Systematic palaeontology (Palaeobotany)

Three-item analysis: Hierarchical representation and treatment of missing and inapplicable data

René Zaragüeta-Bagils [a,*], Estelle Bourdon [b,c]

[a] *Laboratoire "Informatique et systématique", UMR 5143 CNRS, université Pierre-et-Marie-Curie (Paris-6), 12, rue Cuvier, 75005 Paris, France*
[b] *UMR 7179 CNRS, université Pierre-et-Marie-Curie (Paris-6), 4, place Jussieu, 75005 Paris cedex 05, France*
[c] *Collège de France, 11, place Marcelin-Berthelot, 75231 Paris cedex 05, France*

Abstract

Matrix-based methods, including parsimony programs, represent and treat missing data in the same way, such that every character-state is possible, and inapplicable data, such that every character state is impossible. This is because the hierarchical nature of homology assessments cannot be represented in taxon–character matrices. We show that the hierarchical representation of hypotheses of homology used in three-item analysis permits the accurate treatment of missing and inapplicable data. *To cite this article: R. Zaragüeta-Bagils, E. Bourdon, C. R. Palevol 6 (2007).*
© 2007 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

Résumé

Analyse à trois éléments : représentation hiérarchique et traitement des données manquantes et non applicables. Les méthodes qui utilisent des matrices, dont les programmes de parcimonie, représentent et traitent de la même manière les données manquantes, pour lesquelles tout état de caractère est possible, et les données non applicables, pour lesquelles tout état de caractère est impossible. Ce problème vient du fait que la nature hiérarchique des hypothèses d'homologie ne peut pas être représentée dans une matrice taxons–caractères. Nous montrons que la représentation hiérarchique des hypothèses d'homologie utilisée en analyse à trois éléments permet de traiter correctement les données manquantes et non applicables. *Pour citer cet article : R. Zaragüeta-Bagils, E. Bourdon, C. R. Palevol 6 (2007).*
© 2007 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

---

* Corresponding author.
E-mail addresses: rzb@ccr.jussieu.fr (R. Zaragüeta-Bagils), bourdon@mnhn.fr (E. Bourdon).

1. Introduction

Palaeontology or, more precisely, systematic studies that include fossils, are particular in two respects. The first is that fossils are the unique source of temporal information in phylogenetic studies [32]. The temporal information conveyed by fossils and its measurement will be dealt with elsewhere [13]. The second peculiarity of palaeontology is the intrinsic incompleteness of fossil specimens, which results in the impossibility of observing features not preserved. These features are coded with question marks in matrix-based computer programs. However, other circumstances involve the use of question marks [7,10]. These include:

- ignorance due to either the unavailability of specimens for a given taxon or to the difficulty relating two structures in some specimens as homologous;
- polymorphism, i.e. taxa that show instances of more than one character state;
- inapplicability of character states. This situation appears when a given character cannot be scored due to the absence of the structure in some organisms or taxa. For example, if the relationships among Embryophyta are analysed, the character "flowers" would be inapplicable in e.g., Filicopsida, as flowers are not present.

This study is concerned with true missing data, i.e. characters relating parts that are either not preserved or ignored, and inapplicable character states. In matrix-based methods, these cases are represented by question marks (generally coded as "?", "–", "N" or "*"). As the same representation is used, current phylogenetic methods, i.e. parsimony analysis, treat unknown and inapplicable cases in exactly the same way. The relevance of these identical representation and treatment is examined here.

