Phylogenetic Nomenclature, Three-Taxon Statements, and Unnecessary Name Changes

KEVIN DE QUEIROZ1,* AND MICHAEL J. DONOGHUE2

1Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20560-0162, USA; and 2Department of Ecology and Evolutionary Biology and Peabody Museum of Natural History, Yale University, New Haven, CT 06520-8106, USA;

*Correspondence to be sent to: Division of Amphibians and Reptiles, National Museum of Natural History, Smithsonian Institution, PO Box 37012, NHB, W-200, MRC 162, Washington, DC 20013-7012, USA; E-mail: dequeirozk@si.edu.

Received 4 April 2011; reviews returned 13 June 2011; accepted 16 June 2011

Associate Editor: Frank E. Anderson

Criticisms of phylogenetic nomenclature (see de Queiroz and Gauthier 1994) and of the PhyloCode (Cantino and de Queiroz 2010) have been addressed in the literature (e.g., de Queiroz 1997, 2006; Cantino 2000; de Queiroz and Cantino 2001; Lee 2001; Bryant and Cantino 2002; Laurin et al. 2005; see http://www.phylnames.org/ [literature/replies to critiques] for additional references). Many of the supposed problems result from misunderstandings, whereas others can and have been dealt with by making appropriate adjustments (including changes in the PhyloCode). One recent argument against phylogenetic nomenclature has not yet been addressed, at least not in print. That argument was presented by Platnick in a debate that took place at the Linnean Society of London on 5 December 2005 and was later published in his “Letter to Linnaeus” (Platnick 2009). Although we responded to Platnick’s argument verbally during the debate, we take the opportunity to do so here because some readers of his letter may believe that his conclusions are damaging to phylogenetic nomenclature.

Platnick’s argument involves a numerical comparison of the information content, as measured by implied three-taxon statements (propositions about cladistic relationships), of what he called “Linnaean classification” versus a “node-based system” (we place these terms in quotation marks to indicate that they are misleading, as will be explained below). Here, we show that Platnick’s negative conclusions about phylogenetic nomenclature are incorrect and result from inappropriate comparisons. Moreover, they are based on a criterion that is irrelevant for comparing different approaches to nomenclature. When a nomenclaturally relevant criterion is used, phylogenetic nomenclature outperforms its rank-based counterpart in all of the examples considered by Platnick.

THREE-TAXON STATEMENTS

Platnick’s comparisons are based on an example originally used by de Queiroz and Gauthier (1994) to illustrate the consequences of phylogenetically
defined names in the context of a new hypothesis of phylogenetic relationships. In that example, *Agamidae* and *Chamaeleonidae* (two lizard taxa traditionally ranked as families) are composed of species that are originally thought to be mutually exclusive in terms of their cladistic relationships (Fig. 1a), but a new phylogenetic hypothesis suggests that some of the species referred to *Agamidae* are more closely related to the species in *Chamaeleonidae* than to other species in *Agamidae* (Fig. 1b). This example is simplified from a real case described by Frost and Etheridge (1989), and Platnick’s presentation of it implicitly but incorrectly placed *Agama* in Clade ABC, which he treated as more distantly related to *Chamaeleonidae* than is clade DEF (under the new phylogenetic hypothesis). We have corrected his examples by applying names based on the name *Agama* to Clade DEF, the one that is more closely related to *Chamaeleonidae*, and we have applied a real name (based on the name *Leiolepis*) to the more distantly related clade ABC. We have also added the more inclusive taxon *Acrodonta* (which was not ranked but could have been ranked as an infraorder) as it is relevant to some of the examples. These changes were made for the sake of accuracy, thus better connecting the examples to previous literature; none of them affect Platnick’s or our own conclusions about the number of implied three-taxon statements.

Under standard node-based phylogenetic definitions based on the original hypotheses of composition (Fig. 1), the names “*Agamidae*” and “*Chamaeleonidae*” end up referring to nested rather than mutually exclusive taxa under the new phylogenetic hypothesis (Fig. 1b). According to Platnick, the original “Linnaean classification” allows one to deduce only 18 three-taxan statements: (GH)A, (GH)B, and so forth, representing a reduction in the information content of the classification by about 71%.

