Syst. Biol. 54(3):363–372, 2005

ISSN: 1063-5157 print / 1076-836X online
DOI: 10.1080/10635150590945377

Evaluating Intraspecific "Network" Construction Methods Using Simulated Sequence Data: Do Existing Algorithms Outperform the Global Maximum Parsimony Approach?

INSA CASSENS, PATRICK MARDULYN, AND MICHEL C. MILINKOVITCH

Laboratory of Evolutionary Genetics, Institute of Molecular Biology and Medicine, Université Libre de Bruxelles, rue Jeener et Brachet 12, 6041 Gosselies, Belgium; E-mail: pmarduly@ulb.ac.be (P.M.)

Abstract.—In intraspecific studies, reticulated graphs are valuable tools for visualization, within a single figure, of alternative genealogical pathways among haplotypes. As available software packages implementing the global maximum parsimony (MP) approach only give the possibility to merge resulting topologies into less-resolved consensus trees, MP has often been neglected as an alternative approach to purely algorithmic (i.e., methods defined solely on the basis of an algorithm) "network" construction methods. Here, we propose to search tree space using the MP criterion and present a new algorithm for uniting all equally most parsimonious trees into a single (possibly reticulated) graph. Using simulated sequence data, we compare our method with three purely algorithmic and widely used graph construction approaches (minimum-spanning network, statistical parsimony, and median-joining network). We demonstrate that the combination of MP trees into a single graph provides a good estimate of the true genealogy. Moreover, our analyses indicate that, when internal node haplotypes are not sampled, the median-joining and MP methods provide the best estimate of the true genealogy whereas the minimum-spanning algorithm shows very poor performances. [Intraspecific genealogy; maximum parsimony; median-joining; minimum spanning; network; reticulated graph; statistical parsimony; simulated sequence data.]

Most methods available for (leaf-labeled) tree construction have initially been developed for phylogeny estimation among well-differentiated species. However, trees (i.e., following graph theory, "connected graphs with no circuits") are valid means for portraying genealogical relationships both above and below the species level. For example, as introduced by Avise and colleagues (1987), phylogeography explores the relationship between intraspecific trees (genealogies) and the geographical distribution of haplotypes. Furthermore, it must be emphasized that characteristics specific to population data sets are not automatically in conflict with a general treelike genealogy (but see Posada and Crandall, 2001, for an opposing point of view). For instance, orthologous nonrecombining pieces of DNA form a strictly nonreticulated hierarchical set of relationships across generations. Trees therefore are valid representations of these relationships. Additionally, both multifurcations and ancestral haplotypes (often observed, together with more recently diverged lineages, in intraspecific studies) can easily be represented in a tree by collapsing zero-length branches (e.g., as implemented in PAUP *4.0b10 [Swofford, 2003]). Therefore, a genealogy connecting intraspecific haplotypes can be displayed using different graphical representations (all being "trees") including a cladogram, a phylogram, or a haplotypic tree (Fig. 1). On the contrary, trees are inappropriate graphs when instances of reticulate evolution, such as recombination or horizontal gene transfer, come into play. Then, a less-restrictive graph with cycles ("loops"), rather than a tree, better represents the reticulated relationships among genes.

In the evolutionary genetics literature, reticulated graphs are often called "networks." However, the network is the actual genealogy (with all its complexity), and we prefer to use the term "graph" as the symbolic representation of the genealogy (exactly as a tree is the representation of the phylogeny connectivity and possibly of a few other parameters, such as branch lengths). More importantly, cycles in the reticulated graphs generally reveal ambiguities such that the unreticulated true genealogy is contained within the reticulated graph. The ambiguities are due to homoplasious character changes, and the loops in the reticulated graph indicate alternative genealogical pathways. On the other hand, a strict consensus tree is a (generally polytomic) tree compatible with all most parsimonious (MP) trees (only clades present in all MP trees are included). It represents a very conservative (i.e., less resolved) estimate of the genealogical relationships among the closely related sequences under study. Therefore, reticulated graphs are useful tools as they can convey more information (especially at the population level) than a strict consensus tree.

In the last decade, methods dedicated to the estimation of intraspecific genealogies (with or without reticulations) have been developed and some of these approaches are widely used in population genetic studies (Brant and Orti, 2003; Chenoweth and Hughes, 2003; Collevatti et al., 2003; Contreras-Diaz et al., 2003; Drew et al., 2003; Lawton-Rauh et al., 2003; Lloyd, 2003; Michaux et al., 2003; Printzen et al., 2003). Notwithstanding the diversity of methods that have been described in a recent review (Posada and Crandall, 2001), our recent empirical study (Cassens et al., 2003) demonstrated the need for analyzing the reliability, accuracy, and limitations of different algorithms. In this first comparative analysis, we inferred the genealogical relationships among 36 mitochondrial cytochrome *b* sequences using four different "network" approaches and could show substantial differences among the resulting graph topologies. Given that scientific hypotheses may strongly depend on the topology of the inferred genealogy, extensive comparative analyses are warranted to better

Insa Cassens and Patrick Mardulyn contributed equally to this study.

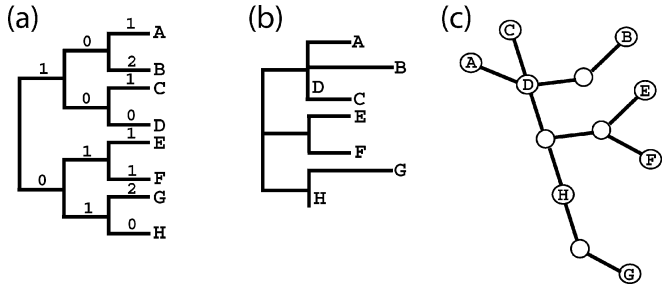


FIGURE 1. Three different graphical representations of the same evolutionary relationships among intraspecific haplotypes labeled A to H. (a) A cladogram shows the branching order of nodes, but branch lengths are not represented. The observed numbers of changes that are indicated above the branches have been added for a better comparison with the phylogram. (b) A phylogram shows branching order of nodes as well as branch lengths (proportional to the number of mutations that have occurred). (c) A haplotypic tree in which missing intermediate haplotypes are represented as open circles. Here, we distinguish "tip haplotypes" such as A or F, connected to the tree by a single branch, "node haplotypes" with three or more connections, and "branch haplotypes" with two connections. "Node haplotypes" and "branch haplotypes" can be sampled or missing (e.g., E and F are connected to a missing node haplotype, whereas H is a sampled branch haplotype).

understand the assumptions underlying some of the existing methods and to test whether the differences can be generalized as indicative of systematic artifacts.

