

Phyletic groups on networks

Ward C. Wheeler*

Division of Invertebrate Zoology, American Museum of Natural History, Central Park West @ 79th Street, New York, NY, 10024-5192, USA

Accepted 22 September 2013

Abstract

Three additional phyletic group types, “periphyletic,” “epiphyletic,” and “anaphyletic” (in addition to Hennigian mono-, para-, and polyphyletic) are defined in terms of trees and phylogenetic networks (trees with directed reticulate edges) via a generalization of the algorithmic definitions of Farris. These designations concern groups defined as monophyletic on trees, but with additional gains or losses of members from network edges. These distinctions should be useful in discussion of systems with non-vertical inheritance such as recombination between viruses, horizontal exchange between bacteria, hybridization in plants and animals, as well as human linguistic evolution. Examples are illustrated with Indo-European language groups.

© The Willi Hennig Society 2013.

Introduction

Hennig (1950, 1966) defined a monophyletic group as a set of taxa containing *all and only* the descendants of a common ancestor, characterized and identified by synapomorphy (shared, derived features). He further defined a paraphyletic group as one based on plesiomorphy (shared, but primitive features), hence containing *some* but not all descendants of a common ancestor, and a polyphyletic group as one based on convergence (features mistakenly thought to be shared, derived features), hence containing descendants of multiple ancestors.

Although monophyly was defined in a tree-based manner, Hennig defined parphyly and polyphyly as alternative forms of error in character interpretation. This inconsistency of definitional form created uncertainty in the discussion of phyletic groups. To address this, Farris (1974) formulated a set of consistent definitions based on an algorithmic process applied to trees.

In short, Farris’s procedure assigns a binary character state to each terminal taxon (operational taxonomic unit, OTU), “1” if a member of the group in question and “0” if not. By group, we mean set of ter-

minals. Any set of terminals, however arrived at, can constitute such a group. The root of the tree is also assigned “0”. Character states are then assigned to the internal vertices (hypothetical taxonomic units, HTUs) such that the overall number of 0/1 transitions between connected vertices is minimized (parsimony). If there is a single 0→1 transformation on the tree, the group is monophyletic; a single 0→1 transformation and at least one 1→0 transformation, the group is paraphyletic; and more than one 0→1 transformation, the group is polyphyletic (detailed in Wheeler, 2012). This set of definitions is formal and clear, if not easily applied in all cases (i.e. unresolved trees).

The trees Farris posited were directed, acyclic graphs (DAGs) with three types of vertices: (i) those with in-degree 0 and out-degree 2 (the root); (ii) those with in-degree 1 and out-degree 2 (internal vertices or HTUs); and (iii) those with in-degree 1 and out-degree 0 (terminal taxa, leaves, or OTUs). Here, the question of how these definitions extend to graphs with other sorts of vertices (i.e. in-degree 2, out-degree 1), generally termed phylogenetic “networks”, is examined. These distinctions should be useful in discussion of systems with non-vertical inheritance such as recombination between viruses, horizontal exchange between bacteria, hybridization in plants and animals, as well as human cultural evolution.

*Corresponding author:

E-mail address: wheeler@amnh.org

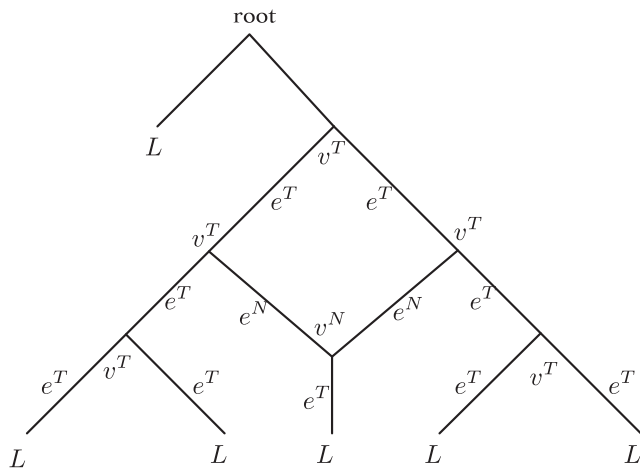


Fig. 1. Elements of a phylogenetic network.

