



doi: 10.15407/ukrbotj74.02.103

A simple parsimony-based approach to assess ancestor-descendant relationships

Damien AUBERT

Académie de Clermont-Ferrand, Ministère de l'Éducation Nationale
3, avenue Vercingétorix 63033
Clermont-Ferrand Cedex 1, France
damien.aubert@ac-clermont.fr

Aubert D. **A simple parsimony-based approach to assess ancestor-descendant relationships.** Ukr. Bot. J., 2017, 74(2): 103–121.

Abstract. One of the main goals of systematics is to reconstruct the tree of life. Half a century ago, the breakthrough of cladistics was a major step towards this objective because it allowed us to assess relatedness patterns among species, an abstract kind of relationship. Unfortunately, the philosophy of cladism forbade to go further and to seek more realistic relationships, like the ancestor-descendant relationship, which is the expected fundamental kind of relationship of the tree of life according to Darwinian evolution. Here, I describe a simple parsimony-based procedure which can be used to transform a classical cladogram into a genuine phylogenetic tree, i.e. a caulogram. It consists in deleting as many unobserved and unnamed nodes as possible and replacing them with observed and named species. A new Bayesian non-stochastic weighting scheme is used to assess character reliability for both this procedure and classical cladogram construction. I illustrate the whole process by assessing the relationships between the species of the moss genus *Didymodon* sensu lato (Pottiaceae) and discuss the resulting caulogram by confronting it with the previous methodology from the evolutionary literature. I finally argue that strictly adhering to cladist epistemology is untenable and that we must seek new formal methods to find ancestral species as well as ancestral higher taxa.

Keywords: ancestor, Bayesian analysis, Bremer support, evolutionary systematics, parsimony, weighting

Introduction

Background and Motivation

Since the sixties, cladistics has aimed to resolve the relationships between species (Hennig, 1950, 1966). These relationships are depicted by cladograms, i.e. hierarchical treelike diagrams where clusters show which two of any three species are more closely related to each other than either is to the third one (Hennig, 1966; Hull, 1979). This type of relationship is characterized as “relatedness” and is supposed to represent the relative recency of the “hypothetical last common ancestors” (or more rigorously the order of emergence of evolutionary novelties). It is, however, dubious that such a relationship corresponds to a true natural process. Unless one believes that a mother species always disappears when it speciates (Lee, 1995), relatedness may ambiguously refer either to a true sister-group relationship (SGR) or

to an ancestor-descendant relationship (ADR) (Aubert 2015). In the first case, the two species are descendants of another unobserved third species (so, a true SGR is in fact two hidden ADRs), while in the second case one of them is a descendant of the other one. This second type of relationship is nonetheless misleadingly represented as a false SGR through the artificial introduction of unobserved species on internal nodes (see **Fig. 1**) if a cladogram is interpreted as a true phylogenetic tree (here “true” means *sensu* Hennig, see below) or caulogram (the word “caulogram” designates a tree that emphasizes serial macroevolutionary transformations, i.e. ADRs, see Zander, 2013). The internal nodes may be avoided by representing a cladogram as a set of nested parentheses, but the introduction of unobserved species is still logically implied by the assumption that none of the known species is actually an ancestor (Podani, 2013). Cladograms are thus ambiguous, not faithful pictures of evolutionary history.

© D. AUBERT, 2017

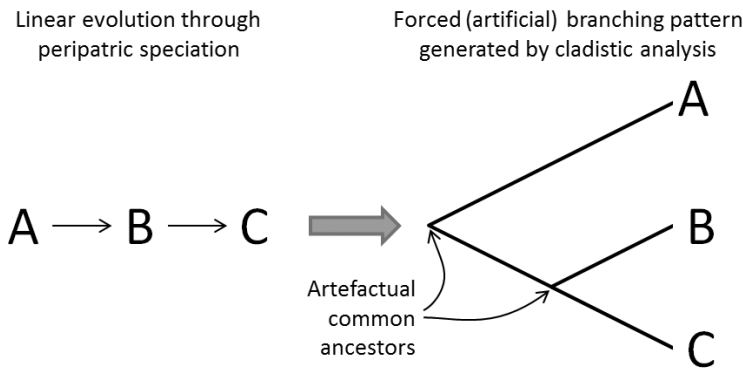


Fig. 1. The artefactual entities introduced by cladistic analysis. The evolutionary model (dichotomous splitting) used by cladistic analysis forces us to hypothesize many unobserved entities in order to optimize the number of transformation events.

Furthermore, even in the case of a true SGR, unnamed nodes would imply that relatedness is a fundamental phylogenetic relationship. This poses epistemological problems because sisters have always been independent entities; they may for example be born or speciated at very different times. The fact that one sister would have never existed does not necessarily imply the non-existence of the other one, while the non-existence of a mother species necessarily implies the non-existence of all its daughter species. In fact, a true SGR (contrary to a false SGR) only means that both species share a ADR with the same third species. ADRs are thus far more fundamental evolutionary (i.e. truly “phylogenetic”) relationships than SGRs. From a more biologically grounded perspective, ADRs generally represent peripatric (or “budding”) speciations. This means that a mother species tends to disperse and invade geographically isolated locations. From there, local populations evolve new traits through genetic bottlenecks and directional selection. Thus, they transform into daughter species while the principal population remains morphologically unchanged. This is because the latter benefits from the stabilizing effects of a larger genetic pool and purifying selection. One may call this phenomenon “phylogenetic niche conservatism” (Pyron et al., 2015). It is very unlikely that the mother species transforms entirely into two daughter species through the gradual divergence of two subspecies at the same time (Levin, 1993). Of course, budding speciation is especially widespread in islands, but is not limited to such. As an example, the endemic liverwort *Leptoscyphus azoricus* of Macaronesia has been shown to have originated through this

process from a parental population of continental *Leptoscyphus porphyrius* (Vanderpoorten and Long, 2006). Moreover, we must point out that this mode of speciation is not restricted to plants. For instance, a recent palaeontological study of the pterocephaliid trilobites have shown through the implementation of a probabilistic model that the main mode of speciation (if not the sole one) is indeed budding cladogenesis, neither bifurcating cladogenesis nor pure anagenesis (Bapst and Hopkins, 2017). This result is largely consistent with the literature in other fields like the foraminifera (Aze et al., 2011). Last but not least, the cladistic axiom that two of any three species must be closer is false since a single mother species may give birth to more than two daughters. Unfortunately, cladistic algorithms force the data to fit a dichotomous tree, which is like trying to hammer a square peg into the round hole of an ideal (Zander, 2013). Such a propensity to use axiomatized synchronous (ahistorical) structures as a fundamental framework is referred to as structuralism (Zander, 2011; Aubert, 2015).

All of the above reasons motivate the research for new methods able to transform cladograms, i.e. Hennig’s “phylogenetic diagrams”, into true phylogenetic trees (or caulograms) reflecting ADRs, i.e. the real genealogical relationships between species (Prothero and Lazarus, 1980; Paul, 1992; Alroy, 1995; Crawford, 2010; Aze et al., 2011; Tsai and Fordyce, 2015). Contrary to what is sometimes assumed (Gee, 2000), the probability of encountering an ancestral species in the fossil record or among extant species is far from being negligible (Rieseberg and Brouillet, 1994; Crisp and Chandler, 1996; Foote, 1996; Funk and Omland, 2003; Aldous

and Popovic, 2005; Aldous et al., 2011; Ross, 2014). A phylogenetic analysis that only results in the publication of a new cladogram is therefore merely a preliminary work, which demands a post-cladistic treatment in order to eliminate the *ad hoc* virtual ancestors and to clarify the nature of the relationships. The feasibility of this objective will be demonstrated by the study of the North American species of the moss genus *Didymodon* sensu lato (Pottiaceae), which has been recently divided into six segregate genera: *Vinealobryum*, *Didymodon* sensu stricto, *Trichostomopsis*, *Geheebia*, *Exobryum* and *Fuscobryum* (Zander, 2016).

This work builds on the studies of Zander (2013, 2014a, b, c, 2016) which introduce means of diagramming serial evolution of taxa as caulograms, and suggest support values in terms of decibans. Although Zander mentioned that variation in occurrence of shared traits affects credibility, he did not detail explicit means of formally measuring and incorporating variation. This paper introduces the consistency index and successive weighting in cladistic analysis as a means of evaluating variability of traits, with those less variable being more important. This study is restricted to morphological traits.

Rationale for the Post-Cladistic Analysis

In the cladist framework synapomorphies are considered the only evidence of common ancestry. Morphological character mapping over the resulting cladogram allows us to infer the phenotype of this common ancestor. If a branch connecting such an internal node to a terminal species bears no character transformation, and hence has a length of zero, then the phenotype of the common ancestor is exactly the same as the terminal species. However, cladists generally do not regard this as evidence that the terminal species and the common ancestor are the same entity, and prefer to systematically hypothesize that they are different (note that cladism is not the same thing as cladistics; see Aubert, 2015). They argue that only shared character transformations can provide evidence of relationship, and that the lack of a transformation is only a lack of evidence, not evidence *per se*. At best, a species characterized by the absence of autapomorphy is termed “metaphyletic” (Donoghue, 1985; de Queiroz and Donoghue, 1988). This means that we do not know whether this species is holophyletic or paraphyletic (respectively, all descendants included or not; see Ashlock, 1971).

I would think however that this interpretation is unscientific. It is argued that since one cannot positively

disprove that the putative common ancestor and the terminal species are distinct entities, then they are (or at least one cannot decide). Here, the burden of proof has just been unjustifiably reversed. Indeed, the burden of proof lies upon a person making scientifically unfalsifiable claims. The very existence of an unobserved common ancestor is an unfalsifiable claim because even if we would have sampled a species matching its reconstructed phenotype, one could argue that since the species has been observed then it is not the common ancestor we were looking at. In this framework, common ancestors are not only unobserved, but also unobservable entities. Yet, the principle of Occam's razor tells us that we should minimize the number of such *ad hoc* entities. On the contrary, the claim that an observed species is the same entity as the predicted common ancestor is a falsifiable claim. It would be theoretically sufficient to find a single autapomorphy in order to disprove it. The existence and the observation of common ancestors are both expected and likely from the theory of evolution. Science must therefore always favour the simplest explanation: if an observed species matches the phenotype of a predicted species, then both species are the same entity. In other words, this is the null hypothesis we must test against alternatives. The very concept of “metaspecies” is therefore unneeded; all so-called unresolved species must be considered paraspecies.

