Syst. Biol. 58(1):100–113, 2009
Copyright © Society of Systematic Biologists
DOI:10.1093/sysbio/syp013
Advance Access publication on May 25, 2009

Measuring Branch Support in Species Trees Obtained by Gene Tree Parsimony

SIMON JOLY[1,*] AND ANNE BRUNEAU[2]

[1] *Allan Wilson Centre for Molecular Ecology and Evolution, Massey University, Palmerston North, New Zealand;*
[2] *Institut de recherche en biologie végétale and Département de Sciences biologiques, Université de Montréal, Montréal, QC, Canada H1X 2B2;*
[*] *Correspondence to be sent to: Department of Biology, McGill University, 1205 Docteur Penfield, Montréal, QC, Canada H3A 1B1; E-mail: simon.joly@mail.mcgill.ca.*

Abstract.—Several methods have recently been developed that allow the reconstruction of species trees from gene trees, an important achievement in our ongoing quest to obtain reliable species phylogenies. However, considerably less attention has been given to evaluating the accuracy of species trees' estimates. Four methods for measuring branch support of species trees are tested in this study in a gene tree parsimony framework: 1) bootstrap lineages (BL) (sequences) within species, 2) bootstrap characters (BC) within genes (i.e., the standard nonparametric bootstrap), 3) bootstrap lineages and characters (BLC), and 4) posterior probability gene tree sampling (PPGTS) (where, for each resampled data set, gene trees are sampled according to their posterior probability). For each method, *n* species trees are reconstructed from *n* resampled data sets and the branch support consists in the percentage of the *n* species trees in which a branch is recovered. The 4 methods were tested for several species trees and for different sampling efforts (i.e., number of genes and individuals sampled) using coalescent simulations. PPGTS performed best overall with lowest Type I and II error rates, followed by BLC. The BL and BC methods had higher error rates. This suggests that in order to properly measure branch support in a species tree context, it is important to account for the uncertainty involved in reconstructing gene trees from DNA sequences as well as that involved in reconstructing the species tree from individual gene trees. With the parameters used in the simulations, sampling more individuals per species resulted in similar improvements in support values as when sampling more genes. Moreover, sampling more individuals per species appeared to be important for escaping the anomaly zone present when only 1 sequence was sampled. We also apply the 4 methods to obtain branch supports for the species phylogeny of diploid wild roses (*Rosa*) in North America. [Branch support; coalescent theory; gene trees; gene tree parsimony; incomplete lineage sorting; nonparametric bootstrap; species trees.]

Reconstructing the evolution of species from genes represents a great challenge, considering that species and genes may have incongruent evolutionary histories (Pamilo and Nei 1988; Takahata 1989; Wu 1991; Doyle 1992; Hudson 1992; Maddison 1997; Nichols 2001; Rosenberg 2002). Simply interpreting gene trees as an estimate of species phylogenies—a common practice—can indeed be misleading. In some situations, the most likely gene tree can even be one that is incongruent with the species tree (Degnan and Rosenberg 2006). Although these problems seem difficult to overcome, important improvements have been made recently toward obtaining reliable species phylogenies. Several methods from diverse philosophical backgrounds (parsimony, likelihood, and Bayesian) have been described for estimating the species trees from 1 or more incongruent gene trees (Maddison 1997; Page and Charleston 1997; Jennings and Edwards 2005; Maddison and Knowles 2006; Carstens and Knowles 2007; Liu and Pearl 2007). Although these recent developments are promising, there has been considerably less attention given to evaluating the accuracy of these species trees. Although full Bayesian approaches (e.g., Liu and Pearl 2007) address the uncertainty of the species tree inference, species trees obtained from parsimony and likelihood criteria are rarely tested (but see Buckley et al. 2006, for an exception).

The species tree methods mentioned above are philosophically and methodologically different from gene tree analyses. Whereas gene tree analyses can be thought of as a 1-level approach where the phylogeny is estimated from DNA sequences, species tree analyses can be viewed as a 2-level approach: gene treesare first estimated from DNA sequences, and these gene trees are then used to reconstruct the species phylogeny. Common methods for estimating branch support on gene trees, such as the nonparametric bootstrap, strictly address the uncertainty involved with the first level of analysis, which evaluates how the sequence data support the gene tree. They do not account for the uncertainty involved with the second level of analysis—finding the best species tree for a given set of gene trees. Consequently, methods for measuring branch support on gene trees and their standard interpretations may not be appropriate for species trees and new methods may be required.

The uncertainty involved in estimating species trees from gene trees can be significant. One source of uncertainty occurs when a limited number of genes are sampled for reconstructing the species tree. The consequence of this uncertainty on species trees' estimates can be assessed by, for instance, gene tree bootstrapping (Burleigh et al. 2006). Yet another source of uncertainty is introduced when sampling individuals within species (Maddison and Knowles 2006). Indeed, the actual sampling of individuals can affect the outcome of a species tree search if a species has individuals that are most closely related to individuals from different species. The impact of such sampling errors—number of genes and lineages sampled—is expected to be more severe when the time between speciation events (t) is short relative to the number of gene copies in the species (i.e., $t < 2Ne$ for diploid organisms; Rosenberg and Nordborg 2002). Because adaptive radiations may be frequent in nature (Seehausen 2004), such circumstances could be

common. In those instances, it might be important to account for these uncertainties when evaluating branch support in species tree analyses.

In this study, we compare the performance of 4 methods, 2 of which are described for the first time, for measuring branch support for species trees obtained by gene tree parsimony. The idea of gene tree parsimony is to select a species phylogeny that minimizes the incongruence between the gene trees and the species tree (Slowinski et al. 1997). Here, we use the approach of Maddison (1997), which is to minimize the number of deep coalescences (i.e., occurrences of incomplete lineage sorting). The gene tree parsimony approach has several advantages for testing branch support methods. First, it is implemented in computer programs such as Mesquite (Maddison and Maddison 2006) and GeneTree (Page 2001). Second, it does not restrict the number of individuals per species to be included. Finally, Maddison and Knowles (2006) have shown that it can recover accurate phylogenies despite widespread incomplete lineage sorting. The different support methods were tested using coalescent simulations on different species trees that represent different degrees of difficulty in recovering the true species tree. The 4 methods were also applied to an empirical data set to assess the robustness of phylogenetic relationships among diploid species of roses (*Rosa* L.) in North America.

Methods

Four Methods for Measuring Branch Support

All 4 methods follow the same idea, which is that n species trees are reconstructed from n resampled data sets. For all methods, the branch support for a specific branch corresponds to the percentage of times that this branch occurs among the n species trees. The differences among methods depend on how data sets are resampled.

Bootstrap lineages.—Sampling different gene copies (i.e., lineages) per species can potentially affect the result of a species tree search. One way to assess the robustness of a phylogeny with respect to the actual sample of lineages is to bootstrap (resample with replacement) lineages (BL) within species. For each gene (data set), resampled data sets are obtained by BL within species to obtain the same number of lineages per species as in the original data set.

Bootstrap characters.—One legitimate question is whether the standard nonparametric bootstrap provides reliable branch supports for species trees, even though this procedure only assesses the uncertainty involved in reconstructing gene trees. In fact, different resampled matrices could result in different gene trees that may support different groups of species. With this approach, resampled data sets are obtained by bootstrapping characters (BC) within each gene until data sets of the same size as the original are obtained.

Bootstrap lineages and characters.—This approach is a combination of the 2 previous ones. Characters are bootstrapped within genes to assess how gene tree reconstruction is affected by the DNA sequences at hand and lineages are bootstrapped within species to evaluate how lineage sampling per species affects the species phylogeny. This approach therefore investigates uncertainty at both levels of analysis involved in reconstructing species trees.

Posterior probability gene tree sampling.—With this approach, the species tree for each replicate is reconstructed by using, for each gene, 1 gene tree that is sampled according to its posterior probability. The posterior distribution of gene trees for each data set (gene) is determined a priori by a Bayesian phylogenetic analysis. For example, if a given gene tree has a posterior probability of 0.2, then it has a 20% chance to be sampled for a given replicate. This method only accounts for the uncertainty in reconstructing genes trees from DNA sequences. It was first proposed by Buckley et al. (2006), but its performance has never been assessed using simulations.

