Kubatko et al. 2009). Species trees were also estimated from the estimated gene trees using the minimize deep coalescence method (Maddison and Knowles 2006) implemented in Mesquite (Maddison and Maddison 2004). This method heuristically searches tree space for the species tree topology that minimizes the number of times gene coalescence occurs prior to speciation (i.e., deep coalescence). Given that a primary focus of phylogenetic studies is to recover the relationship between species (i.e., species-tree topology), accuracy was assessed by calculating topological similarity between the original species tree with the estimated species tree using TreeDist (part of the PHYLIP statistical package [Felsenstein 1993] via symmetric distance [Robinson and Foulds 1981]). Symmetric distance for rooted trees assesses the number of clades found in 1 tree, but not the other, with lower accuracy indicated by scores increasing in intervals of 2, from zero (=all clades the same) to twice the number of internal branches (=12 for rooted tree with 8 terminal taxa). This accuracy metric was converted to a proportion with 1.0 as the most accurate score.

Regression analysis was used to investigate how the accuracy of species-tree estimates was influenced by the underlying shape of the species tree. The standard deviation (SD) of branch lengths in the original species trees was used as a metric to characterize species-tree shape. Because all species trees were the same total depth, the SD is expected to be greater for trees with many short internodes and a few long branches (rapid radiation) compared with trees with speciation events at regular time intervals.

RESULTS AND DISCUSSION

Our results show that for the recent and rapid divergence scenarios we examined (i) the accuracy of species-tree estimates depends on the shape of the underlying evolutionary history and (ii) explicit consideration of the probability of deep coalescence in the estimation procedure provides an improvement over a method that relies on summary statistics, although some species histories are especially challenging and are not correctly estimated.

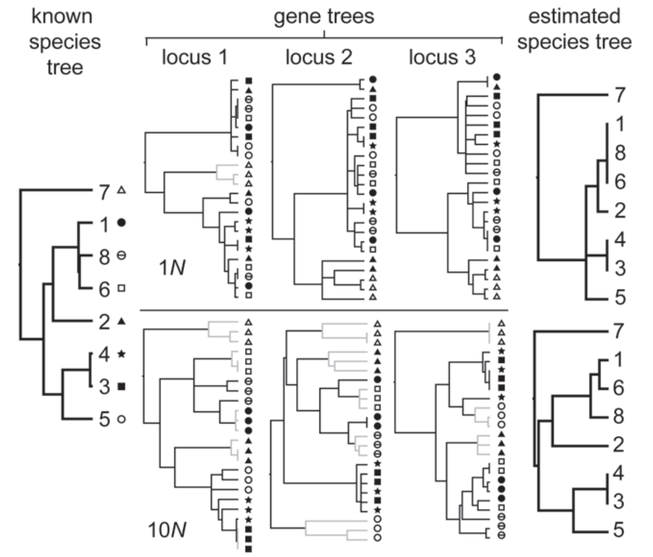

FIGURE 2. Discordant gene trees estimated from DNA sequence data from 3 loci simulated from 1 of the 500 starting species trees. In this example, each of the 8 species is represented by 3 individuals; gray lines identify monophyletic clades in the gene tree. At 1*N*, only in one instance do the individuals of a species form a clade, whereas at 10*N*, incomplete lineage sorting was less severe. Nevertheless, STEM estimated the species tree for this replicate with relatively high accuracy at 1*N* (0.67) and 10*N* (0.83).