We do not dispute the number of three-taxan statements that can be deduced from the two taxonomies; however, the comparison itself is inappropriate in that it matches a “Linnaean classification” under the original phylogenetic hypotheses against a “node-based classification” under the new phylogenetic hypothesis. Appropriate comparisons would use either the original phylogenetic hypothesis or the new hypothesis, not a mixture of the two. If we use the original hypothesis, the “node-based classification” allows one to deduce the same 63 three-taxan statements as does the equivalent “Linnaean classification”; no information is lost. Note that considering the taxa to be unranked is irrelevant as phylogenetic nomenclature does not prohibit the use of ranks. The only relevant consideration is that application of the names is determined by phylogenetic definitions. Conversely, if we use the new phylogenetic hypothesis, then a “Linnaean classification” based on uniting (lumping) the two families (as in the actual proposal of Frost and Etheridge 1989) for the same two nested clades considered in the “node-based classification” allows one to deduce the same 18 three-taxan statements as under the corresponding “node-based classification.” Thus, regardless of whether the application of taxon names is governed by rank-based or phylogenetic definitions, the taxonomies contain exactly the same amount of information in the form of implied three-taxan statements.

![Diagram](image.png)

**Figure 1.** Example used by de Queiroz and Gauthier (1994) to illustrate the application of names under phylogenetic nomenclature in the context of different hypotheses of phylogenetic relationships. a) Original phylogenetic hypothesis, b) new phylogenetic hypothesis. Open and filled circles in the original example have been replaced with the letters A–F and G–I (respectively), as was done by Platnick (2009), and the more inclusive taxon *Acrodonta* has been added. For the examples involving phylogenetically defined names in the sections Three-Taxon Statements and Unnecessary Name Changes, the following standard node-based definitions are used: *Acrodonta* = the smallest clade containing both A and G; *Agamidae* = the smallest clade containing both A and D; *Chamaeleonidae* = the smallest clade containing G, H, and I; *Leiolepidinae* = the smallest clade containing A, B, and C; *Agaminae* = the smallest clade containing D, E, and F.
The “Linnaean classification” based on lumping the two families, as used in de Queiroz and Gauthier’s original example, is not the only possible way to make all names refer to monophyletic taxa under the new phylogenetic hypothesis and rank-based nomenclature. Thus, Platnick considered two alternative “Linnaean classifications” under the new phylogenetic hypothesis: first, a taxonomy based on dividing (splitting) the family Agamidae, rather than lumping it with Chamaeleonidae, but leaving the more inclusive clades (ABCDEFGH and DEFGHI) unnamed:

- Leiolepididae (ABC)
- Agamidae (DEF)
- Chamaeleonidae (GHI)

And second, a fully subordinated taxonomy based on both splitting and lumping in which all five hypothesized clades are named:

- Chamaeleonoidea (ABCDEFGH)
- Leiolepididae (ABC)
- Chamaeleonidae (DEFGHI)
- Agaminae (DEF)
- Chamaeleoninae (GHI)

According to Platnick, the taxonomy based on splitting but leaving the inclusive clades unnamed implies 54 three-taxon statements, whereas the fully subordinated taxonomy implies 81 three-taxon statements, both of which he considered to compare favorably with the 18 three-taxon statements implied by the “node-based taxonomy.”

Again, we do not dispute Platnick’s conclusions about the numbers of three-taxon statements that can be deduced from the alternative taxonomies; however, once again, they involve inappropriate comparisons. In this case, all of the comparisons are based on the new phylogenetic hypothesis, but this time they involve taxonomies with different numbers of named clades: three in the case of the “Linnaean classification” based on splitting but leaving the inclusive clades unnamed versus five in the case of the fully subordinated “Linnaean classification” versus two in the case of the “node-based classification.” The appropriate comparison with the “Linnaean classification” based on splitting and leaving the inclusive clades unnamed is a “node-based classification” that recognizes the same three clades, but with phylogenetically defined names:

- Leiolepidae (ABC)
- Agaminae (DEF)
- Chamaeleonidae (GHI)

This taxonomy implies exactly the same number of three-taxon statements (54) as does the corresponding “Linnaean classification”. Similarly, the appropriate comparison with the fully subordinated “Linnaean classification” is a “node-based classification” that recognizes the same five clades, but with phylogenetically defined names:

- Acrodonta = Agaminae (ABCDEFHGI)
- Leiolepidae (ABC)

[New name required] (DEFGHI)

Agaminae (DEF)
Chamaeleonidae (GHI)

This taxonomy implies exactly the same number of three-taxon statements (81) as does the fully subordinated “Linnaean classification.” In this case, both Acrodonta and Agaminae are considered candidate names for the largest clade because the PhyloCode is not yet in operation and therefore precedence between those names has not been established.

