WILEY-BLACKWELL

Cladistics 29 (2013) 1–6

Cladistics

10.1111/j.1096-0031.2012.00430.x

Letter to the Editor

More on absences

Accepted 24 July 2012

Sir,

"Myths and legends die hard in America." (Hunter S. Thompson, *The Great Shark Hunt*, 1992)

In the letter by Platnick (2012), the author seeks to illustrate how symplesiomorphies are not homologous character states with examples from spiders. None of the arguments presented has any relevance to the issue of homology, its definition in common usage by biologists, or the way we do cladistic analysis, using parsimony or any other mainstream method. Ironically, Platnick merely reinforces the final point of Nixon and Carpenter (2012c). No-one would score a desk as lacking spinnerets—because there is no reasonable expectation that the desk and say, wasps, lack spinnerets because their common ancestor also lacked spinnerets. Simply put, one does not score a desk as lacking (or having) spinnerets, because that would not be a reasonable hypothesis of homology—we must assume evolutionary relationships among terminals *before* we can create a matrix. On the other hand, if we are doing one of the amply funded "tree-of-life" analyses, and we wish to include presence of spinnerets as a character, then we must also score the other taxa in our analysis as lacking spinnerets, be they arachnids or plants (or we may wish to score some of them as inapplicable, in which case we are not making a comparison). Once we have constructed such a matrix, we then must either explain a pair of taxa *that are scored as lacking spinnerets* as lacking them due to common ancestry; or by the alternatives, that their common ancestor had spinnerets and they both lost them, or one or both had ancestors with spinnerets younger than their common ancestor. This explanation follows from the results of an analysis, and does not presuppose the analysis, as suggested by Platnick. Things not being compared—not in the matrix—do not have to be explained. It seems that Platnick does not think in terms of constructing matrices, a task that requires selection of taxa and characters on the basis of homology, and explanation of those scored character state distributions via repeatable analyses and characters optimized on cladograms. Instead, he thinks in terms of "what-if" scenarios with unjustified premises (why would someone include a desk?) and unsupported claims, reminiscent of the laboured argumentation found in previous works (e.g. Nelson and Platnick, 1981). In contrast, those who actually construct matrices select taxa on the basis of their relevance, and characters on the basis of whether they are "good" potential characters—i.e. they can be hypothesized to be the same state due to common ancestry, not on the basis of whether they can be asserted to be synapomorphies to support a preconceived phylogeny. Characters without genetic basis, such as plastic environmental variation, or features of desks, are rejected. If we then analyse these characters with parsimony, the analysis minimizes the number of steps—errors in our homology assessment (i.e. hypotheses of homology, or shared conditions due to ancestry). We do not have to explain why a desk has no spinnerets unless we include the desk in the analysis and score it as lacking them. And if we do so, we must defend that decision on the basis of whether that is a good hypothesis of homology. Of course, these concepts apply even outside the context of matrices. In order to declare spinnerets as a synapomorphy, Platnick must compare the spinnerets of spiders with something that lacks them. After asserting the synapomorphous nature of spinnerets, whenever he compares two or more of these things that lack them, he is asserting (not testing) that the absence is homologous—i.e. they are not nested phylogenetically within the clade of spiders bearing spinnerets. If he wishes to compare spinnerets to absurdities, so be it, but if he compares to nothing outside, the character state explains nothing. What he wishes to explain is a matter of context. Platnick's spinneret example, instead of calling into question the connection between homology, evolution, and parsimony, shows how lack of such grounding creates absurd and unnecessary conundrums. The resolution of Platnick's contrived dilemma is to evaluate scored states and reject character state scores (and terminals bearing them) that cannot be considered hypotheses of homology. Platnick has merely provided a clear counterpoint to modern cladistic analysis, which instead uses all character states (be they zeros or ones), and his

desk example illustrates the pitfalls of the ''pattern'' view that characters are not hypotheses of evolutionary homology.

We could stop here. But it is necessary to comment further on additional examples and the underlying confusion presented by Platnick, in order to illustrate the way in which cladistic theory and practice have changed since Hennig published his landmark work in English (Hennig, 1966). Just as evolution as a field of study did not stagnate after Darwin (1859), phylogenetics did not stagnate after Hennig, and the entire area we now call cladistics is largely based on post-Hennigian concepts and algorithms.