2. Representation of hypotheses of homology

Systematics aims to establish hypotheses of homology and combine them in order to infer relationships among organisms and taxa. These relationships, and the taxa themselves, can be seen as the same [18], homologies being parts of taxa. Homology is the relationship that links homologs [2,17,18,31]. A homolog is, as defined by Owen [31], "the same organ in different animals or plants under every variety of form and function". It is clear from Owen's definition that a homolog is a class of objects. Homology is, formally, a relationship among classes. Homology and character, on the one side, and homolog
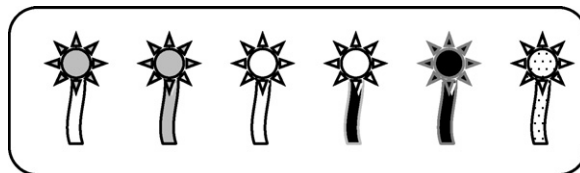

Fig. 1. Theoretical taxonomic sampling including terminal taxa or operational taxonomic units (OTUs). Hypotheses of primary homology concern grouping OTUs into classes, called here homologues, and defining relationships between these classes.
Fig. 1. Échantillonnage taxonomique théorique incluant des taxons terminaux ou unités taxonomiques opérationnelles (UTOs). Les hypothèses d'homologie primaire consistent à grouper les UTOs dans des classes appelées ici homologues, et à définir les relations entre ces classes.

and character state, on the other side, can be regarded as synonyms [2]. However, formalisation must not impose a constraint on the way systematists think. On the contrary, it is intended to express the idea of the systematist in a formal language, i.e. a language devoid of ambiguity [12].

A taxonomic sample of, say, plants is illustrated in the theoretical example of Fig. 1. A hypothesis of primary homology [24] may be expressed by the systematist as "*I think that there are some starry forms that are grey, and some others that are white, among the sample. I think that these colours may be a relevant argument for grouping some of the plants of the sample*." This hypothesis is illustrated in Fig. 2. Some of the operational taxonomic units (OTUs) are grouped into homologues, which correspond to the concepts defined by the colour of the starry form. The white starry form of a given OTU is a concrete object and constitutes an instance of the concept "white starry form". The systematist has no theory of relationships concerning organisms with a black or dotted starry form, which are excluded from the homologues.
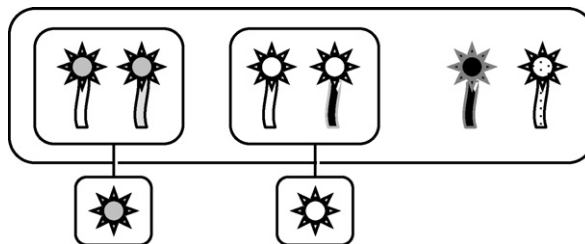

Fig. 2. The hypothesis of primary homology corresponds to the theory of the systematist. The formal classificatory structure that best represents this hypothesis is the hierarchy. Here, the theory is that the colour of the starry form is a relevant argument for grouping some organisms among the sample.
Fig. 2. L'hypothèse d'homologie primaire constitue la théorie du systématicien. La structure classificatoire formelle qui représente le mieux cette hypothèse est la hiérarchie. Ici, la théorie est que la couleur de la forme étoilée est un argument pertinent pour grouper certains organismes au sein de l'échantillon.

3. Representation vs. coding

The problem of generalising a series of observations into hypotheses of homology is often considered a matter of coding. Systematists usually speak of coding characters into a matrix, and think about the most relevant way to code missing characters (e.g., [27]). Kitching et al. [10] define coding as "the conversion of original observations into a discrete alphanumeric format suitable for cladistic analysis". Thus, coding refers to the simple transcription of some information by changing the usual language into the conventional signs of a code. However, there is a stage apparently absent in Kitching et al.'s definition and in other theoretical works dealing with the subject [1,6]. This stage concerns the representation of the knowledge that a systematist extracts from a series of observations. A representation is, in the use given here, "the act of making sensible an absent object or a concept via an image, a figure or a sign" (definition translated from the Petit Robert™ French dictionary). Systematists express hypotheses of homology, not series of observations. A hypothesis of homology is the representation of a concept that relates different objects as being the same [19,26,31]. Only after this stage, a character and its character states are coded using alphanumeric symbols. Coding is thus limited to the choice of symbols used to fill data matrices. Incidentally, "data matrices" are not matrices, but tables; moreover, they do not contain data, i.e. statements about unique objects, but concepts representing the "sameness" of different objects. In a "data matrix", homologues are usually coded as a number. In the case illustrated here, OTUs with a grey starry part might be given the code "1". The confusion between coding and representing and, in general, between objects and concepts, is probably at the source of a number of mistakes and misunderstandings in systematics and other disciplines [5].