Thus, when appropriate comparisons are made between taxonomies (i.e., those based on the same phylogenetic hypothesis and involving the same named clades), those taxonomies contain exactly the same amount of information in terms of implied three-taxon statements regardless of whether the taxon names are governed by rank-based or phylogenetic nomenclature. This conclusion should be neither controversial nor surprising because the information is contained in the taxonomies, not in the names. The taxonomies tell us which species are included in particular clades, and that is the information from which three-taxon statements are deduced. By contrast, taxon names are simply shorthand symbols for referring to monophyletic groups of species that, by themselves, contain minimal information about which species are included in the named clades (e.g., species of Agama are included in Agaminae). These considerations reflect the distinction between taxonomy and nomenclature (see de Queiroz 2006), which was largely ignored by Platnick. When that distinction is taken into consideration, it is clear that the information content of taxonomies is irrelevant to the choice between rank-based versus phylogenetic nomenclature.

UNNECESSARY NAME CHANGES

A property that is relevant to the choice between those approaches (i.e., one that involves nomenclatural rather than taxonomic considerations) is how well the different systems preserve the relationships between names and clades. Unfortunately, this issue has been confused by discussions of stability. For example, Platnick (2009, p. 180) quoted a statement by de Queiroz and Gauthier (1994) about the general goal of stability as if to suggest first, that advocates of phylogenetic nomenclature are trying to achieve constancy of taxon composition at all costs, and second, that they fail to do so as shown by the case of Agaminae in the examples considered. Contrary to Platnick’s implication, the proposition is not that constancy of taxon composition should be maintained at all costs. Instead, the proposition is that taxon names should change their references to different sets of species only if required to do so by a change in the phylogenetic hypothesis and, therefore, that the name of a taxon should not change if hypotheses about the monophyly and composition of that taxon have not changed. Traditional rank-based nomenclature clearly does not achieve this goal in particular cases, as demonstrated by the original taxonomic proposal involving lumping,
in which the name of clade ABCDEFGHI changed from *Acrodonta* to *Chamaeleonidae* and the name of clade GHI changed from *Chamaeleonidae* to *Chamaeleoninae*. To advocates of phylogenetic nomenclature, these sorts of changes are not only unnecessary but also counterproductive.

Considering the issue more generally, we can compare the numbers of unnecessary name changes (i.e., those in which the hypothesized monophyly and composition of the clade is unchanged) in the context of the three taxonomies considered by Platnick (those based on the new phylogenetic hypothesis) under rank-based versus phylogenetic nomenclature. It is useful as background first to consider the original taxonomy and set of names (i.e., under the earlier phylogenetic hypothesis), which would have been the same under both rank-based and phylogenetic nomenclature and which provides a basis for comparison when assessing name changes:

\[
\begin{align*}
\text{*Acrodonta* (ABCDEFGHI)} &
\text{*Agamidae* (ABCDEF)} \\
\text{*Leiolepidinae* (ABC)} &
\text{*Agaminae* (DEF)} \\
\text{*Chamaeleoninae* (GHI)}
\end{align*}
\]

The first case concerns the actual proposal under rank-based nomenclature (Frost and Etheridge 1989), which involved lumping the two original families (we will consider the full set of recognized names rather than only those with -idae endings, although the qualitative conclusions are the same in both cases):

\[
\begin{align*}
\text{*Chamaeleonidae* (ABCDEFGHI)} &
\text{*Leiolepididae* (ABC)} \\
\text{*Agaminae* (DEF)} &
\text{*Chamaeleoninae* (GHI)}
\end{align*}
\]

and the same taxonomy (recognized clades) with names applied according to phylogenetic nomenclature (using the standard node-based definitions given in the legend of Fig. 1):

\[
\begin{align*}
\text{*Acrodonta* = *Agamidae* (ABCDEFGHI)} &
\text{*Leiolepidinae* (ABC)} \\
\text{*Agaminae* (DEF)} &
\text{*Chamaeleoninae* (GHI)}
\end{align*}
\]

In this case, rank-based nomenclature causes two unnecessary name changes (Clades ABC and DEF), whereas phylogenetic nomenclature again causes none.

The next case concerns the taxonomy based on split-hypothesis, which involves lumping the two original families (we will consider the full set of recognized names rather than only those with -idae endings, although the qualitative conclusions are the same in both cases):

\[
\begin{align*}
\text{*Chamaeleonidae* (ABCDEFGHI)} &
\text{*Leiolepididae* (ABC)} \\
\text{*Agaminae* (DEF)} &
\text{*Chamaeleoninae* (GHI)}
\end{align*}
\]

and the same fully subordinated taxonomy with names applied according to phylogenetic nomenclature:

\[
\begin{align*}
\text{*Acrodonta* = *Agamidae* (ABCDEFGHI)} &
\text{*Leiolepidinae* (ABC)} \\
\text{[New name required] (DEFGHI)} &
\text{*Agaminae* (DEF)} \\
\text{*Chamaeleonidae* (GHI)}
\end{align*}
\]