Simulations

Simulations were performed on rooted species trees of 4 taxa representing scenarios that vary with respect to the difficulty of recovering the true tree (Figs 1–4). The depth of species trees was always 100 000 generations, and the population sizes were maintained at 100 000 haploid individuals throughout the tree. At the easy end of the spectrum was a symmetric species tree with external:internal branch length ratio of 1:9. The second and third species trees were also symmetric, but this time with an external:internal branch lengths of 1:1 and 9:1, respectively. The fourth situation, referred to as the "misleading" topology, consisted of an asymmetric species tree with short internal branches of 5000 generations (or 0.05 coalescent unit, where a coalescent unit = number of generations/number of gene copies in the population), which places the species tree into the anomaly zone (sensu Degnan and Rosenberg 2006) when a single sequence is sampled per species. In the anomaly zone, the most likely gene tree is incongruent with the species tree. Indeed, it can be shown (equations 1–3, Degnan and Rosenberg 2006) that with the asymmetric species tree $(A,(B,(C,D)))$ used in the simulations, gene trees with topology $((a,b),(c,d))$, $((a,c),(b,d))$, and $((a,d),(b,c))$ are all more likely than a gene tree with topology $(a,(b,(c,d)))$. Finally, the last situation consisted of a star topology and was used for investigating support values obtained for branches that do not exist (i.e., false branches). Because no true relationships exist in the star phylogeny, support values obtained for a branch necessarily support a false relationship.

Simulations were performed with 4 different sampling schemes that represent a subset of those used by Maddison and Knowles (2006): 3 genes and 1 individual sampled per species, 3 genes and 3 individuals

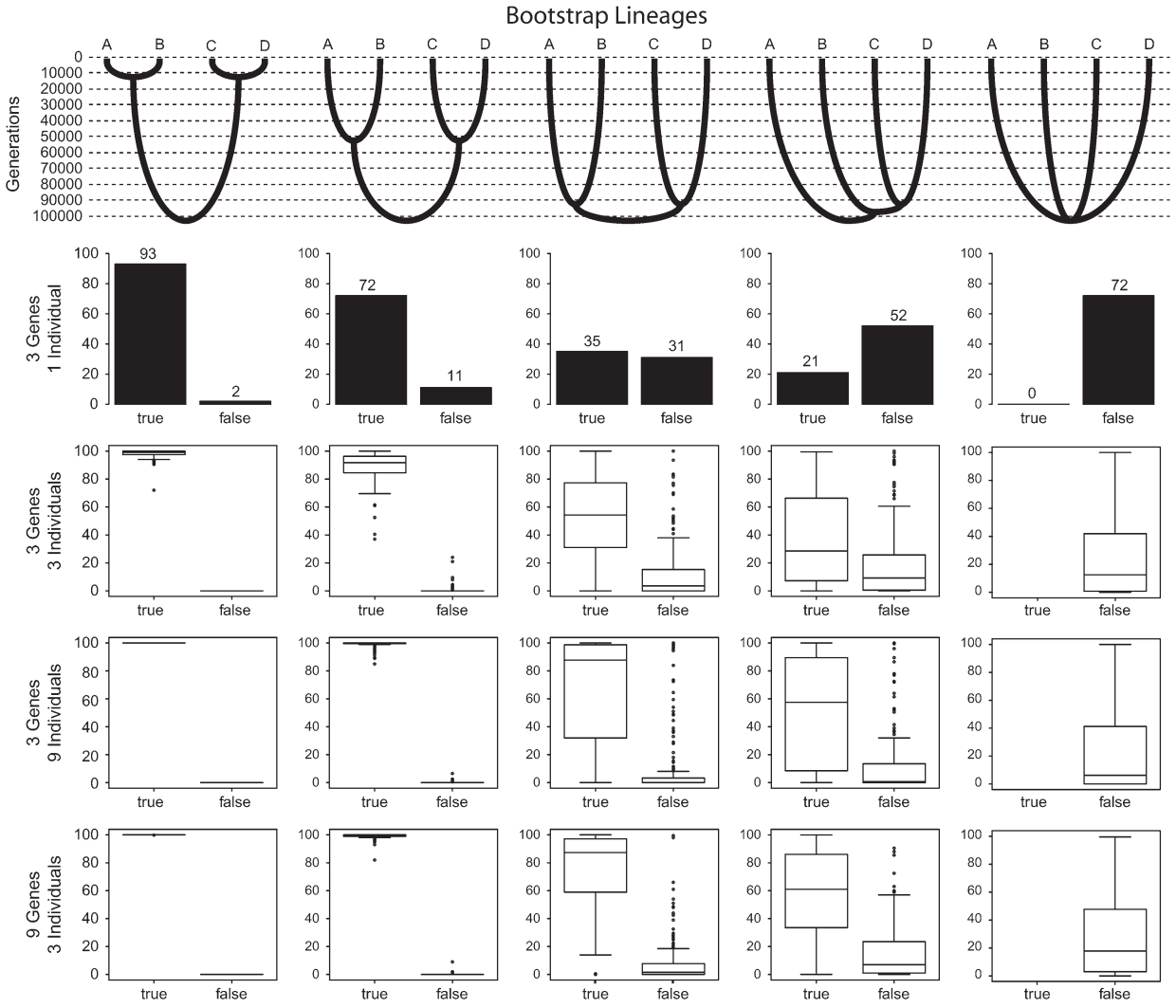

FIGURE 1. Support values obtained by BL (individuals) within species. Values are given for 5 different species trees and for 4 different sampling strategies: 3 genes and 1 individual sampled per species, 3 genes and 3 individuals, 3 genes and 9 individuals, and 9 genes and 3 individuals. For the simulations with only 1 sequence sampled per species, it is not possible to report support values per se because the bootstrapped data sets will always be identical to the original. In these situations, the number of times true and false species tree topologies were recovered is presented instead. The sum of these need not be equal to 100 because unresolved species trees are possible.

per species, 3 genes and 9 individuals per species, and 9 genes and 3 individuals per species. The theory of coalescence (Kingman 1982a, 1982b) was used to simulate gene trees according to these species trees using the neutral coalescent package (Maddison 2005) of Mesquite (Maddison and Maddison 2006). One hundred gene trees were simulated for each sampling scheme and for each species tree.

DNA sequences were simulated on the gene trees using the Genesis package of Mesquite. Sequences of 1000 characters were simulated for all individuals for each locus using an HKY85 + Γ substitution model with a transition/transversion ratio of 3; $\alpha = 0.8$ for the gamma distribution of rates across sites; equilibrium nucleotide state frequencies of 0.3 A, 0.2 C, 0.2 G, and 0.3 T; and a mutation rate of 3×10^{-8}. The same seeds for the random number generators were used for the different methods in simulations. These simulation settings are very similar to those used by Maddison and Knowles (2006) in their evaluation of the gene tree parsimony approach. Their settings were replicated here because they result in gene trees with a reasonable amount of incomplete lineage sorting and variability (see Results). Branch supports for each set of simulated gene trees were measured from 200 resampled matrices.

Phylogenetic Analyses

Gene tree analyses.—For the resampling methods BL, BC, and bootstrap lineages and characters (BLC), maximum parsimony was used to reconstruct the gene trees used for the gene tree parsimony search. Heuristic searches were performed in PAUP* (Swofford 2002) and

