

Experimental systematics: sensitivity of cladistic methods to polarization and character ordering schemes

Valentin Rineau¹, Anaïs Grand², René Zaragüeta³, Michel Laurin^{1,4}

¹ UMR 7207 CR2P, Sorbonne Universités, CNRS/MNHN/UPMC, Muséum national d'Histoire naturelle, Bâtiment de Géologie, Case postale 48, 43 rue Buffon, F-75231 Paris cedex 05, France

² Service du patrimoine naturel, Muséum national d'Histoire naturelle, Bâtiment reptiles et amphibiens, 25 rue Cuvier, 75005 Paris, France

³ UMR 7205 ISYEB, CNRS/MNHN/UPMC/EPHE, Muséum national d'Histoire naturelle, Bâtiment de Géologie, Case postale 48, 43 rue Buffon, F-75231 Paris cedex 05, France

⁴ E-mail: laurin@mnhn.fr

Key words: branch length, evolutionary model, hierarchy, ordered characters, outgroup polarization, parsimony, reversals, three-taxon analysis, three-taxon statements, unordered characters

Abstract

Phenotypic characters are essential to study the evolution of extant and extinct life forms and to reconstruct the tree of life. Inside the cladistics theory, parsimony is used by a large majority of systematists working on phenotypic characters, whereas 3ta is much less widespread but has triggered important debates. Many important differences in the interpretation of the cladistic theory exist between these methods, *e.g.* meaning and treatment of reversals, character representation as 'data-matrices' in parsimony (ordered and unordered), and as rooted trees (hierarchies) in 3ta. Although 3ta has received severe criticism, mostly focused in the use of software intended to be used in parsimony, only a few empirical studies have compared these methods so far. We present the results of simulations of the evolution of phenotypic traits under a Brownian motion model to characterize differences in sensitivity between parsimony and 3ta to (1) outgroup branch length, which affects the reliability of ancestral character state estimates, (2) character state ordering scheme, and (3) ingroup branch lengths that reflect the geological age of studied taxa. Our results show that the 'nihilistic' attitude of leaving multistate characters unordered when criteria to order are available (*e.g.*, similarity, ontogeny, etc...) can decrease resolving power of the method (by 13.4% to 29.3%) and increase the occurrence of artefactual clades (by 5% to 15.6%). Increasing outgroup branch length significantly decreases resolving power and increases artefactual resolution, at least for paleontological trees. All simulations show that ordered parsimony is always superior to 3ta in tested parameter space. These results depend on the assumption in parsimony that reversals (as implied by the Brownian motion, as in most other models) can be evidence for the support of a clade a posteriori from an analysis or a priori on simulations with a known pattern. We discuss implications of these points of view compared to the assumption inherent in 3ta (*i.e.*, that reversals should not support a clade as other synapomorphies do) on evolutionary models.

Contents

| | |
|---|-----|
| Introduction | 129 |
| Material and methods | 135 |
| <i>Simulated character sets</i> | 135 |
| <i>Tree searches</i> | 137 |
| <i>Tree comparisons</i> | 137 |
| <i>Testing the results</i> | 137 |
| Results | 138 |
| Discussion | 139 |
| <i>Brownian motion in cladistics</i> | 139 |
| <i>Reversals in phylogenetics</i> | 139 |
| <i>Uncertain polarization</i> | 143 |
| <i>Character states and ordering schemes</i> | 143 |
| <i>Tree shape and branch length</i> | 143 |
| <i>Implications on simulation-based studies and evolutionary models</i> | 144 |
| Conclusion | 145 |
| Acknowledgements | 146 |
| References | 146 |

Introduction

The reconstruction of the Tree of Life is one of the most important quests in natural sciences, and it is essential for various biological and geological fields. Phylogenies are widely used in comparative and evolutionary biology (Hennig, 1966; Felsenstein, 1985) and even in conservation biology (Faith, 1992). Even though most contemporary phylogenetic studies of extant taxa are based on DNA sequence characters, phenotypic traits continue to play an important role because molecular sequences are unavailable for most long-extinct organisms, given that

the oldest ‘ancient’ DNA extracted so far is less than 1 Ma old (Callaway, 2011). According to various estimates of extant and extinct biodiversity, and depending on how species are conceptualized, approximately 99% to 99.9% of life forms that have existed on Earth are now extinct (Miller and Foote, 1996; Newman and Sibani, 1999; Alroy, 2001). Thus, it is important to continue using and improving methods of phylogenetic inference from phenotypic data. The dominant phylogenetic theory for phenotypic characters is cladistics. This theory is based on the assumptions that the epistemic access we have to discover monophyletic taxa is through phylogenetic arguments called characters. Characters are based on hypotheses of homology, *i.e.* hypotheses that relate the observation of different features as being the same. This identity is explained by their evolutionary origin as parts evolved from the same part in the exclusive (last) common ancestor. The method of argumentation is different in parsimony analysis and three-taxon analysis (hereafter, 3ta). In parsimony analysis, characters-states transform into one another, which is consistent with Hennig’s original description of cladistic theory. In 3ta, character-states differentiate from one another, which is consistent with the relationships of taxa represented as hierarchies: if homologs are parts of taxa and taxa differentiate, then character-states should also differentiate. Theoretically, the criterion of choice among contradictory hypotheses is maximizing congruence. Congruence has been interpreted as minimizing a particular tree-distance (in parsimony method) or as maximizing compatibility among characters (as in the compatibility method; not tested in this paper). In 3ta, the search for optimal trees maximising parsimony or compatibility of 3ts gives exactly the same results (Wilkinson, 1994): this shows that maximizing the treatment of characters is contained in the method of treatment of homology and homoplasy of 3ta and parsimony, and that the representation of characters as matrices, graphs or hierarchies is only a consequence of this treatment.

According to various authors (Chappill, 1989; Stevens, 1991; Thiele, 1993; Wiens, 2001), most morphological characters are fundamentally quantitative. Continuous characters are usually discretized because most cladistic software cannot handle them directly. Several methods have been proposed for discretizing characters, such as segment coding (Simon, 1983; Farris, 1990) and gap coding (Mickeyvich and Johnson, 1976; Almeida and Bisby, 1984; Archie, 1985) to limit the arbitrariness of the state delimitation, even though these do not usually capture all the phylogenetic infor-

mation (Laurin and Germain, 2011) unless ordering schemes are used to order the series of states or set transition costs, and a distinct state is recognized for each taxon (Wilkinson, 1992; Wiens, 2001). Thus, the coding, as a formalization of character hypothesis of the analysis (ordination, discretization, etc...) is of primary importance for phylogenetic inference. Characters can be separated into at least two classes (Fig. 1). The first class comprises parsimony characters, treated as partitioned graphs, oriented or not, with cycles or not. Fig. 1 shows to ‘extreme’ cases that will be dis-

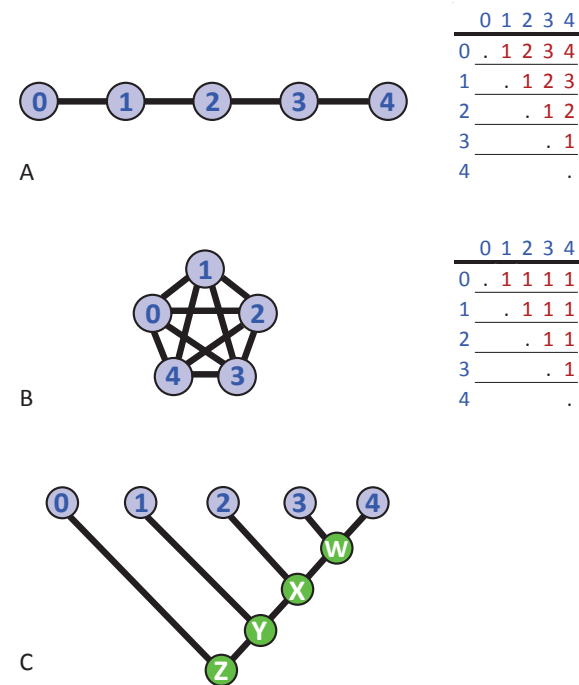


Fig. 1. Three examples of character representation: ordered in parsimony (A; additive) and unordered in parsimony (B; non-additive) form the class of parsimony characters, associated with their stepmatrix (Sankoff and Rousseau, 1975) necessary in a matrix representation, and the other classes of character trees, hierarchies treated in 3ta (C). A symbolizes ordered character states, where transformation costs are given by the stepmatrix to the right. B represents unordered states as an unrooted tree and where all transformations between states have the same cost and C represents a character tree, where nodes are homology hypotheses under the hierarchy. The differences in their representations (graph, unrooted tree, hierarchy) traduce both their content and the way in which corresponding methods treat them (how they deal with incongruence they can generate; for example, in parsimony graphs of Figs 1A and B, a step can correspond to transformation from 1 to 2 or from 2 to 1, which make no sense in the hierarchical representation 1C, symbolizing a 3ta treatment). Letters and numbers (0-5, Z-W) represent named nodes as homologs.

cussed subsequently, as unordered characters (Fig. 1A), which imply equal transition cost between all states (also called ‘maximally connected’; Slowinski, 1993), and ordered characters (Fig. 1B), in which relationships between states are typically depicted by linear series (as ‘minimally connected’ characters in Slowinski’s terminology) or unrooted trees. Ordered characters were firstly labelled as ‘additive characters’ because the number of steps are added with the number of transformations (Fitch, 1971). The second class of character is hierarchies (Cao et al., 2007), where inclusion relationships between states can be represented by rooted trees, which are used in 3ta as implemented in LisBeth (Zaragüeta et al., 2012). The example in Fig. 1C is close to the ordered character shown in Fig. 1B: it is a completely dichotomous hierarchy, with the maximum number of internal nodes for the tree. The results of phylogenetic analysis of quantitative characters, as well as of qualitative characters, is impacted upon by the coding of characters (Laurin and Germain, 2011). Knowing that the analysis of different characters can yield different results (*e.g.*, a different number of optimal trees, different node support by synapomorphies or bootstrapping, different topologies, different retention indices, etc...), differences in coding approaches may hamper some comparisons between trees obtained by various studies (Hauser and Presch, 1991). The advantages and drawbacks of the various ordering schemes have not received as much attention by systematists as they deserve (Wilkinson, 1992; Brazeau, 2011). The assumptions of ordered (*e.g.*, with the criteria of ontogeny or similarity) and unordered parsimony (Hauser and Presch, 1991), and of searching trees with the fewest transformations are widely known and need not be repeated here. Slowinski (1993) noted that in the 1970s and early 1980s, most multistate characters were ordered, with some authors viewing unordered states as ‘agnostic’ or ‘nihilistic’ (Mickeyvich, 1982). Since then, the situation has reversed and ordered characters have been used only in a minority of recent studies (Grand et al., 2013).

