Software for Systematics and Evolution

Dendroscope 3: An Interactive Tool for Rooted Phylogenetic Trees and Networks

DANIEL H. HUSON[1,*] AND CELINE SCORNAVACCA[1,2]

[1]*Department of Computer Science, Center for Bioinformatics (ZBIT), University of Tübingen, 72076 Tübingen, Germany; and*
[2]*Institut des Sciences de l'Evolution Montpellier (ISEM), CNRS UMR 5554, Université Montpellier II, 34095 Montpellier, France;*
*Correspondence to be sent to: Center for Bioinformatics (ZBIT), University of Tübingen, Sand 14, 72076 Tübingen, Germany;
E-mail: daniel.huson@uni-tuebingen.de.

Abstract.—Dendroscope 3 is a new program for working with rooted phylogenetic trees and networks. It provides a number of methods for drawing and comparing rooted phylogenetic networks, and for computing them from rooted trees. The program can be used interactively or in command-line mode. The program is written in Java, use of the software is free, and installers for all 3 major operating systems can be downloaded from www.dendroscope.org. [Phylogenetic trees; phylogenetic networks; software.]

The evolutionary history of a set of species or genes is usually represented by a phylogenetic tree, and many different methods exist for the computation of such trees (Felsenstein 2004). When reticulate events such as horizontal gene transfer, hybridization, or recombination play a significant role in the history of a set of taxa, or when there are major incompatibilities in the given data, then a phylogenetic network may provide a more accurate evolutionary scenario (Doolittle 1999; Huson et al. 2010; Sneath 1975).

There are a number of established tools for computing unrooted phylogenetic networks. One popular program is SplitsTree (Huson and Bryant 2006), which contains implementations of a number of different methods for computing unrooted trees and networks, such as split decomposition (Bandelt and Dress 1992) and neighbor-net (Bryant and Moulton 2004). Another widely used method is median-joining (Bandelt et al. 1995), which is implemented in a program called Network, as well as in SplitsTree.

For rooted phylogenetic networks, the situation is slightly different. Although much work has been done on developing concepts and theoretical methods for computing rooted phylogenetic networks (for an overview, see Huson and Scornavacca 2011), the application of such ideas in biological studies has been hampered by the lack of robust and easy to use tools for their computation. Although a number of tools for calculating rooted phylogenetic networks do exist, such as Lott et al. (2009), Than et al. (2008) and Cardona et al. (2008b), they appear to be of limited utility for biologists.

In this article, we present a new user-friendly program for working with rooted phylogenetic trees and networks called Dendroscope 3, which is based on the popular tree drawing program Dendroscope 1 (Huson et al. 2007).

The program provides numerous methods for drawing and comparing rooted phylogenetic networks, and for computing such networks (including hybridization networks) from rooted trees. With this new software, we hope to provide a standard tool for computing rooted phylogenetic networks.

ROOTED PHYLOGENETIC TREES AND NETWORKS

The evolutionary history of a set of taxa (i.e., species or genes) is usually depicted as a rooted phylogenetic tree. Rooted phylogenetic networks provide a generalization of rooted phylogenetic trees that can be used to explicitly represent reticulate events or to visualize incompatibilities in a data set. By definition, a *rooted phylogenetic network* is a directed acyclic graph in which each leaf is labeled by a unique taxon and that has precisely one node that is the ancestor of all other nodes, called the *root* (Huson et al. 2010). Any node with more than 1 parent is called a *reticulation* and the edges leading into such a node are called *reticulate* edges.