More on states

Clearly, Platnick is confused, as apparently many workers have been, about the distinction between a scored character state (a 0 or 1; or perhaps an A, C, G or T) and the concrete structure or lack of a structure that the encoding represents. As we stated in our discussions (Nixon and Carpenter, 2012a,b,c), characters, and character states, are constructs of the systematist undertaking the analysis, and as such have meaning only in that context. Characters must be defined before we can compare their distribution among taxa, and these definitions determine the field of comparison. As pointed out by Farris (2012), and apparently necessary to repeat here, Hennig accepted absence of a feature as homology in the context of comparison with taxa possessing the feature:

> ''In general we speak only of the homology of organs, but a character may also be the absence of an organ ... The absence of wings in the Anoplura [sucking lice] and Mallophaga [biting lice] is a synapomorphous character.'' (Hennig, 1966, p. 95)

Some commonly used and encoded characters, such as DNA sequence data, may have features that are best considered (or perhaps inescapably considered) as presence–absence data. The most obvious of these are referred to as ''indels''—which may be due either to an evolutionary insertion or an evolutionary deletion of a fragment of DNA sequence into/from a longer sequence. If one understands the nature of indels, it is easy to see (and resolve) the issue of presence–absence data as indicative of a particular event, whereby each state (presence or absence of a sequence fragment) indicates the occurrence or non-occurrence of that event. As in all other characters, without divine revelation it is impossible to tell the direction of change with certainty prior to a cladistic analysis. But we may still score these two states—presence and absence—prior to our analysis while constructing a matrix of characters that might be analysed with parsimony (or some other preferred method). Because parsimony minimizes extra steps (homoplasy, or independent occurrences) over both the presence and absence states equally, it does not distinguish between presence and absence. In each case, our character-state score is thus based on the concept that we wish to minimize the number of times (over all characters and states) that we were ''wrong'' in making such scores—i.e. that the state (either 0 or 1) occurs on the tree more than once independently. If a state occurs only once on the tree, then tracing the optimized character state from any two taxa that bear it in the original matrix will lead to a common ancestor that also bears the state. Not surprisingly, this corollary of non-homoplasy in a character state is homology—the character state in two taxa is homologous when it is due to descent from a common ancestor. This is true for both symplesiomorphy and synapomorphy, independent of whether the states are defined as presence–absence alternatives, or if they are defined as ''positive'' alternatives, such as A–C–G–T, or short–intermediate–long tails. If the author of a matrix defines a character on the basis of presence–absence, we may differ in our confidence in the definition of that character, and whether that particular encoding was a wise encoding, and whether the method deals well with presence–absence data—none of which negates that fact that if a state (as defined in a matrix) is shared in two or more taxa on a tree due to descent from a common ancestor, that *state* is homologous in those two taxa. Thankfully, this simple attribute of parsimony—the fact that it minimizes errors in our original homology assessment is what connects parsimony to evolution, allows us to interpret cladograms as phylogenetic diagrams, and avoids the pitfalls of wrongly assuming that only synapomorphies are homologous. It also prevents us from coding desks as lacking spinnerets.

Perhaps the difficulty that some workers have with these simple concepts is metaphysical. Certainly, it is difficult to conceptualize shared absence as a ''feature'' if one is just looking at two organisms lying dead on the dissecting table. Obviously, without any context, myriad shared absences might be imagined, and these are not concrete features in the corporeal world. However, homology in the context of phylogenetic inference of necessity requires that states be defined and grouped into characters in order to make comparisons, as clearly outlined by Hennig (1966). This is what scientists do in order to construct matrices—conceptualize features that are observable (in the scientific, not ocular, sense). The coding and the feature are not the same thing—one merely represents our interpretation of the information (observation) that we can extract from the other (a condition). Our coding may be good or bad, easy or hard, elegant or complex, independently of the nature and complexity of the feature. Eliminating or reducing such interpretive issues is one of the appeals of molecular sequence data. However, even with such data we may be presented with situations in which we cannot

