

## Classifying species versus naming clades

Philip D. Cantino

*Department of Environmental and Plant Biology, Ohio University, Athens, Ohio 45701, U.S.A. cantino@ohio.edu*

The February 2004 issue of *Taxon* included two articles and an opinion piece that either (Jørgensen, 2004) opposed development of the *PhyloCode* (Cantino & de Queiroz, 2003) or (Barkley & al., 2004a, b) argued that it was unnecessary to meet the needs of phylogenetic classification. Barkley & al. (2004a: 159) conducted “an article by article vetting” of the *ICBN* and concluded that it “does not contain any rules that actually prevent phylogenetic classification” but that clarification of a few articles is needed. To this end, Moore & al. (2004) proposed adding notes to Articles 3.1, 22.3, and 26.3 of the *ICBN* (Greuter & al., 2000). However, a major problem that motivated development of the *PhyloCode*—the inability of rank-based nomenclature to provide stable, unambiguous names for clades—was not given serious consideration by Barkley & al. (2004a, b), and the *ICBN* rules that cause the greatest problems (e.g., Articles 11.2, 11.3 and the various articles that mandate terminations at particular ranks) were completely ignored.

In my view, the primary goals of modern systematics are to document biological diversity and place this diversity in a phylogenetic context. Stated simply, we are trying to reconstruct the tree of life and provide names for the parts of the tree to facilitate communication and information retrieval. The “parts” that are of interest to phylogenetic systematists are species and clades. It is important to discover, characterize, and name both of these kinds of entities, and the names we provide for them should be stable and unambiguous. I consider nomenclatural clarity to be just as important for clades as it is for species. While other *PhyloCode* supporters may have different reasons for preferring phylogenetic to rank-based nomenclature (the “Linnaean” nomenclature of Barkley & al., 2004a, b), my rejection of the latter is due largely to practical considerations: its inability to provide stable, unambiguous clade names, and the difficulty of naming clades as they are discovered without developing or revising a classification. These aspects of rank-based nomenclature discourage systematists from naming clades and are an impediment to communication about phylogeny (Hibbett & Donoghue, 1998; Cantino, 2000).

**The problems with rank-based nomenclature.** — The instability of clade names governed by rank-based nomenclature has been pointed out repeated-

ly (de Queiroz & Gauthier, 1994; Cantino & al., 1997; de Queiroz, 1997; Kron, 1997; Hibbett & Donoghue, 1998; de Queiroz & Cantino, 2001; Pleijel & Rouse, 2003). To cite just one example, the subclade of Lamiaceae that is named Ajugoideae when ranked as a subfamily must be named Teucrieae if ranked as a tribe due to application of priority within rank (*ICBN* Art. 11.2); the oldest name at the tribal rank based on a type within this clade is Teucrieae Dumort., while the oldest such name at the subfamilial rank is Ajugoideae Kostel. (*Index Nominum Supragenericorum Plantarum Vascularium Project DataBase*; <http://matrix.nal.usda.gov:8080/star/supragenericname.html>). Nothing about the clade has changed other than its arbitrary rank assignment, and the same information is provided by the name regardless of its rank. If comparably unnecessary changes were occurring in species names, the taxonomic community would not stand for it and the nomenclatural system would be changed to eliminate the instability.

Not only may a clade have different names under rank-based nomenclature, but the same name is frequently applied to different clades. This may occur because different systematists choose to apply the same name to different nodes of a cladogram. For example, Urticaceae *sensu* Zomlefer (1994) applied to a more inclusive clade than Urticaceae *sensu* Judd & al. (2002), though there was no apparent disagreement about the phylogeny (Bryant & Cantino, 2002). It may also occur because *ICBN* Arts. 19.4 and 22.1 mandate the names that must be used for subdivisions of a family or genus that include the type. As a result, the name Lamioideae has been applied to various clades that happen to be classified as subfamilies (Cantino & al., 1997). The clades so named may vary greatly in inclusiveness; the only thing they must have in common is subfamilial rank and inclusion of *Lamium*. Phylogenetic nomenclature would prevent the application of the same name to different clades by determining priority relative to the clade rather than the rank.