Rooted phylogenetic networks can be computed using a number of different methods. Here, we briefly mention some of the approaches that build such networks from rooted phylogenetic trees or from the clusters that they contain. A cluster network is a rooted phylogenetic network that displays a given set of clusters; for example, all clusters present in a set of rooted phylogenetic trees. Cluster networks are easy to compute (Huson and Rupp 2008). Similarly, a *level-k network* can be used to represent a set of clusters and aims at minimizing the number of reticulations in any biconnected component of the network (Choy et al. 2005; van Iersel et al. 2010) and a galled network can also be used to represent

incompatible clusters (Huson et al. 2009). Given 2 rooted phylogenetic trees, a minimum *hybridization network* is a rooted phylogenetic network that contains both trees and has a minimum number of reticulations. The idea here is that reticulations correspond to speciation-by-hybridization events (Koblmueller et al. 2007; Linder and Rieseberg 2004). Such networks can be computed from bifurcating trees (Albrecht et al. 2011; Baroni et al. 2006; Bordewich et al. 2007; Chen and Wang 2010) or from multifurcating trees on unequal taxon sets (Linz and Semple 2009; Huson D.H. and Linz S, unpublished data). A rooted phylogenetic network may also be used to summarize a *multilabeled* rooted phylogenetic tree (Huber et al. 2006).

Existing Software

There exists a large number of programs for computing phylogenetic trees and for visualizing them (Felsenstein 2012). The original version of our program, Dendroscope 1 (Huson et al. 2007), was designed as a viewer for rooted phylogenetic trees and was based on a survey of existing programs in an attempt to provide all the most useful features in 1 program, in particular including the ability to draw large trees with up to 1 million nodes.

There has been much interest in recent years in developing methods for computing rooted phylogenetic networks and this has led to a number of software packages such as PhyloNet (Than et al. 2008), PADRE (Lott et al. 2009), or the Perl package Bio::PhyloNetwork (Cardona et al. 2008b). More detailed lists of existing software can be found in Huson et al. (2010) and Gambette (2012). Although these programs contain some sophisticated algorithms, they do not appear to have been extensively engineered so as to provide robust, fast, and GUI-based user-friendly tools.

Dendroscope 3

The aim of Dendroscope 3 is to provide a user-friendly tool for working with rooted phylogenetic trees and networks. To this end, all tree visualization and interactive manipulation methods of Dendroscope 1 (Huson et al. 2007) have been extended so as to also work for rooted networks. As a platform for computing rooted phylogenetic networks, our program provides a choice of algorithms for computing consensus trees and consensus networks, such as galled networks or hybridization networks, from rooted phylogenetic trees. To simplify multi-tree analyses, the program provides a number of distance calculations and can display a direct comparison of two trees or networks in terms of a tanglegram. Moreover, the program is able to show any number of trees and networks from the same data set simultaneously in the same window and a find-and-replace tool is provided that operates across all trees or networks that are shown in the same window.

All features of the program are also accessible in command-line mode and the program can be run on a cluster or cloud as part of a larger analysis pipeline. The program is fully multithreaded and a number of the computationally more demanding algorithms are implemented in a parallel fashion.

In Dendroscope 3, phylogenetic trees and networks can be loaded from a file in Newick format, in the case of trees, extended Newick format (Cardona et al. 2008a), in the case of networks, or in Nexus format. Moreover, an input dialog is provided for entering trees or networks by hand. The program uses the NeXML format (Vos et al. 2012) to save and reopen trees or networks that have been edited. In this case, the layout and formatting (colors, line width, fonts, etc.) are saved along with the trees or networks. Trees and networks can also be saved in other formats (Newick, extended Newick, or Nexus) and can be exported in a number of different graphics formats.

The main window of Dendroscope 3 can be configured to show a grid of $n \times m$ trees or networks simultaneously, thus making it easier to work with data sets that contain multiple trees or networks, such as obtained from multiple genes or by using multiple methods. Trees and networks can be displayed in a number of ways, namely as a circular, radial, or rectangular phylograms or as (an internal or external) circular, radial, rectangular, or slanted cladograms (Huson 2009) and a magnifier can be used to enlarge a part of a tree or of a network. Nodes, edges, and labels of trees and networks can be interactively formatted and edited and the program supports operations such as reshaping, rerooting, subtree/subnetwork reordering, or subtree/subnetwork extraction. One can attach images to taxa and these are then displayed next to the corresponding nodes.