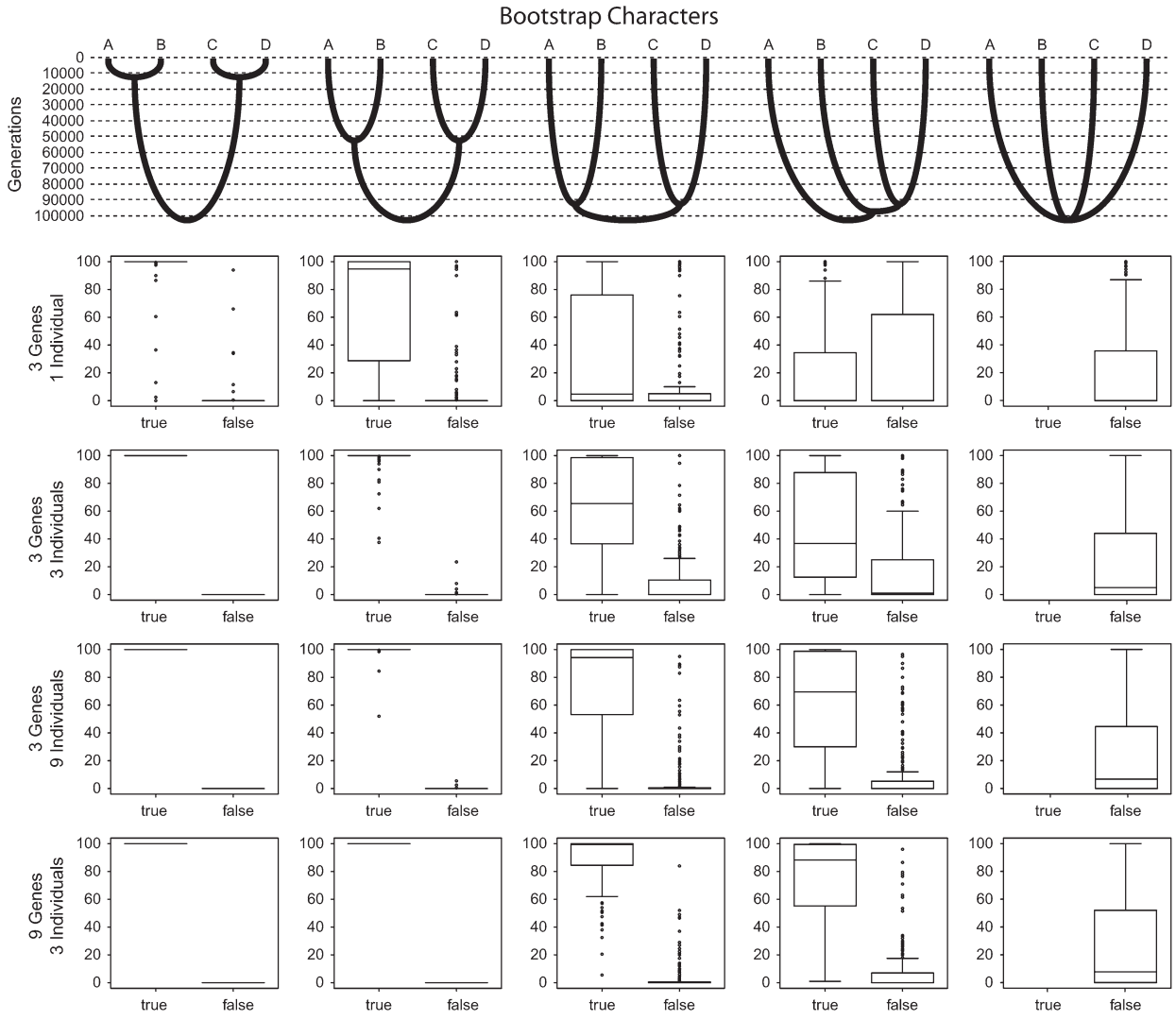

FIGURE 2. Support values obtained by BC within genes. Values are given for 5 different species trees and for 4 different sampling strategies: 3 genes and 1 individual sampled per species, 3 genes and 3 individuals, 3 genes and 9 individuals, and 9 genes and 3 individuals.

consisted of 1 replicate with "closest" addition sequence and TBR branch swapping, saving a maximum of 1000 most-parsimonious trees. For the posterior probability gene tree sampling (PPGTS) method, Bayesian analyses were performed in MrBayes (Version 3.1.2; Ronquist and Huelsenbeck 2003) to estimate the posterior distribution of gene trees for each resampled gene data set. Four chains of 200 000 generations were performed for each run with the GTR + I + Γ substitution model, discarding the first 20 000 generations as "burn-in" and then sampling trees every 100th generation. Given the small size and the nonsaturated signal of the data sets, these settings always gave good effective sample sizes and convergence (determined using Tracer Version 1.3; Rambaut and Drummond 2005).

Species tree analyses.—The species tree analyses were all performed with Mesquite, minimizing the number of deep coalescences over all loci. Deep coalescences were calculated by autoresolving polytomies in gene trees and by assuming that the gene trees were unrooted, which means that for each gene tree, the deep coalescence cost was the smallest count for any rooting of that gene tree within the proposed species tree. The search used an "as is" taxon addition sequence, sub-tree pruning and regrafting branch swapping, and saved all most-parsimonious trees. For the methods BL, BC, and BLC, strict consensus trees of all most-parsimonious trees were used for each gene in the species tree search (Maddison and Knowles 2006). For PPGTS, a fully bifurcating gene tree, sampled according to its posterior probability, was used for each gene.

To assess whether the decision of using parsimony as an optimization criteria and of using consensus trees for the species tree searches could have influenced the results obtained for the methods BL, BC, and BLC, some simulation scenarios were reanalyzed using 2 other settings. In the first, only one of the most parsimonious trees was randomly chosen per gene for the species tree searches instead of a consensus of all most parsimonious

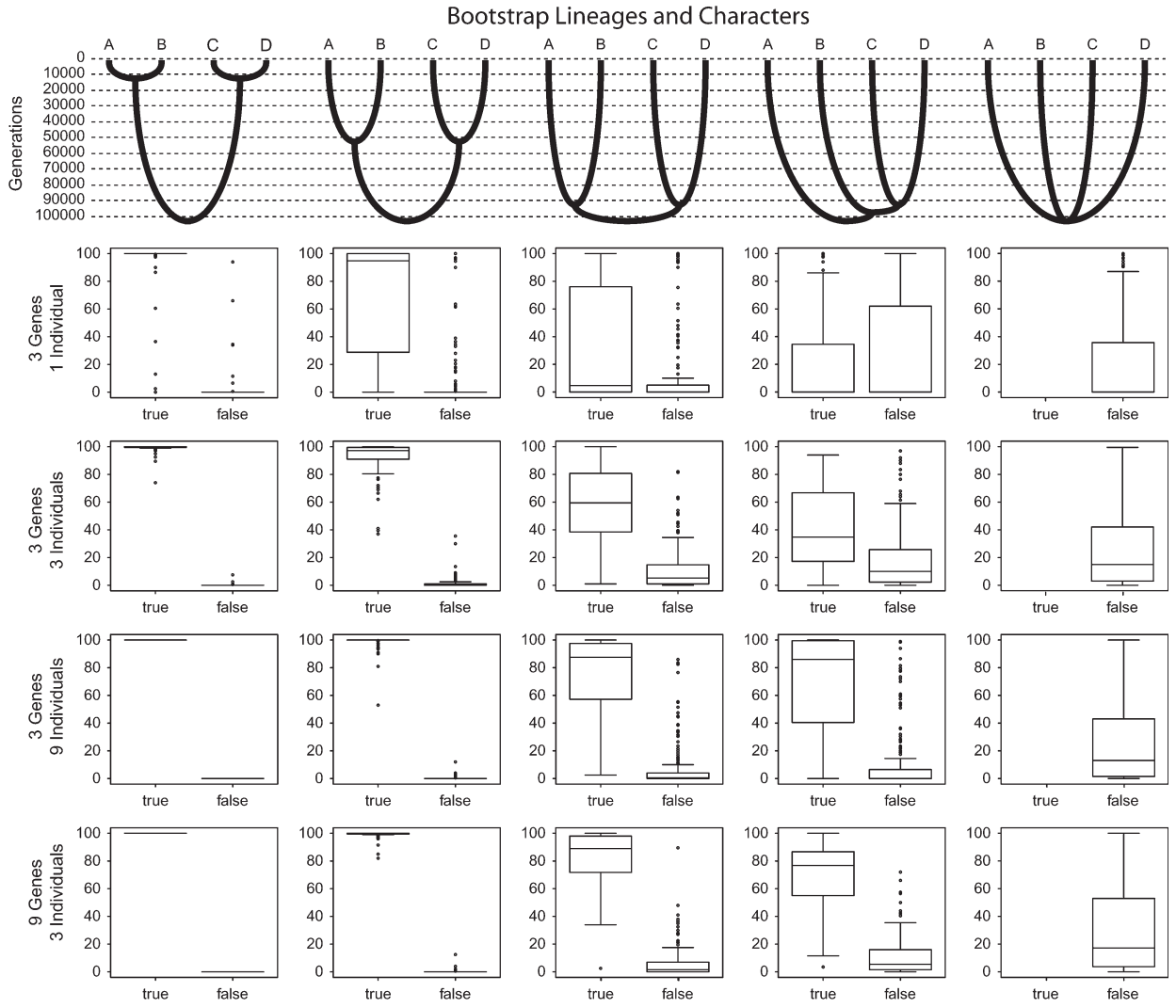

FIGURE 3. Support values obtained by BC within genes and lineages (individuals) within species. Values are given for 5 different species trees and for 4 different sampling strategies: 3 genes and 1 individual sampled per species, 3 genes and 3 individuals, 3 genes and 9 individuals, and 9 genes and 3 individuals. Note that when there is only 1 sequence per species, it is meaningless to BL, and in such instances, the results are identical to those obtained when characters alone are bootstrapped (Fig. 2).

trees. In the other, maximum likelihood, rather than parsimony, was used for reconstructing the gene trees, where the likelihood settings were the same as those used for simulating the sequences. Pearson correlation coefficients were calculated between the original support values and the ones obtained with the new settings for all 100 replicated data sets for the 5 scenarios with 3 genes and 3 individuals for BLC.