If names depend on rank assignment, there must be enough ranks to accommodate all the levels of the hierarchy at which one might want to name clades. Barkley & al. (2004b: 157) noted that “a taxonomist is allowed to use as many informal ranks as needed and may circumscribe them as deemed appropriate”. It is not clear what

is meant by “informal”. If this means a rank that is not governed by the *ICBN*, how will application of the name be determined? (Perhaps by a phylogenetic definition, such as is used in the *PhyloCode*?) If by “informal”, Barkley & al. meant the standard subdivisions of the seven principal ranks (i.e., the ranks listed in *ICBN* Art. 4.2), it will become difficult to accommodate all the hierarchical levels of the tree of life that one might want to name as phylogenetic resolution continues to improve (Hibbett & Donoghue, 1998). For example, to name all of the nested clades that one might want to name from Eukarya to (for example) a distinctive pair of angiosperm sister species, would the 18 supraspecific ranks provided in Art. 4.2 be sufficient, or will we eventually have to resort to ranks such as supersubtribe? Will doing so introduce confusion and thus be counter to Art. 4.3?

Not only does rank-based nomenclature fail to provide stable names for clades, it makes it difficult to name clades one at a time as they are discovered without developing a new classification, thereby changing the names of other clades (Hibbett & Donoghue, 1998). This is because the naming of clades and assembling taxa into nested hierarchies (classification) are part of the same process in rank-based nomenclature, whereas these processes are uncoupled in phylogenetic nomenclature (Bryant & Cantino, 2002). In this regard, the separation of nomenclature and classification is more complete under the *PhyloCode* than under the *ICBN* and other rank-based codes.

Interestingly, Barkley & al. (2004b: 155, 157) implied that phylogenetic nomenclature less clearly separates nomenclature from classification because names are applied “by circumscription in the context of a given phylogeny” rather than by the type method. This statement may confuse readers who don’t understand how phylogenetic nomenclature works. Phylogenetic definitions specify a clade, but the content of the clade is not “circumscribed” in the sense that this term is usually used in plant systematics (i.e., direct determination of the content of a taxon). Rather, clade content is determined by the interaction of the definition with a phylogenetic hypothesis (Bryant & Cantino, 2002). With two of the three basic kinds of phylogenetic definitions used under the *PhyloCode* (stem-based and apomorphy-based), only a single species is required to belong to the clade, thus the clade’s content is circumscribed to the same degree as when a type-based name is used. With node-based definitions, a minimum of two species are required to belong to the clade, a slightly greater degree of “circumscription” than with type-based names. The remaining composition of the clade is not determined directly by the definition.

Barkley & al. (2004b: 155) maintained that basing names on rank has benefits that more than compensate for the “inconvenience” of name changes. They claimed

that ranks “play a valuable role in information retrieval by users”. For example, “someone seeking information on the names *Fagus* and *Quercus* can recognize that these labels are names of genera because they are uninomials with none of the standardized higher level endings. The user can then conclude that the circumscriptions of the two taxa are mutually exclusive, and any information obtained on one taxon is not necessarily applicable to the other”. Although the circumscriptions are mutually exclusive, the utility of this from the standpoint of predictiveness is reduced by the continuing existence of paraphyletic plant genera, many of which are large. This situation will be slow to change because revising the boundaries of large genera will require numerous changes in species names. As long as some genera are paraphyletic, one cannot assume mutual exclusivity of *ancestry*, and this will have negative consequences for predictiveness. The rest of the claim (that “any information obtained on one taxon is not necessarily applicable to the other”) is irrelevant; this is true of any two taxa under either system of nomenclature. The most important thing to facilitate information retrieval about clades is stable, unambiguous names. These are better provided by the *PhyloCode*, which avoids the many rank-based problems outlined above.