4. Missing data

The problem of the inclusion of fossil taxa has been discussed [23] and there exists wide agreement about the importance of including fossils in phylogenetic studies in spite of their incompleteness. Homology statements cannot be assessed if the parts of organisms concerned in the homology relationship are unknown [9]. This has largely been seen as a source of problems [1,5,6,19,24,26,27], mainly because large amounts of missing data tend to increase the number of most parsimonious trees found, often decreasing the resolution in consensus trees. New algorithms may only obscure the procedure in order to hide misrepresentations, however, and remain unable to
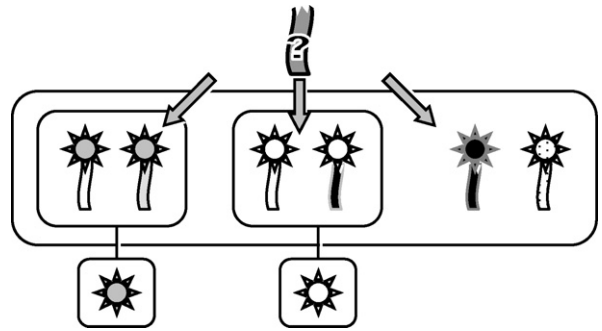

Fig. 3. An organism which lacks its upper part is considered here. It is impossible to assign it either to one of the homologs or to the root, i.e. the class containing all the specimens assigned to the homologs plus the specimens not belonging to any homolog. However, we cannot exclude that this organism belongs to any of the classes of the hierarchy.

Fig. 3. Un organisme avec la partie supérieure manquante est considéré ici. Il est impossible de l'attribuer à l'un des homologues ou à la racine, c'est-à-dire la classe contenant tous les spécimens attribués aux homologues ainsi que les spécimens n'appartenant à aucun homologue. Cependant, on ne peut exclure que cet organisme appartienne à l'une des classes de la hiérarchie.

propose solutions when the relevant information is not there. As Strauss et al. [29] write, "despite the ample literature on missing-value estimation, there is still little empirical guidance for researchers", and the situation will remain, because if empirical evidence is not available, empirical guidance cannot be given in historical biology.

Fig. 3 shows a specimen, which lacks its upper part. It is impossible to know if there was a starry part in the living organism. Consequently, the question of its assignment to a homolog is pointless. However, the incomplete specimen cannot be excluded from any of the classes of the hypotheses of homology that concern the starry part. If a more complete specimen was found, it might fall under any of the concepts of the hierarchy, in Frege's terminology [5]. Thus, missing features must be regarded as potentially belonging to all states because they cannot be excluded from any of them. In other terms, missing data represent the idea that *every state is possible*.

5. Inapplicable data

Inapplicable character states appear in matrices when some character states are sub-concepts of a more inclusive state, called a super-concept [12] (Fig. 4). The hypothesis of primary homology that the systematist wishes to represent may be expressed as "I think that the presence of a starry form is a relevant argument for grouping some of the organisms of my sample. Among them, and only among them, some may be grouped
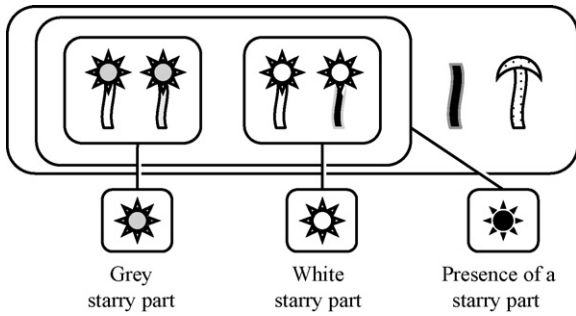
Fig. 4. Inapplicable data appear when some character states depend on more inclusive ones. In this example, the states "grey" or "white" depend on the state "starry part".
Fig. 4. Les données non applicables apparaissent lorsque des états de caractère sont dépendants d'états de caractère plus inclusifs. Dans cet exemple, les états de caractère « couleur grise » ou « couleur blanche » dépendent de l'état de caractère « partie étoilée ».

