

Nodes, Branches, and Phylogenetic Definitions

KEVIN DE QUEIROZ*

Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20560, USA

*Correspondence to be sent to: Division of Amphibians & Reptiles, National Museum of Natural History, Smithsonian Institution, P.O. Box 37012, NHB, WG-7, MRC 162, Washington, DC 20013-7012, USA; E-mail: dequeirozk@si.edu

Received 5 January 2013; reviews returned 21 February 2013; accepted 16 April 2013

Associate Editor: Frank (Andy) Anderson

Adopting the perspective of graph theory, [Martin et al. \(2010\)](#) described two kinds of phylogenetic trees, which they termed *node-based* and *stem- or branch-based*, that differ with respect to the biological interpretations of their component nodes and branches. After establishing equivalency between the two different kinds of trees in terms of encoded information regarding taxa and their phylogenetic relationships, [Martin et al. \(2010\)](#) argued that node-based names should be applied only in the context of node-based trees, and that branch-based names should be applied only in the context of branch-based trees, because node-based names cannot exist on branch-based trees and vice versa. They also suggested that the *International Code of Phylogenetic Nomenclature* or *PhyloCode* ([Cantino and de Queiroz 2010](#)) confuses the two kinds of names and trees and should therefore be amended to adopt one or the other of the two kinds of trees and the corresponding kind of names.

In this contribution, I accept the distinction between the two kinds of phylogenetic trees described by [Martin et al. \(2010\)](#), as well as their equivalency with regard to contained phylogenetic information. Nevertheless, I argue that the names applied to the different kinds of trees by [Martin et al. \(2010\)](#) are inappropriate, and I propose more appropriate names for them. I then argue that [Martin et al.'s \(2010\)](#) conclusions concerning the applicability of node- and branch-based names in the context of what those authors called “node-” and “branch-based” trees are erroneous and result from assuming inappropriate taxon compositions (i.e., ones that would not be deduced from [standard] node- and branch-based phylogenetic definitions). Contrary to [Martin et al. \(2010\)](#), I argue that it is appropriate to apply both node- and branch-based phylogenetic definitions in the context of both “node-” and “branch-based” trees. Next, I show that the *PhyloCode* explicitly adopts trees similar to those that [Martin et al. \(2010\)](#) call “branch-based” and that the inappropriate circumscriptions adopted by those authors are related to their overlooking that fact. Nevertheless, I argue that the term “branch-based definition” is inappropriate in the context of “node-based” trees, and I propose alternative names for both kinds of definitions that are applicable in the context of both kinds of trees. Finally, I argue that what have previously been called “node-based” definitions are most appropriate for defining the names of crown clades while definitions previously termed

“branch-based” are most appropriate for defining the names of total clades, and I describe situations in which the two kinds of definitions have different consequences in the context of alternative phylogenetic hypotheses. These considerations indicate that both “node-” and “branch-based” definitions can and should be used in a system of nomenclature based on phylogenetically defined names.

TREE TERMINOLOGY

Based on a distinction implied by [Hennig \(1966\)](#), [Martin et al. \(2010\)](#) described two kinds of phylogenetic trees that differ with respect to the significations of their component nodes (vertices) and branches (edges). (Hereafter, I will use the terms *nodes* and *branches*, which are more familiar in biology, rather than their graph theoretic synonyms *vertices* and *edges*.) In their *node-based trees*, the nodes represent biological entities (e.g., species, genes), whereas the branches represent relationships between those entities (e.g., ancestor-descendant relationships). In contrast, in their *branch-based (= stem-based) trees*, the branches represent lineages, whereas the (internal) nodes represent lineage splitting events.

Although the distinction between these two kinds of trees is useful, the names applied to them by [Martin et al. \(2010\)](#) do not effectively convey that distinction. The reason is that both kinds of trees (and indeed all trees in the graph theoretic sense) are composed of both nodes and branches. Although the terminology of [Martin et al. \(2010\)](#) could be modified to distinguish more effectively between the two kinds of trees by calling them *taxon-[as]-node trees* and *taxon-[as]-branch trees* (*TN trees* and *TB trees* for short), I consider it more straightforward to call them *relationship trees* and *lineage trees* (*R trees* and *L trees* for short), based on the difference regarding what the branches represent, and I will adopt the latter terminology in the remainder of this contribution. Following [de Queiroz \(2007\)](#), I will also use *branch-based* rather than *stem-based*, except when quoting [Martin et al. \(2010\)](#).

TREES AND PHYLOGENETIC DEFINITIONS

According to [Martin et al. \(2010, p. 1\)](#), “applying a node-based name to a stem-based [lineage] tree is

