Zander, R.H. 2014. Classical determination of monophyly, exemplified with *Didymodon* s. lat. (Bryophyta). Part 2 of 3, concepts. Phytoneuron 2014-79: 1–23. Published 6 August 2014. ISSN 2153 733X

CLASSICAL DETERMINATION OF MONOPHYLY, EXEMPLIFIED WITH *DIDYMODON* S. LAT. (BRYOPHYTA). PART 2 OF 3, CONCEPTS

RICHARD H. ZANDER
Missouri Botanical Garden
P.O. Box 299
St. Louis, Missouri 63166
richard.zander@mobot.org

ABSTRACT

Classical systematists infer evolutionary monophyly by using clues to adaptive or relatively neutral transformative radiation. If such clues are on a logarithmic scale they may be added to yield a probability for direction of evolution of one taxon to another. Such logarithmic clues are the decibans used by World War II code breakers in England. For a group, transformative traits are those convergent among disparate taxonomic groups, while conservative traits or trait combinations occur in multiple species and environments. Stem-based evolutionary trees (caulograms) are generated by models of serial evolutionary change. Direction of macroevolutionary transformation on a caulogram is determined by general morphological cladogram position, and maximum Bayes factor or deciban differential except when an intermediate taxon may be proposed, either from the extant set of terminal taxa or as an unknown shared ancestor that minimizes Bayes factor differences.

First, reassurance. This paper is mainly intended for classical taxonomists or the interested student. It attempts to explain what classical taxonomists do intuitively to generate evolutionary classifications. All relevant mathematics and statistics are simple and here thoroughly explained. Formalization in this paper means determining the statistical or logical basis for organizing related species according to increasing derivation away from some central apparently ancestral species. This paper is not primarily phylogenetic in that identification of shared homologous traits is only part of the method. This is because adaptive or relatively neutral, rare or unique, specialized, or otherwise divergent traits are also examined and evaluated in classical taxonomy to create predictive classifications. Suggestions are made here as to exactly how we use clues to direction of evolution in the taxonomic analytic and synthetic process. Prediction in evolutionary systematics involves placement in an evolutionary diagram that shows both shared ancestry and serial macroevolutionary derivation.

Second, definitions. Some terms appropriate for modern evolutionary systematics (Zander 2013) need a short explanation because of a different manner of use or because they are new.

- *Ancestor* in the present context is used for a taxon, as in "ancestral taxon," not an individual.
- *Bayes' Formula* is a simple statistical method of updating a previously accepted chance of something being true in light of additional information to provide a new ("posterior") probability of it being true. Sequential Bayes analysis simply uses a number of sets of data, one after the other, to continually update the degree of truth about something (such as a process in nature).
- *Clade* is a group consisting of an ancestor and all its descendants, and it is thus indicative of monophyly, but see Figs. 1, 2, 3 and 4.
- *Closed causal group* means that if one relationship is true between two elements, then the relationships of all the other elements are immediately deduced. That is, if one and only one species in a genus or infra group is determined to be the ancestral taxon for another in that

group, all the other species of the group must be also descendants, either immediate or secondarily, assuming again that this group has only one ancestral species.

- *Dissilient genus* refers to a group with a single generalized species with a cloud of clearly derived species associated with it. The derived species are more similar to the ancestral species than they are to each other. Some derived species (*stirps* sensu Zander 2013) may be so specialized as to appear to be dead ends in evolution, while others may prove advanced but generalist, capable of generating a number of specialized, derived species of their own and thus found a new genus.
- *Heterophyly* is either phylogenetic paraphyly or polyphyly, with the same taxon distant on a molecular cladogram and no other evidence of different origin or convergence. Heterophyly implies that intermediate cladogram nodes are of that same taxon.
- *Heuristic* is a short-cut or rule-of-thumb that provides an approximate answer sufficiently exact for everyday purposes.
- *Macroevolution* concerns taxa generating taxa, serially. This may be diagrammed with a *caulogram*, or stem-tree of both serial and branching relationships.
- *Superoptimization* is the process of intelligently assigning names to cladogram nodes — usually these are the names of exemplars or terminal taxa. Otherwise one may create a fully *natural key* (see example by Zander 2013: 80) involving such names, which may be holochotomous (serial/nested, one-branched), dichotomous, or polychotomous. Commonly, information that provides clues to direction of evolution is not phylogenetically informative (i.e., is not about shared ancestors).

Monophyly is determined by both shared ancestors and serial derivation of one taxon from another. Determination of monophyly through cladistic principles has been the object of thirty years of Hennigian phylogenetic analysis (Farris 2012; Felsenstein 2001; Pennisi 2003; Rieppel 2006; Vernon 1993; Williams 2012). Significant changes based on such have been made in classifications and in the way evolution is modeled. I have pointed out (Zander 2013) that phylogenetics cannot alone determine monophyly to any useful degree of accuracy. This is due to *apophenia*, seeing patterns in random data. In morphological cladograms, multiple descendants from one ancestral taxon may have some parallel traits, creating false synapomorphies, while reversals force a descendant lower in the cladogram. All this is due to the few expressed traits involved in speciation in any one part of a cladogram. In molecular studies, random survival of otherwise paraphyletic or phylogenetically polyphyletic molecular strains of the same ancestral taxon confounds interpretation of branch order of taxa. Non-phylogenetic information can correct phylogenetic apophenia to a large extent.

First, ask yourself if any of the taxa in a group qualify as ancestral to some or all of the rest of the taxa, ignoring the possibly misdirective cladogram. Divide your species into group of one potentially ancestral species and its associated derived species. This may seem foreign to students used to cladistic thought, i.e., "tree thinking," but one may imagine purposefully identifying a set of multifurcations. The Hennigian principle that of any three taxa of the same rank two are more closely related fails when the progenitor of both survives with expressed traits in stasis. A way to test this principle is to ask oneself if the group being studied is easily conceived as having one (or more) generalist species closely associated with a cloud of derived species more similar to the generalist species than to each other. If such multifurcations are seen as fundamental, then cladistic analysis is inappropriate, but evolutionary analysis remains possible. Relative stasis of the progenitor taxon is theoretically expected when the progenitor population is much larger than that of the descendants, in which case reduced rate of change by differential swamping of mutations can occur or there may be strong stabilizing selection (Haller & Hendry 2013; Pearman et al. 2007; but see Peterson et al. 1999). A recent, detailed, independent condemnation of Hennigian formalism was presented by Cavalier-Smith (2010).

Persistent molecular strains are implied by phylogenetic paraphyly. Molecular systematics assumes that molecular strains must quickly develop into new species when isolated. On the other hand, because non-coding and trivial genetic mutations occur and are fixed in both time and space, molecular strains are doubtless common. Isolated (in time or space or both) molecular strains of the same taxon may diverge with continued mutation of non-coding traits but without species level change in expressed traits. Abundant molecular paraphyly in cladograms of published phylogenetic papers demonstrates that surviving molecular strains of the same taxon may occur both before and after generation of one or more descendant species differing in expressed traits. Extinction of some molecular strains and survival of others in the same taxon implies that molecular cladogram nodes cannot be named with surety (Zander 2013: 51). In addition, there is lumping of taxa embedded in other taxa of the same rank in both morphological and molecular analyses, as "strict phylogenetic monophyly." Thus, monophyly is poorly discerned because branch order of taxa at any rank can be dubious. Although Mooi and Gill (2010) have contributed a recent, detailed, independent criticism of molecular systematics along similar lines, they make the mistake of assuming that "sequences of DNA and RNA are simply morphology writ small." The main problem is the false assignment of each molecular strain to separate taxa. There may be many parallel molecular strains somewhat distant on a cladogram because some have speciated, and many of these strains are extinct or unsampled.

Some phylogenetic methods are informative of evolutionary monophyly. Certain methods commonly used in phylogenetic analysis are acceptable as informative of serial ancestor-descendant relationships. Morphologically based cladistic analysis is a cluster analysis based on trait transformations, and as such has general utility, though limited by the stochastically based resolution of many groups of three or more species (or higher taxa) of which only one is surviving progenitor. Taxa in short or unitary lineages at the base of cladograms may be either advanced but with intermediate taxa extinct, or primitive (similar to ancient progenitors); but, if two or more such basal taxa are similar in morphology and also in different clades, their basal propinquity implies a primitive status relative to the taxa in the remainder of the cladogram (Zander 2013: 104, 165). Other than this information, a process of naming each node must be effected (parsimony through superoptimization) to collapse the OTU's into coherent progenitor-descendant groups (Zander 2013: 75).

Heterophyly in molecular cladograms is informative. Molecular analysis does reveal branching order of the molecular strains represented by exemplars because the molecular strains studied apparently do split in a dichotomous fashion and all are expected (or hoped) to have somewhat the same rates of mutation of tracking DNA bases. Because extinction or other non-sampling of molecular strains masks true progenitor-descendant relationships, the cladogram is restricted to branching order of the strains studied, which may grossly misrepresent species relationships. On the other hand, cladogram branches of strains of the same taxon that are distant on a cladogram do imply that the different taxa branching off between them are descendants of a deep progenitor of the taxon to which the distant strains belong. This heterophyly (paraphyly or phylogenetic polyphyly) is then informative of taxa that are in a serial ancestor-descendant relationship (Zander 2008) and do indicate what evolutionary direction (represented by changed expressed traits) that transformation took. A second value of molecular systematics is when two taxa are farther apart on a gene tree than expected by possible future informative heterophyly, such as strains of two species or two genera surprisingly occurring in two different families, at which time it may be concluded that there is no deep ancestral connection, and the two taxa are rightly separated. Using molecular heterophyly in determining order of serial taxic transformation is discussed in detail by Zander (2008, 2010a).

There are underlying bases for systematic analysis in addition to shared ancestry. Monophyly in classical systematics can be diagrammed as a caulogram (Besseyan cactus or caulogram) in which all parts of the evolutionary tree are named, if possible. The exception is when

two closely related taxa are each apparently characterized by equally advanced and specialized traits, and a shared ancestor may be postulated. The postulated shared ancestor minimizes credulity necessary for traits having apparent Dollo irreversibility as a group (i.e., macroevolutionarily, at the taxon level) (Atkinson et al. 2014; Gould 2002; Grant 1985: 329; Levinton 1988: 217). This paper is an attempt to formalize that process in classical systematics of intellectually and intuitively generating a caulogram from experiential data informed by process-based theory. Formalization to reveal underlying physical and statistical bases for systematic decision is important to justify credibility in scientific study using heuristics as opposed to the mechanical taxonomy associated with structuralism in phylogenetics (Zander 2010b). Of course, following Giere (2006), the "notions of reference and truth" developed for mathematics and physics may not be the only valid or practical guideposts to understanding a complex universe, but they suffice for this study.

