WHEN BIODIVERSITY STUDY AND SYSTEMATICS DIVERGE

Van Valen’s (1976) ecological species concept (with taxa occupying differential niches or adaptive zones and evolving separately) was a major advance in science, but has been superseded for the past 30 years in systematics by “tree-thinking” and the phylogenetic species concept (taxa being the smallest group of individuals sharing a unique set of advanced traits). Phylogenetics, as the new systematics, has replaced the “modern evolutionary synthesis” (Mayr & Provine 1980; Smocovitis 1996) in the form of a paradigm shift. The present Biodiversity Forum article calls for a similar shift re-emphasizing as basic to systematics certain aspects of that prior synthesis. Phylogenetics basically consists of inferring ancestral state changes from morphological or molecular data (or both at once) on a dichotomously branching tree. Rather than generating degrees of differences between taxa, as in the diagnostic systematics of yore, it instead provides degrees of similarity, essentially limited to genealogy. Concepts of species and higher taxa in phylogenetics are bounded by (1) an unfortunate ideology (conflation of theory and fact) of discovery of phylogenetic patterns in nature that, because they are supposed to be real and not models, are nameable, (2) the stricture of required monophyly (a monophyletic group consists of a common ancestor and all its descendants) and (3) the assumption that surviving ancestors are lacking or very rare. Phylogenetic analysis does not directly include information on phenome/environment interaction other than mere mapping of geography, for instance, post hoc on a molecular tree.

The molecular tree infers, at best, only probabilistic genealogies of (1) populations when nuclear genes are used, or (2) pedigrees of individuals when haplotypes provide the data. Haplotypes are matrilineally inherited DNA from organelle mitochondria and chloroplasts (and the analytically less commonly used patrilineally inherited DNA from Y chromosomes), and are passed without genetic recombination to progeny solely from the parental female (or male in the case of Y chromosomes). Phylogenetics uses both nuclear and organellar DNA when inferring tracks through time of presumed evolutionarily neutral base changes. Given the difference between nuclear and organellar DNA inheritance, there are then two kinds of phylogenetic species concepts, one that generates probabilistic lineages of genetically isolated populations, and another that infers branching pedigrees of individuals whether these lines are in only one population or in several. When the two kinds of data are analyzed together, as in “total evidence” studies, a mix of inferred population and individual genealogies results. Exactly what is represented by the tree and the

exemplar specimens as OTUs (Operational Taxonomic Units are any groups of organisms used in a cluster or phylogenetic tree analysis without concern for taxonomic rank) is further confounded when morphological traits are added to the data set.

Monophyly

Phylogenetic analysis cannot infer from molecular data any terminal entities (that which exemplars are supposed to represent as OTUs) larger than a genetically isolated population (or in the case of haplotypes, one individual), and to this extent phylogenetics enforces the biological species concept (species are genetically different and isolated populations). The biological species concept has been severely criticized by Rieseberg and Burke (2001) whose studies support the idea that that which holds species together is not gene flow across all loci (which is seldom high enough) but instead the dispersal of advantageous alleles (see also Hendry et al. 2001; Syring et al., 2007). An allele is any one of a number of viable DNA codings that occupies a given locus (position) on a chromosome. As long as the selective advantage of an allele is fairly high, even low levels of gene flow will spread the allele though semi-isolated populations. This, coupled with parallel balancing or stabilizing selection, or simply phyletic constraint, is taken here as theoretical justification for a semi-isolated or isolated multipopulation species circumscription as may occur with the ecological species concept. Strict monophyly is a substitute for a lack in phylogenetics of inferable species limits beyond that of the isolated population, when data is restricted to presumably neutrally evolving DNA. This is especially true in light of Reiseberg and Burke’s assertion that just a very few advantageous genes can effect and maintain species integrity, and gene flow. Modifying the whole genome is theoretically then not required. This is supported by the mutation theory of Nei (2005), who proposed that mutation resulting in adaptively important codons is the driving force of evolution, and by the neutral theory of molecular evolution in general, which predicts that rates of phenotypically expressed change are largely independent of mutations in the genotype (Davies & Savolainen 2006).