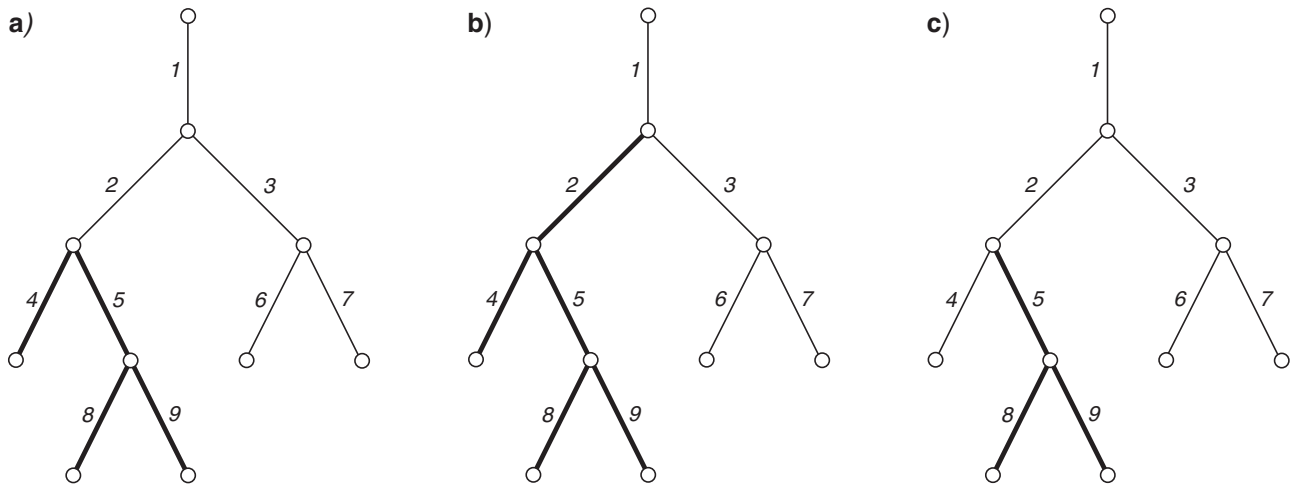


FIGURE 1. Erroneously and correctly inferred clade compositions under node-based definitions on a lineage tree. a) Example of [Martin et al. \(2010\)](#) in which the circumscribed group supposedly includes species 4, 5, 8, and 9 but not 2; the group is polyphyletic because the MRCA (2) of the included species is not itself included. b) Example illustrating that using a standard node-based definition to infer the composition of a clade that includes species 4, 5, 8, and 9 will also include species 2. Here, the name is defined as designating the clade originating in the MRCA of 4 and 9. The MRCA of 4 and 9 is 2, which is therefore included in the clade. c) Example illustrating that using a standard node-based definition to infer the composition of a clade that excludes species 2 will also exclude at least one of the species from the set {4, 5, 8, 9}. Here, the name is defined as designating the clade originating in the MRCA of 8 and 9. The MRCA of 8 and 9 is 5, and therefore the clade includes 5, 8, and 9 but excludes not only 2 but also 4. In all three diagrams, the branches of the trees represent lineage segments (species), and the nodes represent speciation events. Species included in the relevant group or clade are represented by thicker branches.

not logical because node-based names cannot exist on a stem-based [lineage] tree and visa [sic] versa [i.e., branch-based names cannot exist on a node-based (relationship) tree].” These conclusions are incorrect and result from [Martin et al. \(2010\)](#) assuming compositions (circumscriptions) that are inconsistent with those that would be inferred from phylogenetic definitions. Thus, according to ([Martin et al., 2010](#), p. 9): “Applying node-based circumscription to the stem-based [tree] *U* results in the polyphyletic group of 4, 8, 9, and the inferred ancestor 5 . . . because inferred ancestor 2 remains excluded” (Fig. 1a). Although I agree that a group composed of 4, 8, 9, and ancestor 5 on the tree in question is polyphyletic if no part of ancestor 2 is included, that composition would not be inferred from a node-based phylogenetic definition.

Phylogenetic definitions are the methods used to infer taxon composition (circumscription) under phylogenetic nomenclature in general ([de Queiroz and Gauthier 1990, 1992, 1994](#)) and the *PhyloCode* ([Cantino and de Queiroz 2010](#)) in particular. The forms of the two relevant kinds of phylogenetic definitions, which were not mentioned by [Martin et al. \(2010\)](#), are as follows (e.g., *PhyloCode* Art. 9.3, although the second definition has been modified for brevity):

- (1) *Node-based definition*: the clade originating in the most recent common ancestor (MRCA) of A and B.
- (2) *Branch-based definition*: the clade originating in the earliest ancestor of A that is not an ancestor of C.

Following these formulations, any node-based definition that designates a taxon including 4, 8, 9,

and ancestor 5 on the tree in question (Fig. 1) will also include ancestor 2, or at least part of it (see ‘PHYLOGENETIC DEFINITIONS TERMINOLOGY’ section below). For example, the clade originating in the MRCA of 4 and 9 necessarily includes ancestor 2, which is the inferred ancestor in which the clade originated (here I adopt the same position as did [Martin et al. \(2010\)](#), following [Hennig \(1966\)](#), in considering the ancestor in which a given clade originated to be a member of that clade) (Fig. 1b). Conversely, any standard node-based definition that designates a taxon excluding ancestor 2 cannot include all of the other species listed in [Martin et al.’s](#) circumscription (4, 5, 8, and 9). For example, the clade originating in the MRCA of 8 and 9 excludes ancestor 2, but it also excludes 4 (Fig. 1c). Although I have only presented two examples, the general conclusions can be demonstrated by exhaustive enumeration—that is, by substituting all possible combinations from the set {1, 2, 3, 4, 5, 6, 7, 8, 9} for A and B in definition (1) in the context of the tree in Figure 1.

According to [Martin et al. \(2010, p. 10\)](#), “a planted subtree [where the rootward-most branch presumably forms the basis of a branch-based circumscription] in a node-based tree (such as the tree spanned by the black vertices 1, 3, 6, 7 in [their] Fig. 4A) is not monophyletic but paraphyletic, because it includes only one child (3) of its root vertex while excluding child 2 and the children of 2” (Fig. 2a). On the contrary, according to the standard formulation (definition [2]), any branch-based definition that designates a taxon including 1, 3, 6, and 7 on the tree in question (Fig. 2) will also include 2 and its descendants. For example, the clade originating in the earliest ancestor of 7 that is not an ancestor of 10 (where