As already mentioned, the program provides a choice of algorithms for computing consensus trees and consensus networks from rooted phylogenetic trees, such as the strict, majority, and loose consensus (Bryant 2003), as well as the lowest stable ancestor (LSA) consensus (Huson 2009). It also provides implementations of algorithms for computing cluster networks (Huson and Rupp 2008), minimum galled networks (Huson et al. 2009), and level-k networks with small k (van Iersel et al. 2010). If the input trees are on overlapping, but nonidentical taxon sets, then the program uses the *Z-closure* algorithm (Huson et al. 2004) to infer missing data. So, in particular, the program is able to compute the strict or majority consensus of a set of trees even if the trees do not all have the same taxon sets, although in the latter case, the result may be a rooted phylogenetic network rather than a tree. Two recent methods are available for computing hybridization networks, the first can be applied to any 2 (possibly multifurcating) rooted phylogenetic trees with overlapping, but not necessarily identical taxon sets (Huson D.H. and Linz S., unpublished data), whereas the second method is optimized to run on a pair of bifurcating trees that contain exactly the same taxa (Albrecht et al. 2011; Scornavacca C. et al., unpublished

FIGURE 1.   A tanglegram computed by Dendroscope 3 for the cpDNA and ITS trees that are reported in Figure 1 of Pirie et al. (2009).
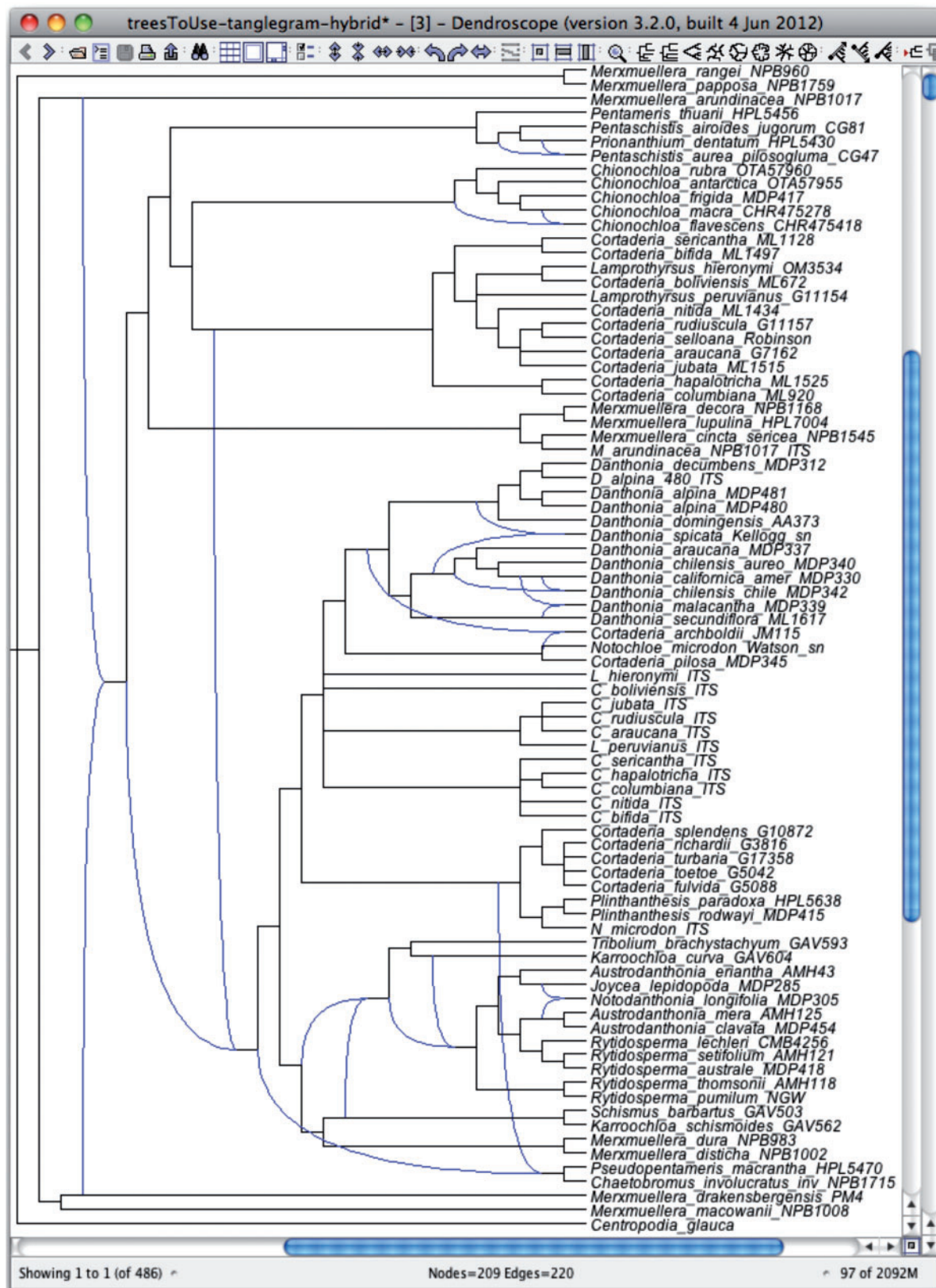
FIGURE 2. A minimum hybridization network computed by Dendroscope 3 for the cpDNA and ITS trees that are reported in Figure 1 of Pirie et al. (2009). It has 12 reticulations and is one of 486 networks calculated by the program.
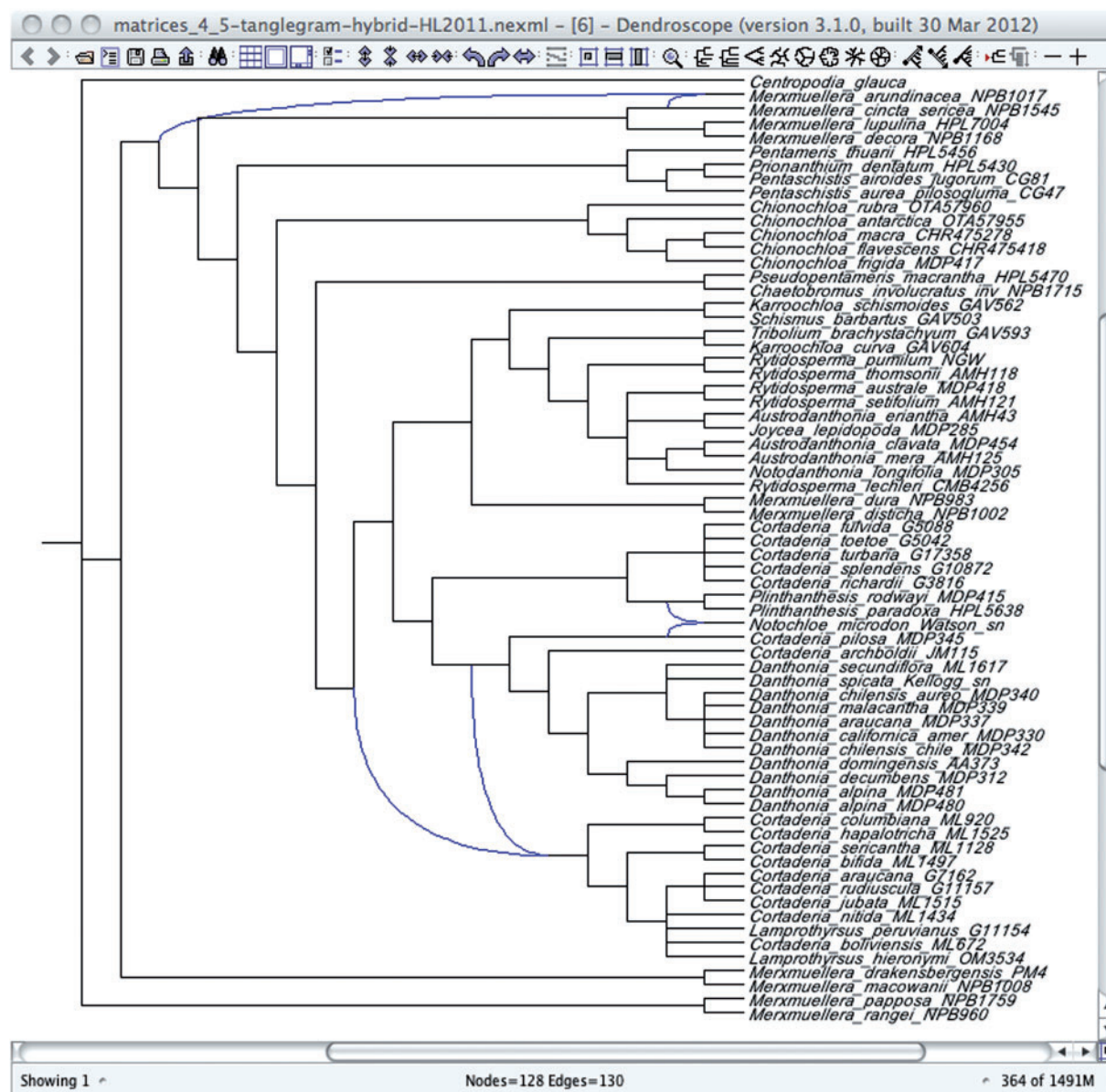
FIGURE 3.   A hybridization network computed from the 2 simplified trees that were used as input for Figure 3(b) of Pirie et al. (2009).