Definitions

Consider a graph $G = (V, E)$, defined in the typical fashion as a tuple consisting of a set of vertices V and set of edges (branches) E . The vertex set contains both¹ leaf (L , terminal or OTU) as well as internal ($V \setminus L$, HTU) vertices. Furthermore, assume that this is a connected graph with directed edges and no cycles (DAG). We will define vertices of in-degree 1, and out-degree 2 as “tree” vertices (v^T), edges that terminate in tree vertices as “tree” edges (e^T), and a character change along such an edge a “tree” change (T_δ). We further define vertices of in-degree 2 and out-degree 1 as “network” vertices (v^N), edges that terminate in a network vertex as “network” edges (e^N), and a character change along such an edge a “network” change (N_δ) (see Fig. 1). Pendant edges (leading to leaf vertices) are considered tree edges. If there are no network vertices, G is a rooted tree as commonly defined in systematics.

Groups

As in Farris (1974), groups are defined as sets of terminal taxa (labelled with “1” if in the set and “0” otherwise), however arrived at. The purpose of the procedure here is to determine phyletic status, not delimitation.

Given that Farris (1974) dealt with trees, his definitions can be expressed as: monophyly = $\{T_{0 \rightarrow 1} = 1, T_{1 \rightarrow 0} = 0\}$; paraphyly = $\{T_{0 \rightarrow 1} = 1, T_{1 \rightarrow 0} \geq 1\}$; and polyphyly = $\{T_{0 \rightarrow 1} > 1\}$.

In the network case, there are other situations that can occur. As with tree edges, there are network

Table 1

Designation of phyletic groups based on the number and polarity of tree ($T_{0 \leftrightarrow 1}$) and network ($N_{0 \leftrightarrow 1}$) transformations

Phyletic group designation				
$T_{0 \rightarrow 1}$	$T_{1 \rightarrow 0}$	$N_{0 \rightarrow 1}$	$N_{1 \rightarrow 0}$	Designation
=1	=0	=0	=0	Monophyletic
			≥ 1	Periphyletic
		≥ 1	=0	Epiphyletic
			≥ 1	Anaphyletic
	≥ 1	=0	=0	Paraphyletic
		≥ 1	≥ 1	
		≥ 1	=0	
			≥ 1	
>1	=0	=0	=0	Polyphyletic
		≥ 1	≥ 1	
		≥ 1	=0	
	≥ 1	=0	=0	
		≥ 1	≥ 1	
		≥ 1	=0	
			≥ 1	

changes that lead to the origination of a group ($N_{0 \rightarrow 1}$) or its loss ($N_{1 \rightarrow 0}$). This results in 16 possible combinations of tree and network, gain and loss, each potentially defining a unique type of phyletic group (Table 1).

However, because we do not define as unique all the four possible tree-based groups (i.e. there is no special name for *all* versus *some* descendants of multiple ancestors), we are not required, and I believe it would be unuseful, to follow such naming combinatorics. I believe that we should focus only on those aspects of monophyly that are affected by network changes. This restricts us to the four network combinations of the origination of a group ($N_{0 \rightarrow 1}$), loss ($N_{1 \rightarrow 0}$), neither, or both. These coupled with the definition of monophyly ($T_{0 \rightarrow 1} = 1, T_{1 \rightarrow 0} = 0$) yield four group types. Clearly, we would maintain monophyly for the original case when we have only tree group origination. I suggest the names “periphyly” for the case of network loss, “epiphyly” for network gain, and “anaphyly” for the situation where both occur (Fig. 2). I further suggest we maintain the Farris (1974) definitions of paraphyly and polyphyly for all network combinations added to the tree definition.²

The determination of vertex (tree and network) state assignments cannot be accomplished as easily for those of trees, which can be accomplished in linear time ($O(n)$, for n leaf taxa). However, this labeling can be determined in a naive manner for n leaves with r reticulate vertices in $O(n \cdot 2^r)$ by simple examination of all network vertex assignment combinations. Although in principle exponential, such an

¹Assuming $V > 2$, which is safe for phylogenetically interesting graphs.