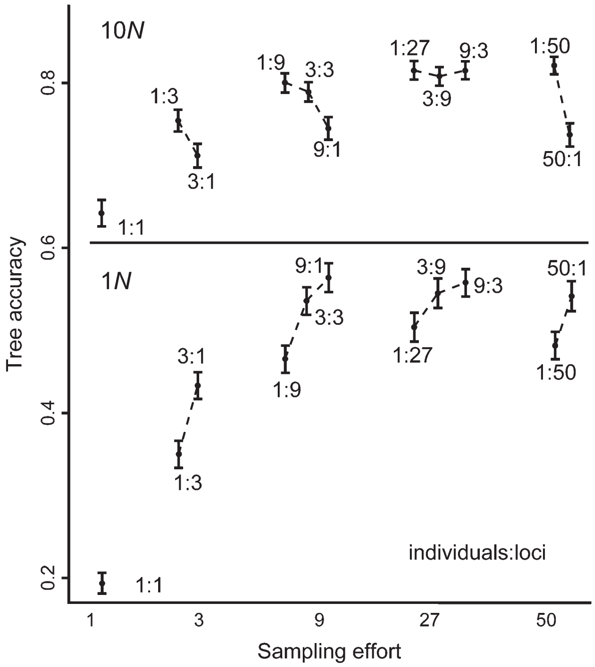

FIGURE 3. Change in tree accuracy with sampling effort for recent (1*N*) and older (10*N*) divergence (averages and standard errors for the multiple simulated data sets are shown). For both time depths, steep gains are made in tree accuracy up to 9 individuals and/or loci after which accuracy plateaus. For recent divergence, a higher ratio of individuals to loci results in more accurate species trees, whereas for older divergence, more loci compared with individuals, results in higher accuracy (positive vs. negative slopes for dotted lines in 1*N* vs. 10*N*, respectively).

Factors Influencing the Accuracy of Species-Tree Estimates

Very high levels of incomplete lineage sorting characterized the species histories we studied, which resulted in considerable genealogical discord among loci, as well as pervasive deep coalescence for any given locus (Fig. 2). Such genealogical discord was expected given the exceedingly high rate of speciation we were modeling, which was equivalent to 1 lineage originating every 12,500 or every 125,000 years, respectively, for the 1*N* and 10*N* total species-tree depths, assuming an effective population size of 100,000 individuals and a generation time of 1 year. This speciation rate corresponds to a net diversification interval (NDI) of 0.05 and 0.5 for the simulated species trees at 1*N* and 10*N* total tree depth, respectively. This range is within the magnitude of NDIs of some of the most impressive empirical examples of evolutionary radiations, including the African cichlids from Lake Malawi (NDI = 0.1–0.3), Galápagos finches (NDI = 0.8–1.1), and Hawaiian silverswords (NDI = 1.8) (see table 12.1, Coyne and Orr 2004, and the references therein). The overall rate of speciation (i.e., the number of species per total tree depth) had an obvious effect on the accuracy of species-tree estimates (i.e., accuracy scores at 10*N* were always higher than those at 1*N*) (Fig. 3). However, there was also considerable variation among the species trees simulated within each tree depth (Table 1), indicating that the shape of the species tree (i.e., the relative lengths of internal branches) influenced the accuracy of its estimation (Fig. 4). Lastly, the method itself also had a clear impact on the accuracy of the species-tree estimates. In general, irrespective of sampling effort and design (exceptions are discussed below), maximum-likelihood estimates were more accurate than those based on minimizing the number of deep coalescents (Table 1).

Divergence history's impact on species tree estimates.—Results from STEM clearly show that it is not only the recency of speciation but also the relative timing of species divergence (i.e., lengths of internal branches) that determines whether a species tree is accurately estimated. Within a particular sampling effort and design, the accuracy of the species tree estimates shows a wide range across the 500 simulated histories (as illustrated by the interquartile ranges, Table 1). Species trees with a high SD of branch lengths (indicative of many short internal branches) were less accurate; this metric alone explained a sizeable portion of the variance in accuracy scores (Fig. 4). The fact that this relationship varied little across sampling designs for the older species divergence times of 10*N* (i.e., similar slopes for each sampling design; solid lines in Fig. 5) highlights the inherent difficulty of obtaining accurate estimates of species relationships for some species trees, and no doubt contributes to the observed plateau in accuracy scores with increasing sampling effort (Fig. 3). However, for the shallower species divergence histories of 1*N*, there is a pronounced effect of sampling design on

TABLE 1. Average accuracy of species tree estimates for the 500 different evolutionary histories under the maximum-likelihood procedure and summary-statistic approach of minimizing deep coalescence for differing sampling efforts and sampling designs