The assumptions underlying the 3ta are less well known and probably need to be presented briefly. Initially described as a ‘more precise use of parsimony’ (Nelson and Platnick, 1991), the three-taxon analysis can be understood as a completely different method (Nelson et al., 2003; Cao et al., 2007; Zaragüeta and Bourdon, 2007; Zaragüeta et al., 2012). The method is available in LisBeth (Zaragüeta et al., 2012), but can be performed using any parsimony software after the parsimony matrix has been converted into a three-taxon

statements (3ts) matrix, an operation that is automated in LisBeth. Within 3ta, characters are represented as hierarchies (Cao et al., 2007). The phylogenetic trees are reconstructed after the hierarchical characters are treated as sets of 3ts. 3ts (*e.g.*, (A, B) C) are statements about the phylogenetic affinities between three taxa (*e.g.*, A is more closely related to B than to C). The reconstructed cladograms are those that contain the largest subsets of compatible 3ts; two 3ts are incompatible if the three terminals considered are identical, but the relationship is different: *i.e.* (A(BC) and (B(AC))). Systematists generally store the character states in matrices; each parsimony-state in 3ta is placed in a terminal node of a hierarchy (Cao et al., 2007; Zaragüeta and Bourdon, 2007) (Fig. 1C), which represents the character. The plesiomorphic state is also defined, as the root. This representation differs from the step matrices and their assumptions of ‘proximity’ (Fig. 1A, B; their pattern of transformation, and the transformation costs). Some authors consider that hierarchical representation is a more relevant and intuitive way to express homologies (Williams and Ebach, 2006; Cao et al., 2007) in a cladistic analysis, where hierarchies of homologs thus leads to a hierarchy of taxa. The relationship homolog/taxa is thus equivalent to the relationship character/cladogram: the first is the part and the second is the whole.

3ta involves a second analytical stage that has sometimes been misunderstood. Characters, *i.e.* hierarchies of character-states, may be converted into 3ts, *i.e.* minimal hierarchies relating two terminals that are closer to each other than any is to a third. This decomposition may imply some redundancy, *i.e.* there are more 3ts produced than necessary to recompose the character. In these cases, Nelson and Ladiges (1992) proposed what they called a fractional weighting (FW hereafter). FW quantifies the contribution of each 3ts to the character. In the example shown in Fig. 2A, only two 3ts out of three in each group of three 3ts are necessary because if taxon C is closer to D than to A, and if C is closer to E than to A, it follows that D is closer to E than to A. Hence, the third triplet (shown in grey) is redundant and could be omitted. However, in the absence of external criteria, it is impossible to objectively determine which 3ts is redundant (in the above example, any of the two other triplets could be omitted, instead of the one in grey). FW quantifies the contribution of each 3ts to the character. 3ts are not characters, contrary to what has often been stated in the literature (Kluge and Wolf, 1993). Thus, even when using parsimony software, the constraint of the logical independ-

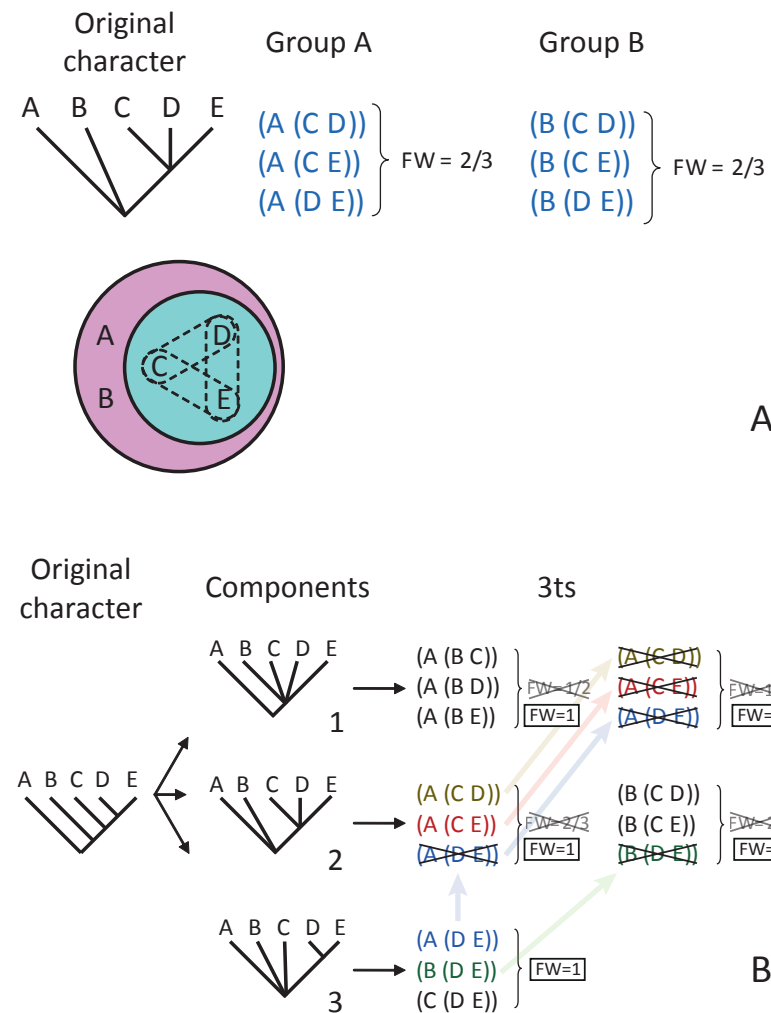


Fig. 2. Fractional weighting (FW) and dependence among 3ts derived from the same character. A. A character relating five terminal taxa is decomposed into its elementary relationships, expressed as a list of 3ts. The decomposition results into a list of six 3ts. However, only two of the three 3ts [e.g. (A (C D)) and (A (C E))] that relate three terminals (C D, E) to other, more distant ones (e.g. A) are necessary to build the initial hierarchical character, as shown in the set diagram. FW assigns an information content/weight of 2/3 to each 3ts of the two groups. B. A multi-state character is analysed into components and then into 3ts. 3ta takes into account the fact that each component is not independent. Thus, the 3ts derived from different components of the same character may be redundant. Consider the 3ts (A (D E)): it appears in the three components. However, it represents a single node in the original character. The 3ts derived from the third component have a FW=1, i.e. they are non-redundant. The same 3ts in the first and second components are thus redundant and can be removed. The FW of the remaining 3ts of the component is then recalculated from FW=2/3 to FW=1 in the second component, and FW<1 in the first, etc... Note that the matrix treatment of the 3ts is immaterial with their status of dependent or independent, contrary to characters in parsimony analysis.

ence (Wilkinson, 1995) of the different columns of the matrix does not apply to the different 3ts derived from the same character. On the contrary, FW reflects the logical mutual dependency (Williams and Siebert, 2000), as shown on Fig. 2B (but see Wilkinson *et al.*, 2004).

With perfectly congruent data, parsimony and 3ta give always the same tree. Operationally, the only difference between them is how they deal with incongruence in the data (e.g. in (Nelson *et al.*, 2003)). The history of wings in stick insects (phasms) highlighted by (Whiting *et al.*, 2003) can illustrate the differences and similarities between parsimony and 3ta (Fig. 3). The cladogram of phasms shows evidence for many losses and independent reacquisitions of a complex structure (Fig. 3A), the pterygote wing (Fig. 3B). In Fig. 3C, this hierarchy (character tree drawn a posteriori from the

analysis) is simplified so as to illustrate the different points of view. In parsimony analysis, the character 'wing' is traditionally coded as binary with the states absent and present. In 3ta, the character is coded as an inclusive hierarchy with the state present nested within the state absent (Fig. 3A). The states 'absent' and 'present' are both considered as complementary homologues (de Pinna, 1991) by parsimony proponents whereas only the acquisition is considered to be a homologue by the 3ta proponents (Cao *et al.*, 2007): the class 'absent' includes 'everything else' and is not considered as an homologue (the plesiomorphic state in 3ta). 3ta proponents consider 'homologs' the nodes of a character state tree (Fig. 1C, 3C). For instance, in Fig. 1C, W is a homologue (i.e., states 3 and 4 are considered homologous), X is another homologue (i.e., 2, 3 and 4 are homologous), etc. The character tree shows 'polarized'

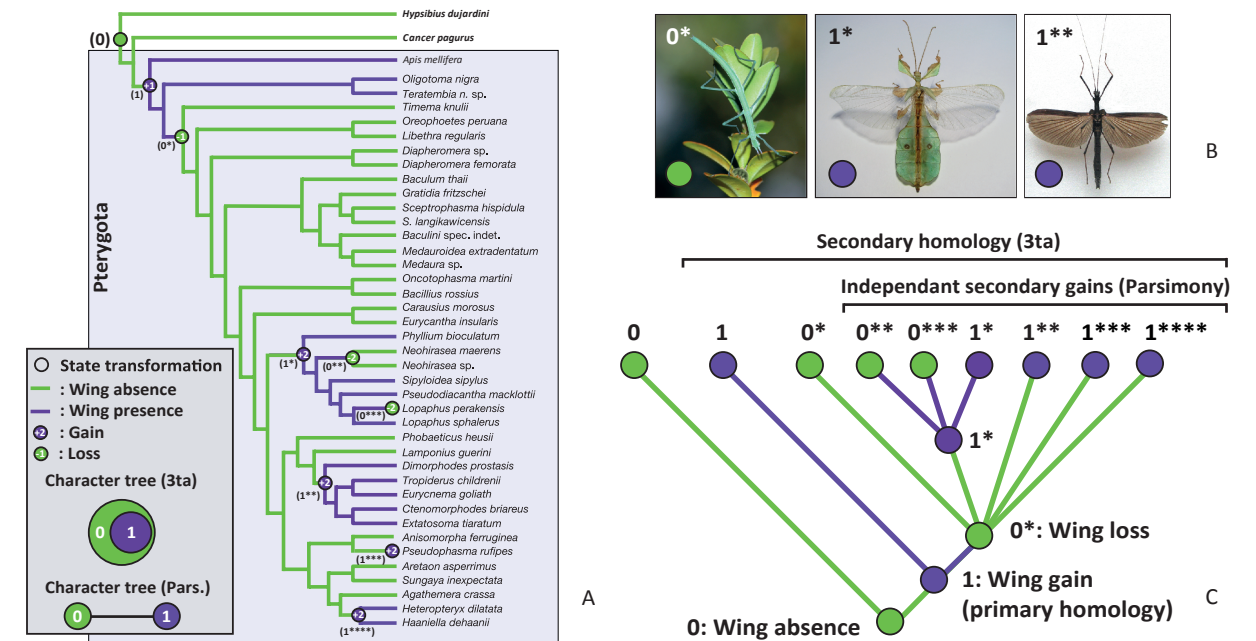


Fig. 3. An example of character history according to cladistics. A is a modified phylogeny from Whiting *et al.* (2003) with transformations (obtained using parsimony under acctran optimization) including reversals (denoted by a '-' sign) and acquisitions or reacquisitions ('+') from a binary character absence/presence (the 3ta character tree is represented as a hierarchy). The degree of gains and losses is denoted by numbers (1 for primary, 2 for secondary, etc.) in the circles placed at the corresponding nodes. The degree of the resulting state (presence or absence, appearing in parentheses below and to the left of the symbols for gains or losses) is denoted by the asterisks; primary gains or losses have none; secondary gains or losses have one; ternary ones have two, etc. B shows the wingless Phasmatodea *Leptynia hispanica* (Bolivar, 1878), *Phyllium bioculatum* (Gray, 1832) with wings and *Pseudophasma acanthotum* (Redtenbacher, 1906) with wings and *Pseudophasma acanthotum* (Redtenbacher, 1906) with convergently re-acquired wings (copyrights, respectively: Fritz Geller-Grimm and Felix Grimm - CC BY-SA 3.0, Dräguš - GFDL and Dräguš - PD-self). C represents with a character tree the history reconstructed a posteriori from the analysis (with the mapping of the character on the optimal tree) of the character 'wing' of phasms. Each node represents a parsimony apomorphy, contrary to the 3ta for which the only apomorphy is the node 1 (as only secondary homology because the only primary homology postulated was the homology between all wings).

