

## Visualizing Conflicting Evolutionary Hypotheses in Large Collections of Trees: Using Consensus Networks to Study the Origins of Placentals and Hexapods

BARBARA R. HOLLAND,<sup>1</sup> FRÉDÉRIC DELSUC,<sup>1,3</sup> AND VINCENT MOULTON<sup>2</sup>

<sup>1</sup>The Allan Wilson Centre for Molecular Ecology and Evolution, Massey University, Private Bag 11 222, Palmerston North, New Zealand;  
E-mail: b.r.holland@massey.ac.nz

<sup>2</sup>The Linnaeus Centre for Bioinformatics, Uppsala University, Sweden

<sup>3</sup>Present Address: Département de Biochimie, Faculté de Médecine, Université de Montréal, Québec, Canada

**Abstract.**— Many phylogenetic methods produce large collections of trees as opposed to a single tree, which allows the exploration of support for various evolutionary hypotheses. However, to be useful, the information contained in large collections of trees should be summarized; frequently this is achieved by constructing a consensus tree. Consensus trees display only those signals that are present in a large proportion of the trees. However, by their very nature consensus trees require that any conflicts between the trees are necessarily disregarded. We present a method that extends the notion of consensus trees to allow the visualization of conflicting hypotheses in a consensus network. We demonstrate the utility of this method in highlighting differences amongst maximum likelihood bootstrap values and Bayesian posterior probabilities in the placental mammal phylogeny, and also in comparing the phylogenetic signal contained in amino acid versus nucleotide characters for hexapod monophyly. [Bayesian posterior probabilities; consensus trees; hexapods; median networks; nonparametric bootstrap; placental mammals; phylogenetic trees; quartet puzzling.]

The estimation of phylogenetic trees remains a central task in evolutionary biology. Many methods have been developed for constructing trees (Swofford et al., 1996). However, a single point estimate of the optimal tree with no measure of reliability is not always useful (Penny and Hendy, 1986). For this reason it is common practice to employ methods that generate a tree together with some measure of support for individual edges (branches). These methods include nonparametric bootstrap resampling with replacement (Felsenstein, 1985), Bayesian approaches based on Markov chain Monte Carlo (MCMC) methods (Huelsenbeck et al., 2002), and quartet puzzling (Strimmer and von Haeseler, 1996).

A common feature of all these methods is that, rather than producing a single tree, they produce a large set of trees. In the case of nonparametric bootstrapping, this is accomplished by resampling columns of the sequence alignment and estimating a tree for each sample (Felsenstein, 1985). Bayesian approaches explore the phylogenetic landscape by using MCMC, a numerical method where trees are sampled in proportion to their posterior probability (Huelsenbeck et al., 2002). Quartet puzzling provides a fast estimate of the maximum likelihood (ML) tree by computing all ML quartets and then puzzling them together (Strimmer and von Haeseler, 1996). This puzzling operation is order dependent; so many random orderings of the quartets are used, each producing a tree.

In order to summarize the large collections of trees that these methods produce, consensus trees are usually constructed in which the weights assigned to the edges reflect their level of support. Bryant (2003) recently provided a comprehensive overview of consensus tree methods and their classification. Consensus trees display only those signals that are supported by a large proportion of the trees. However, although a consensus tree is a concise way to summarize the information contained in a large collection of trees, by definition it cannot display conflicting signals. For instance, the following two situations would be impossible to distinguish using a con-

sensus tree: (1) One grouping of the taxa has slightly more than 50% support and another conflicting grouping has slightly less than 50% support; (2) one grouping of the taxa has slightly more than 50% support and no other conflicting grouping has more than 10% support. In addition, a lack of resolution in the consensus tree usually implies conflict, but a polytomy could be resolved in many ways, only a few of which may have support amongst the input trees.

We present a method that generalizes the notion of consensus trees to allow conflicting evolutionary hypotheses to be displayed within a network. This expands upon a method proposed by Bandelt (1995). We illustrate the use of consensus networks by focusing on two biological examples that have recently generated controversies. First, we explore the differences between Bayesian posterior probabilities and bootstrap proportions in the case of placental mammal phylogeny (Murphy et al., 2001b; Suzuki et al., 2002), and second, we examine the extent of the conflict between amino acid and nucleotide characters for the origins of hexapods (Nardi et al., 2003a; Delsuc et al., 2003).

### THE CONSENSUS NETWORK METHOD

We begin with an example to illustrate the way in which consensus trees can be generalized to consensus networks. Consider the three trees shown in Figure 1A. We begin by making a list of the splits that are displayed by the internal edges of these trees, together with a weight indicating the number of trees displaying each split (Figure 1B). The strict consensus tree is the tree that displays only those splits which occur in all **three** of the trees and, because none of the splits in our list have weight 3, only the splits corresponding to the pendant edges of the trees are displayed (Figure 1C). The majority-rule consensus tree is the tree that displays only those splits which occur in **two** or more of the trees (Figure 1D). In this example, the required splits are those with weight

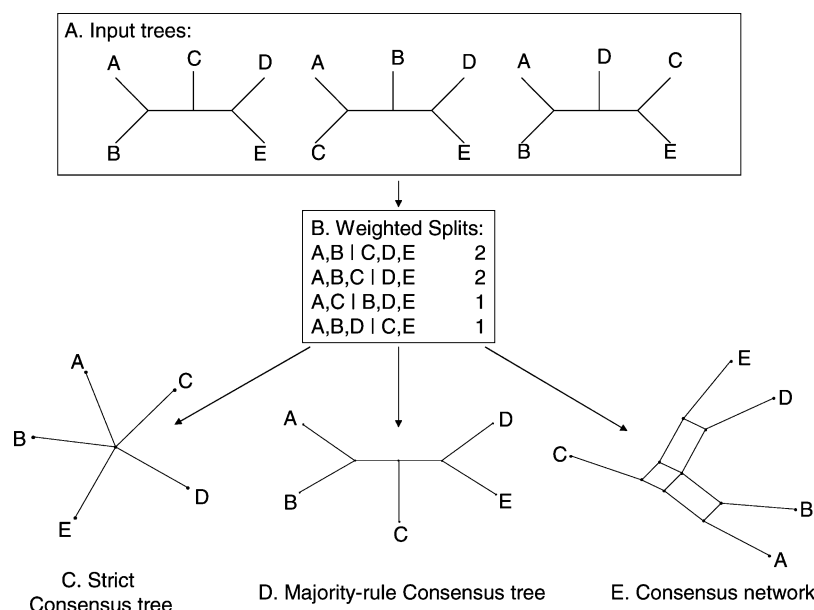


FIGURE 1. The strict consensus tree (C), majority consensus tree (D), and consensus network (E) for three trees (A). Each edge in a phylogenetic tree corresponds to a split or bipartition of the taxa set, as removing an edge from a tree divides it into two subtrees, which also partitions its labels. The pendent edges always appear in a phylogenetic tree—these correspond to trivial splits that separate one taxon from the rest. All the nontrivial splits displayed in the trees are listed in the weighted splits box (B).

2 (AB | CDE and ABC | DE) plus the splits corresponding to the pendant edges of the input trees, and so the majority rule tree is identical to the first of the input trees. Finally, Figure 1E shows the consensus network, which displays those splits that occur in **one** or more trees, of which there are 9. This network may be interpreted as follows: Each split is represented by a class of parallel edges, whose lengths are proportional to the weight assigned to the split. Note that even though, for example, the splits AB | CDE and AC | BDE cannot be displayed simultaneously by any tree, they are both displayed by the network.