together because they share a grey starry form, while others may be grouped because they share a white starry form." It has been shown that matrices, i.e. tables, cannot represent this kind of hierarchical information [2]. In order to bypass this inadequacy, systematists try to represent something similar using a complicated procedure (Table 1). The hierarchical hypothesis of primary homology is broken up into series of independent binary characters (Fig. 4). The first one is intended to represent the presence of a starry form. The second and third characters represent the presence of a starry form of a particular colour. Question marks occur when organisms which lack a starry part have to be represented for the characters that require the presence of a starry part. In this case, the systematist is certain that both the character state "grey starry form" and the character state "white starry form" are absent for organisms devoid of starry part. In other words, inapplicable character states correspond to the case for which *every state is impossible*.

6. Representation of missing and inapplicable data in matrix-based methods

Matrix-based programs in general, and parsimony in particular, represent and code in the same way missing data, for which *every state is possible*, and inapplicable character states, for which *every state is impossible*. In addition, matrix-based methods apply the same treatment in order to minimise steps, no matter what the source of a question mark is. Some authors have suggested that a solution to this problem cannot be found until new algorithms are available [8,14,15]. However, the problem does not concern better algorithms, but the

Table 1
Matrix representation of a hypothesis of homology including a hierarchical structure of character states
Tableau 1
Représentation en matrice d'une hypothèse d'homologie incluant une structure hiérarchique d'états de caractère

	starry form	s.f. grey	s.f. white
	0	?	?
	0	?	?
	1	1	0
	1	1	0
	1	0	1
	1	0	1

The complex hypothesis is broken up into series of independent binary characters. The hierarchical dependence is lost, and question marks appear where there was none.
L'hypothèse complexe est transformée en une série de caractères binaires indépendants. La dépendance hiérarchique est perdue, et des points d'interrogation apparaissent alors qu'il n'y en avait aucun initialement.

relevance of representation of hypotheses of homology in standard parsimony.

Lee and Bryant [14] have accurately identified the core problem of the defective representation of inapplicable states in matrix-based methods, e.g., parsimony analysis, even if the idea had already been suggested before [8,26]. They affirm that inapplicable character

states entail "a character hierarchy". However, they do not manage to find a relevant solution, because their rationale is based on a transformational viewpoint. After Lee and Bryant, "the recognition of a character with two or more character states implies that transformations have occurred between those states [. . .] The state that becomes a synapomorphy is determined by rooting the tree" (p. 374). On the other hand, Lee and Bryant state in the same page that "in cladistic analysis it is assumed that organismic diversity forms a nested taxic hierarchy of groups within groups. This hierarchy is inferred empirically using characters, which form a complementary nested hierarchy."

A hierarchy is a classificatory structure, i.e. a collection of non-empty classes such as every individual belongs to at least one class. Moreover, a hierarchy is defined by the following properties: a class called root contains all the individuals; there is a class that contains a single individual, for every individual, called singleton (the OTU of systematists); most importantly, the intersection between any two classes of the hierarchy is either empty or one of the classes [4]. There is an isomorphism, i.e. an equivalence, between a hierarchy and a rooted tree [12].

Transformation processes, however, do not produce hierarchical relationships. This is why, in a cladistic context, taxa are not supposed to transform into one another; they differentiate from more general to more particular, i.e. they show hierarchical relationships. Lee and Bryant clearly show in their Fig. 2 that the representation expressing their idea of character is a rooted tree (see Fig. 5A), or its equivalent hierarchy (Fig. 5B). It has been shown that matrices cannot represent hierarchies without distorting hypotheses of homology [2]. The only way to correctly represent hypotheses of homology is to have at one's disposal a method that *understands* the hierarchical nature of homology assessments. The only method that permits hierarchies is three-item analysis.