In this case, rank-based nomenclature causes four unnecessary name changes (Clades ABCDEFGHI, ABC, DEFGHI, and GHI). Note that application of the name *Chamaeleonidae* to clade DEFGHI, which was not recognized under the earlier hypothesis, is considered to involve an unnecessary name change because that name was previously applied to clade GHI, whose monophyletic status and composition have not changed. By contrast, phylogenetic nomenclature causes no name changes, though it requires one new name (for the clade that was not inferred under the earlier phylogenetic hypothesis), and it might involve one necessary change in hypothesized composition (if *Agaminae* were to have precedence over *Acrodonta*). Thus, for all of the taxonomies considered by Platnick, phylogenetic nomenclature outperforms rank-based nomenclature according to the nomenclaturally relevant criterion of minimizing the number of unnecessary name changes.

### Jumping Spiders, Ranks, and Rank-specific Endings

Platnick presented one other example to illustrate a supposed problem with phylogenetic nomenclature. Wandering around John Murphy’s garden in England, Platnick came across a spider, which he identified as a salticid. Platnick has compiled a catalog of all known spider species. In that catalog, *Salticidae* is one of 108 taxa that Platnick ranks as families, and it includes 5088 species versus 34,794 species that are not salticids. From this information, Platnick inferred that the spider in John Murphy’s garden is more closely related to all of the other 5087 salticid species than to any of the 34,794 nonsalticid species, which would allow him to make 5087 × 34,794 = 176,997,078 three-taxon statements. Up to this point, we have no disagreement with Platnick; however, he then asserted that all of these inferences are possible “in a context provided—solely—
by the Linnaean hierarchy and the mutual exclusivity of equally ranked names it requires” (p. 184). This assertion is false. Simply to calculate the number of implied three-taxon statements requires only that the name Salticidae refers to a clade and that the numbers of included (salticid) and excluded (nonsalticid) species have been specified. It matters not at all whether that clade is ranked as a family. The same number of three-taxon statements would be implied if the name Salticidae had been applied using a phylogenetic definition to a clade ranked as something other than a family, or even to one that was not ranked at all.

It should also be noted that under rank-based nomenclature, there is no guarantee that the name Salticidae applies to a clade. Even knowing the rank-based definition of the name (i.e., Salticidae = the family containing the species Salticus senicus) does not rule out the possibility that the name refers to a paraphyletic or polyphyletic group (though knowing something about Platnick’s taxonomic philosophy, we infer that he did not apply the name to such a group in his catalog). By contrast, under phylogenetic nomenclature, a phylogenetic definition could be used to tie the name Salticidae explicitly to a hypothesized clade (e.g., Salticidae = the largest clade of spiders possessing enlarged and tube-shaped anterior median eyes with strip-shaped and tiered retinnae synapomorphic with those of Salticus senicus). And if that definition were known, it would be indisputable that the name referred to a clade (though its precise composition might vary depending on the hypothesized phylogeny).

Continuing with the same example, Platnick stated that under “the NB [node-based] system” . . . “one can no longer infer anything about the status of species included in other groups” (p. 184). He pointed out that Liphistiidae, the first family listed in his catalog (Salticidae is the last) could be a subgroup of Salticidae. Therefore, the fact that the spider that he encountered in John Murphy’s garden is a salticid would not allow him to make even a single three-taxon statement because (under the types of definitions used in the lizard example) it would not allow him to infer that the salticid from John Murphy’s garden must be more closely related to another salticid than to a liphistiid. Although the premise of this argument is true, the conclusion does not follow from it. To make an actual three-taxon statement (i.e., as opposed to calculating the number of possible three-taxon statements), one needs to know which particular species are assigned to the two taxa. To make this determination, one needs to consult a taxonomy, or better still, a tree with named clades. Such a consultation would immediately reveal not only the species included in Liphistiidae and Salticidae but also that the names refer to mutually exclusive taxa. And once one has that information, the same actual three-taxon statements (and thus the same total number of them) can be made regardless of whether the names are governed by rank-based or phylogenetic definitions. Again, this situation should not be surprising given that implied three-taxon statements and their numbers are taxonomic rather than nomenclatural issues.