(nested) homologies. With the practical example of our Fig. 3 (Whiting *et al.*, 2003), all wings can be considered homologous at a specific level of inclusiveness (Fig. 3C, first appearance of state 1, represented as slightly above the root). In parsimony, this binary character supports six clades (on Fig. 3C, six classes are synapomorphic: 1, 0*, 1*, 0**, 1**, 1****, and one class is autapomorphic: 1***), whereas in 3ta, the character is one homology, accepted by the analysis, and support only one clade (in Fig. 3A, the sister group of *Cancer pagurus*, the class 1 in Fig. 3C, is synapomorphic). In 3ta, only the primary homology hypotheses proposed by the systematist are tested: the hypothesis was that the wings were homologous. Either the character is accepted as secondary homology or it is rejected as a hole. No new hypotheses generated by the algorithm are considered, especially homoplasies, i.e., 'synapo-

morphic' reversals with multiple reappearances are not considered as evidence for support of clades. The algorithm can reject (or not) the proposed hypothesis, given the general framework. In this case, the hypothesis that all phasm wings are just pterygote wings cannot be rejected. The generation by the algorithm of new hypotheses about multiple transformations, appearances and disappearances, is considered spurious. Also, 3ta characters are inherently polarized because this information is necessary to the analysis, while the polarization of characters or tree can be made before or after the analysis in parsimony. The assessment of plesiomorphy is considered in 3ta as necessary to establish phylogenetic relationships between states, and is part of the homology hypothesis. The treatment as 3ts is not equivalent to unrooted quartets, which are partitions, and parsimony allows partitions to represent relationships,

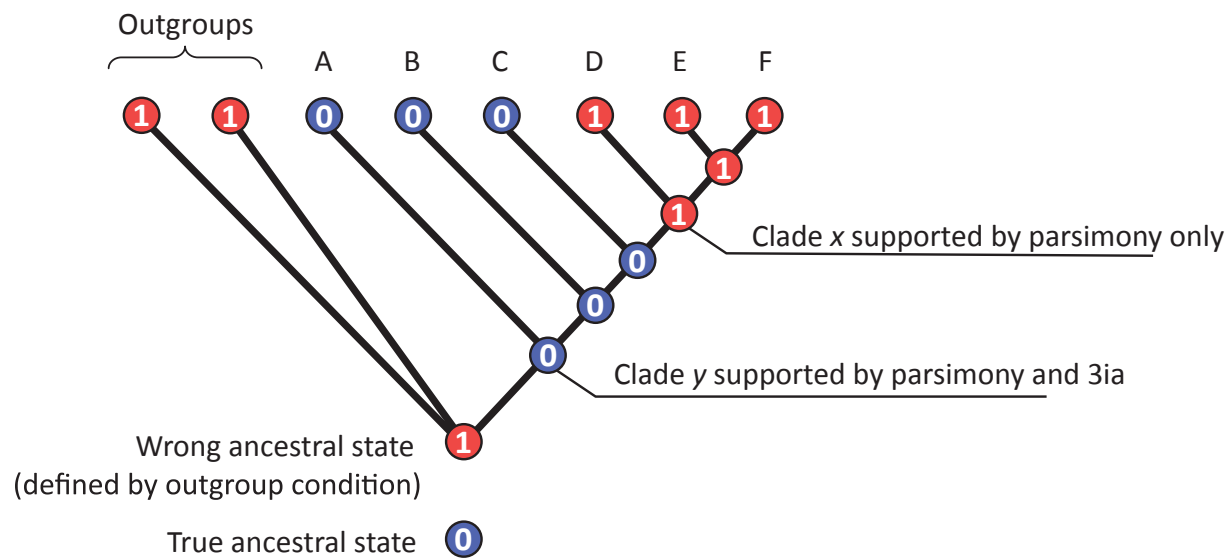


Fig. 4. An example of optimal cladogram obtained both in parsimony and 3ta with effect of polarization error with an outgroup criterion. The true ancestral state is 0, and 1 is a true synapomorphy of D, E and F, but the outgroup criterion leads to an erroneous ancestral condition (1). Parsimony may be less sensitive to errors in polarization than 3ta, which cannot support a clade with what appears to be a reversal from 0 to 1.

instead of 3ta. A long debate has led proponents on both sides to provide arguments on whether parsimony or 3ta was the best cladistic method. This debate has mostly relied on theoretical considerations and contrived examples based on very small datasets (Farris and Kluge, 1998; Farris, 2012; Harvey, 1992; Nelson and Ladiges, 1996; Platnick *et al.*, 1996; Zaragüeta and Bourdon, 2007), although a few studies have compared the methods with empirical datasets (Patterson and Johnson, 1995; Udovicic *et al.*, 1995; Williams, 1996; Bourdon, 2006; Corvez, 2012).

Recently, Grand *et al.* (2013) studied the relative merits of 3ta and parsimony with ordered and unordered states through simulations (100 matrices of six characters and eight taxa each) and two empirical examples. Their results suggested that 3ta has a greater efficiency to retrieve correct relationships (resolving power) than ordered parsimony, but that it also yields more incorrect relationships (artefactual resolution). They also clearly demonstrated that, under the conditions investigated in the simulations, unordered parsimony had the least resolving power and that it had greater artefactual resolution than ordered parsimony (*i.e.*, unordered parsimony can be considered the least effective of the three methods for treating multistate characters reflecting underlying continuous data). The results of the empirical studies of Grand *et al.* (2013)

are difficult to interpret because of the absence of an outgroup in its reference phylogeny and the small taxonomic and character sample. Thus, it would be interesting to perform a more thorough study (based on more extensive simulations) with data generated especially to better discriminate between the three character coding schemes.

In this study, we test various hypotheses in order to refine characterization of the behaviour of parsimony (ordered and unordered) and 3ta on discretized continuous characters. Thus, we test the following hypotheses:

1. Given that our simulated data are intrinsically continuous, we expect that ordered methods (parsimony and 3ta) will outperform unordered parsimony. The relative performance between 3ta and ordered parsimony is more difficult to predict, but a greater accuracy of ordered parsimony could be expected for model-based simulations that allow genuine reversals.
2. We expect that errors in character polarity affect differently the methods, with the performance of 3ta deteriorating faster with increasing polarity errors than ordered parsimony because 3ta uses rooted character state trees, whereas ordered parsimony uses unrooted character state trees (Fig. 1). The transformations are multidirectional in parsimony, contrary to 3ta, which should result in more synapomorphies being recognized in parsimony. 3ta might

be more affected by polarization errors because 3ta disregards reversals as synapomorphies, and if a character is wrongly polarized, a genuine synapomorphy would be disregarded (Fig. 4). In parsimony, if the polarity is wrong, all changes would be used, even though their interpretation might be erroneous (Fig. 4). The performance of unordered parsimony with polarization errors is more difficult to predict. These variable levels of polarization errors are obtained by changing outgroup branch lengths.

3. Tree balance and branch lengths of the real ingroup phylogeny have an effect on the behaviour of the methods: performance of phylogenetic methods can be influenced by terminal and internal branch length and tree symmetry. Based on the ratios of terminal to internal branch lengths (which can be considered a ratio of homoplasy to synapomorphy), then for ultrametric trees, performance should be best for the symmetrical tree, which has the lowest terminal/internal branch length ratio, and worst for a fully pectinate tree, which has the highest ratio. This should prevail in all three methods (throughout this paper, the ‘three methods’ will always refer to ordered parsimony, unordered parsimony, and 3ta). If we consider 3ts as the minimal phylogenetic information (or if we use a metric, such as ITRI, which relies on 3ts), we also expect that a long branch leading to a clade of several taxa is more important than a long branch length leading to a clade of two taxa, because fewer 3ts are involved in the latter. Thus, the structure of a phylogenetic tree may be expected to have an impact.

Material and methods

Simulated character sets

Reference topologies. Data was simulated on three topologies of 20 ingroup OTUs and one outgroup OTU and a total of 18 branch length settings (Fig. 5). One of the reference topologies is fully pectinate (*i.e.*, fully asymmetrical), a second is fully symmetrical for the ingroup (balanced) and a third is an equiprobable (randomly generated) tree, intermediate in asymmetry between the two others (Fig. 5A-F). In this paper, branch lengths represent evolutionary time, but given that we simulated characters using Brownian motion (see below), branch lengths also reflect expected character variance (Felsenstein, 1985).

Reference trees. Three outgroup branch lengths were used on each of the reference topology, leading to

nine ultrametric reference trees (Fig. 5A-C). Nine other branch lengths differing in both ingroup and outgroup were specified on the equiprobable tree with taxa of various geological ages, leading to 9 additional non-ultrametric reference trees (Fig. 5D-G). For six out of these nine trees, all terminal and internal branch lengths of the ingroup were set to one, and the outgroup branch was set at zero, one, three, five, six tenths, or the full tree depth (Fig. 5D). Three out of the nine trees (Fig. 5E-G) were generated by modifying the ingroup branch lengths on the equiprobable tree with the outgroup branch set at zero length (actual ancestor).

Simulated matrices. From each of the nine reference trees illustrated in Fig. 5A-C, 100 matrices of 100 characters \times 10 states were simulated (*i.e.*, a total of 900 matrices). Similarly, 100 matrices of 100 characters \times 10 states were simulated for each of the nine paleontological trees illustrated in Fig. 5 D-G (*i.e.*, a total of 900 paleontological matrices). We thus simulated a total of 1800 matrices, which were produced using Mesquite and the scripts in Supplementary Online Materials 1 (S1) on data that were discretized using Excel spreadsheets (S2 for parsimony; S3 for 3ta) and that were compiled into S4 (which are the matrices in the parsimony format).

