VIEWPOINT

To network or not to network, that is the question

PÉTER POCZAI*

Department of Biological and Environmental Sciences, University of Helsinki, FIN-00014, PO Box 65, Helsinki, Finland

Polyploidization can be regarded as one of the most important evolutionary processes. It is well known that the condition of polyploidy is common in plant species, and it has been suggested that approximately 70% of angiosperms have undergone at least one ancient genome duplication (Masterson 1994). Moreover, recent studies assume that palaeopolyploidy is widespread throughout all eukaryotic lineages, having occurred, for example, during the evolution of baker's yeast (*Saccharomyces cerevisiae* Meyen ex E.C. Hansen), in the early ancestor of vertebrates, and repeatedly during the evolution of teleost fishes (Makalowski 2001). In polysporangiophytes ('vascular plants' or tracheophytes), polyploidy is recognized as an important phenomenon, with several lines of evidence suggesting that 15% of angiosperm and 31% of fern speciation events are accompanied by ploidy increase (Wood *et al.* 2009). Considering that a single polyploid species can have multiple origins and genome duplications can occur multiple times during the history of a lineage, it can be assumed that polyploidization events are very significant sources of genome and/or species diversity. But what can this phenomenon tell when viewed through an evolutionary lens? Besides the fact that polyploidy is one of the most interesting events in genome biology, it also sheds light on evolutionary processes, which may differ from those typically represented in tree-like phylogenies (Poczai and Hyvönen 2011). Moreover, polyploidy is not the only biological process that may interfere with phylogenetic reconstructions aiming to find a single representative tree (Doolittle 1999). Other process, including (i) gene conversion or horizontal gene transfer (HGT), (ii) interspecific recombination, and (iii) hybrid speciation (both diploid and polyploid) may also result in complex evolutionary relationships that are better represented as a network rather than a strictly bifurcating tree. These processes together are termed reticulate evolutionary events (Huson *et al.* 2010).

Reticulate evolutionary pathways characteristically differ from those traditionally explored by biologists confronted with the complexity of the 'tree of life'. Phylogenetic trees have been the main tools for representing evolutionary relationships, although analysis of the growing body of data made available by molecular systematics increasingly highlights the need to search for reticulating events. The principal reason for this is that reticulating events can distort phylogenetic signals to such an extent that cladistic methods, which assume a bifurcating tree-like structure, cannot perfectly represent the structure of the data. Vriesendorp and Bakker (2005) correctly pointed out that the botanical literature is all but devoid of examples of reticulating patterns. The question, however, is whether this is because currently developed methods are inadequate or because the phenomenon itself is completely ignored. Researchers continue to use currently available algorithms and solutions without being fully aware of their limitations, or undertake projects which are effectively experiments with the available computational methods. Although other problems also remain in phylogenetics (e.g. sequence alignment, missing sequence data, fragmented morphological information), the phenomenon of reticulate evolution is now receiving growing attention. This is often because of observation of conflicting signals in datasets, leading to the development of phylogenetic networks. Such methods are at the forefront of phyloinformatics today, because they allow the modelling of these increasingly and frequently recognized nontree-like evolutionary histories.

There are a number of recently developed phylogenetic network methods that will not be reviewed here, although these are comprehensively discussed by Huson *et al.* (2010). All methods are similar in that they generalize phylogenetic trees and depict networks that allow the representation of several trees simultaneously. In other words, any drawn network can be termed a phylogenetic network when it represents the evolutionary relationships of a set of terminals divided by nodes and with evolutionary pathways

*E-mail: peter.poczai@gmail.com.

Keywords. reticulate evolution; phylogenetic networks; split graphs; hybridization; polyploidy.

represented by edges (Huson and Bryant 2006). Under this extended view, phylogenetic trees are the simplest networks as they can be regarded as undirected acyclic graphs. To construct any kind of network, it is essential to represent conflicting signals, but not to any greater extent than they are found in the dataset. Otherwise, the network methods will build a phylogeny of conflicting characters rather than depicting the reticulate evolutionary scenario. In a reticulated graph, ambiguities should arise due to homoplasious character changes, and parallel edges should designate alternative genealogical paths. However, this is not always the case, as ambiguities due to other factors (e.g. lineage sorting) may be represented on the graph. More importantly, this property of networks can be highly useful when incongruence between different sorts of datasets is being explored. Exceptionally, when sequence information originating from different genes is analysed and found to be incongruent, it may be preferable to construct a single phylogeny representing this incongruence based on separate analyses of individual genes rather than conduct simultaneous analysis of the data (e.g. Bell and Hyvönen 2010). Network methods such as splits graph have the ability to highlight the predominant signal in the data and the extent to which that signal may or may not be tree-like. This also means that an ideal network method should return a perfect bifurcating tree—or a network shape close to a tree—when the data is perfectly consistent with a tree-like structure.

