Evolutionary inferences from non-monophyly on molecular trees

Richard H. Zander

Missouri Botanical Garden, P.O. Box 299, St. Louis, Missouri 63166-0299 U.S.A. richard.zander@mobot.org

Traditional taxa that are non-monophyletic (paraphyletic or polyphyletic) on a molecular phylogenetic tree may be interpreted as surviving ancestors that are evolutionarily static in expressed traits but accumulating DNA changes used to track genetic continuity. In cases in which re-evolution (convergence) of such taxa is deemed improbable, non-monophyly may be used to infer evolutionary trees of virtual fossils reflecting macroevolution. A taxon tree or Besseyan cactus of the family Pottiaceae (Bryophyta) is presented as an example. Non-monophyly may be evolutionarily informative at the taxon level in the same way synapomorphies are phylogenetically informative at the molecular trait level.

KEYWORDS: evolution, macroevolution, molecular systematics, morphology, Pottiaceae, taxon tree, virtual fossil

INTRODUCTION

It has been pointed out (e.g., Farjon, 2007) that “taxa must evolve from other taxa” yet phylogenetic analysis of both morphological and molecular data produces trees of character state changes, not changes of taxa. Inference of the evolution of taxa is at best indirect, commonly mediated by rules and assumptions involving, e.g., phylogenetic definitions of taxa involving monophyly, and mapping of morphological traits on molecular trees. At present, traditional taxonomists are alarmed by the burgeoning primacy of studies involving largely non-selective molecular traits in systematics although fields of evolution and ecology continue to emphasize phenotypic traits or their genotypic basis, emphasizing selection and other environmental interactions. Molecular phylogenetics have failed to provide a direct guide to evolutionary relationships as determined through selection and drift acting on expressed traits (Hörandl, 2006; Zander, 2007a, b, c), although molecular lineages themselves are often well supported. This paper is an attempt to return morphology to a primary role in systematics without deemphasizing the molecular contribution, and hinges on a new interpretation of non-monophyly as a way to progress beyond a tree of trait changes to a taxon tree.

Simply put, non-monophyly of exemplars of the same traditional phenotype-based taxon (hereinafter a “phenotaxon”) on a molecular tree implies a shared ancestor at that taxonomic level. That inferred phenotaxon is interpreted as an ancestor of two apparently evolutionarily static surviving populations. There is a large literature on whether paraphyletic taxa should be recognized in classification, and particularly valuable are the papers of Brummitt (e.g., 2003, 2006). Brummitt (2006) indicated that “paraphyly is the most important issue debate in taxonomy today.” The question of whether paraphyly is to be recognized as acceptable in classification is central to the present conflict between traditionalists and phylogeneticists, and I here suggest that not only is paraphyly acceptable, but non-monophyly in general may be evolutionarily informative. Non-monophyly of taxa is satisfied by either a paraphyletic group (a phenotypically coherent molecular lineage except for a different phenotaxon embedded in the lineage) or a polyphyletic group (phenotypically coherent groups presented as molecular lineages broken and scattered among other lineages on a molecular tree). Exemplars of the same phenospecies that are phylogenetically distant on a molecular tree (i.e., at minimum not sister groups) imply that the species has two surviving ancestral populations (the exemplars) in phenotypic stasis (e.g., through stabilizing selection) while continuing to undergo gradual accumulation of mutations in the genotype. An “isophenotype” is here defined in the taxonomic context as one or a number of specimens involved in a molecular analysis that share the same phenetic traits at some particular taxonomic level (no specimens are, of course, exactly alike). Exemplars of different phenotaxa bracketed on the molecular tree by the non-monophyletic isophenotypic exemplars (as lineages branching from the arc on the tree connecting the isophenotypes) are derived from the ancestral phenotaxon implied by the isophenotypic exemplars, and are nested terminally on a taxon tree of ancestors.