Surprisingly, one aspect of the *ICBN* that Barkley & al. (2004a) identified as potentially problematic is the fact the principle of priority is not mandatory above the family rank. On the contrary, I view this as beneficial. At least above the family level, one can avoid rank-based name changes provided that the names are not based on the names of genera (so-called “descriptive names”, though some of them such as *Plantae* do not refer to a character). However, there is still the problem of determining to which clade a name should be applied. Some suprafamilial names have been applied to more than one clade. For example, *Plantae* is applied with roughly equal frequency to the clade that is also known as Embryophyta and the larger clade that comprises Embryophyta and Chlorophyta. Furthermore, there is sometimes disagreement whether a name should be given a node-based, stem-based, or apomorphy-based definition, each of which may specify a different clade (identical in their extant members but potentially different in their extinct members). The *ICBN* cannot resolve these inconsistencies in the use of clade names, but the *PhyloCode* does so through the principle of priority—specifically, priority of publication of a phylogenetic definition.

**Getting down to basics.** — The rank-based system, as manifested in the *ICBN*, was not designed to name clades. It was designed to classify organisms into a hierarchy of ranked taxa. Because the species rank is nomenclaturally basic (*ICBN* Art. 2.1; Barkley & al., 2004a), users of the rank-based system, in effect, name species and classify them into supraspecific taxa. With

the progressive acceptance of a more thoroughly phylogenetic perspective has come the widespread (but not universal) expectation that these supraspecific taxa will be monophyletic. However, there is a fundamental difference between (1) naming species and classifying them into higher taxa, even monophyletic ones, and (2) *naming clades*, as entities that are of interest in their own right and just as deserving of stable, unambiguous scientific names as are species. This perspective, which I think is widely shared by proponents of phylogenetic nomenclature, was not addressed by Barkley & al. (2004a, b). Instead, they emphasized the compatibility of “Linnaean” nomenclature, and the *ICBN* in particular, with phylogenetic *classification* rather than its suitability for naming clades.

From this perspective, most of the *ICBN* articles on which Barkley & al. (2004a) focused are marginally relevant. Their failure even to consider the problems caused by *ICBN* Arts. 11.2 and 11.3 (application of priority within rank) is difficult to understand, given that these were the primary problems that motivated development of the *PhyloCode*, but it is easy to understand why they would not want to change these rules. Rank-based priority of names is central to the *ICBN*. It could not be changed without fundamentally altering the system upon which this code is based. A type alone cannot determine the application of a clade name. The type must be combined with a rank or some other information to determine which of many nested clades containing a particular type should bear the corresponding name. Unfortunately, basing names on rank is a major source of instability in clade names, because shifting a clade to a different rank requires that its name be changed. This problem was largely ignored by Barkley & al. (2004a, b), and when mentioned, it was written off as a mere “inconvenience” (2004b: 155). But if one’s goal is to name clades rather than to classify species, the instability introduced by basing priority on rank is not inconsequential.

Jørgensen (2004) expressed disapproval that “the forces behind the *PhyloCode* appear determined not to try to incorporate their needs in the present international Codes”, and that we are not presenting proposals for change in the *ICBN* at the next *International Botanical Congress*. To do so would be pointless. It should be clear from my comments above that the changes that would be needed in the *ICBN* to facilitate the unambiguous naming of clades (elimination of rank-based priority, rank-based terminations, and related rules) are so fundamental that a code in which they were implemented would bear little resemblance to the current *ICBN*. Such proposals would have no chance of being accepted, thus making them a waste of time for everyone involved. If a few simple changes in the *ICBN* would eliminate the problems discussed above, I or others would have proposed them a decade ago. Developing the *PhyloCode* has been

a difficult, tedious task, which we would not have undertaken if there had been easier alternatives.