We now discuss our method in general. Consider as input a collection of phylogenetic trees, all on the same set of taxa. These could be generated by any of the methods mentioned above or might be the result of heuristic or exact searches that return many equally well-supported trees. Next, as in the example, the list of splits corresponding to edges in the trees is generated and weighted according to the frequency with which they occur. The weighted splits are then displayed in a *median network* (Bandelt, 1994). These networks can be used to display any collection of weighted splits and have some attractive properties. For example, in the case where the collection of splits corresponds to a tree (i.e., corresponds to the edges of a given phylogenetic tree), the associated median network is that tree. Moreover, median networks are straightforward to generate using an algorithm first introduced in Bandelt et al. (1995), which has been implemented in the freely available programs Spectronet (Huber et al., 2002) and jsplits ([http://www-ab.informatik.uni-tuebingen.de/software/jsplits/welcome\\_en.html](http://www-ab.informatik.uni-tuebingen.de/software/jsplits/welcome_en.html)).

Even so, median networks can in general be rather complex, depending on the extent to which the displayed splits deviate from fitting into a tree (Huber et al., 2001). To control this complexity, it is convenient to quantify this deviation as follows. We say that two splits are *incompatible* if they cannot be displayed simultaneously by any tree. Moreover, we say that a collection of splits is *k-compatible* if it contains no subcollection of  $k + 1$  pairwise incompatible splits. For instance, any collection of splits that can be displayed in a tree is 1-compatible, as it contains no pair of incompatible splits. Incompatible collections of splits are represented by high-dimensional hypercubes in median networks, which are difficult to visualize. Thus the parameter  $k$  is important since, for small  $k$ , collections of  $k$ -compatible splits are easier to visualize with median networks.

If the collection of splits has a high degree of incompatibility, we can apply a filter to produce a subcollection of splits that has a lower degree of incompatibility, and hence results in a less complex network visualization. As with the strict and majority consensus methods, we discard those splits which are not displayed by more than some threshold proportion,  $0 < x \leq 1$ , of the trees. A crucial observation is that a collection of splits so generated is  $\lceil 1/x \rceil - 1$  compatible (where  $\lceil \cdot \rceil$  indicates rounding up to the nearest integer) (Holland and Moulton, 2003). For example, if  $x$  is 1 then the resulting consensus network will be the strict consensus tree, if  $x$  is 1/2 then the resulting consensus network will be the majority-rule tree, whereas if  $x$  is 1/3 then the consensus network will be at most 2 dimensional; that is, it may contain 2-cubes but no 3-cubes. For convenience, this result is stated with proof in the Appendix.

Note that  $k$ -compatible split systems have been extensively studied (Dress et al., 2001) and are strongly related to circular split systems and weakly compatible split systems (Bandelt and Dress, 1992). Such split systems and related phylogenetic networks—such as split graphs (Huson, 1998)—could in principle be used as alternative methods for generating consensus networks (Holland and Moulton, 2003).

A Python script has been written that reads a list of trees in Newick format (bracket, comma notation) and produces the corresponding weighted split system in Nexus format. This file can be read by Spectronet (Huber et al., 2002), which displays the corresponding consensus network. Consensus networks can also be constructed using jSplits ([http://www-ab.informatik.uni-tuebingen.de/software/jsplits/welcome\\_en.html](http://www-ab.informatik.uni-tuebingen.de/software/jsplits/welcome_en.html)).

#### APPLICATIONS

The consensus network method has a wide range of potential applications. Here, we will focus on two biological examples that demonstrate its utility in displaying conflicting hypotheses existing in large collections of trees which cannot be displayed using standard consensus tree analysis.

##### *Indices of Phylogenetic Reliability and the Phylogeny of Placental Mammals*

The sequencing and analysis of both mitochondrial and nuclear genes during the last decade has completely redefined our vision of the placental tree (de Jong, 1998; Waddell et al., 1999). Recently, two large-scale studies robustly identified for the first time four major placental clades—Afrotheria, Xenarthra, Euarchontoglires, and Laurasiatheria—and a close relationship between the two latter ones (Madsen et al., 2001; Murphy et al., 2001a). The subsequent combination of these two data sets proposed a fairly resolved picture of placental mammal relationships (Murphy et al., 2001b).

This latest study has been criticized on the grounds that Bayesian posterior probabilities (PP) might provide, in this case and others, overestimation of phylogenetic reliability (Suzuki et al., 2002). In fact, large discrepancies between PP and other measures of statistical support such as nonparametric bootstrap proportions (BP) exist for a wide range of empirical studies (see Douady et al., 2003 and references therein), and the nature of the relationship between the two indices is still poorly understood from a theoretical point of view (Huelsenbeck et al., 2002). These observations have recently generated a bloom of papers aiming at understanding the relation between the PP and BP indices (Waddell et al., 2002; Wilcox et al., 2002; Alfaro et al., 2003; Cummings et al., 2003; Douady et al., 2003; Erixon et al., 2003; Simmons et al., 2004). Credibility values of the Bayesian phylogenetic inference seem to be excessively high under some circumstances (Suzuki et al., 2002; Cummings et al., 2003; Douady et al., 2003; Erixon et al., 2003; Simmons et al., 2004), whereas bootstrap percentages might be too conservative (Hillis and Bull, 1993; Wilcox et al., 2002; Alfaro

et al., 2003). Sensitivity to model misspecification has been suspected to be an important factor in the behavior of PP (Waddell et al., 2002) and the importance of model adequacy in Bayesian phylogenetic analyses has been recently outlined (Lemmon and Moriarty, 2004). A recent paper by Huelsenbeck and Rannala (2004) attempted to put this controversy to rest. They concluded that the posterior probability of a tree can be interpreted as the probability that the tree is the true tree, with the very important proviso that the substitution model is correct. When the model is misspecified they found that posterior probabilities were biased upwards and that bootstrap values were more robust. As current substitution models are frequently likely to be misspecified (Goldman, 1993), these results suggest that both posterior probabilities and bootstrap values will continue to be of interest to practitioners.

Some authors have also independently suggested computing Bayesian BP or bootstrapped PP ( $BP_{Bay}$ ) as an alternative (Waddell et al., 2002; Douady et al., 2003). Although time-consuming, this approach leads to Bayesian support values closer to the bootstrap percentages and allows the identification of excessively supported edges in a Bayesian framework. The debate on the relationships between BP and PP is reminiscent of the one engendered by the introduction of the quartet puzzling method (Strimmer and von Haeseler, 1996) and its associated measure of edge support expressed as reliability percentages (RP). Indeed, it has been suggested that RP are generally higher than the corresponding BP (Cao et al., 1998).

We use the consensus network method to visualize the differences between the indices PP,  $BP_{ML}$ ,  $BP_{Bay}$ , and RP by analyzing the collections of trees produced by each method for the Murphy et al. (2001b) data set. The application of this method also allows the visualization of the remaining uncertainties in the placental mammal phylogeny.

The data set of Murphy et al. (2001b) consists of the concatenation of 22 genes (19 nuclear and three mitochondrial) representing a total of 16,397 unambiguously aligned nucleotide sites for 44 taxa. This includes 42 placental distributed in Afrotheria (8 taxa), Xenarthra (3), Euarchontoglires (11), and Laurasiatheria (20), plus two marsupial outgroups. A schematic overview of the experimental design we used to compare the four indices of phylogenetic reliability is presented in Figure 2.

