



For peer review only. Do not cite.

Nodes, Branches, and Phylogenetic Definitions.

Journal:	<i>Systematic Biology</i>
Manuscript ID:	USYB-2013-005.R2
Manuscript Type:	Point of View
Date Submitted by the Author:	n/a
Complete List of Authors:	de Queiroz, Kevin; Smithsonian Institution, Vertebrate Zoology
Keywords:	branch, graph theory, lineage tree, maximum clade definition, minimum clade definition, node, phylogenetic definition, relationship tree

SCHOLARONE™
Manuscripts

Peer Review Only

NODES, BRANCHES, AND PHYLOGENETIC DEFINITIONS

Nodes, Branches, and Phylogenetic Definitions

Kevin de Queiroz

Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian
Institution, Washington, DC 20560, dequeirozk@si.edu

Corresponding Author:

Kevin de Queiroz

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

Phone: (202) 633-0727; FAX: (202) 633-0182

E-mail: dequeirozk@si.edu

de Queiroz Nodes, Branches, and Phylogenetic Definitions - 2

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Keywords: Branch, graph theory, lineage tree, maximum clade definition, minimum clade definition, node, phylogenetic definition, relationship tree.

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-

de Queiroz

Nodes, Branches, and Phylogenetic Definitions - 3

1
2
3 based phylogenetic definitions in the context of both “node-“ and “branch-based” trees. Next, I
4
5 show that the *PhyloCode* explicitly adopts trees similar to those that Martin et al. (2010) call
6
7 “branch-based” and that the inappropriate circumscriptions adopted by those authors are related
8
9 to their overlooking that fact. Nevertheless, I argue that the term “branch-based definition” is
10
11 inappropriate in the context of “node-based” trees, and I propose alternative names for both
12
13 kinds of definitions that are applicable in the context of both kinds of trees. Finally, I argue that
14
15 what have previously been called “node-based” definitions are most appropriate for defining the
16
17 names of crown clades while definitions previously termed “branch-based” are most appropriate
18
19 for defining the names of total clades, and I describe situations in which the two kinds of
20
21 definitions have different consequences in the context of alternative phylogenetic hypotheses.
22
23 These considerations indicate that both “node-“ and “branch-based” definitions can and should
24
25 be used in a system of nomenclature based on phylogenetically defined names.
26
27
28
29
30
31
32
33

34 TREE TERMINOLOGY

35
36 Based on a distinction implied by Hennig (1966), Martin et al. (2010) described two kinds of
37
38 phylogenetic trees that differ with respect to the significations of their component nodes
39
40 (vertices) and branches (edges). (Hereafter, I will use the terms *nodes* and *branches*, which are
41
42 more familiar in biology, rather than their graph theoretic synonyms *vertices* and *edges*.) In their
43
44 *node-based trees*, the nodes represent biological entities (e.g., species, genes) while the branches
45
46 represent relationships between those entities (e.g., ancestor-descendant relationships). By
47
48 contrast, in their *branch-based (=stem-based) trees*, the branches represent lineages, while the
49
50 (internal) nodes represent lineage splitting events.
51
52
53
54
55
56
57
58
59
60

de Queiroz

Nodes, Branches, and Phylogenetic Definitions - 4

1
2
3 Although the distinction between these two kinds of trees is useful, the names applied to
4 them by Martin et al. (2010) do not effectively convey that distinction. The reason is that both
5 kinds of trees (and indeed all trees in the graph theoretic sense) are composed of both nodes and
6 branches. Although the terminology of Martin et al. (2010) could be modified to distinguish
7 more effectively between the two kinds of trees by calling them *taxon-[as-]node trees* and *taxon-*
8 *[as-]branch trees* (*TN trees* and *TB trees* for short), I consider it more straightforward to call
9 them *relationship trees* and *lineage trees* (*R trees* and *L trees* for short), based on the difference
10 regarding what the branches represent, and I will adopt the latter terminology in the remainder of
11 this contribution. Following de Queiroz (2007), I will also use *branch-based* rather than *stem-*
12 *based*, except when quoting Martin et al. (2010).
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

29 TREES AND PHYLOGENETIC DEFINITIONS

30
31 According to Martin et al. (2010:1), “applying a node-based name to a stem-based [lineage]
32 tree is not logical because node-based names cannot exist on a stem-based [lineage] tree and visa
33 [sic] versa [i.e., branch-based names cannot exist on a node-based (relationship) tree].” These
34 conclusions are incorrect and result from Martin et al. (2010) assuming compositions
35 (circumscriptions) that are inconsistent with those that would be inferred from phylogenetic
36 definitions. Thus, according to Martin et al. (2010: 9): “Applying node-based circumscription to
37 the stem-based [tree] *U* results in the polyphyletic group of 4, 8, 9, and the inferred ancestor 5 ...
38 because inferred ancestor 2 remains excluded” (Fig. 1a). Although I agree that a group
39 composed of 4, 8, 9, and ancestor 5 on the tree in question is polyphyletic if no part of ancestor 2
40 is included, that composition would not be inferred from a node-based phylogenetic definition.
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

de Queiroz

Nodes, Branches, and Phylogenetic Definitions - 5

1
2
3 Phylogenetic definitions are the methods used to infer taxon composition (circumscription)
4 under phylogenetic nomenclature in general (de Queiroz and Gauthier 1990, 1992, 1994) and the
5
6
7
8 *PhyloCode* (Cantino and de Queiroz 2010) in particular. The forms of the two relevant kinds of
9
10 phylogenetic definitions, which were not mentioned by Martin et al. (2010), are as follows (e.g.,
11
12 *PhyloCode* Art. 9.3, although the second definition has been modified for brevity):

- 13
14
15
16
17
18 1) *node-based definition*: the clade originating in the most recent common ancestor of A and B.
19
20 2) *branch-based definition*: the clade originating in the earliest ancestor of A that is not an
21
22 ancestor of C.
23
24
25
26

27 Following these formulations, any node-based definition that designates a taxon including 4, 8, 9,
28 and ancestor 5 on the tree in question (Fig. 1) will also include ancestor 2, or at least part of it
29 (see PHYLOGENETIC DEFINITIONS TERMINOLOGY, below). For example, the clade originating in
30 the most recent common ancestor of 4 and 9 necessarily includes ancestor 2, which is the
31 inferred ancestor in which the clade originated (here I adopt the same position as did Martin et al.
32 (2010), following Hennig (1966), in considering the ancestor in which a given clade originated to
33 be a member of that clade) (Fig. 1b). Conversely, any standard node-based definition that
34 designates a taxon excluding ancestor 2 cannot include all of the other species listed in Martin et
35 al.'s circumscription (4, 5, 8, 9). For example, the clade originating in the most recent common
36 ancestor of 8 and 9 excludes ancestor 2, but it also excludes 4 (Fig. 1c). Although I have only
37 presented two examples, the general conclusions can be demonstrated by exhaustive
38 enumeration—that is, by substituting all possible combinations from the set {1, 2, 3, 4, 5, 6, 7, 8,
39 9} for A and B in definition (1) in the context of the tree in Figure 1.
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

de Queiroz

Nodes, Branches, and Phylogenetic Definitions - 6

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54

According to Martin et al. (2010: 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 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.