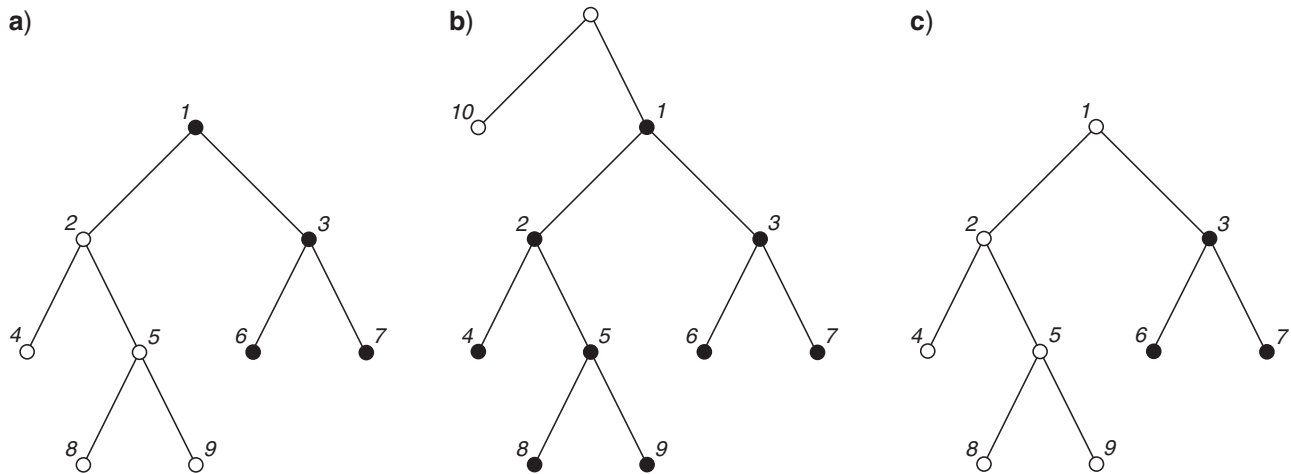


FIGURE 2. Erroneously and correctly inferred clade compositions under branch-based definitions on a relationship tree. a) Example of [Martin et al. \(2010\)](#) in which the circumscribed group supposedly includes species 1, 3, 6, and 7; the group is paraphyletic because it does not include all of the descendants of ancestor 1 (2, 4, 5, 8, and 9 are excluded). b) Example illustrating that using a standard branch-based definition to infer the composition of a clade that includes species 1, 3, 6, and 7 will also include the other descendants of species 1 (2, 4, 5, 8, and 9). Here, the name is defined as designating the clade originating in the earliest ancestor of 7 that is not an ancestor of 10. The earliest ancestor of 7 that is not an ancestor of 10 is 1, so the designated clade contains 1 and all of its descendants (2–9). c) Example illustrating that using a standard branch-based definition to infer the composition of a clade that excludes species 2, 4, 5, 8, and 9 will also exclude species 1. Here, the name is defined as designating the clade originating in the earliest ancestor of 7 that is not an ancestor of 9. The earliest ancestor of 7 that is not an ancestor of 9 is 3, and therefore the clade includes 3, 6, and 7 and excludes 2, 4, 5, 8, and 9 as well as 1. In all three diagrams, the branches of the trees represent ancestor-descendant relationships, and the nodes represent species. Species included in the relevant group or clade are represented by black nodes.

10 is the sister species of the clade composed of 1–9) includes 1, 3, 6, and 7 as well as 2, 4, 5, 8, and 9 (Fig. 2b). Conversely, any standard branch-based definition that designates a taxon excluding 2, 4, 5, 8, and 9 will also exclude ancestor 1. For example, the clade originating in the earliest ancestor of 7 that is not an ancestor of 9 includes 3, 6, and 7 but excludes 2, 4, 5, 8, and 9 as well as 1 (Fig. 2c). Again, although I have only presented two examples, the general conclusions can be demonstrated by exhaustive enumeration—that is, by substituting all possible combinations from the set {1, 2, 3, 4, 5, 6, 7, 8, 9} for A and C in definition (2) in the context of the tree in Figure 2.

It makes sense that definitions (defining formulas, definitia) taking the general forms described above (see definitions [1] and [2]), which (apart from their names) make reference neither to nodes nor to branches, would be applicable in the context of any relevant phylogenetic hypothesis (i.e., one including the designated specifiers) regardless of how that hypothesis is represented (e.g., as a relationship tree vs. a lineage tree). [Martin et al.'s \(2010\)](#) conclusions about the equivalence between relationship trees and lineage trees only reinforce that expectation.

REASONS FOR THE DISCREPANCIES

Presumably, [Martin et al. \(2010\)](#) did not deliberately adopt inappropriate clade circumscriptions (compositions). It is therefore useful to consider why they might have adopted those circumscriptions.

Although [Martin et al. \(2010\)](#) did not mention node-based and branch-based phylogenetic definitions, they did cite the following from *PhyloCode* Art. 2.2:

- (3) “A node-based clade is a clade originating with a particular node on a phylogenetic tree, where the node represents a lineage at the instant of a splitting event.”
- (4) “A branch-based clade is a clade originating with a particular branch (internode) on a phylogenetic tree, where the branch represents a lineage between two splitting events.”

Although [Martin et al. \(2010, p. 10\)](#) suggested that the *PhyloCode* is unclear as to whether its rules are intended to be applied in the context of lineage versus relationship trees, the above definitions clearly demonstrate a lineage tree perspective (although the meaning of a node in a lineage tree as described in the *PhyloCode* is not identical to that described by [Martin et al. 2010](#); see below). Thus, (3) explicitly equates a node with a lineage at the instant of a lineage splitting event (rather than an entire species = a lineage between two splitting events, at least under a bifurcating model of speciation [as opposed to budding and phyletic transformation models; see [Foote 1996](#)]), and (4) explicitly equates a branch with a lineage between two splitting events (rather than an ancestor-descendant relationship). If one were to overlook those statements, as [Martin et al. \(2010\)](#) apparently did, and consider only the characterizations “a node-based clade is a clade originating with a particular node on a

phylogenetic tree" and "a branch-based clade is a clade originating with a particular branch on a phylogenetic tree," one might be tempted to apply characterizations (3) and (4) in the context of relationship (rather than lineage) trees, which clearly was not intended by the authors of the *PhyloCode* (Cantino and de Queiroz 2010).