Time depth	Total sampling effort	Individuals	Loci	Maximum-likelihood species tree[a]	Minimizing deep coalescence[a]
1N	1	1	1	0.16 (0.00–0.33)	0.15 (0.00–0.17)
	3	1	3	0.34 (0.17–0.50)	0.19 (0.00–0.33)
	9	1	9	0.47 (0.33–0.67)	0.27 (0.17–0.33)
	27	1	27	0.52 (0.33–0.67)	0.35 (0.17–0.50)
	50	1	50	0.49 (0.33–0.67)	0.38 (0.17–0.50)
	3	3	1	0.43 (0.33–0.67)	0.35 (0.17–0.50)
	9	3	3	0.55 (0.33–0.67)	0.53 (0.33–0.67)
	27	3	9	0.56 (0.33–0.83)	0.68 (0.50–0.83)
	9	9	1	0.58 (0.33–0.83)	0.58 (0.50–0.67)
	27	9	3	0.58 (0.50–0.67)	0.73 (0.67–0.83)
	27	27	1	0.57 (0.33–0.67)	0.68 (0.50–0.83)
	50	50	1	0.56 (0.33–0.67)[b]	0.69 (0.50–0.83)[c]
10N	1	1	1	0.67 (0.50–0.83)	0.49 (0.33–0.67)
	3	1	3	0.80 (0.67–1.00)	0.56 (0.33–0.67)
	9	1	9	0.86 (0.83–1.00)	0.59 (0.50–0.67)
	27	1	27	0.87 (0.83–1.00)	0.60 (0.50–0.67)
	50	1	50	0.88 (0.83–1.00)	0.60 (0.50–0.67)
	3	3	1	0.75 (0.67–0.83)	0.57 (0.50–0.67)
	9	3	3	0.84 (0.83–1.00)	0.68 (0.50–0.83)
	27	3	9	0.87 (0.83–1.00)	0.75 (0.67–0.83)
	9	9	1	0.79 (0.67–1.00)	0.63 (0.50–0.83)
	27	9	3	0.87 (0.83–1.00)	0.73 (0.67–0.83)
	27	27	1	0.79 (0.67–0.83)	0.64 (0.50–0.83)
	50	50	1	0.78 (0.67–0.83)	0.64 (0.50–0.83)[d]

[a] Average accuracy score with interquartile range (25–75th percentile) in parentheses.

[b] Calculated from 470 species trees.

[c] Calculated from 167 species trees due to computational time.

[d] Calculated from 214 species trees due to computational time.

the relationship between accuracy and the underlying species divergence history (i.e., the slope of the relationship changes from one design to another in Fig. 5). These results confirm the findings of Maddison and Knowles (2006) that for recent histories, sample design, and in particular, increasing the number of individuals, as opposed to loci, can significantly improve the accuracy of species-tree estimates. Our results further indicate that this difference in slope reflects the increased accuracy of estimates for species trees with longer internal branches (for dashed lines, the difference is mainly in the position of the y-intercept in these plots; Fig. 5). In contrast, the accuracy of species trees with a high SD of branch lengths (indicative of many short internal branches) remains low, despite shifts in sampling effort and design.

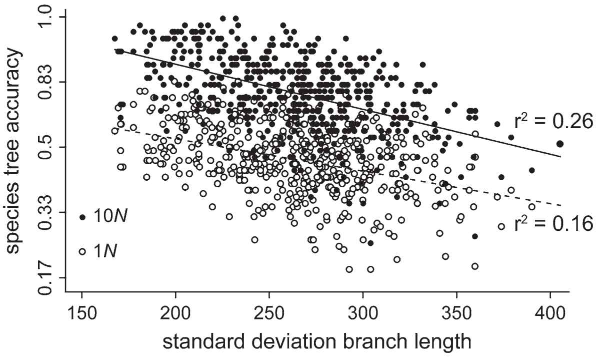

FIGURE 4. Negative relationship between SD of branch length and accuracy of species-tree estimation; simulated data sets for the recent (1N) and older (10N) divergence are shown by open and closed circles, respectively. Original species trees with short internodes (high SD of branch lengths) are harder to estimate accurately compared with those with more equivalent branch lengths.