55
56
57
58
59
60

It makes sense that definitions (defining formulas, definiencia) 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 versus 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

de Queiroz

Nodes, Branches, and Phylogenetic Definitions - 7

1
2
3 Presumably, Martin et al. (2010) did not deliberately adopt inappropriate clade
4
5 circumscriptions (compositions). It is therefore useful to consider why they might have adopted
6
7 those circumscriptions. Although Martin et al. (2010) did not mention node-based and branch-
8
9 based phylogenetic definitions, they did cite the following from *PhyloCode* Art. 2.2:

10
11
12
13
14
15 3) “A node-based clade is a clade originating with a particular node on a phylogenetic tree,
16
17 where the node represents a lineage at the instant of a splitting event.”

18
19
20 4) “A branch-based clade is a clade originating with a particular branch (internode) on a
21
22 phylogenetic tree, where the branch represents a lineage between two splitting events.”
23
24
25
26

27 Although Martin et al. (2010:10) suggested that the *PhyloCode* is unclear as to whether its rules
28
29 are intended to be applied in the context of lineage versus relationship trees, the above
30
31 definitions clearly demonstrate a lineage tree perspective (although the meaning of a node in a
32
33 lineage tree as described in the *PhyloCode* is not identical to that described by Martin et al. 2010;
34
35 see below). Thus, (3) explicitly equates a node with a lineage at the instant of a lineage splitting
36
37 event (rather than an entire species = a lineage between two splitting events, at least under a
38
39 bifurcating model of speciation [as opposed to budding and phyletic transformation models; see
40
41 Foote 1996]), and (4) explicitly equates a branch with a lineage between two splitting events
42
43 (rather than an ancestor-descendant relationship). If one were to overlook those statements, as
44
45 Martin et al. (2010) apparently did, and consider only the characterizations “a node-based clade
46
47 is a clade originating with a particular node on a phylogenetic tree” and “a branch-based clade is
48
49 a clade originating with a particular branch on a phylogenetic tree”, one might be tempted to
50
51
52
53
54
55
56
57
58
59
60

de Queiroz

Nodes, Branches, and Phylogenetic Definitions - 8

1
2
3 apply characterizations (3) and (4) in the context of relationship (rather than lineage) trees, which
4
5 clearly was not intended by the authors of the *PhyloCode* (Cantino and de Queiroz 2010).
6
7

8 Martin et al. (2010) erroneously concluded that a branch-based circumscription on a
9
10 relationship tree is paraphyletic because it includes only one of the two descendants of the
11
12 ancestor in which the group originated. In fact, the inconsistency is worse than they suggested,
13
14 because the concept of a branch-based clade (a clade originating in a particular branch on a
15
16 phylogenetic tree) does not make sense in the context of a relationship tree. The reason is that on
17
18 a relationship tree, the branch in which the clade is supposed to have originated is a relationship
19
20 rather than an ancestor (i.e., an association rather than an entity). Nevertheless, Martin et al.
21
22 (2010) inappropriately applied the concept of a branch-based clade in the context of a
23
24 relationship tree, which lead them to the erroneous conclusion that both the branch (the one
25
26 between nodes 1 and 3) and the node at its base (1) are to be included in the circumscription but
27
28 the other descendants of 1 are not (see Fig. 2a)—the branch because the clade in question is
29
30 branch-based (originates in a branch), and its basal node because according to “the mathematical
31
32 definition of a graph ... one cannot have an edge without both its endpoints” (Martin et al.
33
34 2010:10).
35
36
37
38
39

40
41 Martin et al. (2010) also erroneously concluded that a node-based circumscription on a
42
43 lineage tree is polyphyletic because it does not include the ancestor in which the group originated.
44
45 This erroneous conclusion results not from applying the node-based concept in an inappropriate
46
47 context but from a subtle difference between the meanings of nodes in the lineage trees of Martin
48
49 et al. (2010) as opposed to those of the *PhyloCode* (Cantino and de Queiroz 2010). For Martin et
50
51 al. (2010:4), a node in a lineage tree represents a speciation event; by contrast, in the *PhyloCode*
52
53 (see definition [3]), a node represents a lineage at the instant of a speciation event (obviously a
54
55
56
57
58
59
60

de Queiroz

Nodes, Branches, and Phylogenetic Definitions - 9

1
2
3 simplification, given that speciation is a process). In this context, Martin et al. (2010)
4
5 presumably concluded that the circumscribed group is polyphyletic because its basal node (the
6
7 one between branches 2, 4, and 5) is a speciation event rather than an ancestor, and the ancestor
8
9 (represented by branch 2) is not included in the circumscription (see Fig. 1a). By contrast, a
10
11 lineage at the time of a speciation event is an ancestor (even if it is only part of the entire
12
13 ancestral species), and therefore the concept of a node-based clade as defined in the *PhyloCode*
14
15 includes the ancestor so that the group is monophyletic.
16
17
18
19

20 Here it is also worth noting that there is a contradiction between the designations of the
21
22 entities as node-based and branch-based *clades* and Martin et al.'s conclusions that the
23
24 circumscribed groups are not monophyletic. A clade is monophyletic *by definition*.
25
26
27
28

29 PHYLOGENETIC DEFINITIONS TERMINOLOGY

30
31
32 Contrary to Martin et al. (2010:10), the terms “node-based” and “branch-based” as applied to
33
34 phylogenetic definitions are not intended to imply different concepts of clades or monophyly.
35
36 Instead, the clades designated by different kinds of phylogenetic definitions “are fundamentally
37
38 similar in being conceptualized and defined as parts of phylogeny each of which is composed of
39
40 an ancestor and all of its descendants” (de Queiroz 2007:957). The different categories of clades
41
42 simply represent convenient ways of partitioning the idealized representation of phylogeny as a
43
44 tree in the graph theoretic sense (de Queiroz 2007). Thus, the most recent common ancestor of
45
46 A and B in a node-based definition (1) is equated with the node in which the ancestral lineages of
47
48 A and B coalesce (Fig. 3a), while the earliest ancestor of A that is not an ancestor of C in a
49
50 branch-based definition is equated with the branch representing the first species (lineage
51
52 segment) in the ancestral lineage of A after it diverged from the ancestral lineage of C (Fig. 3b).
53
54
55
56
57
58
59
60

de Queiroz

Nodes, Branches, and Phylogenetic Definitions - 10

1
2
3 This distinction corresponds roughly to Martin et al.'s (2010) distinction between a rooted but
4 non-planted tree (in which the root node has two "children", represented by the two internal
5 specifiers in a fundamental node-based definition [1]) versus a planted tree (in which the root has
6 a single "child", represented by the single internal specifier in a fundamental branch-based
7 definition [2]), except that in the case of branch-based definitions the node at the base of the
8 specified branch is not included.
9
10
11
12
13
14
15
16