²After all, how many ways do we need to express “not Greek”?

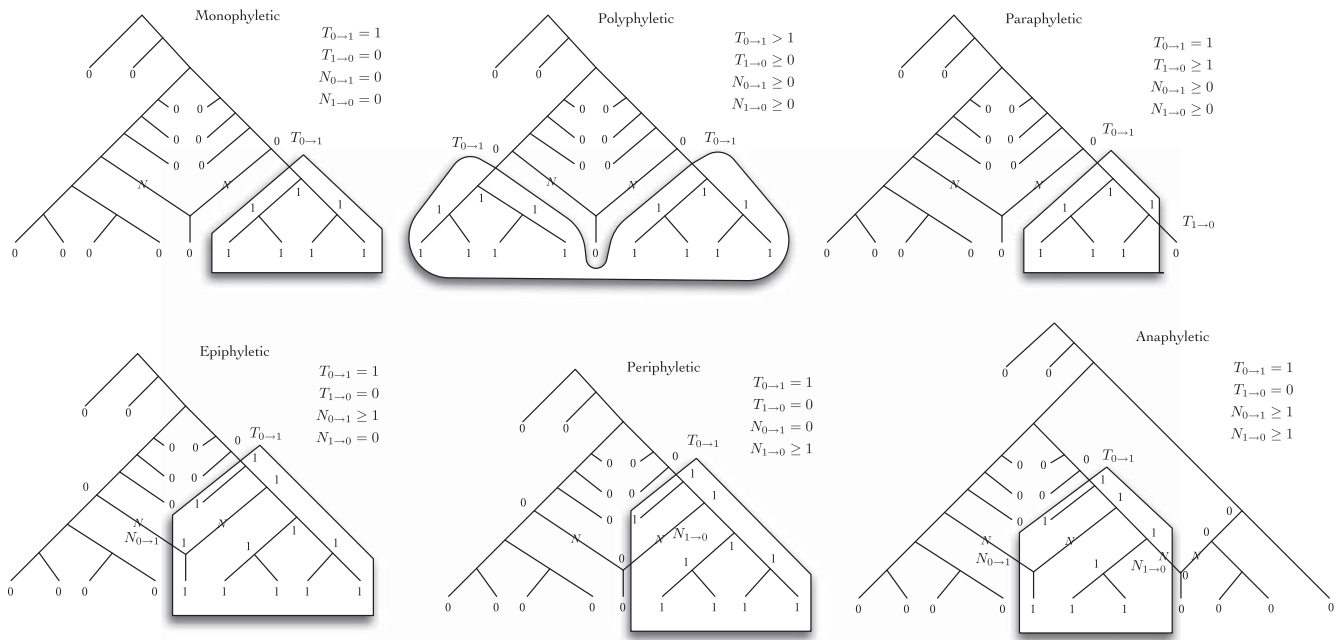


Fig. 2. Groups in trees and networks. Network edges are labelled *N*; all others are tree edges.

optimization procedure should be tractable for most real-world cases.

Examples

Cases where these network-based phyletic groups are most likely to occur are systems where reticulation has been previously proposed. These include the evolution of viruses (e.g. origins of pandemic flu), horizontal gene transfer among pathogenic bacteria, symbiotic capture and loss (e.g. mitochondria in eukaryotes), and human linguistic systems (e.g. pidgin formation). Certain phyletic group types may be associated with

specific forms of non-vertical transformation, and identification of these should promote communication and investigation of general phenomena. Figure 3 shows a case of Indo-European language groups with Modern English having descent from both the Italic and the Germanic language groups. Membership of Modern English in the Germanic group renders it epiphyletic and the Italic group periphyletic. If English were shifted to the Italic group, the status of the groups would exchange. The placement of Sranan (Sranan Tongo) as sister to English (with network connections to African languages) would render Germanic anaphyletic if Sranan were not considered “Germanic”.