Comparing the Resampling Methods

For each resampling method, we represented the distribution of support values for the 100 simulated data sets for branches present (the true branches) or not (the false branches) in the species tree used for simulating the data. Because the star species tree did not contain any branch, all branches were considered to be false in that case. To further compare the different methods, we calculated the cutoff support value to give a Type I error rate of 5% on the star topology phylogeny, where no true branch exists. The Type II error was estimated on the misleading species tree topology and on the symmetric topology with short internal branch lengths using the threshold support value to obtain a Type I error of 5% on the star tree. The relationship of branch support to phylogenetic accuracy was also assessed (sensu Hillis and Bull 1993): branch support values were divided into 10 categories and the accuracy was determined for each.

Diploid Roses in North America

The wild diploid roses of section *Cinnamomeae* in North America represent a good group to further compare the 4 methods. Several individuals from the 6 species of this section have been sequenced for 3 single-copy nuclear genes (*GAPDH*, *TPI*, and *MS*; Joly and Bruneau

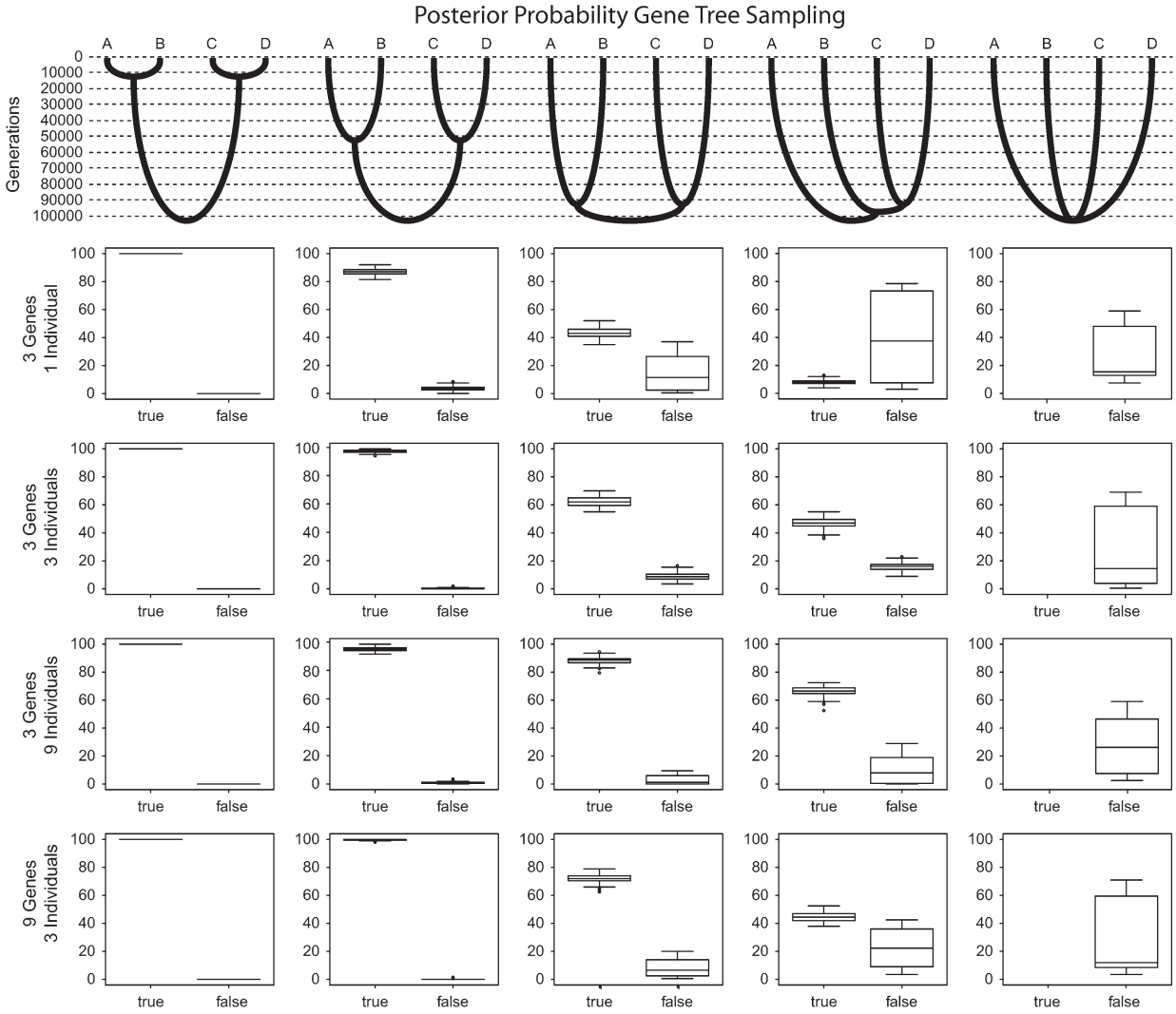

FIGURE 4. Support values obtained using a PPGTS. Values are given for 5 different species trees and for 4 different sampling strategies: 3 genes and 1 individual sampled per species, 3 genes and 3 individuals, 3 genes and 9 individuals, and 9 genes and 3 individuals.

2006). The use of species tree approaches is legitimate in this group, as species boundaries have been critically assessed for most species (Bruneau et al. 2005; Joly and Bruneau 2007). Two species of section *Synstylae*, also found in North America, were used as outgroup taxa. The data set used here is the same as that used by Joly and Bruneau (2006), except for the addition of 1 individual of *Rosa foliolosa*, 2 of *Rosa gymnocarpa*, 2 of *Rosa pisocarpa*, and 2 of *Rosa woodsii*. Sequences from these newly sampled individuals were obtained as previously described (Joly and Bruneau 2006; Joly et al. 2006) (GenBank accessions EU315075–EU315109). For all analyses, *R. woodsii* was treated as a synonym of *Rosa blanda* (Joly and Bruneau 2007). Information on the individuals sampled is given in Supplementary Appendix 1 (http://www.sysbio.oxfordjournals.org/). Gene sequences of homozygous individuals were duplicated in the data sets because these sequences are effectively present in 2 copies in these individuals and because gene copies present in individuals are considered to be a random sample of the allele pool of the previous generation. Parsimony gene tree analyses were performed in PAUP* using a heuristic search with TBR branch swapping and 10 random addition sequence replicates, saving a maximum of 1000 trees for each replicate. Indels were included in the analysis by recoding them as binary characters using the simple coding method (Simmons and Ochoterena 2000) implemented in GapCoder (Young and Healy 2003). The strict consensus trees of all most-parsimonious trees obtained for each gene were used in a gene tree parsimony search in Mesquite. Five independent searches were performed, each time adding species randomly to increase the chance of obtaining the global optima.

Branch supports were calculated for the different methods as described in the simulations. These calculations were performed with the help of the program Bottine (available from http://www.allanwilsoncentre.ac.nz/researchResources.htm), which generates Mesquite and PAUP* scripts that automate

the different steps for obtaining support values for the 4 resampling methods. Bayesian analyses of gene trees were performed in MrBayes (Version 3.1.2; Ronquist and Huelsenbeck 2003) using GTR + Γ_4+ I. A complex model was preferred to minimize the influence of the model on posterior distribution of gene trees (Lemmon and Moriarty 2004). Indels were not included in the Bayesian analyses because treating the recoded gap characters as restriction site characters in the analyses led to acceptance ratio and convergence problems. Two independent runs of 1×10^7 generations were performed with samples taken every 250 generations. The program Tracer was used to confirm appropriate mixing and convergence of the chains as well as the recovery of reasonable estimated sample sizes for all parameters. The first 200 000 generations were discarded as burn-in and the remaining samples from the 2 independent runs were combined prior to the resampling procedure.