Two different species of the same non-monophyletic genus imply an ancestor with phenotype resolvable only at the genus level. Ditto subfamilies and families. Surviving ancestors (as species pairs with one derived from the other) have been inferentially demonstrated in the past through biosystematic and cytogenetic analysis, for instance, by Lewis (1962, 1966), Lewis & Roberts (1956)
and Vasek (1968) in *Clarkia* (Onagraceae). Persistence of evolutionarily static populations or species (Guillaumet & al., 2008; Leschen & al., 2008; Shen & al., 2008) is explained by the increasingly supported punctuated equilibrium theory of Gould & Eldredge (1993), which may be valid for many or most taxa, though apparently rare stepwise transitions have been demonstrated (e.g., De-méré & al., 2008).

### CONSTRUCTING TAXON TREES

A taxon tree can easily be constructed from inspection of published molecular trees that have many exemplars of species in each of several genera or of many genera in several families. The resolution of ancestral taxa related by descent with modification (Hall, 2003) is scarcely as good as the apparent resolution of changes on molecular phylogenetic trees, but the results are more direct, and are more in consonance with the aim of basing systematics on evolution of taxa. Future phylogenetic analyses that sample many populations of the same phenotaxon from distant areas or unusual habitats are most apt to uncover evolutionarily informative non-monophyly.

Inferring ancestral taxa from phenotaxa exhibiting non-monophyly depends on certain new assumptions and rejection of some past assumptions. Presently, non-monophyly is commonly attributed to convergence, and for higher taxa, is commonly lumbered with the unsupported ascription of “massive homoplasy”. Phenotypically essentially identical exemplars representing species, genera or even families, if not monophyletic on a molecular tree, are commonly labeled as “cryptic”, and one of them may be renamed to fit a phylogenetic classification enforcing monophyly (see e.g., Shaw, 2001; Hebert & al., 2004; Hebert & Gregory, 2005; Kress & al., 2005; Newmaster & al., 2006; Elmer & al., 2007; Hedenäs & Eldenäs, 2007; Zander, 2007c; Hedenäs, 2008). This will hide, however, evolutionary information about the phenotype of ancestral taxa. Complete convergence (or in the case of closely related but non-sister group taxa, parallelism) is commonly assumed at some taxonomic level, but one might argue that, except in cases of closely related taxa under similar environmental selective regimes, it is improbable.

For example, consider two phylogenetically distant exemplars that match phenotypically. The chance of, say, 10 evolutionarily neutral or randomly evolving traits shared by the two exemplars to converge to identity is the reciprocal of 2 to the tenth power, or 1/1024. Even five neutral traits match at only 1/32 probability. That neutral traits, or spandrels (sensu Gould & Lewontin, 1979), or nearly neutral traits exist, among other evidence, demonstrated by the co-appearance of similar species of different genera with apparently the same biorole in the same habitat. Given that morphological taxonomy has found that most genera have clear-cut traits and have gaps between, there is no spectral gradation and lack of taxonomic distinction between higher taxa that might be caused by rampant convergence of taxa through random modification of neutral traits or even those subject to selection. Convergent selection of phenotype (and linked traits) should in most cases be readily associated with selection into the same biorole. Complete identity is doubtful involving conservative traits resistant to change or at least major change (phylectic constraint) that might be involved in convergence across lines of genera or higher taxa, such as might be involved in phylogenetically distant isophenotypic exemplars.