Now, not all morphological characters are equally reliable. Characters in a phylogenetic data set that transform as shared traits (synapomorphies) only once in a cladogram are reputed quite stable, and so are reliable indicators of relationships. But characters that transform many times are rather labile and create many homoplastic misleading relationships. What if then, an observed species *nearly* matches the phenotype of a predicted species? Are they the same? Here we must leave naïve Popperian hypothetico-deductivism, i.e. unweighted parsimony optimization. It is rather obvious that if the observed autapomorphies are several stable characters, then the null hypothesis must be rejected in favour of the alternative one. But if the observed autapomorphy is only a single very labile character, then the null hypothesis cannot be convincingly rejected. The objective evaluation of the null hypothesis therefore demands a probabilistic quantification of characters' reliability. As we will see, weighted parsimony can be interpreted as a form of non-parametric (i.e. not “model-based”) Bayesian approach.

A Bayesian Interpretation of the Consistency Index

We consider a morphological binary character x in a matrix of OTUs. We note s the actual number of changes of this character occurring in the most parsimonious dichotomous unrooted cladogram (or at least the chosen one) and m the minimum number of changes that it may require in any such cladogram (i.e. one, in this case). The consistency index is thus equal to $c = m/s$. Let us now consider four OTUs A, B C, and D. We know the fact $F =$ "A and B share the same state of x , while C and D share another state of this character". The reliability of x can be regarded as the increase in probability that the relationship $R =$ "{A, B} and {C, D} are two mutually exclusive monophyletic groups" is true. We are only interested in monophyly, not holophyly, because rooting a topology is a different task from reconstructing it ("monophyletic" means that the most recent common ancestor is a member of the group, this can be tested without rooting; and then different rootings of the tree can make this group "holophyletic" or "paraphyletic", i.e. containing all its descendants or not, respectively).

The prior probability of R , i.e. not knowing F , is theoretically $p = 1/3$ because there are exactly three unrooted possible four-taxon trees and only one is compatible with R . The posterior probability q , knowing F , would be one if and only if both character states are homologous for A and B, and for C and D respectively. If either of the two states is homoplastic in these pairs (for example it evolved independently in A and B) then R would be true only by chance, so its probability would be only $1/3$. We must therefore evaluate the probability z that the first case occurs.

The character x clusters the whole tree into $s + 1$ monophyletic parts. There may be u monophyletic groups with state 1 and v monophyletic groups with state 0, so that $u + v = s + 1$. The probability z that A has been randomly picked in the same monophyletic as B and C in the same as D is therefore $z = 1/(uv)$. Any u and v are theoretically possible, but since convergences and reversions are here considered equally probable, we should generally get $u \approx v$. More accurately, this is like tossing a coin $s - 1$ times because of the constraint that $u \geq 1$ and $v \geq 1$. Consequently, we have a simple binomial distribution:

$$z = \frac{1}{2^{s-1}} \cdot \sum_{u=1}^s \binom{s-1}{u-1} \cdot \frac{1}{u \cdot (s+1-u)}$$

Once z has been computed, we can easily estimate q as the weighted sum of both cases, i.e. $q = z + (1-z)/3 = 2z/3 + 1/3$. Therefore the posterior odds (ratio of probabilities), knowing F , are:

$$q/(1-q) = (2z/3 + 1/3)/(2/3 - 2z/3) = (2z + 1)/(2 - 2z)$$

Since without knowing F , the relationship R can only be true by chance, the prior odds were $(1/3)/(2/3) = 1/2$ (i.e., 0.5:1). Hence, the evidence provided by F can be evaluated as the ratio of odds, also known as Bayes factor:

$$k = [(2z + 1)/(2 - 2z)]/(1/2) = (2z + 1)/(1 - z)$$

This Bayes factor is independent of the prior probability of R , which means that $q/(q - 1) = kp/(1 - p)$ is always true even if $p \neq 1/3$ because of some other sources of knowledge (stratigraphy or biogeography for example). Thus, k measures the amount of knowledge that F adds to our previous knowledge. This evidence provided by the consistency index is more appropriately expressed in the logarithmic unit of bans or decibans (dB) because this allows us to interpret evidence in an intuitive manner and makes it possible to literally add units of knowledge. Thus, if we get several independent sources of evidence from different characters, we can mentally add up units of evidence instead of doing complicated computations. A deciban (or decihartley) is a tenth of a ban, a unit used by Bayesian statisticians to represent Bayes factors in hypothesis testing. The deciban scale is here calculated with the formula: $w = 10 \times \log_{10} k$ (the letter w stands for "weight of evidence"). This scale goes from 0 to infinity, but 13 dB can be interpreted as a strong evidence (> 95% chances to be true, see **Table I**). Let us mention that the smallest intuitively detectable evidence is roughly 1 dB, which approximately corresponds to the difference we perceive between the odds 5:4 (around 55-56%) and the totally equivocal 1:1 (exactly 50%) (Good, 1979, 1985).

The computation of z is a little more complicated for multistate characters, but the problem can be reduced to a weighted average of the reliability of each possible pair of states. There are $\binom{m+1}{2}$ such pairs. For example, if we consider a three-state character, A and B may be in state 0 and C and D in state 1 or 2, or A and B may be in state 1 and C and D in state 2, or inversely. We can therefore evaluate independently the three possible pairs 0/1, 0/2 and 1/2. If $s = 3$, then one of the states is represented by two separated monophyletic groups instead of just one, so two of the three pairs have a z equal to 0.5 whereas the third one have a z equal to 1, hence a global value $z = 2/3$.

If $s = 4$, then either two states are represented by two separated monophyletic groups, or one state is represented by three such groups. In the first case, two pairs have $z = 0.5$ and the third one $z = 0.25$, while in the second case two pairs have $z = 1/3$ and the third one $z = 1$. Provided that these supplementary monophyletic partitions are distributed randomly among the different states, the first case has a probability of $2/3$ while the second one has a probability of $1/3$. Hence, the global value of $z = 2/3 \times (1/2 + 1/2 + 1/4)/3 + 1/3 \times (1/3 + 1/3 + 1)/3 = 25/54$. Here, we observe that $(m, s) = (2, 4)$ implies $z \approx 0.463$, which is slightly inferior to the case $(m, s) = (1, 2)$ where $z = 0.5$, although $c = 0.5$ in both cases. Thus, the consistency index does not accurately take into account the number of distinct states. In the general case we have:

$$z = \frac{\sum_{v=1}^{s-m+1} \sum_{u=1}^{s-m-v+2} \binom{s-m-v+1}{u-1} \cdot \binom{s-m}{v-1} \cdot \binom{m}{2} \cdot (m-1)^{s-m-u-v+1} \cdot \frac{1}{u \cdot v}}{\binom{m+1}{2} \cdot (m+1)^{s-m-1}}$$

If $s - m = 1$ then $z = c = m/s$, but the value of z is generally strictly lower than that of c . The values of z are presented in the **Table II**, as well as the corresponding values of w . Note that the consistency index cannot reach 0 even if the character is cladistically completely uninformative. We could note g the maximum number of transformations that the character x may undergo among all possible cladograms in order to explain its state distribution, i.e. the minimum between the number of 0 and the number of 1. By definition, c

Table I. The deciban scale of the Bayesian weight of evidence. Only the most salient values are psychologically interpreted.

Probability	Odds	Evidence (dB)	Interpretation
1.000	$+\infty$	$+\infty$	Certain
0.990	100:1	20	Decisive (Nearly three-sigma)
0.952	20:1	13	Strong (Nearly two-sigma)
0.929	13:1	11.14	
0.889	8:1	9.03	
0.863	6.3:1	8	Substantial
0.800	4:1	6.02	
0.760	3.2:1	5	Believable
0.750	3:1	4.77	
0.666	2:1	3	Small hint
0.636	1.75:1	2.43	
0.571	1.33:1	1.25	
0.557	1.26:1	1	Barely worth mentioning
0.529	1.12:1	0.50	
0.5	1:1	0	Totally equivocal

cannot go below m/g . This has led some systematists to the conclusion that c must be rescaled between 0 and 1 (Farris, 1989). However, this would mean that the amount of evolution needed for the transformation of a cladistically uninformative character is exactly zero, i.e. is equivalent to no transformation at all. Thus, I do not recommend the use of the rescaled consistency index (RCI) to calculate the evidence provided by a character to assess the nature of a transformation of shared traits.

Table II. The consistency index understood as Bayesian evidence. The letters m and s indicate respectively the minimal and the actual number of transformations of a character on a particular cladogram. The consistency index is defined as $c = m/s$. The values of z are probabilities while the values of w are expressed in decibans (see text for formulas).

Values of z :

s	m				
	1	2	3	4	5
1	1				
2	0.5	1			
3	0.292	0.667	1		
4	0.188	0.463	0.75	1	
5	0.129	0.333	0.573	0.8	1
6	0.094	0.248	0.445	0.647	0.833
7	0.071	0.189	0.352	0.528	0.699
8	0.055	0.148	0.282	0.435	0.59
9	0.044	0.119	0.23	0.362	0.502
10	0.036	0.097	0.189	0.304	0.429
11	0.03	0.08	0.158	0.258	0.369
12	0.025	0.068	0.133	0.22	0.319

Values of w :

s	m				
	1	2	3	4	5
1	13.00				
2	6.02	13.00			
3	3.49	8.45	13.00		
4	2.28	5.55	10.00	13.00	
5	1.60	3.98	7.01	11.14	13.00
6	1.17	2.98	5.33	8.12	12.04
7	0.89	2.31	4.20	6.39	9.01
8	0.70	1.82	3.38	5.20	7.26
9	0.57	1.47	2.77	4.32	6.04
10	0.47	1.21	2.31	3.64	5.12
11	0.39	1.01	1.94	3.10	4.40
12	0.33	0.86	1.65	2.66	3.81

Table III. The matrix of characters of *Didymodon* s. l. The species *Barbula unguiculata* is the outgroup. The 42 characters used are presented in the annex.