Martin et al. (2010) erroneously concluded that a branch-based circumscription on a relationship tree is paraphyletic because it includes only one of the two descendants of the ancestor in which the group originated. In fact, the inconsistency is worse than they suggested, because the concept of a branch-based clade (a clade originating in a particular branch on a phylogenetic tree) does not make sense in the context of a relationship tree. The reason is that on a relationship tree, the branch in which the clade is supposed to have originated is a relationship rather than an ancestor (i.e., an association rather than an entity). Nevertheless, Martin et al. (2010) inappropriately applied the concept of a branch-based clade in the context of a relationship tree, which lead them to the erroneous conclusion that both the branch (the one between nodes 1 and 3) and the node at its base (1) are to be included in the circumscription but the other descendants of 1 are not (see Fig. 2a)—the branch because the clade in question is branch-based (originates in a branch), and its basal node because according to "the mathematical definition of a graph . . . one cannot have an edge without both its endpoints" (Martin et al. 2010, p. 10).

Martin et al. (2010) also erroneously concluded that a node-based circumscription on a lineage tree is polyphyletic because it does not include the ancestor in which the group originated. This erroneous conclusion results not only from applying the node-based concept in an inappropriate context but also from a subtle difference between the meanings of nodes in the lineage trees of Martin et al. (2010) as opposed to those of the *PhyloCode* (Cantino and de Queiroz 2010). For Martin et al. (2010, p. 4), a node in a lineage tree represents a speciation event; in contrast, in the *PhyloCode* (see definition [3]), a node represents a lineage at the instant of a speciation event (obviously a simplification, given that speciation is a process). In this context, Martin et al. (2010) presumably concluded that the circumscribed group is polyphyletic because its basal node (the one between branches 2, 4, and 5) is a speciation event rather than an ancestor, and the ancestor (represented by branch 2) is not included in the circumscription (see Fig. 1a). In contrast, a lineage at the time of a speciation event is an ancestor (even if it is only part of the entire ancestral species), and therefore the concept of a node-based clade as defined in the *PhyloCode* includes the ancestor so that the group is monophyletic.

It is also worth noting that there is a contradiction between the designations of the entities as node-based and branch-based *clades* and Martin et al.'s conclusions that the circumscribed groups are not monophyletic. A clade is monophyletic *by definition*.

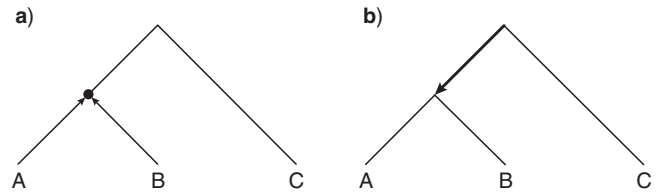


FIGURE 3. Rationale for the names "node-based" and "branch-based" phylogenetic definitions. a) "Node-based" refers to the fact that definitions of this kind specify the ancestor in which the clade being named originated as the point in the tree where the ancestral lineages of the internal specifiers (in this case, A and B) coalesce. It is natural to equate such a point with a node, represented by the black dot in the diagram. b) "Branch-based" refers to the fact that definitions of this kind specify the ancestor in which the clade being named originated as the first segment in a lineage that is ancestral to the internal specifier (in this case, A) but not to the external specifier (in this case, C). It is natural to equate such a lineage segment with a branch, represented by the thick line with an arrow in the diagram. In both diagrams, the trees are lineage trees in the sense described in the *PhyloCode* (where a node represents an ancestor at the instant of a speciation event rather than the speciation event itself); the specifiers are positioned at the ends of branches to indicate that the clades may include other species that are not shown (e.g., species that are more closely related to A than to B).

PHYLOGENETIC DEFINITIONS TERMINOLOGY

Contrary to Martin et al. (2010, p. 10), the terms "node-based" and "branch-based" as applied to phylogenetic definitions are not intended to imply different concepts of clades or monophyly. Instead, the clades designated by different kinds of phylogenetic definitions "are fundamentally similar in being conceptualized and defined as parts of phylogeny each of which is composed of an ancestor and all of its descendants" (de Queiroz 2007, p. 957). The different categories of clades simply represent convenient ways of partitioning the idealized representation of phylogeny as a tree in the graph theoretic sense (de Queiroz 2007). Thus, the MRCA of A and B in a node-based definition (1) is equated with the node in which the ancestral lineages of A and B coalesce (Fig. 3a), while the earliest ancestor of A that is not an ancestor of C in a branch-based definition is equated with the branch representing the first species (lineage segment) in the ancestral lineage of A after it diverged from the ancestral lineage of C (Fig. 3b). This distinction corresponds roughly to Martin et al.'s (2010) distinction between a rooted but nonplanted tree (in which the root node has two "children," represented by the two internal specifiers in a fundamental node-based definition [1]) versus a planted tree (in which the root has a single "child," represented by the single internal specifier in a fundamental branch-based definition [2]), except that in the case of branch-based definitions the node at the base of the specified branch is not included.