Character coding. The characters were simulated with continuous Brownian motion in Mesquite (Maddison and Maddison, 2014) to represent data inherently ordered as morphoclines (such as size or shape characters). Simulations were made using this evolutionary model because it is one of the simplest and most widely used in evolutionary biology to study the evolution of continuous phenotypic characters. For example, phylogenetic independent contrasts (Felsenstein, 1985) and squared-change parsimony (Maddison, 1991) assume this model. Characters simulated through Brownian motion are continuous; they were then discretized into 10 equal intervals representing character states, in order to simulate morphoclines following the simple procedure described in Laurin and Germain (2011). Because Brownian motion has no tendency, the resulting distribution is Gaussian; thus, gap coding cannot be used, and the limits between states are arbitrary. The primitive condition is determined by the outgroup criterion. The variable outgroup branch lengths allow us to assess the influence of polarization errors, whereas the variable ingroup branch lengths allow assessment of the impact of geological age of ingroup taxa on tree resolution (presumably by altering support of the clade subtended by the various branches), thus enabling a comparison of paleontological and neontological data-

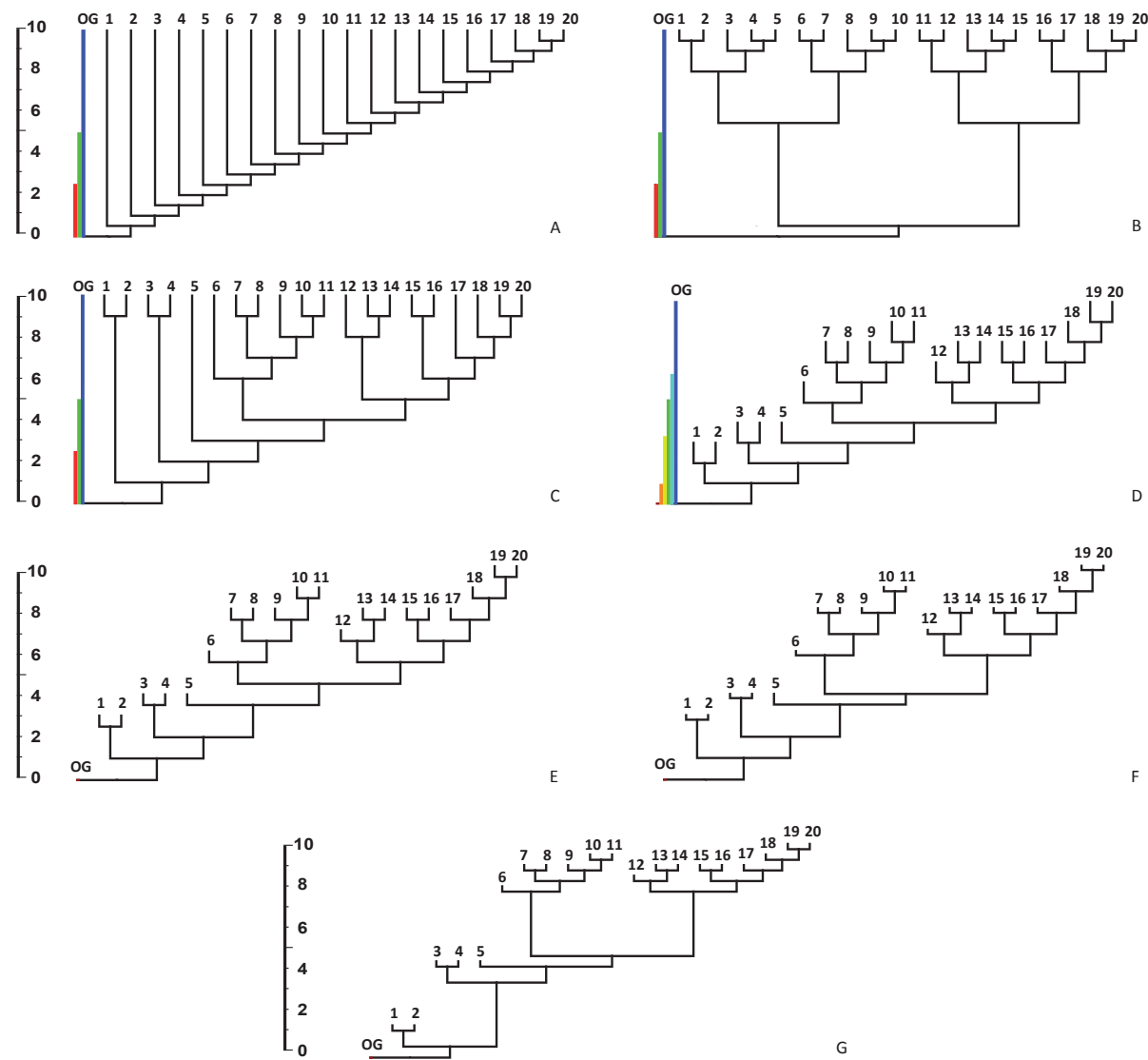


Fig. 5. Trees used for our simulations: A, pectinate; B, symmetric; C, equiprobable; D, equiprobable with branch length set to 1; E-G, equiprobable with steady increase of internal/external branch length ratio. Each color represent a specific outgroup branch length expressed as a proportion of total tree depth (A-C: blue, 1; green, 1/2; red, 1/4; D: dark blue, 1; light blue, 2/3; green, 1/2; yellow, 1/3; orange, 1/10; red, 0). A-C represent trees with a neontological ingroup; trees D-G represent paleontological trees (with diachronous tips).

sets. Each of the 100 characters (in the 1800 matrices) was coded in three different ways corresponding to unordered parsimony, ordered parsimony (with linear character states) and 3ta.

Tree searches

Characters were analysed with the three methods. Thus, 5400 phylogenetic analyses were performed

(three analyses for each of the 1800 matrices).

Cladistic analyses were performed with PAUP* 4.0b10 (Swofford, 2003) for both ordered and unordered parsimony. Three-item analyses were also performed on PAUP* 4.0b10 but parsimony matrices were transformed into three-item statements matrices with fractional weighting (Mickevich and Platnick, 1989; Nelson and Ladiges, 1992) using LisBeth 1.3 (Zaragüeta *et al.*, 2012). All analyses were done with a heuristic

search of 50 replicates (a number of searches that our preliminary analyses suggested was sufficient to recover all optimal trees for our matrices), using the TBR algorithm. A strict consensus (Sokal and Rohlf, 1981) of all optimal trees was then constructed using PAUP*.

Tree comparisons

The optimal tree yielded by the data analyses (or strict consensus tree when several optimal trees were found) were compared with the reference phylogenies, to assess accuracy of the results. We compared the behaviour (resolving power and artefactual resolution) of ordered parsimony, unordered parsimony and three-item analysis (3ta) in phylogenetic inference, as recently done by Grand *et al.* (2013).

Several methods are available to compare unrooted trees (Robinson and Foulds, 1981; Estabrook *et al.*, 1985), but we prefer using rooted trees because they are the only classificatory structures which can convey unambiguous information about phylogenetic relationship (Rohlf, 1982). An unrooted tree conveys information, but it is ambiguous in the sense that it is compatible with several alternative relationships between taxa. Tree comparisons were performed with the ‘Inter-Tree Retention Index’ (ITRI) proposed by Grand *et al.* (2013) to measure the degree of congruence between two trees. The ITRI is based on the proportion of relationships (*i.e.* 3ts relationships between three OTUs, irrespective of their relationships with other OTUs) that are common to two trees. An advantage of this method is that the computed tree-to-tree similarities are not symmetrical (*i.e.*, the proportion of relationships is either the proportion for one tree or for the other). It is thus possible to discriminate between resolving power (power to find correct relationships) and artefactual resolution (incorrect relationships), by comparing trees obtained from simulated data with the tree used to generate these data. ITRI is equivalent to the retention index (RI) (Farris, 1989; Archie, 1989), calculated as a proportion of 3ts (Kitching, 1998). The ITRI is defined as:

$$RI = \frac{(2n - (X + 2(n - X)))}{(2n - n)} = \frac{X}{n}$$

Where X stands for the sum of fractional weighting (FW) of 3ts implied by a character compatible with a given tree; n is the sum of FW of all 3ts obtained from the character. In each pairwise tree comparison, strict consensus trees were used to summarize congruence between results of an individual analysis (on a single

matrix of 100 characters). See Grand *et al.* (2013) for a more detailed explanation. A strict consensus tree was built whenever analysis of a matrix produced more than one tree.

Our results are presented without mention of the number of optimal trees because the resolving power and artefactual resolution are the direct expression of data congruence and of the ratio homology/homoplasy, and that the ITRI expresses it well. Generally, a strict consensus tree of a thousand trees will show lower resolving power and artefactual resolution than a consensus of ten trees, because of data ambivalence. But there is always the possibility that an analysis will result in one false tree versus another analysis with one hundred slightly better trees. Kearney (2002) argued that ‘resolution of relationships is obviously a goal of phylogenetic analysis’, but her results show that phylogenetic performance cannot be measured by the number of trees. To avoid confusion with artefactual resolution, we complete her sentence by stating that it is *correct* resolution which is a goal of phylogenetic analysis. We think it is a necessary adjunction to remove the ambiguity between resolution and number of trees. Correct resolution does not display a simple relationship with the number of optimal trees.

Testing the results

Each consensus tree summarizing a phylogenetic analysis is compared with the reference tree using the ITRI to assess resolving power and artefactual resolution yielded by each method, each set of outgroup branch length, and each topology. Resolving power is calculated as the proportion of FW for 3ts of the reference tree that are also present in the consensus tree, and can be understood as the proportion of ‘true information’ the analysis has retrieved. Artefactual resolution is calculated as the proportion of FW for 3ts of the consensus tree obtained from an analysis that is not present in the reference tree, and this represents artefactual resolution yielded by the analysis. We calculated means of ITRI (*i.e.*, we got a mean value for resolving power artefactual resolution on each reference topology). The statistical significance of differences in resolving power and artefactual resolution of the various methods of analysis was tested with the Wilcoxon signed-rank test (for paired samples) to compare means of ITRI, because there was no sample with normal distribution (Shapiro-Wilk test) and homoscedasticity of variances (Fisher test). Differences in topologies and branch lengths were tested using a Mann-Whitney test, and linear regres-

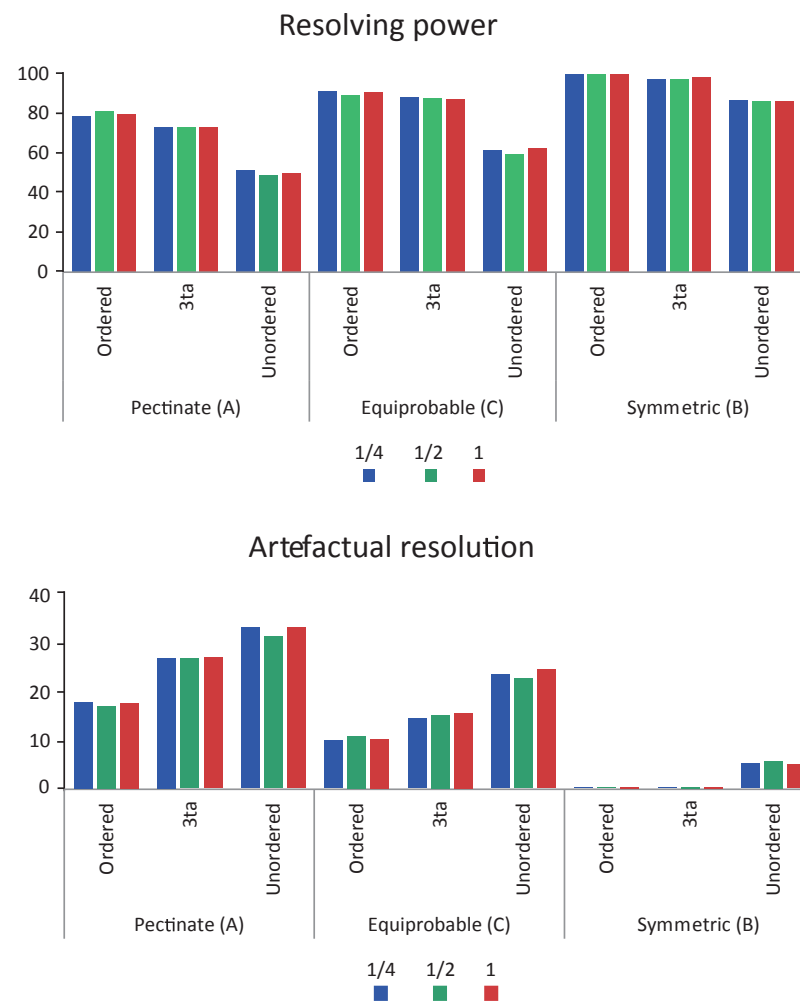


Fig. 6. Histograms showing resolving power and artefactual resolution obtained by ITRI mean percentages sorted by outgroup branch length (blue, 1/4; green, 1/2; red, 1), by method (ordered parsimony, 3ta and unordered parsimony), and by topology (pectinate, equiprobable and symmetric; see S8). Each ITRI mean represents 100 matrices of 100 characters and 21 taxa (outgroup included).

sions were performed to test tendencies were the external/internal branch length ratio varies, as in trees D, E, F and G (Fig. 5). Because many comparisons were made, we use the false discovery rate procedure (Benjamini and Hochberg, 1995) to assess the statistical significance of the differences in performance.