**Different Kinds of Phylogenetic Definitions**

As noted above, the premise in Platnick’s salticid argument is true: under the types of phylogenetic definitions that were used in the lizard example, it would be possible for two nested clades to have names that both end in –idae. And if that were the case, it would not be possible, from the names alone (i.e., without a taxonomy or tree), to infer correctly that all species assigned to one of those two taxa (e.g., Salticidae) are more closely related to one another than they are to the species assigned to the other taxon (e.g., Liphistiidae). We have shown above that this situation does not prevent a person from making the same actual three-taxon statements as would be possible under traditional nomenclature. Nonetheless, Platnick, and perhaps others, may object to the very fact that two names with the same ending do not necessarily refer to mutually exclusive sets of species (though this situation exists for numerous names associated with higher ranks under traditional nomenclature—e.g., Araneae, Opisthothelae, and Mygalomorphae, names with identical endings that are applied to nested clades in some of Platnick’s classifications [e.g., Platnick 1977]).

In response to this criticism, we want to point out that the situation in question is not a general property of phylogenetic nomenclature but results instead from the particular type of node-based definitions that were used in the lizard example. The characterization of phylogenetic nomenclature as a “node-pointing” or “node-based” approach by Platnick and others is misleading, because node-based definitions are only one of the several types of phylogenetic definitions, and different types of definitions can have different consequences. For example, let us consider the same lizard example using branch-based rather than node-based definitions. Specifically, let us adopt the following branch-based definitions for the two original names with –idae endings: Agamidae = the largest clade containing D but not G; Chamaeleonidae = the largest clade containing G but not D. Under those definitions, the two names in question would be applied to clades under the new phylogenetic hypothesis as follows:

**Agamidae (DEF)**

**Chamaeleonidae (GHI)**

Again, no unnecessary name changes are involved, though the name Agamidae is associated with a necessary change in hypothesized composition (ABC is no longer included, as in the ranked taxonomy based on splitting). Moreover, the two names with –idae endings continue to refer to mutually exclusive clades.

Even using node-based definitions, the application of names with the same (traditionally rank-specific) endings to nested taxa can be avoided. For example, qualifying clauses can be used to nullify certain names under certain phylogenetic hypotheses (Lee 1998;
Cantino and de Queiroz 2010). Thus, suppose that the names were defined as follows: Agamidae = the smallest clade containing A (Leiolepis) and D (Agama) provided that it does not include G (Chamaeleo); Chamaeleonidae = the smallest clade containing G (Chamaeleo), H, and I, provided that it does not include D (Agama). Under those definitions, the two names in question would be applied under the new phylogenetic hypothesis as follows:

Agamidae (nullified)
Chamaeleonidae (GHI)

Again, no unnecessary name changes are involved. In this case, unlike the original one involving unqualified node-based definitions, the name Agamidae is nullified in the context of the second phylogenetic hypothesis because the smallest clade that contains both Agama and Leiolepis also contains Chamaeleon, violating the qualifying clause. Consequently, Agamidae would not be a synonym of Acrodonta, which would then be the only name applied to Clade ABCDEFGHI. Moreover, given that the name Agamidae would be nullified, it would not refer to a clade that includes Chamaeleonidae—that is, the two names that end in -idae would not refer to nested clades.

CONCLUSION

In sum, Platnick’s criticisms of phylogenetic nomenclature are misguided. His conclusions about the inferiority of phylogenetic nomenclature based on implied three-taxon statements are based on inappropriate comparisons and confuse nomenclature with taxonomy. Taxonomies involving the same named clades imply exactly the same three-taxon statements whether the names of those clades are determined by rank-based or phylogenetic definitions. The approaches thus perform identically in terms of three-taxon informativeness, which is not surprising given that this property is determined by the clades recognized rather than the rules governing their names. By contrast, phylogenetic definitions outperform their rank-based counterparts when the approaches are compared using the nomenclaturally relevant criterion of unnecessary name changes. Phylogenetically defined names result in fewer unnecessary name changes in the context of new phylogenetic hypotheses. Although names having the same endings can end up referring to nested clades under phylogenetic nomenclature, that does not prevent the derivation of three-taxon statements, which requires consultation of a taxonomy or a tree regardless of the system of nomenclature adopted. Finally, Platnick’s criticisms result from considering only unqualified node-based definitions from among the several possible kinds of phylogenetic definitions that have been developed. Other types of phylogenetic definitions can be used to prevent names (such as those with the same rank-signifying endings under rank-based nomenclature) from applying to nested clades. As a tree-based rather than a rank-based approach to nomenclature, phylogenetic nomenclature better preserves the associations between names and clades and has the flexibility to accommodate diverse phylogenetic and nomenclatural considerations.

FUNDING

The Yale Institute for Biospheric Studies awarded a fellowship to K.d.Q. that facilitated interactions with M.J.D. in the preparation of this article. The fellowship was funded by Edward P. Bass.

ACKNOWLEDGMENTS

We thank Jacques Gauthier and an anonymous referee for comments on an earlier version of this article.

REFERENCES