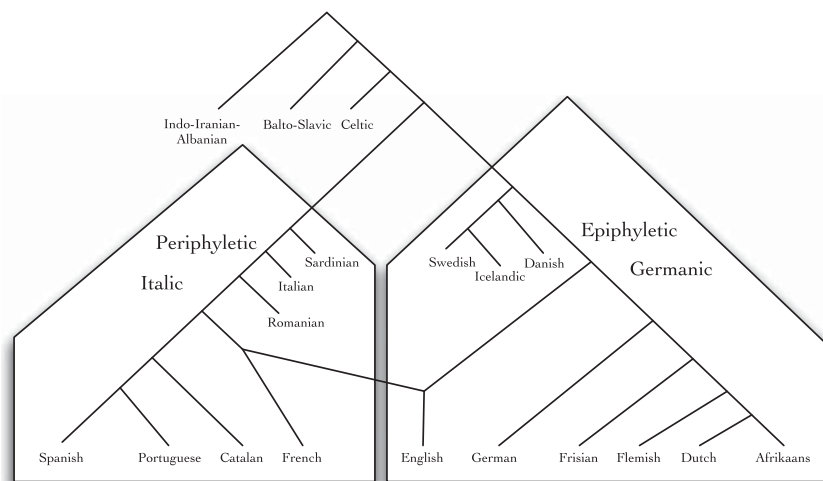


Fig. 3. Examples of periphyletic and epiphyletic groups in Indo-European languages (tree adapted from Atkinson and Gray, 2006).

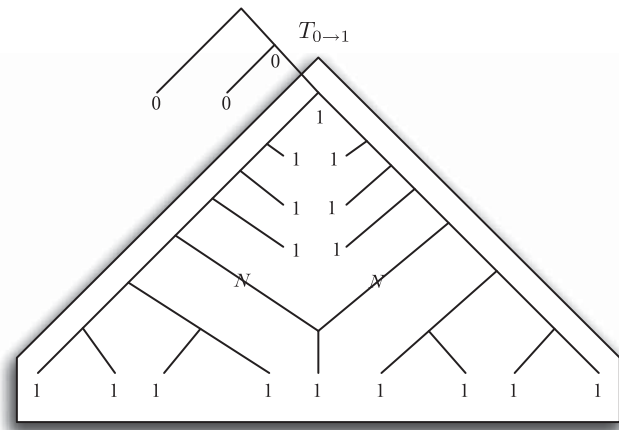


Fig. 4. Minimum size monophyletic group (size ≥ 2) including the network vertex.

A biological example may be found in the butterfly genus *Heliconius*. Traditionally, wing mimicry patterns have been regarded as cases of convergence. Recently, the suggestion has been made that, at least in some cases, shared mimicry patterns are derived from horizontal exchange of alleles (Nadeau et al., 2013; Smith and Kronforst, 2013). Smith and Kronforst (2013) specifically suggest that *Heliconius amaryllis* (sister taxon of *H. aglaope*) has shared haplotypes with *H. timareta*. This would render the monophyletic *amaryllis* + *aglaope* periphyletic (network loss) with respect to *timareta*.

Networks and species

One effect of network edges on monophyly is that the minimum size of monophyletic groups (as defined here) becomes larger (Fig. 4). The root node of a monophyletic clade is pushed back towards the root. If we were to have sexually reproducing individual organisms as leaves (as in Vrana and Wheeler, 1992), the notion of monophyly would only be applicable with larger sets of individuals. Such a graph would be complex and probably not planar, but could (at least in principle) be constructed. This is in contrast to the concept of “nothospecies” (Wagner, 1983), where an individual species arises from a hybridization event between two extant species. Monophyly would still be identifiable in such a situation, but might then become more concordant with clusters of individuals exhibiting patterns of ancestry and descent. In principle, these larger groups of individuals could be a point of convergence between monophyletic species (e.g. Rosen, 1979; Mishler and Donoghue, 1982; Vrana and Wheeler, 1992) and diagnostic/phylogenetic definitions (e.g. Cracraft, 1983; Nixon and Wheeler, 1990).