The Bayesian analysis of the original data set was conducted using MrBayes version 3.0b4 (Ronquist and Huelsenbeck, 2003). Four incrementally heated Metropolis-coupled Markov chains Monte Carlo (MCMCMC) were simultaneously run for 1,000,000 generations, using the program's default uninformative priors as starting values for trees, edge lengths, and model parameters. We used the GTR+I+ $\Gamma_4$  model of sequence evolution estimated to be the best fitting model by Modeltest 3.06 (Posada and Crandall, 1998). Trees were sampled every 100 generations, resulting in a collection of 10,000 MCMCMC sampled trees. We discarded the first 5000 trees as a conservative burn-in

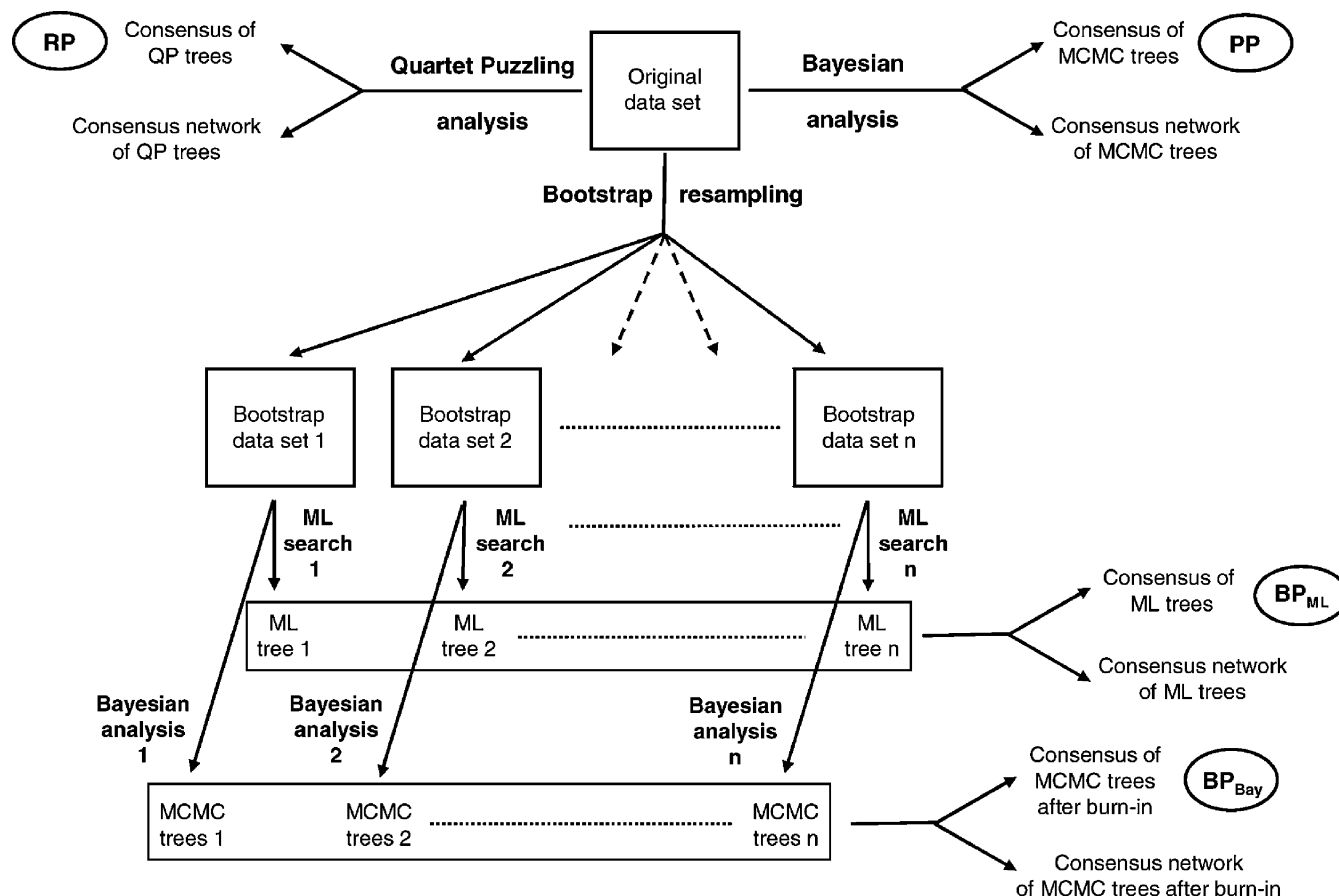


FIGURE 2. Experimental design used to analyze the collections of trees issued from the four phylogenetic methods investigated in this study. This scheme illustrates the relationships between the indices of phylogenetic reliability associated with the four methods: Bayesian posterior probabilities (PP), maximum likelihood nonparametric bootstrap proportions (BP<sub>ML</sub>), Bayesian nonparametric bootstrap proportions (BP<sub>Bay</sub>), and reliability percentages (RP).

to minimize the risk of making inferences before the convergence of the chains. A consensus network of the remaining 5000 trees was constructed using a 10% threshold, which implies that only splits that appeared in more than 500 of the 5000 trees were included (Fig. 3A).

For the bootstrap-based analyses, we generated 100 nonparametric bootstrap pseudoreplicates of the original data set using the program SEQBOOT of the PHYLIP version 3.6a3 package (Felsenstein, 2003). These bootstrap data sets were then subsequently analyzed under the maximum likelihood criterion with PAUP\* version 4.0b10 (Swofford, 2002) using the parameters of the GTR+I+ $\Gamma_4$  model previously estimated by Modeltest on the original data set. The ML heuristic searches were conducted with a neighbor-joining (NJ) starting tree and the tree bisection reconnection branch-swapping algorithm. A consensus network of the 100 resulting ML trees was then constructed using a 10% threshold (Fig. 3B). The same 100 bootstrap data sets were also analyzed under the Bayesian approach as described in Douady et al. (2003). For each data set the Bayesian analysis was performed under exactly the same conditions as the original data set described above. These analy-

ses resulted in 10,000 MCMC-sampled trees for each bootstrap data set; the first half of each run was discarded as the burn-in. A consensus network of the  $100 \times 5000 = 500,000$  remaining trees was constructed using a 10% threshold (Fig. 3C).

The ML quartet puzzling method was implemented using the program Tree-Puzzle version 5.0 (Schmidt et al., 2002). Quartets were analyzed under the TN+I +  $\Gamma_4$  model of sequence evolution (the most general model available). The model parameters were estimated during the search and the analysis resulted in 10,000 quartet puzzling trees. A consensus network of these 10,000 was constructed using a both 10% and 25% thresholds. However, because the consensus network obtained with the 10% threshold was rather complex we chose to present the 25% threshold network (Fig. 3D).

The consensus networks computed for the collections of trees obtained by the four strategies are presented in Figure 3. The consensus network built using the trees resulting from the Bayesian analysis is very treelike, displaying only two 2-cubes (Fig. 3A). All internal edges are fully resolved except the relationships among Paenungulata (elephant, hyrax, and sirenian) within Afrotheria and the interrelationships of Ostentoria (pangolin and