According to Martin et al. (2010, p. 10), "the mathematical definition of a graph does not permit such a structure; one cannot have an edge without both its endpoints." If that convention were to be used to argue against the exclusion of the basal node generally, such an argument would be overly formalistic in placing more importance on a graph theoretic convention than on the accurate representation of phylogeny. In a lineage tree,

the basal node cannot be part of a specified branch-based clade, because it represents either a speciation event (under the representational system of Martin et al.) or the last part of a more remote ancestral species than the one specified by the definition (under the representational system of the *PhyloCode*), neither of which are members of the specified (branch-based) clade. However, Martin et al. were referring to the interpretation of a branch-based clade in the context of a relationship tree and the fact that the specified clade supposedly includes only one of the two immediate descendants of its basal node. On the contrary, as I argued earlier, any branch-based definition specifying a clade that includes a particular ancestor (in this case, the one represented by node 1 in Fig. 2) would also include all the descendants of that ancestor (Fig. 2b), while any branch-based definition specifying a clade that excludes any descendants of a particular ancestor (e.g., node 1) would also exclude that ancestor (Fig. 2c). The reason for the discrepancy is that Martin et al. (2010) applied the concept of a branch-based clade in a context in which it was not intended to be applied (i.e., that of a relationship tree) and assumed inclusion of the node at the beginning of the basal branch (as dictated by a graph theoretic convention) rather than considering which ancestor would have been specified by an actual branch-based definition.

Despite the fact that both node-based and branch-based definitions are applicable in the context of both relationship and lineage trees, some of the distinctions made both by Martin et al. (2010) and in this contribution reveal that the current terminology regarding phylogenetic definitions is potentially confusing. In the case of the term “node-based definition,” the potential confusion involves the conceptualization of the inferred ancestor. For those people who would allow the last common ancestor of A and B to be only part of a species (see fig. 2 of de Queiroz and Donoghue 1990), it seems reasonable to represent that ancestor as a node representing the last “instant” that a lineage existed before diverging into two lineages (obviously an abstraction) on a lineage tree (Fig. 3a). By contrast, as discussed by Gauthier and de Queiroz (2001), other people might conceptualize the last common ancestor of A and B as an entire ancestral species. For those people, the ancestor in question should be represented as a branch rather than a node on a lineage tree, in which case it makes little sense to designate the definition “node-based.”

In addition, as noted earlier, the term “branch-based definition” makes little sense if applied in the context of a relationship tree. The reason is that the name “branch-based” derives from representation of the designated ancestor (the one in which the named clade originated) as a branch on a phylogenetic tree, but that representation applies only to lineage trees. On relationship trees, where the branches represent relationships rather than ancestral species, clades always originate in nodes rather than branches.

To avoid confusion with respect to how the designated ancestors are conceptualized or represented, it would seem preferable to adopt a terminology that avoids equating those ancestors specifically with either nodes or branches. Therefore, I here propose alternative terms for the two kinds of definitions in question based on ideas put forward by Schander and Tholleson (1995). Those authors provided alternative wordings (but not names) for node-based and branch-based definitions (defining formulas, *definientia*) that take the following forms (here modified slightly for brevity):

- (5) *Node-based definition*: the smallest clade containing both A and B.
- (6) *Branch-based definition*: the largest clade containing A but not C.

Based on these formulations, the term “node-based definition” may be replaced with “minimum-clade definition” and the term “branch-based definition” with “maximum-clade definition” (but note that the new names are intended to apply to both the original [1, 2] and the alternative [5, 6] wordings of the definitions). I will abbreviate the former definitions (1, 5) “min (A and B)” and the latter definitions (2, 6) “max (A not C).” These alternative names should eliminate potential confusions related to particular conceptualizations or representations of the designated ancestors and would also help to avoid the misconception (e.g., Martin et al. 2010) that it is only appropriate to apply a minimum-clade (“node-based”) definition in the context of a relationship (“node-based”) tree and a maximum-clade (“branch-based”) definition in the context of a lineage (“branch-based”) tree.

These new terms also suggest that there is no need to recognize corresponding categories of clades (i.e., modifications of definitions [3] and [4]). Under the minimum/maximum terminology, a distinction is no longer being made regarding the nature of the specified ancestors (i.e., between ancestors as the terminal parts of species [nodes] and ancestors as entire species [branches] in the context of one type of lineage tree). Therefore, although it is useful to distinguish between minimum-clade and maximum-clade definitions based on differences in the structures of their defining formulas, it is not important to distinguish between the corresponding categories of clades, which do not differ with respect to any intrinsic properties but only with respect to how their names are defined. A related consequence of the new terminology is that if an author proposing a minimum-clade definition wishes to adopt a particular concept of the specified ancestor (i.e., as an entire species vs. the terminal part thereof) for the future application of the definition (analogous considerations make no difference in the case of maximum-clade definitions), he or she would have to do so with an explicit statement (given that it would no longer be implicit though definition [3], which I am suggesting should be eliminated).

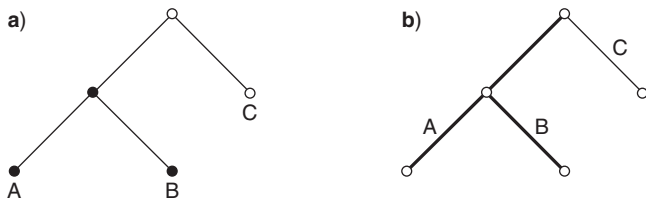


FIGURE 4. If the ancestor in which a clade originates is considered to be an entire species (rather than the last part of a species), it might seem that minimum- and maximum-clade definitions are simply different methods for specifying the same clades and ancestors. For the relationship tree illustrated in a) the last common ancestor of A and B (minimum-clade definition) and the earliest ancestor of A that is not an ancestor of C (maximum-clade definition) refer to the same clade, the members of which are represented by black nodes. The same is true for the lineage tree illustrated in b), where the members of the specified clade are represented by thicker branches. Although this conclusion of equivalency would be correct if all species that have ever existed and their phylogenetic relationships were known with certainty, its limitations in the context of incomplete and provisional knowledge are illustrated in Figures 5–7.