The rapid development of network methods has resulted in a range of new algorithms (Huson *et al.* 2004; Than *et al.* 2008; Eslahchi *et al.* 2010) with different putative applications, strengths and weaknesses (Linder and Rieseberg 2004). Whether or not these graphs provide an accurate way to resolve true genealogies, some problems still remain with complex evolutionary pattern reconstructions. There is also a critical lack of solutions for defining the order of possible hybridization or recombination events in an accurate and optimal manner for assessing the reliability of defined scenarios. Development of such methods would make reticulate evolutionary studies more tractable, allowing researchers to choose from multiple resulting hypotheses. This is a critical step of network analysis, as many ways exist to explain nontree-like evolution. Phylogenetic network reconstruction methods have moved on from an early stage of development to a progressive phase, with the appearance of promising new approaches (see Esser *et al.* 2004; Than *et al.* 2008; Huson *et al.* 2010). However, several important questions remain, such as how to distinguish truly reticulate events from phylogenetic noise and incongruence in analysed datasets. It is essential to note that other factors such as lineage sorting, inadequate data and insufficient analysis may also lead to reticulate patterns, and the major challenge of future software development should be finding ways to resolve such ambiguity. Correspondingly, researchers using the currently available methods should not interpret the results uncritically. The growing number of developed methods certainly provides new approaches for building reticulate phylogenies,

but these should be regarded as the preliminary tools of well-developed methods. Current methods do, however, provide the heuristic means and in some cases the objective analytical frameworks for studying complex evolutionary processes and even for visualization of incongruence in datasets. In summary, these methods are currently under intensive development and the solutions offered for phylogenetic inference are becoming more sophisticated, such that there is less ground for ignoring reticulation issues in evolutionary analysis. Although with the currently developed methods, we have the potential to observe and determine reticulation events such as recombination, we can hardly determine the number and/or direction of these patterns. It is even more difficult to characterize these events in a powerful and accurate way which is also phylogenetically sound. The ideal solution to this problem would be the development of user friendly software that is able to apply optimality criteria for reticulations to produce an explicitly interpretable collection of events and represent the scenario that best fits the analysed data. Otherwise, researchers will need to keep on guessing what the optimal solutions to their questions are. In this respect, the developed methods represent only the first step rather than the desired goal, although they are certainly on the correct path.

References

Bell N. E. and Hyvönen J. 2010 Phylogeny of the moss class Polytrichopsida (Bryophyta): generic-level structure and incongruent gene trees. *Mol. Phylogenet. Evol.* **55**, 381–398.

Doolittle W. F. 1999 Phylogenetic classification and the universal tree. *Science* **284**, 2124–2129.

Eslahchi C., Habibi M., Hassanzadeh R. and Mottaghi E. 2010 MC-Net: a method for the construction of phylogenetic networks based on the Monte-Carlo method. *BMC Evol. Biol.* **10**, 254.

Esser C., Ahmadinejad N., Wiegand C., Rotte C., Sebastiani F., Gelius-Dietrich G. *et al.* 2004 A genome phylogeny for mitochondria among a-proteobactera and predominantly eubacterial ancestry of yeast nuclear genes. *Mol. Biol. Evol.* **21**, 1643–1660.

Huson D. H. and Bryant D. 2006 Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267.

Huson D. H., Deulian T., Klöpper T. and Steel M. A. 2004 Phylogenetic super-networks from partial trees. *IEEE/ACM Trans. Comp. Biol. Bioinf.* **1**, 151–158.

Huson D. H., Rupp R. and Scornavacca C. 2010 *Phylogenetic networks. Concepts, algorithms and applications*. Cambridge University Press, Cambridge, UK.

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

Makalowski W. 2001 Are we polyploids? A brief history of one hypothesis. *Genome Res.* **11**, 667–670.

Masterson J. 1994 Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. *Science* **264**, 421–424.

Poczai P. and Hyvönen J. 2011 On the origin of *Solanum nigrum*: can networks help? *Mol. Biol. Rep.* **38**, 1171–1185.

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

Vriesendorp B. and Bakker F. 2005 Reconstructing patterns of reticulate evolution in angiosperms: what can we do? *Taxon* **54**, 593–604.

Wood T. E., Takebayashi N., Barker M. S., Mayrose I., Greenspoo P. B. and Rieseberg L. H. 2009 The frequency of polyploid speciation in vascular plants. *Proc. Natl. Acad. Sci. USA* **106**, 13875–13879.