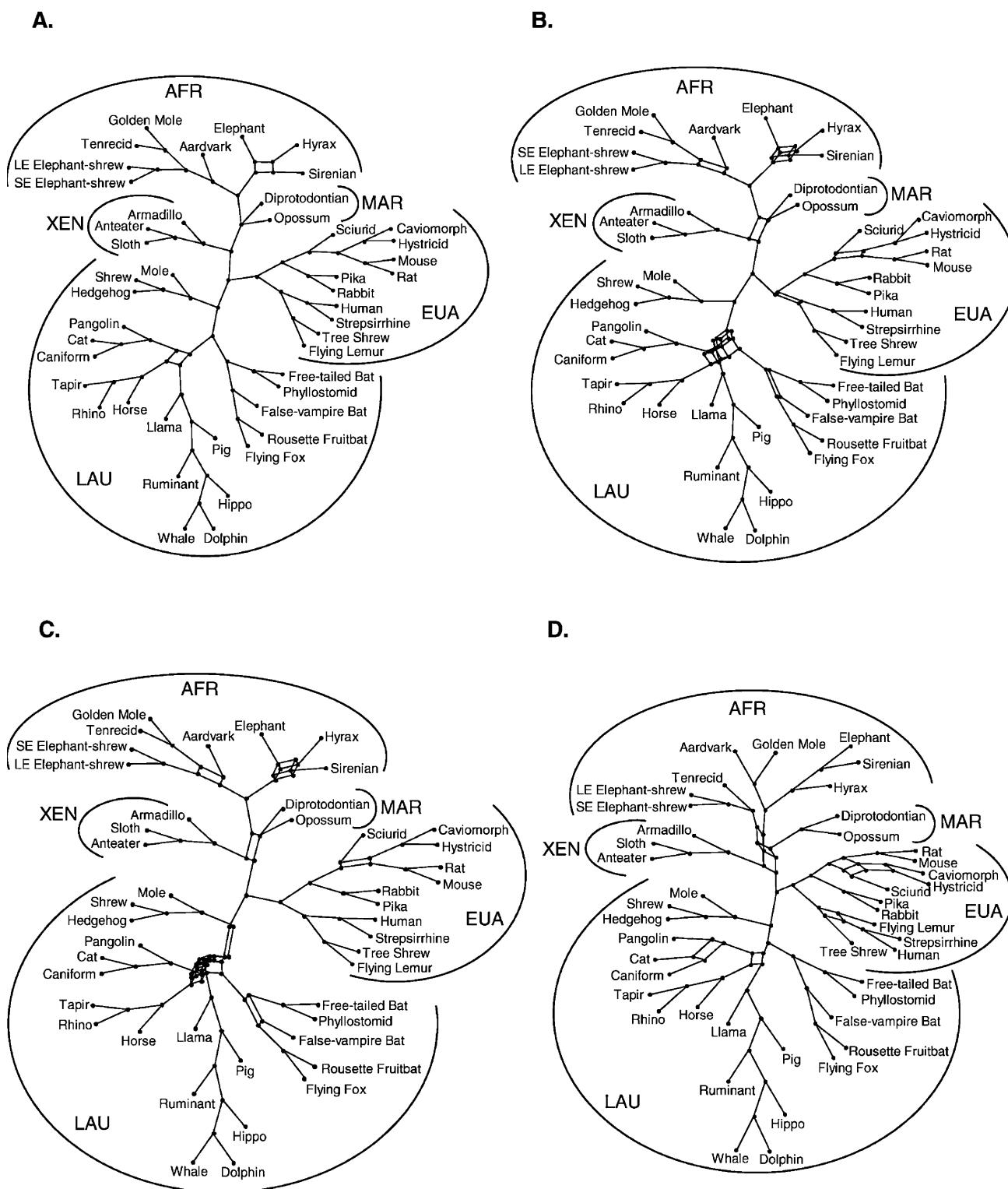


FIGURE 3. Consensus networks obtained from the analysis of the collections of trees produced by the four different methods for the placental mammal data set of Murphy et al. (2001b). The name and composition of the four major placental groups are indicated: Afrotheria (AFR), Xenarthra (XEN), Euarchontoglires (EUA), and Laurasiatheria (LAU), as well as the two marsupial outgroups (OUT). (A) Consensus network (10% threshold) of the 5000 MCMCMC-sampled trees computed in the Bayesian analysis of the original data set. These trees are usually used to derive Bayesian posterior probabilities (PP). (B) Consensus network (10% threshold) of the 100 trees computed in the maximum likelihood analyses of the 100 nonparametric bootstrap-resampled data sets. These trees are usually used to compute maximum likelihood bootstrap proportions (BP<sub>ML</sub>). (C) Consensus network (10% threshold) of the 500,000 trees computed in the Bayesian analyses of the 100 nonparametric bootstrap-resampled data sets. These trees can be used to calculate Bayesian bootstrap proportions (BP<sub>Bay</sub>). (D) Consensus network (25% threshold) of the 10,000 trees computed in the quartet puzzling analysis of the original data set. These trees are usually used to obtain reliability percentages (RP).

carnivores), Perissodactyla (horse, rhino, and tapir), and Cetartiodactyla (llama, pig, ruminant, hippo, and whale) within Laurasiatheria. Unlike the 50% majority-rule consensus tree usually computed to obtain PP, the consensus network allows the identification of two main competing hypotheses among the three possible schemes of interrelationships for each of the two polytomies. As pointed out in the original publication (Murphy et al., 2001b), the Bayesian analysis seems to resolve the central question of the position of the root by supporting the early emergence of Afrotheria within placentals.

In contrast, the ML bootstrap consensus network reveals a congruent but less resolved picture of the placental mammal's phylogenetic relationships with the occurrence of both 2-cubes and 3-cubes (Fig. 3B). Some of the 2-cubes such as the one observed within Chiroptera (bats) are elongated rectangles, indicating that one of the two alternative hypotheses is only marginally supported by the data. This is also the case in the 3-cube depicting the relationships among paenungulates where the hypothesis of a close relationship between the elephant and the sirenian appears much less likely than the other two, which are almost equally supported. However, the occurrence of multiple 3-cubes within Laurasiatheria indicates that the relationships among Chiroptera, Ostentoria, Perissodactyla, and Cetartiodactyla are left unresolved by the ML bootstrap analysis. Interestingly, the ML bootstrap consensus network reveals that two alternative hypotheses for the position of the root coexisted in this data set. The early emergence of Afrotheria remains the most likely, but the hypothesis of a close relationship between Afrotheria and Xenarthra, named the Atlantogenata hypothesis (Waddell et al., 1999), appears in some of the bootstrap trees. The third hypothesis of an early emergence of Xenarthra within placentals, the classical morphologically based Epitheria hypothesis (McKenna, 1975), is apparently not supported by this data. Given the contrasting levels of resolution observed between the Bayesian and ML bootstrap consensus networks, the consensus network visualization tends to support the observation that posterior probabilities are generally higher than the corresponding bootstrap values. This illustrates the fact that the collection of trees sampled in the Bayesian analysis are more restricted in the treespace than the bootstrap trees.

The application of the bootstrap to the Bayesian method yielded a consensus network very similar to the ML bootstrap consensus network (Fig. 3C). Minor differences between the two networks involve the appearance in the Bayesian bootstrap consensus network of an additional 2-cube for the position of Eulipotyphla (hedgehog, mole, and shrew) within Laurasiatheria and the disappearance of the 2-cube linking primates to the grouping of tree shrew and flying lemur. Apart from these differences, the two networks agree on the lack of resolution for the relationships within Laurasiatheria and among Paenungulata. The same hierarchy of hypotheses is observed for the position of the root, with the early emergence of Afrotheria being favored over Atlantogenata. These results support the fact that the

application of the bootstrap resampling to the Bayesian method leads to values comparable to the ML bootstrap ones. This confirms the utility of this approach for computing bootstrap support values in a Bayesian framework, allowing the implementation of complex models of sequence evolution (Waddell et al., 2002; Douady et al., 2003).

In striking contrast to the other methods, the quartet puzzling-based consensus network computed with a 10% threshold was highly unresolved and complicated by the occurrence of hypercubes (data not shown). Therefore, we present the consensus network obtained with a more stringent 25% threshold (Fig. 3D). This led to a more treelike consensus network roughly compatible with the ones obtained for the other methods. However, some of the 2-cubes do not appear in the other consensus networks, such as the one disrupting the otherwise strongly supported monophyly of Afrotheria or two additional 2-cubes within rodents. It is noteworthy that the position of the root appears particularly unresolved with this method compared to the others. This might reflect the fact that phylogenetic reconstruction based on quartets of species is highly sensitive to rate variation between the taxa involved in the quartet (Philippe and Douzery, 1994; Adachi and Hasegawa 1996a; Ranwez and Gascuel, 2001). Indeed, the marsupials used here as outgroups are very divergent from the placentals in terms of base composition and substitution patterns, likely leading to tree-reconstruction artefacts in the quartets in which they are involved. The presence of quartets that support conflicting positions of marsupials relative to the placentals might be responsible for the observed 2-cubes at the root of the placental tree. The extreme sensitivity of quartet-based methods of phylogenetic reconstruction to systematic biases such as long-branch attraction phenomena (Felsenstein, 1978; Hendy and Penny, 1989) and their tendency to lead to too-symmetrical topologies have been shown to be responsible for the general limitations of this kind of method (Ranwez and Gascuel, 2001).