data). Dendroscope 3 provides 2 approaches to computing a single-labeled rooted phylogenetic network that represents a multilabeled phylogenetic tree, namely one based on nested labels (Huber et al. 2006) and another that consists of extracting all clusters from the tree and then representing them by a cluster network (Huson et al. 2010, Section 11.7).

Dendroscope 3 provides a number of distance calculations for comparing 2 rooted phylogenetic trees or networks based on the contained clusters or trees and other concepts, such as the *hardwired cluster distance* (Huson et al. 2010), the *softwired cluster distance* (Huson et al. 2010), the *displayed trees distance* (Huson et al. 2010), the *tripartition distance* (Moret et al. 2004), the *nested labels*

distance (Cardona et al. 2009; Nakhleh 2009), and the *path multiplicity distance* (Cardona et al. 2007). None of these measures is a proper metric because there exist cases in which they return a distance of zero for 2 distinct networks. Nevertheless, they are useful in practice and some of them have been shown to be proper metrics on a suitably restricted set of networks. For 2 rooted phylogenetic trees on the same taxon set, another option is to compute their hybridization distance (Albrecht et al. 2011; Bordewich et al. 2007; Huson D.H. and Linz S., unpublished data), or their *rSPR distance*.

One way to visualize similarities and differences between 2 rooted phylogenetic trees or networks is to display a *tanglegram* in which the 2 trees or networks

are drawn opposite each other and corresponding taxa are connected by lines, possibly changing the layout order of taxa in an attempt to minimize the number of line crossings. Dendroscope 3 contains a very general algorithm that can compute a tanglegram for 2 rooted trees or networks that may contain multifurcations and may have different taxon sets (Scornavacca et al. 2011).

EXAMPLES

In Pirie et al. (2009), the authors explore the potential impact of conflicting gene trees on inferences of evolutionary history above the species level, in particular studying grasses and using both chloroplast DNA (cpDNA) and nuclear ribosomal DNA (internal transcribed spacer (ITS) regions). The authors argue that the contradictions can be explained by past hybridization events, which have linked gains of complex morphologies with unrelated chloroplast lineages and have erased evidence of dispersals from the nuclear genome.

In Figure 1 of Pirie et al. (2009), the authors show a tanglegram involving a rooted phylogenetic tree computed from cpDNA and one computed from ITS sequence data. The 2 trees can be downloaded from http://ab.inf.uni-tuebingen.de/software/dendroscope/data/Pirie2009-cpDNA-ITS.txt. Although time-consuming to produce by hand, such a visualization is easily produced in Dendroscope 3 simply by loading the 2 trees and then selecting the "Tanglegram" menu item. The resulting tanglegram is shown in Figure 1 of this article.