easily define an alternative state—which throws us into the realm of presence–absence data. As a consequence, with the act of defining a presence–absence character, and entering scores for such a character into a matrix, such states then must be explained in the context of homology and homoplasy. When we use such characters, absence is treated the same as presence in a parsimony analysis. The most parsimonous trees are those that minimize extra steps across all states. In all cases, whether presence or absence, if a state has no homoplasy then the best explanation is that all bearers of the state (in this context—this matrix—this analysis) have that state (as defined by the author of the matrix) due to common ancestry, and thus that state, *as defined*, is homologous in those taxa.

More on recent history

A corollary premise of our apparently controversial paper on homology (Nixon and Carpenter, 2012a) is the dividing line between Hennigian and Farrisian phylogenetics. Although Hennig clearly held an evolutionary concept of homology (as illustrated in Nixon and Carpenter, 2012a,b,c), he did not provide an explicit, repeatable, or consistent method for constructing phylogenetic trees. Just as Darwin had enormous influence in providing a new direction for biology, Hennig was a pioneer in developing concepts that provided the foundation for modern phylogenetics. However, to project ideas onto Hennig that he never held is no more productive than to do so for Darwin. Hennig developed the principles of phylogenetics, and later workers, most notably and influentially James S. Farris, developed the modern methodological approaches that are still commonly used—modern phylogenetics, or just cladistics if one prefers.

To illustrate the difference between these two approaches, it is necessary first to contrast tree-construction methods. Hennig proposed that phylogenetic trees could be produced by inspection of homologies, selection of synapomorphies from those homologies, and using those synapomorphies to group taxa together into monophyletic groups, thereby constructing rooted phylogenetic trees. Hennig was not explicit on numerous aspects of this endeavour: for example, issues such as how to deal with conflicting synapomorphies were not adequately addressed. An explicit method that required a priori polarization of characters was proposed by Wagner (1961), and eventually a computer algorithm was implemented by Farris (see discussion in Farris, 2012). By the time the ''Wagner method'' was formally presented it no longer required a priori polarization of characters (Farris, 1970). However, soon Farris (and others) realized that the trees found were not necessarily the most parsimonious, and parsimony, as an optimality criterion, became the goal of subsequent methods, most prominently branch-swapping (e.g. Mickevich, 1978). Although we now find ourselves in a world replete with other optimality criteria, such as maximum likelihood, the construction algorithms that are implemented in these methods are basically those developed for parsimony—branch-swapping, and various more devious ways of doing branch-swapping more efficiently or more effectively (e.g. Goloboff, 1999; Nixon, 1999). Thus the basis of all modern tree-building has its roots in the papers, algorithms, and computer programs developed by Farris (and a few others) from 1969 onward (beginning with Kluge and Farris, 1969).

How does all this relate to homology? Hopefully, at this point it should be clear that these modern tree-construction methods work in an unrooted tree space (at least for parsimony) and are not ''grouping by synapomorphy'' in the sense of Platnick (2012), that is by a priori synapomorphy. Characters are not polarized before analysis in Farrisian cladistics—although the user may designate a root or outgroup prior to the analysis, this has no bearing on the unrooted topology, or the length of the tree. Thus it can be stated clearly that the methods of constructing most-parsimonious trees do NOT ''group by (a priori) synapomorphy''—and for any given data set, there may be more or fewer ''synapomorphies'' depending on the position of the root.

If you maximize synapomorphy only, you accordingly must allow more steps in plesiomorphic character states. This deviation from parsimony is a goal of three-taxon analysis (3ta), and when selectively ''grouping by synapomorphy'' it might be considered collateral damage. In contrast, if you minimize steps across the entire tree (indifferently as to whether a character state is apomorphic or plesiomorphic), you are optimizing on the basis of homology. This is what we have been doing since the 1970s in parsimony analysis.