RESULTS

Because species trees obtained by gene tree parsimony are rooted, support values can be evaluated from either rooted or unrooted trees. Because gene tree parsimony does not always position the root accurately with a deep coalescence criterion (Maddison and Knowles 2006), all results presented here are based on support values calculated on unrooted trees. This does not alter the conclusions reached here as support values from rooted trees showed the same trends as those from unrooted trees, although rooted trees resulted in lower average support for true branches and greater variance among replicates (data not shown).

Simulated Gene Trees

The simulation settings used here resulted in moderate amounts of incongruence among gene trees as evaluated by the deep coalescence score of the most parsimonious species trees reconstructed from the simulated data sets (Supplementary Table 1). In general, shorter terminal branches of the species tree and larger numbers of individuals sampled per species resulted in greater gene tree incongruence (Supplementary Table S1). The character matrices were moderately variable, containing from 1.7% to 5.8% variable characters on average (Supplementary Table 2). As expected, sampling more individuals per species resulted in greater variability. These levels of variability are similar to those observed in empirical data sets. With the variability obtained in simulated data sets, it is impossible to obtain strong bootstrap support for most of the branches in the gene trees. This is also evident when looking at the number of trees contained, on average, in 50%, 95%, and 100% credible sets in Bayesian analyses of the simulated data sets (Supplementary Table 3).

Comparison of Methods for Measuring Branch Support

Distribution of support values.—As expected, average support values for true branches decreased for all methods when going from the symmetric species tree with long internal branches to the star phylogeny for a given sampling scheme (Figs 1–4). The median support values obtained were similar across methods, but the distribution of support values differed. PPGTS had the narrowest distributions of support values overall, whereas the distribution of support values was similar among the other methods. This implies that PPGTS often showed a clear distinction between support values obtained for true and false branches (Fig. 4). A clear distinction between values obtained for true and false branches was not always evident for BL, BC, and BLC, especially in situations where the species trees had short internal edges. All methods also performed better with increasing sampling of either individuals or genes, and both strategies resulted in similar improvements. The importance of increasing the sampling effort was particularly evident for the "misleading" species tree topology. Indeed, whereas false branches received higher average support values than true branches when a single sequence was sampled per species, average support values for true branches were greater than those for false branches when at least 3 sequences were sampled (Figs 1–4), a trend observed with all methods. Overall, apart from the "misleading" topology scenario with 1 sequence sampled per species and apart from the star species topologies where there were no true branches, the average support for true branches was always greater than that for false branches.

The results suggest that the methods fall into 2 distinct groups (PPGTS vs. BL, BC, and BLC) that differ by several aspects. PPGTS resamples gene trees from a posterior distribution, uses a likelihood criterion for gene tree reconstruction and dichotomous gene trees in the species search. In contrast, BL, BC, and BLC resample lineages or characters before reconstructing the gene trees and use a parsimony criterion for gene tree reconstruction and consensus gene trees in the species search. To investigate whether using a parsimony criterion affected the results obtained with BL, BC, and BLC, 5 simulation schemes were also analyzed using a likelihood criterion. The branch support obtained from these new simulations was identical to the original ones ($r = 1$, data not shown). Similarly, the use of a single most-parsimonious tree rather than a consensus for each gene in the species tree search did not alter the results for the same 5 simulation schemes ($r = 1$, data not shown). This probably reflects the fact that very little homoplasy was present in the simulated data sets, a consequence of the simulation settings. Yet, these low levels of homoplasy are not unrealistic because the low levels of variation generally found in sequences from closely related species are unlikely to harbor much convergent and parallel evolution.

Type I and II error rates.—Type I error represents the probability of falsely rejecting the null hypothesis (of no relationship), or in the present case, of considering that a nonexisting branch is true. For each method, we

TABLE 1. Threshold support values (%) required for obtaining a Type I error rate (α) of 5% for 4 resampling methods for the star species phylogeny simulations

Sampling effort	BL	BC	BLC	PPGTS
3 genes | 3 individuals	92	98	85	64.5
3 genes | 9 individuals	98	94.5	89.5	51
9 genes | 3 individuals	93.5	100	93	66
Average value	94.5	97.5	89.2	60.5

Abbreviations: BL = bootstrap lineages; BC = bootstrap characters; BLC = bootstrap lineages and characters; PPGTS = posterior probability gene tree sampling.

calculated the cutoff support values below which the Type I error rate (α) would be greater than 5%. These were calculated on the star species phylogeny (Table 1), which is expected to be the most challenging scenario as there are no true relationships. The cutoff values for the PPGTS were lower than those for the other methods (Table 1). Indeed, in no simulations did PPGTS give support >71% for false branches (Fig. 4). BLC had the second lowest Type I errors on average, followed by BL and BC (Table 1).

By itself, Type I error is not sufficient for comparing the different methods because a method with a high Type I error might nevertheless have a lower Type II error (e.g., the probability of accepting the null hypothesis of no relationship when it is false). The Type II error of each method was thus calculated for a Type I error of 5% as calculated on the star species phylogeny to assess the power (i.e., 1 minus the Type II error) of each method for a given Type I error (Table 2). Although PPGTS had high Type II error rates under the misleading topology, it performed much better than the other methods under the symmetric tree topology with short branches (Table 2). Of the remaining methods, BLC had the lowest Type II error overall, followed by BL and then BC. Clearly, the high Type II error of BL and BC is a consequence of their high cutoff values required to have a Type I error of 5% (Table 1).

Accuracy.—Accuracy followed similar trends for the different methods (Fig. 5), apart from PPGTS, which generally showed a sigmoid curve shape with an overestimation of accuracy for low support values and an underestimation of accuracy for support values above approximately 40%. For the other methods, support values generally gave a reasonable estimate of phylogenetic accuracy, with perhaps a small underestimation of accuracy overall. Yet, the small sample sizes available do not allow us to reach strong conclusions. Although this trend must be generalized and confirmed with additional simulations that cover a wider portion of the parameter space, it does suggest that branch support of species trees does not systematically underestimate phylogenetic accuracy, as previously suggested for gene trees (Hillis and Bull 1993). Yet, because the correspondence between support values and accuracy is not linear and appears to vary for different simulation settings (Fig. 5), one should be cautious when interpreting support values as a measure of accuracy.

TABLE 2. Type II error rates (power) obtained when applying a threshold support value that implies a Type I error rate of 5% for the star species phylogeny simulations (Table 1)

Species tree	Sampling effort	BL	BC	BLC	PPGTS
Misleading	3 genes | 3 individuals	98/96	100/85	99/94	100/100
	3 genes | 9 individuals	85/81	100/72	56/51	35/7
	9 genes | 3 individuals	97/92	100/66	83/77	100/100
Short internal branches	3 genes | 3 individuals	97/92	100/73	92/90	88/38
	3 genes | 9 individuals	73/62	100/57	63/54	0/0
	9 genes | 3 individuals	78/66	100/41	62/52	7/0

Notes: Because different threshold values were obtained for the different sampling efforts, Type II error rates are given for the maximum and the average threshold values following the format: maximum/average. Abbreviations: BL = bootstrap lineages; BC = bootstrap characters; BLC = bootstrap lineages and characters; PPGTS = posterior probability gene tree sampling.

Phylogeny of Diploid North American Roses

The number of individuals included per species for the North American roses' data set is reported in Table 3. The number of lineages sampled per species equals twice the number of individuals, because all species are diploids. More detailed information regarding the data, such as homozygous and heterozygous individuals for each gene, is given in Supplementary Appendix 1. Two most-parsimonious trees were obtained with a deep coalescence score of 53 and another topology had a score of 54 (Fig. 6). All other species tree topologies had scores $\geq$ 57. The 3 best trees showed that species of section *Cinnamomeae* in North America are resolved as monophyletic relative to species of section *Synstylae*, a relationship strongly supported (100%) by all methods. Another consistent pattern among these trees is the sister relationship of *Rosa nitida* and *Rosa palustris* that received considerable support (Fig. 6). Finally, the 3 best species trees suggested that *R. gymnocarpa* is sister to all remaining North American diploid species. The remaining clades received very little support and suggest that the present data do not provide enough information to resolve the phylogenetic position of *R. blanda*, *R. foliolosa*, and *R. pisocarpa*.