7. Representation of missing and inapplicable data in three-item analysis

Cladistic analysis is not called analysis by chance. The analytical method has roots very deep in time [3]. In his Discours de la méthode pour bien conduire sa raison et chercher la vérité dans les sciences, plus la Dioptrique, les Météores et la Géométrie qui sont des essais de cette méthode, René Descartes gives the methodological principles that constitute the base of the analytical reasoning. His second principle states that, in front of a complex problem, one has "to divide each of the difficulties under examination into as many parts as possible, and as might be necessary for its adequate solution"; the third principle states that "to conduct my thoughts in such order that, by commencing with objects the simplest and easiest to know, I might ascend by little and little, and, as it were, step by step, to the knowledge of the more complex." The analytical method can be reformulated as follows: prior to solving a complex problem, one must decompose it into a set of simpler problems, each admit-
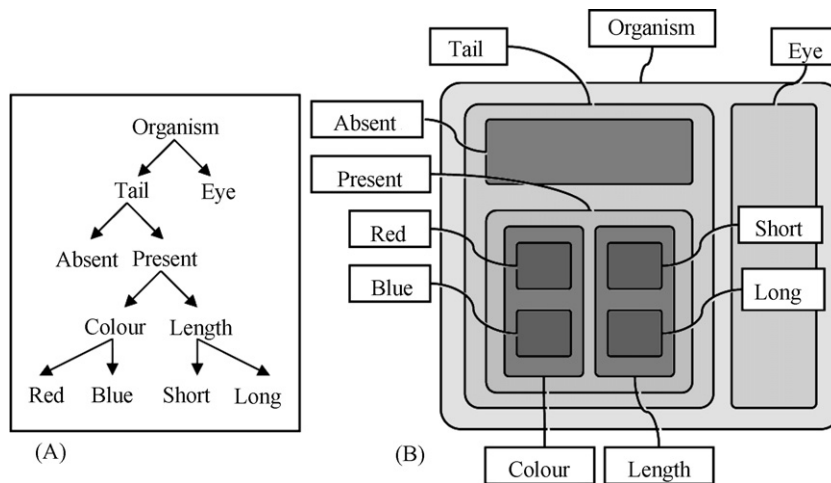

Fig. 5. . (**A**) Hierarchical relationships between "a part and its character variables", following Lee and Bryant [14], Fig. 2, modified; this representation corresponds to a rooted tree. (**B**) Hierarchical representation isomorphic of the tree shown in (**A**). The character is the hierarchy and the character states are the classes of the hierarchy [2].

Fig. 5. (**A**) Relations hiérarchiques entre « une partie et ses variables de caractère », selon Lee et Bryant [14], Fig. 2, modifiée ; cette représentation correspond à un arbre raciné. (**B**) Représentation hiérarchique isomorphe de l'arbre montré en (**A**). Le caractère correspond à la hiérarchie et les états de caractère constituent les classes de la hiérarchie [2].
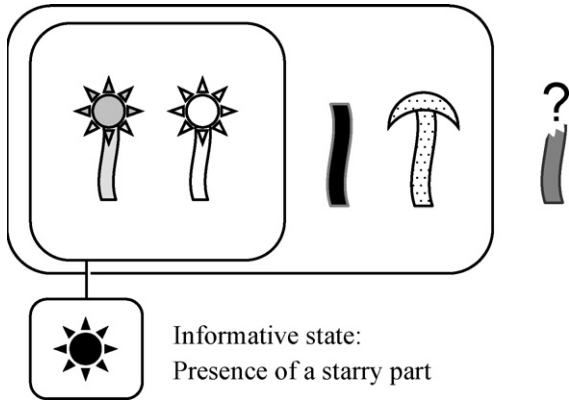
Fig. 6. Hierarchical representation of a hypothesis of primary homology in three-item analysis.