	1	2	3	4	
B. un	01000 00000	01100 00000	00010 01000	00000 02100	20
V. vi	01011 00000	01111 01010	00110 01000	00001 01100	20
V. br	01011 10020	20211 01100	00110 01000	00001 01100	00
V. ne	01013 00020	20211 01010	11122 01000	11001 1????	??
V. nl	01010 10010	03111 01100	00120 11000	00011 02100	20
V. mu	01010 10000	00124 01100	00010 11000	10012 ?????	??
T. au	11013 20012	01111 12020	10111 00000	00011 01100	11
T. um	21013 20002	11012 12020	10011 00101	02?11 01100	11
T. re	10012 20022	00210 12020	10120 01100	00?01 010?2	00
D. ac	01011 20001	20111 11101	00110 11100	00210 02101	10
D. ri	01011 20001	00113 11101	00110 11000	00210 02101	10
D. jo	01010 10002	00114 10101	00?00 01?10	12112 00011	02
D. ic	01011 20001	10112 11101	00110 11000	10210 02101	10
D. an	01010 10002	00114 10011	00112 01000	12010 1????	??
E. as	32210 22010	11211 01100	00110 02110	02102 00011	01
G. fa	02010 11010	01111 10100	00000 12010	00101 01010	20
G. to	01010 11011	11110 01100	00010 120?0	02101 01012	01
G. le	02011 31000	11112 00100	00000 02100	12102 1????	??
G. ms	21011 31020	?0111 11100	00010 12000	00101 1????	??
G. fe	02010 12020	01111 11100	00000 12010	12101 01010	10
G. mx	02010 12020	02111 11100	00000 03110	12302 1????	??
G. gi	03010 11010	03111 11100	00020 03110	12302 1????	??
F. ni	31110 10022	00101 12100	00000 02101	01002 01100	10
F. pe	30012 20022	00100 12100	00100 01000	01002 1????	??
F. su	31110 10122	00100 12100	00100 12000	01002 1????	??

The Bayesian pieces of evidence provided by putative characters' transformations can be used as weights in a weighted parsimony cladistic analysis and as a branch length scale in a phylogrammatic representation of amounts of evolution separating species. Indeed, additivity is an expected property of distances on a phylogenetic tree, and contrary to the raw consistency index, Bayesian evidence measured in decibans is additive. This makes sense intuitively: if a stable character transformation is as probable as several more common character transformations, then they must be represented by the same length. Therefore, labile character transformations should be represented by shorter branches. The (patristic) distances on such a phylogram would represent the probability that any character transform. It is an intuitive measure of the "amount of evolution" between any pair of species.

Material and Methods

Data Source

As an example to demonstrate the feasibility of the analysis, I have used the matrix of 20 characters of 23 OTUs from Zander (1998), including 22 species from North America of the genus *Didymodon* sensu lato, plus an outgroup species *Barbula unguiculata*. Two additional species and 22 supplementary characters were included after reviewing the most recent literature (Zander, 2013, 2014a, b, c, 2016), plus the Internet website <www.efloras.org>. The data are entirely neontological, all species being extant (see **Table III** and **Annex**).

Weighted Parsimony Cladistic Analysis

Character transformations were generally considered as unordered and of equal weight, except for discretized

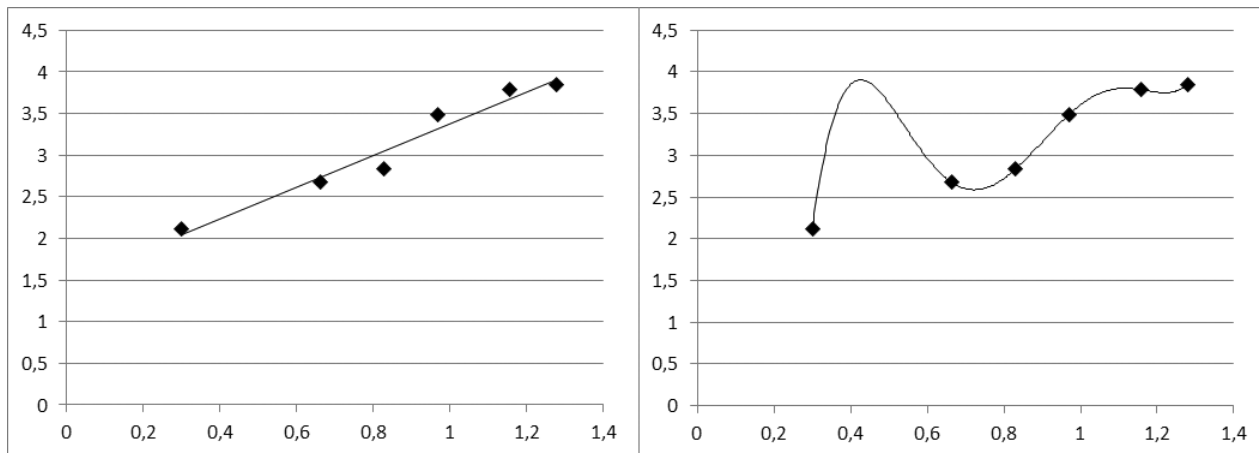


Fig. 2. The concept of overfitting. With only two parameters, a linear function can only approximate the six depicted points. With six parameters, a polynomial function can go exactly through each of the six points. The function is thus more precise, but obviously less accurate. In our case, each unnecessary ancestor can be regarded as a supernumerary parameter.

morphometric characters and for those containing intermediary, variable or ambiguous states. These were considered ordered characters and were rescaled so that their portion of the transition over their full range represents a single transformation (see **Annex**). Heuristic search of the most parsimonious cladograms were carried on with PAUP* version 4.0a150, with 100 replicates starting with random trees, holding 10 trees at each step, swapping on all trees with TBR algorithm, letting reconnection limit at 8 by default and saving multiple trees. After the initial search non-homoplastic characters were considered equal to 13 dB while others were reweighted according the Bayesian interpretation of the mean consistency index (see **Table II**) of all retained trees. The weights were used in the next steps to search again the most parsimonious trees and then compute again new weights, and so on iteratively (Farris, 1969). Computed weights were used at a precision of 2 digits after the point. The branches were systematically collapsed (creating polytomies) when the minimum length was zero (parameter “amb-”).

Bremer Support

The most parsimonious tree is not always the true tree. In fact, optimization of the data over a model can result in overfitting. This is a serious bias (see **Fig. 2** for a simple illustration of this notion). In classical cladistic analysis, Bremer support of a clade in the most parsimonious tree is the minimum number of extra steps required to draw a near-most-parsimonious tree that does not contain this clade (Bremer, 1988, 1994). In order to evaluate the support of the putative clades,

the trees that are less than 12.99 dB away from the most parsimonious one were sampled and a strict consensus tree was built. Branches whose loss is inferior to this cost were therefore not retrieved, which means that their Bremer support is strictly inferior to 13 dB (i.e. < 95%). On the other hand, branches that were retrieved have a Bremer support superior or equal to 13 dB.

Post-Cladistic Analysis

Character changes can be mapped using either ACCTRAN or DELTRAN algorithms, so that the mean length of every branch is estimated. Branch length is simply the sum of weights of all character transformations (measured in dB as assigned by the weighted parsimony cladistic analysis, PAUP* can do it automatically, see above) along this branch. Nodes that are joined by very short branches correspond to clades supported by very labile characters. Intuitively, the support of such branches is not very strong. In fact, if the length is inferior to 13 dB, we should generally conclude that the branch does not exist and the two nodes represent the same entity. Such a deletion would have consequences on neighbouring branches (see **Fig. 3**). Any procedure of elimination of unnecessary entities may be called superoptimization (Zander, 2013). However, we cannot proceed directly to these deletions directly on the most parsimonious cladogram, because we would rely on the assumption that its topology is not strongly distorted by cladistic overfitting (it could be possible to do so only if each and every ancestral species gave birth to only one or two derived species, which is a very strong assumption I do not hold).

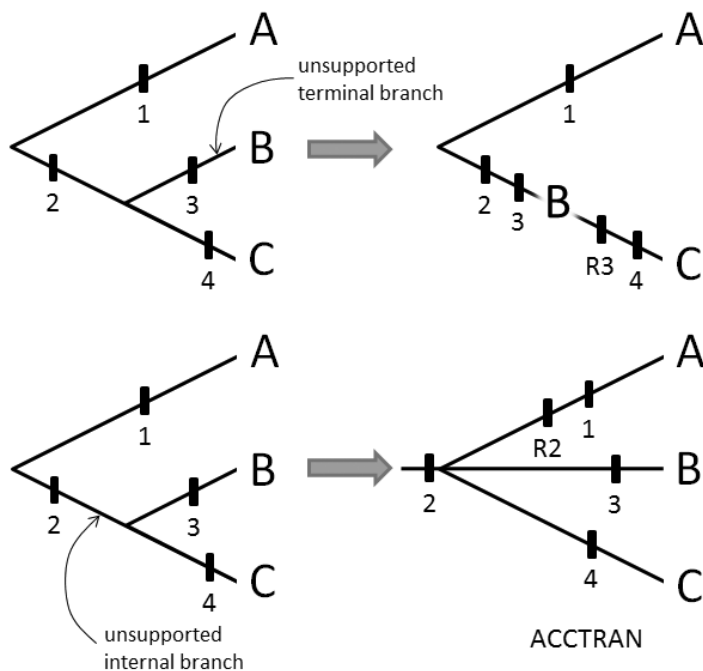
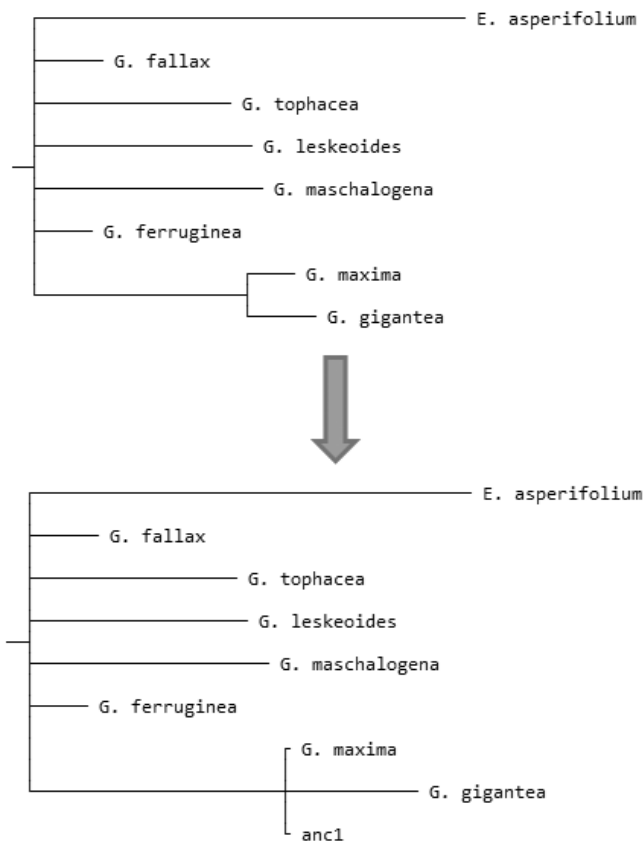


Fig. 3. The principle of superoptimization. The goal of superoptimization is to remove unnecessary entities by deleting insufficiently supported terminal or internal branches. Like in classical character mapping, several solutions may exist (ACCTRAN or DELTRAN). The letters “A”, “B” and “C” indicate species, the numbers indicate characters being transformed, and the letter “R” indicates a reversion.