The simplest case (Fig. 1) of the problem of treating monophyly as a first principle—and thus all deductions must be correct—is a taxon of two genetically isolated populations (say, A in Jamaica and B in western Mexico). From A there arises a new species C with a different morphotype, reproductive strategy, and habitat, but A does not “go away” so C is a daughter species not a sister species though it would be treated as the latter in phylogenetic analysis. All these populations, A, B, and C, then gradually and separately accumulate neutral DNA
mutations (Fig. 2) and become molecularly different, but A and B retain the same phenome and environmental traits. To conserve monophyly, all three must be named as different species, with A and B having reciprocally cryptic status. If morphology is ignored (and this is becoming more and more the case in modern systematics), all molecularly different populations otherwise contrary to monophyly can be named as different species when nuclear loci are analyzed, and all individuals can be named as species when matrilineally inherited loci (mitochondria and chloroplasts) are used and monophyly is required. The way to attack this problem is to consider the case where all three above populations were essentially identical in expressed traits. Would they all be named as cryptic species or all considered merely one species with complex and varying genotype? Doubtless the latter, as in the real case of a phylogenetically complex species of moss (Shaw 2000).

**Barcoding**

DNA barcoding (Hebert & Gregory 2005) is the use of a particular short DNA sequence, mutating at approximately the rate of speciation, to quickly identify species; the species are expected to mostly differ in sequence only between each other. Changes in the sequences are random and are not caused by genetic isolation, but instead we assume a fairly constant molecular clock. The mitochondrial COI gene is usually used (Hebert et al. 2004) though the chloroplast trnH-psbA intergenic spacer (Kress et al. 2005) and rbcL region (Newmaster et al. 2006) have been recommended for plants. There is a large literature pro and con, with emphasis on exactly what value is received given costs in time and funding (e.g., DeSalle 2006; Goldstein et al. 2000; Lambert et al. 2005; Moriz & Cicero 2004; Rubinoff 2006; Wheeler 2005; Will & Rubinoff 2004). Most particularly relevant is the critical work of Funk and Omland (2003) on the heterogeneous nature of genome change over time, resulting in a mosaic of loci reflecting relationships to different organisms (see also comments of Rieseberg & Burke 2001). On the other hand, one would think that if DNA barcoding were done in conjunction with traditional methods on particularly amenable groups in light of its limitations by those with funding and time to do so, there should be little problem and some advantages. In groups with known species-level thresholds for mutation rate of the particular sequence chosen for barcoding, specimens of potential new species can be tentatively segregated with barcode analysis and confirmed as representing new species by determination of characteristic morphological and other traits expressed in nature, a fail-safe procedure.

A classic example of the interplay of molecular and morphological study resulting in taxa based largely on the ecological species concept is the study of Hebert, Penton et al. (2004), who divided Astraptes fulgerator, a neotropical skipper butterfly known as the Two-barred Flasher, into 10 species following DNA barcoding of a large series of specimens, in the context of years of suspicion by taxonomists that this was a complex of more than one species. The new species are not fully cryptic in that, although adults are nearly identical, the larvae are generally distinctive, have different browsing habits, and rather different ecosystem preferences. (Cryptic species can only be separated using non-morphological data, such as from DNA sequence analysis, chemistry, bioacoustics, or life history studies.) But, such joint analysis cannot be expected to work effectively when fully cryptic taxa are identified by molecular traits alone. Surveys (e.g. Clare et al. 2007; Hebert et al. 2004) have found many potentially new species that may prove fully cryptic, distinguished from known species by molecular traits alone, and there is some chance that these will be presented to science as acceptable.

---

Figures 1-2. 1, Diagram of an hypothetical ecological species producing, over time, two genetically isolated or semi-isolated populations, B then C. From each of the three are sampled three specimens, A, B, and C. The population (shaded) represented by Specimen C (but not Specimen B) has a somewhat different ecology and associated different set of adaptively advantageous traits. All populations accumulate evolutionarily neutral mutations over time. 2, Phylogenetic structure of the evolving set of populations in Fig. 1. The morphological and environmental study of Specimen C allows inference of a new and different ecological species derived from the population represented by Specimen A. Specimens A and B belong to the same ecological species but may be treated as different cryptic species under other (biological, phylogenetic, evolutionary, barcode) species concepts, in particular because they differ molecularly from each other.
following the phylogenetic species concept (as required by strict monophyly), or an ad hoc barcode species concept as some pre-selected level of sequence difference (Lambert et al. 2005), or simply a relatively large genetic distance, exemplified by the isozyme work of Bell et al. (1998) on New Zealand frogs.