17 According to Martin et al. (2010:10), "the mathematical definition of a graph does not permit
18 such a structure; one cannot have an edge without both its endpoints" (p. 10). If that convention
19 were to be used to argue against the exclusion of the basal node generally, such an argument
20 would be overly formalistic in placing more importance on a graph theoretic convention than on
21 the accurate representation of phylogeny. In a lineage tree, the basal node cannot be part of a
22 specified branch-based clade, because it represents either a speciation event (under the
23 representational system of Martin et al.) or the last part of a more remote ancestral species than
24 the one specified by the definition (under the representational system of the *PhyloCode*), neither
25 of which are members of the specified (branch-based) clade. However, Martin et al. were
26 referring to the interpretation of a branch-based clade in the context of a relationship tree and the
27 fact that the specified clade supposedly includes only one of the two immediate descendants of
28 its basal node. On the contrary, as I argued above, any branch-based definition specifying a
29 clade that includes a particular ancestor (in this case, the one represented by node 1 in Fig. 2)
30 would also include all of the descendants of that ancestor (Fig. 2b), while any branch-based
31 definition specifying a clade that excludes any descendants of a particular ancestor (e.g., node 1)
32 would also exclude that ancestor (Fig. 2c). The reason for the discrepancy is that Martin et al.
33 (2010) applied the concept of a branch-based clade in a context in which it was not intended to
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

de Queiroz

Nodes, Branches, and Phylogenetic Definitions - 11

1
2
3 be applied (i.e., that of a relationship tree) and assumed inclusion of the node at the beginning of
4
5 the basal branch (as dictated by a graph theoretic convention) rather than considering which
6
7 ancestor would have been specified by an actual branch-based definition.
8
9

10
11 Despite the fact that both node-based and branch-based definitions are applicable in the
12
13 context of both relationship and lineage trees, some of the distinctions made both by Martin et al.
14
15 (2010) and in this contribution reveal that the current terminology regarding phylogenetic
16
17 definitions is potentially confusing. In the case of the term “node-based definition”, the potential
18
19 confusion involves the conceptualization of the inferred ancestor. For those people who would
20
21 allow the last common ancestor of A and B to be only part of a species (see Fig. 2 of de Queiroz
22
23 and Donoghue 1990), it seems reasonable to represent that ancestor as a node representing the
24
25 last “instant” that a lineage existed before diverging into two lineages (obviously an abstraction)
26
27 on a lineage tree (Fig. 3a). By contrast, as discussed by Gauthier and de Queiroz (2001), other
28
29 people might conceptualize the last common ancestor of A and B as an entire ancestral species.
30
31 For those people, the ancestor in question should be represented as a branch rather than a node
32
33 on a lineage tree, in which case it makes little sense to designate the definition “node-based.”
34
35
36
37
38

39
40 In addition, as noted above, the term “branch-based definition” makes little sense if applied
41
42 in the context of a relationship tree. The reason is that the name “branch-based” derives from
43
44 representation of the designated ancestor (the one in which the named clade originated) as a
45
46 branch on a phylogenetic tree, but that representation applies only to lineage trees. On
47
48 relationship trees, where the branches represent relationships rather than ancestral species, clades
49
50 always originate in nodes rather than branches.
51
52

53
54 In order to avoid confusion with respect to how the designated ancestors are conceptualized
55
56 or represented, it would seem preferable to adopt a terminology that avoids equating those
57
58
59
60

de Queiroz

Nodes, Branches, and Phylogenetic Definitions - 12

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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 Thollesson (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

de Queiroz

Nodes, Branches, and Phylogenetic Definitions - 13

[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 versus 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).

THE UTILITY OF DIFFERENT KINDS OF DEFINITIONS

Martin et al. (2010:10) 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,

de Queiroz

Nodes, Branches, and Phylogenetic Definitions - 14

1
2
3 although correct in the context of complete and error-free knowledge of phylogeny, is over-
4
5 simplified in the context of real (i.e., incomplete and at least potentially inaccurate) phylogenetic
6
7 hypotheses, which is to say that there are pragmatic reasons for using both kinds of definitions
8
9 even in the context of a single kind of tree. In the rest of this section, I will assume, for the sake
10
11 of simplifying the discussion, that the ancestor in which a clade originates is an entire species.
12
13

14
15 Crown and total clades are two significant categories of clades defined with respect to the
16
17 extant biota (reviewed by de Queiroz 2007). A *crown clade* is a clade in which both (all)
18
19 lineages originating from a particular splitting event have extant descendants (which is not to say
20
21 that they do not also have extinct ones); a *total clade* is a clade composed of a crown clade and
22
23 all extinct species (stem group species) that are more closely related to that crown clade than to
24
25 any other crown clade or extant species (see Hennig 1965, 1969, 1981; Jeffreys 1979;
26
27 Lauterbach 1989; Meier and Richter 1992; de Queiroz 2007). (Given that the stem species is
28
29 part of [and equivalent to] the clade that originates in it [Hennig 1966], a species that has not yet
30
31 given rise to any daughter species can be considered a clade composed of that one species, and
32
33 thus a single extant species that has not given rise to any daughter species can be considered a
34
35 crown clade. Based on the same premise, an *ancestor per se definition* is a phylogenetic
36
37 definition taking the form “the clade originating in D”, where D is a species known from
38
39 specimens [as opposed to one inferred solely from a tree], whether extinct or extant.)
40
41
42
43
44

45
46 A fact of real phylogenetic hypotheses (as opposed to hypothetical ones, as in Fig. 4) is that
47
48 they are almost invariably based on incomplete species sampling, particularly with regard to
49
50 extinct species. However, to define the name of a crown clade using a maximum clade (“branch-
51
52 based”) definition, it would be necessary to know at least one species (and in some cases there
53
54 will have been only one) of the immediate (often extinct) sister group of the crown clade in
55
56
57
58
59
60

de Queiroz

Nodes, Branches, and Phylogenetic Definitions - 15

1
2
3 question (Fig. 5). Such a species (X in Fig. 5) would be a necessary (and sufficient) external
4 specifier in a maximum clade definition of the name of the crown clade—for instance, max (A
5 not X) in the example (Fig. 5). Therefore, such a species would not only have to have been
6 discovered, it would also have to have been inferred to be the sister group of the crown clade to
7 be named. If a different external specifier (such as Y or Z in Fig. 5) were used instead, the name
8 would not apply to the crown clade. By contrast, knowledge of the immediate sister species of a
9 crown clade is not necessary to define the name of that crown clade using a minimum clade
10 (“node-based”) definition. Thus, in Figure 5, the following minimum clade defining formula
11 would apply the name to the crown clade regardless of knowledge of any extinct species: min
12 (A and B). Given that the immediate (often extinct) sister species of a crown clade commonly
13 will not be known, it will generally be more feasible to define the names of crown clades using
14 minimum clade (“node-based”) definitions.