Taxa evolve, not just traits. Many phylogenetic papers have deprecated standard evolutionary theory as contrary to cladistic results. Cladistic results are obtained by mapping of trait transformations on a morphological or molecular cladogram., that criticism is based the Hennigian fallacy, which is contrary to well-established theory (Bowler 1989: 346; Mayr 1981). This is in part because an algorithm expecting two of three taxa to be more closely related will in fact generate a fully resolved diagram by misinterpreting randomly congruent state changes as synapomorphies or a molecular strain as the complete taxon, when a multifurcation better represents macroevolutionary transformations. In fact, the opposite may be expected to be so in the majority of cases, that *immediate descendants, if two or more, will be more similar to their progenitor than to each other*. This last is a clear heuristic that I believe is much used in classical systematics. It is similar to the K statistic of Blomberg et al. (2003), but does not involve a measurement of phylogenetic signal. Sober (2008: 264) discussed at length the topic of shared ancestry, even invoking the Bayes formula to distinguish which hypothesis of Hennigian-style shared ancestry is more probable, but his rationale is throughout limited by a reliance on the Hennigian two-out-of-three principle.

Together with selected information from molecular and morphological cladistics, the modern heuristics of classical systematics can devise an acceptable caulogram that represents serial macroevolutionary transformations of monophyly. With formalization, the heuristics can be put on a mathematical and statistical methodological basis.

SERIAL MONOPHYLY VERSUS CLADES

Radiative evolution is the key to recognition of transformation of one taxon to another. When asking a classical taxonomist to estimate monophyly of a group in which he or she is expert, one can expect that taxonomist to sort and polarize the taxa into successive but also commonly branching groups each modified away from some central set of features identifiable as "general" or "primitive" for all taxa. Then each sub-group is evaluated as a kind of radiant circle away from a generalized ancestral taxon towards a set of often highly adaptive or at least neutral but unique descendants, representing radiative evolution into new environments, centrifugal from a generalized ancestral taxon. The result is a caulogram, or commagram, or Besseyan cactus. This sort of analysis is often done intuitively as a function of a mysterious facility of taxonomists called "expertise," not presently duplicable in software. Is there an intellectual structure to taxonomic expertise in determining serial monophyly? To what extent is experience a kind of frequentist statistic or generative of Bayesian expectation?

The classical systematist uses clues to both shared and serial relationships. Phylogenetics has formalized the grouping of taxa using clustering methods based on successive trait transformations, resulting in dichotomous trees of terminal taxa quite like standard cluster analysis but with more information beyond raw or massaged similarity. Each node in a cladogram represents the beginning of a supposed monophyletic group, called a *clade*. Tree-thinking methods are criticized

at length in "Framework for Post-Phylogenetic Systematics" (Zander 2013), and a caulistic (stem-thinking) alternative to cladistics is there proposed. The use of heuristics in classical systematics was treated (Zander 2013), but only one aspect was *formalized* (i.e., given a clear physical or mathematical structure or explanation), namely, the geometric mean basis for the paradigm (a–)b–c(–d) in descriptive measurements.

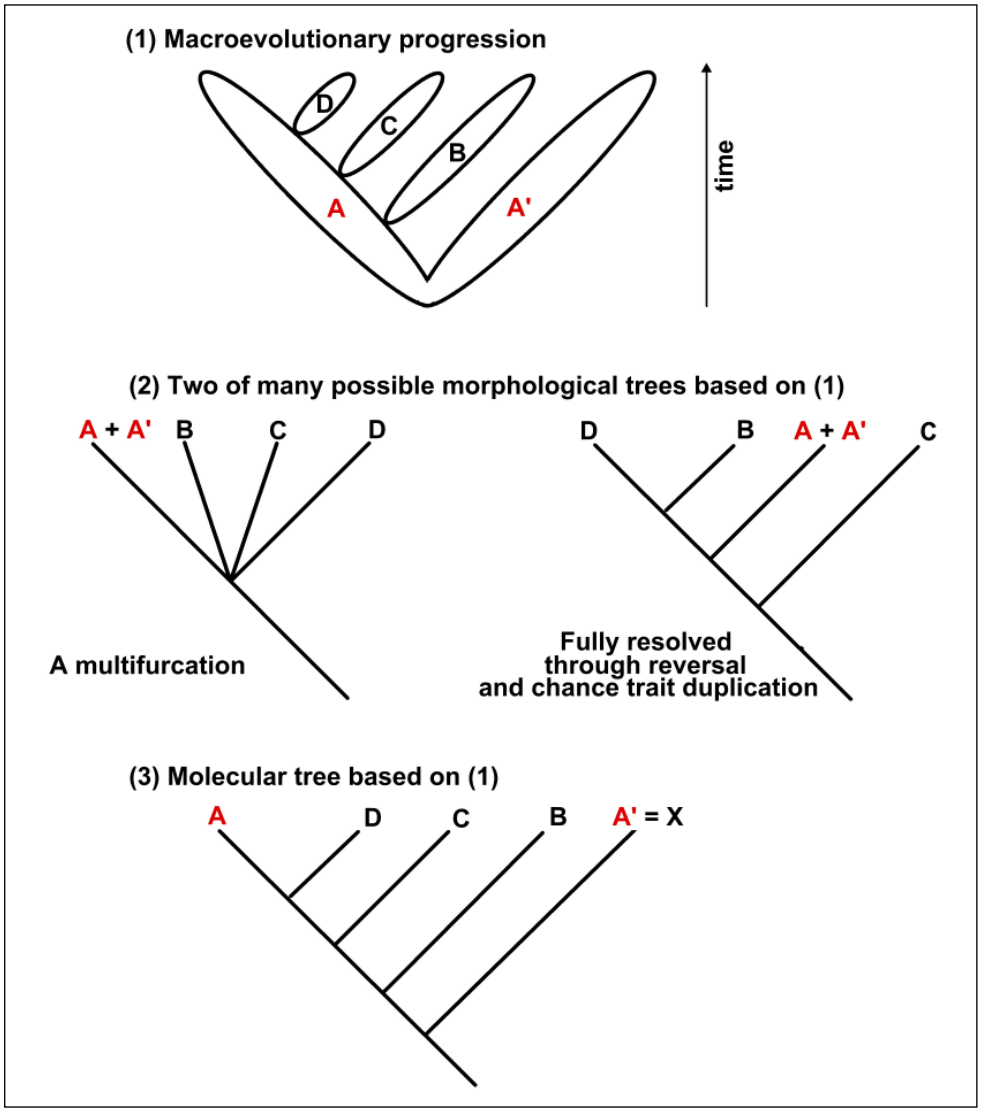

Figure 1. Comparison of contrived trees of the same evolutionary scenario. **(1)** Macroevolutionary progression of three derived species B, C and D, in that order, from species A. Species A has previously split into two isolated but morphologically static and identical populations, A and A′. **(2)** Cladograms of parsimonious analysis of morphological traits. Left is a multifurcation. Right is a fully resolved morphological cladogram with chance duplication of traits in B and D, and reversal of a trait in C. **(3)** Molecular tree showing A as terminal having generated B, C and D in the past while itself mutating but A′ is treated as a new cryptic species "x".

Cladistic analysis is accepted as valuable for preliminary clustering of taxa, but it must be carefully evaluated because the central Hennigian thesis that of every three taxa two *must* be more closely related can be quite wrong for estimated order of branch splitting. In the present paper, the ability of classical systematists to evaluate evolutionary monophyly intuitively is examined and

formalized as actually a combination of cladistic-style evaluation of shared conservative tracking traits and a simple sequential Bayes analysis (explained by Kachiashvili 2012) done through assignment of coarse likelihoods as clues in the manner of World War Two code breakers.

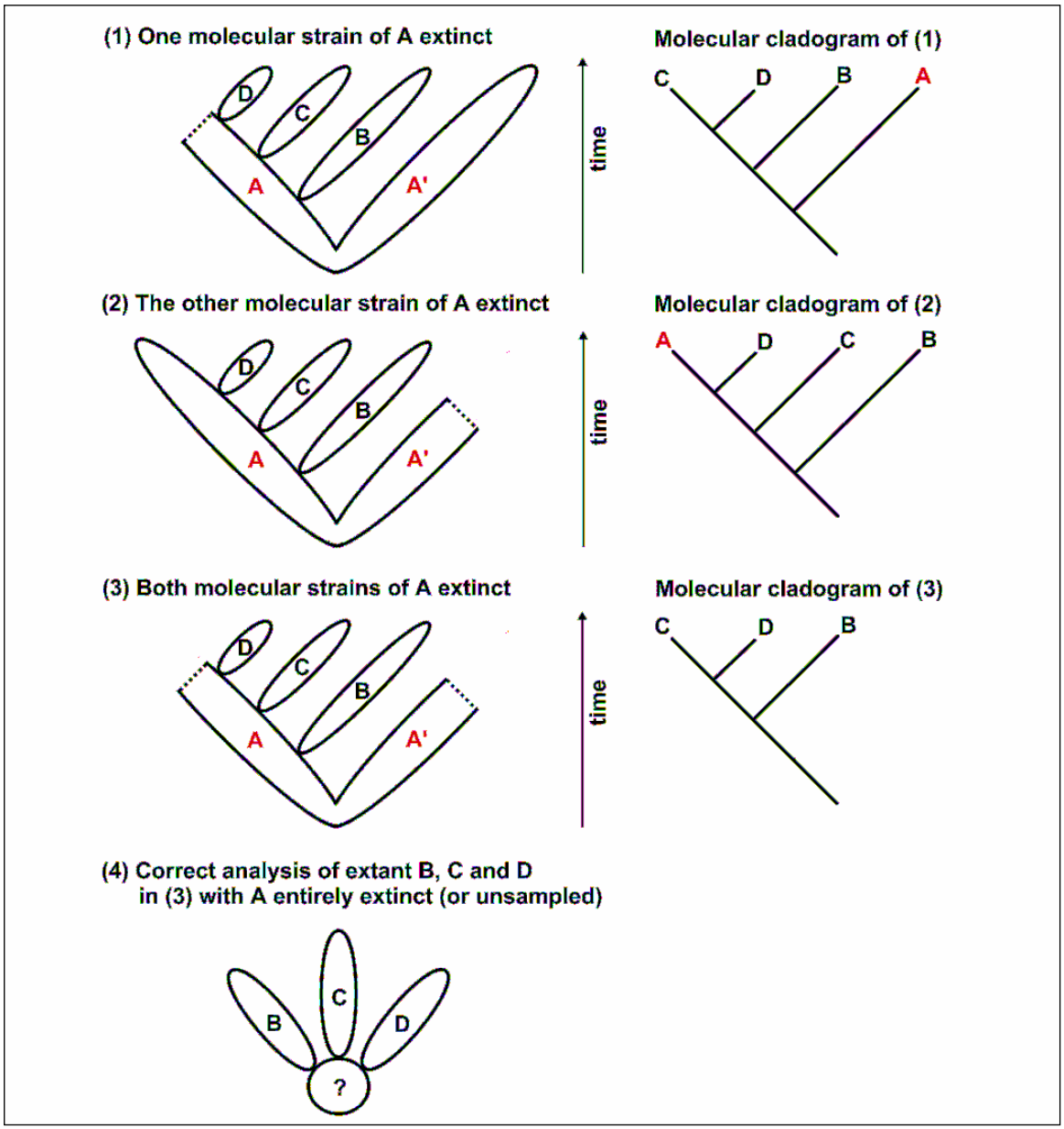

Figure 2. Effect of extinction (or non-sampling) of molecular strains of ancestral species. **(1)** A (one of the ancestral molecular strains) is missing, molecular cladogram at right has species A basal. **(2)** A′ (the other ancestral strain) is missing, species A is terminal. **(3)** Both molecular strains of species A are missing, and molecular cladogram is restricted to B, D, and D. **(4)** This is the correct caulogram for the extinct ancestor and its descendants (3).