947.25 steps

954.87 steps

Fig. 4. How to force PAUP* to draw a caulogram. These are parts of phylograms drawn by PAUP*. This software is able to map characters' transformations according to several optimization algorithms (ACCTRAN, DELTRAN or MINF) but a particular branch cannot be directly forced to have a length of zero. However, if we add a new OTU (here «anc1» has exactly the same character states as *G. maxima*) in an unresolved trichotomy then PAUP* is obliged to infer that the last common ancestor of these three OTUs had the same character states as the majority of them (that is, «anc1» and *G. maxima*), thus drawing two branches with a length of zero and increasing the length of the remaining branches. The tests must be conducted in the context of the tree, not in isolation. Because the result may change according to this very context, huge polytomies (like the one including *G. fallax*, *G. ferruginea*, etc.) necessitate trying many rearrangements of outgroups (not just pairwise tests like those of **Table V**) in order to find the best superoptimization.

Table IV. The stable weights obtained after successive weighting. The best estimate of weight corresponds to the consistency index of the only one most parsimonious tree found at the end of the iterative search. The minimum, mean and maximum weights correspond to the minimum, mean and maximum consistency indices found among all trees that are less than 12.99 dB away from the most parsimonious one (the mean weight is computed by rounding the mean value of *s* to the closest integer). Only two iterations were necessary to obtain stable weights. The five highlighted characters are those whose best estimates differ from their mean estimates.

Char.	Weight				Char.	Weight				Char.	Weight			
	best	min.	mean	max.		best	min.	mean	max.		best	min.	mean	max.
1	7.01	7.01	7.01	7.01	15	3.64	3.10	3.64	4.32	29	3.49	2.28	3.49	6.02
2	7.01	7.01	7.01	10.00	16	1.17	0.89	1.17	2.28	30	6.02	6.02	6.02	6.02
3	8.45	8.45	8.45	8.45	17	2.98	2.31	2.98	3.98	31	1.60	0.89	1.17	2.28
4	13.00	13.00	13.00	13.00	18	3.49	2.28	3.49	3.49	32	3.98	2.31	2.98	3.98
5	4.20	3.38	4.20	5.33	19	8.45	5.55	8.45	8.45	33	10.00	10.00	10.00	10.00
6	3.38	2.77	3.38	4.20	20	13.00	6.02	13.00	13.00	34	6.02	2.28	6.02	6.02
7	5.55	5.55	5.55	8.45	21	13.00	13.00	13.00	13.00	35	2.98	1.82	2.98	5.55
8	13.00	13.00	13.00	13.00	22	13.00	13.00	13.00	13.00	36	1.17	0.89	0.89	1.17
9	0.86	0.72	0.86	1.01	23	1.17	1.17	1.17	1.60	37	3.98	3.98	3.98	5.55
10	2.31	1.82	2.31	2.31	24	1.21	1.21	1.47	1.82	38	3.49	3.49	3.49	6.02
11	2.31	1.82	2.31	2.98	25	3.98	3.98	3.98	3.98	39	6.02	6.02	6.02	13.00
12	2.77	2.31	2.77	2.77	26	2.28	0.89	1.60	2.28	40	5.55	3.98	5.55	5.55
13	3.98	3.98	3.98	5.55	27	10.00	10.00	10.00	10.00	41	1.82	1.47	1.82	1.82
14	13.00	8.45	13.00	13.00	28	1.60	0.89	1.60	1.60	42	5.55	3.98	5.55	5.55

For each putative ADR, the strict consensus cladogram computed from all near-most-parsimonious cladograms was modified by pruning it from the other competing ADR hypotheses so as to compare its total length if the two taxa retained are in a sister-group relationship (with an unobserved common ancestor) or in an ancestor-descendant relationship (without any unobserved *ad hoc* entity). The ADR was forced in PAUP* by copying several times the putative ancestral species in a basal polytomy (see Fig. 4). If the total length difference was inferior to 13 dB, then it was considered that the null hypothesis (ADR) could not be rejected, and so was accepted. Since there can exist only one mother species (unless we assume that hybridization is likely), in the case where several possible ancestors could not be rejected, the less costly competing hypothesis would be accepted, however only with a credibility corresponding to the difference of the two costs (i.e. unfortunately necessarily inferior to 13 dB).

Results

Cladistic Analysis

In the initial step of the analysis only one most parsimonious tree was found, at a length of 155.67 steps. The consistency indices were computed for each character and accordingly reweighted using the Bayesian interpretation described above. The most

parsimonious trees were searched again, and then new weights were computed again. This second iteration had also only one most parsimonious tree with a length of 774.56 steps. The third iteration led to the tree and the same weights (see Table IV). The tree is described in Fig. 5A. I obtained 1474 trees with a score inferior to 787.55 steps. A strict consensus tree has been built from them (see Fig. 5B). Clades that appear on the first tree but not on the second have a Bremer support strictly inferior to 13 dB (i.e. < 95%) and so are not retained.

If one compares the two trees obtained in Fig. 5, the data could seem rather noisy. However, the evaluation of 10 million random trees with PAUP* show that none of them approaches the score of the most parsimonious tree. The mean score was 1422.71 steps, with a standard deviation of 46.34 steps and a skewness index $g_1 = -0.5873$ (or -0.4365 without weighting), which is far more negative than the critical values needed for such amounts of taxa and characters (Hillis and Huelsenbeck, 1992). According to Table IV, the best estimates of characters' weights may be a little doubtful for only five characters: 24, 26, 31, 32 and 36. However, eliminating them completely does not substantially alter the most parsimonious topology nor the consensus tree obtained from all near-most-parsimonious trees (data not shown).

The instability of the cladogram can thus be attributed to the unstable phylogenetic positions of

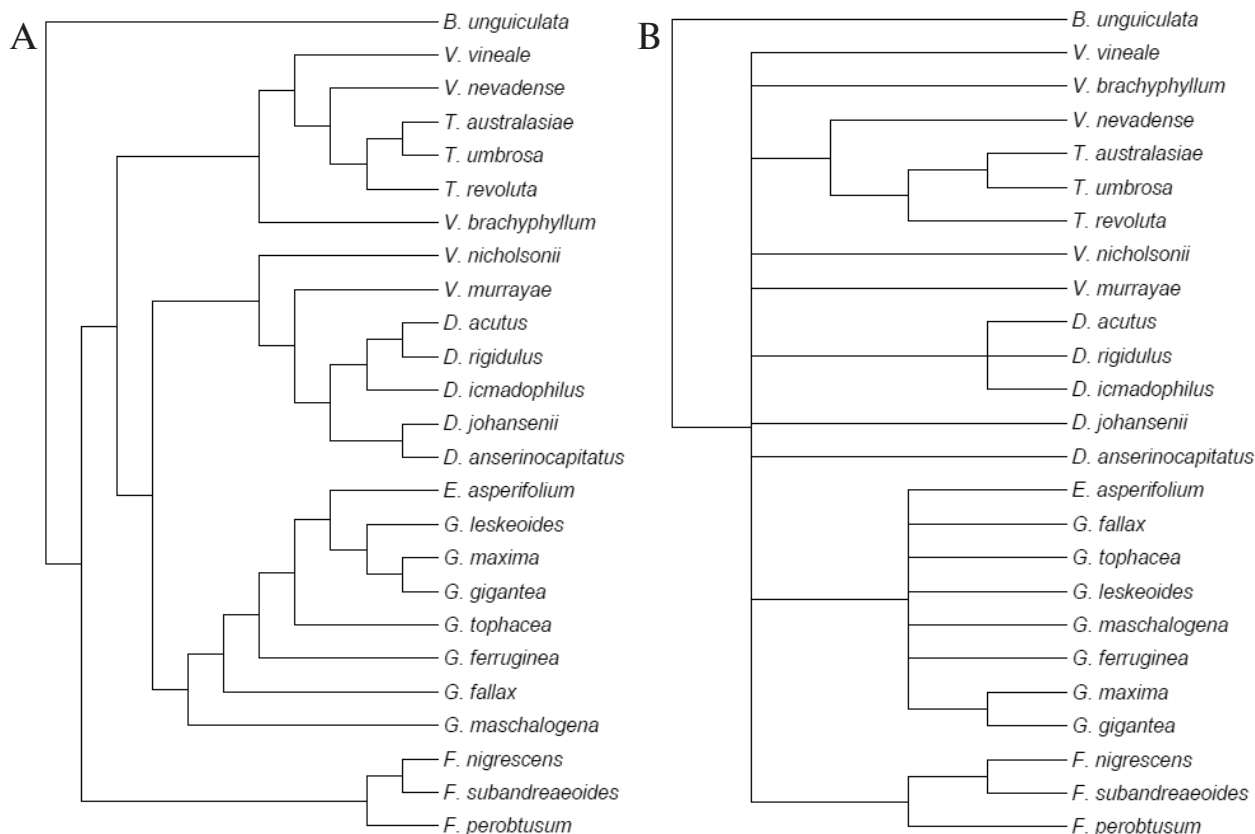


Fig. 5. The consensus trees. (A) Most parsimonious tree found after a heuristic search with successive weighting (stable after two iterations). (B) Strict consensus tree built from all the 1474 trees that are less than 12.99 dB away from the most parsimonious score.

mother species relative to their daughter species which may be wrongly grouped together because of convergent evolution or reversions. In other words, these are hard polytomies; they are not resolvable because of the clear implication that a single ancestor gave birth to several derived relatives. They are not soft polytomies that could be solved by using more and more data. The less resolved tree of **Fig. 5B** is therefore certainly more accurate, i.e. closer to truth, than is the more precise tree of **Fig. 5A**. One should not force the data into an artificial dichotomous scheme (Hull, 1979). The data are well structured, but not cladistically so.