15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32 Conversely, to define the name of a total clade using a minimum clade (“node-based”)
33 definition, it would be necessary to know at least one species (and in some cases there will have
34 only been one) representing the extinct clade resulting from the very first divergence within the
35 total clade in question (Fig. 6). Such a species (Z in Fig. 6) would be one of the necessary
36 internal specifiers in the minimum clade definition—for instance: min (A and Z) in the example
37 (Fig. 6). Therefore such a species would not only have to have been discovered, it would also
38 have to have been inferred to be a representative of the extinct clade resulting from the very first
39 divergence within the total clade to be named. If a different internal specifier (such as X or Y in
40 Fig. 6) were used instead, the name would not apply to the total clade. By contrast, knowledge
41 of a species representing the extinct clade resulting from the very first divergence within a total
42 clade is not necessary to define the name of that total clade using a maximum clade (“branch-
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

de Queiroz

Nodes, Branches, and Phylogenetic Definitions - 16

1
2
3 based”) definition. Thus, in Figure 6, the following maximum clade defining formula would
4
5 apply the name to the total clade regardless of knowledge of any extinct species: $\max (A \text{ not } C)$.
6
7 Given that a species representing the extinct clade resulting from the very first divergence within
8
9 a total clade often will not be known, it will generally be more feasible to define the names of
10
11 total clades using maximum clade (“branch-based”) definitions.
12
13

14
15 Naming crown versus total clades is not the only context in which minimum clade (“node-
16
17 based”) and maximum clade (“branch-based”) definitions exhibit important and useful
18
19 differences. In addition, the two kinds of phylogenetic definitions have different consequences
20
21 in the context of revisions to phylogenetic hypotheses that apply to clades more generally. In the
22
23 following examples, the same conclusions apply regardless of whether the clades are crown
24
25 clades, total clades, both, or neither. Consider the following definitions: $Athena := \min (A \text{ and}$
26
27 $B)$; $Minerva := \max (A \text{ not } C)$. In the context of Figure 7a, the two names apply to the same
28
29 clade (at least if the ancestor[s] in which the specified clades originated are considered to be
30
31 entire species rather than parts thereof): the one composed of A and B and their inferred
32
33 common ancestor. In the context of this phylogenetic hypothesis, the names *Athena* and *Minerva*
34
35 are (heterodefinitional) synonyms. By contrast, in the context of Figures 7b and 7c, the names
36
37 apply to different clades and therefore are no longer synonyms. In the case of Figure 7b, *Athena*
38
39 is composed of A and B and their inferred common ancestor, while *Minerva* is composed of A, B,
40
41 and D, and two inferred ancestors. By contrast, in the case of Figure 7c, *Athena* is composed of
42
43 A, B, C, and two inferred ancestors, while *Minerva* is composed of only A, thus reversing the
44
45 hierarchical relationships of the named clades from the previous case (Fig. 7b).
46
47
48
49
50
51
52

53 These examples demonstrate that different kinds of phylogenetic definitions have different
54
55 consequences in the context of alternative phylogenetic hypotheses. Which kind of definition is
56
57
58
59
60

de Queiroz

Nodes, Branches, and Phylogenetic Definitions - 17

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).

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 [Internet].
Version 4c. Available from: <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.” Pages 7–41 in *New perspectives on the origin and early evolution of birds*
(J. Gauthier and L. F. Gall, eds.). Peabody Museum of Natural History, Yale University,
New Haven, Connecticut.

1 de Queiroz

Nodes, Branches, and Phylogenetic Definitions - 19

2
3 Hennig W. 1965. Phylogenetic systematics. *Annu. Rev. Entomol.* 10:97–116.

4
5 Hennig W. 1966. Phylogenetic systematics. University of Illinois Press, Urbana.

6
7
8 Hennig W. 1969. Die Stammesgeschichte der Insekten. Kramer, Frankfurt.

9
10
11 Hennig W. 1981. *Insect phylogeny*. Wiley, Chichester, UK.

12
13 Jefferies R.P.S. 1979. The origin of chordates—A methodological essay. Pages 443–477 in *The*
14
15 *origin of major invertebrate groups* (M. R. House, ed.). Academic Press, London.

16
17
18 Lauterbach K.-E. 1989. Das Pan-Monophylum—Ein Hilfsmittel für die Praxis der
19
20 phylogenetischen Systematik. *Zool. Anz.* 223:139–156.

21
22
23 Lee M.S.Y. 1998. Phylogenetic uncertainty, molecular sequences, and the definition of taxon
24
25 names. *Syst. Biol.* 47:719–726.

26
27
28 Martin J., Blackburn D., Wiley E.O. 2010. Are node-based and stem-based clades equivalent?
29
30 Insights from graph theory. *PLOS Currents Tree of Life* doi: 10.1371/currents.RRN1196.

31
32
33 Meier R., Richter S. 1992. Suggestions for a more precise usage of proper names of taxa:
34
35 Ambiguities related to the stem lineage concept. *Z. Zool. Syst. Evol.* 30:81–88.

36
37
38 Schander C., Thollesson M. 1995. Phylogenetic taxonomy—Some comments. *Zool. Scr.*
39
40 24:263–268.

41
42
43 Sereno P.C. 1999. Definitions in phylogenetic taxonomy: Critique and rationale. *Syst. Biol.*
44
45 48:329–351.

de Queiroz

Nodes, Branches, and Phylogenetic Definitions - 20

Figure Captions

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 most recent common ancestor (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 most recent common ancestor (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.