ILLUSTRATED COMPARISON OF CAULOGRAMS AND CLADOGRAMS

It is untrue that two of every three taxa must be more closely related. After thirty years of viewing evolutionary relationships diagrammed with cladograms, the reader may find difficulty assimilating a way of presenting both serial and lateral evolutionary relationships, the caulogram (also known as the Besseyan or Bessey's cactus). The restriction of cladograms to showing shared

ancestry alone can bias the presentation of information in various ways. Such differences need explication, which is given here in a series of illustrations (Figs. 1–4).

Figure 1 is an exemplary analysis comparing various contrived trees of the same evolutionary scenario.

Figure 1(1): macroevolutionary progression of three derived species B, C and D, in that order, from species A. The last previously split into two isolated but morphologically static populations, A and A′. Note the time bar on the right. This is a caulogram.

Figure 1(2): shows cladograms of parsimonious analysis of morphological traits. Left is a multifurcation as expected if traits of derived species did not reverse or duplicate. A and A′ are correctly treated as the same. On right is a fully resolved morphological cladogram with chance duplication of traits in B and D, and reversal of a trait in C. All three species, B, C and D, remain derived from A, however, no matter where on the cladogram they appear.

Figure 1(3): presents a molecular tree showing A as terminal having generated B, C and D in the past while itself mutating (a "self-nesting tree"), but A′ is treated by molecular phylogeneticists as a new cryptic species "x". Although (2) and (3) show cladistic (lateral) relationships they are wrong because the serial relationship is ignored. Note no time bars for (2) and (3). The macroevolutionary formula for all three, (1), (2) and (3), is (**A**, A′) > ^{1}B, ^{2}C, ^{3}D.

Figure 2 shows effect of extinction (or non-sampling) of molecular strains of ancestral species.

Figure 2(1): A (an ancestral molecular strain) to be missing, and molecular cladogram at left has species A basal.

Figure 2(2): A′ (non-ancestral strain) is missing, species A is terminal. Given that only one molecular strain of species A is known, the macroevolutionary formula is **A** > ^{1}B, ^{2}C, ^{3}D for both (1) and (2) yet the molecular cladograms are not congruent.

Figure 2(3): Both molecular strains of species A are missing, and the molecular cladogram is restricted to B, D, and D. It is fully resolved although from other information B, C and D are apparently equally derived from some unknown ancestral species.

Figure 2(4): This is the correct caulogram for (3). Only when there is no known candidate ancestral species for two or more equally derived species can an unknown shared ancestor be postulated.

Figure 3 asks the question that given that we do not know the true macroevolutionary relationships, what is the best we can do in determining branch order? This figure shows what we can infer from minimal data on a molecular tree. Remember that the ancestral nature of A and derived natures of B and C are determined in large part in superoptimization by non-phylogenetic (non-shared ancestry) information. Even this basic information, however, is not available from the phylogenetic molecular analysis of *Didymodon* s.lat. by Werner et al. (2005) because the segregate genera are scattered and in very short branches (Zander 2013: 89–90). But if such information were available for the contrived example in Fig. 3, then:

Figure 3(1): Heterophyly implies that B and C are derived from species A, and the more terminal of the two is last in order of speciation, e.g. **(A**, A′) > ^{1}C, ^{2}B.

Figure 3(2): If a species judged ancestral by non-phylogenetic information is terminal, species clearly derived from it are in order, e.g., **A** > ^{1}C, ^{2}B.

Figure 3(3): If only two derived species are more terminal than the ancestral species, no order is discoverable because they are presented as sister groups. Thus, **A** > B, C.

Figure 3(4): But if three or more derived species are more terminal, the lowermost are in discernable order, e.g., **A** > ^{1}D, ^{2}E, ^{3}F, ^{4}(C, D).

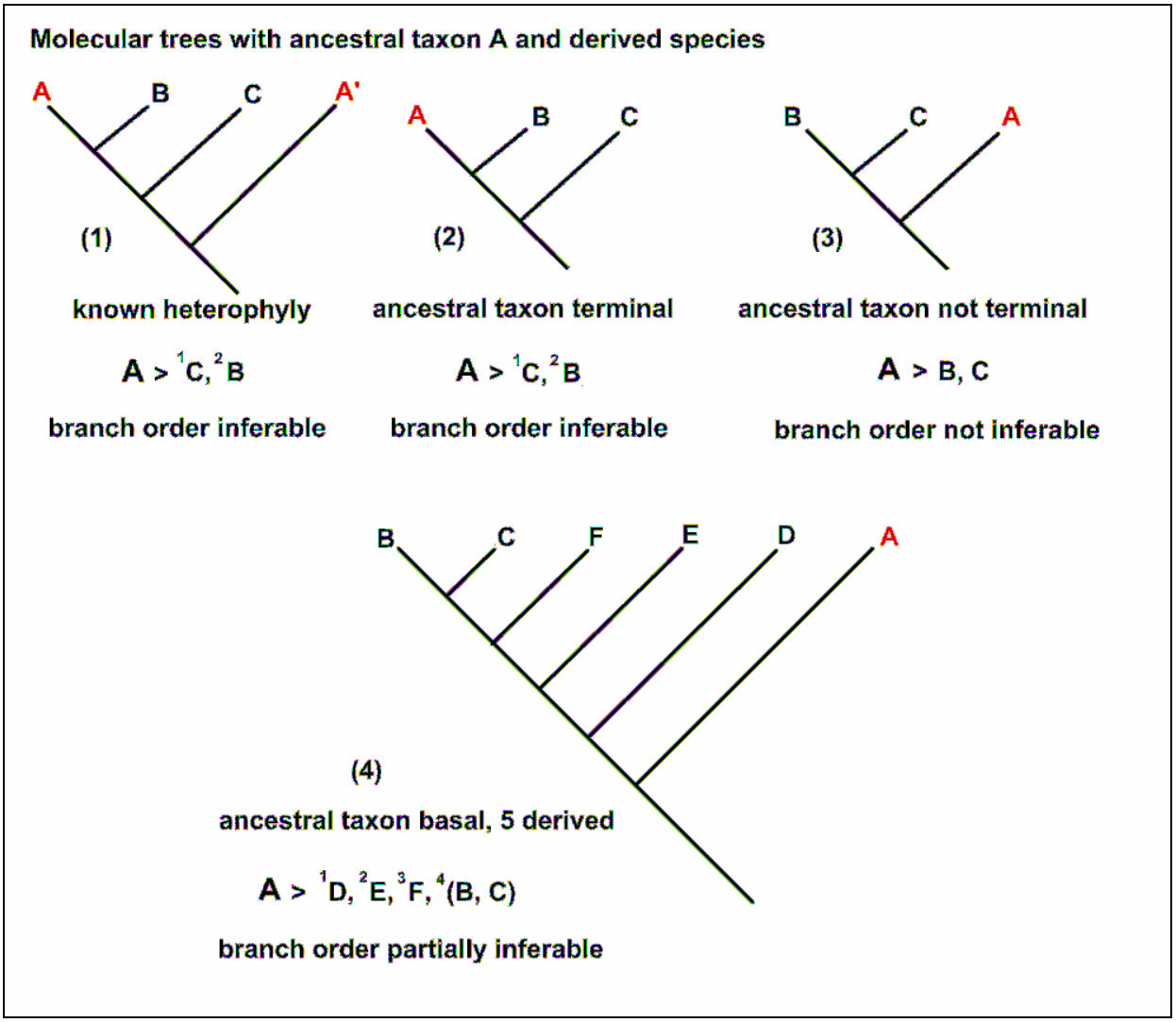

Figure 3. Inference from minimal data on a molecular tree. **(1)** Heterophyly implies that B and C are derived from species A, and the more terminal of the two is last in order of speciation, e.g. **(A**, A′) > ^{1}C, ^{2}B. **(2)** If a species judged ancestral by non-phylogenetic information is terminal, species clearly derived from it by non-phylogenetic information are in order. **(3)** If only two derived species are more terminal than the ancestral species, no order is discoverable because they are presented as sister groups. **(4)** But if three or more clearly derived species are more terminal, the lowermost are in discernable order.

Figure 4 explains how the principle of strict phylogenetic monophyly does not reflect generation of a taxon of higher rank from another. Genusation, for example, is generation of a new genus from a species in another genus, or just from the other genus if which species is ancestral cannot be readily determined. Can you see the two genera in Fig. 4? Both species A and B generate derived species based on superoptimization information evaluating ancestral and derived species. B, like A, is identifiable as central to a particular dissilient (exploding) genus concept. The macroevolutionary formula of this caulogram is (**A**, A′) > ^{1}(**B** > ^{1}E, ^{2}F, ^{3}G), ^{2}C, ^{3}D. In this contrived example, we know both serial evolutionary direction and the branch order. In practice, the formula is usually incomplete.