More on spiders

Platnick also provided an example of how one might recode characters to avoid ''unordered multistate 'pseudocharacters'', but what his example actually illustrates is that with creative coding one can craft whatever outcome one wishes. He referred to a matrix for oonopid spiders from Platnick et al. (2012), reproduced here as Table 1. Two variables, numbers 2 and 6, were used to score the tarsal organ receptor pattern. This is because, Platnick stated, ''we wanted to emphasize our treatment of the evidence''. He explained ''Some reviewers of that manuscript suggested that the presence of a 4-4-3-3 receptor pattern could be construed as evidence uniting *Orchestina* and *Kapitia*. In other words, those reviewers suggested that in a matrix, *Orchestina*

Table 1
Data matrix from Platnick et al. (2012)

Dysdera	0000000
Orsolobus	0000000
Orchestina	1110000
Cortestina	1101000
Sulsula	1101000
Dalmasula	1101000
Xiombarg	1101000
Unicorn	1101000
Kapitia	1100100
Oonops	1100111
Gamasomorpha	1100111

Character list
1. Tarsal organ with proximal longitudinal ridge
2. Tarsal organ with raised receptors only, in serially dimorphic pattern (either 4-4-3-3 or a modified, reduced form of that pattern, i.e. 3-3-2-2 or 2-2-1-1)
3. Femur IV enlarged
4. Tarsal organ at least partly capsulate
5. Male palp without heavily sclerotized sperm duct
6. Tarsal organ receptor pattern reduced to 3-3-2-2
7. Ocular group clumped

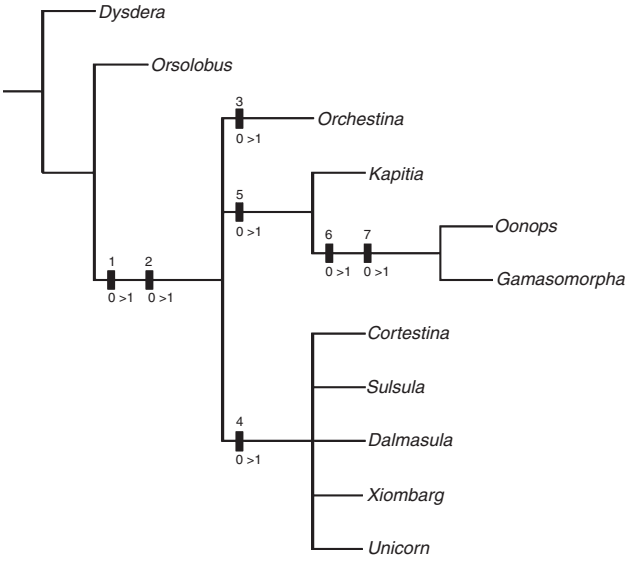

Fig. 1. First spider tree. The length is seven steps for the data matrix of Table 1, and nine steps for the data matrix of Table 2 with character 2 treated additively. Characters are plotted for the data of Table 1; character numbers are placed above hash marks, with the state numbers below; state changes are denoted by ">."

and *Kapitia* should be coded as sharing a state that is not found in any other taxa." It is therefore interesting to note that this is precisely what is not done in the matrix—a state 4-4-3-3 is not scored. Instead, variable 2 has a state defined as "with raised receptors only, in serially dimorphic pattern (either 4-4-3-3 or a modified, reduced form of that pattern, i.e. 3-3-2-2 or 2-2-1-1)", which lumps three conditions, while variable 6 has a state defined as "receptor pattern reduced to 3-3-2-2", which is one of the conditions lumped in variable 2. Neither 4-4-3-3 nor 2-2-1-1 is scored as such. Analysis of that data matrix with the program TNT (Goloboff et al., 2008), using implicit enumeration, does yield a single cladogram, here as Fig. 1, which does correspond to the classification of Platnick et al. (2012).