Most of the available intraspecific "network" construction methods are algorithmic (sensu Swofford et al., 1996; i.e., a method defined solely on the basis of an algorithm). Although the use of an optimality criterion is sometimes made—for example, for the definition of the "parsimony limit" (cf. the statistical parsimony approach implemented in TCS [Clement et al., 2000]) or for the inference of node haplotypes to reduce the overall length of a constructed graph (cf. the median-joining network method (MJN) implemented in NETWORK [Bandelt et al., 1999])—that optimality function is not used for searching the space of possible topologies. For three main reasons, the global maximum parsimony has sometimes been considered inappropriate for inferring intraspecific gene genealogies (Posada and Crandall, 2001). First, the low levels of genetic variation found in population data can lead to an enormous number of equally most parsimonious trees. Second, the sometimes very large number of haplotypes in intraspecific studies can render even heuristic parsimony searches computationally impractical. And third, the global maximum parsimony approach only provides the possibility to merge all constructed trees into a less resolved strict consensus tree. We tend to disagree with the latter point, as this restriction is not inherent to the parsimony approach but simply due to the fact that software packages do not implement the possibility to combine trees into a reticulated graph.

In this study, using computer simulations, we report on a comparative analysis of different graph construction methods. Evolution of sequence data was simulated covering a range of intraspecific tree topologies and branch lengths. Besides using approaches both implemented in freely available software packages and widely used in population genetic studies (minimum-spanning network (MSN)/ARLEQUIN, statistical parsimony/TCS, and median-joining/NETWORK), we also developed and applied an algorithm for constructing a (possibly reticulated) graph that contains all equally parsimonious trees. Topologies produced by the four methods are then compared focusing on their compatibility with the true genealogy and level of ambiguity (the latter being represented by the number of loops). In particular, we tested whether the MP method is less suited than the algorithmic "network" construction methods, as sometimes suggested in the literature (e.g., Posada and Crandall, 2001), for inferring relationships at the intraspecific level.

MATERIAL AND METHODS

Simulation of Data

Using the program SEQ-GEN (Rambaut and Grassly, 1997), we generated 100 sequence data sets with up to 36 sampled haplotypes along each of the four different template genealogies portrayed in Figure 2. For 100 variable nucleotides, sequence evolution was simulated using the Hasegawa, Kishino, and Yano nucleotide substitution model (Hasegawa et al., 1985) with a transition/transversion ratio of 4 and a high frequency for A-T nucleotides (A:0.35, G:0.15, C:0.15, T:0.35); these conditions reasonably reflect some animal mitochondrial DNA sequences. Note that, given the stochastic substitution process, mutations are distributed along the branches of the template genealogy with a probability proportional to the length of each branch. The exact process that generated the possibly unique simulated data set yielded realized branch lengths that can be larger or smaller than the branch lengths of the template genealogy. When no mutation occurs along a branch of the template genealogy, the two haplotypes on either side of that zero-length branch are, by definition, identical in the realized genealogy. The template genealogies were chosen such that they present features, as identified in a previous study (Cassens et al., 2003), making them difficult to reconstruct. More specifically, they were selected for the following characteristics. Tree a (Fig. 2): all 36 haplotypes are sampled and connected to a maximum of four other haplotypes by single-step branches (the probability of occurrence of one mutation along a one-step branch is 0.01 for each site). This tree should be easy to reconstruct by most methods and is used as a reference for comparison to more complex topologies. Tree b (Fig. 2) is very similar to tree a: it exhibits the same topology among 36 sampled haplotypes but branch length is increased for many of the connections. This will allow testing the impact of long branches on genealogy reconstruction. Tree c (Fig. 2) represents a "starlike genealogy" with eight sampled tip haplotypes connected to a single central sampled haplotype through relatively long (two- and four-step) branches. The starlike pattern is often encountered in empirical intraspecific studies and the relatively