Results

All of the 2,700 analyses with equal branch lengths yielded strict consensus trees that are at least partly resolved; they all carry phylogenetic information. For ordered parsimony resolving power has a mean of 99.8% (sd: 1.1%) on the symmetric tree (Fig. 5B), 78.8% (sd: 8.8%) on the fully pectinate tree (Fig. 5A) and 91.1% (sd: 8%; Fig. 6) on the equiprobable tree (Fig. 5C)

with blue outgroup branch length). Artefactual resolution is respectively about 0.2% (sd: 0.7%), 17.9% (sd: 7.2%) and 9.9% (sd: 6.3%). Resolving power is always a little lower and artefactual resolution greater in 3ta than in ordered parsimony ($p < 0.0001$; Table 1). Unordered parsimony has the lowest resolving power ($p < 0.0001$), and produces the greatest artefactual resolution ($p < 0.0001$). The outgroup branch length does not have a significant effect on resolving power or artefactual resolution for the first three ingroup trees (Fig. 5A-C). Among the three reference topologies, the symmetrical topology seems to be the easiest to retrieve whatever cladistic method is used (greatest resolving power and lowest artefactual resolution), followed by the equiprobable topology, and by the fully pectinate topology (Table 1). With the 1,800 analyses performed on the paleontological tree (Fig. 5D) and whatever method used,

resolving power decreases and artefactual resolution increases with outgroup branch length ($p < 0.0001$; Table 2; Fig. 7). Ordered parsimony seems to be less sensitive than 3ta and unordered parsimony to polarization errors (reflecting outgroup branch lengths).

From the 2,400 analyses with different branch lengths (Fig. 5D-G), the differences between the resolving power and artefactual resolution values calculated on trees from a same reference topology cannot be explained by a linear model; results do not even vary consistently with the ratio between internal and terminal branch lengths (S5).

The impact of outgroup branch lengths was generally non-significant when we used trees with contemporary ingroup taxa (Fig. 5A-C). In 54 tests based on these trees, only 4 rejected the null hypothesis that outgroup branch length has no impact on performance (Fig. 5; Table 1). All 4 significant results concern the trees built from the symmetrical reference topology. Differences in the other 50 comparisons on the same trees (Fig. 5A-C) do not show any tendency of increasing or decreasing resolving power or artefactual resolution when the outgroup branch length increases. Tests on trees built from the equiprobable paleontological reference tree (Fig. 5D) with a variable outgroup branch length show that the performance of all methods decreases when outgroup branch length increases (Fig. 7; Table 2). Results of ordered parsimony and 3ta are very similar, with unordered parsimony yielding much lower performance.

Finally, we have compared the performance for each method on two trees with the same topology but in which the taxa were extant (Fig. 5C) or extinct (Fig. 5D), to simulate neontological and paleontological trees, respectively. The results suggest that the paleontological tree is far easier to reconstruct than the neontological tree for ordered parsimony and 3ta in terms of resolution power and artefactual resolution ($p = 0.0004$ for resolution power in 3ta and $p < 0.0001$ for resolution in ordered parsimony, as well as for artefactual resolution with both methods). Differences between trees in unordered parsimony were not significant.

Discussion

Brownian motion in cladistics

Under the conditions examined in our simulations, ordered parsimony performed best, thus confirming our first hypothesis. Simulating characters on a phylogeny

has the advantage that the reference phylogeny is known without error. However, the Brownian motion, evolutionary model that is widely used in simulations of evolution, and which we used here, impacts upon the results, and reflects theoretical assumptions. Under Brownian motion, character evolution is stochastic and unpredictable, as are many historical events, but follows a general pattern that reflects the phylogeny, which can be inferred by analyzing character state data. Here, because of the generation of discretized continuous characters, the distribution of character states is unimodal (S6). They are intrinsically ordered and thus represent morphoclines. Brownian motion, like most other models of molecular evolution, such as GTR+I+ Γ (Tavaré, 1986) or the speciation model, is not directional (*i.e.*, it implies no trends). Thus, the relationships between character states can be represented by unrooted trees (Fig. 1A). Brownian motion thus leads to an intrinsically ordered and unpolarized modeling (when the root condition is not specified in the simulation, as is the case here), contrary to models implying trends or irreversible evolution, which are intrinsically polarized. Under Brownian motion, the probability for a character state 0 to evolve into 1 is greater than the probability for 0 to evolve into 2 in a short time; it was thus expected to favour ordered parsimony over unordered parsimony, and to a lesser extent, 3ta. However, this model, one of the simplest, seems applicable to various characters, such as ontogenetic sequence data (Poe and Wake, 2004; Poe, 2006).

Reversals in phylogenetics

A particularly controversial issue in cladistics concerns the treatment of reversals. Proponents of parsimony (Kluge, 1994; Farris *et al.*, 1995; Farris, 1997; Farris and Kluge, 1998) and 3ta (De Laet and Smets, 1998; Siebert and Williams, 1998) have been deeply divided on this particular issue. In parsimony, a transformational approach to homology using the Wagner (Farris *et al.*, 1970) and Fitch (Fitch, 1971) parsimony algorithms treats characters from the perspective of unrooted character-transformation trees (Slowinski, 1993). Reversals provide information and can serve for clade support because they are evidence of secondary homology with the appropriate test of maximizing congruence. This maximization of congruence leads to search the pattern with the minimum of *ad hoc* hypotheses that are convergences and reversals. For Farris (2012), reversals can be inferred a priori in the inference of primary homologies but also from an analysis: 'More

Table 1. Statistical significance of differences in resolving power and artefactual resolution (calculated through mean ITRI) between methods, outgroup branch lengths, and topology of the reference tree. The probabilities were calculated using Wilcoxon signed-rank

RESOLVING POWER**Significance of differences between methods**

| | Symmetric | | | Pectinate | | | Equiprobable | | |
|-----|-------------------|-------------|---------------|-------------------|-------------|---------------|-------------------|--------------|---------------|
| | Ordered/Unordered | Ordered/3ia | Unordered/3ia | Ordered/Unordered | Ordered/3ia | Unordered/3ia | Ordered/Unordered | Ordered/3ia | Unordered/3ia |
| 1/4 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 |
| 1/2 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | 0.036 | < 0.0001 |
| 1 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 |

Significance of differences between polarizations

| Outgroup b.l. | Symmetric | | | Pectinate | | | Equiprobable | | |
|---------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | 0.25/0.5 | 0.25/1 | 0.5/1 | 0.25/0.5 | 0.25/1 | 0.5/1 | 0.25/0.5 | 0.25/1 | 0.5/1 |
| Ordered | 0.005 | 0.492 | 0.303 | 0.097 | 0.279 | 0.583 | 0.054 | 0.311 | 0.239 |
| Unordered | 0.063 | 0.849 | 0.636 | 0.941 | 0.517 | 0.394 | 0.706 | 0.988 | 0.101 |
| 3ia | 0.348 | < 0.0001 | < 0.0001 | 0.125 | 0.651 | 0.756 | 0.131 | 0.317 | 0.363 |

Significance of differences between topologies

| | Ordered | | | Unordered | | | 3ia | | |
|-----|-----------------------|--------------------------|------------------------|-----------------------|--------------------------|------------------------|-----------------------|--------------------------|------------------------|
| | Symmetrical/Pectinate | Symmetrical/Equiprobable | Pectinate/Equiprobable | Symmetrical/Pectinate | Symmetrical/Equiprobable | Pectinate/Equiprobable | Symmetrical/Pectinate | Symmetrical/Equiprobable | Pectinate/Equiprobable |
| 1/4 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 |
| 1/2 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | 0.000 | < 0.0001 | < 0.0001 | < 0.0001 |
| 1 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 |

ARTIFACTUAL RESOLUTION**Significance of differences between methods**

| | Symmetric | | | Pectinate | | | Equiprobable | | |
|-----|-------------------|-------------|---------------|-------------------|-------------|---------------|-------------------|-------------|---------------|
| | Ordered/Unordered | Ordered/3ia | Unordered/3ia | Ordered/Unordered | Ordered/3ia | Unordered/3ia | Ordered/Unordered | Ordered/3ia | Unordered/3ia |
| 1/4 | < 0.0001 | 0.089 | < 0.0001 | < 0.0001 | < 0.0001 | 0.002 | < 0.0001 | < 0.0001 | < 0.0001 |
| 1/2 | < 0.0001 | 0.002 | < 0.0001 | < 0.0001 | < 0.0001 | 0.028 | < 0.0001 | < 0.0001 | < 0.0001 |
| 1 | < 0.0001 | 0.013 | < 0.0001 | < 0.0001 | < 0.0001 | 0.001 | < 0.0001 | < 0.0001 | < 0.0001 |

Significance of differences between polarizations

| Outgroup b.l. | Symmetric | | | Pectinate | | | Equiprobable | | |
|---------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | 0.25/0.5 | 0.25/1 | 0.5/1 | 0.25/0.5 | 0.25/1 | 0.5/1 | 0.25/0.5 | 0.25/1 | 0.5/1 |
| Ordered | 0.554 | 0.868 | 0.668 | 0.348 | 0.913 | 0.403 | 0.403 | 0.800 | 0.296 |
| Unordered | 0.837 | 0.426 | 0.330 | 0.400 | 0.632 | 0.593 | 0.694 | 0.676 | 0.439 |
| 3ia | 0.007 | 0.209 | 0.227 | 0.624 | 0.545 | 0.970 | 0.095 | 0.179 | 0.815 |

Significance of differences between topologies

| | Ordered | | | Unordered | | | 3ia | | |
|-----|-----------------------|--------------------------|------------------------|-----------------------|--------------------------|------------------------|-----------------------|--------------------------|------------------------|
| | Symmetrical/Pectinate | Symmetrical/Equiprobable | Pectinate/Equiprobable | Symmetrical/Pectinate | Symmetrical/Equiprobable | Pectinate/Equiprobable | Symmetrical/Pectinate | Symmetrical/Equiprobable | Pectinate/Equiprobable |
| 1/4 | < 0.0001 | < 0.0001 | 0.009 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 |
| 1/2 | < 0.0001 | < 0.0001 | < 0.0001 | 0.001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 |
| 1 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | 0.001 | < 0.0001 | < 0.0001 | < 0.0001 |

tests for the comparisons between methods and with a Mann-Whitney test for the outgroup branch length and the shape of the reference tree. Non-significant results with $\alpha=0.05$ are shown in blue, and non-significant results after correction for multiple tests using the false discovery rate (Benjamini and Hochberg, 1995) are shown in red. Individual values can be found in S7. Abbreviations: b.l., branch length.