Almost monophyletic

When systematists discuss groups on trees, monophyletic groups are clearly favoured as complete and most closely related. In the network case, monophyletic groups still retain the greatest interpretive clarity. Of the three new types of phyletic groups proposed here, anaphyly seems the least desirable with contributions from and losses to other groups. Situations with rampant horizontal exchange among multiple lineages (e.g. promiscuous gene transfer) would tend to lead to scenarios in which many identified groups were anaphyletic. In essence, lineages would be heterogeneous amalgams of information from diverse sources. This is the antithesis of monophyly, and if it were to occur, might lead to situations where monophyly would be unrecognizable.

Between epiphyly and periphyly, only an epiphyletic group contains all descendants of a given common ancestor and is closer in some sense to our intuitive feel of a natural (i.e. monophyletic) group. In this sense, epiphyletic groups are “almost” monophyletic and, in some situations, that may be as good as it gets.

Acknowledgements

I would like to thank Andrew Brower, Louise Crowley, Darrel Frost, Gonzalo Giribet, Brent Mishler, Norman Platnick, Peter Whiteley, and an anonymous reviewer for suggesting examples as well as helping to clarify this presentation and identifying surprisingly many errors in such a short manuscript. This material is based upon work supported by, or in part by, the US Army Research Laboratory and the US Army Research Office under contract/grant number W911NF-05-1-0271.

References

- Atkinson, Q.D., Gray, R. 2006. How old is the Indo-European language family? Illumination or more moths to the flames? In: Forster, P., Renfrew, C. (Eds.), *Phylogenetic Methods and the Prehistory of Languages*, MacDonald Institute Monographs. MIT Press, Cambridge, MA, pp. 91–109.
- Cracraft, J., 1983. Species concepts and speciation analysis. *Curr. Ornithol.* 1, 159–187.
- Farris, S.J., 1974. Formal definitions of paraphyly and polyphyly. *Syst. Zool.* 23, 548–554.
- Hennig, W. 1950. *Grundzüge einer Theorie der Phylogenetischen Systematik*. Deutscher Zentralverlag, Berlin.
- Hennig, W. 1966. *Phylogenetic Systematics*. University of Illinois Press, Urbana.
- Mishler, B.D., Donoghue, M.J., 1982. Species concepts: a case for pluralism. *Syst. Biol.* 31, 491–503.
- Nadeau, N.J., Martin, S.H., Kozak, K.M., Salazar, C., Dasmahapatra, K.K., Davey, J.W., Baxter, S.W., Blaxter, M.L., Mallet, J., Jiggins, C.D., 2013. Genome-wide patterns of

- divergence and gene flow across a butterfly radiation. *Mol. Ecol.* 22, 814–826.
- Nixon, K., Wheeler, Q.D., 1990. An amplification of the phylogenetic species concept. *Cladistics* 6, 211–223.
- Rosen, D.E., 1979. Fishes from the uplands and intermontane basin of Guatemala: revisionary studies and comparative geography. *Bull. Am. Mus. Nat. Hist.* 162, 267–376.
- Smith, J., Kronforst, M.R., 2013. Do *Heliconius* butterfly species exchange mimicry alleles. *Biol. Lett.* 9, 20130503.
- Vrana, P., Wheeler, W.C., 1992. Individual organisms as terminal entities: laying the species problem to rest. *Cladistics* 8, 67–72.
- Wagner, W.H. 1983. Reticulistics: the recognition of hybrids and their role in cladistics and classification. In: Platnick, N.I., Funk, V.A. (Eds.), *Advances in Cladistics*, volume 2. Columbia University Press, New York, pp. 63–79.
- Wheeler, W.C. 2012. *Systematics: A Course of Lectures*. Wiley-Blackwell, Oxford.