Fig. 6. Représentation hiérarchique d'une hypothèse d'homologie primaire en analyse à trois éléments.
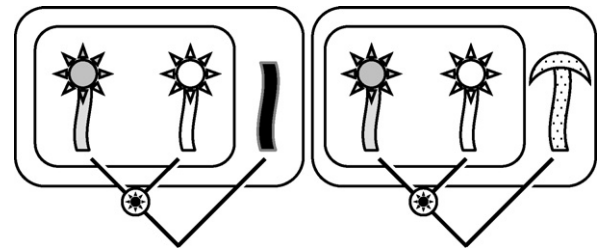

Fig. 7. Three-item statements (3is) resulting from the character shown in Fig. 6.

Fig. 7. Assertions à trois éléments (3is) résultant du caractère montré en Fig. 6.

ting of solution, and combine these partial solutions in order to obtain a general solution.

Finding relationships of a set of taxa is the complex problem that has to be solved. Analysis is performed in order to decompose this global problem into a suite of characters or primary hypotheses of homology [2]. This means that the parts that compose taxic relationships are the characters [17,18]. One of the main constraints of the analytical method is that, because the whole, i.e. relationships among taxa, results from the combination of the partial solutions, i.e. secondary homologies, nothing can be learnt about the latter from the former. However, Lee and Bryant claim that "the state that becomes a synapomorphy is determined by rooting the tree." This assertion violates the analytical principle, because hierarchical relationships among character states are established from the relationships among taxa, which are themselves built from the relationships among character states.

This circularity in Lee and Bryant's reasoning is not their only misunderstanding. In parsimony and other matrix-based methods, character states define neither hierarchies, nor other formal kind of classificatory structure. This is due to the existence of reversals considered as synapomorphies. In this case, the same state is regarded as plesiomorphous in a part of the tree and apomorphous in another. This leads to a non-sense in hierarchical terms. As an analogy, France is a class inside Europe characterised by French people; France itself contains Paris, a class characterised by Parisians; the class "Paris" would then include a class characterised by French people again!

Three-item analysis (3ia) is a phylogenetic method that uses hierarchical hypotheses of homology [20,21] and thus respects analytical principles. In Fig. 6, the conjecture of the existence of a clade is based on the presence of a starry part. No hypothesis is made concerning organisms devoid of a starry part. Concerning the
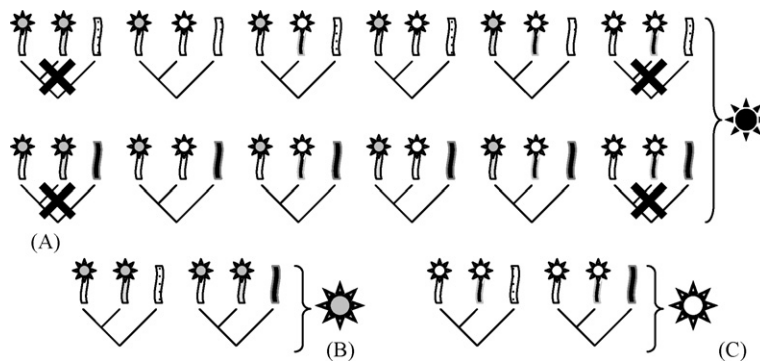

Fig. 8. 3is deduced from the character shown in Fig. 4. (**A**) 3is deduced from the state "presence of a starry form". (**B**) 3is deduced from the state "grey starry form". (**C**) 3is deduced from the state "white starry form". Dependence among character states permits to remove part of the logical redundancy in the most inclusive state.

Fig. 8. 3is déduits du caractère montré dans la Fig. 4. (**A**) 3is déduits de l'état de caractère « présence de partie étoilée ». (**B**) 3is déduits de l'état de caractère « partie étoilée grise ». (**C**) 3is déduits de l'état de caractère « partie étoilée blanche ». La dépendance entre états de caractère permet d'éliminer une partie de la redondance logique au niveau de l'état le plus inclusif.