Reconstruction of ADRs

As a first example, let us discuss the case of the relationship between *G. maxima* and *G. gigantea*. The total length of the strict consensus tree of all near-most-parsimonious trees is 947.25 steps. If we suppose that *G. maxima* is the ancestor of *G. gigantea* then we get a length of 954.87 steps (see **Fig. 4**), whereas we

get a length of 958.24 steps if we force *G. gigantea* to be the ancestor of *G. maxima*. In the first case, our ADR hypothesis only costs 7.62 dB, while in the second case it costs 10.99 dB. Neither hypothesis exceeds the threshold, but the first one is less costly and is therefore accepted while the second one is rejected. The fact that the resulting phylogenetic tree seems less parsimonious is an illusion caused by the lack of penalty accounting for unobserved ancestors. We should actually subtract 13 dB and realize that we have just won 5.38 dB. Simple ADRs like the above one should always be resolved first, before tackling more complex cases.

The second example I am now going to detail is the genus *Fuscobryum*, comprising three species. This case is simple to resolve because there are no polytomies, all dichotomies are supported at 13 dB or more. There are therefore only three tests to conduct: is any these three species the same as the node it is supposed to derive from? The three trees corresponding to *F. perobtusum*, *F. nigrescens*, and *F. subandreaeoides* have respectively

a length of 969.46 steps, 963.66 steps and 970.15 steps. Compared to the previous best tree of 954.87 steps, these hypotheses have a cost of 14.59 dB, 8.79 dB and 15.28 dB. The first and the third hypotheses exceed the threshold and are therefore rejected, but the second hypothesis is well below and is accepted. This means that *F. nigrescens* is the extant ancestor of *F. subandreaeoides*, but the last common ancestor of the three species remains unknown. It appears that this scenario is not exactly the one favoured by Zander (2014c) who inferred that *F. nigrescens* was the last common ancestor of the other two species.

The case of the genus *Trichostomopsis* is also simple to resolve. Only four tests are needed to assess potential ancestral species. All of them were rejected at around 25 dB except *T. australasiae* whose status of ancestor of *T. umbrosa* costs nothing at all. Zander (2014c) concluded that *T. australasiae* is the ancestor of both *T. umbrosa* and *T. revoluta*, but my test rejected this hypothesis at 24.34 dB. Unless the cladogram was misleadingly distorted by an artefact of long branch attraction we must conclude that *T. australasiae* and *T. revoluta* are derived from a shared unknown common ancestor. The case of *D. acutus*, *D. rigidulus* and *D. icmadophilus* is more ambiguous. Any of the three species may be the ancestor of the other two, with the respective costs 5.51 dB, 3.64 dB and 9.15 dB. There is however a small hint in favour of *D. rigidulus*, so we accept this hypothesis. These three species may be better considered subspecies rather than distinct species, but more data on morphology and possible reproductive isolation is needed to conclude definitively.

The clade *Geheebia-Exobryum* is a quite big polytomy and needs many tests in order to resolve it. I estimated the cost of ADR for each pair of species (excluding *G. gigantea* which we already know is directly derived from *G. maxima*). The results presented in **Table V** show us that *G. ferruginea* is certainly the ancestor of *G. maxima* (with a cost of 0 dB). In fact, both *G. ferruginea* and *G. fallax* could be the ancestor of all other species. However, *G. ferruginea* seems to be the direct ancestor of *E. asperifolium* while *G. fallax* would be the one of *G. leskeoides* and *G. maschalogenae*. The species *G. tophacea* seems slightly closer to *G. ferruginea* but has clear affinities with both *G. leskeoides* and *G. maschalogenae*, which leads to the suspicion of convergent evolution. Both *G. ferruginea* and *G. fallax* were tested as a potential direct ancestor of all remaining ones, but these hypotheses were rejected at 26.41 dB and 20.76 dB respectively. This means that there are

Table V. The evaluation of ADRs among species of the genera *Geheebia* and *Exobryum*. Each ADR hypothesis is tested against the corresponding SGR hypothesis. Their rejections are expressed in decibans, the threshold of 13 dB (> 95%). ADRs that were not rejected are highlighted.

	<i>E. as</i>	<i>G. fa</i>	<i>G. to</i>	<i>G. le</i>	<i>G. ms</i>	<i>G. fe</i>	<i>G. mx</i>
> <i>E. as</i>	-	8.78	12.96	17.94	23.48	2.46	26.4
> <i>G. fa</i>	61.78	-	28.19	25.23	29.89	8.01	36.53
> <i>G. to</i>	46.55	8.78	-	18.77	23.37	8.01	36.53
> <i>G. le</i>	34.61	3.98	9.32	-	21.63	6.41	29.18
> <i>G. ms</i>	32.22	2.98	5.95	13.66	-	7.15	29.18
> <i>G. fe</i>	56.23	8.78	28.19	26.61	33.01	-	28.52
> <i>G. mx</i>	34.75	6.96	15.27	15.86	31.84	0	-

certainly at least two independent lineages. *G. fallax* may be the most primitive species in the genus *Geheebia*. Indeed, if *E. asperifolium* is assumed to be descended from *G. ferruginea* then *G. fallax* as potential ancestral species is not rejected anymore, at 8.89 dB.

We need however to keep in mind that a hypothesis that is not rejected is not necessarily the best solution. Specifically, the pairwise tests are carried out in particular phylogenetic contexts, so that any modification in the neighbouring topology may change the results of the tests. All possible rearrangements were tried to place *G. maschalogenae*, *G. tophacea* and *G. leskeoides* in the right phylogenetic positions. Surprisingly, it appeared that the best score was obtained with *G. tophacea* as the sister species of *G. fallax*, both descended from an unknown founding mother species of the genus. The species *G. maschalogenae* is finally best considered a direct descendant of *G. tophacea* while *G. leskeoides* is probably a direct descendant of *G. fallax*, just like *G. ferruginea*. The genera *Vinealobryum* and *Didymodon* sensu stricto have also been studied but the detailed calculations are not shown here since the approach is exactly the same as above. Many rearrangements were tried and the best caulogram found is presented in **Fig. 6**.

Discussion

The Meaning of Parsimony

The length of the best caulogram found is 873.11 steps, which is 98.55 more steps than the most parsimonious cladogram at 774.56 steps. However, there are only 7 unobserved species instead of the 23 necessary internal nodes of the cladogram. Thus, we economized by eliminating 16 *ad hoc* entities. Since the procedure we used is equivalent to the fact of considering each additional entity as having a value of 13 dB, we can say that we economized 208 dB, which compensate

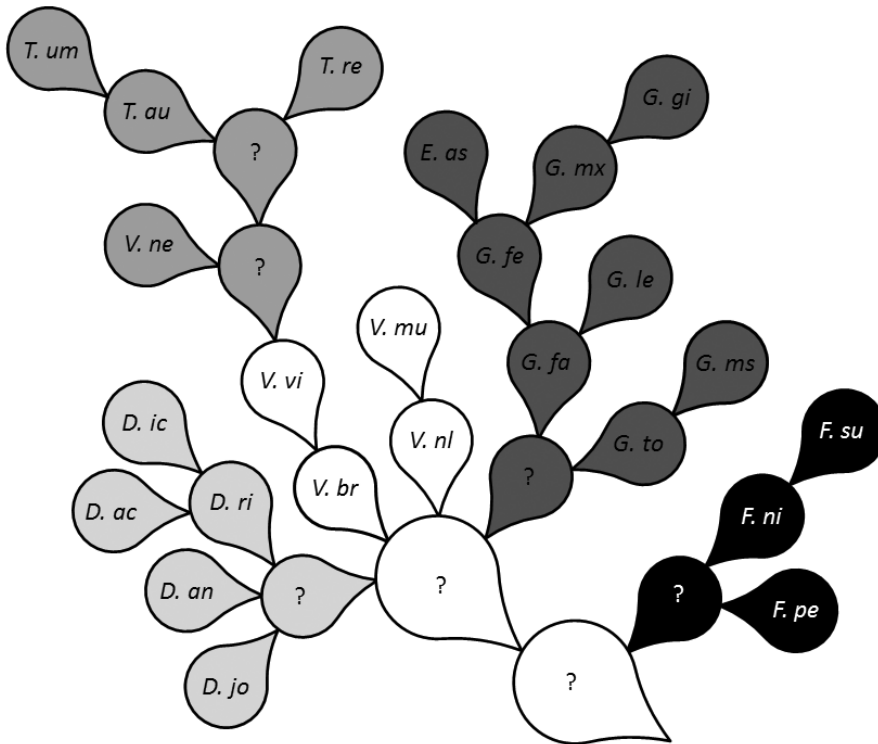


Fig. 6. The commagram depicting the ancestor-descendant relationships among the species of *Dydimodon sensu lato*. The corresponding caulogram has a length of 873.11 steps. The seven unknown predicted ancestral species are represented by question marks.

the previous loss. Our caulogram is therefore 109.45 steps more parsimonious than the most parsimonious cladogram.

Character reliability

In the dataset, although many characters may be considered quite reliable at about 8 or 10 dB, very few seem to be extremely reliable. Indeed, only three non autapomorphic characters have a weight of 13 dB (see Table IV). These are characters 14, 20 and 21, or respectively margin ornamentation of the leaf, the presence or absence of a bulge on the abaxial face, and the presence or absence of a thin-walled pad of cells on the adaxial face. The first one is a synapomorphy defining the large clade made of the genera *Vinealobryum*, *Didymodon sensu stricto*, *Geheebia*, *Exobryum* and *Trichostomopsis*, but not *Fuscobryum*. The second one defines the clade of the genus *Didymodon sensu stricto*. And the third one defines the clade made of the genus *Trichostomopsis* and the species *V. nevadense*. Thus, this species may be better considered a member of this genus (morphological convergence seems unlikely in this context because this trait has only evolved once with a support of at least 13 dB).