Methodological influence on species-tree estimates.—The procedure used to estimate a species tree has a significant impact on the accuracy of the inference (Table 1), validating the theoretical prediction that explicit incorporation of a model of gene coalescence should improve species-tree estimation (Felsenstein 1993; Maddison 1997). Although the magnitude of this effect differs depending on the details of the species history and particular sampling design and effort, the impact of the procedure on the accuracy of a species-tree estimate can be substantial. For example, the relative increase in accuracy achieved by using the maximum-likelihood procedure implemented in STEM (Kubatko 2008) compared with the summary statistic approach of minimizing the number of deep coalescent implemented in Mesquite (Maddison and Maddison 2004) is similar to the gains in accuracy achieved by increasing the sampling effort from 1 to 27 gene copies per species (Table 1). Such gains in accuracy highlight how computational improvements can rival, and perhaps surpass, the increases in accuracy that can be achieved by modifying the sampling effort and design of a study, which has been the subject of previous work (Tajima 1983;

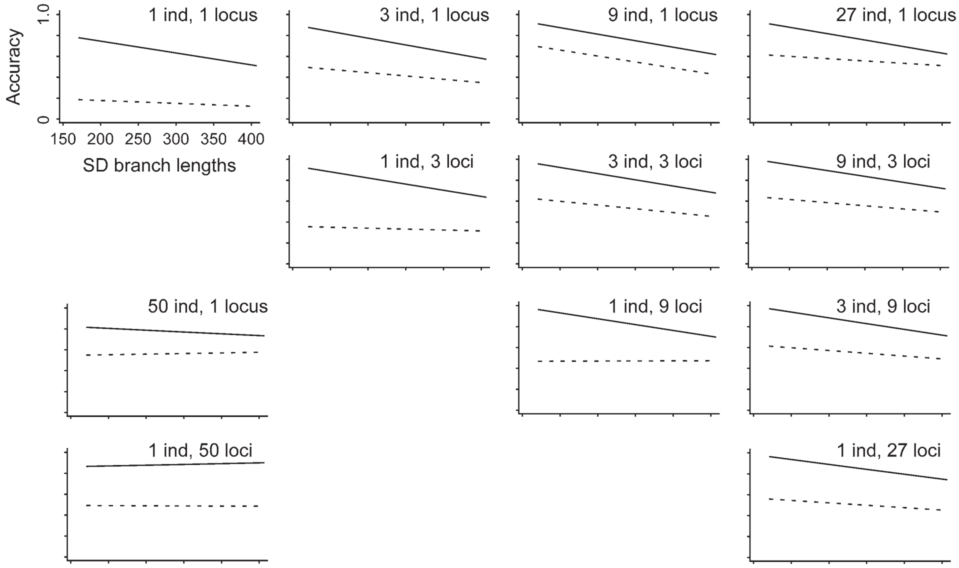

FIGURE 5. The effect of sampling design on the relationship between the standard deviation (SD) of branch lengths and accuracy of species-tree estimation. Among different sampling designs, there is a consistent negative relationship between the shape of the species trees and the accuracy at 10*N* (solid line); however, at 1*N* (dotted line), the relationship changes depending on sampling design, with gains usually achieved in species trees with low SD of branch lengths (i.e., relatively long internal branches).

Takahata and Nei 1985; Pamilo and Nei 1988; Maddison and Knowles 2006; Carstens and Knowles 2007; Edwards et al. 2007; Knowles and Chan 2008).

The gains in accuracy afforded by the maximum-likelihood approach, while impressive, did not mean that species relationships were always accurately estimated. At 1*N*, accuracy ranged from 0.16 to 0.58, and at 10*N*, accuracy ranged from 0.67 to 0.87, with accuracy generally increasing with sampling effort until reaching a plateau between 9 and 27 individuals and/or loci (Fig. 3). Nevertheless, for many of the species histories and sampling efforts and designs, the maximum-likelihood estimation procedure is quite accurate. For example, an accuracy score of 0.67 in this case corresponds to the misplacement of a single species (moving it from 1 clade to another).