### RATIONALE

The null hypothesis (state of nature) for tapping isophenotypic non-monophyly for evolutionary information is that any exemplar may be evolutionarily derived from any other exemplar that is isophenotypic (e.g., same morphology) of the same taxon at any taxonomic level. At taxonomic levels of genus or higher, the included subunits of each are considered sufficiently phenotypically similar such that the phenotype of any subunit may be derived, as a null hypothesis, from any other subunit of that one taxon. Rejection of the null allows the alternative hypothesis of recognition of taxa included in the genus or higher group as ancestors. This evolutionary concept of higher taxa requires at least some distinctive and identifying combinations of phenotypic traits apparently resistant to random evolutionary change, quite along the lines of traditional taxonomy. Rejection of the null is the simple demonstration that in fact two (or more) exemplars isophenotypic at the same taxonomic level are phylogenetically disjunct (non-monoptyhic) on a molecular tree. Classifications based on phylogeny should be changed to accommodate non-monophyly-based taxon trees unless convergence of traits can definitely be demonstrated through identification of identical selection pressures and absence of improbable multiple converged neutral traits (which might be the case among species of the same genus but such would not be expected to converge into the same biorole), or the phenotaxon may be split along phylogenetic lines for some better reason than simply enforcing phylogenetic monophyly.

Given the assumption of a null hypothesis that in one taxon at any particular taxonomic level, any of the included taxa of the next lower category might be derived phenotypically from any other, trees of molecular state changes are poor guides to taxon evolution. Even inferred
morphological state changes mapped on a molecular tree simply support the idea that ancestral phenotaxa should be similar to the nearest exemplars, but no combinations of states suitable for interpretation as taxa are given. The idea that there are no or few surviving ancestors, exemplified by the maxim that a molecular dichotomy must be a speciation event and also implies disappearance of an ancestor and development of two different taxa, is artificial at best. This is because there is much evidence of evolutionary stasis of phenotypes, which might better explain isophenotypic non-monophyly than appeal to convergence. The assumption of modern systems that species are not phylogenetically complex must be discarded, not only because of the existence of non-monophyletic isophenotypes at the species level, but by examination of published robust phylogenies at the species level, for instance the considerable complexity of the internal lineage of the domestic cat species (Driscoll et al., 2007). The assumption that species disappear when molecular lines bifurcate is true only in the sense that molecular mutations continue whether or not the morphology changes, and the molecular “ancestor” of two molecular sister lineages is, of course, molecularly different. That a molecular split is necessarily or even often a speciation event and that ancestral phenospecies actually disappear in a bifurcation of a molecular tree are burdensome assumptions, and the extinct species that is assumed for every two bifurcating lineages is wholly imaginary and non-parsimonious, particularly poignant when non-monophyly is real evidence for surviving ancestors.

Inferential demonstration of the phenotypes of an ancestral taxon (in absence of fossil information) involves identification of surviving ancestors. This can be done (1) by biosystematic and cytogenetic studies, particularly in the case of “quantum” or local evolution (Lewis, 1962; Grant, 1971; Levin, 2001), which is the budding of a daughter species from a peripheral ancestral population, and including the more recent method of Theriot (1992), who inferred a surviving ancestor in a group of diatoms by evaluating a morphologically based cladogram and biogeographical information; or (2) the somewhat more simplistic and problematic selection of a surviving ancestor as one lacking autapomorphies on a morphological tree; or (3) the method of isophenotypic non-monophyly introduced here.

**INTERPRETING PREVIOUSLY PUBLISHED PHYLOGENIES**

The method of interpreting isophenotypic non-monophyly on molecular trees as evidence of ancestral taxa involves the following generalizations. For this discussion, we here assign a phenotaxon a letter A, B, C, etc. Exemplars (sampled specimens of particular taxa usually named as species) of each phenotaxon are given subscripts A1, A2, etc. Consider the terminal lineage (A1, B1) A2… X, where “… X” is the remainder of the cladogram. The exemplars A1 and A2 are paraphyletic, and if A is a species, then the ancestor of A1, A2 and species B is species A. A nested example is (((A1, B1) A2) C1) C2) … X, with two paraphyletic species A and C, and this cladogram implies species A is the ancestor of A1, A2 and species B, while species C is ancestor of species A. Exemplars of paraphyletic taxa are in this manner interpreted as surviving ancestors with other taxa evolutionarily budding off their respective lineages.