The results obtained from the application of the consensus network method highlight the remaining uncertainties in the placental mammal phylogeny by allowing the visualization of competing hypotheses coexisting in the Murphy et al. (2001b) data set for different edges (Fig. 3). Hence, one of the most puzzling questions remains the branching order between Ostentoria (pangolins and carnivores), Perissodactyla (odd-toed ungulates), and Cetartiodactyla (even-toed ungulates) with two competing hypotheses appearing as equally likely even based on the Bayesian analysis: Ostentoria + Perissodactyla or Cetartiodactyla + Perissodactyla. Similarly, the tricky problem of paenungulates relationships is left unresolved by the sequence data with the classical morphologically based Tethytheria hypothesis (McKenna, 1975) of a close relationship between elephants and sirenians in fact appearing to be the only one to be strongly rejected (Amrine-Madsen et al., 2003). However, the central question remaining is the position of the root. Indeed, the rooting of the placental tree is still unstable mainly

because of the high degree of divergence observed between marsupials and placentals (Delsuc et al., 2002). Recent additional sequence data still fail to statistically resolve the question (Amrine-Madsen et al., 2003). The study of indels in protein sequences has been shown to prove useful by constraining the rooting possibilities (de Jong et al., 2003). The study of rare genomic changes such as protein sequence signatures or indels and SINEs insertions might also be required to distinguish between the alternative hypotheses revealed by the consensus network method in order to firmly resolve the remaining uncertainties in the phylogeny of placental mammals.

#### *Amino Acids versus Nucleotides and the Origin of Hexapods*

Recent analyses of mitochondrial data have raised a controversy on the origins of six-legged arthropods (hexapods). Analyzing a concatenation of the four most conserved genes of the mitochondrial genome (COXI, COXII, COXIII, CYTB), Nardi et al. (2003a) found strong support for the paraphyly of hexapods due the emergence of two collembolan species before crustaceans and the other hexapod taxa. These results imply that the colonization of land by hexapods happened at least twice, and that the developmental and morphological features shared by collembolans and others hexapods consequently arose by convergence (Nardi et al., 2003a; Thomas, 2003).

However, it has been shown that nucleotide-based phylogenetic analyses of the initial 35-taxa data set of Nardi et al. (2003a) can support the retrieval of the respective monophyly of insects and hexapods when RY (purine-pyrimidine) coding is used to alleviate the compositional bias at 1st and 3rd codon positions (Delsuc et al., 2003). Furthermore, the analysis of nucleotides with RY-coded 3rd codon positions from a 25-taxa data set—excluding both closely related and rapidly evolving taxa—confirmed the previous results in supporting hexapod monophyly with a high Bayesian posterior probability (0.99) but with an ML bootstrap value of only 57 (Delsuc et al., 2003). The Bayesian analysis of the amino acids of this 25-taxa data set yielded a tree where hexapods were still paraphyletic but with a moderate posterior probability value of 0.70 (Nardi et al., 2003b). Although not statistically decisive, these analyses showed that two contradictory signals concerning the origin of extant hexapods seemed to coexist in this mitochondrial data set and that their relative strengths appeared to depend upon the use of amino acid or nucleotide characters. Here, we have used the consensus network representation to examine the extent of the conflicting signals contained in this data set by analyzing the collection of trees produced by ML bootstrap and Bayesian analyses of the nucleotides and corresponding amino acids of the 25-taxa data set of Delsuc et al. (2003).

To make things comparable, the 25-taxa data set of Delsuc et al. (2003) containing 3777 aligned nucleotide sites was translated into the 1259 corresponding amino acid sites using MEGA version 2.1 (Kumar et al., 2001). The Bayesian analysis of the nucleotide data set with

RY-coded 3rd codon positions was conducted with MrBayes, using a partitioned-likelihood model attributing a GTR+I+ $\Gamma_8$  model to the 1st and 2nd codon positions and a two-state substitution model I+ $\Gamma_8$  to the RY-coded 3rd codon positions. For the amino acid data set, we used the empirical mitochondrial MtRev substitution matrix I+ $\Gamma_8$  (Adachi and Hasegawa, 1996b). For both data sets, five incrementally heated MCMCMC chains were simultaneously run for 1,000,000 generations, using the program's default uninformative priors as starting values for trees, edge lengths, and model parameters. Trees were sampled every 10 generations, leading to a collection of 100,000 MCMCMC-sampled trees whose first 50,000 were discarded as a the burn-in period of the chains. Consensus networks of the remaining 50,000 trees were constructed using a 10% threshold, which implies that only splits that appear in more than 5000 of the 50,000 trees were displayed.

ML bootstrap analyses of the two data sets were performed by analyzing 100 bootstrap-resampled data sets generated by SEQBOOT using PHYLML version 2.0.3 (Guindon and Gascuel, 2003). The use of this very fast ML program allowed parameters to be estimated for each replication in a reasonable computation time. As 3rd codon positions were RY-coded in the nucleotide data set, we independently generated the bootstrap pseudoreplicates from each of the three codon positions using the same random seed in SEQBOOT. This ensured that the proportion of each codon position was kept constant in each bootstrap-resampled data set. These concatenated bootstrap data sets were then analyzed under the GTR+I+ $\Gamma_8$  model with PHYLML. The amino acid bootstrap-resampled data sets were analyzed under the MtRev+I+ $\Gamma_8$  model. Consensus networks of the 100 ML bootstrap trees obtained for each of the two data sets were constructed using a 10% threshold.

The consensus networks obtained for the ML bootstrap and Bayesian analyses of the amino acids and nucleotides of the 25-taxa data set are compared in Figure 4. The Bayesian analysis of both amino-acid and nucleotide data sets yielded notably treelike consensus networks (Fig. 4A, B), compared to the ML bootstrap analyses of the same data sets that led to consensus networks with many 3-cubes (Fig. 4C, D). As seen in the previous example, this highlights the differences between the two methods, the collections of trees sampled by the Bayesian analysis being much more restricted in tree space than the ML bootstrap ones. The question of the monophyletic or paraphyletic origin of hexapods relies on the interrelationships between crustaceans (CRU), collembolans (COL), and insects (INS), for which classical phylogenetic analyses of nucleotides and amino acids appears to conflict in supporting hexapod monophyly (Delsuc et al., 2003) or paraphyly (Nardi et al., 2003b), respectively. The consensus network representation sheds light on this apparent conflict. It reveals that if the Bayesian analysis of nucleotides indeed supports hexapods monophyly with a split clearly separating crustaceans from collembolans and insects (Fig. 4B), the collection of trees coming from the Bayesian analysis of