OTU that lacks its upper part, the presence of a starry part cannot be set aside, nor can its absence.

In order to maximize analysis, each hypothesis of primary homology is decomposed into three-item statements (3is), which correspond to the minimum hierarchical information possible [21]. The character shown in Fig. 6 produces two 3is, which are illustrated in Fig. 7. Note that the OTU which lacks its upper part is not included in the resulting 3is: missing data are just treated as missing in three-item analysis.

Fig. 8 shows the 3is deduced from the character represented in Fig. 4. This character has two states, grey starry form (Fig. 8B) and white starry form (Fig. 8C) that are dependent on the presence of a starry form (Fig. 8A). Nelson and Ladiges [22] showed that logical redundancy among 3is can be treated with fractional weighting. The value of the fractional weight (FW) is expressed by the formula: 2/(number of terminals included in the state). In Fig. 8A, the 3is deduced from the state "presence of a starry form" bear a fractional weight of ½, because only half of the 3is are necessary to deduce the others. 3ia, as illustrated in Fig. 8, distinguishes a multi state character from a series of independent binary characters through the removal of redundancy: the four 3is of the most inclusive state that are also deduced from the two less inclusive states are eliminated. The final fractional weight of the 3is deduced from the state "presence of a starry form" is thus 3/4.

8. Criticisms

3ia has been criticised because of the addition of question marks [11]. Kluge [11] writes: "The matrix resulting from the three-taxon transformation has considerable missing data, where none existed before. This amounts to the unscientific exercise of throwing away observations". As recently stated, a computer program for 3ia is being developed that uses no matrix [2]. Moreover, it has been shown that 3ia finds the correct answer even if a matrix is used (e.g., [16,25]). The reason is that parsimony optimises missing data as if it was potentially any of the states. In equivalent compatibility terms [30], questions marks are compatible with any state. They lack any empirical content because they do not forbid anything [28]. On the other hand, parsimony creates question marks were none existed before, so as to deal with the hierarchical dependence among character states of a multistate character. Parsimony then treats these artificial question marks as missing data. However, no missing data were present in the original hypothesis, as specified by Lee and Bryant [14]. Moreover, inapplicable data have a very rich empirical content, because they forbid the assignment to any of the dependent states. We conclude that parsimony (or, in general, matrix-based methods) creates ambiguity and distortion of hypotheses where none existed before, as opposed to Kluge's statement.

References

[1] A.V.Z. Brower, V. Schawaroch, Three steps of homology assessement, Cladistics 12 (1996) 265–272.

[2] N. Cao, R. Zaragüeta Bagils, R. Vignes-Lebbe, Hierarchical representation of hypotheses of homology, Geodiversitas 29 (2007) 5–15.

[3] R. Descartes, Discours de la méthode pour bien conduire sa raison et chercher la vérité dans les sciences, plus la dioptrique les météores et la géométrie qui sont des essais de cette méthode, Imprimerie de I. Maire, Leyde, 1637.

[4] E. Diday, Une représentation visuelle des classes empiétantes : les pyramides, Rapport de recherche, INRIA, Rocquencourt, 1984, p. 75.

[5] G. Frege, Écrits Logiques et Philosophiques, Éditions du Seuil, Paris, 1971.

[6] J.V. Freudenstein, Characters, states, and homology, Syst. Biol. 54 (2006) 965–973.

[7] L. Grande, W.E. Bemis, A comprehensive phylogenetic study of amiid fishes (Amiidae) based on comparative skeletal anatomy. An empirical search for interconnected patterns of natural history, Society of Vertebrate Paleontology Memoir 4, J. Vertebr. Paleontol. 18 (Suppl.) (1998) 1–690.

[8] J.A. Hawkins, C.E. Hughes, R.W. Scotland, Primary homology assessment, characters and character states, Cladistics 13 (1997) 275–283.