DISCUSSION

Comparison of the Methods

Although many approaches can be considered for measuring branch support for species trees obtained by gene tree parsimony, only 4 were evaluated here. These methods were selected either to determine whether common approaches used for gene tree analyses are also appropriate for measuring branch support on species phylogenies, or because they were the ones that are expected to be the most appropriate for empirical data sets that usually consist of only a few genes. For instance, gene tree bootstrapping (Burleigh et al. 2006) was not considered here because, when only a small number of genes are sampled, the number of possible combinations is small (for 3 data sets, there are only 9

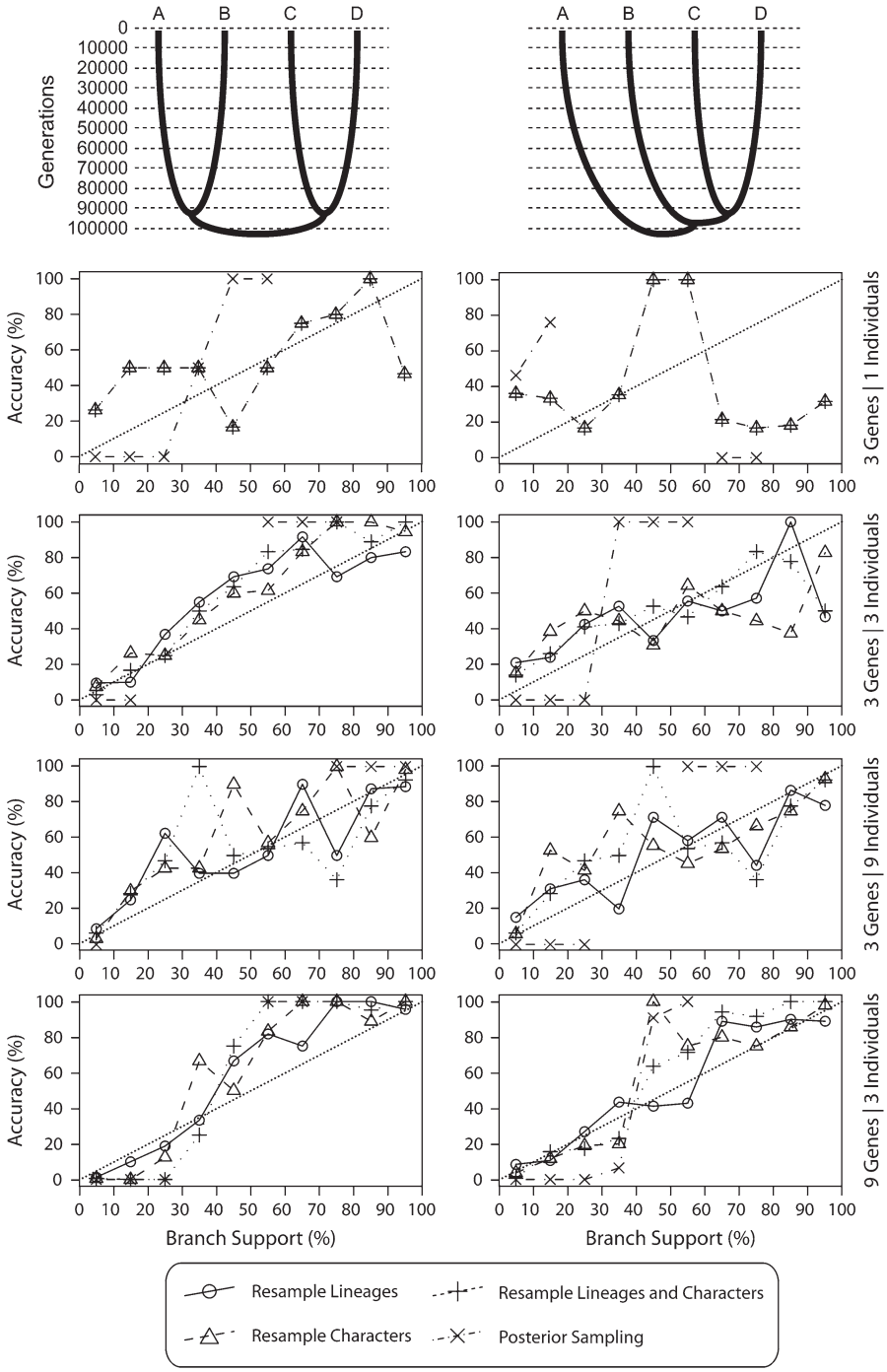

FIGURE 5. Phylogenetic accuracy obtained for different ranges of support values for the 4 resampling methods tested. Note that BL is not plotted where there is a single sequence per species, as this resampling method becomes meaningless in these situations.

possible combinations) and its efficiency is expected to be limited.

The results showed that among the bootstrap procedures, bootstrapping both lineages and characters (BLC) together resulted in lower Type I and II error than bootstrapping either only lineages (BL) or only characters (BC). If we consider species tree analyses as a 2-step procedure (estimation of gene trees followed by the species tree reconstruction), the results suggest that assessing the uncertainty at both levels of analysis gives better results than taking only one of these into account. It also further reinforces the difference between gene tree and species tree analyses and suggests that the standard nonparametric bootstrap is not a very effective method for estimating branch support of species trees derived from multiple gene trees. Although these results were obtained in a gene tree parsimony framework, the importance of accounting for the uncertainty

TABLE 3. Number of individuals sampled per species for the phylogenetic analysis of North American roses

Species	Number of individuals
Rosa blanda Ait.	22
Rosa foliolosa Nutt.	3
Rosa gymnocarpa Nutt. ex Torr. & A. Gray	5
Rosa multiflora Thunb.	1
Rosa nitida Willd.	4
Rosa palustris Marsh.	6
Rosa pisocarpa A. Gray	4
Rosa setigera Michx	1

present at both levels of analysis is likely to hold with other methods of species tree reconstruction. Moreover, although gene tree bootstrapping was not evaluated here, it might nevertheless represent an approach to consider when more genes are included, as it assesses another source of uncertainty present in species tree analyses.

Compared with the other methods, PPGTS had interesting properties, such as a low support value threshold for a 5% Type I error and narrow distributions that result in values that do not overlap between true and false branches. This performance is surprising given that PPGTS only assesses the uncertainty at the gene tree level of analysis. It is possible that accounting for the uncertainty involved in intraspecific sampling in such a Bayesian framework would give even better results, but this was not investigated. The contrasting results between PPGTS and the standard nonparametric bootstrap (BC), which had the highest Type I and II error, further highlights the philosophical differences between these 2 commonly used methods for estimating branch support of gene trees (Alfaro et al. 2003; Cummings et al. 2003; Douady et al. 2003; Erixon et al. 2003). Although PPGTS appears to be a good method for estimating branch support in species tree analyses, the same cannot be said of BC, at least with the present simulation settings.

Interpretation of support values in empirical studies.—Despite difficulties in comparing results obtained from simulations with those from empirical studies, the threshold values for obtaining a Type I error rate of 5% on the star phylogeny simulations give useful guidelines in interpreting results from real data sets. Whenever a support value above such a threshold is obtained for a branch, one might conclude that it is well supported by the data. Again, it is probably dangerous to interpret this as an indication of accuracy. The simulations suggested that these thresholds were 98% for BL, 100% for BC, 93% for BLC, and 66% for PPGTS. These values are conservative, as they assume a star topology, an improbable pattern in nature. In fact, considerably lower support values were obtained for false branches with bifurcating species trees, even when the species topology was challenging, such as with the asymmetric tree with short internal edges. Yet, because of the simplicity of the simulations performed here (small range of parameters covered, no population structure within species, etc.), these values should be interpreted cautiously. Furthermore, undetected hybridization or the presence of other evolutionary events that are not modeled by gene tree parsimony and that could be present in empirical data sets could invalidate these results.

Application of methods to an empirical data set.—Because simulation studies do not capture the full complexity present in empirical data sets, it is useful to compare the behavior of the methods using biological examples. The analysis of the eastern North American roses shows that the relationships present in a strict consensus of all most parsimonious trees (not shown) received considerable support (>50%). In contrast, the clades not supported by the strict consensus tree received very little support from all methods, suggesting that the data do not strongly support these relationships. All methods gave very similar support values for all groups, except for the group that consists of eastern *R. blanda*, *R. foliolosa*, *R. nitida*, and *R. palustris*, which received >79% support from BL, BC, and BLC but only 22% support from PPGTS. To a certain extent, this might be caused by differences in the way methods are implemented. Yet, the simulations performed here showed that the use of consensus trees or the choice of a parsimony criterion for reconstructing gene trees cannot account for the discrepancies obtained between PPGTS and the other methods. Clearly, the differences observed between the results from simulations and those from the rose example also illustrate the need for evaluating the resampling methods using more complex simulation schemes and more empirical data sets in order to better understand their behaviors. Nevertheless, this empirical example shows that support values obtained with all resampling methods are useful for interpreting results from species tree searches.