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

de Queiroz

Nodes, Branches, and Phylogenetic Definitions - 21

1
2
3 name is defined as designating the clade originating in the earliest ancestor of 7 that is not an
4
5 ancestor of 10. The earliest ancestor of 7 that is not an ancestor of 10 is 1, so the designated
6
7 clade contains 1 and all of its descendants (2–9). c) Example illustrating that a using standard
8
9 branch-based definition to infer the composition of a clade that excludes species 2, 4, 5, 8, and 9
10
11 will also exclude species 1. Here the name is defined as designating the clade originating in the
12
13 earliest ancestor of 7 that is not an ancestor of 9. The earliest ancestor of 7 that is not an
14
15 ancestor of 9 is 3, and therefore the clade includes 3, 6 and 7 and excludes 2, 4, 5, 8, and 9 as
16
17 well as 1. In all three diagrams, the branches of the trees represent ancestor-descendant
18
19 relationships and the nodes represent species. Species included in the relevant group or clade are
20
21 represented by black nodes.
22
23
24
25
26
27
28

29 Figure 3. Rationale for the names “node-based” and “branch-based” phylogenetic definitions. a)
30
31 “Node-based” refers to the fact that definitions of this kind specify the ancestor in which the
32
33 clade being named originated as the point in the tree where the ancestral lineages of the internal
34
35 specifiers (in this case, A and B) coalesce. It is natural to equate such a point with a node,
36
37 represented by the black dot in the diagram. b) “Branch-based” refers to the fact that definitions
38
39 of this kind specify the ancestor in which the clade being named originated as the first segment
40
41 in a lineage that is ancestral to the internal specifier (in this case, A) but not to the external
42
43 specifier (in this case, C). It is natural to equate such a lineage segment with a branch,
44
45 represented by the thick line with an arrow in the diagram. In both diagrams, the trees are
46
47 lineage trees in the sense described in the *PhyloCode* (where a node represents an ancestor at the
48
49 instant of a speciation event rather than the speciation event itself); the specifiers are positioned
50
51
52
53
54
55
56
57
58
59
60

de Queiroz

Nodes, Branches, and Phylogenetic Definitions - 22

1
2
3 at the ends of branches to indicate that the clades may include other species that are not shown
4
5 (e.g., species that are more closely related to A than to B).
6
7
8
9

10 Figure 4. If the ancestor in which a clade originates is considered to be an entire species (rather
11 than the last part of a species), it might seem that minimum and maximum clade definitions are
12 simply different methods for specifying the same clades and ancestors. For the relationship tree
13 illustrated in (a) the last common ancestor of A and B (minimum clade definition) and the
14 earliest ancestor of A that is not an ancestor of C (maximum clade definition) refer to the same
15 clade, the members of which are represented by black nodes. The same is true for the lineage
16 tree illustrated in (b), where the members of the specified clade are represented by thicker
17 branches. Although this conclusion of equivalency would be correct if all species that have ever
18 existed and their phylogenetic relationships were known with certainty, its limitations in the
19 context of incomplete and provisional knowledge are illustrated in Figures 5–7.
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35

36 Figure 5. Limitations of maximum clade (“branch-based”) and utility of minimum clade (“node-
37 based”) definitions for defining the names of crown clades. In order to define the name of the
38 crown clade composed of species A and B using a maximum clade definition, a member of the
39 immediate sister group of the crown clade—in this case, extinct species X—must be used as an
40 external specifier in the defining formula (e.g., max [A not X]). If species X has not been
41 discovered and the name of a different member of the stem group (e.g., Y or Z) is used instead,
42 the name will not refer to the crown clade. By contrast, the minimum clade defining formula
43 min (A and B) can be used to apply a name to the crown clade without knowledge of X or any
44 other member of the stem group. The tree illustrated here is a lineage tree and the ancestors in
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

de Queiroz

Nodes, Branches, and Phylogenetic Definitions - 23

1
2
3 which clades originate are considered to be entire species (rather than parts thereof). The vertical
4
5 component of the diagram represents time, with the present represented by the level of the
6
7 terminal nodes of branches A, B, and C. Members of the crown clade to be named are indicated
8
9 by thicker branches.
10
11

12
13
14
15 Figure 6. Limitations of minimum clade (“node-based”) and utility of maximum clade (“branch-
16
17 based”) definitions for defining the names of total clades. In order to define the name of the total
18
19 clade corresponding to the crown clade composed of species A and B using a minimum clade
20
21 definition, a member of the extinct group representing the first divergence within the total
22
23 clade—in this case, extinct species Z—must be used as an internal specifier in the defining
24
25 formula (e.g., min [A and Z]). If species Z has not been discovered and the name of a different
26
27 member of the stem group (e.g., X or Y) is used instead, the name will not refer to the total clade.
28
29 By contrast, the maximum clade defining formula max (A not C) can be used to apply a name to
30
31 the total clade without knowledge of Z or any other member of the stem group. The tree
32
33 illustrated here is a lineage tree and the ancestors in which clades originate are considered to be
34
35 entire species (rather than parts thereof). The vertical component of the diagram represents time,
36
37 with the present represented by the level of the terminal nodes of branches A, B, and C.
38
39 Members of the total clade to be named are indicated by thicker branches.
40
41
42
43
44
45
46
47

48
49 Figure 7. Different consequences of minimum clade (“node-based”) and maximum clade
50
51 (“branch-based”) definitions in the context of alternative phylogenetic hypotheses. The
52
53 following definitions are applied in the context of the three phylogenetic hypotheses illustrated in
54
55 a–c: *Athena* := min (A and B); *Minerva* := max (A not C). In the context of phylogenetic
56
57
58
59
60