Any potential new species identified as such by DNA barcoding must stand alone on their expressed traits because the barcode molecular traits are generated randomly. There are too few molecular data in the barcode sequence to statistically demonstrate that particular DNA changes are not to be expected by chance alone. Any suggestion that DNA barcoding actually supports a potential new species (rather than merely flagging the possibility) may be a multiple test problem (Benjamini & Hochberg 1995; Miller 1981) in statistics, because it is tempting to accept poor morphological traits as acceptably diagnostic under the apprehension that they are supported by barcoding. In my own field, phylogenetic analyses using mainly chloroplast DNA have turned up 14 cryptic or nearly cryptic species of bryophytes (Shaw 2001); the problem is now with the level of sophistication in statistics needed by systematists and biodiversity specialists to intelligently accept or reject them. Although fully cryptic taxa (with no expressed traits) can be rejected across the board, as encouraged here, the evaluation of the taxonomic utility of a few weak morphological traits associated with the molecular differences may be difficult. For instance, multiple test problems were identified (Zander & Eckel 2007) as cause to reject a molecular species of the Tortula subulata complex where one morphological trait had been discovered by the authors from a set of possible traits post hoc as supporting a clade not having particularly high molecular support. In another case, when I asked an expert on the Scrophulariaceae whether the recent molecular splitting of that family was justified, the response was that the group was not well justified morphologically; but, ambivalence about morphology cannot increase statistical support for a molecular split, which must stand on its own.

**Fractal Speciation Versus Conspecific Paraphyly**

Whether or not to allow paraphyly, the case when not all descendents are included in a taxon, such as when a surviving ancestor develops a new morphospecies from one of two or more isolated populations, is a presently a major item of contention in classification (e.g., Brummitt 2003, 2006; Hörandl 2006). An example is the fairly well-known fact that polar bears are more closely related to certain populations of brown bears than those populations are to other brown bear populations (Talbot & Shields 1996; see photos 1 and 2). Following the requirement of monophyly in classification, one could lump them all, but lose recognition of the polar bear’s unique morphotype and biorole; or, recognize three species, but create two reciprocally cryptic brown bear species. Accepting two species with paraphyly of one by recognition that polar bears are derived from one of several populations of brown bears is a better and more natural solution. An example of paraphyly at a higher taxonomic level is the group “reptiles” when birds are excluded. When paraphyly is rejected, the paraphyletic group must be split into differently named taxa. When working with nuclear DNA, monophyly requires a name for each split off genetically isolated molecularly different extant population, and with haplotypes, a name for each extant molecularly different exemplar individual. This may be termed fractal speciation. A hyperatomistic species concept will in practice gradually result in same-named organisms that react much the same in experimental circumstances, a good thing, but increasing numbers of differently named organisms also reacting the same, a source of uncertainty. Some limit to splitting is needed based on a different or additional criterion. The insistence on monophyly is epidemic in systematics today, but fractal speciation is clearly deadly to biodiversity analysis. For instance, in Figs. 1 and 2, the genetic distance between exemplar specimens A and B may be greater than between A and C, or B and C; and, under biodiversity triage preserving greatest genetic distance, the taxon represented by Specimen C would be given short shrift. By extension, such problems may occur at higher taxonomic levels. Given that all populations of a multipopulation species continue to generate new neutral molecular traits, Cronquist’s (1975) urging the acceptance of conspecific parallelisms in closely related species complexes is still relevant. Also, important here is a call by Page (1998) to map molecular changes on species trees to examine alternative phylogenies. Hörandl (2007) recently came to much the same conclusion: that evolution is poorly represented in phylogenetic analysis.