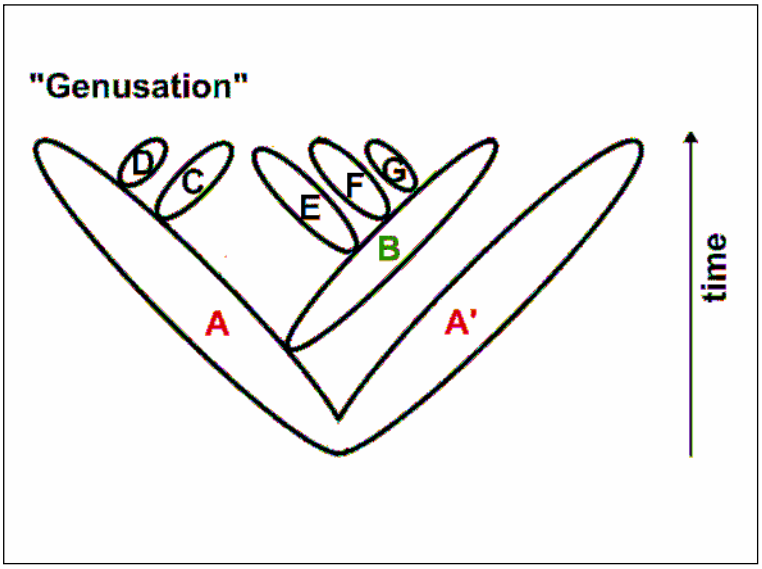

Figure 4. Genusation is generation of a new genus from a species in another genus. Can you see the two genera in Fig. 4? B, like A, is identifiable as a species central to a particular dissilient (exploding) genus concept.

CONVERGENCE ANALYSIS

Evolutionary stasis of ancestral taxa does not mean they do not speciate. That progenitor-descendant series exist and are often abundant is demonstrable in the evolutionary literature, and in the associated phenomenon of phylogenetic paraphyly in both morphological and molecular analyses. If any population is split and isolated in two or more unequal parts (e.g., peripatric speciation, see Futuyma 2009: 484), including founder events, genetic stability through time is expected to be greater in the larger portion. It is commonly acknowledged that allopatric speciation (Futuyma 2009: 472) is quite common or even more common than sympatric (Barraclough & Nee 2001; Mayr 1954, 2001). Species may remain static for millions of years, whether of large distribution or not, but smaller or founder isolates (including sympatric isolates) may speciate rapidly, escaping the homogenizing effects of gene flow, and more rapidly drifting in traits or changing through selection (Via 2001). There is great evidence from fossil studies that stasis in expressed traits (Haller & Hendry 2013) associated with punctuated evolution is common (Benton & Pearson 2001). Surviving progenitor species with two or more immediate descendant species are to be expected.

Recognition of the difference between generalized ancestral taxa and specialized descendant taxa is often easy for an expert in the group. The phylogenetically assumed pseudoextinction (rapid post-speciation anagenetic change on the part of a progenitor species) is thus theoretically uncommon. Figure 2(4) shows similar branching from an unknown shared ancestor but the ancestor may be pseudoextinct or simply unsampled. For evolutionary science to advance, theory

is used to create models to explain process-wise the relationships of organisms. By evaluating sets of closely related species, experienced taxonomists can usually identify a surviving progenitor from its surviving descendants by reference to rules of thumb (heuristics) that are widely accepted. *Convergence analysis* is simply the heuristic that if a trait (or set of traits) is scattered about an accepted classification, these are convergent (or parallel if from the same ancestral taxon), and are therefore may be either radiatively adaptive or are at least an element in transformation away from a generalized species.

There are well-established clues to direction of evolutionary radiation. Progenitors are taken here to generally have comparatively broad distributions, occur in older habitats, occupy less specialized niches, are morphologically generalized, have all expressed features (are not much reduced), are polymorphic with many subspecies, varieties, biotypes or cytotypes, have a distinctive morphological trait combination that may be variously modified or reduced, and lack asexual reproduction as primary. In the present paper the likelihoods are clues about direction of evolutionary transformation along the lines delineated by Grant (1949), Simpson (1953) and Mayr (1954) in the context of the New Synthesis, and of others who have discussed adaptive trends and orientations in detail (e.g., Futuyma 2009: 595; Gavrilets & Vose 2005; Seehausen 2006). According to Schluter (2000: 2), "Adaptive radiation is the evolution of ecological and phenotypic diversity within a rapidly multiplying lineage. It occurs when a single ancestor diverges into a host of species that use a variety of environments and that differ in traits used to exploit those environments." It is possible that some new traits associated with radiation are not adaptive (Gittenberger 1991; Gould & Lewontin 1979; Rundel & Price 2009), so one should identify putative descendants as simply transformative radiation, including both changes associated with adaptive radiation and new neutral traits whether refractory to selective pressures or not.

Convergence is a clue to adaptability of traits. Neutral morphological traits are only neutral as evolutionarily "local" conservative traits. According to Simpson (1953: 174, 179) conservative traits almost always are have adaptive significance for higher categories. Adaptive traits, if identifiable, might thus be informative of direction of macroevolution. They converge across taxonomic boundaries as different taxa adapt to the same evolutionary pressures. Conservative traits do not converge except at high taxonomic levels when associated with phyletic constraint, and are not immediately informative at the species level. Convergence analysis distinguishes adaptive from conservative traits and weights adaptive traits by level of confidence in distinguishing direction of evolution.

There are two principles of convergence analysis in the present paper.

First is that **for any closely related group, taxa with advanced specialized traits have a more generalized taxon as shared ancestor**. Thus, a possible dead end will have an entrance somewhere. Such a generalized ancestral taxon may be extinct or simply unknown but it can be "described" (and searched for) as at least having the common traits of the remaining species. Given this method of analysis, monophyly with named or at least describable ancestral taxa is possible. This can replace the "clade" of phylogenetics in which every node is necessarily an unnamed and unnamable ancestor of ultimately all distal branches, which is palpably incorrect because many nodes are easily named and many of these are of the same taxon (Zander 2008).

Second, **post-dissilience descendant-descendant transformation series with no reversals are preferred**. That is, after an ancestral taxon has generated a set of descendants, those descendants may be expected to telescope outwards (nest) by generating descendants of their own with increasing specializations. Sometimes, a new generalized taxon of higher rank may be a descendant (Fig. 4). Takhtajan (1997: 4) emphasizes this point: "Every new stage of evolution, and consequently every new taxon, differs from the ancestral taxon by an acquisition of some new, derived characters. The

ancestral taxon, on the other hand, will differ from its descendants by the absence of these derived characters." The simplest case is when no reversals are required and the putative secondary descendant is simply a sub-set of its immediate secondarily progenitor species (see discussion of *Vinealobryum brachyphyllum* and *V. nevadense*, and of *Geheebia fallax, G. ferruginea, G. maxima*, and *G. gigantea*, below).

Conservative traits may be weighted by the number of different habitats they tolerate. The transformation from one genus to another theoretically involves fixation of a *set* of different conservative traits in the progenitor that occur in a multiplicity of somewhat different environments. Phyletic constraint (restriction of a taxon to environments survivable with that set of conservative traits) both restrains the increase in number of conservative traits and the survival of organisms without the full complement. Thus, existence of at least one generalized species in a genus demonstrates long-term (multiple speciation events) survival of that conservative trait combination in the most ideal of environments. Conservative traits, as is the case with non-coding or genetically trivial DNA bases, can thus be used to track evolution.

Conservative traits are those refractory to selection because they are in combination evolutionarily neutral or neutral enough for that group in its particular range of environments given phylogenetic constraint of other expressed traits. They are identified as those tolerating a range of selective regimes. Such can be judged by the number of species of a group in which the traits occur, given that each species has or probably has a distinctive adaptive range. Thus, cladistic logic works fine to cluster conservative traits that track evolution, yet fails when highly adaptive traits are automatically assigned synapomorphy status simply because they appear in two or more species that morphological and ecological evaluation may deem more probably all joint descendants of one other generalized species. The present analysis began with a cladistic (Zander 1998, 2001) study of *Didymodon* Hedw., a genus of mosses (Bryophyte, Pottiaceae) that used equal weighting of traits but happened to matched past classical groupings (Zander 1993).

Adaptive or transformative radiation is considered a standard view of macroevolutionary change, is implicit in classical systematics (e.g., generation of a caulogram or Besseyan cactus), and comprises the data used in classical heuristics relating to monophyly. This paper studies the backbone of those heuristics, namely the mathematical and statistical structure that allows evaluation of the significance of the data.

The equivalent of Hennigian pseudoextinction is possible, but probably rare. Dichogamy (equal splitting and isolation of a progenitor population) may give rise to two descendants that gradually diverge. This is equivalent to Hennigian pseudoextinction. Another equivalent is extinction of ancestral species or those of intermediate morphology such that two species are so different but equally specialized as to be equivocal in estimation of a serial transformation series. Both scenarios must be included in estimation of progenitor-descendant series and signaled as involving an unknown shared ancestor, but there should be a justification based on process-based evolutionary theory, not an axiom for reliance on cladistic splits.

DETECTIVE WORK IN CRYPTANALYSIS

Sequential Bayesian analysis is a powerful tool. This paper makes use of sequential Bayesian analysis. The essence of Bayesian analysis is to combine a prior (chance of the hypothesis given prior knowledge) with a likelihood (chance of the data given the hypothesis) to calculate with the Bayes Formula a posterior probability (chance of the hypothesis given the data and the prior). Details of Bayesian analysis as used in phylogenetic analysis is presented well by Sinsheimer et al. (2003) and by others, and will not be discussed here. Kruschke (2011) has recently produced a

software-oriented (R and BUGS) manual for Bayesian analysis that is potentially highly flexible although the de novo programming of a MrBayes equivalent would be daunting and superfluous.

Sequential Bayesian analysis was developed by A. Wald and others (Kachiashvili 2012; Wald 1947), and was kept secret by the United States because of its value during World War 2. About the same time, it was separately developed in Britain by Alan Turing, in a somewhat different form. Again, applications to the British war effort and later the Cold War kept it secret until about 1980 (Good 1979). It is a form of empiric Bayes (McGrayne 2011: 134, 168, 205)—the posterior probability that was obtained from the first calculation is used as the prior for a second Bayesian calculation with another additional likelihood from additional data, and so on with added data until a stopping rule is triggered or one runs out of data. A formal equation for Bayesian causal induction involving a stopping rule is given by Bonawitz et al. (2013), but amount of data, in the present study, is the limiting feature. The assumption is that the shape of the distribution curve for the data of each sequential implementation is the same, i.e., conjugate priors (McGrayne 2011: 149).