THE UTILITY OF DIFFERENT KINDS OF DEFINITIONS

Martin et al. (2010, p. 10) proposed that the *PhyloCode* adopt only a single kind of tree and the corresponding kind of definition. I have demonstrated that their proposal was based on an incorrect proposition that “node-based” names are incompatible with “branch-based” trees and vice versa. Nonetheless, given that minimum-clade (“node-based”) and maximum-clade (“branch-based”) definitions are used to apply names to entities of the same fundamental kind (clades), and that both kinds of definitions can be applied in the context of both relationship (“node-based”) and lineage (“branch-based”) trees, one might still be tempted to conclude that a single kind of definition is sufficient. At least for people who equate the ancestors specified by phylogenetic definitions with entire species (rather than parts of species), it might seem that minimum- and maximum-clade definitions are simply different methods for specifying the same clades and ancestors (Fig. 4) (see Frost and Kluge 1994; Sereno 1999). Such a conclusion, although correct in the context of complete and error-free knowledge of phylogeny, is over-simplified in the context of real (i.e., incomplete and at least potentially inaccurate) phylogenetic hypotheses, which is to say that there are pragmatic reasons for using both kinds of definitions even in the context of a single kind of tree. In the rest of this section, I will assume, for the sake of simplifying the discussion, that the ancestor in which a clade originates is an entire species.

Crown and total clades are two significant categories of clades defined with respect to the extant biota (reviewed by de Queiroz 2007). A *crown clade* is a clade in which both (all) lineages originating from a particular splitting event have extant descendants (which is not to say that they do not also have extinct ones); a *total clade* is a clade composed of a crown clade and all extinct species (stem group species) that are more closely related to that crown clade than to any other crown clade or extant species (see Hennig 1965, 1969, 1981;

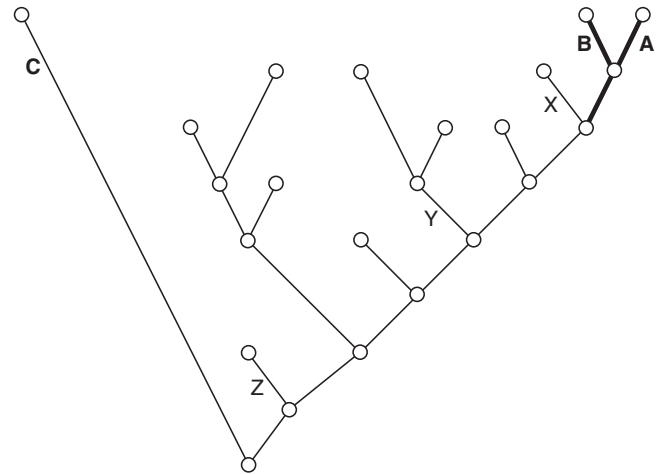


FIGURE 5. Limitations of maximum-clade (“branch-based”) and utility of minimum-clade (“node-based”) definitions for defining the names of crown clades. To define the name of the crown clade composed of species A and B using a maximum-clade definition, a member of the immediate sister group of the crown clade—in this case, extinct species X—must be used as an external specifier in the defining formula (e.g., max [A not X]). If species X has not been discovered and the name of a different member of the stem group (e.g., Y or Z) is used instead, the name will not refer to the crown clade. By contrast, the minimum-clade defining formula min (A and B) can be used to apply a name to the crown clade without knowledge of X or any other member of the stem group. The tree illustrated here is a lineage tree, and the ancestors in which clades originate are considered to be entire species (rather than parts thereof). The vertical component of the diagram represents time, with the present represented by the level of the terminal nodes of branches A, B, and C. Members of the crown clade to be named are indicated by thicker branches.

Jefferies 1979; Lauterbach 1989; Meier and Richter 1992; de Queiroz 2007). (Given that the stem species is part of [and equivalent to] the clade that originates in it [Hennig 1966], a species that has not yet given rise to any daughter species can be considered a clade composed of that one species, and thus a single extant species that has not given rise to any daughter species can be considered a crown clade. Based on the same premise, an *ancestor-per-se definition* is a phylogenetic definition taking the form “the clade originating in D,” where D is a species known from specimens [as opposed to one inferred solely from a tree], whether extinct or extant.)

A fact of real phylogenetic hypotheses (as opposed to hypothetical ones, as in Fig. 4) is that they are almost invariably based on incomplete species sampling, particularly with regard to extinct species. However, to define the name of a crown clade using a maximum-clade (“branch-based”) definition, it would be necessary to know at least one species (and in some cases there will have been only one) of the immediate (often extinct) sister group of the crown clade in question (Fig. 5). Such a species (X in Fig. 5) would be a necessary (and sufficient) external specifier in a maximum-clade definition of the name of the crown clade—for instance, max (A not X) in the example (Fig. 5). Therefore, such a species would not only have to have been discovered, it would also have to have been inferred to be the sister group of