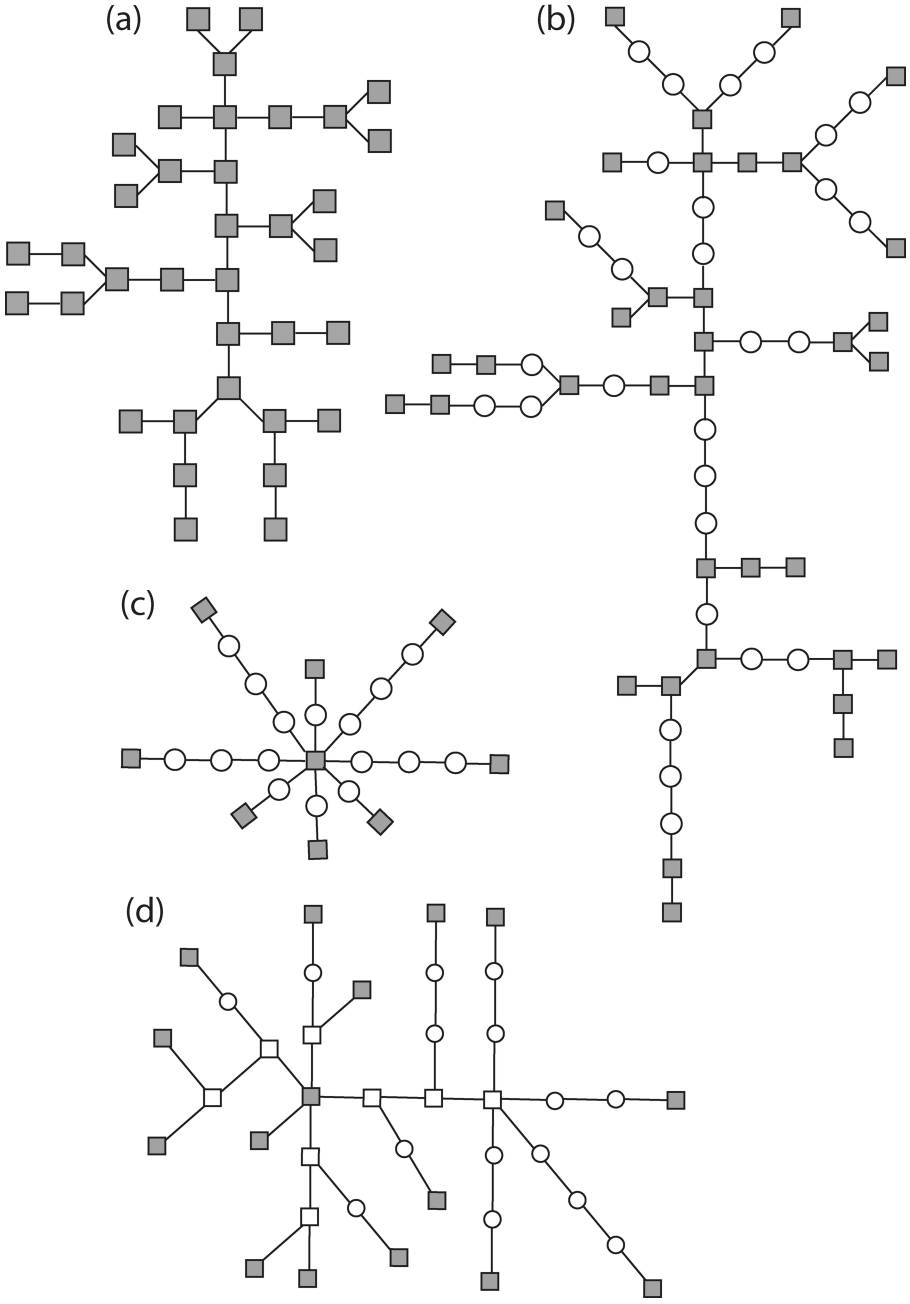


FIGURE 2. Template tree topologies along which the evolution of DNA sequence data sets have been simulated. Along each of the one-step connections between any two haplotypes, sampled or missing, the probability of occurrence of one mutation is 0.01 per site. (a) The topology with single-step branches and all haplotypes sampled should be easy to reconstruct by most methods. The more complex topologies in b to d were chosen as they allow testing the impact of (b) the presence of long branches, (c) a starlike pattern, and (d) missing node haplotypes on tree construction (see Material and Methods for more details). Grey squares, sampled haplotypes; open squares and circles, missing node and branch haplotypes, respectively.

long branches have been chosen to test whether methods would incorrectly connect some tip haplotypes together instead of connecting them independently to the central haplotype. Finally, in tree d (Fig. 2), eight of the nine node haplotypes (open squares in Fig. 2d) were excluded from the data set before genealogical reconstruction in order to investigate the influence of missing node haplotypes on graph construction. Indeed, the presence of missing node haplotypes was shown in Cassens et al. (2003) to be problematic for the MSN method because they are not inferred by this approach.

For all 400 simulated data sets (100 independent simulations along each of the four template topologies), the genealogical relationships among haplotypes have been estimated using (i) three widely used "network" construction methods and (ii) our newly developed

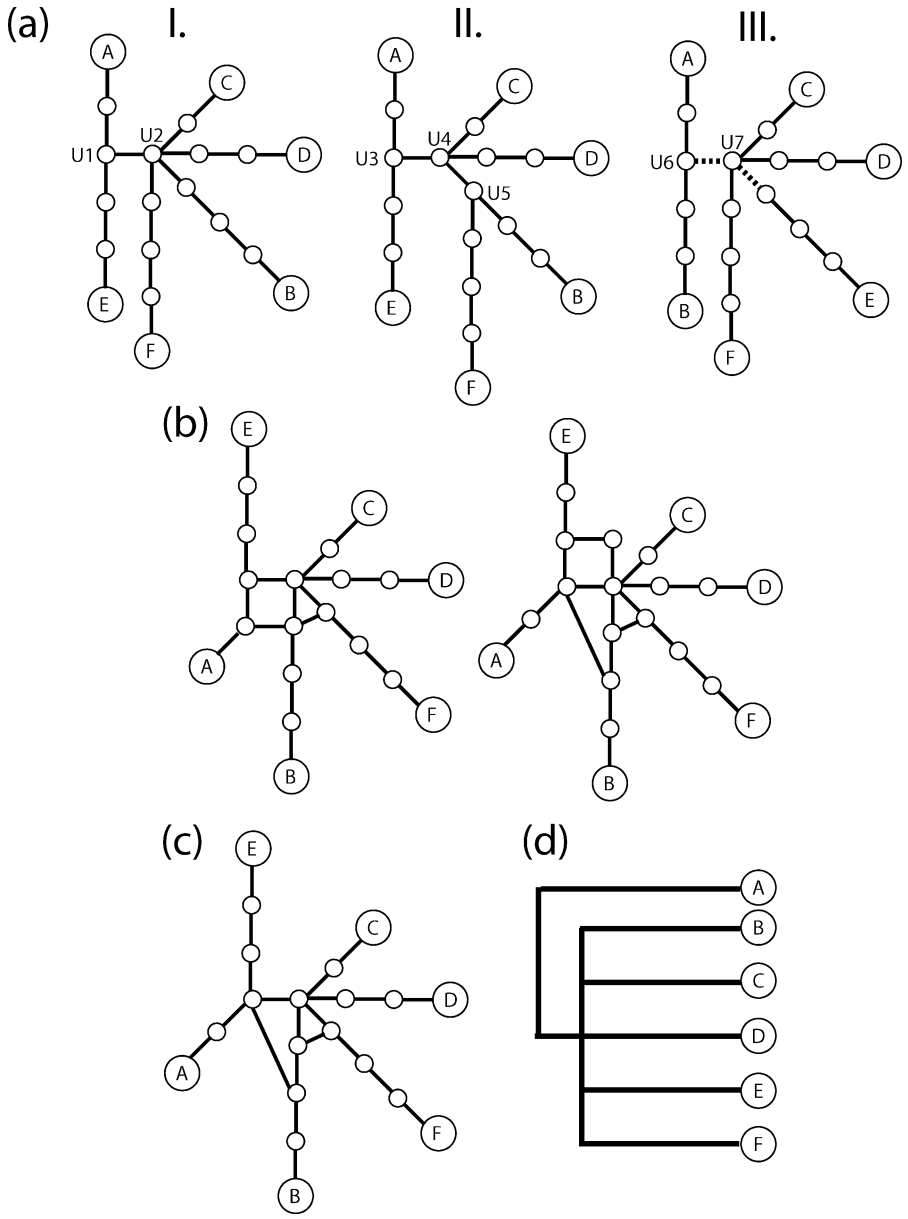


FIGURE 3. An example of how alternative tree topologies are combined into a single graph using the algorithm described in the Appendix. (a) Three trees suggesting different genealogical topologies among haplotypes A to F. (b) Two reticulated graphs, both including all of the three topologies in (a) (this can be verified by cutting cycles at different positions). For reticulated graph construction, single-step connections are compared among trees to decide whether they can be merged (when identical among all trees) or have to be maintained (when unique in some tree topologies). As demonstrated in this example, constructed reticulated graphs can differ in the number and/or placement of loops depending on the order in which the connections are tested. The left and right graphs exhibit two and three loops, respectively. The former is therefore selected as the maximum parsimony graph (UMP). (c) An illustration of a graph that does not include all genealogical pathways. Indeed, two branches of the third topology (dashed lines) must be merged to fit the graph in (c). (d) The strict consensus among the three trees in (a) which conveys less information than the reticulated graphs in (b) (see introductory section of text).

approach that simply combines all maximum parsimony trees into a single (possibly reticulated) graph.