#### **Co-existence of two nomenclatural systems.**

— Barkley & al. (2004b:154) discussed the concept of a general-purpose classification, including that it “must be “practical”, and should work under a variety of circumstances”. They stated (p. 156) that “Linnaean nomenclature has a utility that extends beyond the reflection of evolution”, and that a system that abandons ranks and insists on monophyly is a special-purpose classification. In view of the fundamental philosophical differences within the systematics community, I don’t think that either system qualifies as general-purpose at this time. Rank-based nomenclature meets the needs of those whose priority is preserving information about ranks, even though it entails changes in the names of clades and species. Phylogenetic nomenclature meets the needs of those whose priority is maximizing the stability and clarity of names for the historical entities resulting from evolution (clades and species), even though the names do not preserve information about rank.

With each system failing to meet the needs of a sizable group of users, the coexistence of two systems may be the only viable solution. (The impression one gets from Barkley & al. (2004a, b) that only a few people are dissatisfied with rank-based nomenclature is heavily influenced by the composition of the group that participated in the workshop on which the papers are based.) Barkley & al. (2004b: 158) are willing to accept the *PhyloCode* as a special-purpose system, but Jørgensen (2004) expressed the concern that having parallel systems will lead to confusion. Fortunately, this concern is overblown. The vast majority of clade names would refer to the same set of species under both systems. When this is not the case, confusion can easily be avoided through the use of a symbol or other convention to indicate which system governs a name (e.g., Baum & al., 1998), as recommended in the *PhyloCode* (Rec. 6.1B). In contrast, Jørgensen’s (2004) proposal that the *PhyloCode* “invent a quite dissimilar set of names for their clades” would complicate information retrieval and require people to learn two different names for each clade.

Barkley & al. (2004b: 157) assert that adopting phylogenetic nomenclature could be extremely costly, saying “the actual monetary costs of changing the nomenclatural system, for example in the U.S.A. Endangered Species Act, could range in the millions to billions of dollars”. In the absence of an explanation of how they derived this fear-inspiring and highly questionable figure, it cannot be taken seriously. Furthermore, there are also costs to maintaining the current system, such as the time (and hence money) required for systematists to revise classifications and change the names of previously named clades and species simply to be able to name newly discovered clades (Hibbett & Donoghue, 1998).

**Abandonment of the “PEST” convention.** —

Although Barkley & al. (2004a, b) ignored the largest problem that the rank-based system poses for the naming of clades, they highlighted a way in which the application of the *ICBN* to phylogenetic classification could be improved without changing any rules. As they correctly pointed out, the only mandatory rank is genus. However, many other ranks are treated as mandatory by most systematists. This is true of secondary as well as primary ranks due to the convention known as the Principle of Exhaustive Subsidiary Taxa (de Queiroz & Gauthier, 1992). This convention (which I like to refer to by its acronym, PEST) dictates that if a nonmandatory rank is used at all, every species must belong to some taxon at that rank. For example, if a family is divided into subfamilies, it is expected that every species in that family be assigned to a subfamily.

The PEST convention is problematical for those who would like to avoid naming paraphyletic groups. (Although I am not one of those who would advocate public flogging for anyone recognizing a paraphyletic group, I have no interest in naming them myself.) For example, if there is only one well supported clade within a family, adherence to the PEST convention requires that one either name the other (poorly supported) clades as subfamilies or accommodate the genera composing these poorly supported clades in single subfamily, which in many cases will be paraphyletic. Happily, this convention is gradually being abandoned, and I am pleased to see Barkley & al. (2004a) and Moore & al. (2004) lending support to its abandonment.

Abandonment of the PEST convention will make it possible to avoid naming paraphyletic groups except at the genus level, where the kind of taxonomic dilemma described by Cantino & al. (1999) will continue to exist. However, it will not eliminate the more serious problems that the rank-based system presents to those who would like to give stable, unambiguous names to clades as they are discovered, without developing or revising a classification. By eliminating these problems, the *PhyloCode* will facilitate the naming of clades, storage and retrieval of information pertaining to these clades, and communication about phylogeny.

**ACKNOWLEDGEMENTS**

I am grateful to Kevin de Queiroz and Michael Donoghue for their helpful comments on an earlier draft of this paper.