Limits to the BL approach.—The results suggest that BL is a worthwhile approach to consider for evaluating branch support of species phylogenies. But this might not always be the case. With the current simulation settings, the terminal branches of the species trees were relatively short in coalescent units. This means that several sequences will not coalesce within the species lineage and that coalescent events will occur along internal branches of the species tree, leading to incongruent gene trees. When this happens, sampling more alleles per species increases dramatically the chance of sampling more ancestral lineages, that is, lineages that were present at species formation. Because different ancestral lineages represent independent assessments of species relationships (Maddison and Knowles 2006), sampling more ancestral lineages is likely to result in better phylogenetic accuracy. This explains the relevance of BL within species because different samples of lineages may support different species trees. It also explains why sampling more sequences (individuals) per species increased support values for true branches. Indeed, the results showed that for an equivalent amount of data (in terms of number of nucleotides) and with the present simulation settings, increasing the number of individuals yielded improvements in support values

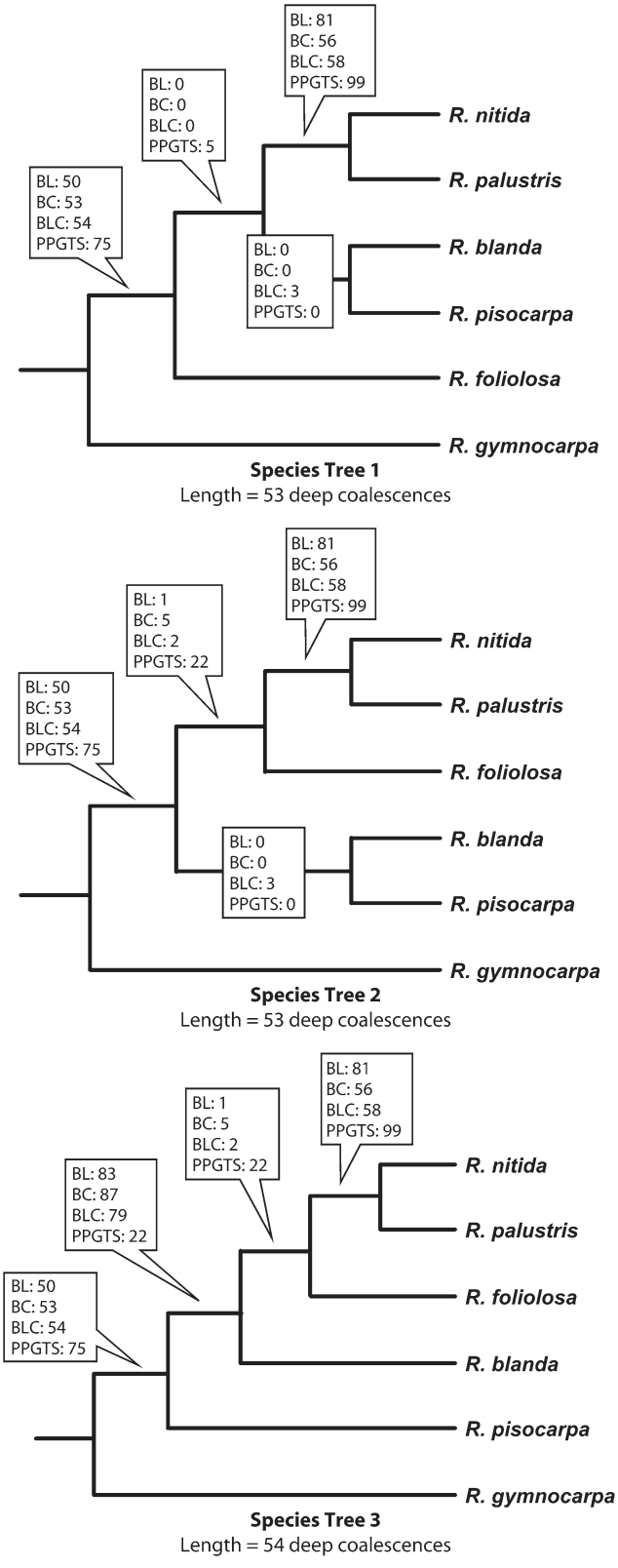

FIGURE 6. Representation of the 3 most-parsimonious species trees obtained by gene tree parsimony showing branch supports obtained with all 4 resampling methods for diploid eastern North American roses. Only the clade with ingroup species is shown; this clade received 100% support from all methods.

similar to those obtained when increasing the number of loci sampled (for further discussion on the relative advantage of sampling more alleles or more genes, see Rosenberg [2002] and Maddison and Knowles [2006]). In contrast, when the terminal branches of the species tree are very long (>>1 coalescent units), species are expected to be monophyletic at most loci. In such situations, sampling more lineages will not provide more information for species tree reconstruction because it will not increase the number of ancestral lineages sampled. Consequently, BL in these situations is not expected to be an effective approach for evaluating branch support of species trees.

One easy way to determine if it is pertinent to BL for a given data set is simply to look at the gene trees. If species in gene trees are not reciprocally monophyletic, and therefore show several instances of incomplete lineage sorting, then the uncertainty involved in sampling lineages is likely to be important and should be taken into account when assessing branch support. Of course, to be able to detect nonmonophyletic species, more than 1 individual needs to be sequenced. We therefore strongly recommend sampling multiple individuals per species in empirical studies that aim at inferring species relationships, as this tends to give not only higher support values but also more accurate species trees (Maddison and Knowles 2006).

The BL procedure may have limited efficiency when few sequences are available per species, as it occurs for gene tree bootstrapping when there are only a few gene trees. Indeed, bootstrap methods are not expected to be very effective when sample sizes are small (Good and Hardin 2006). However, this problem becomes less important if the complete gene tree is considered instead of the species itself. For example, in the current simulation procedure with 3 alleles sampled per species, the number of possible combinations of lineages for 1 species is 9 but that for 1 gene with 4 species is 9^4. And if 3 genes are sampled, the total number of possible combinations of lineages is $(9^4)^3 = 2.8 \times 10^{11}$. Therefore, even with few lineages sampled per species, the BL approach should not be affected strongly by small sample sizes for moderate numbers of species and genes sampled.

Support in the Anomaly Zone

Degnan and Rosenberg (2006) showed that for some species tree topologies, the most likely gene tree does not reflect the species tree for some combination of branch lengths of the species tree (the anomaly zone). This problem occurs because in a population, symmetric gene trees are more likely than asymmetric ones because they can be obtained in more ways. For example, in a population of 4 sequences (a,b,c,d), a gene tree ((a,b),(c,d)) can be obtained if (a,b) coalesce first and (c,d) second and vice versa. In contrast, there is only 1 combination of coalescent events that could result in the gene tree (((a,b),c),d). Therefore, in such a population, the gene tree ((a,b),(c,d)) is twice as likely to be observed than gene tree (((a,b),c),d). If internal branches are very short on a species tree, the probability that lineages coalesce along them is small and consequently most coalescent events will occur along branches that are ancestral to many species, which will make symmetric gene trees more likely than asymmetric ones. As such, if the species tree is asymmetric, there are some combinations of branch lengths that result in most likely gene trees that are incongruent with the species tree.

This has important implications for species tree reconstruction, as it means that branches found most often in an infinite number of gene trees might not represent the relationships of the true species tree. Until now, the anomaly zone of species trees has been characterized only for 4 (Degnan and Rosenberg 2006) and 5 (Rosenberg and Tao 2008) species, and only with a single sequence per species. If more alleles are sampled per species, the probability of a coalescent event along the branches of the species tree increases importantly, even along short internal branches. This means that with additional samples per species, the anomaly zone will be reduced for a given species tree. The results presented here support this because although, on the misleading topology with 1 sequence sampled per species, the support for false branches was greater than that for true branches, when 3 or more sequences were obtained per species, true branches were better supported. This suggests that with the "misleading" species tree scenario used here, sampling at least 3 sequences per species might be sufficient for this topology to escape the anomaly zone. However, the beneficial effect of increasing the number of individuals sampled within species will decrease as the time between speciation events increases. As discussed above, with longer terminal branches, there is more time for species to become monophyletic at most genes, and therefore, increasing the number of sequences sampled will not increase the sampling of ancestral lineages. This also implies that ancient rapid radiations are likely to be more difficult to recover than recent ones.