Network Construction

Minimum-spanning network.—An algorithm for constructing minimum-spanning trees (MSTs) from a matrix of pairwise distances (absolute number of differences) among haplotypes (Prim, 1957; Rohlf, 1973) has been modified in order to include all possible MSTs within a single graph, the MSN (Excoffier and Smouse, 1994). The connections in MSTs are only formed among sampled haplotypes. The inference of missing node haplotypes (see definition in Fig. 1 legend) is therefore not possible. Each of our 400 simulated data sets was analyzed using the software ARLEQUIN, v. 2.000 (Schneider et al., 2000).

Median-joining network.—Under this approach (Bandelt et al., 1999), all MSTs are first combined within a single network (MSN) following an algorithm analogous to that proposed by Excoffier and Smouse (1994). Then, using the parsimony criterion, inferred missing node haplotypes are added to the graph in order to reduce its overall length. Each of the 400 data sets was analyzed with the program NETWORK, v. 2.0 (available at http://www.fluxus-engineering.com/sharenet.htm), with parameter $\varepsilon = 0$. Additionally, some of the data sets were reanalyzed with an increased value of epsilon to test whether it improves the performances of the method (with $\varepsilon > 0$, less parsimonious pathways are also included in the graph).

Statistical parsimony network.—The method (Templeton et al., 1992; implemented in TCS, v. 1.13 [Clement et al., 2000]) first defines the uncorrected distance above which the parsimony principle is violated with more than 5% probability ("parsimony limit"). Then, all connections are iteratively established among haplotypes starting with the smallest distances and ending either when all haplotypes are connected or the distance corresponding to the parsimony limit has been reached. Although missing node haplotypes can be inferred using the TCS program, the exact algorithm is not described in the literature yet. For graph construction using the TCS program, we added 900 constant characters (monomorphic nucleotide sites) to each of the 400 simulated data sets in order to increase the parsimony limit (see above) and force the program to connect all haplotypes into a single figure.

A new simple approach: Union of maximum parsimonious trees (UMP).—This method requires two consecutive steps. First, maximum parsimony analyses are performed for each data set and all most parsimonious trees are saved with their respective branch lengths; we used the TBR branch swapping (1000 replicates with random sequence addition) heuristic search option in the program PAUP *4.0b10 (Swofford, 2003). Second, all the saved MP trees are combined into a single figure (see Fig. 3a and b for an example) using an algorithm that (i) combines all connections from all MP trees into a single reticulated graph, and (ii) merges branches, node haplotypes, and branch haplotypes (see definition in Fig. 1), sampled or missing, that are identical among different trees (see Appendix 1 for a detailed description of the algorithm, available at the Society of Systematic Biologists website, http://systematicbiology.org; executables for Windows, Mac OS, and Linux, as well as the source code are available at www.ulb.ac.be/sciences/ueg/html_files/softwares.html). Hence, during step (ii), some cycles are maintained (i.e., some branches/haplotypes from different trees are not merged) where unique genealogical pathways are suggested in one or several (but not all) MP trees (Fig. 3). This second step is algorithmic, as it builds one of different possible graphs that include all the saved MP trees (i.e., each MP tree, including branch lengths, can be reconstructed from the final reticulated graph by removing a number of edges). As we found (unpublished data) that the placement and number of loops constructed by this algorithm can depend on the order with which connections are compared among trees (see Fig. 3b for an example), the result with the lowest number of cycles was selected among the 10 graphs produced with 10 different orders of connection comparisons.

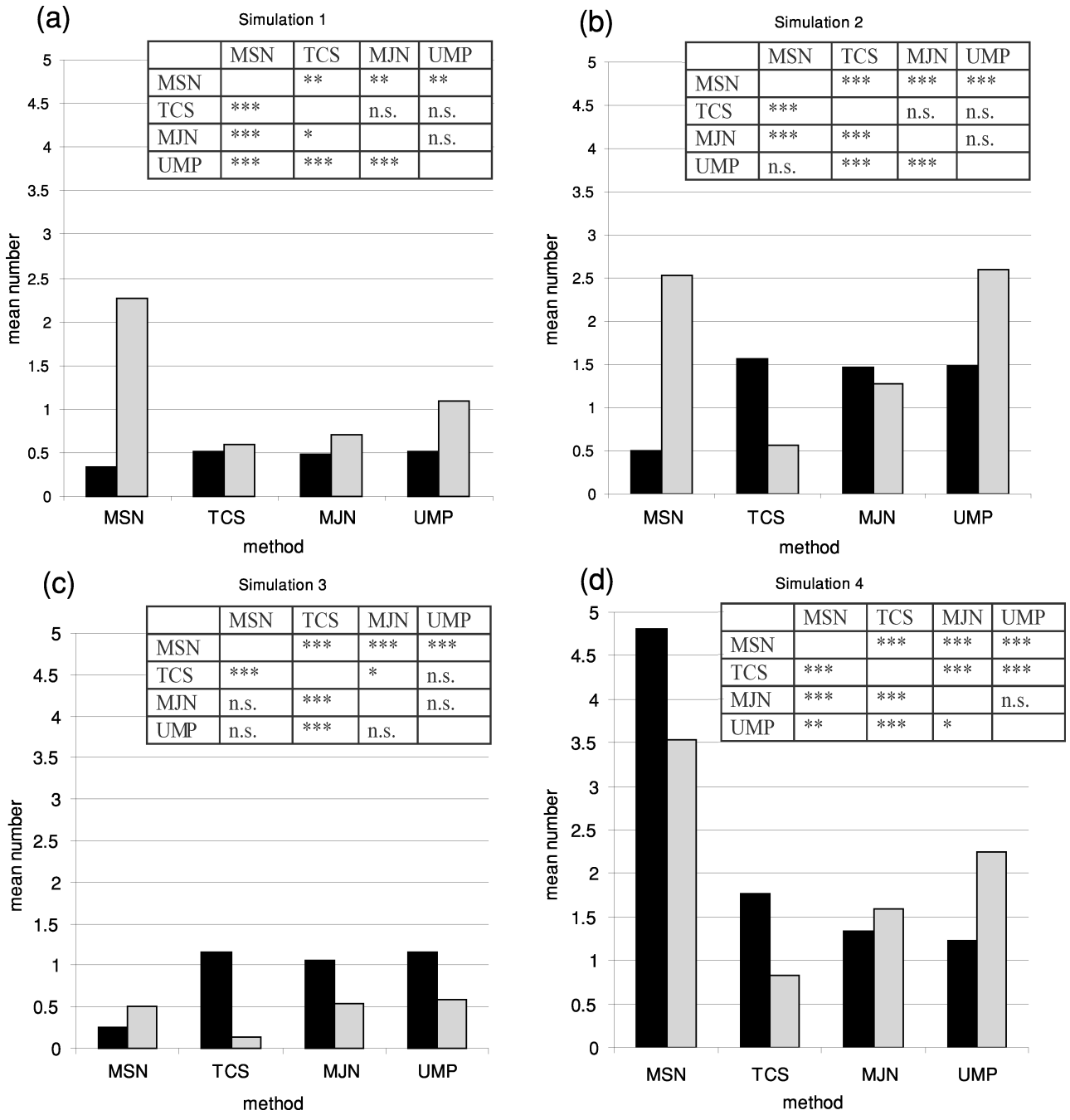
Comparison of Graphs and Statistical Analyses