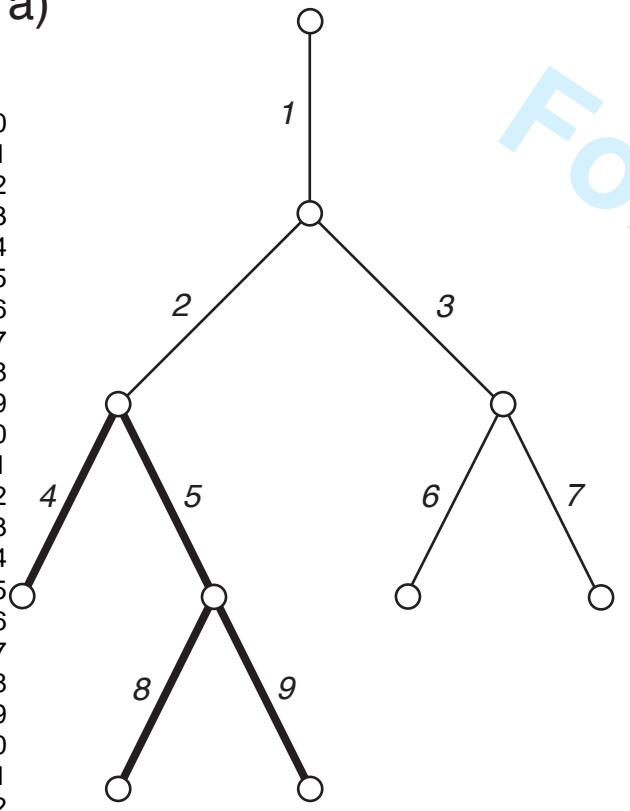
de Queiroz

Nodes, Branches, and Phylogenetic Definitions - 24

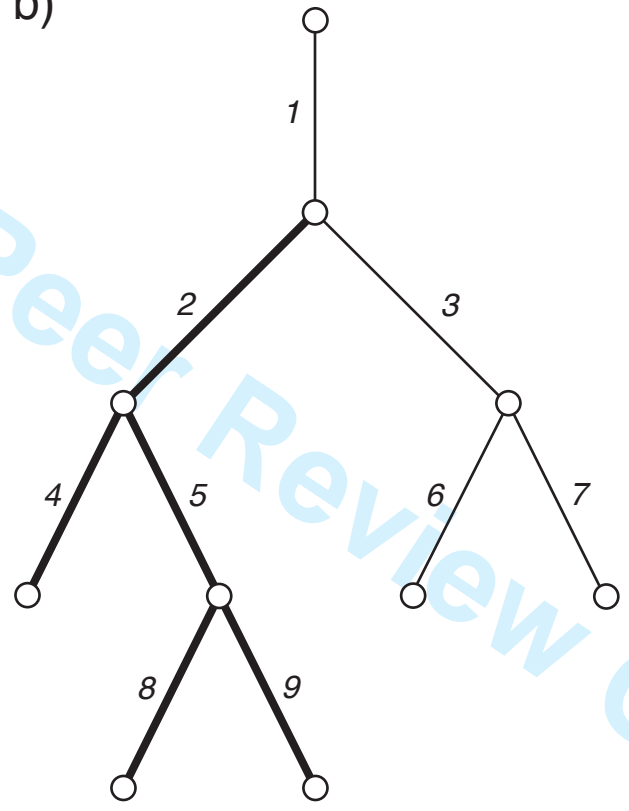
1
2
3 hypothesis (a), both *Athena* and *Minerva* refer to the same clade (the names are synonymous). In
4
5 the context of hypothesis (b), *Minerva* refers to a more inclusive clade than does *Athena*. In the
6
7 context of hypothesis (c), *Athena* refers to a more inclusive clade than does *Minerva*. All three
8
9 trees are relationship trees (branches represent relationships; nodes represent species). Grey-
10
11 filled circles = members of *Athena* only; black-filled circles = members of *Minerva* only; grey-
12
13 and black-filled circles = members of both *Athena* and *Minerva*; unfilled circles = members of
14
15 neither *Athena* nor *Minerva*.
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

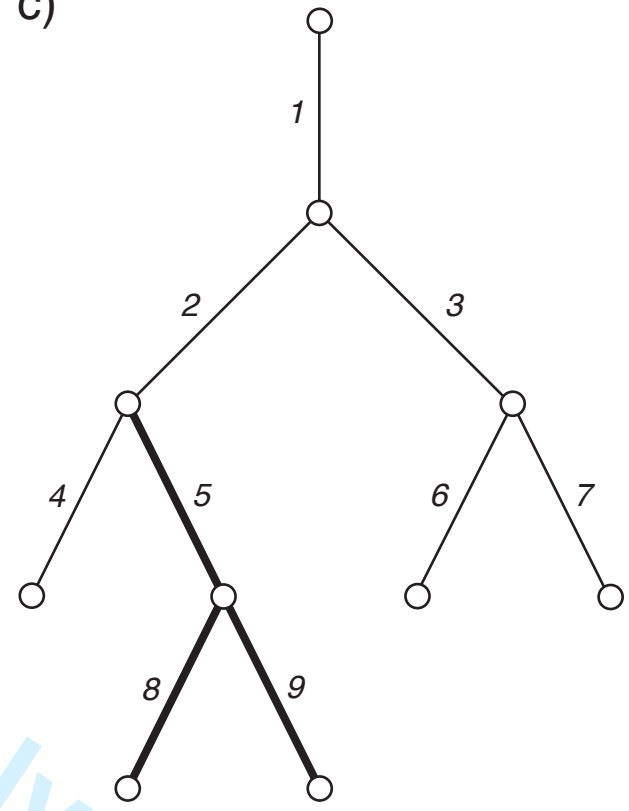
a)