To go beyond the analyses performed in Pirie et al. (2009), one can take the 2 trees and use Dendroscope 3 to compute a minimum hybridization network for them. In less than 5 s, the program establishes that the minimum number of reticulations is 12 and it returns a "representative set" of 486 such networks (Albrecht et al. 2011). This type of computation can form the basis of a more refined analysis. In Figure 2, we display the first network in the list.

In Figure 3(b) of Pirie et al. (2009), the authors show an unrooted hybridization network for 2 "combined gene trees," computed using SplitsTree (Huson and Bryant 2006). Reportedly, the first input tree is based on ITS data, with the cpDNA data recoded as missing data, whereas the second input tree is based on cpDNA data with ITS recoded as missing data. The trees were computed using maximum parsimony and then subjected to a 70% bootstrap threshold. They can be downloaded from http://ab.inf.uni-tuebingen.de/software/dendroscope/data/Pirie2009-Matrix4-5.txt.

A rooted version of that network is easily computed using Dendroscope 3. One must first load the 2 trees into the program and then select the "Hybridization Network" menu item. The resulting rooted phylogenetic network is shown in Figure 3 of this article.

AVAILABILITY

Dendroscope 3 is written in Java. Use of the software is free and installers for all 3 major operating systems (MacOS, Windows, and Linux) are available from http://www.dendroscope.org.

REFERENCES

Albrecht B., Scornavacca C., Cenci A., Huson D.H. 2011. Fast computation of minimum hybridization networks. Bioinformatics 28:191–197.

Bandelt H.-J., Dress A.W.M. 1992. A canonical decomposition theory for metrics on a finite set. Adv. Math. 92:47–105.

Bandelt H.-J., Forster P., Sykes B. C., Richards M.B. 1995. Mitochondrial portraits of human population using median networks. Genetics 141:743–753.

Baroni M., Semple C., Steel M. 2006. Hybrids in real time. Sys. Biol. 55:46–56.

Bordewich M., Linz S., St. John K., Semple C. 2007. A reduction algorithm for computing the hybridization number of two trees. Evol. Bioinform. 3:86–98.

Bryant D. 2003. A classification of consensus methods for phylogenies. In: Janowitz M., Lapointe F.-J., McMorris F., Mirkin B., Roberts F. editors. Providence (RI): BioConsensus, AMS, DIMACS. p. 163–184.

Bryant D., Moulton V. 2004. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. Mol. Biol. Evol. 21:255–265.

Cardona G., Llabrés M., Rosselló F., Valiente G. 2009. On Nakhleh's metric for reduced phylogenetic networks. IEEE/ACM Trans. Comput. Biol. Bioinform. 6:629–638.

Cardona G., Rosselló F., Valiente G. 2007. Comparison of tree-child phylogenetic networks. IEEE/ACM Trans. Comput. Biol. Bioinform. 6:552–569.

Cardona G., Rosselló F., Valiente G. 2008a. Extended Newick: it is time for a standard representation of phylogenetic networks. BMC Bioinformatics 9:532.

Cardona G., Rosselló F., Valiente G. 2008b. A perl package and an alignment tool for phylogenetic networks. BMC Bioinformatics 9:175.

Chen Z.-Z., Wang L. 2010. HybridNET: a tool for constructing hybridization networks. Bioinformatics 26:2912–2913.

Choy C., Jansson J., Sadakane K., Sung W.-K. 2005. Computing the maximum agreement of phylogenetic networks. Theor. Comput. Sci. 335:93–107.

Doolittle W.F. 1999. Phylogenetic classification and the universal tree. Science 284:2124–2128.

Felsenstein J. 2004. Inferring Phylogenies. Sunderland (MA) Sinauer Associates, Inc.

Felsenstein J. 2012. Tree software. Available from: http://evolution.genetics.washington.edu/phylip/software.html.

Gambette P. 2012. Who is who in phylogenetic networks—articles, authors and programs. Available from: http://www.atgc-montpellier.fr/phylnet/programs.