On the other hand many characters are very labile and are accordingly weighted at less than 2 dB (see Table IV). These are the characters 9, 16, 23, 24,

28, 31, 36, and 41. Three of them (24, 28, 31) concern the shape of different cells. This indicates that cell shape is generally not a good phylogenetic marker. The general shape of the leaf (9) is also to be considered a poorly reliable character. However, in both cases it is hard to know if this comes from a real tendency to evolve frequently or from the lack of a precise and reproducible morphometric measure (in which case the dataset should be corrected and re-analysed). The absence of a sporophyte (36) is not a reliable character either; therefore it seems that the loss of sexuality is very easy to evolve in Pottiaceae.

Finally, two character transformations are unique to one species each (autapomorphies). These are character 8 for *F. subandreaeoides* which uniquely have dimorphic leaves, and character 22 for *V. nevadense* which uniquely have multi-layered photosynthetic cells on the ventral costal surface. These two unique traits strongly indicate that these two species cannot be an ancestral to another one, which is also the case in Zander's analysis (see below).

Comparison with Zander's Results and Methods

These results only slightly differ from those of Zander (1998, 2013, 2014a, b, c, 2016). For example, the phylogenetic position of *E. asperifolium* in these previous studies, basal to the genus *Geheebia*, is arguably due a longbranch artefact. The construction of a UPGMA tree

indeed revealed that this species does not cluster within the genus *Geheebia* (data not shown), which is probably due to an accelerated evolution. The species *V. vineale* is here revealed as being derived, not ancestral to all the other species, but its rather conservative morphology explains the previous conclusion. The prime ancestor of this complex seems extinct or pseudoextinct, i.e. anagenetically transformed into another species. It is really not surprising given that this species is supposed to be quite old: the more time passes, the less likely a species remains unchanged. On the contrary, the more recent ancestors of this complex of species are still alive. In fact, exactly half of them (12) have unobserved direct ancestor.

The other differences with Zander's results are minor and are certainly due to the different set of data I used. They may merit a careful re-examination but I shall not comment them any further since the purpose of this study is primarily methodological. I will therefore not make any formal taxonomic decision. However, it seems now unclear that *E. asperifolium* deserves its own genus. The results also suggest that *V. nevadense* may be better treated as belonging to *Trichostomopsis*. A patrocadistic analysis may be conducted in order to test the consistency of the remaining genera (Stuessy and König, 2008). It seems that they fit more or less the definition of dissilience (Zander, 2013), that is a core species with several radiative species bursting from this core. The scheme seems however more complex than previously thought, including not only distinct lineages or stirps radiating from the same core, but also stirps arising from other stirps.

The main difference between Zander's methodology (Zander, 2014a, b, c) and mine is that he attempts to assess ADRs by seeking among available species which one is the more likely candidate to the status of ancestral species. The contrast between SGR and ADR hypotheses is not done explicitly. However, the ancestral species are not found directly, but through the successive elimination of the less probable candidates, that is those with obvious derived traits. The weighting of the different traits is also done in a Bayesian framework; however, he follows an intuitive scheme which ultimately relies on expertise, i.e. a long-standing experience with regard to trait stability. Moreover, he does not use only morphological data, but also other kinds of information like distribution or environment. For example *T. umbrosa* occurs in human environments contrary to the case with the other *Trichostomopsis* species, which indicates that it is probably not the ancestral species we are seeking (as it is

confirmed here). In its spirit, Zander's methodology is quite similar to *continuous track analysis* (Alroy, 1995), but it is far more holistic.

We must always seek a way to formalize our implicit (expert) reasoning into an explicit one so as to make it reproducible by others. My methodology necessitates no aprioristic expertise because it is more "mechanical", i.e. more algorithmically constrained. It may thus be qualified as more reductionist because I do not use some kind of informations like distribution or environment, and also because the measures of lability cannot be nuanced by some kind of a priori complex knowledge. It is therefore perhaps more reproducible, but above all completely doable by a computer. Yet, expertise is still needed a posteriori in order to interpret the results and evaluate their plausibility. Indeed, an expert can suspect a bias if for example the results are nonsensical even though they are numerically strongly supported (Hołyński, 2010).

Perspectives on Post-Phylogenetic Systematics (aka Modern Evolutionary Systematics)

Phylogenetic reconstruction methods are classically classified as either model-based or not. In the first case we have maximum likelihood (ML) and Bayesian inference (BI), while in the second case there exist maximum parsimony and compatibility technique, for example (Felsenstein, 1978, 1984; Farris, 1983). However, the term "model-based" is really ill-chosen. It misleadingly suggests that the classical cladistic analysis does not assume any evolutionary hypothesis and is therefore model-free as opposed to other techniques. This is certainly false (Friday, 1989), but we may still distinguish between those that explicitly specify a parametrized evolutionary dynamics and call them stochastic models, and those that do not and therefore call them non-parametric methods.

The main assumptions that all the above cited techniques share are that no ancestor was sampled and that speciation is strictly dichotomous, which are very strong assumptions. Even if the latter is not always lucidly claimed, it is a rather obvious consequence of the cladistic algorithm: since one cannot (in this framework) distinguish between a genuine polytomy and a lack of resolution then the data are always forced into artificial dichotomies (Hull, 1979). The method presented in this paper, as well as Zander's, may be considered non-parametric like the classical cladistic analysis. Even if weights are used here, they do not quantify a part of the evolutionary process but our confidence in our inferences. However, the two assumptions that no

ancestor was sampled and that speciation is strictly dichotomous are explicitly rejected because they are not realistic. Our post-phylogenetic analyses are therefore based on a distinct evolutionary model which is more empirically grounded (Zander, 2013). A stochastic approach that would also reject these two assumptions is conceivable and should actively be sought.

Some cladists have insisted that ancestral species cannot be recognized as such, and that ADRs are therefore unknowable (Nelson, 1973; Farris, 1976). These claims seem to be fundamentally based on a particular version of nominalist epistemology (Aubert, 2015). As a consequence, anybody that does not endorse this very philosophy could simply refuse these assertions without further justification, scientific realists for example (Sankey, 2001). Hull (1979) warned: "In general, I think it is very bad strategy for proponents of a particular scientific research program to stake their future on epistemological considerations, especially on our inability to know something." Indeed, epistemology should not be seen as an a priori set of constraints that dictate what science can do or cannot do. Epistemology should be handled in a more empirical manner, mainly in order to take a global view a posteriori on the achievements of science. A one-way relationship between epistemology and science is really a kind of sterilizing metaphysics which has its modern origins in German Idealism (especially Kant's *Critique of Pure Reason*). Only a genuine dialogue between both can be fruitful: this is dialectics, not metaphysics.

As a consequence of their hypothetico-deductivist framework, cladists seem afraid of type I errors, i.e. not being able to reject the null hypothesis (the relationship is an ADR) when it is wrong, so they always reject it by default (all relationships are SGRs), which logically leads to many type II errors (many of the supposed SGRs are in fact ADRs). It is necessary to overcome this by proposing new methods of caulogram reconstruction, and also new methods in taxonomy. Indeed, as opposed to cladograms it is widely recognized that caulograms cannot be directly translated into a hierarchical classification because of the named ancestors it includes. Contrary to SGRs, ADRs highlight that "Evolution is paraphyly all the way" (Brummitt, 2002). This is obviously true for species: paraphyletic species (like *G. fallax*) are not less real than holophyletic ones, paraphyly does not preclude them from being "natural entities" (i.e. biological species). This is also true for higher taxa which cannot all be made holophyletic (Brummitt, 2002; Aubert, 2015). More importantly, paraphyly or holophyly have really no influence on the

evolutionary dynamics of a genus or a family (Sepkoski, 1996).

Conclusion

The results of this study support the following major conclusions:

- Ancestor-descendant relationships (ADRs) are knowable;
- *Didymodon*, *Geheebia*, *Fuscobryum*, *Vinealobryum* and *Trichostomopsis* are dissiliant genera (from Latin *dissiliens*, which means "bursting apart", see Zander, 2013);
- Consistency index can be interpreted in a Bayesian framework, measured in decibans and used as a weighting scheme for cladistic parsimony analysis;
- Bremer support can also be interpreted in a Bayesian manner and measured in decibans;
- There exist simple and objective ways to transform a cladogram depicting only putative SGRs into a caulogram depicting real ADRs (and real SGRs).

Another obvious consequence of this study is that autapomorphies or "uninformative" cladistic characters and character states should never be pruned from matrices. This would strongly bias the data for any post-cladistic analysis. In the same manner, labile and even very labile characters should be studied and added to matrices so that the data are as complete as possible. These requirements are also needed in order to not bias stochastic cladogram reconstructions such as likelihood methods or Bayesian inference.

My procedure, as well as Zander's, are limited heuristics and are not guaranteed to find optimal solutions. They are rather constraining guides that help organize the data, so that the systematists can reconstruct an evolutionary scenario and make taxonomic decisions accordingly. My study clearly revealed that the length of a cladogram is not the sole parameter we need to minimize, but that the minimization of unobserved entities is also an important parameter to take into account. This naturally led to an equivalence relationship between these two parameters, which can be translated into a new objective function that a specialized algorithm could minimize by trying many rearrangements of the possible topologies. This function is simply as it follows:

$$S = L + 13n$$

S is the score we want to minimize, L is the total length of the tree (on the deciban scale) and n is the number of unobserved ancestral species required by the tree topology. The minimization of this function is not

feasible manually by a human being because of the huge number of possible topologies. The best way to proceed for now is to reduce the problem by selecting small areas in a cladogram and resolve them individually as it was done in this study:

1. Collect a morphological data set as complete as possible, including autapomorphies and other “cladistically uninformative” characters;
2. Find the most parsimonious cladogram through an iterative procedure which uses Bayesian weights measured in decibans (dB);
3. Collapse all the branches whose Bremer support is not at least 13 dB (> 95%);
4. In the distinct parts of this strict consensus cladogram, try all possible pairwise arrangements by contrasting the shorter SGR trees with the corresponding longer ADR trees;
5. ADR hypotheses should be rejected only if the trees of corresponding SGR hypotheses are at least 13 dB shorter;
6. In order to address huge polytomies, all possible arrangements (or at least all likely ones) should be tried to find the shortest configuration, but taking into account the number of unnamed ancestors eliminated (see above formula).