Platnick also asserted "We could, of course, have achieved the same result, computationally, by using one ordered, multistate character instead", but evidently did not bother to check. In Table 2, we have replaced the two variables of Table 1 with a single variable, number 2, which is now multistate. The states are "without same type of tarsal organ" (state 0), "4-4-3-3" (state 1), "3-3-2-2" (state 2) and "2-2-1-1" (state 3). The scores for that variable are based on: Platnick's statement that the outgroups do not have the same type of tarsal organ (state 0), that *Orchestina* and *Kapitia* have the condition 4-4-3-3 (state 1), variable 6 of Table 1 scoring 3-3-2-2 for *Oonops* and *Gamasomorpha* "etc." (state 2), and the inference that the remaining ingroup taxa therefore have the condition 2-2-1-1 (state 3). Treating this variable as additive, analysis with implicit enumeration of the data in Table 2 yields the tree of Fig. 1—but it also yields an additional tree, Fig. 2. *Orchestina* and *Kapitia* are now unresolved at the base of oonopids. Actually scoring their state 4-4-3-3 does have an effect. Amusingly, if the multistate variable is treated as nonadditive, just the tree of Fig. 1 results from analysis. This is Platnick's desired tree, but he dismissed nonadditive variables, stating "To me, such coding allows potentially erroneous groupings." Perhaps he should reconsider.

Table 2
Modified data matrix, with variable two now multistate

Dysdera	000000
Orsolobus	000000
Orchestina	111000
Cortestina	130100
Sulsula	130100
Dalmasula	130100
Xiombarg	130100
Unicorn	130100
Kapitia	110010
Oonops	120011
Gamasomorpha	120011

Conclusion

Whether or not Platnick wishes to admit it, when he declares that spinnerets are a synapomorphy of spiders, he is also declaring that giraffes and hippos lack spinnerets because their common ancestor also lacked them—not because an ancestor had spinnerets and they have been lost. This cannot be claimed as Hennigian vs

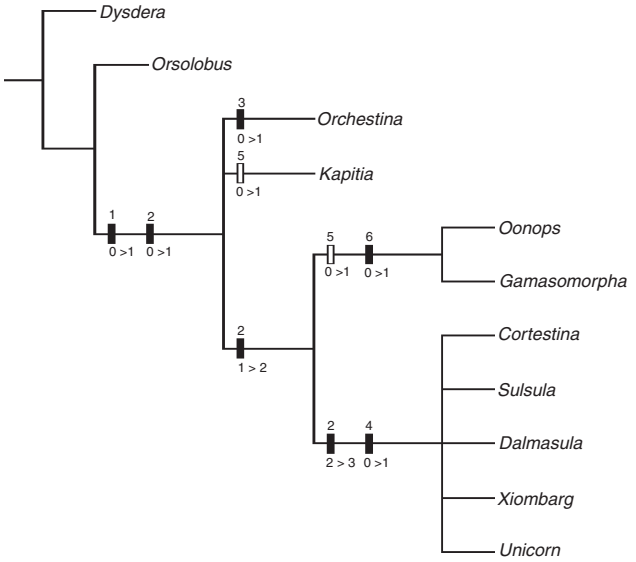

Fig. 2. Second spider tree. The length is nine steps for the data matrix of Table 2 with character 2 treated additively. Characters are plotted for the data of Table 2; character numbers are placed above hash marks, with the state numbers below; state changes are denoted by "> ." A filled hash mark indicates an uncontroverted change; an open one indicates homoplasy.

Farrisian, as Hennig sided with Farris in the matter. Instead, contrary to Platnick's lamentations, this absence phobia seems to be part of the "pattern" school of thought, and can be traced to the first and most influential pattern/3ta manifesto (Nelson and Platnick, 1981).

It is of no surprise to us that three of the four most prominent proponents of 3ta have now come forward to complain about our explanation of homology in the context of parsimony (see also Williams and Ebach, 2012). As these authors are closely tied to the broader community of "pattern" cladistics, it is also no surprise that the other most vocal complaint has come from that direction (Brower and de Pinna, 2012). To the extent that the complaints provide the opportunity to clarify the way in which modern cladistic analysis is undertaken, they serve a useful purpose.