The ecological species concept, employing the useful idea of niche, continues to prove popular as shown by its frequency in articles in Biodiversity and other journals in the field. Of course, because the concept more fully describes nature, it is not as simple or mechanical as the phylogenetic species concept. For instance, an ecological species may be a phenotype or a range of phenotypes filling or partly filling a realized niche or a range of niches. (A phenotype is a set of measurable expressed traits that may or may not be variably expressed when affected by the environment.) A diagnosis of a species that includes elements of population biology and ecotype, no matter how abbreviated or rudimentary, is however, essential for biodiversity studies. The niche, as the environmental dimension, is not tracked or measured by phylogenetic analysis, no matter how consistent or robust the analysis is internally. The concept of niche applies whether expressed genes are selected for or are evolutionarily neutral and simply fixed after, e.g. bottlenecks of population size. Although morphologically intergrading species and
Adaptations by the polar bear to life on sea ice include a white coat with water repellent guard hairs and dense underfur, short furred snout, short ears, teeth specialized for a carnivorous rather than an omnivorous diet, and hair nearly completely covering the bottom of the feet.

The neutral theory of biodiversity (Hubbell & Lake 2003) suggests that the niche is irrelevant in certain situations. This theory postulates that patterns of species abundance, diversity and distribution may be explained without reference to fitness and adaptation but merely by factors of dispersal, speciation rate, and numbers of individuals. This theory, however, applies to macroevolutionary theory and community structure, not to comparing organisms in a taxonomic context; it is expected to work well only with similar sized and similarly trophic, sympatric organisms in tight competition (e.g. tropical trees); and it is basically a null hypothesis in which even if common garden experiments demonstrate no difference between variance of all traits and that of traits expected to be subject to environmental selection (Hendry et al. 2001), still no refutation of evolutionary selection in the context of the niche is allowed. These and other objections are discussed more fully by Gaston and Chown (2005) and Stockwell et al. (2006).

What an exemplar actually represents depends on the species concept employed in the analysis. A species name, say species A, is applied to a specimen, which is treated as an OTU exemplar, say Specimen A. After the analysis, the OTU, which may have represented a traditional species at first, is now redefined by whatever restrictions are involved in the data used and the analytical process. Figure 1 is a simplistic representation of an ecological species budding off over time two isolated or semi-isolated populations, with sampled specimens A, B, and C representing each population as exemplars in analysis, and all sharing evolutionarily neutral mutation set number 1 and a basic morphology (Fig. 2) characteristic of the, say, genus. The population represented by Specimen B has no particular expressed traits different from those of Specimen A. The population represented by Specimen C colonizes a different environment in a different manner, and does develop new expressed traits that are adaptively advantageous for its particular niche. All populations gradually accumulate mutations in evolutionarily neutral loci of DNA.

Figure 2 is a skeletonized representation of the phylogenetic structure after analysis of the three specimens. Autapomorphies are defined as the unique (at least locally), nonshared, traits, molecular or morphological, found in only one terminal OTU. The molecular changes associated only with Specimen A are treated as autapomorphies (see also Fig. 5) in maximum parsimony analysis, and exemplar Specimen A is then treated as representing a different entity than the ancestor that is represented by the node shared by populations A and C. This would justify treating population B as a cryptic species, with its own molecular autapomorphies (neutral mutation set 4). This follows the phylogenetic species concept but not the ecological species concept, which unites populations A and B in one species even though Specimen A differs from Specimen B by neutral mutation sets 2 and 3. In Figs. 1 and 2, the surviving ancestor is defined as that population (here with Specimen A in it) that buds off a new ecological species (here with Specimen C in it) with no change in itself except for gradual accumulation of evolutionarily neutral molecular traits. Various different neutral traits occur in exemplars A, B, and C. These molecular differences, in fact, may be expected to accumulate in all extant populations of any significant age. Both exemplars A and B here represent different populations of the same ecological species, and there is no real separate ancestor represented in the cladogram by the node shared by lineages A and C.