Note that an automatized program could test many more topologies than a human being and do so very quickly. Since several decades, whatever the analyses conducted, the amount of data has become huge and phylogenetic studies therefore rely heavily on computers. As a consequence, phylogeneticists tend to do only what available software allows them to do. That is why the methods for assessing ADRs needed to be first formalized – as it was done in this study, or by Zander (2013, 2014a, b, c, 2016) – and then must be implemented in new software or in new packages for already existing software. I cannot stress enough the importance for evolutionary systematics to head towards informatics.

Acknowledgements

I would like to sincerely thank Richard Zander who provided me the initial files and also some insightful comments on an early draft of this paper. I would also like to thank the two reviewers who brought many useful suggestions.

Compliance with Ethical Standards

The author Damien Aubert declares that he has no conflict of interest.

Annex

Gametophyte Characters:

1. Colour of plants:
 0. green to red-brown
 1. green, often blackish (never reddish)
 2. glossy green or very bright green
 3. yellow- or orange-brown to red- or black-brown.
2. Stem height (**Ordered**):
 0. less than 1 cm.
 1. 1–2 cm.
 2. 2.5–3.5 cm.
 3. more than 4 cm.
3. Stem central strand (**Ordered**):
 0. always present, strong.
 1. variable, present or absent.
 2. weak or absent.
4. Axillary hairs:
 0. with all cells hyaline.
 1. with brown basal cells.
5. Specialized asexual reproduction:
 0. absent, or deciduous leaf apex.
 1. present, axillary, multicellular.
 2. present, axillary, unicellular.
 3. tubers present on rhizoids.

Leaf Characters:

6. Leaf stance when dry:
 0. spirally twisted, weakly spreading.
 1. erect-appressed, not twisted.
 2. appressed-incurved.
 3. catenulate.
7. Leaf stance when moist and removed (**Ordered**):
 0. spreading to weakly recurved, commonly lying flat.
 1. spreading to spreading-recurved and weakly keeled.
 2. strongly recurved and keeled.
8. Leaf polymorphism:
 0. monomorphic.
 1. dimorphic, the smaller leaves strongly concave in series in some parts of the plant.
9. Leaf shape (**Ordered**):
 0. short- to long-lanceolate.
 1. variable or intermediate.
 2. deltoid to ovate-lanceolate.
10. Leaf channelling above mid-leaf (**Ordered**):
 0. with a narrow groove along ventral surface of costa.
 1. variable or intermediate.
 2. broadly concave to nearly plane.
11. Leaf base:

- 0. scarcely differentiated to oblong, gradually or quickly narrowed to the insertion.
 - 1. well-differentiated ovate, possibly winged or auriculate.
 - 2. well-differentiated rectangular, squared.
12. Leaf length (**Ordered**):
- 0. 0.9–1.5 mm.
 - 1. 1.2–2.5 mm.
 - 2. 2.0–3.5 mm.
 - 3. 3.5–5.0 mm.

Margin and Apex Characters:

13. Margin flexion:
- 0. plane or nearly so.
 - 1. recurved in lower 1/2 or 3/4.
 - 2. recurved or revolute to apex.
14. Margin ornamentation:
- 0. minutely crenulate.
 - 1. entire (smooth or papillose).
 - 2. often notched laterally.
15. Apex shape:
- 0. rounded or obtuse (only occasionally broadly acute).
 - 1. broadly to narrowly acute, variable.
 - 2. very long acuminate, whiplike.
 - 3. narrowly acute, thickened, but not deciduous.
 - 4. swollen as a propagulum, deciduous.
16. Apex ending:
- 0. of a clear conical cell or mucro.
 - 1. rounded and obscure.

Costa Characters:

17. Costa ending:
- 0. clearly reccurrent, usually as a short mucro.
 - 1. percurrent or ending to 2 cells before the apex.
 - 2. ending more than 2 cells before the apex.
18. Costa hydroids:
- 0. present.
 - 1. absent.
19. Well-developed specimens with ventral (adaxial) costal stereid band (**Ordered**):
- 0. present.
 - 1. weak, variable.
 - 2. absent.
20. Costa abaxial shape:
- 0. clear dorsal bulge present.
 - 1. dorsally flattened.
21. Adaxial thin-walled pad of cells:
- 0. absent.
 - 1. present.

22. Multi-layered photosynthetic cells on ventral surface of mid-costa:
- 0. absent.
 - 1. present.
23. Costa superficial ventral cells:
- 0. elongate, more than 2:1.
 - 1. quadrate or very short-rectangular (at least near apex).
24. Costa width at midleaf (**Ordered**):
- 0. 2–3(–4) cells.
 - 1. 4–5(–6) cells.
 - 2. (5–)6–9 cells.
25. Guide cells (**Ordered**):
- 0. nearly always one layer.
 - 1. variable (one or two layers).
 - 2. two (or three) layers.

Basal Characters:

26. Basal laminal cells:
- 0. differentiated medially or across the leaf.
 - 1. weakly differentiated.
27. Basal laminal cell walls (**Ordered**):
- 0. very thin, hyaline, possibly perforated by transverse slits.
 - 1. thin to weakly thickened, not perforated.
 - 2. usually thick, not perforated.
 - 3. always thick and porose.
28. Basal laminal cell shape:
- 0. quadrate or very short-rectangular.
 - 1. rectangular.

Upper Laminal Characters:

29. Upper laminal cell width:
- 0. (7–)8–10(–12) μm .
 - 1. (11–)13–15 μm .
30. Upper laminal cell shape:
- 0. quadrate or very short-rectangular, rarely transversely elongate.
 - 1. in rather distinct rows, often longitudinally elongate.
31. Upper laminal cell lumens:
- 0. mostly rounded.
 - 1. mostly angular.
32. Papillae:
- 0. usually present, simple, possibly bifid or multifid, (1–)2–3 per lumen.
 - 1. low, massive and lens-shaped or scablike.
 - 2. essentially absent, rarely simple, weak and domelike, one per lumen

33. Upper laminal cell walls:
0. thin to evenly thickened, convex on both sides of lamina.
 1. irregularly thickened, convex on both sides of lamina.
 2. evenly thickened, moderately bulging on both sides or only abaxially.
 3. irregularly thickened and trigonous, convex on both sides.
34. Upper laminal cell layering:
0. unistratose, rarely bistratose in patches.
 1. bistratose along margins, at apex, or often bistratose in patches.
35. KOH colour reaction of upper laminal cells:
0. yellow or orange, occasionally negative, never brownish or deep red.
 1. variously negative or yellow- or orange- or red-brown.
 2. brick-red, occasionally dark red-brown, never yellow or orange or negative.

Sporophyte Characters:

36. Sporophyte:
0. present.
 1. absent.
37. Seta length (**Ordered**):
0. 0.3–0.6 cm.
 1. 0.6–1(–1.2) cm.
 2. (0.8–)1–1.7(–2.5) cm.
38. Capsule size:
0. 0.5–1.5 mm.
 1. 1.25–2.5 mm.
39. Number of peristome teeth, when present:
0. 32 teeth, or possibly 16 cleft to base.
 1. 16 teeth.
40. Tooth shape:
0. usually twisted
 1. usually straight.
 2. essentially absent.
41. Tooth length, when present (**Ordered**):
0. 50–300 μm .
 1. 300–800 μm .
 2. 800–1500 μm .
42. Spore size (**Ordered**):
0. 8–12 μm .
 1. 13–15 μm .
 2. 16–20 μm .

REFERENCES

- Aldous D., Popovic L. A critical branching process model for biodiversity. *Adv. Appl. Probab.*, 2005, 37: 1094–1115.
- Aldous D.J., Krikun M.A., Popovic L. Five statistical questions about the tree of life. *Syst. Biol.*, 2011, 60: 318–328. doi: 10.1093/sysbio/syr008
- Alroy J. Continuous track analysis: a new phylogenetic and biogeographic method. *Syst. Biol.*, 1995, 44: 152–178. doi: 10.2307/2413704
- Ashlock P.D. Monophyly and associated terms. *Syst. Biol.*, 1971, 20(1): 63–69. doi: 10.1093/sysbio/20.1.63
- Aubert D. A formal analysis of phylogenetic terminology: Towards a reconsideration of the current paradigm in systematics. *Phytoneuron*, 2015, 2015-66: 1–54.
- Aze T., Ezard T.H.G., Purvis A., Coxall H.K., Stewart D.R.M., Wade B.S., Pearson P.N. A phylogeny of Cenozoic macroperforate planktonic foraminifera from fossil data. *Biol. Rev. Camb. Philos. Soc.*, 2011, 86: 900–927. doi: 10.1111/j.1469-185X.2011.00178.x
- Bapst D.W., Hopkins M.J. Comparing cal3 and other a posteriori time-scaling approaches in a case study with the pteroccephaliid trilobites. *Paleobiology*, 2017, 43: 49–67. doi: 10.1017/pab.2016.34
- Bremer K. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution*, 1988, 42: 795–803. doi: 10.2307/2408870
- Bremer K. Branch support and tree stability. *Cladistics*, 1994, 10: 295–304. doi: 10.1111/j.1096-0031.1994.tb00179.x
- Brummitt R.K. How to chop up a tree. *Taxon*, 2002, 51: 31–41.
- Crawford D.J. Progenitor-derivative species pairs and plant speciation. *Taxon*, 2010, 59: 1413–1423.
- Crisp M.D., Chandler G.T. Paraphyletic species. *Telopea*, 1996, 6: 813–844.
- de Queiroz K., Donoghue M.J. Phylogenetic systematics and the species problem. *Cladistics*, 1988, 4: 317–338. doi: 10.1111/j.1096-0031.1988.tb00518.x
- Donoghue M.J. A critique of the biological species concept and recommendations for a phylogenetic alternative. *The Bryologist*, 1985, 88: 172–181. doi: 10.2307/3243026
- Farris J.S. The retention index and the rescaled consistency index. *Cladistics*, 1989, 5: 417–419. doi: 10.1111/j.1096-0031.1989.tb00573.x
- Farris J.S. A successive approximations approach to character weighting. *Syst. Zool.*, 1969, 18: 374–385. doi: 10.2307/2412182
- Farris J.S. The logical basis of phylogenetic analysis. In: *Advances in Cladistics, II*. Eds Platnick N.I., Funk V.A. New York: Columbia University Press, 1983, pp. 7–36.
- Farris J.S. Phylogenetic classification of fossils with recent species. *Syst. Biol.*, 1976, 25: 271–282. doi: 10.2307/2412495
- Felsenstein J. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.*, 1978, 27: 401–410. doi: 10.2307/2412923
- Felsenstein J. The statistical approach to inferring evolutionary trees and what it tells us about parsimony and compatibility. In: *Cladistics: Perspectives on the*