Sequential Bayesian analysis is increasingly used with sequential sampling, but it may also be used when dealing with individual "particles" of information. When the data change with time, this quasi-recursive method is known as Sequential Bayesian Updating (Lauritzen 2009) and is used for control in robotics, speech recognition, political polling, target tracking, and steering/control, for example of large ships, airplanes, and space ships. In taxonomy, it has been examined as a method for identification of bacteria (Gyllenberg & Koski 2002), but that paper is largely of mathematical proofs of certain very general assumptions. The heuristic use of sequential Bayesan analysis (as "updating") in day-to-day human affairs was investigated by Bonawitz et al. (2013). They found that a simple Win-Stay, Lose-Shift sampling algorithm, in which a learner keeps a particular hypothesis until receiving evidence that is inconsistent with the hypothesis, approximates Bayesian inference, and does so efficiently. Sequential Bayesian analysis in the present paper is suggested as a formal, previously unrecognized basis for heuristic evaluation of monophyly in classical systematics.

Adding decibans together may be used as a substitute for using Bayes' formula. Alan Turing's work in breaking German war codes during the 1940's (McGrayne 2014: 67) led to his use of a kind of sequential Bayesian analysis. Given that computers were then primitive, being hand-operated, logarithms were extensively relied on. Turing, with I. J. Good and others in the group of code breakers at Bletchley Park, used clues, often tiny clues, to narrow down particular settings of the Enigma machines the Germans used. Statistically, the unit they used was the ban, which indicates that one hypothesis is 10 times as likely as an alternative hypothesis. The basic unit for a clue was the deciban (abbreviated dB), defined casually as the minimal level quantifiable as a measure of belief in a hypothesis, somewhat more precisely as an change in odds ratio from 1:1 to about 5:4. (Remember that an odds ratio of, say, 2:1 is actually the fraction 2/3, where the denominator must be increased by the value of the numerator. The odds ratio of 1:1 is 1/2, and 5:4 is 5/9.)

A deciban is technically 10 times the base 10 log of the odds, or $10^{n/10}$:1, or a ratio of tenths of a power of 10 to one. It is a logarithmic unit of probability that measures information (or entropy). It is a decimal digit as opposed to a bit, which is a binary digit. One ban corresponds to about 3.32 bits, and a deciban is about 0.33 bits. A change of 1 deciban changes the odds by a factor of approximately 5:4. A change of 10 decibans changes the odds by a factor of 10, 20 decibans changes the odds by a factor of 100. Most systematic analysis is restricted to a range of 1 to 20 decibans (0.55 to 0.99 probability values), whether given as exact or informal probabilities.

Turing and his group found that by combining clues (some in hundredths of a deciban) enough relevant information could be gathered together to break codes. The process was essentially Bayesian, and can be easily matched using today's computational conveniences (e.g., a spreadsheet as

discussed below) with sequential Bayes calculation. In fact, the probabilities associated with decibans are exactly duplicated by sequential Bayes analysis. To review the equivalencies, an *odds ratio* of approximately 5:4 is equal to the *fraction* 5/(4+5), and that is the decimal fraction 0.05555..., an approximation of one deciban. The exact deciban calculation of $10^{n/10}$:1 as odds ratio is equal to the fraction $10^{1/10}/(10^{1/10}+1)$, which is the decimal fraction 0.5573.... With a prior of 0.50 and a likelihood of $10^{1/10}/(10^{1/10}+1)$, the Bayesian posterior probability is the exact same decimal fraction 0.5573.... If one does have posterior probabilities that include priors that are not 0.50, then convert by multiplying by the reciprocal of the priors (Kruschke 2011: 253), but this is not necessary with deciban calculations.

Using decibans does not require computers. Two and five decibans add (logarithmically, as with a slide rule) to seven decibans, and from the formula $10^{7/10}/(10^{7/10}+1)$ we get the decimal probability 0.8337...., which can also be read off a chart (see Table 1). Bayesian sequential analysis with a spreadsheet yields the same results, with more complex but more flexible calculation. With a prior of 0.50 and seven likelihoods of $10^{1/10}/(10^{1/10}+1)$ each of the seven being analyzed with sequential Bayes (using each posterior as the prior of the next calculation), the same decimal 0.8337 is also obtained. Clearly, using decibans is a short cut in Bayesian sequential analysis and can have heuristic value as a simplifying tool.

CLADOGRAM ERROR

The resolution of molecular trees of branch order of taxa is not high even if branch order of molecular strains is well determined. In phylogenetics, each node defines the start of a clade. Yet in morphological cladograms, chance matching of new traits during parallel speciation from one ancestral taxon results in false synapomorphies, thus false nodes. In molecular cladograms, the potential for hidden paraphyly (or heterophyly) caused by extinct or unsampled molecular strains, and for generation of multiple descendants from one ancestral taxon makes any node uncertain as to whether or not it is the beginning of a clade.

If any molecular analysis is liable to uncertainty because of the above potential problems, one cannot use details of the analysis for classification purposes. The branch order resolution of a molecular cladogram cannot be better than the average distance of known heterophyly. One might expect an average resolution of at least three or four nodes for widespread taxa and often up to 10 nodes for certain taxa (e.g., *Brachyglottis*, *Ligularia* and *Senecio* in Senecioneae, Asteroideae, Pelser et al. 2007). This applies to any rank exhibiting paraphyly, species, genus, or family. Of course there is a limit to uncertainty due to expectation of hidden heterophyly, since one might not expect it to cross established higher ranks.

If known molecular heterophyly is largely, say, two nodes, as modeled in Figure 5, then an error bar showing this uncertainty might be inserted for each phylogenetically postulated "shared ancestor." All postulated shared ancestors are then affected by overlapping error bars, see Fig. 5(1). Molecular heterophyly as in Fig. 5(2) and superoptimization as in Fig. 5(3) largely eliminate uncertainty due to paraphyly. Note in Fig. 5(3), even though nodes with known ancestral taxa crowd the end of the cladogram, the error bars remain and represent predictive uncertainty for any new taxa that might be inserted into the cladogram. Such a problem is obviated if new taxa are inserted into the equivalent caulogram of Fig. 5(4), which is why caulograms are ultimately better than cladograms.

Why such an involved and complicated introduction? Lakatos (1978) proposed that research papers do not need a justification of their theoretical basis for each publication in those cases when a firmly established intra-disciplinary research program is understood. Even if the researcher is not fully familiar with every theoretical nuance, a paper observing standard protocols indeed contributes to science. In systematics, for instance, a paper that is simply a check list of species for

an area is accepted as part of a 250-year research program documenting and explaining when possible the historical biogeography of the earth's life. The present taxonomic paper requires a detailed theoretical justification, however, because the appropriate research program is new, and is, in Lakatos' sense, progressive in predicting novel facts. Besides, although the human brain is slow to work out complex problems, its own complexity and power can deal with that complexity, slowly but surely.

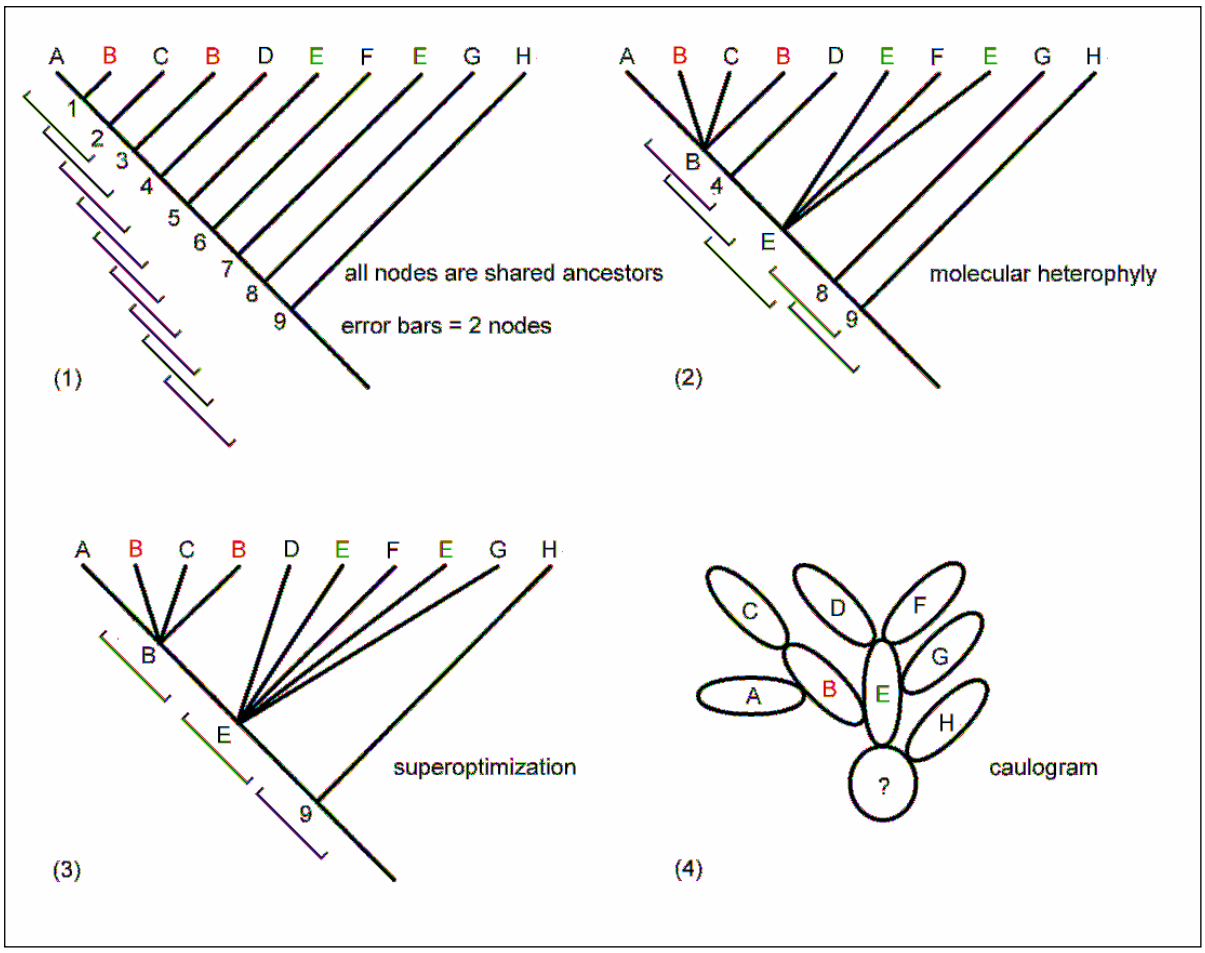

Figure 5. Analysis leading to a caulogram, based on a terminal portion of a molecular cladogram. Heterophyletic terminal exemplars (of species B and E) are in color, they are of the same taxon but each distant by two nodes. (1) Overlapping error bars for all nodes when all are each treated as a shared ancestor. (2) Molecular heterophyly eliminates much uncertainty by identifying one taxon giving rise to another. (3) Superoptimization through identification of ancestor-descendant relationships on the basis of non-phylogenetic information reduces uncertainty more. Node 9 in this contrived example cannot be eliminated because exemplar H cannot be easily assigned in this example to an extant group, and remains an unknown shared ancestor. (4) Caulogram showing stem relationships.