Table 2. Impact of outgroup branch length (values of 0, 1/10, 1/3, 1/2, 2/3, 1, expressed as a proportion of total tree depth, were tested) on resolving power and artefactual resolution. Probabilities of the null hypothesis (no impact) were tested through linear regressions of ITRI means on outgroup branch length. These results were obtained from the tree shown in Fig. 5D. Resolving power, slope: -7.2 (ordered), -10.1 (unordered), -13.5 (3ta); intercept (1000 indicates complete power or error): 991.9 (ordered), 675.3 (unordered), 989 (3ta). Artefactual resolution, slope: 4.6 (ordered), 18.3 (unordered), 19.8 (3ta); intercept: 28.9 (ordered), 160 (unordered), 30.8 (3ta). Individual values can be found in S7.

Linear regressions on polarization (tree D)

| | Ordered | Unordered | 3ta |
|------------------------|---------|-----------|-----------|
| Resolving power | 0.02306 | 0.001758 | 0.0008212 |
| Artefactual resolution | 0.02505 | 0.008793 | 0.0009794 |

fundamentally, even if (as I have seen other authors suggest) Hennig would have preferred to distinguish apomorphies from plesiomorphies before starting to construct the tree, he was obviously willing to revise assessments of plesiomorphy during tree construction, for Hennig did in fact recognize reversals and apply them as synapomorphies'. This is the most widespread point of view in cladistics, which prevails in systematic paleontology, and the only point of view represented in probabilistic methods, which prevail in molecular systematics and are starting to be used on phenotypic characters as well (Müller and Reisz, 2006). Assumptions of 3ta are much less familiar. In 3ta, hierarchical hypotheses of homology, *i.e.* a nested set of character states, are submitted to a test of congruence. The test either accepts or rejects the relevance of the hypothesis. Convergence is one of the multiple explanations of rejection. 'Parsimony-like reversals', *i.e.* the generation of hypotheses of homology not proposed by the systematist but generated by the method, violate hierarchical classifications. Thus, they cannot be justified in 3ta rationale. Evolutionary reversals, *i.e.*, losses of instances of character-states, are not used in 3ta to support nodes; they represent only homoplasies, *i.e.* mistaken hypotheses of homology.

For instance, the evolutionary hypothesis (generated after an initial analysis by inferring character history on a tree) involving three conditions deduced a posteriori from an analysis: 0 ('absent'), 1 ('present') and 0* ('secondary absence'; scored the same in a matrix but interpreted differently from primitive absence on a tree) is interpreted differently under parsimony and 3ta. The secondary absence can be explicitly represented as an apomorphy in the primary homology hypothesis (0(1(0*))), under parsimony. Another interpretation (3ta) consists in disregarding secondary absence as synapomorphic but to consider it as a particular case of

absence: (0,1(0*)). Here, neither the absence nor the reversal is considered as a state (neither plesiomorphy, nor apomorphy) because the absence is not a state in 3ta (in Fig. 3, 0*, 0** and 0*** are not considered in 3ta). Parsimony proponents favour the first option, which yields support for six clades in the phasmatodea phylogeny (Fig. 3A). To summarize, the first interpretation (parsimony) considers a loss as an homology and a synapomorphy (because it supports a clade), an homoplasy (because the primary hypothesis is falsified by the distribution of the other characters) and a plesiomorphy (as defined in the matrix), according to Brower and de Pinna (2014). The second interpretation considers a loss as uninformative: it is neither an homology nor a synapomorphy (because it supports no clade), it is not an homoplasy (because the primary hypothesis is in agreement with the distribution of the other characters) and it is not a plesiomorphy (because the absence is not a state in 3ta; it is the root, including all). 3ta proponents favour this interpretation: only one clade in the Phasmatodea phylogeny is supported, and the only synapomorphy is the homology reflecting the first appearance of wings. These two interpretations are thus in perfect opposition. Here we emphasize that Brownian motion is only coherent with the assumptions entailed by the first interpretation: reversals (*i.e.*, secondary absence) are treated as apomorphies in the primary homology hypotheses (as an order with parsimony, or as a hierarchy with 3ta). Our simulations produce informative reversals under Brownian motion, which can be exploited only under a parsimony viewpoint of these reversals: our results present a quantification of the loss in resolving power and artefactual resolution in 3ta if true and informative reversals are present (*i.e.* if true reversals are simulated and 'hidden' into the same state as the plesiomorphy but which support a clades of the known tree). Thus, our results must be interpreted ac-

cordingly. Irreversible characters might yield different results and will be tackled in another study.

We take this opportunity to propose a nomenclatural clarification about reversals (based on the example in Fig. 3A) as secondary homology hypotheses; thus, this clarification is valid both for parsimony and for 3ta. First rounds of reversals are generally called 'secondary losses' (*e.g.* (Carine and Scotland, 1999), when in fact, only the absence should be considered secondary and the loss in itself should be considered as an event that appeared for the first time (*i.e.*, primary). Thus, a character state is primitively absent (primary absence; state 0 on Fig. 3). It can then appear; this is a primary appearance (of state 1), denoted +1 on Fig. 3A. It can be subsequently lost (-1, reversal to state 0, but identified as 0* on Fig. 3a, for greater clarity); this should be called a primary loss, which results in a secondary absence. After this, a secondary gain (+2) can lead to secondary presence (1* in Fig. 3A), and a secondary loss (-2) can lead to ternary absence (0** in Fig. 3), etc.

Uncertain polarization

We failed to find significant results on the impact of outgroup branch lengths and uncertainty in polarization on the neontological trees (Fig. 5A-C), except on some trees built from the symmetrical reference topology.

The effect of outgroup branch length appears to be much stronger on paleontological trees, perhaps because of the shorter branches in the ingroup. Tests on a non-ultrametric version of the equiprobable tree (Fig. 5D) with a variable outgroup branch lengths show that the performance of all methods decreases when outgroup branch length increases (Fig. 7; Table 2). Results of ordered parsimony and 3ta are very similar, with unordered parsimony performing much more poorly. 3ta is more sensitive to outgroup branch length, an effect that might be linked to reversal treatment, but that in any case confirms our second hypothesis. It is however surprising to see that an incorrect polarization has so little effect on resolving power and artefactual resolution.

Character states and ordering schemes

Our simulations clearly show that unordered parsimony performs far worse than the two other methods when reliable criteria for character state ordering are ignored (Fig. 6). All states were ordered (except for unordered parsimony) using a similarity criterion, where-

by transition costs (step-matrices in ordered parsimony) or state hierarchy (3ta) reflect similarity (and outgroup condition, for 3ta). Other ordering criteria exist (Hauser and Presch, 1991), but our results clearly indicate that ordering character states is preferable when characters can be shown to form morphoclines. These results suggest that the current tendency not to order characters in phylogenetic analyses is suboptimal, and shows that important benefits could arise from considering ordering schemes when it appears biologically justified. Note that such ordering requires no prior knowledge of the phylogeny; only knowledge of the character distribution or likely evolutionary model (which can be gained from genetic or developmental data, among others) is required. Ordered parsimony and 3ta share more similarities on this particular point than unordered parsimony.

Tree shape and branch length

Not all topologies are equally difficult to recover accurately. The pectinate topology (Fig. 5A), which shows the longest terminal branches (when its internal branches are all of about the same lengths), is most difficult, followed by the equiprobable (Fig. 5C) and the symmetrical topology (although this may be linked with the fact that internal branch lengths of that tree were roughly proportional with the number of descendant taxa, except for the branch below the ingroup; Fig. 5B; Fig. 6; Table 1). This confirms our second hypothesis. Modifying the distribution of branch lengths yielded results that are more difficult to interpret (S5). Our first assumption was that the results would improve (*i.e.*, the resolving power would be greater and the artefactual resolution lower) as the terminal/internal branch length ratio decreases. This assumption is corroborated for the trees D, E and F (Fig. 5) for resolving power under ordered parsimony and 3ta, but only for trees D and E for unordered parsimony. Tree G yielded worse results, perhaps because some internal branches are shorter. Results on artefactual resolution are more complicated to interpret. Along with the branch lengths of a tree, the 3ts content of clades is another important parameter to consider when performance is assessed using the ITRI. It is directly connected to tree shape, *i.e.* the number of terminal taxa inside and outside each clade (Nelson and Ladiges, 1992). As a consequence, some clades and characters they support will affect more the ITRI than others. More specifically, clades with few taxa within them or with few taxa outside them have less 3ts content than clades

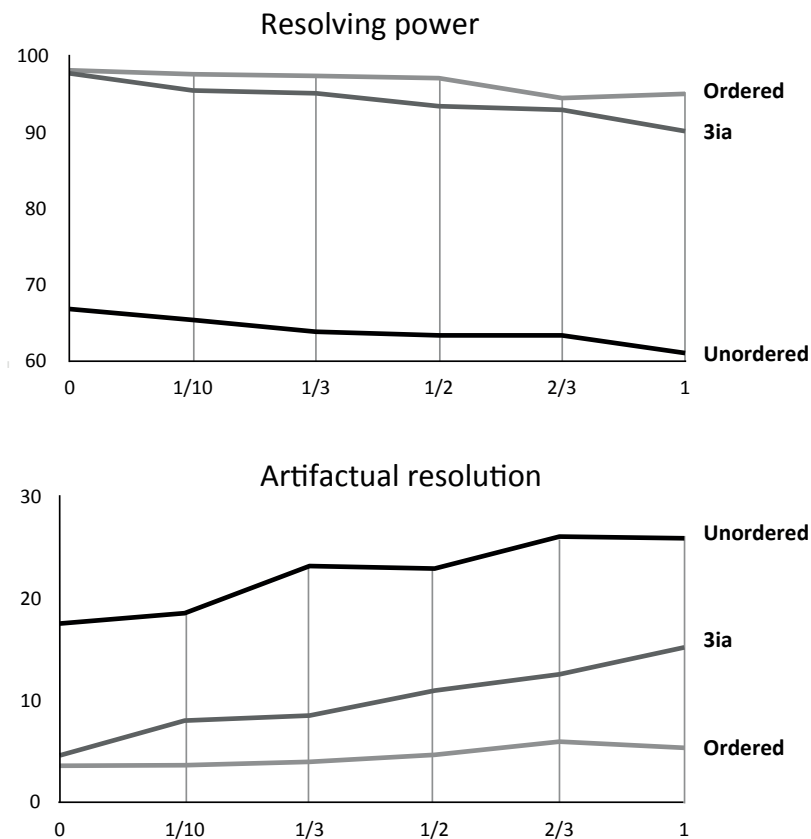


Fig. 7. Impact of outgroup branch length expressed as a proportion of total tree depth (0, 1/10, 1/3, 1/2, 2/3, 1) on resolving power and artifactual resolution. This was obtained through ITRI mean percentages from the tree shown in Fig. 5D (also see S8).

containing an intermediate number of taxa (in our trees, clades with the maximal 3ts content have ten taxa inside and eleven taxa outside). These imbalanced clades will impact results only slightly compared to balanced clades. This may explain partly why tree shape influences phylogenetic reconstruction.