Determination of non-monophyly is dependent on reliable molecular analysis (Zander, 2007a). The reliability of any pair of lineages or exemplars implying a surviving ancestor depends on the joint probability of the two being indeed phylogenetically disjunct by at least one internode. For example, in ((A1,B1)A2) … X, for exemplars A1 and A2, of species A, to be reliably non-monophyletic depends on the probability of A1 and B1 being sister groups on a molecular tree, while A1 and A2 must also be involved in calculation of joint probability if these are lineages of multiple species and there is uncertainty that these are not monophyletic.

Such simplicity and excellent resolution cannot be expected in practice given the few exemplars of each taxonomic category in most published molecular studies, and there are problematic scenarios with overlapping inferred ancestors. For example, the cladogram (((A1, B1) A2) B3) … X implies that species B may be the ancestor of all other exemplars and the species they represent, yet A is the ancestor of exemplar B3. This “double ancestor” (B overlapping part of the lineage of A) may have various explanations: (1) There may be lacking an exemplar not yet studied (or now extinct) the molecular traits of which would rearrange the cladogram to an acceptably nested form. For instance, a superimposition of ancestral morphotypes implied by, e.g., (((A1, B1)(A2, B2)) … X can be resolved to the fully nested (((A1, D1)A2)B1)B3) … X by addition of another exemplar D1 that shares many advanced traits with A1 and A2 but few or none with B1 and B2. (2) There may be, in fact, a single phenotypic ancestor (say, species B) which retains traits of the other species (say, species A) through balanced polymorphism, or epigenetically, such as a suppressed gene complex (Caporale, 1999, 2003; Bachmann, 2001; Le Rousic & Carlborg, 2007; Zander, 2007c, d). These and other scenarios (e.g., punctuated evolution) might be investigated with modern techniques, but isophenotypic non-monophyly should not be administratively discounted as convergence when alternative evolutionary explanations are plausible or even more likely.
EXAMPLES

The reader is encouraged to examine published molecular cladograms in his or her field to find cogent examples of phylogenetically distant isophenotypes interpretable thusly. For instance, Zander (2008) found after reanalysis of previously published research of other authors that the moss family Dicranaceae was the paraphyletic ancestor of two phylogenetically disjunctive and well-supported reciprocally monophyletic clades, the Dicranaceae s.str. and a lineage of generally small-sized taxa commonly referred to in phylogenetic literature as the “Rhabdoweisiaceae” but with no agreed morphological diagnosis to distinguish this phantom taxon from the Dicranaceae. Following the line of thought presented here, the Dicranaceae is then the phenotypic ancestor of the embedded families Calympereaceae, Erpodiaceae, Leucobryaceae and Pottiaceae.

Figure 1 is a composite taxon tree combining information from two large molecular cladograms of the moss family Pottiaceae, that of Werner & al. (2004) as reinterpreted with reliable (95 percent or more Bayesian credibility) branch arrangements by Zander (2007a), and of Werner & al. (2005), the branch arrangements of which are mostly reliable as published. From the cladogram of Zander (2007a), it can be immediately seen that the Pottiaceae subfam. Timmielloideae (exemplar *Timmiella crassinervis*) is phylogenetically disjunct from the remainder of the Pottiaceae. The inferred ancestral phenotaxon is not resolvable as anything but Pottiaceae. The non-monophyly of the genus *Pseudocrossidium* provides a deep ancestral taxon for much of the remainder of the Pottiaceae at the genus level. The family Ephemereaceae is inserted with best resolution in Fig. 1 directly on the genus *Pseudocrossidium*, while the family Cinclidotaceae and the subfamily Barbuloideae may both be interpreted as derived from the subfamily Pottioidae.