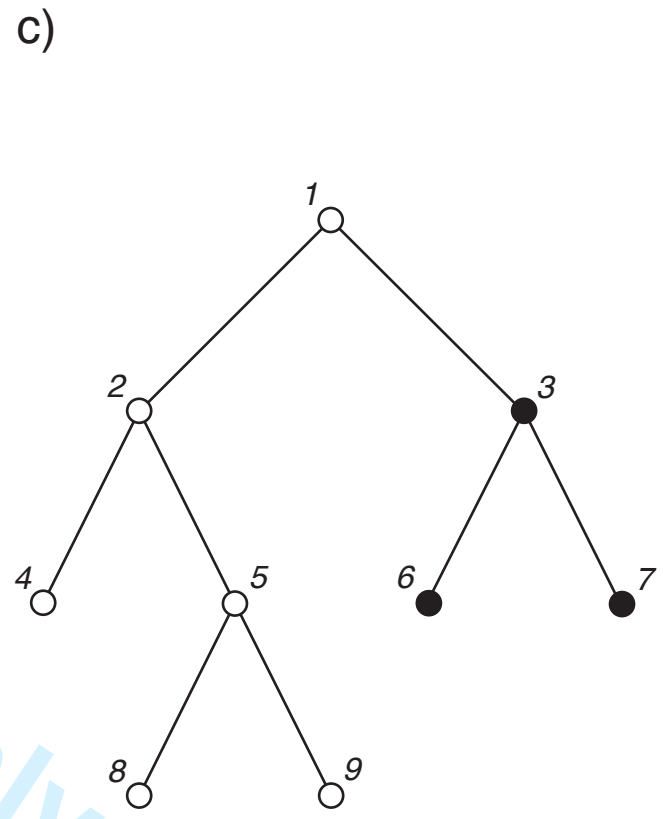
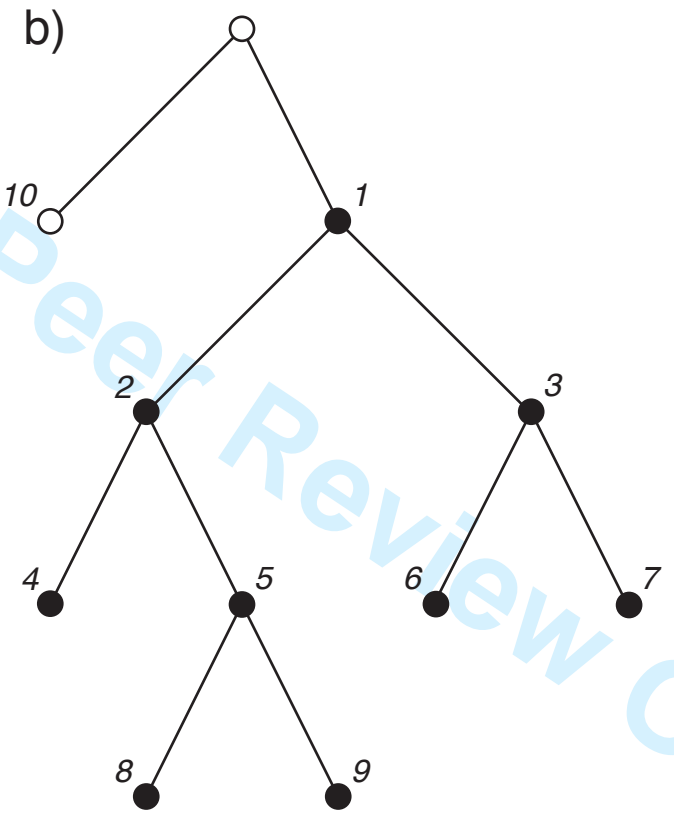
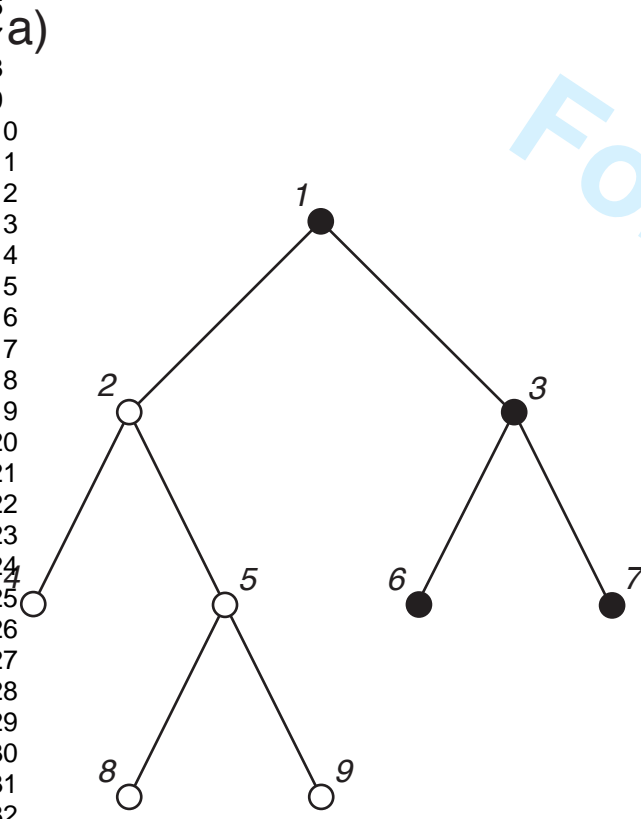
b)



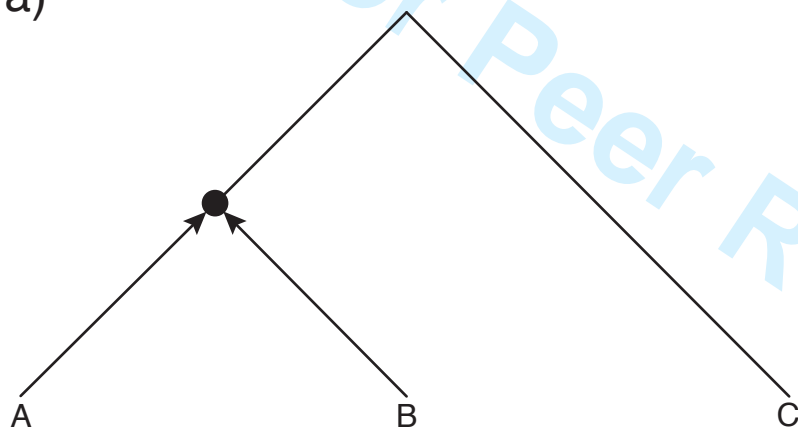
c)



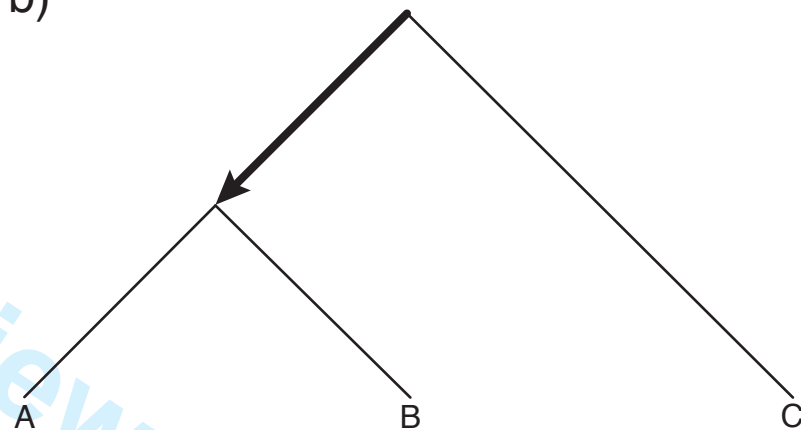
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47



a)



b)

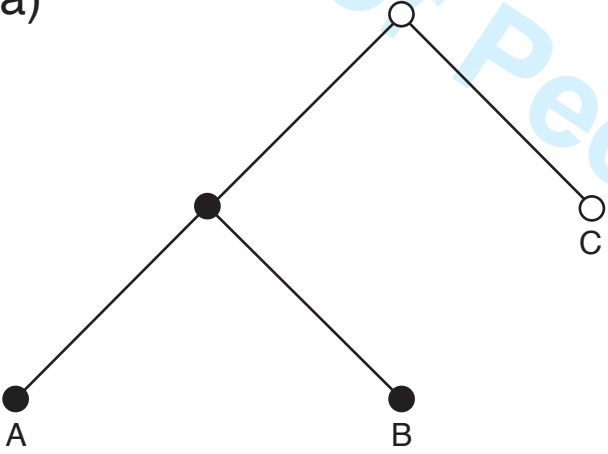


For Peer Review Only

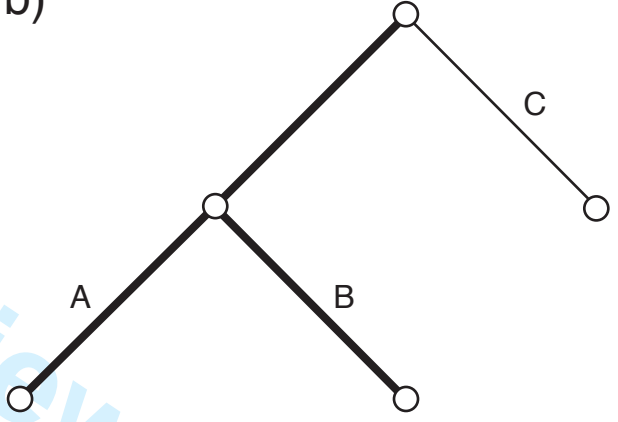
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