Comparisons of results between an ultrametric tree (Fig. 5C) and a paleontological tree of the same topology (Fig. 5D) show that the latter features better resolving power and less artifactual resolution for ordered parsimony and 3ta. This result is congruent with the claim that adding extinct taxa breaks long branches, which results in important improvement on the resolution of the optimal trees. This claim is supported both on empirical data (Gauthier *et al.*, 1988) and simulations (Huelsenbeck, 1991).

Implications on simulation-based studies and evolutionary models

Our results highlight the advantages of models under which the relationships between character states can be

represented by an unrooted tree (all molecular models) over 3ta hierarchical coding, if reversals can be simulated as in our study. Our simulations under Brownian motion can be represented as in Fig. 8. Firstly, characters are generated under a priori assumptions (here, the evolutionary model represented in Fig. 8A and the reference phylogeny shown on Fig. 8B). In a second step, these characters are interpreted as primary homology hypotheses by discretization and (for 3ta) conversion into hierarchical structures. Our procedure for simulation of characters, and by extension the use of evolutionary models, reflects quantitative characters that display informative reversals.

In parsimony, states that can be hypothesized to form morphoclines should be ordered if there is evidence. In this case, reversals can be represented by primary homology hypotheses. Empirical studies suggest that Brownian motion is a reasonable model for several types of characters, such as body size (Laurin, 2004), bone microanatomy (Canoville and Laurin, 2010), etc. Some characters do not seem to follow a strictly Brownian motion model, but instead may follow a spe-

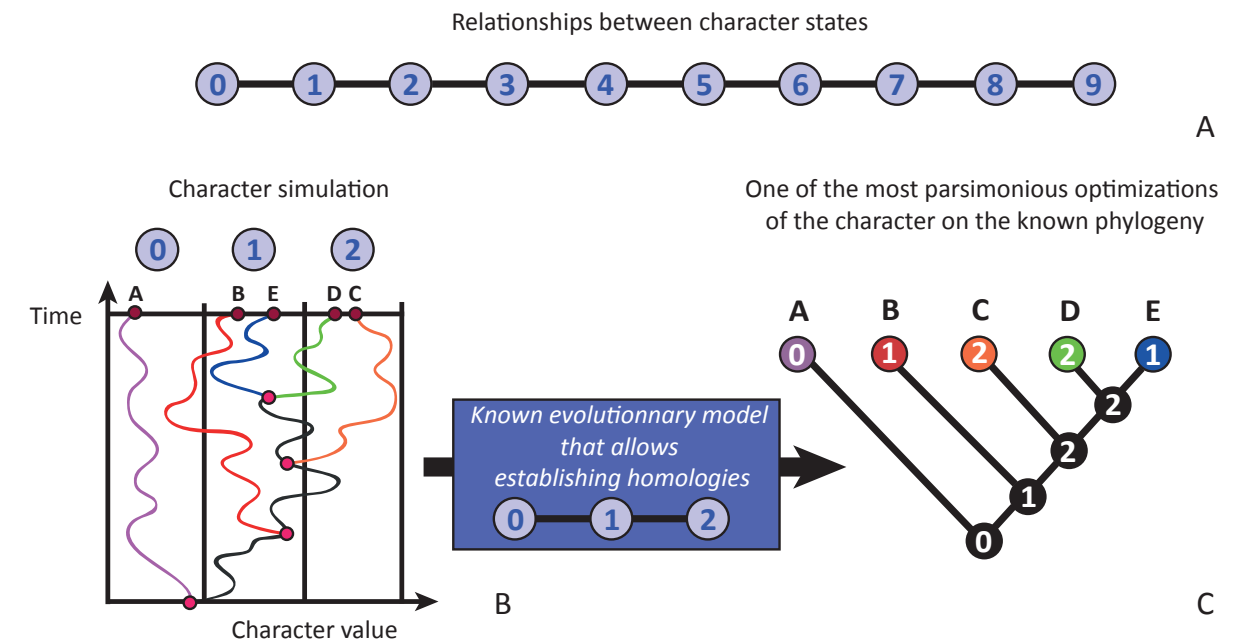


Fig. 8. Character evolution under Brownian motion. In A, the unrooted tree represents the relationships between states of a character simulated under Brownian motion and discretized into 10 states. B shows the simulation of characters on a known phylogeny under Brownian motion. The left graphic represents fluctuations of a continuous character, with cladogenesis and terminal taxa shown as red nodes. After discretization through states 0, 1 and 2, the known evolutionary model and a topology (known or inferred) allow inferring character history. Here, the example treats characters under the parsimony method, with acctran optimization.

ciational model (in which change occurs in both daughter-lineages at or near speciation, but no anagenesis takes place), such as morphological shape data in ratites (Laurin *et al.*, 2012), a punctuational model (similar to the speciation model, but change occurs only on one branch) or an Ornstein-Uhlenbeck model (Uhlenbeck and Ornstein, 1930; Felsenstein, 1988). Some of these alternative models can be obtained from Brownian motion by modifying the branch lengths of a tree (Garland *et al.*, 1993). Hence, data produced using such models might give results close to those that we report below, which should be applicable to much biological data.

In 3ta, the same representation of change is used for groups of parts (homologues) and for groups of whole organisms (taxa). In a transformation, the transformed part is considered a different object. In differentiation or modification, the differentiated or modified part is a particular form of the unmodified part. Many systematists accept this difference for taxa (*e.g.* birds as a differentiated taxon nested in Dinosauria) but not for parts (*e.g.* feathers as a modified part,

nested in scales). In this context, reversal as a loss is perfectly acceptable. However, reversal as synapomorphy is viewed as pointless.

All this suggests that irreversible characters (Nopcsa, 1923; Goldberg and Igić, 2008; Kohlsdorf *et al.*, 2010) are 'more compatible' with 3ta. Further work should be done to compare resolving power and artifactual resolution yielded by parsimony and 3ta when an irreversible evolution model is used (which can be understood as hierarchic). Another interesting field of research consists in developing methods that better simulate Darwinian evolution applied to digital life forms, as in the software Avida (Adami and Brown, 1994) to understand better the behaviour of parsimony and 3ta.

Conclusion

Our simulated characters evolved gradually; they are continuous and have a unimodal distribution (S6). They were discretized because phenotypic data are often

available in discrete form (in descriptions, for instance), and most available parsimony and all 3ta programs require this. Thus, these reflect the data routinely analysed by systematists. In this case, we created 10 states per character to retain much of the original information while producing data that can easily be analysed by almost any phylogenetic package (Laurin and Germain, 2011), and to allow detection of the spurious relationships yielded by homoplasy in the data (Bardin et al., 2013). This procedure yielded many significant results (Table 1; Table 2). We show that unordered parsimony performs far worse than ordered methods on such data, with a loss in resolving power between 13.4% and 29.3% compared to ordered parsimony and with between 4.7% and 15.6% more artefactual resolution (results always statistically significant). Thus, a significant decrease in performance is expected when characters are not ordered as morphoclines. This result highlights the information content of character ordering schemes (Wilkinson, 1992), and we infer that coding continuous characters as fully unordered significantly decreases resolving power and increases artefactual resolution in empirical datasets. We also quantified the differences in resolving power between ordered parsimony and 3ta, which differ in the way they handle reversals as evidence for the support of a clade (among others). These differences vary between 1.3% (non-significant) and 6.3% (Table 1) for the resolving power, and between 0.0% (non-significant) and 9.4% for the artefactual resolution. We also have quantified resolving power when polarization errors, produced by varying outgroup branch length, are introduced (*i.e.*, when polarization is based on an outgroup including a variable proportion of plesiomorphies). Further work could be done to quantify the effect of incorrect state ordering schemes as incorrect morphocline assumptions.

Recently, Grand et al. (2013) found that 3ta yielded significantly greater resolving power and more artefactual resolution than ordered parsimony. In our study, based on an extended set of matrices, taxa and characters, ordered parsimony yielded the greatest resolving power and the fewest artefactual resolutions. The use of exact hypothetical ancestors to root trees in Grand et al. (2013) may explain the contrasting results. Moreover, the present work yields a better understanding of the impact of evolutionary assumptions about character state order and reversals. Further simulations could be done, using different evolutionary models, such as irreversible characters, to see if they yield different results about the performance of parsimony (with ordered states or not) and 3ta.

Acknowledgements

We are grateful to Jérémy Tissier, Laetitia Carrive and Eduardo Ascarrunz for fruitful discussions. We also thank Jacques Ducasse who developed modules of the LisBeth program (on which the 3ta analyses were ran) according to our needs, and to Régine Vignes-Lebbe who thought of more efficient algorithms to store and compare sets of 3ts deduced from the 3ta characters. Comments by Mark Wilkinson and an anonymous referees improved this paper. This work was financially supported by the UMR 7207 CR2P.

References

- Adami C, Brown CT. 1994. Evolutionary learning in the 2D artificial life system "Avida". In: *Artificial life IV*. Cambridge, MA: MIT Press. 377-381.
- Almeida MT, Bisby FA. 1984. Simple method for establishing taxonomic characters from measurement data. *Taxon* 33: 405-409.
- Alroy J. 2001. A Multispecies Overkill Simulation of the End-Pleistocene Megafaunal Mass Extinction. *Science* 292: 1893-1896.
- Archie JW. 1985. Methods for Coding Variable Morphological Features for Numerical Taxonomic Analysis. *Systematic Biology* 34: 326-345.
- Archie JW. 1989. Homoplasy Excess Ratios: New Indices for Measuring Levels of Homoplasy in Phylogenetic Systematics and a Critique of the Consistency Index. *Systematic Biology* 38: 253-269.
- Bardin J, Rouget I, Yacobucci M, Cecca F. 2013. Increasing the number of discrete character states for continuous characters generates well-resolved trees that do not reflect phylogeny. *Integrative Zoology*. 9: 531-541.
- Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: A practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society* 57: 289-300.
- Bourdon E. 2006. L'avifaune du Paléogène des phosphates du Maroc et du Togo: diversité, systématique et apports à la connaissance de la diversification des oiseaux modernes (Neornithes). Doctoral thesis. Paris, Muséum National d'Histoire Naturelle.
- Brazeau MD. 2011. Problematic character coding methods in morphology and their effects. *Biological Journal of the Linnean Society* 104: 489-498.
- Brower AVZ, de Pinna MCC. 2014. About nothing. *Cladistics* 30: 330-336.
- Callaway E. 2011. Ancient DNA reveals secrets of human history. *Nature News* 476: 136-137.
- Canoville A, Laurin M. 2010. Evolution of humeral microanatomy and lifestyle in amniotes, and some comments on palaeobiological inferences. *Biological Journal of the Linnean Society* 100: 384-406.
- Cao N, Zaragüeta R, Vignes-Lebbe R. 2007. Hierarchical representation of hypotheses of homology. *Geodiversitas* 29: 5-15.
- Carine MA, Scotland RW. 1999. Taxic and transformational homology: different ways of seeing. *Cladistics* 15: 121-129.
- Chappill JA. 1989. Quantitative Characters in Phylogenetic Analysis. *Cladistics* 5: 217-234.