The second study relevant to Fig. 1 is that of Werner & al. (2005) and, though more complex, is readily available for examination in the pages of *Taxon*. The tribe Leptodontieae (exemplars in genus *Triquetrella*) is clearly embedded in the subfam. Barbuloideae (exemplars including *Didymodon*, *Hyophila*, and *Aneostangium*). The genus *Chenia* is derived from *Tortula* (bracketed by *T. muralis* and *T. inermis*, which provide the ancestor resolvable at the genus level). *Trichostomum tenuirostre* has phylogenetically distant exemplars, and these bracket the genus *Chionoloma*, which itself brackets *Pseudosymblepharis*. The genus *Pleurochaete* is embedded among various exemplars of species of *Tortella*, while *T. tortuosa* and *T. fragilis* are both non-monophyletic and seem to overlap.

Fig. 1. An evolutionary taxon tree of the Pottiaceae combining the Werner & al. (2004) and Werner & al. (2005) data. It shows resolvable paraphyletic or polyphyletic ancestral taxa interpretable as virtual fossils. The remainder are included in the various residual groups of no discernable evolutionary structure. Families buried in the paraphyletic Pottiaceae are the Cinclidotaceae and Ephemereaceae, and (not shown) those arising directly from the Pottiaceae residuum are the Dicranaceae, Fissidentaceae, Grimmiaaceae, Ptychomitriaceae, Seligeriaceae, and Wardiaceae. The ancestral taxon from which the Pottiaceae itself is derived is unclear.
each other’s ancestors. *Weissia controversa* is strongly non-monophyletic and implies an ancestor that can only be derived from the Trichostomoideae, and many other *Weissia* species (not shown) are derived from that species. The groups listed as residua include perfectly acceptable phenotypes represented by exemplars in the cladograms for which no clear resolution of taxon evolution is available.

One may note that the taxon tree of Fig. 1 is much like a Besseyan “cactus” (e.g., Bessey, 1915). This is not unintended, and serves to emphasize that a taxon tree represents inferred descent with modification of taxa, not of traits. This paper does not propose that molecular and morphological evidence are equally valid in the usual sense of equal contribution to a phylogenetic classification, but that molecular evidence can present changes in sequence lineages through time as well-resolved trees, and interpretations from isophenotypic non-monophyly on molecular trees may clothe such barren frameworks with ancestral taxa.

In other literature, a casual survey will commonly reveal well-supported non-monophyly that implies ancestral taxa. For instance, the study of the Knoxiae (Rubiaceae) by Kärched & Bremer (2007) detailed a well-supported phylogenetic disjunction of exemplars of species of *Otomeria* (embedding derived species of *Batopedina* and *Parapentas*), of *Pentanisia* (embedding species of *Neopentanisia* and *Calanda*), and of *Pentas* (embedding species of *Carphalea, Knoxia*, and *Placopoda*). The molecular tree indicated that the phylogenetic relationship of the three ancestral genera is (*Pentanisia, Otomeria* *Pentas*, while the actual evolutionary relationship (one as the ancestor of another) is better conceived as a Besseyan cactus with both *Otomeria* and *Pentanisia* budding off from *Pentas*. Exemplar species of *Parapentas* are widely disjunctive on the molecular tree, separated from *Otomeria* and *Pentanisia* by several genera, and if re-examination supports this as true homoplasy (and not better interpretable as two different genera) then *Parapentas* would be an intermediate on the taxon tree between the rather basal *Pentas* and the other two surviving ancestral genera.

Four well-supported ancestral genera can be perceived at the crown of a fern phylogeny (Schuettpeltz & Pryer, 2007: Eupolypods 1, part 2) with exemplars of surviving ancestors *Ctenopteris* and *Lelingeria* arising from *Terpsichore*, which arises from *Grammitis*. In the Senecioneae (Asteraceae), relationships (Pelser & al., 2007) within the subtribe Othoniinae clearly demonstrate *Othonna* as a surviving ancestor, and *Senecio* the same in the subtribe Senecioninae. The appearance of exemplars of certain genera (e.g., *Curio, Dendrophorbiurn*) with exemplars themselves phylogenetically disjunct within a range of *Senecio* exemplar branches indicates that the molecular lineage may document either parallel development of the same genus from a *Senecio* ancestor, or a switching back and forth among ancestral genus morphologies (as ancestral species of consensus morphotypes), perhaps via silenced gene complexes. These lineage splits are, in some cases, well supported.