The morphology of surviving ancestors, at least those traits that are critical for survival, evolved basal to the splitting...
off of a daughter species, and that daughter species necessarily reversed those traits. Therefore, the length (number of steps) of a true morphological (or combined morphological and molecular) tree with one or more surviving ancestors is longer than the tree of maximum parsimony; this is why surviving ancestors are problematic in phylogenetics though largely eliminated from consideration by the analytic process, maximum parsimony (Fig. 5). There may be many lineages of surviving ancestors, and each surviving lineage may have autapomorphic morphological traits critical for survival that need to be viewed as evolving earlier than calculated or when mapped on a molecular tree. Because autapomorphies of an ancestor become the synapomorphies of derived taxa in a lineage (less reversals), the problem with maximum parsimony not recognizing reversals in daughter species is cladogram-wide. It is not easy to identify which terminal (much less which internal branch) is a surviving ancestor on a cladogram. Exactly how to map morphological traits on a molecular cladogram is problematic, as one has a choice between placing morphological traits as diagnostic of a clade, or as a guess where they might have evolved basal to a surviving ancestor and daughter lineage pair (Fig. 5) if these could be identified. Correct placement on a tree of the evolution of morphological traits is important in gauging expressed evolutionary change in geological time. Given the uncertainty and difficulty of dealing with morphology in phylogenetics, and the futility of establishing species limits by inferred genealogies of neutral mutations, the evolution of expressed traits remains best inferred, observed, and recorded among extant species interacting with each other and the physical environment.

Diagnostic Systematics

The time has come to return to diagnostic systematics that infers evolutionary distinction from a combination of morphology, reproductive strategies, and environmental interaction (niche). These are the traits commonly used today as basic in the literature on biological diversity (as opposed to that of at least theoretical systematics). Although there is a global crisis in biodiversity involving massive differential selection on species, the cutting edge of systematics remains transfixed by genealogy as tracked by neutral mutations, i.e. tree-thinking. Paraphyly (whenever a species or higher group has new and different taxa arising from the inner workings of its complex phylogenetic framework) should be considered acceptable when massive homoplasy on a cladogram is not expected by chance alone (Zander 2006), or there is clearly parallel selection (Cronquist 1975) or at least identical neutral evolution through drift (Ohta & Gillespie 1996) of expressed traits among elements of a phylogenetically complex ecological species (Fig. 6).

There is no information whatever in the population and individual genealogies generated by molecular systematics as now practiced that allows integration of multiple genetically isolated populations into phenetically and environmentally unique species. Attempts to integrate morphological and environmental information is hampered by the stricture of monophyly, the biological species concept inherent in molecular analysis, and general rejection of surviving ancestors; this all results in unjustified taxonomic splitting. The phylogenetic species concept is thus more a feature of the limitations of the analytic method than an attempt to compose a basic unit of taxonomy that reflects best all relevant data. Figure 6 illustrates quandaries not addressed by species concepts that avoid the environmental aspect. A further problem is

Figures 5-6. 5, Morphological traits on an hypothetical cladogram of species X, Y and Z. Because X is a surviving ancestor, the morphological trait marked by the open circle on the terminal branch as autapomorphy (trait unique in the cladogram) actually evolved prior to the budding off of new ecological species Z. For this reason, a reversal of that trait, marked by the shaded circle, must be added to lineage Z. There are other morphological traits associated with Z that distinguish it, but the reversal of the autapomorphy is necessary, and increases the length of the tree by one step; 6, Cladogram of a hypothetical phylogenetically complex group, gradually deriving genetically largely isolated populations represented by exemplar Specimens B through I, including some identical in phenotype and biorole to the surviving ancestor represented by Specimen A. How many ecological species are there? How many species would there be, including cryptics, if a different species concept were used (phylogenetic, barcode, biological)? How distant must cryptic lineages be from each other to deserve a separate name, if ever? Is there another metric not represented that is appropriate for biodiversity studies?
that later users (e.g., biodiversity specialists and biogeographers) of the largely speculative phylogenetic literature have a difficult time distinguishing reliable conclusions of evolutionary relationship (say branching patterns with 95% probability) from those that are merely the best of several alternatives (Zander 2007).

Stakeholders in biological diversity study should encourage systematists to more commonly broaden their analytic methods with techniques of “biosystematics” such as common garden and reciprocal transplant, organismal chemistry, biogeography, population biology, and observation of modes of interspecific competition to approach the ideal of a taxonomy based on process-based comparative evolutionary ecology. The basic unit of systematics would then be the basic unit of biodiversity. Phylogenetic analysis resulting in inferential trees of populations and individuals would contribute to this effort but not overwhelm by excising the ecological dimension.

By Richard H. Zander
Missouri Botanical Garden
PO Box 299
St. Louis, MO 63166-0299, U.S.A.
Email: richard.zander@mobot.org

References