METHODS

PROBABILITIES AND SUPEROPTIMIZATION

Simplifying formal estimation of probabilities uses decibans. Probability is here not just expected intersubjective agreement, but also a measure of how much more predictive or explanatory one model is over another. Informally, the probability imbued in a scientific hypothesis must be tested by real decisions and their aftermath. Classifications are the result of series of heuristic

decisions over time (250 years of Linnaean and 150 of Darwinian taxonomy), much like updating in sequential analysis. Formalization involves assigning probabilities as measures of expectation on the basis of theory; these are coarse measures, but nonetheless not "mere" intuition. There are four ways probability is used in this paper: (1) Sequential Bayes analysis of "clues" to direction of serial transformation using decibans and perceived radiative transformations is done in much the same spirit as by code breakers during the Second World War. (2) Bayes factors are used to evaluate competing hypotheses. (3) Probability that at least one ancestor-descendant transformation is correct (IRCI formula below) uses the concept of a closed causal pool. (4) Deciban differentials allow distinction of the two most likely models when the B.F. of each are very close. The calculations in this paper use simple mathematical concepts suitable for traditionally innumerate taxonomists, and are facilitated by spreadsheets available online at <http://www.mobot.org/plantscience/resbot/evsy/sprsh/>. The spreadsheets can be "unprotected" and modified for use with larger numbers of traits and taxa.

Intuitive expert systems can be explained. The analysis of serial macroevolutionary transformations at the taxon level was discussed at length by Zander (2013). In the present paper, an expert system is exemplified that attempts to formalize (identify physical and mathematical bases of) the scientific intuition approach referred to in the past as Gestalt or omnispection methods. Two data sets are gathered for the group studied, one set for shared traits published by Zander (1998), and another set (see tables in Part 3) for unique and apparently advanced traits. The first set is of homologous traits that may be used in cladistic analysis, the second largely of autapomorphies.

"Superoptimization" means naming cladogram nodes to eliminate invention of unknown shared ancestors. The deeper a taxon is buried in a rooted cladogram (subtended by many nodes) the more likely it is to be advanced in terms of serial transformation, but this is usually masked to a great extent by false resolution due to a methodological requirement that of every three taxa, two are more closely related, resulting in branch order based on chance (parallel) shared traits or reversed traits. This problem is resolved by "superoptimization" (Zander 2013: 75), which is the naming, whenever possible of cladogram nodes. This results in identification of groups of one progenitor taxon and one or more immediate or secondarily descendant taxa. This is usually done informally in classical taxonomy, through omnispection and reliance of a set of informal heuristics that identify primitive-advanced transformations along the lines of evolutionary theory.

This has already been done for *Didymodon* by Zander (2013: 80). That same analysis is continued in this paper but with formalization of the heuristic used in superoptimization. The intuitively superoptimized groups are here re-analyzed by assigning each taxon a set of clues or items of evidence. A dissilient genus (Zander 2013: 83, 92) is often easily identified as a group of similar species with a putative ancestral taxon for the other species. Inasmuch as nature teaches us taxonomic concepts, there may be other definitions of taxonomic groups that are equally effective in prediction when dissilience is not evident. The putative progenitor has a maximum of theoretically primitive (i.e., first of a series) traits vis-à-vis those of the other taxa in the group.

One can assign one deciban as minimal clue, or a higher number of decibans for very convincing clues. With sequential Bayes analysis, with all evidence treated as minimal clues (0.56 probability) and assuming an initial 0.50 prior, 13 Bayesian operations (13 clues, that is, 13 likelihoods) are needed to provide minimal scientifically reliable support (0.95 or more). A single 0.76 (5 dB) likelihood among the sequence reduces the number of clues needed to eight for scientific reliability. Two 0.76 likelihoods reduces the number of clues needed to four. Thus, moderately strong evidence, if convincing, can be quite helpful in supporting a particular hypothesis. With strong evidence, three Bayesian operations at 0.76 (5 dB) likelihood surpass a standard scientific minimum 0.95 at 0.97; four at 0.76 likelihood gives 0.99. Thus, three clues determined to be of moderate not minimal import (0.76) can combine in sequential Bayes analysis (with an initial 0.50 prior) to yield a

scientific minimum reliability for whatever hypothesis is examined, four giving very strong scientific support. Coarseness in assignment of clues to direction of macroevolutionary transformation aids in making these studies repeatable, using, say, only odd numbers of decibans (1, 3, 5, 7) as is done here.

Deciban analysis is like using a slide rule. Table 1 compares the probabilities of evolutionary serial transformation using decibans obtained from numbers of perceived advanced transformative traits. The probabilities are given from 26 to −26 dB because negative decibans are required when evaluating pro and con hypotheses; zero decibans is 0.50 probability. Figure 6 presents a chart of the exact probabilities on the y-axis, and decibans on the (logarithmic) x-axis. A dashed horizontal line is given for 0.95, 0.99, 0.76, 0.56 and 0.50 probabilities, showing their position on the asymptote. Their negative values are also given for dB less than zero (i.e., less than 0.50). One can use this chart for quick estimation of probabilities. Three clues in a sequential Bayesian analysis can be read off the 3 dB bar. Two clues of one deciban each plus one element of moderate support of five decibans adds to seven decibans, or 0.833 posterior probability for that single analysis of the probability that a taxon was derived from another. Note that this method is similar to the use of the logarithmic scales on an analog slide rule. Dealing with complex digital calculation through the mental analogue of a specialized slide rule may partially constitute "intuition" in systematics.

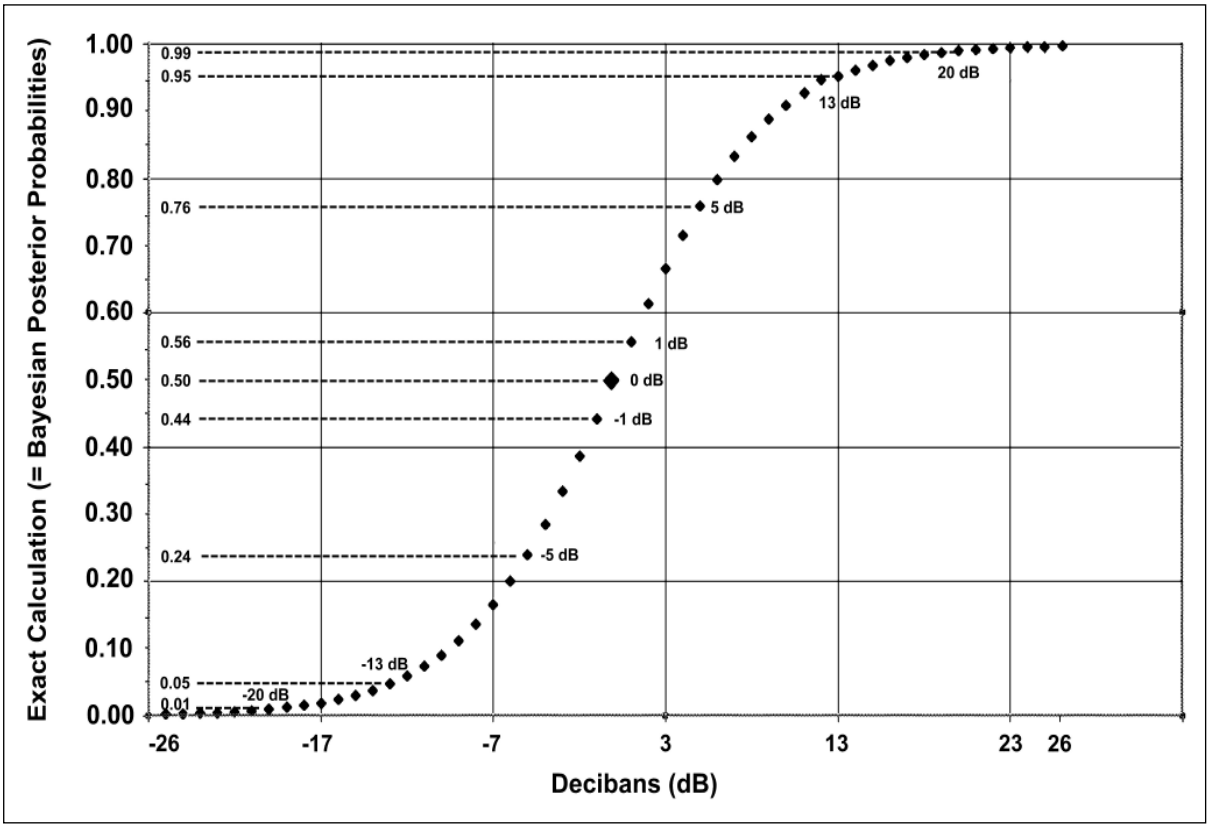

Figure 6. Chart of probabilities showing decibans on the x-axis (logarithmic) and exact probabilities on y-axis. After assigning clues as decibans and adding them, the exact probabilities of this form of sequential Bayes analysis can be read off the y-axis, in the fashion of a slide rule.