We have compared by hand, for each of the 100 simulated data sets, each of the four graphs (constructed by the four different graph construction methods, including ours) to the original (template) topology (making, in total, 1600 comparisons) and computed their level of ambiguity (number of loops), tree length, and compatibility (number of errors). When the constructed graph was reticulated, the "most correct tree" (i.e., the tree with the minimum number of wrong connections) that could be obtained by cutting loops was used for computing tree length. Hence, the tree length we report here is neither necessarily the length of the MP tree (even for the UMP approach) nor the total length of the reticulated graph. An error was defined as a branch that has to be removed from the template genealogy (making it partly unconnected; i.e., each removal of a branch separates the template tree into two subtrees) in order to make it totally compatible with the constructed graph (i.e., all the connections that remain in the subtrees are included in the constructed graph). In all cases, the smallest possible number of errors was calculated. We also computed, for each of the four methods, the mean number of errors and loops (among the 100 constructed graphs). Furthermore, for each simulated data set, the four graph construction methods were sorted according to the number of errors and to tree length; i.e., we computed the relative number of times each individual approach found the most correct and/or the shortest tree. Note that two or more methods can yield trees with identical length and/or compatibility. Significant differences between methods were tested using a *t*-test (pairwise comparisons, two-sided). Bonferroni corrections (Rice, 1989) were used to eliminate the false assignment of significance by chance. Although the level of compatibility between the template genealogy and the constructed graph (i.e., the number of errors calculated) is an important measure of a method's effectiveness, it is also essential to take ambiguity (number of loops) into account. Indeed, a graph with maximum ambiguity (i.e., in graph theory terms: "a complete graph," where each node is connected to all others) has a 100% compatibility (no errors), yet it is of little value since it conveys no genealogical information. We, however, find it artificial to combine these two different measures into a single metric. Indeed, this procedure would require that we give a specific weight to one measure relative to the other (how many loops are equivalent to one error?). As this would involve an arbitrary choice, we prefer to report both measures separately.

RESULTS

Compatibility (Fig. 4, black bars)

When all node haplotypes were present in the sample (topologies in Fig. 2a to c), three of the tested methods (TCS, MJN, and UMP) yielded similar numbers of errors when constructing the graph (Fig. 4): the mean number of errors per constructed graph is approximately 0.5, 1.5, and 1.2 for simulations 1, 2, and 3, respectively. Differences of compatibility among these three approaches



Simulation 1	MSN	TCS	MJN	UMP
MSN		**	**	**
TCS	***		n.s.	n.s.
MJN	***	*		n.s.
UMP	***	***	***	

Simulation 2	MSN	TCS	MJN	UMP
MSN		***	***	***
TCS	***		n.s.	n.s.
MJN	***	***		n.s.
UMP	n.s.	***	***	

Simulation 3	MSN	TCS	MJN	UMP
MSN		***	***	***
TCS	***		*	n.s.
MJN	n.s.	***		n.s.
UMP	n.s.	***	n.s.	

Simulation 4	MSN	TCS	MJN	UMP
MSN		***	***	***
TCS	***		***	***
MJN	***	***		n.s.
UMP	**	***	*	

FIGURE 4. Mean number of errors (black bars) and mean number of loops (grey bars) in the graphs generated by the different methods (MSN, minimum-spanning network; TCS, statistical parsimony network; MJN, median-joining network; UMP, union of maximum parsimonious trees) from 100 simulated data sets (along the template topologies represented in Fig. 2a to d). Significance levels of pairwise *t*-tests (after Bonferroni corrections) are indicated in the tables (for mean number of errors and number of loops, above and below diagonal, respectively). n.s., nonsignificant; $^{*}P < 0.0083$; $^{**}P < 0.0017$; $^{***}P < 0.0002$. Note that highly significant differences between TCS and both the MJN and UMP approaches were found in the simulations under template topology d (Fig. 2d), whereas the differences among the three methods were mostly not significant in simulations under the three other template topologies.

are not significant (with the exception of the comparison between TCS and MJN for topology 3). TCS, MJN, and UMP, however, all built significantly more erroneous connections than the MSN approach, which yielded in average only 0.33, 0.50, and 0.26 errors per constructed graph for simulations 1, 2, and 3, respectively. On the other hand, results were diametrically different for topology d (Fig. 2d): when most node haplotypes are missing, the MSN method yielded a number of wrong connections (in average 4.81) significantly higher ($P < 0.0002$) than those produced by the three other methods (Fig. 4d). Although the differences in absolute numbers of errors are less spectacular, TCS constructed a number of wrong connections significantly higher than MJN and UMP, whereas the difference in compatibility between these two latter approaches is not significant.

Ambiguity (Fig. 4, grey bars)

The differences among methods for the inferred number of loops are striking: TCS inferred a number of alternative connections always significantly lower than the three other methods. Furthermore, with the exception of simulation 3 (where MSN, MJN, and UMP approaches yield results that cannot be differentiated statistically), the MSN and UMP approaches generated numbers of loops significantly higher than MJN.