A study of the Coreopsideae, Asteraceae (Mort & al., 2008) showed non-monophyly of *Coreopsis*, exemplars of the genus appearing in scattered groups throughout the molecular cladogram, implying that this genus is the surviving ancestral morphotype for most genera of the tribe, including *Bidens*.

In an investigation of lava lizard phylogeny, Benavides & al. (2007) demonstrated phylogenetic disjunction on a nuclear DNA tree of *Microlophus quadriovittatus*, which bracketed *M. atacamensis*, and of *M. tigriss*, which bracketed *M. peruviunus*, although this is contradicted by the mtDNA clade. The consensus tree supported the nDNA tree, which, according to the authors, also makes better geographical sense.

### SUMMARY

All the above estimates of ancestral taxa must be viewed as somewhat speculative and a first attempt at inferring taxon trees, in spite of commonly high support values, because of the general paucity of exemplars in all studies. What I hope is new in this paper is an alternative to basing systematics on lineages of trait changes. The taxon tree incorporates advances in molecular systematics but goes beyond phylogenetics to chart descent with modification of living things. Exactly how classification, particularly modern phylogenetic classifications, might incorporate information from a taxon tree is yet to be determined. Papers by other authors contributing relevant ideas essential to the method of isophenotypic non-monophyly include those of Sosef (1997), Brummitt (2003, 2006), Lee (2005), Nordal & Stedje (2005), Alexander (2006), Hörandl (2006, 2007), and Farjon (2007) among others.

The reader is asked to try this method. First, one must find a robust, reliable molecular cladogram with duplicate exemplars representing species, genera or higher taxa. The names assigned to the exemplars must represent traditional phenotypic taxa, not later phylogenetic constructs. Then, draw lines in an arc along the cladogram structure between exemplars that are non-monophyletic (that is, not sister groups or not in a monophyletic lineage) but representing the same species or the same genera (or higher taxon). All along this arc one infers an ancestor at that taxonomic level (species or genus). Those exemplars with the same phenotype of their represented taxon plus the taxa represented by exemplars of lineages diverging from the arc (bracketed by the isophenotypes), are all descendents of the ancestral taxon. Create then a taxon tree.
of the ancestors and derived taxa, plus unresolved “residual” groups of taxa without clear phenotypic ancestral relationships.

Phenoancestors distinguished by the kind of phylogenetic triangulation described above are quite like fossils in that they can be dated (with standard techniques estimating divergence times of molecular lineages rather than, say, radiocarbon). In addition, their descendents can be directly inferred by shared molecular traits. Such “virtual fossils” can be fully described by consensus traits of their surviving populations. It is probable that continued research will find far more of these virtual fossils than fossils and subfossils. This will help in understanding evolution at least among taxa exhibiting phenotypic stasis in populations separated by time and place.

Key points for this syncretic method are: DNA base pairs continue to gradually change, yet phenotype may not significantly change, even when molecular lineages diverge. As a null hypothesis, any subunit phenotaxa of one higher taxon may evolutionarily derive directly from any other subunit phenotaxon of that higher one taxon. Ancestors should have phenotypes similar to nearest exemplars on a tree, but when two exemplars of same phenotype are distant on a tree, either they are surviving ancestors or are phenotypically convergent. Convergence on the level of complete crypsis at a particular taxonomic level is improbable, particularly involving two or more higher taxa. A taxon tree inferred from molecularly non-monophyletic, isophenotypic exemplars (as surviving populations of ancestors) provides a direct representation of evolution that should aid classification, and is in fact direct evidence for macroevolution.

ACKNOWLEDGEMENTS

I appreciate the support of the Missouri Botanical Garden in this continuing work. Relevant communications with R. K. Brummitt, R. Mesibov and A. R. Templeton were much appreciated, as were comments of reviewers.

LITERATURE CITED