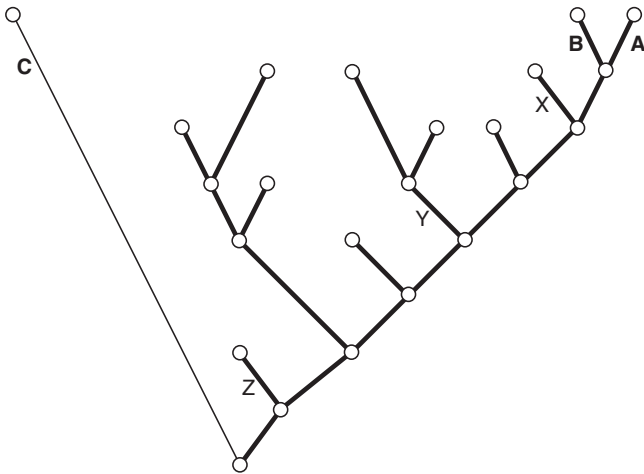


FIGURE 6. Limitations of minimum-clade (“node-based”) and utility of maximum-clade (“branch-based”) definitions for defining the names of total clades. To define the name of the total clade corresponding to the crown clade composed of species A and B using a minimum-clade definition, a member of the extinct group representing the first divergence within the total clade—in this case, extinct species Z—must be used as an internal specifier in the defining formula (e.g., $\min [A \text{ and } Z]$). If species Z has not been discovered and the name of a different member of the stem group (e.g., X or Y) is used instead, the name will not refer to the total clade. By contrast, the maximum-clade defining formula $\max (A \text{ not } C)$ can be used to apply a name to the total clade without knowledge of Z or any other member of the stem group. The tree illustrated here is a lineage tree and the ancestors in which clades originate are considered to be entire species (rather than parts thereof). The vertical component of the diagram represents time, with the present represented by the level of the terminal nodes of branches A, B, and C. Members of the total clade to be named are indicated by thicker branches.

the crown clade to be named. If a different external specifier (such as Y or Z in Fig. 5) were used instead, the name would not apply to the crown clade. By contrast, knowledge of the immediate sister species of a crown clade is not necessary to define the name of that crown clade using a minimum-clade (“node-based”) definition. Thus, in Figure 5, the following minimum-clade defining formula would apply the name to the crown clade regardless of knowledge of any extinct species: $\min (A \text{ and } B)$. Given that the immediate (often extinct) sister species of a crown clade commonly will not be known, it will generally be more feasible to define the names of crown clades using minimum-clade (“node-based”) definitions.

Conversely, to define the name of a total clade using a minimum-clade (“node-based”) definition, it would be necessary to know at least one species (and in some cases there will have only been one) representing the extinct clade resulting from the very first divergence within the total clade in question (Fig. 6). Such a species (Z in Fig. 6) would be one of the necessary internal specifiers in the minimum-clade definition—for instance: $\min (A \text{ and } Z)$ in the example (Fig. 6). Therefore, such a species would not only have to have been discovered, it would also have to have been inferred to be a representative of the extinct clade resulting from the very first divergence within the total clade to be named. If a different internal

specifier (such as X or Y in Fig. 6) were used instead, the name would not apply to the total clade. By contrast, knowledge of a species representing the extinct clade resulting from the very first divergence within a total clade is not necessary to define the name of that total clade using a maximum-clade (“branch-based”) definition. Thus, in Figure 6, the following maximum-clade defining formula would apply the name to the total clade regardless of knowledge of any extinct species: $\max (A \text{ not } C)$. Given that a species representing the extinct clade resulting from the very first divergence within a total clade often will not be known, it will generally be more feasible to define the names of total clades using maximum-clade (“branch-based”) definitions.

Naming crown versus total clades is not the only context in which minimum-clade (“node-based”) and maximum-clade (“branch-based”) definitions exhibit important and useful differences. In addition, the two kinds of phylogenetic definitions have different consequences in the context of revisions to phylogenetic hypotheses that apply to clades more generally. In the following examples, the same conclusions apply regardless of whether the clades are crown clades, total clades, both, or neither. Consider the following definitions: *Athena* = $\min (A \text{ and } B)$; *Minerva* = $\max (A \text{ not } C)$. In the context of Figure 7a, the two names apply to the same clade (at least if the ancestor[s] in which the specified clades originated are considered to be entire species rather than parts thereof): the one composed of A and B and their inferred common ancestor. In the context of this phylogenetic hypothesis, the names *Athena* and *Minerva* are (heterodefinitive) synonyms. By contrast, in the context of Figure 7b and c, the names apply to different clades and therefore are no longer synonyms. In the case of Figure 7b, *Athena* is composed of A and B and their inferred common ancestor, whereas *Minerva* is composed of A, B, and D, and two inferred ancestors. By contrast, in the case of Figure 7c, *Athena* is composed of A, B, C, and two inferred ancestors, whereas *Minerva* is composed of only A, thus reversing the hierarchical relationships of the named clades from the previous case (Fig. 7b).

These examples demonstrate that different kinds of phylogenetic definitions have different consequences in the context of alternative phylogenetic hypotheses. Which kind of definition is selected should therefore depend on how the clade to be named is conceptualized—that is, on the intended meaning (reference) of the name. For example, if one intends for the named clade always to include both A and B regardless of their exact phylogenetic relationships, but if it is acceptable for other taxa, such as C, to be included or excluded, then $\min (A \text{ and } B)$ should be used. By contrast, if one intends for the named clade to include A and to exclude C under any and all hypotheses of relationships, then $\max (A \text{ not } C)$ should be used instead. Other types of definitions are possible, including ones that make particular names inapplicable in the context of specified phylogenetic hypotheses (Lee 1998; de Queiroz and Donoghue 2012).