Although there is a general increase in the accuracy of species-tree estimates with increased sampling effort for both estimation procedures for 10*N* (Table 1), sampling design had a greater influence using the maximum-likelihood approach. This contrasts with the minimal effect sampling design has on species tree accuracy for older species divergences based on the summary-statistic approach (see Fig. 5, Maddison and Knowles 2006). Another interesting contrast in performance between the methods is where the summary-statistic approach apparently out-performs the maximum-likelihood approach. This involves shallow species trees of depth 1*N* and an increased allocation of sampling effort to individuals rather than loci. For example, for a total sampling effort of 27 gene copies per species, when more than 1 individual per species is sampled (e.g., 3 individuals and 9 loci, 9 individuals and 3 loci, and 27 individuals and 1 locus), the accuracy of the maximum-likelihood estimates was lower than those based on minimizing the number of deep coalescents but only for the more recent species tree depth of 1*N* (Table 1). Because the summary-statistic approach relies on the pattern of deep coalescence as information for inferring species relationships, it is not surprising that the method does not perform as well as the maximum-likelihood approach for the deeper species trees (i.e., 10*N*). As the amount of incomplete lineage sorting decreases, as expected with older species divergence times, the information content relevant to a method based on minimizing the number of deep coalescents is necessarily also decreasing (Maddison and Knowles 2006; Knowles and Chan 2008). However, why the maximum-likelihood estimates do not exhibit proportional increases in accuracy at 1*N* when the sampling effort shifts from 9 to 27 gene copies and multiple individuals are sampled is less clear and warrants further investigation (see below).

Intractable Histories versus Methodological Shortcomings

Short internodes are expected to exacerbate the difficulties that mutational and coalescent stochasticity can pose for the accurate estimation of species relationships (Takahata and Nei 1985; Pamilo and Nei 1988; Kubatko and Degnan 2007). In the most extreme case, it is even possible that the most likely gene tree topology will not match the true species trees (i.e., anomolous gene trees, AGTs; Degnan and Rosenberg 2006). Without an analytical solution for the maximum branch length generating AGTs in a species trees with 8 taxa, it is not clear how much AGTs might contribute to inaccurate species-tree estimates in this case. However, a recent study that examines the prevalence of AGTs for estimated gene trees indicates that although AGTs are theoretically possible, they are unlikely to pose a significant danger to empirical phylogenetic study (Huang and Knowles 2009). Specifically, when both mutational and coalescent variance are taken into account, as done in this study (i.e., analyses are based on estimated gene trees, rather than coalescent gene trees), polytomies (i.e., unresolved estimated gene trees)—not AGTs—predominate within the zone of species tree branch lengths where AGTs are possible (Huang and Knowles 2009).

Although the 2 methods explored here explicitly consider coalescent stochasticity, neither takes into account uncertainty in gene tree estimation, as do some other approaches (e.g., Liu and Pearl 2007). Therefore, it is possible that the accuracy could be improved upon by considering this additional source of variance. However, the increased computational demands of such approaches (e.g., species tree estimates from the program BEST; Liu and Pearl 2007) limit exploration of a broad spectrum of species histories and sampling strategies, as we have examined here. Without such analyses, the boundary between intractable history and possible methodological shortcomings remains an open question.

It is worth noting the disproportionately small gains in accuracy with increased sampling for the older species divergences of 10*N* compared with 1*N* (Fig. 3). For example, despite a doubling of sampling effort to 50 loci for the older species divergences (10*N*), the improvement (if any) over 27 loci was negligible (Table 1). This pattern highlights one of the inherent challenges to increasing the accuracy of species-tree estimates (i.e., it is not likely to reflect the failure of methods to consider errors in the gene tree estimates). Likewise, the lack of an increase in accuracy when the sampling effort shifts from 9 to 27 gene copies when multiple individuals are sampled (Table 1) for both recent (1*N*) and older divergences (10*N*) with the maximum-likelihood approach hints at other possible pitfalls. With increased sampling, the number of possible gene tree topologies increases dramatically (Felsenstein 2004). Without sufficient mutations, the potential information contained in the pattern of gene lineage coalescence with additional sampling will be lost. Additional studies that consider the stochasticity of not only the coalescent (e.g., Degnan and Rosenberg 2006; Liu and Pearl 2007) but also mutation are needed to resolve such enigmatic effects of sampling design on the accuracy of species-tree estimates.