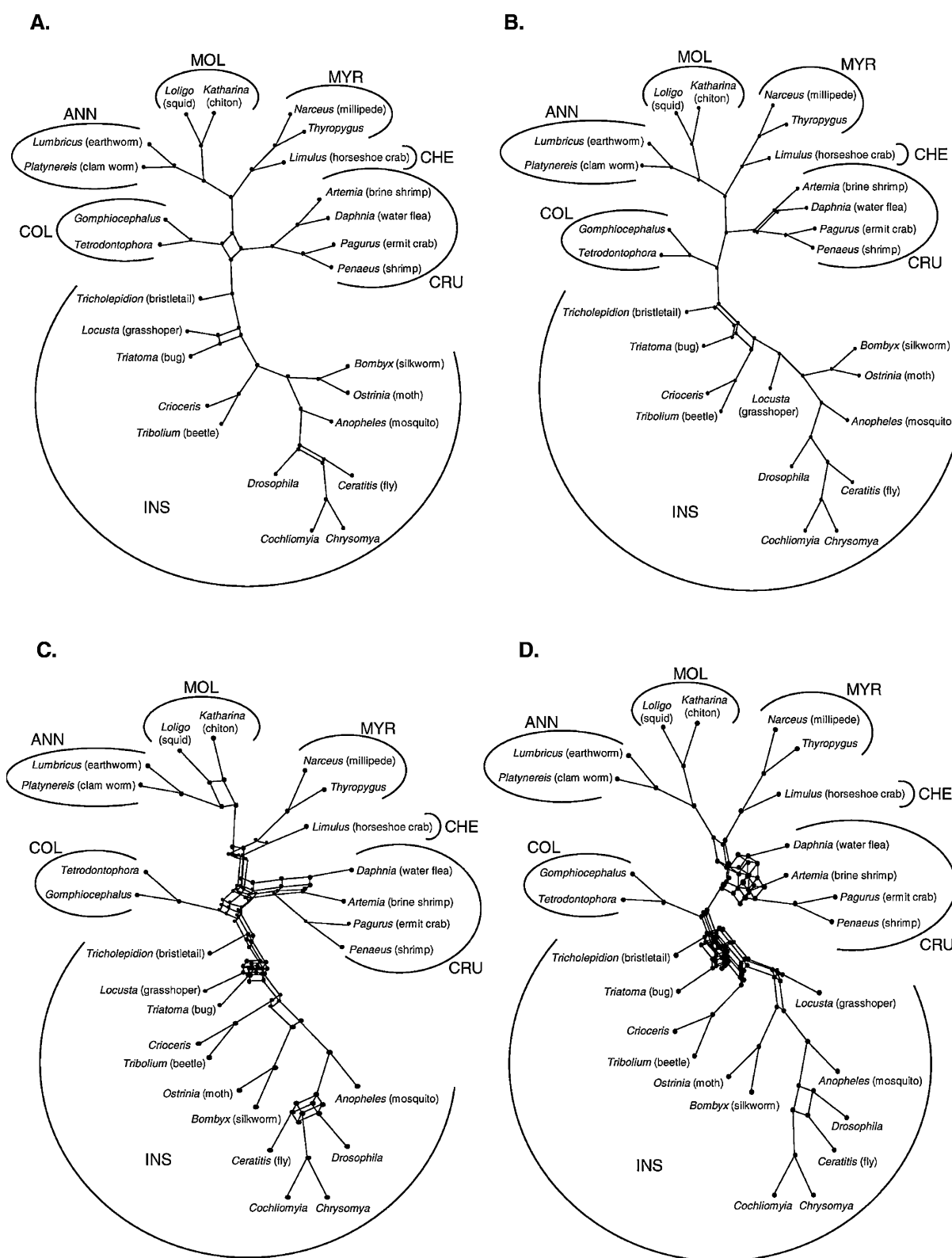


FIGURE 4. Consensus networks obtained using a 10% threshold from the analysis of the collections of trees produced by the ML bootstrap and Bayesian analyses of amino acids and nucleotides of the 25-taxa arthropod data set used in Delsuc et al. (2003). The name and composition of the different groups are indicated. (A) Consensus network of the 50,000 MCMCMC-sampled trees computed in the Bayesian analysis of the amino acid data set. (B) Consensus network of the 50,000 MCMCMC-sampled trees computed in the Bayesian analysis of the nucleotide data set with RY-coded 3rd-codon positions. (C) Consensus network of the 100 trees computed in the maximum likelihood analyses of the 100 nonparametric bootstrap-resampled data sets of the amino acid data set. (D) Consensus network of the 100 trees computed in the maximum likelihood analyses of the 100 nonparametric bootstrap-resampled data sets of the nucleotide data set with RY-coded 3rd-codon positions.



amino acids contains two competing signals pertaining to the origin of hexapods (Fig. 4A). Indeed, the consensus network displays a 2-cube (Fig. 4A) according almost the same credit to the hypothesis of hexapod monophyly (COL+INS) as to their paraphyly (CRU+INS). The comparison of the ML bootstrap consensus networks also shows that there is less signal pertaining to the question of hexapod origins in the amino acid data set (Fig. 4C) than in the nucleotide data (Fig. 4D). In fact, in the amino acid consensus network, the relationships between crustaceans, collembolans, and insects are depicted by a complex network (Fig. 4C), whereas the nucleotide consensus network reveals a split in favor of hexapod monophyly that clearly separates crustaceans from collembolans and insects (Fig. 4D) as in the Bayesian nucleotide consensus network (Fig. 4B).

Overall, the consensus network results on arthropods phylogeny indicate that there is simply not enough signal in these data sets, especially in the amino acid data, to adequately resolve the position of collembolans. This result is not necessarily surprising given the relatively short length of the data sets (3777 nucleotide or 1259 amino acid sites) and the divergence level under scrutiny (Philippe and Laurent, 1998). Indeed, mitochondrial genes are likely to contain sites that have undergone many substitutions, causing difficulties for phylogenetic reconstruction between the major arthropod lineages. Such difficulties have already been observed for much younger divergences such as the ones between placental mammal orders (Springer et al., 2001). These considerations have led to the preferential analysis of amino acid characters for this data set (Nardi et al., 2003a, 2003b). However, it has been demonstrated that amino acid characters do not necessarily outperform nucleotides even at deep phylogenetic levels (Simmons et al., 2002), especially when there is compositional bias in the data (Foster and Hickey, 1999). These results are in agreement with the fact that RY-coding the mtDNA nucleotide data sets of Nardi et al. (2003) produces the more generally accepted result of arthropod monophyly (Delsuc et al., 2003). It is thus worth noting that the consensus network of the relatively conservative ML bootstrap analysis of nucleotides displays a split that supports the classically admitted monophyly of hexapods by clearly separating them from crustaceans (Fig. 4D). For this reason, it appears safe to consider the hypothesis of hexapod paraphyly (Nardi et al., 2003a) as no more than a working hypothesis that needs to be tested further by the future analysis of more slowly evolving nuclear genes. Interestingly, recent improved analyses of the nuclear small subunit rRNA (18S) retrieved the monophyly of hexapods including collembolans, especially when using rRNA specific alignment and substitution models (Kjer, 2004).

## DISCUSSION

We have introduced a consensus method that extends the notion of strict and majority consensus trees to allow the display of conflicting evolutionary hypotheses

within a collection of trees using a network. This method can be used in conjunction with many existing phylogenetic techniques that generate large sets of trees such as bootstrapping, Bayesian inference, and quartet puzzling. It might also be applicable in case the trees are generated from multiple data sets (such as gene trees), but this requires additional analysis, which will be presented elsewhere. Consensus networks are intended as a visualization tool rather than as a confidence set of trees (although they might be a useful display of a confidence set). Their interpretation differs depending on the source of the trees being displayed. For instance, a consensus network of bootstrap trees has a different interpretation to a consensus network of trees from an MCMC Bayesian analysis, which has a different interpretation again to a set of equally scoring parsimony trees.

Although consensus networks, like other network methods such as median networks and split decomposition, are capable of displaying incompatible split systems, unlike median networks they do not become too visually complex. This is because the dimensionality of the consensus network can be controlled by choosing an appropriate threshold value. The worst-case complexity is described in the Appendix, but in practice we found that it was easy to use Spectronet and jSplits to experiment with thresholds that produced 2- or 3-dimensional networks, and that the thresholds used were typically much lower than the theoretical worst case would suggest. For example, many of the networks presented here use a threshold of 10%, which means they could in theory contain 9-dimensional hypercubes; however, in practice this is not observed. One way to quickly choose an appropriate threshold value would be to examine the output produced by PAUP\* bootstrap runs or the MrBayes *sumt* command to see how many splits have above, say, 20% or 10% support.

Consensus networks and split decomposition (Bandelt and Dress, 1992) could be used in combination, as the two methods have different limitations and advantages. Split decomposition represents distance data as a set of weighted weakly compatible splits, where the weight for each split is a function that is minimized over all quartets (subsets of size four). The nature of this function means that for large data sets the method tends to produce star-like topologies with little resolution. In addition, split decomposition will never produce a split system with three mutually incompatible splits. Consensus networks do not have these limitations but they could give a misleading picture when tree estimation suffers from systematic bias. For example, in the "Felsenstein zone" for parsimony (Felsenstein, 1978) you might get 100% bootstrap support for the wrong tree; in this case the consensus network of the bootstrap trees would show no conflict but a split decomposition network could display that the data do have conflicting signal.