Table 1. Decibans (dB) with calculations of probabilities by odds of $10^{n/10}$:1, which is the fraction $10^{n/10}$ divided by $10^{n/10} + 1$. Given that decibans are logarithmic, they may be added, e.g., 2 dB plus 3 dB = 5 dB. Deciban and their exactly equivalent sequential Bayesian probabilities may be read off this table. Negative dBs are for contrary hypotheses. Zero dB = 0.500 probability. Approximate odds ratios are given. Important mileposts are bold-faced. First standard deviation = 3 dB; second S.D. = 13 dB, third S.D. = 26 dB.

dB	**1**	2	3	4	**5**	6	7	8	9	**10**	11	12	**13**
Prob.	**0.557**	0.613	0.666	0.715	**0.759**	0.799	0.833	0.863	0.888	**0.909**	0.926	0.940	**0.952**
Odds	5:4	3:2	2:1	5:2	3:1	4:1	5:1	6:1	8:1	10:1	25:2	15:1	20:1
dB	14	15	16	17	18	19	**20**	21	22	23	24	25	26
Prob.	0.961	0.969	0.975	0.980	0.984	0.987	**0.990**	0.992	0.993	0.995	0.996	0.996	0.997
Odds	25:1	30:1	40:1	50:1	60:1	80:1	100:1	125:1	165:1	200:1	250:1	330:1	400:1
dB	**-1**	-2	-3	-4	**-5**	-6	-7	-8	-9	**-10**	-11	-12	**-13**
Prob.	**0.442**	0.386	0.333	0.284	**0.240**	0.200	0.166	0.136	0.111	**0.090**	0.073	0.059	**0.047**
Odds	4:5	2:3	1:2	2:5	1:3	1:4	1:5	1:6	1:8	1:10	2:25	1:15	1:20
dB	-14	-15	-16	-17	-18	-19	**-20**	-21	-22	-23	-24	-25	-26
Prob.	0.038	0.030	0.024	0.019	0.015	0.012	**0.009**	0.007	0.006	0.005	0.004	0.003	0.002
Odds	25:1	1:30	1:40	1:50	1:60	1:80	1:100	1:125	1:165	1:200	1:250	1:330	1:400

The 95% level of significance does have a good basis. Clues can be "added" as equivalent to numbers of decibans, and the probabilities read off a mental x-axis. If at least one taxon of a closed causal group (all members surely of that group) reaches a high probability of being a descendant, all of them are. One may note for Figure 5 that 0.95 as the standard scientific minimum of confidence is associated with the beginning of rapid logarithmic rise in probabilities, while 0.99 signals a very tight rise. Neither of these two limits to statistical confidence or credibility, the lower for non-critical decisions, the higher for critical applications or very complex problems, is really very arbitrary, as is sometimes suggested.

THE BAYES FACTOR

Bayes factors (Kruschke 2011: 58) have been often used in phylogenetic analysis to determine model selection and species delimitation (Fan et al. 2011; Grummer et al. 2014; Li & Drummond 2012; Suchard et al. 2002; Sullivan & Joyce 2005; Ward 2008). The use of Bayes factors in the present paper is simplified but, as measures of direction of macroevolutionary transformation, Bayes factors are powerful in explaining monophyly.

The Bayes factor is a measure of which species best models the ancestral species versus the other species. The Bayes factor (B.F.) is the ratio of the likelihoods of the data for two models. It is derived from Harold Jeffrey's (1961) concept of relative betting odds (McGrayne 2011: 116). Thus, B.F. = $Pr(D|M_1)$ / $Pr(D|M_2)$, or, the probability of the data given model 1 divided by the probability of the data given model 2. This is the likelihood ratio. (Note: This is an odds ratio because a probability would have the both of the different probabilities of both models in the denominator, and the fraction would not rise above 1.0.) That is, it measures the change in the odds in favor of the hypothesis when going from the prior to the posterior (Lavine & Schervish 1999). For any *one* hypothesis, if the prior is 0.50, the Bayes factor is simply the posterior probability. For *two* hypotheses, if the prior is 0.50, then the ratio of the two posterior probabilities is the Bayes factor (i.e., the same as the likelihood ratio) (Kass & Raftery 1995). The Bayes factor is somewhat better than standard hypothesis testing because the latter cannot provide an evaluation of information in

favor of the null, just a way to see if it can be falsified, while Bayes factor analysis can evaluate both the hypothesis and the alternative.

Deciban analysis is the same as sequential Bayes analysis but is simpler. The Bayes factor as used here measures probabilistically the central hypothesis that species 1 is ancestral to all immediate descendants as evaluated by sequential Bayes analysis. This is done directly (with a spreadsheet using the results of one Bayes formula analysis as the prior for another, see online spreadsheets) or more coarsely by using decibans (see Table 1). This is done against the alternative hypothesis that some other species is ancestor (as null). That is, the chance of species 1 generating species 2, 3, ...n, versus the chance of species 2, 3, ...n generating species 1 and the rest.

According to Jeffries (1961), a Bayes factor (odds ratio in favor) for one hypothesis against a null hypothesis may be evaluated thus:

Table 2. Table for interpreting Bayes factors according to Jeffries (1961). Bayes factors in the text leave off the ":1" indication but remain ratios.

Bayes factor	Value	Probability	Deciban equivalent
1:1– 3:1	trivial	0–0.76	0–5 dB
3:1–10:1	substantial	0.76–0.91	5–10 dB
10:1–100:1	strong	0.91–0.99	10–20 dB
more than 100:1	decisive	more than 0.99	more than 20 dB

Kass and Raftery (1995) also provide significance charts for B.F. expressed to base $\log_{10}$ and $\log_e$, which are scales suitable for certain purposes, but this paper eliminates unnecessary mathematical burdens for the classical systematist readership. To derive a Bayes factor against an alternative hypothesis, the likelihood of the first hypothesis must be divided by that of the second. We first determine the odds ratio for each of the two probabilities, a:b and c:d, which is (a/c)/(b/d) = ad/bc. Here, c and d are commonly the same (i.e., 1). For example, we may have two hypotheses, A of 0.99 or 20 dB, and B of 0.61 or 2 dB. The odds ratio is 100:1/1.5:1, or 67, which may be taken as the Bayes factor. Approximate odds ratios for various positive and negative deciban levels are given in Table 1. Table 2 allows an interpretation of Bayes factors in terms of probabilities and decibans.

GRANGER CAUSALITY AND BAYES FACTORS

Causal connections are determined by predictability as well as correlation. As discussed by Sugiharra et al. (2012), Berkeley (1710, numbered paragraphs 20, 50, 64, and 65) made the observation that simple correlation in time or space is no assurance of a causal connection between one thing and another. This is mainly because there may be a third thing affecting both, and both will change following a causal connection between the third thing and the other two. There may even be a lag time that confounds direct detection of that third causal element. The solution is apparently "Granger causality" (Granger 1969), which promotes predictability rather than correlation for detecting causality. An element is said to "Granger cause" another element if the predictability of that second element declines when the first element is removed from the model, all else being the same (Sugiharra et al. 2012). Information about a causative element must be independent of other elements associated with some particular process-based model.

Predictability is essential in determining monophyly. Speciation may be interpreted as an ecological time series, thus the causal connections of ancestor-descendant relationships may be tested using Granger causality. The assumption used in the present paper is that for any one, two, or more closely related species all with advanced traits, there should be or should have been another with more generalized traits. The null model for the central hypothesis (that species 1—the most

generalized species—is the putative ancestral taxon) is that some other species in the group is the basal ancestor of the group. We can reject the null if species 1 in the group (the causative ancestor), when eliminated, results in very poor predictability of ancestor-descendant relationships among the remaining taxa in the group. Poor predictability would be immediately evident by there being no strong polarization of support towards one of the other species, that is, low Bayes factors. We do not eliminate the putative progenitor completely from the null model, but instead calculate the chance of each of the other species being progenitor of the group. If such a chance is far lower than that of the putative progenitor, and prediction (here actually retrodiction of a serial evolutionary transformation) is much lessened, we can say the "Granger cause" of the ancestor-descendant relationships in the group is species 1, the putative progenitor.

IMPLIED RELIABLE CONFIDENCE INTERVAL (IRCI)

When many models are tested together there may be support between them. IRCI tells you if you have enough data on transformation directions to make any decision at all. Suppose tyou are examining serial species transformation involving one species and a number of other species (1>2, 1>3, 1>4, etc.), and the data on all species transformations support to some extent the same decision that 1>rest. This multiplication of evidence from two or more data sets can be reflected in increased probability that 1>rest. For this the Implied Reliable Credible Interval (IRCI) formula can be used. The IRCI was used by Zander (2006) to evaluate the chance of at least one of several concatenated cladogram branches of moderate credibility support being correct. It uses the fact that there are more than one sources of at least some support. In this case, even 0.10 probability of 1>2 is some support that 1>rest, because if one hypothesis is true, all species transformation directions are true in this closed group.

Unlike Bayes factor analysis, nesting is not necessary since for any set of probabilities of any process, the more processes involved the greater the chance that one is correct. This is a kind of "existence" estimation when probabilities of events are not individually decisive.

For the IRCI formula, because calculating positive support is difficult, calculating negative support and then subtracting from one is easier. The formula is basically 1 minus the multiplied chances that each element is not true ("not true" meaning one minus the probability it is true). This is not the chance of one particular transformation between two species being correct, but the chance that a sufficient number of hypotheses each of less than acceptable probability will support the idea that at least one of them is correct.

If at least one of the models of macroevolutionary transformation is correct (say, 1>2), then the others (1>3, 1>4, etc.) must be because the models are in a closed causal pool. The closed causal pool for Zander's (2006) cladogram analysis was a series of concatenated cladogram internodes, where if one internode is correct then the taxa beyond the series are indeed in a clade of their own. Here the closed causal pool is a set delimited by the decision that they are all related and one of them is a direct or indirect basal ancestral species for all. It does not particularly increase the odds that a particular species is ancestral, but does ensure that the problem is decidable. If the problem is decidable, then the species with the least probable chance of being ancestral are well established as descendants, and the two most likely are the only candidates. If the Bayes factor for those two most likely candidates exceeds 3.1, then the most likely species is well supported as ancestor.

The IRCI is only used when no single Bayesian posterior probability among the results of sequential Bayes' analysis is adequate for a decision. Note that the IRCI deals with probabilities (chance that the hypothesis is correct) not with likelihoods (chance that the data are correct).

Probabilities calculated from different information do not necessarily have to add to 100 percent, but the closed pool ensures they are nested.

Probabilities can be confusing. When used as clues, any probability higher than 0.50 can be added when converted to a deciban. Thus, in sequential Bayes analysis, any probability more than 0.50 (more than zero dB) contributes to total clue decibans, but probabilities less than 0.50 (less than zero dB) reduce the confidence in total clue decibans and therefore in credibility that one species is direct or indirect ancestor of all in the closed causal group.

For IRCI, on the other hand, any probability greater than zero contributes to total confidence that the question of which species is ancestral is decidable. This is because total polarization of clues in the closed causal group contributes to focusing on one or two species as true candidates. Sometimes one species can be singled out to be well supported as ancestral species.

Consider, for example, the contrived situation of species A through E all in one group with A the fairly obvious ancestral species. What is the exact support for A being the ancestral species? Each species has a probability based on various data that it is the direct or indirect ancestral species of the group. An IRCI for the probabilities 0.90, 0.25, 0.20, 0.10, 0.10 for the species A through E series gives 0.95 IRCI, as in IRCI formula (1):

$$(1 - ((1 - 0.90) \times (1 - 0.25) \times (1 - 0.20) \times (1 - 0.10) \times (1 - 0.1))) = 0.95 \qquad (1)$$

and Bayes factor of 3.6 (that is, 0.90 divided by 0.25), which is substantial for species A being the direct ancestral species. This is true even when it does not have the full 0.95 initial probability based on decibans alone. Note, again, that these probabilities do not have to add to 1.00 because somewhat different data is used to calculate each probability.