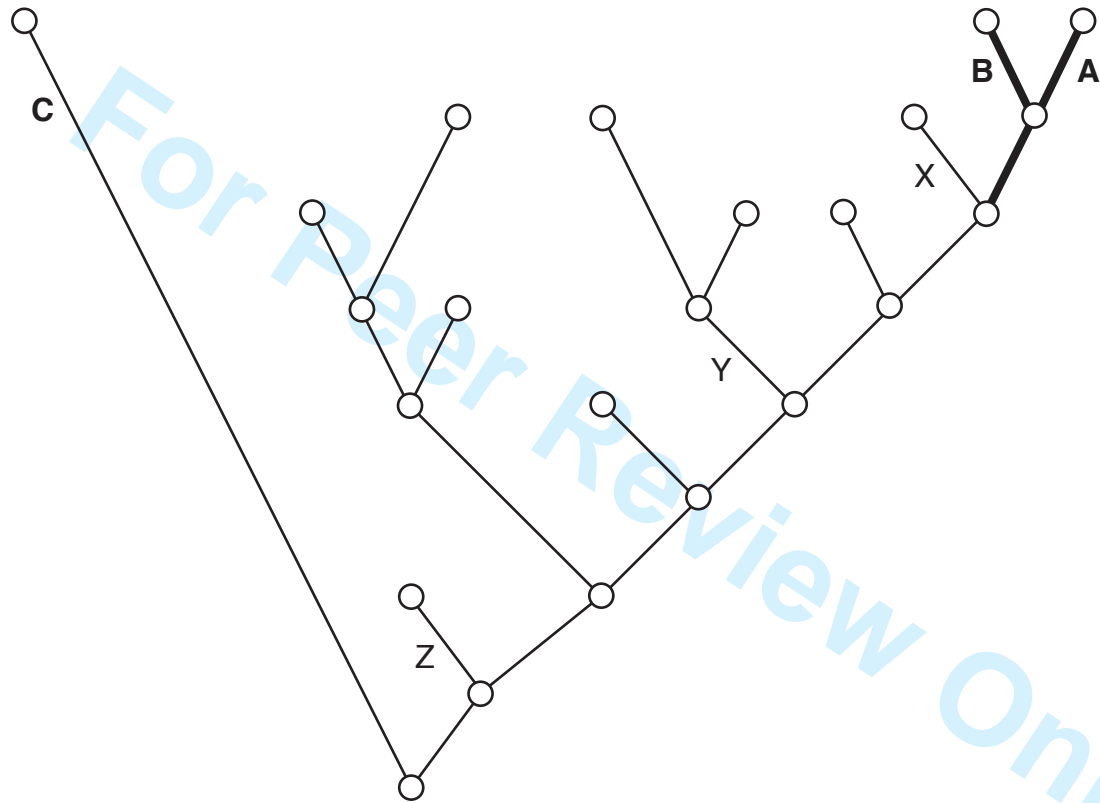
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

a)



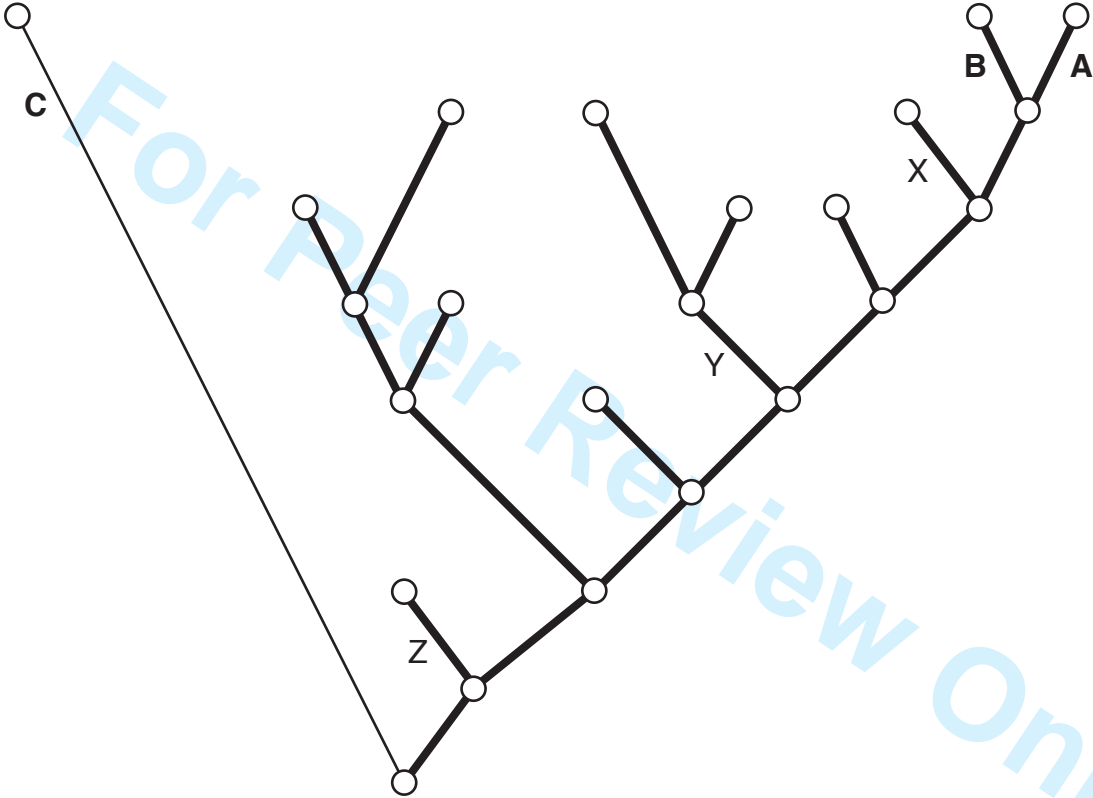
b)





1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