The widespread misconception that modern parsimony methods construct trees using a priori synapomorphy as the actual grouping criterion needs to be abandoned. Parsimony analysis (when successful) produces optimal trees with minimal homoplasy. Homoplasy by definition is non-homology, i.e. shared states that (as defined) are not due to common ancestry. Homoplasy is more precisely an attribute of states, not characters, and although we report it as a number reflecting extra steps in characters, we might also report it as extra origins of states. One state may be homoplastic, while the alternative has no homoplasy (it originates at a single point on the unrooted network and rooted tree). As a corollary, by minimizing homoplasy, parsimony maximizes both types of homology—symplesiomorphy and synapomorphy. The optimization of homology and homoplasy on the same topology remains the same independent of the position of the root. To state this more succinctly, parsimony maximizes our original proposals of homology, which we present as unpolarized character states. Where we place the root is another matter (Nixon and Carpenter, 1993).

Acknowledgements

We thank Steve Farris for suggestions, and Norman for providing illustrative examples.

References

Brower, A.V.Z., de Pinna, M.C.C., 2012. Homology and errors. Cladistics 28, 529–538.
Darwin, C., 1859. On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. John Murray, London.
Farris, J.S., 1970. Methods for computing Wagner Trees. Syst. Zool. 19, 83–92.
Farris, J.S., 2012. Early Wagner trees and "the cladistic redux". Cladistics 28, 545–547.
Goloboff, P.A., 1999. Analyzing large data sets in reasonable times: solutions for composite optima. Cladistics 15, 415–428.
Goloboff, P., Farris, J.S., Nixon, K.C., 2008. TNT, a free program for phylogenetic analysis. Cladistics 24, 774–786.
Hennig, W., 1966. Phylogenetic Systematics. University of Illinois Press, Urbana.
Kluge, A.G., Farris, J.S., 1969. Quantitative phyletics and the evolution of anurans. Syst. Zool. 18, 1–32.
Mickevich, M.F., 1978. Taxonomic congruence. Syst. Zool. 27, 143–158.
Nelson, G.J., Platnick, N.I., 1981. Systematics and Biogeography: Cladistics and Vicariance. Columbia University Press, New York.
Nixon, K.C., 1999. The parsimony ratchet, a new method for rapid parsimony analysis. Cladistics 15, 407–414.
Nixon, K.C., Carpenter, J.M., 1993. On outgroups. Cladistics 9, 413–426.
Nixon, K.C., Carpenter, J.M., 2012a. On homology. Cladistics 28, 160–169.
Nixon, K.C., Carpenter, J.M., 2012b. More on homology. Cladistics 28, 225–226.
Nixon, K.C., Carpenter, J.M., 2012c. More on errors. Cladistics 28, 539–544.
Platnick, N.I., 2012. Less on homology. Cladistics, in press, doi: 10.1111/j.1096-0031.2012.00432.x.
Platnick, N.I., Abrahim, N., Álvarez-Padilla, F., Andriamalala, D., Baehr, B.C., Baert, L., Bonaldo, A.B., Brescovit, A.D., Chousou-Polydouri, N., Dupérré, N., Eichenberger, B., Fannes, W., Gaublomme, E., Gillespie, R.G., Grismado, C.J., Griswold, C.E., Harvey, M.S., Henrard, A., Hormiga, G., Izquierdo, M.A., Jocqué, R., Kranzbaltensperger, Y., Kropf, C., Ott, R., Ramírez, M.J., Raven, R.J., Rheims, C.A., Ruiz, G.R.S., Santos, A.J., Saucedo,

A., Sierwald, P., Szüts, T., Ubick, D., Wang, X.-P. 2012. Tarsal organ morphology and the phylogeny of goblin spiders (Araneae, Oonopidae), with notes on basal genera. Am. Mus. Novitat. 3736, 1–52.
Thompson, H.S., 1992. The Great Shark Hunt (Gonzo Papers Vol. 1). Ballantine Books, New York.
Wagner, W.H. Jr, 1961. Problems in the classification of ferns. Recent Adv. Bot. 1, 841–844.
Williams, D.M., Ebach, M.C., 2012. Confusing homologs as homologies: a reply to ''On homology''. Cladistics 28, 223–224.

Kevin C. Nixon[a] and James M. Carpenter[b]
[a]*Bailey Hortorium, Department of Plant Biology, Cornell University, Ithaca, NY 14853, USA;*
[b]*Division of Invertebrate Zoology, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024, USA*
E-mail address: kcn2@cornell.edu