**LITERATURE CITED**

- Barkley, T. M., DePriest, P., Funk, V., Kiger, R. W., Kress, W. J., McNeill, J., Moore, J., Nicolson, D. H., Stevenson, D. W. & Wheeler, Q. D. 2004a. A review of the *International Code of Botanical Nomenclature* with respect to its compatibility with phylogenetic classification. *Taxon* 53: 159–161.
- Barkley, T. M., DePriest, P., Funk, V., Kiger, R. W., Kress, W. J. & Moore, G. 2004b. Linnaean nomenclature in the 21<sup>st</sup> Century: a report from a workshop on integrating traditional nomenclature and phylogenetic classification. *Taxon* 53: 153–158.
- Baum, D. A., Alverson, W. S. & Nyffeler, R. 1998. A durian by any other name: taxonomy and nomenclature of the core Malvales. *Harvard Pap. Bot.* 3: 315–330.
- Bryant, H. N. & Cantino, P. D. 2002. A review of criticisms of phylogenetic nomenclature: is taxonomic freedom the fundamental issue? *Biol. Rev.* 77: 39–55.
- Cantino, P. D. 2000. Phylogenetic nomenclature: addressing some concerns. *Taxon* 49: 85–93.
- Cantino, P. D. & de Queiroz, K. 2003. *PhyloCode: A Phylogenetic Code of Biological Nomenclature*. Version 2. [<http://www.ohiou.edu/phylocode/>]
- Cantino, P. D., Olmstead, R. G. & Wagstaff, S. J. 1997. A comparison of phylogenetic nomenclature with the current system: a botanical case study. *Syst. Biol.* 46: 313–331.
- Cantino, P. D., Wagstaff, S. J. & Olmstead, R. G. 1999. *Caryopteris* (Lamiaceae) and the conflict between phylogenetic and pragmatic considerations in botanical nomenclature. *Syst. Bot.* 23: 369–386.
- de Queiroz, K. 1997. The Linnaean hierarchy and the evolutionization of taxonomy, with emphasis on the problem of nomenclature. *Aliso* 15: 125–144.
- de Queiroz, K. & Cantino, P. D. 2001. Phylogenetic nomenclature and the *PhyloCode*. *Bull. Zool. Nom.* 58: 254–271.
- de Queiroz, K. & Gauthier, J. 1992. Phylogenetic taxonomy. *Ann. 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.
- Greuter, W., McNeill, J., Barrie, F. R., Burdet, H. M., Demoulin, V., Filgueiras, T. S., Nicolson, D. H., Silva, P. C., Skog, J. E., Trehane, P., Turland, N. J. & Hawksworth, D. L. (eds.). 2000. *International Code of Botanical Nomenclature (Saint Louis Code) adopted by the Sixteenth International Botanical Congress St. Louis, Missouri, July–August 1999*. Koeltz Scientific Books, Königstein. [Regnum Veg. 138.]
- Hibbett, D. S. & Donoghue, M. J. 1998. Integrating phylogenetic analysis and classification in fungi. *Mycologia* 90: 347–356.
- Jørgensen, P. M. 2004. Rankless names in the *Code*? *Taxon* 53: 162.
- Judd, W. S., Campbell, C. S., Kellogg, E. A. & Stevens, P. F. 2002. *Plant Systematics: a Phylogenetic Approach*, ed. 2. Sinauer Associates, Sunderland, Massachusetts.
- Kron, K. A. 1997. Exploring alternative systems of classification. *Aliso* 15: 105–112.
- Moore, G., Barkley, T. M., DePriest, P., Funk, V., Kiger, R. W. & Kress, W. J. 2004. (065–067) Proposals to amend Article 3.1, Article 22.3, and Article 26.3. *Taxon* 53: 214.
- Pleijel, F. & Rouse, G. W. 2003. Ceci n'est pas une pipe: names, clades and phylogenetic nomenclature. *J. Zool. Syst. Evol. Research* 41: 162–174.
- Zomlefer, W. B. 1994. *Guide to Flowering Plant Families*. Univ. North Carolina Press, Chapel Hill.