Huber K.T., Oxelman B., Lott M., Moulton V. 2006. Reconstructing the evolutionary history of polyploids from multilabeled trees. Mol. Biol. Evol. 23:1784–1791.

Huson D.H., Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23:254–267. Software available from: www.splitstree.org.

Huson D.H., Rupp R., Berry V., Gambette P., Paul C. 2009. Computing galled networks from real data. Bioinformatics. 25:i85–i93.

Huson D.H., Rupp R., Scornavacca C. 2010. Phylogenetic networks. Cambridge: Cambridge University Press.

Huson D.H., Rupp R. 2008. Summarizing multiple gene trees using cluster networks. In: Algorithms in bioinformatics, WABI 2008. Crandall K., Lagergren J. editors. Berlin/Heidelberg: Springer, (vol. 5251). p. 296–305.

Huson D.H., Scornavacca C. 2011. A survey of combinatorial methods for phylogenetic networks. Genome Biol. Evol. 3:23–35.

Huson D.H. 2009. Drawing rooted phylogenetic networks. IEEE/ACM Trans. Comput. Biol. Bioinform. 6:103–109.

Huson D., Dezulian T., Kloepper T., Steel M.A. 2004. Phylogenetic super-networks from partial trees. IEEE/ACM Trans. Comput. Biol. Bioinform. 1:151–158.

Huson D., Richter D., Rausch C., Dezulian T., Franz M., Rupp R. 2007. Dendroscope: An interactive viewer for large phylogenetic trees. BMC Bioinformatics 8:460. doi:10.1186/1471-2105-8-460. Software available from: www.dendroscope.org.

Koblmueller S., Duftner N., Sefc K., Aibara M., Stipacek M., Blanc M., Egger B., Sturmbauer C. 2007. Reticulate phylogeny of gastropod-shell-breeding cichlids from lake tanganyika - the result of repeated introgressive hybridization. BMC Evol. Biol. 7:7.

Linder C.R., Rieseberg L.H. 2004. Reconstructing patterns of reticulate evolution in plants. Am. J. Botany 91:1700–1708.

Linz S., Semple C. 2009. Hybridization in nonbinary trees. IEEE/ACM Trans. Comput. Biol. Bioinform. 6:30–45.

Lott M., Spillner A., Huber K.T., Moulton V. 2009. Padre: a package for analyzing and displaying reticulate evolution. Bioinformatics 25:1199–1200.

Moret B.M.E., Nakhleh L., Warnow T., Linder C.R., Tholse A., Padolina A., Sun J., Timme R. 2004. Phylogenetic networks: modeling, reconstructibility, and accuracy. IEEE/ACM Trans. Comput. Biol. Bioinform. 1:13–23.

Nakhleh L. 2009. A metric on the space of reduced phylogenetic networks. IEEE/ACM Trans. Comput. Biol. Bioinform. 99.

Pirie M.D., Humphreys A.M., Barker N.P., Linder H.P. 2009. Reticulation, data combination, and inferring evolutionary history: an example from Danthonioideae (Poaceae). Syst. Biol. 58:612–28.

Scornavacca C., Zickmann F., Huson D. 2011. Tanglegrams for rooted phylogenetic trees and networks. Bioinformatics 27: i248–i256.

Sneath P. 1975. Cladistic representation of reticulate evolution. Sys. Zool. 24:360–368.

Than C., Ruths D., Nakhleh L. 2008. PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. BMC Bioinformatics 9:322.

van Iersel L., Kelk S., Rupp R., Huson D.H. 2010. Phylogenetic networks do not need to be complex: using fewer reticulations to represent conflicting clusters. Bioinformatics 26:i124–i131.

Vos R.A., Balhoff J.P., Caravas J.A., Holder M.T., Lapp H., Maddison W.P., Midford P.E., Priyam A., Sukumaran J., Xia X., Stoltzfus A. 2012. Nexml: rich, extensible, and verifiable representation of comparative data and metadata. Syst. Biol. 61:675–689.