An increase in the number of loops should logically increase compatibility: the addition of alternative connections increases the likelihood to build correct connections (a complete graph will necessarily contain the correct topology). However, this negative correlation between compatibility and loop number (visible in Fig. 4d for TCS, MJN, and UMP) does not seem to hold in all analyses performed here (see, for example, the difference between TCS and UMP in Fig. 4c). Surprisingly, a high proportion of the cycles inferred by ARLEQUIN (e.g., 65% and 50% in simulations 1 and 2, respectively) violate the principle of the minimum-spanning approach as they consist of links between two haplotypes that are already connected within the MSN by a shorter branch.

Tree Length

As described in Material and Methods, we defined "length of most correct tree" as the number of mutational steps on the most correct tree topology (i.e., the tree with the minimum number of wrong connections) that can be found within the constructed reticulated graph. We then estimated how frequently different methods found not only the most correct, but also the shortest tree relative to other methods. In simulations 1 to 3 (cf. Fig. 5a to c), MSN-constructed graphs least often included the shortest topology, but most often included the most compatible topology. On the contrary, in the fourth simulation series (Fig. 5d), there is a good correlation between the number of times a method finds the most correct tree and the number of times it finds the shortest tree. When node haplotypes are missing, the most correct tree topology found is often (but not always) the shortest one. Note, however, that the most correct tree extracted from a UMP graph (by cutting loops, see Material and Methods) is sometimes longer than the individual MP trees used for reticulated graph construction. Indeed, a UMP graph can be compatible with a tree that is not one of the MP trees, but whose topology is nevertheless more correct, i.e., more similar to that of the model tree.

DISCUSSION

General Performance of the UMP Approach

In the recent literature, a clear line has often been drawn between algorithms developed for the reconstruction of evolutionary relationships among intraspecific haplotypes and more traditional methods for phylogeny inference among well-separated species. More specifically, it has been argued that "network" methods such as statistical parsimony (Clement et al., 2000) are more appropriate than the global MP approach for the analysis of intraspecific data sets (Crandall, 1994; Posada and Crandall, 2001). Reticulated graphs undoubtedly are valuable tools for visualization, in a single figure, of all the connections that are observed in a set of alternative genealogies (rarely do loops represent true reticulation, cf. introductory section of this article). However, there is no logical reason we know of that could make purely algorithmic approaches, such as those tested in this study, necessarily more appropriate for building such graphs than methods based on an optimality criterion. A strict consensus tree is usually produced after an MP analysis and conveys less information than what is available in the original MP trees. For this reason, we introduce here a simple method (union of maximum parsimonious trees, UMP) that unites all most parsimonious trees within a single (possibly reticulated) graph. Using simulated data sets, we then compared the performance of UMP against other intraspecific "network construction methods." Despite the low variation among haplotypes and the large number of possible topologies among them, a heuristic MP search could be carried out in practical computing time for all 400 simulations (e.g., we analyzed all 100 data sets simulated from topology d with PAUP* in only 8 min using an iMac G4 700 Mhz), resulting most frequently in 1 to 50, and only rarely in more than 500, MP trees. Neither ancestral nor multiple descendant haplotypes, often found in intraspecific data sets, were problematic in the analysis, as they could be represented in the constructed graph through collapsing of zero-length branches. Most importantly, the analyses presented here show that the UMP approach performs equally well or, in some of the simulations, even better than some of the purely algorithmic "network construction methods." UMP performs especially well in situations where "node haplotypes" have not been sampled and therefore need to be inferred (see below for a detailed discussion on the relative compatibility/ambiguity of different methods). Our analyses indicate that UMP has a tendency to construct more loops than the TCS or MJN methods, but the absolute number of cycles per graph (0.5 to 2.6, cf. Fig. 4) remains reasonable such that graphical representation is usually not hampered. Moreover, given that the order in