[9] M. Kearney, J.M. Clark, Problems due to missing data in phylogenetic analyses including fossils: a critical review, J. Vertebr. Palaeontol. 23 (2003) 263–274.

[10] I.J. Kitching, P.L. Forey, C.J. Humphries, D.W. Williams, Cladistics: The Theory and Practice of Parsimony Analysis, second edn, Oxford University Press, Oxford, UK, 1998.

[11] A.G. Kluge, Three-taxon transformation in phylogenetic inference: ambiguity and distortion as regards explanatory power, Cladistics 9 (1993) 246–259.

[12] J. Lebbe, Représentation des concepts en biologie et médecine. Introduction à l'analyse des connaissances et à l'identification assistée par ordinateur, thèse, université Paris-6, 1991, p. 281.

[13] H. Lelièvre, R. Zaragüeta-Bagils, I. Rouget, Temporal information, fossil record and phylogeny, C. R. Palevol. 6–7 (2007).

[14] D.-C. Lee, H.N. Bryant, A reconsideration of the coding of inapplicable characters: assumptions and problems, Cladistics 15 (1999) 373–378.

[15] D.R. Maddison, Missing data versus missing characters in phylogenetic analysis, Syst. Biol. 42 (1993) 576–581.

[16] G.J. Nelson, Reply, Cladistics 9 (1993) 261–265.

[17] G.J. Nelson, La systématique et l'homologie, in: P. Tassy, H. Lelièvre (Eds.), Caractères, Société française de systématique, Paris, 1994, pp. 5–28.

[18] G.J. Nelson, Homology and Systematics, in: B.K. Hall (Ed.), The Hierarchical Basis of Comparative Biology, Academic Press, San Diego, 1994, pp. 101–149.

[19] G.J. Nelson, N.I. Platnick, Systematics and Biogeography: Cladistics and Vicariance, Columbia University Press, New York, 1981.

[20] G.J. Nelson, P.Y. Ladiges, Three-area statements: standard assumptions for biogeographic analysis, Syst. Zool. 40 (1991) 470–485.

[21] G.J. Nelson, N.I. Platnick, Three-taxon statements: a more precise use of parsimony? Cladistics 7 (1991) 351–366.

[22] G.J. Nelson, P.-Y. Ladiges, Information content and fractional weight of three-item statements, Syst. Biol. 41 (1992) 490–494.

[23] C. Patterson, Significance of fossils in determining evolutionary relationships, Annu. Rev. Ecol. Syst. 12 (1981) 195–223.

[24] M.C.C. de Pinna, Concepts and tests of homology in the cladistic paradigm, Cladistics 7 (1991) 367–394.

[25] N.I. Platnick, Character optimization and weighting: differences between the standard and three-taxon approaches to phylogenetic inference, Cladistics 9 (1993) 267–272.

[26] N.I. Platnick, Philosophy and the transformation of cladistics, Syst. Zool. 28 (1979) 537–546.

[27] F. Pleijel, On character coding for phylogeny reconstruction, Cladistics 11 (1995) 309–315.

[28] K. Popper, Conjectures et Réfutations : la croissance du savoir scientifique, Payot, Paris, 1972.

[29] R.E. Strauss, M.N. Atanasov, J. Alves de Oliveira, Evaluation of the principal-component and expectation-maximization methods for estimating missing data in morphometrics studies, J. Vertebr. Paleontol. 23 (2003) 284–296.

[30] M. Wilkinson, Three-taxon statements: when is a parsimony analysis also a clique analysis? Cladistics 10 (1994) 221–223.

[31] D.W. Williams, Homologues and homology, phenetics and cladistics: 150 years of progress, in: D.W. Williams, P.L. Forey (Eds.), Milestones in Systematics, CRC – The Systematics Association, London, 2004, pp. 191–224.

[32] R. Zaragüeta Bagils, H. Lelièvre, P. Tassy, Temporal paralogy, cladograms, and the quality of the fossil record, Geodiversitas 26 (2004) 381–389.