We applied the consensus network technique to two biological data sets. Consensus networks helped identify areas of uncertainty that remain in the placental mammal phylogeny and understand to what degree the competing hypotheses have support in the data of Murphy

et al. (2001b). The networks make it possible to compare the different levels of support for clades that are conferred using either an ML bootstrap or Bayesian approach. The results support the findings of various authors (Waddell et al., 2002; Wilcox et al., 2002; Suzuki et al., 2002; Alfaro et al., 2003; Cummings et al., 2003; Douady et al., 2003; Erixon et al., 2003; Simmons et al., 2004) that Bayesian posterior probabilities are generally more extreme than the corresponding ML bootstrap values. The consensus network approach also confirms that the method independently suggested by Waddell et al. (2002) and Douady et al. (2003) of computing a bootstrap-based measure of phylogenetic reliability in a Bayesian framework represents a valuable alternative to posterior probabilities. Consensus networks also shed light on apparently conflicting results for hexapod monophyly using either nucleotide or amino acid characters. As the networks are capable of displaying conflicting hypotheses, it was easy to see that the amino acid data has quite strong support for both trees where hexapods are monophyletic and trees where they have paraphyletic origins. This information would have been lost in a standard consensus tree analysis in which the two data sets would have appeared to be in conflict.

Many phylogenetic techniques expend significant computational effort to generate large sets of trees; it is not only wasteful to throw away much of the information by restricting their display to a single tree, but it can also be misleading because it is the conflicting signals that are often of most interest to evolutionary biologists. For example, this kind of approach might be useful for representing multiple gene trees, particularly for organisms such as bacteria and archaea, whose genomes underwent numerous horizontal gene transfer events and for which the adequate representation might be a phylogenetic tree of the "core" untransferred genes nested within a network of all genes (Philippe and Douady, 2003).

#### ACKNOWLEDGMENTS

Calculations were conducted using the Helix supercomputing facility of the Allan Wilson Centre for Molecular Ecology and Evolution, consisting of a cluster of 66 dual-processor AMD Athlon computing nodes (<http://awcmee.massey.ac.nz/helix-service.htm>). Tim White provided invaluable assistance in setting up the calculations on the Helix supercomputer. Mark S. Springer kindly sent the aligned mammalian data set. Christophe Douady and Emmanuel Douzery provided helpful comments on early versions of the manuscript. Suggestions from Chris Simon, Allan J. Baker, and George F. Barrowclough improved the latest version of the manuscript. The authors thank The Swedish Foundation for International Cooperation and Research in Higher Education (STINT), and The Linnaeus Centre for Bioinformatics for their support. FD acknowledges the support of a Lavoisier Postdoctoral Grant from the French Ministry of Foreign Affairs.

#### REFERENCES

Adachi, J., and M. Hasegawa. 1996a. Instability of quartet analyses of molecular sequence data by the maximum likelihood method: The Cetacea/Artiodactyla relationships. *Mol. Phylogenet. Evol.* 6:72–76.  
Adachi, J., and M. Hasegawa. 1996b. Model of amino acid substitution in proteins encoded by mitochondrial DNA. *J. Mol. Evol.* 42:459–468.  
Alfaro, M. E., S. Zoller, and F. Lutzoni. 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.  
Amrine-Madsen, H., K. P. Koepfli, R. K. Wayne, and M. S. Springer. 2003. A new phylogenetic marker, apolipoprotein B, provides compelling evidence for eutherian relationships. *Mol. Phylogenet. Evol.* 28:225–240.  
Bandelt, H.-J. 1994. Phylogenetic Networks. *Verhandl. Naturwiss.* 34:51–71.  
Bandelt, H.-J. 1995. Combination of data in phylogenetic analysis. *Plant Syst. Evol. Suppl.* 9:355–361.  
Bandelt, H.-J., and A. Dress. 1992. A canonical decomposition theory for metrics on a finite set. *Adv. Math.* 92:47–105.  
Bandelt, H.-J., P. Forster, B.C. Sykes, and M.B. Richards. 1995. Mitochondrial portraits of human populations using median networks. *Genetics* 14:743–753.  
Bryant, D. 2003. A classification of consensus methods for phylogenetics. Pages 1–21 in *Bioconsensus* (M. Janowitz, F. J. Lapointe, F. McMorris, B. Mirkin and F. Roberts, eds.). DIMACS-AMS, Providence.  
Cao, Y., J. Adachi, and M. Hasegawa. 1998. Comment on the quartet puzzling method for finding maximum-likelihood tree topologies. *Mol. Biol. Evol.* 15:87–89.  
Cummings, M. P., S. A. Handley, D. S. Myers, D. L. Reed, A. Rokas, and K. Winka. 2003. Comparing bootstrap and posterior probability values in the four-taxon case. *Syst. Biol.* 52:477–487.  
de Jong, W. W. 1998. Molecules remodel the mammalian tree. *Trends Ecol. Evol.* 13:270–275.  
de Jong, W. W., M. A. van Dijk, C. Poux, G. Kappe, T. van Rheede, and O. Madsen. 2003. Indels in protein-coding sequences of Euarchontoglires constrain the rooting of the eutherian tree. *Mol. Phylogenet. Evol.* 28:328–340.  
Delsuc, F., M. J. Phillips, and D. Penny. 2003. Comment on "Hexapod origins: monophyletic or paraphyletic?" *Science* 301:1482d.  
Delsuc, F., M. Scally, O. Madsen, M. J. Stanhope, W. W. de Jong, F. M. Catzefflis, M. S. Springer, and E. J. P. Douzery. 2002. Molecular phylogeny of living xenarthrans and the impact of character and taxon sampling on the placental tree rooting. *Mol. Biol. Evol.* 19:1656–1671.  
Douady, C. J., F. Delsuc, Y. Boucher, W. F. Doolittle, and E. J. P. Douzery. 2003. Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. *Mol. Biol. Evol.* 20:248–254.  
Dress, A., M. Klucznik, J. Koolen, and V. Moulton. 2001.  $2n-1$  choose 2: A note on extremal combinatorics of cyclic split systems. *Seminaire Lotharingien de Combinatoire* 47.  
Erixon, P., B. Svennblad, T. Britton, and B. Oxelman. 2003. Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. *Syst. Biol.* 52:665–673.  
Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* 27:401–410.  
Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783–791.  
Felsenstein, J. 2003. PHYLIP (PHYLogeny Inference Package) version 3.6a3.  
Foster, P. G., and D. A. Hickey. 1999. Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. *J. Mol. Evol.* 48:284–290.  
Goldman, N. 1993. Statistical tests of models of DNA substitution. *J. Mol. Evol.* 36:725–736.  
Guindon, S., and O. Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52:696–704.  
Hendy, M. D., and D. Penny. 1989. A framework for the quantitative study of evolutionary trees. *Syst. Zool.* 38:297–309.  
Hillis, D. M., and J. J. Bull. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol.* 42:182–192.  
Holland, B., K. Huber, V. Moulton, and P.J. Lockhart. 2004. Using consensus networks to visualize contradictory evidence for species phylogeny. *Mol. Biol. Evol.* 21:1459–1461.  
Holland, B., and V. Moulton. Consensus networks: A method for visualising incompatibilities in collections of trees. Pages 165–176 in *WABI 2003*. Springer-Verlag, Berlin, Heidelberg, Budapest.