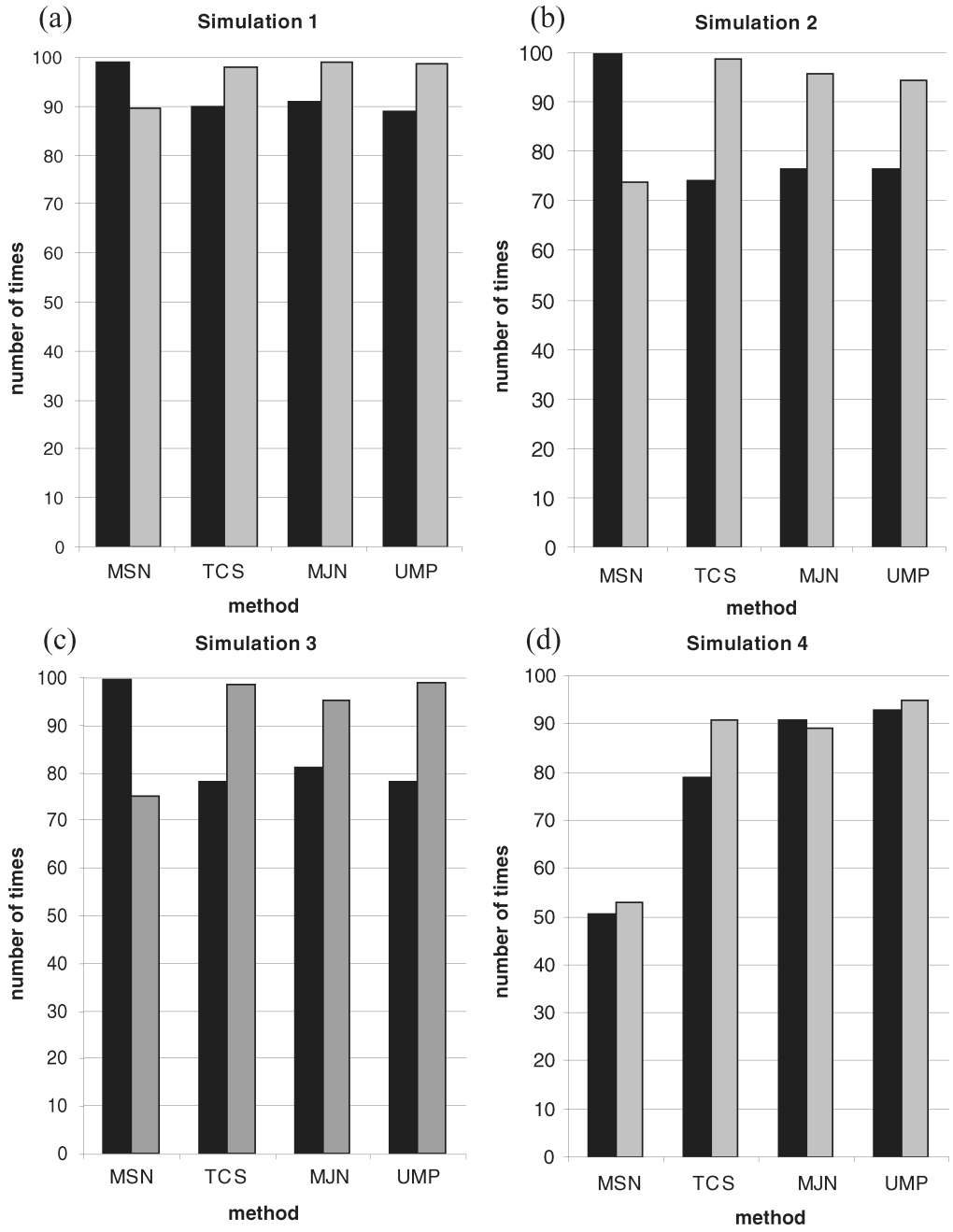


FIGURE 5. The relative proportion of times that the four tested methods (see legend of Fig. 4 for abbreviations) have constructed a graph containing the most correct (black bars) and shortest (grey bars) topology in each of the four simulations sets.

which the UMP algorithm chooses the connections for comparison among the MP trees sometimes influences not only the placement but also the number of inferred loops (as shown in Fig. 3b), it is likely that some of the cycles that have been constructed are not necessary to maintain compatibility with all MP trees. As we selected the graph with the lowest number of cycles among 10 graphs produced with 10 different initial comparisons of connections, it is possible that more repetitions (e.g., 100 or 1000) are necessary to reliably find the minimum number of loops.

Another issue that might be relevant for improving the performance of the UMP approach is not related to the combination of the trees into a (possibly reticulated) graph, but to the reconstruction of ancestral haplotypes on the individual MP topologies (during the MP search). Most phylogeny inference programs that implement maximum parsimony do not display all possible assignments of states to ancestral nodes, and thus all possible distributions of substitutions across the tree. For example, using the program PAUP* 4.0b10 (Swofford, 2003), one chooses one of the subsets of all possible

reconstructions of ancestral states: the ACCTRAN ("accelerated transformation") and DELTRAN ("delayed transformation") options force changes to occur, respectively, as far down or up the tree as possible. It remains to be investigated which, if any, of these biased distributions of substitutions, or the general inclusion of all possible ancestral reconstructions, could lead to a higher level of compatibility of our UMP method.

Comparative Analysis of Graph Construction Methods

Although network construction approaches are widely used in population genetic studies, their general performance had never been analyzed in a systematic fashion. That there is a clear need for a comparative analysis, however, has recently been shown in our empirical study, which revealed substantial differences among methods (Cassens et al., 2003).

The four methods included in our present study can clearly be divided into two groups regarding their general performances. Although the statistical parsimony (TCS), median-joining (MJN), and maximum parsimony (UMP) approaches show relatively similar results in terms of compatibility, the graph topologies constructed using the minimum spanning algorithm (MSN) are frequently very different. For example, trees extracted from minimum-spanning networks are, on average, considerably longer in all simulation sets (Fig. 5a to d), confirming the much higher number of mutational steps that had been inferred for the empirical data set using this method (Cassens et al., 2003). Somewhat counterintuitive (at least when considering the parsimony principle) is that the significantly longer MSN tree topologies are correlated with higher compatibility levels in the first three simulations (Figs. 4 and 5a to c). However, this can be explained by the fact that the node haplotypes in the corresponding model topologies (1 to 3, Fig. 2a to c) were all sampled (while branch lengths and/or branching patterns differed). Indeed, the MSN algorithm searches for minimum length connections between sampled haplotypes only, and is not capable of inferring missing node haplotypes. Consequently, the poor performances (high number of errors and high number of loops) of the MSN method (Figs. 4d and 5d) with data simulated on model tree d (Fig. 2d) must be caused by missing node haplotypes: the average lengths of MSN most correct trees are high and their corresponding topologies contain a large average number (4.8) of incorrect connections (connections are always formed between the genetically nearest sampled haplotypes without inference of missing node haplotypes). Hence, the MSN approach is expected to perform well under the rather unrealistic situation when most or all node haplotypes have been sampled. We feel that the inclusion of the minimum-spanning method within our comparative analysis is important, as the approach is widely used in population genetic studies (more than 10 publications in Volume 12 of *Molecular Ecology* make use of MSN for data analysis; e.g., Chenoweth and Hughes, 2003; Drew et al., 2003; Michaux et al., 2003).

Each of the three other methods (TCS, MJN, and UMP) occasionally failed to reconstruct the correct topology. This seems to be caused by homoplasious characters that force the construction of false node haplotypes (e.g., erroneously linking convergent haplotypes). One finding of our previous empirical mitochondrial study (Cassens et al., 2003) was that the placement of a branch (leading to three sampled haplotypes) considerably differed between the topologies inferred by the TCS and MJN methods. More surprisingly, both graphs were completely resolved trees (i.e., without any reticulation), suggesting the lack of ambiguity regarding genealogical relationships. In the present simulation study, the performance of the two methods could be further tested as well as compared to the new maximum parsimony approach. Although the three methods yield very similar results (in terms of compatibility) on data sets simulated along model topologies with sampled node haplotypes (Fig. 2a to c), we observed a different outcome for the simulations on the fourth model topology characterized by missing node haplotypes (topology d, Fig. 2). Under these conditions, both the MJN and the UMP methods perform equally well and generate significantly less errors than the statistical parsimony approach as implemented in TCS (Fig. 4d). These two methods also generated more ambiguous graphs, as shown by their higher average number of loops (see Fig. 4). In our view, however, it is more important to increase compatibility, even at the price of a slight increase in ambiguity (in Fig. 4, maximum average number of loops reaches only 2.25). Indeed, it is incompatibility, and not ambiguity, that can yield erroneous interpretations of the graphs.