- Corvez A. 2012. L'origine de la mégaphylle chez les Monilophytes. Paris, Muséum National d'Histoire Naturelle.
- Estabrook GF, McMorris FR, Meacham CA. 1985. Comparison of undirected phylogenetic trees based on subtrees of four evolutionary units. *Systematic Zoology* 34: 193-200.
- Faith DP. 1992. Systematics and Conservation: On Predicting the Feature Diversity of Subsets of Taxa. *Cladistics* 8: 361-373.
- Farris JS. 1989. The Retention Index and the Rescaled Consistency Index. *Cladistics* 5: 417-419.
- Farris JS. 1990. Phenetics in Camouflage. *Cladistics* 6: 91-100.
- Farris JS. 1997. Who really is a statistician? In: *Sixteenth meeting of the Willi Hennig Society*. George Washington Univ. Washington, DC.
- Farris JS. 2012. 3ta sleeps with the fishes: Book review. *Cladistics* 28: 422-436.
- Farris JS, Källersjö M, Albert VA, Allard M, Anderberg A, Bowditch B, Bult C, Carpenter JM, Crowe TM, Laet J. 1995. Explanation. *Cladistics* 11: 211-218.
- Farris JS, Kluge AG. 1998. A/The Brief History of Three-Taxon Analysis. *Cladistics* 14: 349-362.
- Farris JS, Kluge AG, Eckardt MJ. 1970. A Numerical Approach to Phylogenetic Systematics. *Systematic Biology* 19: 172-189.
- Felsenstein J. 1985. Phylogenies and the comparative method. *The American naturalist* 125: 1-15.
- Felsenstein J. 1988. Phylogenies and quantitative characters. *Annual Review of Ecology and Systematics* 19: 445-471.
- Fitch WM. 1971. Toward Defining the Course of Evolution: Minimum Change for a Specific Tree Topology. *Systematic Biology* 20: 406-416.
- Garland TJ, Dickerman AW, Janis CM, Jones JA. 1993. Phylogenetic Analysis of Covariance by Computer Simulation. *Systematic Biology* 42: 265-292.
- Goldberg EE, Igić B. 2008. On Phylogenetic Test of Irreversible Evolution. *Evolution* 62: 2727-2741.
- Grand A, Corvez A, Duque Velez LM, Laurin M. 2013a. Phylogenetic inference using discrete characters: performance of ordered and unordered parsimony and of three-item statements. *Biological Journal of the Linnean Society* 110: 914-930.
- Harvey AW. 1992. Three-taxon statements: more precisely, an abuse of parsimony? *Cladistics* 8: 345-354.
- Hauser DL, Presch W. 1991. The effect of ordered characters on phylogenetic reconstruction. *Cladistics* 7: 243-265.
- Hennig W. 1966. *Phylogenetic Systematics*. University of Illinois Press.
- Kearney M. 2002. Fragmentary taxa, missing data, and ambiguity: mistaken assumptions and conclusions. *Systematic biology* 51: 369-381.
- Kitching IJ. 1998. *Cladistics: The Theory and Practice of Parsimony Analysis*. Oxford: Oxford University Press.
- Kluge A. 1994. Moving targets and shell games. *Cladistics* 10: 403-413.
- Kohlsdorf T, Lynch VJ, Rodrigues MT, Brandley MC, Wagner GP. 2010. Data and Data Interpretation in the Study of Limb Evolution: a Reply to Galis et al. on the Reevolution of Digits in the Lizard Genus *Bachia*. *Evolution* 64: 2477-2485.
- De Laet J, Smets E. 1998. On the three-taxon approach to parsimony analysis. *Cladistics* 14: 363-381.
- Laurin M. 2004. The Evolution of Body Size, Cope's Rule and the Origin of Amniotes. *Systematic Biology* 53: 594-622.

- Laurin M, Germain D. 2011. Developmental Characters in Phylogenetic Inference and Their Absolute Timing Information. *Systematic Biology* 60: 630-644.
- Laurin M, Gussekloo SWS, Marjanović D, Legendre L, Cubo J. 2012. Testing gradual and speciation models of evolution in extant taxa: the example of raites: Testing models of evolution. *Journal of Evolutionary Biology* 25: 293-303.
- Maddison WP. 1991. Squared-Change Parsimony Reconstructions of Ancestral States for Continuous-Valued Characters on a Phylogenetic Tree. *Systematic Biology* 40: 304-315.
- Maddison W, Maddison D. 2014. Mesquite: a modular system for evolutionary analysis, version 3.0.
- Mickevich MF. 1982. Transformation Series Analysis. *Systematic Biology* 31: 461-478.
- Mickevich MF, Johnson MS. 1976. Congruence Between Morphological and Allozyme Data in Evolutionary Inference and Character Evolution. *Systematic Biology* 25: 260-270.
- Mickevich MF, Platnick NI. 1989. On the information content of classifications. *Cladistics* 5: 33-47.
- Miller AI, Foote M. 1996. Calibrating the Ordovician Radiation of Marine Life: Implications for Phanerozoic Diversity Trends. *Paleobiology* 22: 304-309.
- Müller J, Reisz RR. 2006. The Phylogeny of Early Eureptiles: Comparing Parsimony and Bayesian Approaches in the Investigation of a Basal Fossil Clade. *Systematic Biology* 55: 503-511.
- Nelson G, Ladiges PY. 1992. Information content and fractional weight of three-item statements. *Systematic biology* 41: 490-494.
- Nelson GJ, Ladiges PY. 1996. Paralogy in cladistic biogeography and analysis of paralogy-free subtrees. *American Museum novitates* 3167: 58.
- Nelson G, Platnick NI. 1991. Three-Taxon Statements: A More Precise Use of Parsimony? *Cladistics* 7: 351-366.
- Nelson G, Williams DM, Ebach MC. 2003. A question of conflict: Three-item and standard parsimony compared. *Systematics and Biodiversity* 1: 145-149.
- Newman MEJ, Sibani P. 1999. Extinction, diversity and survivorship of taxa in the fossil record. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 266: 1593-1599.
- Nopcsa FB. 1923. Reversible and Irreversible Evolution; a Study based on Reptiles. *Proceedings of the Zoological Society of London* 93: 483-1097.
- Patterson C, Johnson GD. 1995. The intermuscular bones and ligaments of teleostean fishes. *Smithsonian contributions to zoology (USA)* 559: 83.
- Platnick NI, Humphries CJ, Nelson G, Williams DM. 1996. Is Farris Optimization perfect?: three-taxon statements and multiple branching. *Cladistics* 12: 243-252.
- Poe S. 2006. Test of Von Baer's Law of the Conservation of Early Development. *Evolution* 60: 2239-2245.
- Poe S, Wake MH. 2004. Quantitative tests of general models for the evolution of development. *The American naturalist* 164: 415-422.
- Robinson DF, Foulds LR. 1981. Comparison of phylogenetic trees. *Mathematical Biosciences* 53: 131-147.
- Rohlf J. 1982. Consensus indices for comparing classifications. *Mathematical Biosciences* 59: 131-144.
- Siebert DJ, Williams DM. 1998. Recycled. *Cladistics* 14: 339-347.

- Simon C. 1983. A New Coding Procedure for Morphometric Data with an Example from Periodical Cicada Wing Veins. In: Felsenstein J ed. *Numerical Taxonomy*. NATO ASI Series. Springer Berlin. 378-382.
- Slowinski JB. 1993. "Unordered" versus "ordered" characters. *Systematic Biology* 42: 155-165.
- Sokal RR, Rohlf FJ. 1981. Taxonomic Congruence in the Leptopodomorpha Re-Examined. *Systematic Zoology* 30: 309-325.
- Stevens PF. 1991. Character states, morphological variation, and phylogenetic analysis: a review. *Systematic Botany* 16: 553-583.
- Swofford D. 2003. *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Version 4. Sinauer Associates.
- Tavaré S. 1986. Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences. In: *Lectures on Mathematics in the Life Sciences*. American Mathematical Society. 57-86.
- Thiele K. 1993. The Holy Grail of the Perfect Character: The Cladistic Treatment of Morphometric Data. *Cladistics* 9: 275-304.
- Udovicic F, Mcfadden GI, Ladiges PY. 1995. Phylogeny of *Eucalyptus* and *Angophora* Based on 5S rDNA Spacer Sequence Data. *Molecular Phylogenetics and Evolution* 4: 247-256.
- Uhlenbeck GE, Ornstein LS. 1930. On the Theory of the Brownian Motion. *Physical Review* 36: 823-841.
- Whiting M, Bradler S, Maxwell T. 2003. Loss and recovery of wings in stick insects. *Nature* 421: 264-267.
- Wiens JJ. 2001. Character analysis in morphological phylogenetics: problems and solutions. *Systematic Biology* 50: 689-699.
- Wilkinson M. 1992. Ordered versus unordered characters. *Cladistics* 8: 375-385.
- Wilkinson M. 1994. Three-taxon statements: when is a parsimony analysis also a clique analysis? *Cladistics* 10: 221-223.
- Williams D. 1996. Characters and cladograms. *Taxon* 45: 275-283.
- Williams D, Ebach MC. 2006. The data matrix. *Geodiversitas* 28: 409-420.
- Zaragüeta R, Bourdon E. 2007. Three-item analysis: Hierarchical representation and treatment of missing and inapplicable data. *Comptes Rendus Palevol* 6: 527-534.
- Zaragüeta R, Ung V, Grand A, Vignes-Lebbe R, Cao N, Ducasse J. 2012. LisBeth: New cladistics for phylogenetics and biogeography. *Comptes Rendus Palevol* 11: 563-566.

Received: 4 June 2014

Revised and accepted: 3 February 2015

Published online: 7 July 2015

Editor: M. Brazeau

Online supplementary information

S1. Vba script used to generate the individual Nexus text files from the columns of discretized simulated data in Excel obtained using *S2* and 7 and compiled into *S4* (the latter is processed using *S1* to get individual Nexus text files).

S2. Spreadsheet using the vba script presented in *S1* to produce Nexus files for parsimony analysis.

S3. Spreadsheet using the vba script presented in *S1* to produce Nexus files for 3ta analysis.

S4. The 1800 PAUP nexus files used in our parsimony analyses. Each sheet presents 100 matrices corresponding to one tree. Each sheet is identified by two numbers; the first number refers to the topology (1, pectinate; 2, symmetrical; 3, equiprobable), and second, branch length (sorted from shortest to longest). Thus, sheet 2.3 presents the matrices obtained from the symmetrical tree (Fig. 5B) with the third set of branch lengths (1). For the third topology, there are 12 branch lengths, corresponding to three branch lengths associated with Fig. 5C, six branch lengths associated with Fig. 5D, and with Fig. 5E-G).

S5. Resolving power and artefactual resolution in trees of various branch lengths.

S6. Frequency distribution of character states, based on 10000 characters in 21 taxa, obtained by discretization of continuous data generated by Brownian motion on the equiprobable tree (Fig. 5C, with outgroup branch length set at full tree depth).

S7. Individual ITRIs of the 1800 matrices (each of which was analysed using the three methods, for 5400 ITRI values for resolving power, and as many for artefactual resolution) used to compute resolving power and artefactual resolutions.

S8. Impact of branch length and topology on resolving power and artefactual resolution. Each cell value represents ITRI means for 100 matrices.