Molecular and morphological datasets

Our recent paper (Scotland & al., 2003) on the role of morphology in phylogeny reconstruction struck a sensitive nerve in the systematics community. In that paper, we posited that difficulties in character conceptualization and coding of morphological characters rendered the use of morphology inadequate for accurate or complete reconstruction of phylogenetic trees, with the caveat that for certain groups, for which morphology is all that is available (e.g., many fossil groups), morphology is still the best available evidence of phylogeny. Despite the fact that many of the monophyletic groups we recognize were identified as such by their morphological homologies, we argued that molecular data today offer more and better data with which to reconstruct the Tree of Life in all of its branches, not just the easily recognized groups. Furthermore, the problems with morphological characters lead to the conclusion that the effort to reconstruct phylogenies should lead to an increased effort to obtain sufficient molecular data to obtain resolution, rather than to efforts to combine morphological datasets with molecular ones, which may lead to reduced accuracy or resolution. We are sensitive to the criticism that our statement of this position has labeled us “anti-morphology” by some systematists, though the label couldn’t be further from the truth. We urged that the paradigm be reframed as one in which the central goal of morphological studies is to understand the origin and evolution of morphology in a phylogenetic framework and, where possible, in the context of developmental genetics. That the practice of phylogenetic systematics bears out our arguments, one needs look no farther than the program of any recent systematics conference or the contents of systematics journals.

The recent paper by Lee (2004) represents one of several responses to our initial paper (see also Jenner, 2004; Wiens, 2004). In it, he made two errors that we think most of our critics have made. (1) Citing an example will resolve the debate. In our paper we used one example in order to make a point, but we realized that the literature was replete with published examples that could be used to argue either way. Consequently, we chose to emphasize the conceptual problems with morphological characters and the difficulties they pose for phylogeny reconstruction. These problems have been largely ignored in the responses to our paper. (2) The use of 20th century DNA datasets (in the case of Lee’s paper, the example used was published in 2001, but long after the data were gathered, and included rbcL and 18S sequences, which are not used much anymore, because they don’t have much information at this level) as if such data constitute the sum total of available DNA data. As we enter an era of genome-scale molecular data gathering, it is inevitable that molecular datasets in the near future will far outstrip those used as recently as the late 1990s, whereas morphological datasets are unlikely to expand in size without encountering the problems with character conceptualization and coding that we identified in our paper. The problem of citing past papers as examples ignores the potential for future molecular studies. The remainder of this letter will respond more explicitly to Lee’s paper.

In the first paragraph Lee agrees with our assertion that “subjectivities in morphological character coding” are “well founded”, but then goes on to talk about morphological characters and DNA characters as if they are both equally reliable. A point we made was that, if you want to resolve difficult parts of a tree, you are going to be able to do so better by going back to the font of DNA data as often as you need (true enough as a rough approximation), whereas atomizing morphology in an attempt to identify more characters is likely to lead to the inclusion of more bad (e.g., non-independent or poorly coded) characters.

We also disagree with branch length as an index of number of informative characters, since a small number of really bad characters is likely to result in longer branch lengths and less accuracy than an equal number of good characters. Unless, of course, they are ‘bad’ due to non-independence, in which case they may yield increased resolution, but not necessarily accuracy. Neither accuracy nor resolution entered into Lee’s argument, since he simply took one of an indefinite number of most parsimonious trees and calculated branch lengths on it, without regard for whether it was accurately depicting phylogeny or not, or whether the data found a single optimal tree or many possible trees. Even with the single-tree statistics that he presented and the relatively old molecular data, the combined DNA data have ca. 4× the branch length and 2× the Bremer support that morphology provides. Comparing morphology to each individual partition of the DNA dataset, as he does, misses the precise point we wanted to make about DNA having the capability to yield more characters than morphology.

Lastly, we made the case that DNA was better suited, at least in part, because it could address all levels in a tree, while he chose to address just one area in the hier-
archy (the family rank in this case), which, in this partic-
ular example, is poorly resolved. He also framed the
debate entirely in a parsimony context, which is neces-
sary for morphological characters, but which is a method
of analysis now considered crude (though perfectly suf-
ficient in many studies) for getting optimal inference
from DNA data.

We maintain that the era of systematics in which sci-
centists conducted broad surveys of morphological traits
simply for the purpose of inferring phylogeny is in the
past, and that the future of the study of morphology lies
in the interpretation of morphology in light of DNA-
based phylogenies and the developmental context of the
traits of interest. In this context it will often be the case
that the traits most surprisingly seen NOT to be evidence
of homology will be the ones meriting the greater inter-
est by morphologists and evolutionary biologists, pre-
cisely because these will be traits that may be more read-
ily linked to the environment or selection.

Richard G. Olmstead
Department of Biology, University of Washington, Seattle,
Washington 98195, U.S.A. olmstead@u.washington.edu

Robert W. Scotland
Department of Plant Sciences, University of Oxford, South Parks
Road, Oxford OX33RB, U.K. robert.scotland@plant-sciences.
ox.ac.uk

Literature cited
Phylogeny reconstruction: the role of morphology. Syst.
Biol. 52: 539–548.

Jenner, R. A. 2004. Accepting partnership by submission?
Morphological phylogenetics in a molecular millennium.

Lee, M. S. Y. 2004. Molecular and morphological datasets
have similar numbers of relevant phylogenetic characters.
Taxon 53: 1019–1022.

Wiens, J. J. 2004. The role of morphological data in phyloge-