Identification of the source of TCS lower compatibility is difficult, as no exact description exists in the literature on how the TCS algorithm works and, in particular, on how missing node haplotypes are inferred. This is important because TCS networks are increasingly used in phylogeography studies (more than 25 publications in Volume 12 of *Molecular Ecology* make use of TCS for data analysis; e.g., Brant and Orti, 2003; Contreras-Diaz et al., 2003; Lawton-Rauh et al., 2003; Printzen et al., 2003), often in connection with the Nested Clade Analysis method (Templeton, 1998). Many of these published TCS networks include missing node haplotypes. Based on our simulation results, we propose that, if haplotypes are relatively distant (such that some node haplotypes are missing), priority in data analysis should be given to either the median-joining approach (which is considerably less often used in intraspecific studies: only two publications in Volume 12 of *Molecular Ecology* make use of MJN for data analysis; Collevatti et al., 2003; Lloyd, 2003) or the maximum parsimony approach (UMP) described here. Furthermore, among the methods tested, only the MJN and our UMP approaches show enough flexibility in the accepted number of inferred loops. With MJN, increasing the value of the parameter epsilon increases the number of reticulations as less-parsimonious solutions are then included (Bandelt et al., 1999). Likewise, we suggest that near-MP trees (i.e., trees with length slightly higher than that of the MP trees) can be saved during

tree search (e.g., in PAUP*) and then combined into a single graph. Preliminary tests show that the inclusion of near parsimonious trees (e.g., length of MP trees + 1) generated much more complex but also more compatible genealogy inference (data not shown).

ACKNOWLEDGEMENTS

We thank Valérie Gillon for providing programming assistance and Colas Schretter for helping us to formally describe the UMP algorithm. We are grateful to Mike Steel, Peter Foster, and two anonymous reviewers for useful suggestions that have greatly improved our manuscript. This work was funded by the National Fund for Scientific Research Belgium (FNRS), the Université Libre de Bruxelles (ULB), the Fonds Van Buuren, the Fonds E. Defay, and the "Communauté Française de Belgique" (ARC 11649/20022770). PM is Research Associate at the National Fund for Scientific Research Belgium (FNRS).

REFERENCES

Avise, J. C., J. Arnold, R. M. Ball, E. Bermingham, T. Lamb, J. E. Neigel, C. A. Reeb, and N. C. Saunders. 1987. Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. Ann. Rev. Ecol. Syst. 18:489–522.

Bandelt, H. J., P. Forster, and A. Röhl. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16:37–48.

Brant, S. V., and G. Orti. 2003. Phylogeography of the northern short-tailed shrew, *Blarina brevicauda* (insectivora: Soricidae): Past fragmentation and postglacial recolonization. Mol. Ecol. 12:1435–1449.

Cassens, I., K. Van Waerebeek, P. B. Best, E. A. Crespo, J. C. Reyes, and M. C. Milinkovitch. 2003. The phylogeography of dusky dolphins (*Lagenorhynchus obscurus*): A critical examination of network methods and rooting procedures. Mol. Ecol. 12:1781–1792.

Chenoweth, S. F., and J. M. Hughes. 2003. Oceanic interchange and non-equilibrium population structure in the estuarine dependent indo-pacific tasselfish, *Polynemus sheridani*. Mol. Ecol. 12:2387–2397.

Clement, M., D. Posada, and K. A. Crandall. 2000. Tcs: A computer program to estimate gene genealogies. Mol. Ecol. 9:1657–1659.

Collevatti, R. G., D. Grattapaglia, and J. D. Hay. 2003. Evidences for multiple maternal lineages of *Caryocar brasiliense* populations in the brazilian cerrado based on the analysis of the chloroplast DNA sequences and microsatellite haplotype variation. Mol. Ecol. 12:105–115.

Contreras-Diaz, H. G., O. Moya, P. Oromi, and C. Juan. 2003. Phylogeography of the endangered darkling beetle species of *Pimelia* endemic to gran canaria (canary islands). Mol. Ecol. 12:2131–2143.

Crandall, K. A. 1994. Intraspecific cladogram estimation: Accuracy at higher levels of divergence. Syst. Biol. 43:222–235.

Drew, R. E., J. G. Hallett, K. B. Aubry, K. W. Cullings, S. M. Koepf, and W. J. Zielinski. 2003. Conservation genetics of the fisher (*Martes pennanti*) based on mitochondrial DNA sequencing. Mol. Ecol. 12:51–62.

Excoffier, L., and P. E. Smouse. 1994. Using allele frequencies and geographic subdivision to reconstruct gene trees within a species: Molecular variance parsimony. Genetics 136:343–359.

Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating the human-age splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22:160–174.

Lawton-Rauh, A., R. H. Robichaux, and M. D. Purugganan. 2003. Patterns of nucleotide variation in homoeologous regulatory genes in the allotetraploid hawaiian silversword alliance (Asteraceae). Mol. Ecol. 12:1301–1313.

Lloyd, B. D. 2003. The demographic history of the new zealand short-tailed bat *Mystacina tuberculata* inferred from modified control region sequences. Mol. Ecol. 12:1895–1911.

Michaux, J. R., E. Magnanou, E. Paradis, C. Nieberding, and R. Libois. 2003. Mitochondrial phylogeography of the woodmouse (*Apodemus sylvaticus*) in the western palearctic region. Mol. Ecol. 12:685–697.

Posada, D., and K. A. Crandall. 2001. Intraspecific gene genealogies: Trees grafting into networks. Trends Ecol. Evol. 16:37–45.

Prim, R. C. 1957. Shortest connection networks and some generalizations. Bell Syst. Tech. J. 36:1389–1401.

Printzen, C., S. Ekman, and T. Tonsberg. 2003. Phylogeography of *Cavernularia hultenii*: Evidence of slow genetic drift in a widely disjunct lichen. Mol. Ecol. 12:1473–1486.

Rambaut, A., and N. C. Grassly. 1997. SEQ-GEN: An application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. Comput. Appl. Biosci. 13:235–238.

Rice, W. R. 1989. Analyzing tables of statistical tests. Evolution 43:223–225.

Rohlf, F. J. 1973. Algorithm 76. Hierarchical clustering using the minimum spanning tree. Comp. J. 16:93–95.

Schneider, S., D. Roessli, and L. Excoffier. 2000. ARLEQUIN version 2.000: A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva.

Swofford, D. L. 2003. PAUP*, phylogenetic analysis using parsimony (*and other methods), Sinauer Associates, Sunderland, Massachusetts.

Swofford, D. L., G. J. Olsen, P. J. Waddell, and D. M. Hillis. 1996. Phylogeny reconstruction. Pages 407–514 *in* Molecular systematics, 2nd ed. (D. M. Hillis, C. Moritz, and B. K. Mable, eds.). Sinauer, Sunderland, Massachusetts.

Templeton, A. R. 1998. Nested clade analyses of phylogeographic data: Testing hypotheses about gene flow and population history. Mol. Ecol. 7:381–397.

Templeton, A. R., K. A. Crandall, and C. F. Sing. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. II. Cladogram estimation. Genetics 132:619–633.

First submitted 3 March 2004; reviews returned 6 May 2004; final acceptance 14 December 2004
Associate Editor: Mike Steel