Conclusions

Recent developments in phylogenetics have opened the door to a suite of methods that represent a fundamental shift in how we go about reconstructing species histories. Discord among gene trees, once an intractable problem, is now widely understood as a natural biological phenomenon (explained by the coalescent; Kingman 1982), and as such, can be explicitly incorporated into the estimation procedure (Maddison and Knowles 2006), similar to the process of mutation (Felsenstein 2004). Attention is now focused on how to extract the information content of multiple discordant gene histories (Degnan and Salter 2005; Carstens and Knowles 2007; Edwards et al. 2007; Knowles and Carstens 2007; Kubatko and Degnan 2007; Liu and Pearl 2007). Here, we show that the maximum-likelihood species tree calculated in STEM (Kubatko et al. 2009) can offer substantial improvements in accuracy compared with a summary-statistic approach that minimizes the number of deep coalescents (Maddison and Knowles 2006). By examining a broad array of species histories, as opposed to focusing on the accuracy of species-tree estimates for a few case histories (Carstens and Knowles 2007; Liu and Pearl 2007), we are able to demonstrate both the general feasibility and possible limits of obtaining accurate species-tree estimates, as well as properties of the species trees that make them difficult to estimate. Our results highlight the incredible promise of approaches for estimating species trees, including those species histories that defy analysis under traditional phylogenetic analysis because of widespread incomplete lineage sorting. Yet, species histories marked by recent rapid radiation continue to pose a major challenge. Although these new methods are beginning to be applied to empirical questions (Carstens and Knowles 2007;
Edwards et al. 2007; Knowles and Carstens 2007; Belfiore et al. 2008; Brumfield et al. 2008; Carling and Brumfield 2008; Liu et al. 2008; Niemiller et al. 2008), sensitivity to aspects of the species divergence history and sampling design highlight the need for more study to understand how the accuracy of species-tree estimates might be improved.

Funding

This project was funded by National Science Foundation DEB 0447224 and DEB 0715487 to L.L.K.

Acknowledgements

We would like to thank L. Kubatko for assistance with implementing STEM and P. Klimov for kindly loaning processor time for Mesquite runs.

References

Avise J.C. 1989. Gene trees and organismal histories: a phylogenetic approach to population biology. Evolution. 43:1192–1208.

Belfiore N.M., Liu L., Moritz C. 2008. Multilocus phylogenetics of a rapid radiation in the genus *Thomomys* (Rodentia: Geomyidae). Syst. Biol. 57:294–310.

Brumfield R., Liu L., Lum D., Edwards S.V. 2008. Comparison of species tree methods for reconstructing the phylogeny of bearded manakins (Aves: Pipridae, *Manacus*) from multilocus sequence data Syst. Biol. 57:719–731.

Buckley T., Cordeiro M., Marshall D., Simon C. 2006. Differentiating between hypotheses of lineage sorting and introgression in New Zealand alpine cicadas (*Maoricicada* Dugdale). Syst. Biol. 55:411–425.

Carling M.D., Brumfield R.T. 2008. Integrating phylogenetic and population genetic analyses of multiple loci to test species divergence hypotheses in *Passerina* buntings. Genetics. 178:363–377.

Carstens B.C., Knowles L.L. 2007. Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: an example from *Melanoplus* grasshoppers. Syst. Biol. 56:400–411.

Coyne J.A., Orr H.A. 2004. Speciation. Sunderland (MA): Sinauer Associates.

Degnan J.H., Rosenberg N.A. 2006. Discordance of species trees with their most likely gene trees. PLoS Genet. 2:762–768.

Degnan J.H., Rosenberg N.A. 2009. Gene tree discordance, phylogenetic inference, and the multispecies coalescent. Trends Ecol. Evol. 24:332–340.

Degnan J.H., Salter L.A. 2005. Gene tree distributions under the coalescent process. Evolution. 59:24–37.

Dolman G., Moritz C. 2006. A multilocus perspective on refugial isolation and divergence in rainforest skinks (*Carla*). Evolution. 60:573–582.

Edwards S.V., Liu L., Pearl D.K. 2007. High-resolution species trees without concatentation. Proc. Natl. Acad. Sci. USA. 104:5936–5941.

Felsenstein J. 1993. PHYLIP (phylogeny inference package). Version 3.5c. Distributed by the author. Seatle (WA): Department of Genetics, University of Washington.