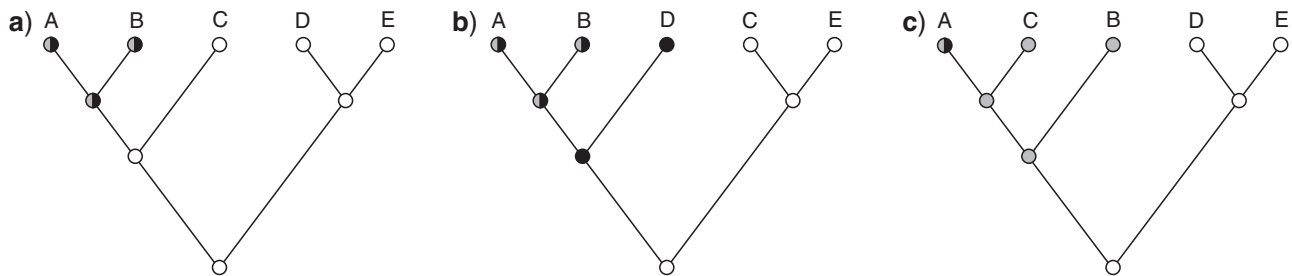


FIGURE 7. Different consequences of minimum-clade (“node-based”) and maximum-clade (“branch-based”) definitions in the context of alternative phylogenetic hypotheses. The following definitions are applied in the context of the three phylogenetic hypotheses illustrated in a–c: *Athena* = min (A and B); *Minerva* = max (A not C). In the context of phylogenetic hypothesis a), both *Athena* and *Minerva* refer to the same clade (the names are synonymous). In the context of hypothesis b), *Minerva* refers to a more inclusive clade than does *Athena*. In the context of hypothesis c), *Athena* refers to a more inclusive clade than does *Minerva*. All three trees are relationship trees (branches represent relationships; nodes represent species). Gray-filled circles = members of *Athena* only; black-filled circles = members of *Minerva* only; gray- and black-filled circles = members of both *Athena* and *Minerva*; unfilled circles = members of neither *Athena* nor *Minerva*.

CONCLUSIONS

A graph theoretic perspective is useful for clarifying issues concerning both the representation of phylogeny using trees (minimally connected graphs) and methods for applying taxon names in the context of such trees (de Queiroz 2007; Martin et al. 2010). Consideration of alternative significations of the nodes and branches of phylogenetic trees suggests that the terminology applied to different kinds of phylogenetic trees by Martin et al. (2010), as well as that applied to different kinds of phylogenetic definitions in the *PhyloCode* (Cantino and de Queiroz 2010), are ambiguous or potentially confusing. Nevertheless, both minimum-clade (“node-based”) and maximum-clade (“branch-based”) definitions are not only applicable in the context of both relationship (“node-based”) and lineage (“branch-based”) trees, both kinds of phylogenetic definitions are also necessary components of a nomenclatural system that must necessarily function in the context of incomplete taxon sampling and changing hypotheses about phylogeny.

ACKNOWLEDGMENTS

Phil Cantino and Ed Wiley provided comments on an earlier version of this contribution.

REFERENCES

- Cantino P.D., de Queiroz K. 2010. International code of phylogenetic nomenclature. Version 4c [Internet]. Available from: URL <http://www.ohio.edu/phylocode/>.
- de Queiroz K. 2007. Toward an integrated system of clade names. *Syst. Biol.* 56:956–974.
- de Queiroz K., Donoghue M.J. 1990. Phylogenetic systematics or Nelson’s version of cladistics. *Cladistics* 6:61–75.
- de Queiroz K., Donoghue M.J. 2012. Phylogenetic nomenclature, hierarchical information, and testability. *Syst. Biol.* 62: 167–174.
- de Queiroz K., Gauthier J. 1990. Phylogeny as a central principle in taxonomy: phylogenetic definitions of taxon names. *Syst. Zool.* 39:307–322.
- de Queiroz K., Gauthier J. 1992. Phylogenetic taxonomy. *Annu. Rev. Ecol. Syst.* 23:449–480.
- de Queiroz K., Gauthier J. 1994. Toward a phylogenetic system of biological nomenclature. *Trends Ecol. Evol.* 9:27–31.
- Foote M. 1996. On the probability of ancestors in the fossil record. *Paleobiology* 22:141–151.
- Frost D.R., Kluge A.G. 1994. A consideration of epistemology in systematic biology, with special reference to species. *Cladistics* 10:259–294.
- Gauthier J., de Queiroz K. 2001. Feathered dinosaurs, flying dinosaurs, crown dinosaurs, and the name “Aves.” In: Gauthier J., Gall L.F., editors. *New perspectives on the origin and early evolution of birds*. New Haven (CT): Peabody Museum of Natural History, Yale University. p. 7–41.
- Hennig W. 1965. Phylogenetic systematics. *Annu. Rev. Entomol.* 10: 97–116.
- Hennig W. 1966. *Phylogenetic systematics*. Urbana: University of Illinois Press.
- Hennig W. 1969. *Die Stammesgeschichte der Insekten*. Frankfurt: Kramer.
- Hennig W. 1981. *Insect phylogeny*. Chichester, UK: Wiley.
- Jefferies R.P.S. 1979. The origin of chordates—a methodological essay. In: House M.R., editor. *The origin of major invertebrate groups*. London: Academic Press. p. 443–477.
- Lauterbach K.-E. 1989. Das Pan-Monophylum—Ein Hilfsmittel für die Praxis der phylogenetischen Systematik. *Zool. Anz.* 223: 139–156.
- Lee M.S.Y. 1998. Phylogenetic uncertainty, molecular sequences, and the definition of taxon names. *Syst. Biol.* 47:719–726.
- Martin J., Blackburn D., Wiley E.O. 2010. Are node-based and stem-based clades equivalent? Insights from graph theory. *PLOS Curr.* 2: pages 1–12.
- Meier R., Richter S. 1992. Suggestions for a more precise usage of proper names of taxa: ambiguities related to the stem lineage concept. *Z. Zool. Syst. Evol.* 30:81–88.
- Schander C., Thollesson M. 1995. Phylogenetic taxonomy—some comments. *Zool. Scr.* 24:263–268.
- Sereni P.C. 1999. Definitions in phylogenetic taxonomy: critique and rationale. *Syst. Biol.* 48:329–351.