- Reconstruction of Evolutionary History*. Eds Duncan T., Stuessy T.F. Columbia University Press, New York, 1984, pp. 169–191.
- Footo M. On the probability of ancestors in the fossil record. *Paleobiology*, 1996, 22: 141–151. doi: 10.1017/S0094837300016146
- Friday A. Quantitative aspects of the estimation of evolutionary trees. *Folia Primatol.* (Basel), 1989, 53: 221–234. doi: 10.1159/000156418
- Funk D.J., Omland K.E. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annu. Rev. Ecol. Evol. Syst.*, 2003, 34: 397–423. doi: 10.1146/annurev.ecolsys.34.011802.132421
- Gee H. *Deep Time: Cladistics, the Revolution in Evolution*. Fourth Estate, London, UK, 2000.
- Good I.J. Studies in the History of Probability and Statistics. XXXVII A.M. Turing's statistical work in World War II. *Biometrika*, 1979, 66: 393–396. doi: 10.2307/2335677
- Good I.J. Weight of evidence: A brief survey. *Bayesian Stat.*, 1985, 2: 249–270.
- Hennig W. *Grundzüge einer Theorie der phylogenetischen Systematik*. Deutscher Zentralverlag, Berlin, 1950.
- Hennig W. *Phylogenetic Systematics*. University of Illinois Press, Urbana, 1966.
- Hillis D.M., Huelsenbeck J.P. Signal, noise, and reliability in molecular phylogenetic analyses. *J. Hered.*, 1992, 83: 189–195.
- Holyński R.B. Is paraphyly indication of poor taxonomy? – Open letter to Drs. Carvalho and Ebach. *Munis Ent. Zool.*, 2010, 5 (Suppl.): 825–829.
- Hull D.L. The limits of cladism. *Syst. Biol.*, 1979, 28: 416–440. doi: 10.2307/sysbio/28.4.416
- Lee M.S.Y. Species concepts and the recognition of ancestors. *Hist. Biol.*, 1995, 10: 329–339. doi: 10.1080/10292389509380528
- Levin D.A. Local speciation in plants: The rule not the exception. *Syst. Bot.*, 1993, 18: 197–208. doi: 10.2307/2419397
- Nelson G.J. "Monophyly again?" – A reply to P.D. Ashlock. *Syst. Biol.*, 1973, 22: 310–312. doi: 10.1093/sysbio/22.3.310
- Paul C.R.C. The recognition of ancestors. *Hist. Biol.*, 1992, 6: 239–250. doi: 10.1080/10292389209380433
- Podani J. Tree thinking, time and topology: comments on the interpretation of tree diagrams in evolutionary/phylogenetic systematics. *Cladistics*, 2013, 29: 315–327. doi: 10.1111/j.1096-0031.2012.00423.x
- Prothero D.R., Lazarus D.B. Planktonic microfossils and the recognition of ancestors. *Syst. Zool.*, 1980, 29: 119–129. doi: 10.2307/2412642
- Pyron R.A., Costa G.C., Patten M.A., Burbrink F.T. Phylogenetic niche conservatism and the evolutionary basis of ecological speciation. *Biol. Rev. Camb. Philos. Soc.*, 2015, 90: 1248–1262. doi: 10.1111/brv.12154
- Rieseberg L.H., Brouillet L. Are many plant species paraphyletic? *Taxon*, 1994, 43: 21–32. doi: 10.2307/1223457
- Ross H.A. The incidence of species-level paraphyly in animals: A re-assessment. *Mol. Phylogenet. Evol.*, 2014, 76: 10–17. doi: 10.1016/j.ympev.2014.02.021
- Sankey H. Scientific Realism: An Elaboration and a Defence. *Theor. J. Soc. Polit. Theory*, 2001, 35–54.
- Sepkoski J.J. Competition in macroevolution: the double wedge revisited. In: *Evolutionary Paleobiology*. Eds Jablonski D., Erwin D.H., Lipps J.H. University of Chicago Press, Chicago, USA, 1996, pp. 211–255.
- Stuessy T.F., König C. Patrocladistic classification. *Taxon*, 2008, 57: 594–601.
- Tsai C.-H., Fordyce R.E. Ancestor-descendant relationships in evolution: origin of the extant pygmy right whale, *Caperea marginata*. *Biol. Lett.*, 2015, 11: 20140875. doi: 10.1098/rsbl.2014.0875
- Vanderpoorten A., Long D.G. Budding speciation and neotropical origin of the Azorean endemic liverwort, *Leptocycphus azoricus*. *Mol. Phylogenet. Evol.*, 2006, 40: 73–83. doi: 10.1016/j.ympev.2006.02.013
- Zander R.H. A phylogrammatic evolutionary analysis of the moss genus *Didymodon* in North America north of Mexico. *Bull. Buffalo Soc. Nat. Sci.*, 1998, 36: 81–115.
- Zander R.H. Structuralism in Phylogenetic Systematics. *Biol. Theory*, 2011, 5: 383–394.
- Zander R.H. *A Framework for Post-Phylogenetic Systematics*. Zetetic Publications, St. Louis, 2013.
- Zander R.H. Classical determination of monophyly, exemplified with *Didymodon* s. lat. (Bryophyta). Part 1 of 3, synopsis and simplified concepts. *Phytoneuron*, 2014a, 2014-78: 1–7.
- Zander R.H. Classical determination of monophyly, exemplified with *Didymodon* s. lat. (Bryophyta). Part 2 of 3, concepts. *Phytoneuron*, 2014b, 2014-79: 1–23.
- Zander R.H. Classical determination of monophyly, exemplified with *Didymodon* s. lat. (Bryophyta). Part 3 of 3, analysis. *Phytoneuron*, 2014c, 2014-80: 1–19.
- Zander R.H. Macrosystematics of *Didymodon* sensu lato (*Pottiaceae*, *Bryophyta*) using an analytic key and information theory. *Ukr. Bot. J.*, 2016, 73: 319–332. doi: 10.15407/ukrbotj73.04.319

Recommended for publication
by S.L. Mosyakin

Submitted 29.03.2017

Обер Д. Простий підхід до оцінки зв'язків предок-нащадок на основі парсимонії. Укр. бот. журн., 2017, 74(2): 103–121. Академія Клермон-Феран, Міністерство національної освіти Франції

Однією з головних цілей систематики є реконструкція Древа життя. Півстоліття тому проривом стала кладистика, яка виявилась важливим кроком на шляху до цієї мети, що дозволив нам визначити зв'язки спорідненості між видами, абстрактний вид спорідненості. На жаль, філософія кладизму не дозволила просунутися далі в пошуку більш реальних взаємозв'язків, таких як зв'язок предок-нащадок, який, відповідно до дарвінівської еволюційної концепції, ймовірно, є основним типом зв'язків на Дереві життя. У статті описана проста методика на основі парсимонії, яка може бути використана для перетворення класичної кладограми у справжнє філогенетичне дерево, тобто в каулограму. Ця методика полягає у видаленні якомога більшого числа неспостережених і неіменованих вузлів та в їх заміні реально визначеними та іменованими видами. Як в цій методиці, так і при побудові класичної кладограми, для оцінки надійності ознаки використовується нова байєсівська нестохастична схема зважування. Для ілюстрації проводиться оцінка взаємозв'язків між видами мохів роду *Didymodon sensu lato (Pottiaceae)* з обговоренням отриманої каулограми шляхом її співставлення з попередньою методологією відповідно до еволюційних публікацій. На закінчення робиться висновок про недоцільність суворого додержання кладистичної епістемології та необхідність пошуку нових формальних методів для встановлення предкових видів, а також і предкових таксонів вищих рангів.

Ключові слова: предок, байєсівський аналіз, індекс Бремера, еволюційна систематика, парсимонія, зважування

Обер Д. Простой подход к оценке связей предок-потомок на основании парсимонии. Укр. бот. журн., 2017, 74(2): 103–121.

Академія Клермон-Ферран, Міністерство національного образования Франції

Одной из главных целей систематики является реконструкция Древа жизни. Полвека тому назад прорывом стала кладистика, которая оказалась важным шагом на пути к этой цели, позволившим нам оценить относительные родственные связи между видами, абстрактный тип родства. К сожалению, философия кладизма не позволила продвинуться дальше в поиске более реальных родственных взаимосвязей, таких как связь предок-потомок, которая, согласно дарвинистской эволюционной концепции, является, очевидно, основополагающим типом связей на Древе жизни. В статье описана простая методика на основе парсимонии, которая может быть использована для преобразования классической кладограммы в настоящее филогенетическое дерево, то есть в каулограмму. Эта методика состоит в удалении как можно большего числа ненаблюдаемых и именованных узлов и в замене их наблюдаемыми и именованными видами. Как в этой методике, так и при построении классической кладограммы, для оценки надежности признака используется новая байесовская нестохастическая схема взвешивания. В качестве иллюстрации проводится оценка взаимосвязей между видами мохов рода *Didymodon sensu lato (Pottiaceae)* с обсуждением итоговой каулограммы путем сопоставления ее с ранее применявшейся методологией по материалам эволюционных публикаций. В заключение делается вывод о несостоятельности строгого следования кладистической эпистемологии и необходимости поиска новых формальных методов для установления предковых видов, а также предковых таксонов более высоких рангов.

Ключевые слова: предок, байесовский анализ, индекс Бремера, эволюционная систематика, парсимония, взвешивание