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

Huang H., Knowles L.L. 2009. What's the danger of the anomaly zone for empirical phylogenetics? Syst. Biol., in press.

Hudson R. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics. 18:337–338.

Jennings W.B., Edwards S.V. 2005. Speciational history of Australian grass finches (*Poephila*) inferred from thirty gene trees. Evolution. 59:2033–2047.

Kingman J.F.C. 1982. The coalescent. Stoch. Process Appl. 13:235–248.

Knowles L.L. 2000. Tests of Pleistocene speciation in montane grasshoppers (genus *Melanoplus*) from the sky islands of western North America. Evolution. 54:1337–1348.

Knowles L.L., Carstens B.C. 2007. Estimating a geographically explicit model of population divergence. Evolution. 61:477–493.

Knowles L.L., Chan Y.-H. 2008. Resolving species phylogenies of recent evolutionary radiations. Ann. Mo. Bot. Gard. 95:224–231.

Kubatko L.S., Carstens B.C., Knowles L.L. 2009. STEM: Species tree estimation using maximum likelihood for gene trees under coalescence. Bioinformatics. 25:971–973.

Kubatko L.S., Degnan J.H. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56:17–24.

Liu L., Pearl D.K. 2007. Species trees from gene trees: reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. Syst. Biol. 56:504–514.

Liu L., Pearl D., Brumfield R., Edwards S. 2008. Estimating species trees using multiple-allele DNA sequence data. Evolution. 62: 2080–2091.

Maddison W.P. 1997. Gene trees in species trees. Syst. Biol. 46:523–536.

Maddison W.P., Knowles L.L. 2006. Inferring phylogeny despite incomplete lineage sorting. Syst. Biol. 55:21–30.

Maddison W.P., Maddison D.R. 2004. Mesquite: a modular system for evolutionary analysis. Version 1.01. Available from URL http://mesquiteproject.org.

Mossel E., Roch S. 2007. Incomplete lineage sorting: consistent phylogenetic estimation from multiple loci. Available from URL http://arXiv:0710.0262v2.

Niemiller M.L., Fitzpatrick B.M., Miller B.T. 2008. Recent divergence with gene flow in Tennessee cave salamanders (Plethodontidae: *Gyrinophilus*) inferred from gene genealogies. Mol. Ecol. 17:2258–2275.

Page R.D.M., Charleston M.A. 1997. From gene to organismal phylogeny: reconciled trees and the gene tree/species tree problem. Mol. Phylogenet. Evol. 7:231–240.

Pamilo P., Nei M. 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5:568–583.

Pollard D.A., Iyer V.N., Moses A.M., Eisen M.B. 2006. Widespread discordance of gene trees with species tree in *Drosophila*: Evidence for incomplete lineage sorting. PLoS Genet. 2:e173.

Rambaut A., Grassly N. 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. Comput. Appl. Biosci. 13:235–238.

Robinson D.F., Foulds L.R. 1981. Comparison of phylogenetic trees. Math. Biosci. 53:131–147.

Rokas A., Williams B.L., King N., Carroll S.B. 2005. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature. 425:798–804.

Sato A., O'hUigin C., Figueroa F., Grant P.R., Grant B.R., Tichy H., Klein J. 1999. Phylogeny of Darwin's finches as revealed by mtDNA sequences. Proc. Natl. Acad. Sci. USA. 96:5101–5106.

Slowinski J.B., Page R.D.M. 1999. How should species phylogenies be inferred from sequence data? Syst. Biol. 48:814–825.

Swofford D.L. 2000. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4.0b. Sunderland (MA): Sinauer Associates.

Tajima F. 1983. Evolutionary relationships of DNA sequences in finite populations. Genetics. 105:437–460.

Takahata N. 1989. Gene genealogy in three related populations: consistency probability between gene and population trees. Genetics. 122:957–966.

Takahata N., Nei M. 1985. Gene geneaology and variance of interpopulational nucleotide differences. Genetics. 110:325–344.

Zwickl D.J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion [Ph.D. dissertation]. [Austin (Texas)]: University of Texas.

Received 24 November 2008; reviews returned 5 March 2009; accepted 10 July 2009

Guest Associate Editor: Scott Edwards