CONTINUED IN PART 3, THE ANALYSIS

SUPPLEMENTARY MATERIAL

Spreadsheets for calculating Bayes sequential analysis, decibans, and IRCI are available at <http://www.mobot.org/plantscience/resbot/evsy/sprsh>.

ACKNOWLEDGEMENTS

The comments and editorial expertise of Guy Nesom are much appreciated. I thank the Missouri Botanical Garden for excellent support and superb laboratory space over the past 11 years. The users of the listservers Taxacom and Bryonet are again saluted for their patience and fortitude in fielding my repeated queries and opinions on issues in systematics.

LITERATURE CITED

Atkinson, G.C., A. Kuzmenko, I. Chicherin, A. Soosaar, T. Tenson, M. Carr, P. Kamenski, and V. Hauryliuk. 2014. An evolutionary ratchet leading to loss of elongation factors in eukaryotes. BMC Evolutionary Biology 14: 35. [doi: 10.1186/1471-2148-14-35]

Barraclough, T. and S. Nee. 2001. Phylogenetics and speciation. Trends Ecol. Evol. 16: 391–399.

Benton, M.J. and P.N. Pearson. 2001. Speciation in the fossil record. Trends Ecol. Evol. 16: 405–411.

Berkeley, G. 1710. A Treatise on Principles of Human Knowledge. Dublin.

Blomberg, S.P., T. Garland, and A.R. Ives. 2003. Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. Evolution 57: 717–745.

Bonawitz, E., S. Denison, A. Chen, A. Gopnik, and T.L. Griffiths. 2013. A Simple Sequential Algorithm for Approximating Bayesian Inference. <http://cocosci.berkeley.edu/tom/papers/wslsadults.pdf>

Bowler, P. 1989. Evolution: The History of an Idea (rev. ed.). Univ. of California Press, Berkeley.

Cavalier-Smith, T. 2010. Deep phylogeny, ancestral groups and the four ages of life. Philos. Trans. Roy. Soc. B. 365: 111–132.

Fan, Y., R. Wu, M.H. Chen, L. Kuo, and P.O. Lewis. 2011. Choosing among partition models in Bayesian phylogenetics. Molec. Biol. Evol. 28: 523–532.

Farris, J.S. 2012. Early Wagner trees and the "cladistic redux." Cladistics 27: 1–3.

Felsenstein, J. 2001. The troubled growth of statistical phylogenetics. Syst. Biol. 50: 465–467.

Futuyma, D.J. 2009. Evolution (ed. 2). Sinaurer Associates, Inc., Sunderland, Massachusetts.

Gavrilets, S. and A. Vose. 2005. Dynamic patterns of adaptive radiation. Proc. Natl. Acad. Sci., U.S.A. 102: 18040–18045.

Giere, R.N . 2006. Modest evolutionary naturalism. Biol. Theory 1: 52–60.

Gittenberger, E. 1991. What about non-adaptive radiation? Biol. J. Linnean Soc. 43: 263–272.

Good, I.J. 1979. Studies in the history of probability and statistics. XXXVII. A.M. Turing's statistical work in World War II. Biometrika 66: 393–396.

Gould, S.J. and R.C. Lewontin. 1979. The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. Proc. Roy. Soc., London B 205: 581–598.

Granger, C.W.J. 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37: 424–438.

Grant, V. 1949. Pollination systems as isolating mechanisms in flowering plants. Evolution 3: 83–97.

Grant, V. 1989. The Evolutionary Process: A Critical Review of Evolutionary Theory. Columbia Univ. Press, New York.

Grummer, J.A., R.W. Bryson, Jr., and T.W. Reeder. 2014. Species delimitation using Bayes factors: simulations and application to the *Sceloporus scalaris* species group (Squamata: Phrynosomatidae). Biol. Syst. 63: 119–2014.

Gyllenberg, M. and T. Koski. 2002. Bayesian predictiveness, exchangeability and sufficientness in bacterial taxonomy. Math. Biosci. 177–178: 161–184.

Haller, B.C. and A.P. Hendry. 2013. Solving the paradox of stasis: Squashed stabilizing selection and the limits of detection. Evolution 68: 483–500.

Jeffreys, H. 1961. The Theory of Probability. Third edition. Oxford Univ. Press, Oxford.

Kachiashvili, K.J. 2012. Comparison analysis of the Wald's and the Bayes type sequential methods for testing hypotheses. World Academy of Science, Engineering and Technology 70: 1168–1172.

Kruschke, J.K. 2011. Doing Bayesian Data Analysis: A Tutorial with R and BUGS. Elsevier, Amsterdam, New York.

Lakatos, I. 1978. The Methodology of Scientific Research Programmes: Philosophical Papers. Volume 1. Cambridge Univ. Press, Cambridge.

Lauritzen, S. 2009. Sequential Baysian Updating. University of Oxford. BS2 Statistical Inference. <http://www.stats.ox.ac.uk/~steffen/teaching/bs2HT9/kalman.pdf>

Lavine, M. and M. J. Schervish. 1999. Bayes factors: What they are and what they are not. Amer. Statistician 53: 119–122.

Levinton, J. 1988. Genetics, Paleontology, and Macroevolution. Cambridge Univ. Press, Cambridge.

Li, W.L. S. and A.J. Drummond. 2012. Model averaging and Bayes factor calculation of relaxed molecular clocks in Bayesian phylogenetics. Mol. Biol. Evol. 29: 751–761.

Mayr, E. 1954. Pp. 157–180, *in* Huxley, J. et al. (eds.). Change of a genetic environment and evolution. Evolution as a Process. Allen and Unwin Press, London.

Mayr, E. 1981. Biological classification: Toward a synthesis of opposing methodologies. Science 214: 510–516.
Mayr, E. 2001. What Evolution Is. Basic Books, New York.
McGrayne, S.B. 2011. The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code. Yale Univ. Press, New Haven, Connecticut.
Mooi, R.D. and A.C. Gill. 2010. Phylogenies without synapomorphies—A crisis in fish systematics: Time to show some character. Zootaxa 2450: 26–40.
Pearman, P.B., A. Guisan, O. Broenniman, and C.F. Randin. 2007. Niche dynamics in space and time. Trends Ecol. Evol. 23: 149–158.
Pelser, P.B., B. Nordenstam, J.W. Kadereit, and L. E. Watson. 2007. An ITS phylogeny of tribe Senecioneae (Asteraceae) and a new delimitation of *Senecio* L. Taxon 56: 1077–1104.
Pennisi, E. 2003. Modernizing the Tree of Life. Science 300: 1692–1697.
Peterson, A.T., J. Soberón, and V. Sánchez-Cordero. 1999. Conservatism of ecological niches in evolutionary time. Science 285: 1265–1267.
Rieppel, O. 2006. Willi Hennig on transformation series: metaphysics and epistemology. Taxon 55: 377–385.
Schluter, D. 2000. The Ecology of Adaptive Radiation. Oxford Series in Ecology and Evolution. Oxford Univ. Press, New York.
Seehausen, O. 2006. African cichlid fish: A model system in adaptive radiation research. Proc. Roy. Soc. B 273: 1987–1998.
Simpson, G.G. 1953. The Major Features of Evolution. Columbia Univ. Press, New York.
Sinsheimer, J.S., M.A. Suchard, K.S. Dorman, F. Fang, and R.E. Weiss. 2003. Are you my mother? Bayesian phylogenetic inference of recombination among putative parental strains. Applied Bioinformatics 2003, 2(3): 131–144.
Sober, E. 2008. Evidence and Evolution: The Logic Behind the Science. Cambridge Univ. Press, Cambridge.
Suchard, M.A., R.E. Weiss, K.S. Dorman, and J.S. Sinsheimer. 2002. Oh brother, where art thou? A Bayes factor test for recombination with uncertain heritage. Syst. Biol. 51: 715–728.
Sugihara, G., R. May, Hao Ye, Chih-hao Hsieh, E. Deyle, M. Fogarty, and S. Munchs. 2012. Detecting causality in complex ecosystems. Science 338: 496–500.
Sullivan, J. and P. Joyce. 2005. Model selection in phylogenetics. Ann. Rev. Ecol. Evol. Syst. 36: 445–466.
Takhtajan, A. 1997. Diversity and Classification of Flowering Plants. Columbia Univ. Press, New York.
Vernon, K. 1993. Desperately seeking status: Evolutionary systematics and the taxonomists' search for respectability 1940–60. Brit. J. Hist. Sci. 26: 207–227.
Via, S. 2001 . Sympatric speciation in animals: The ugly duckling grows up. Trends Ecol. Evol. 16: 381–390.
Wald, A. 1947. Sequential Analysis. John Wiley and Sons, New York.
Ward, E.J. 2008. A review and comparison of four commonly used Bayesian and maximum likelihood model selection tools. Ecol. Modeling 211: 1–10.
Werner, O., J.A. Jiménez, R.M. Ros, M.J. Cano, and J. Guerra. 2005. Preliminary investigation of the systematics of *Didymodon* (Pottiaceae, Musci) based on nrITS sequence data. Syst. Bot. 30: 461–470.
Williams, D. 2012. Willi Hennig and the cladistic revolution. Systematist 34: 6–11.
Zander, R.H. 1993. Genera of the Pottiaceae: Mosses of harsh environments. Bull. Buffalo Soc. Nat. Sci. 32 1–378.
Zander, R.H. 1998. A phylogrammatic evolutionary analysis of the moss genus *Didymodon* in North America North of Mexico. Bull. Buffalo Soc. Nat. Sci. 36: 81–115.
Zander, R.H. 2001. A conditional probability of reconstruction measure for internal cladogram branches. Syst. Biol. 50: 425–437.

Zander, R.H. 2006. The Pottiaceae s. str. as an evolutionary Lazarus taxon. J. Hattori Bot. Lab. 100: 581–602.
Zander, R.H. 2008. Evolutionary inferences from non-monophyly of traditional taxa on molecular trees. Taxon 57: 1182–1188.
Zander, R.H. 2010a. Taxon mapping exemplifies punctuated equilibrium and atavistic saltation. Pl. Syst. Evol. 286: 69–90.
Zander, R.H. 2010b. Structuralism in phylogenetic systematics. Biol. Theory 5: 383–394.
Zander, R.H. 2013. A Framework for Post-Phylogenetic Systematics. Zetetic Publications, St. Louis. CreateSpace Independent Publishing, Amazon.