- Huber, K. T., M. Langton, D. Penny, V. Moulton, and M. Hendy. 2002. Spectronet: A package for computing spectra and median networks. *App. Bioinformatics* 1:159–161.
- Huber, K. T., V. Moulton, P. J. Lockhart, and A. Dress. 2001. Pruned median networks: A technique for reducing the complexity of median networks. *Mol. Phylogenet. Evol.* 19:302–310.
- Huelsenbeck, J. P., B. Larget, R. E. Miller, and F. Ronquist. 2002. Potential applications and pitfalls of Bayesian inference of phylogeny. *Syst. Biol.* 51:673–688.
- Huelsenbeck, J. P., and B. Rannala. 2004. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Syst. Biol.* In press.
- Huson, D. 1998. SplitsTree: A program for analyzing and visualizing evolutionary data. *Bioinformatics* 14:68–73.
- Kjer, K. M. 2004. Aligned 18S and insect phylogeny. *Syst. Biol.* 53:506–514.
- Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei. 2001. MEGA2: Molecular evolutionary genetics analysis software. *Bioinformatics* 17:1244–1245.
- Lemmon, A. R., and E. C. Moriarty. 2004. The importance of proper model assumption in Bayesian phylogenetics. *Syst. Biol.* 53:265–277.
- Madsen, O., M. Scally, C. J. Douady, D. J. Kao, R. W. DeBry, R. Adkins, H. M. Amrine, M. J. Stanhope, W. W. de Jong, and M. S. Springer. 2001. Parallel adaptative radiations in two major clades of placental mammals. *Nature* 409:610–614.
- McKenna, M. C. 1975. Toward a phylogenetic classification of the Mammalia. Pages 21–46 in *Phylogeny of the primates* (W. P. Luckett, and F. S. Szalay, eds.). Plenum Press, New York.
- Murphy, W. J., E. Eizirik, W. E. Jonhson, Y. P. Zhang, O. A. Ryder, and S. J. O'Brien. 2001a. Molecular phylogenetics and the origins of placental mammals. *Nature* 409:614–618.
- Murphy, W. J., E. Eizirik, S. J. O'Brien, O. Madsen, M. Scally, C. J. Douady, E. C. Teeling, O. A. Ryder, M. J. Stanhope, W. W. de Jong, and M. S. Springer. 2001b. Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* 294:2348–2351.
- Nardi, F., G. Spinsanti, J. L. Boore, A. Carapelli, R. Dallai, and F. Frati. 2003a. Hexapod origins: Monophyletic or paraphyletic? *Science* 299:1887–1889.
- Nardi, F., G. Spinsanti, J. L. Boore, A. Carapelli, R. Dallai, and F. Frati. 2003b. Response to Comment on "Hexapod Origins: Monophyletic or Paraphyletic?" *Science* 301:1482e.
- Penny, D., and M. Hendy. 1986. Estimating the reliability of evolutionary trees. *Mol. Biol. Evol.* 3:403–417.
- Philippe, H., and C. J. Douady. 2003. Horizontal gene transfer and phylogenetics. *Curr. Opin. Microbiol.* 6:498–505.
- Philippe, H., and E. Douzery. 1994. The pitfalls of molecular phylogeny based on four species, as illustrated by the Cetacea/Artiodactyla relationships. *J. Mammal. Evol.* 2:133–151.
- Philippe, H., and J. Laurent. 1998. How good are deep phylogenetic trees? *Curr. Opin. Genet. Dev.* 8:616–623.
- Posada, D., and K. A. Crandall. 1998. MODELTEST: Testing the model of DNA substitution. *Bioinformatics* 14:817–818.
- Ranwez, V., and O. Gascuel. 2001. Quartet-based phylogenetic inference: Improvements and limits. *Mol. Biol. Evol.* 18:1103–1116.
- Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.
- Schmidt, H. A., K. Strimmer, M. Vingron, and A. von Haeseler. 2002. TREE-PUZZLE: Maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18:502–504.
- Simmons, M. P., H. Ochoterena, and J. V. Freudenstein. 2002. Amino acid vs. nucleotide characters: challenging preconceived notions. *Mol. Phylogenet. Evol.* 24:78–90.
- Simmons, M. P., K. M. Pickett, and M. Miya. 2004. How meaningful are Bayesian support values? *Mol. Biol. Evol.* 21: 188–199.
- Springer, M. S., R. W. DeBry, C. J. Douady, H. M. Amrine, O. Madsen, W. W. de Jong, and M. J. Stanhope. 2001. Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. *Mol. Biol. Evol.* 18:132–143.
- Strimmer, K., and A. von Haeseler. 1996. Quartet puzzling: A quartet maximum-likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* 13:964–969.
- Suzuki, Y., G. V. Glazko, and M. Nei. 2002. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. *Proc. Natl. Acad. Sci. USA* 99:16168–16173.
- Swofford, D. L. 2002. PAUP\*. Phylogenetic analysis using parsimony (\* and other methods), version 4.0b10. Sinauer Associates, Sunderland, Massachusetts.
- Swofford, D. L., G. J. Olsen, P. J. Waddell, and D. M. Hillis. 1996. Phylogenetic inference. Pages 407–514 in *Molecular systematics* (D. M. Hillis, C. Moritz, and B. K. Mable, eds.). Sinauer Associates, Sunderland, Massachusetts.
- Thomas, R. H. 2003. Wingless insects and plucked chickens. *Science* 299:1854–1855.
- Waddell, P. J., H. Kishino, and R. Ota. 2002. Very fast algorithms for evaluating the stability of ML and Bayesian phylogenetic trees from sequence data. *Genome Inform.* 13:82–92.
- Waddell, P. J., N. Okada, and M. Hasegawa. 1999. Towards resolving the interordinal relationships of placental mammals. *Syst. Biol.* 48:1–5.
- Wilcox, T. P., D. J. Zwickl, T. A. Heath, and D. M. Hillis. 2002. Phylogenetic relationships of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. *Mol. Phylogenet. Evol.* 25:361–371.

First submitted 17 December 2003; reviews returned 3 August 2004;

final acceptance 12 September 2004

Associate Editor: Allan Baker

## APPENDIX

We briefly describe the result indicated in Consensus Network Method. Suppose that  $X$  is a finite set. A *split*  $A|B$  of  $X$  is a bipartition of  $X$ ; i.e., a partition of  $X$  into two nonempty sets  $A$  and  $B$  with  $A \cup B = X$  and  $A \cap B = \emptyset$ . We call a collection of splits a *split system* for short. A split system is called *compatible* if there is a phylogenetic tree that displays every split in the system (see Consensus Network Method). We say that a split system is *incompatible* if it contains no subset of cardinality two that is compatible. Note that a split system that is not compatible need not be incompatible. For  $k$  a positive integer, we say that a split system is *k-compatible* if it contains no incompatible subset of splits with cardinality  $k + 1$ .

The concept of  $k$ -compatibility was introduced and studied in Dress et al. (2001). By definition, a  $k$ -compatible split system is compatible if and only if  $k = 1$ , in which case the median network associated to it must be a tree, but, as  $k$  increases, the associated median network can become progressively more complex. We now state the main result of Holland and Moulton (2003).

**Theorem.**—Given  $N$  phylogenetic trees each with leaf set  $X$  and some  $0 < x \leq 1$ , let  $S_x$  denote the split system containing those splits of  $X$  that are displayed in  $\lceil Nx \rceil$  or more of these trees. Then  $S_x$  is  $\lfloor 1/x \rfloor$ -compatible.

**Proof.**—Suppose that  $S_x$  contains  $\lfloor 1/x \rfloor + 1$  incompatible splits. Then, because each of these splits is displayed by at least  $\lceil Nx \rceil$  of the trees, it follows by the Pigeonhole Principle that one of the trees must display at least two of the incompatible splits. But this is impossible.

This result allows control of the complexity of the median network associated to the split system: Larger values of  $x$  imply a lower extent of incompatibility in the split system  $S_x$ . Note that in case  $x = 1$ , the median network associated to split system  $S_x$  will be the strict consensus tree associated to the phylogenetic trees, and that if  $x = 1/2$ , the associated median network will be the majority-rule consensus tree.