Using unrooted gene trees in the species tree search, as done here, necessarily reduces the difficulty of recovering the true species tree. Taking the misleading scenario as an example, using unrooted gene trees implies that gene tree (a,(b,(c,d))), which is congruent with the species tree, is identical with other gene trees such as ((a,b),(c,d)) that are incongruent with the species tree. Consequently, fewer gene trees are incongruent with the species tree. This tree search setting was chosen to mimic empirical studies in which it is often difficult to accurately position the root of a gene tree (Castelloe and Templeton 1994). Nevertheless, with the misleading scenario, the gene trees ((a,c),(b,d)) and ((a,d),(b,c)) remain incongruent with the congruent gene tree (a,(b,(c,d))) and are more likely than the latter. Thus, the asymmetric species tree used here remains misleading even when unrooted gene trees are used in the species tree search.

The present results show that it will be important to further characterize the anomaly zones for species trees

with more than 1 sample per species to evaluate to what extent the anomaly zone can be reduced. This has important repercussions for understanding our ability to recover species trees in empirical studies.

Phylogeny of Diploid Roses in North America

The species tree analysis of the North American diploid roses confirmed certain relationships found in previous analyses but also highlighted the ambiguous position of some species that require further study in the future. An example of the former is the sister relationship between *R. nitida* and *R. palustris* that received moderate to strong support from all 4 methods, a relationship that was also found in previous studies (Joly and Bruneau 2006, 2007; Joly et al. 2006). Clearly, this relationship is unlikely to be challenged in the future. Another consistent finding is the sister position of *R. gymnocarpa* to the other North American species (Joly and Bruneau 2006). In contrast, the phylogenetic positions of *R. blanda*, *R. foliolosa*, and *R. pisocarpa* are a source of confusion.

Because most species tree approaches have been developed recently, additional studies are needed to improve and test them and to determine the most appropriate methods for measuring the amount of uncertainty surrounding the species tree inference. Although full Bayesian methods were previously available for measuring branch support of species trees, this study provides useful guidelines for the development of effective methods applicable in a parsimony or likelihood framework. The present study is far from being comprehensive for evaluating the accuracy of species trees, but we hope that it will stimulate further studies to improve the quality of phylogenies obtained with species tree approaches.

Supplementary Material

Supplementary material can be found at http://www.oxfordjournals.org/our journals/sysbio/.

Funding

This work has been possible by a Natural Sciences and Engineering Research Council (NSERC) postdoctoral fellowship to S.J. and an NSERC grant to A.B.

Acknowledgments

The authors thank Wayne Maddison for making modifications to Mesquite that facilitated the simulations performed in this study, and Lacey Knowles, Karen Cranston, and an anonymous reviewer for useful comments on a previous version of the manuscript. The authors also thank Sugir Selliah for her help in the laboratory, and colleagues of the Allan Wilson Centre, who kindly accepted to have us run simulations on their computers.

References

Alfaro M.E., Zoller S., Lutzoni F. 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. Mol. Biol. Evol. 20:255–266.

Bruneau A., Joly S., Starr J.R., Drouin J.-N. 2005. Molecular markers indicate that the narrow Québec endemics *Rosa rousseauiorum* and *Rosa williamsii* are synonymous with the widespread *Rosa blanda*. Can. J. Bot. 83:386–398.

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

Burleigh J.G., Driskell A.C., Sanderson M.J. 2006. Supertree bootstrapping methods for assessing phylogenetic variation among genes in genome-scale data sets. Syst. Biol. 55:426–440.

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

Castelloe J., Templeton A.R. 1994. Root probabilities for intraspecific gene trees under neutral coalescent theory. Mol. Phylogenet. Evol. 3:102–113.

Cummings M.P., Handley S.A., Myers D.S., Reed D.L., Rokas A., Winka K. 2003. Comparing bootstrap and posterior probability values in the four-taxon case. Syst. Biol. 52:477–487.

Degnan J.H., Rosenberg N.A. 2006. Discordance of species trees with their most likely gene trees. PLoS. Genetics. 2:e68.

Douady C.J., Delsuc F., Boucher Y., Doolittle F.W., Douzery E.J.P. 2003. Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. Mol. Biol. Evol. 20:248–254.

Doyle J.J. 1992. Gene trees and species trees: molecular systematics as one-character taxonomy. Syst. Bot. 17:144–163.

Erixon P., Svennblad B., Britton T., Oxelman B. 2003. Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. Syst. Biol. 52:665–673.

Good P.I., Hardin J.W. 2006. Common errors in statistics (and how to avoid them). 2nd ed. Hoboken (NJ): Wiley.

Hillis D.M., Bull J.J. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42:182–192.

Hudson R.R. 1992. Gene trees, species trees and the segregation of ancestral alleles. Genetics. 131:509–512.

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

Joly S., Bruneau A. 2006. Incorporating allelic variation for reconstructing the evolutionary history of organisms from multiple genes: an example from *Rosa* in North America. Syst. Biol. 55:623–636.

Joly S., Bruneau A. 2007. Delimiting species boundaries in *Rosa* sect. *Cinnamomeae* (Rosaceae) in Eastern North America. Syst. Bot. 32:819–836.

Joly S., Starr J.R., Lewis W.H., Bruneau A. 2006. Polyploid and hybrid evolution in roses east of the Rocky Mountains. Am. J. Bot. 93:412–425.

Kingman J.F.C. 1982a. The coalescent. Stochastic process. Appl. 13:235–248.

Kingman J.F.C. 1982b. On the genealogy of large populations. J. Appl. Prob. A 19:27–43.

Lemmon A.R., Moriarty E.C. 2004. The importance of proper model assumption in Bayesian phylogenetics. Syst. Biol. 53:265–277.

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.

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

Maddison W.P. 2005. Coalescence package for Mesquite. Version 1.06. Available from: http://mesquiteproject.org.

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

Maddison W.P., Maddison D.R. 2006. Mesquite: a modular system for evolutionary analysis. Version 1.12. Available from: http://mesquiteproject.org.

Nichols R. 2001. Gene trees and species trees are not the same. Trends. Ecol. Evol. 16:358–364.

Page R.D.M. 2001. GeneTree. Version 1.3.0. Glasgow (UK): University of Glasgow. Available from: http://taxonomy.zoology.gla.ac.uk/rod/genetree/genetree.html.

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.

Rambaut A., Drummond A.J. 2005. Tracer. Version 1.3. Oxford (UK): University of Oxford. Available from: http://evolve.zoo.ox.ac.uk/software/tracer/.

Ronquist F., Huelsenbeck J.P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19: 1572–1574.

Rosenberg N.A. 2002. The probability of topological concordance of gene trees and species trees. Theor. Popul. Biol. 61:225–247.

Rosenberg N.A., Nordborg M. 2002. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. Nat. Rev. 3:380–390.

Rosenberg N.A., Tao R. 2008. Discordance of species trees with their most likely gene trees: the case of five taxa. Syst. Biol. 57:131–140.

Seehausen O. 2004. Hybridization and adaptive radiation. Trends. Ecol. Evol. 19:198–207.

Simmons M.P., Ochoterena H. 2000. Gaps as characters in sequence-based phylogenetic analyses. Syst. Biol. 49:369–381.

Slowinski J.B., Knight A., Rooney A.P. 1997. Inferring species trees from gene trees: a phylogenetic analysis of the Elapidae (Serpentes) based on the amino acid sequences of venom proteins. Mol. Phylogenet. Evol. 8:349–362.

Swofford D.L. 2002. PAUP*: Phylogenetic analysis using parsimony (*and other methods). Version 4.0b10. Sunderland (MA): Sinauer Associates.

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

Wu C.-I. 1991. Inferences of species phylogeny in relation to segregation of ancient polymorphisms. Genetics. 127:429–435.

Young N.D., Healy J. 2003. GapCoder automates the use of indel characters in phylogenetic analysis. BMC. Bioinformatics. 4:6.

First submitted 18 April 2008; reviews returned 18 July 2008; final acceptance 5 January 2009

Associate Editor: L. Lacey Knowles