Dance of the Cave Bear: Honouring the Scientific Legacy of Björn Kurtén

### The corrected parsimony phylogenetic method following Hennig's principle for analyzing morphological characters

#### Yue Zhang

Institute of Biotechnology, P.O. Box 56, FI-00014 University of Helsinki, Finland (e-mail: yzhangfossil@gmail.com)

Received 2 Mar. 2024, final version received 17 Sep. 2024, accepted 18 Sep. 2024

Zhang, Y. 2024: The corrected parsimony phylogenetic method following Hennig's principle for analyzing morphological characters. — *Ann. Zool. Fennici* 61: 455–464.

In phylogenetic analysis using morphological characters, probability-based methods are increasingly employed compared to the parsimony methods. Due to the high diversity of morphological data, however, realistic models for among-states transitional rates have not been found, unlike in the case of molecular sequences. Parsimony, as a classical philosophical thinking framework, remains a powerful tool for inferring phylogenetic using diversified data. In parsimony-based phylogenetic analysis, how to deal with character conflicts is essential. The long-standing numerical method, searching for the tree(s) with the overall fewest evolutionary steps, however, brings in the issue of weighting characters in an unjustifiable way. By standardizing step number changing ranges among different characters, I proposed the corrected parsimony approach and restate it here. How to properly weight characters *a priori* is a separate issue that requires carefully scrutinizing the biological evidence case by case, as previously emphasized by numbers of scholars, explicitly or implicitly, including Hennig himself.

#### Introduction

Phylogenetic reconstruction is essential to the studies of evolution. The only illustration in *On the origin of species by means of natural selection* (Darwin 1859: 177) was one of the earliest undisputed phylogenies. Phylogeny was made for inferring genealogical relationships (Darwin 1859), and phylogenetic analysis has long been the key component in studies of evolution. In recent decades, phylogenetic analysis has also been used in various domains far beyond biology, such as humanity, social sciences, and fine arts (e.g., Toussaint 2003, Cochrane 2015, Le Bomin *et al.* 2016, Minocher *et al.* 2019). There exist various kinds of methods for inferring phylogenies

(Felsenstein 2004). In the field of paleontology, the most used methods are parsimony methods and probability-based methods, including maximum likelihood methods and Bayesian methods. Parsimony means to minimize *ad hoc* hypotheses, also known as Ockham's razor, named after the medieval philosopher William of Ockham. Parsimony as a reasoning principle, however, has influenced generations of natural scientists and philosophers during a long historical span since the age of ancient Greece if not earlier, including figures like Aristotle, Descartes, Newton, and Kant, among others (Sober 2015).

For phylogenetic construction, parsimony methods minimize *ad hoc* evolutionary steps. To distinguish it from the corrected parsimony

Edited by Ian Corfe, Geological Survey of Finland, and Oscar E. Wilson, University of Helsinki

approach as presented in this paper and in Zhang (2016), the parsimony phylogenetic method is, more straightforwardly, referred to as the minimum step method here. During the mid-20th century, Willi Hennig (1913-1976) explicitly proposed the procedure of reconstructing phylogenetic trees using the auxiliary principle, which is to minimize ad hoc evolutionary steps when only one character is given (Hennig 1950, 1966). Hennig's significant contribution to phylogenetics was broadly acknowledged, and he was even regarded as the father of cladistics (Wheeler et al. 2013). In contrast, until his final years, when the minimum step method was already widely adopted, Hennig never endorsed the method (Felsenstein 2004). This is because, the minimum step method is fundamentally different from Hennig's own auxiliary principle (Felsenstein 1983), as elaborated further in this paper.

Beyond parsimony, probability-based methods are also widely used for phylogenetic reconstruction. Probability-based methods, including maximum likelihood methods and the Bayesian methods, which calculate the posterior probability with predefined prior probabilities, can be applied to infer phylogenetic trees based on models specifying transitional rates among different character states (e.g., among A, T, G, and C), and return the tree with the highest probability (Felsenstein 1981, 2004). For molecular data, such as nucleotide sequences or amino acids, due to the homogeneity of the properties of characters and moderate to fast evolutionary rates, probability-based methods perform reasonably well and are broadly regarded as the preferred approaches over the parsimony methods (Felsenstein 2004, Yang & Rannala 2012). In contrast, for morphological data, the scale of character diversity is much larger, and thus it is extremely difficult to find realistic parameters to model evolutionary rates. This also explains why until about two decades ago, when probability-based phylogenetic methods had long been popular in tree constructions for molecular data, parsimony phylogenetic approach was still the default method in the field of paleontology, with only sporadic attempts using the Bayesian phylogenetic method (Müller & Reisz 2006). For the recent one and half decades, however, even for morphological data, probability-based methods

have become the preferred phylogenetic tools by more and more researchers compared to parsimony methods (Wright & Hillis 2014, Puttick et al. 2017, 2019). In this paper, I further discuss this issue and show that the parsimony principle, if used correctly, should remain a powerful tool for inferring phylogenies using morphological data and beyond. First, I argue why parsimony methods are still needed in phylogenetic analysis of morphological data. This is followed by an overview of the properties of morphological characters and character state delimitation issues for metric data, along with a brief discussion on the distinction between character resolution and accuracy. Then, I revisit Hennig's position on character conflicts and present a corrected parsimony approach following my earlier advocation of the method (Zhang 2016), as well as issues with character weighting. Finally, the corrected parsimony approach and the minimum step method are contrasted by drawing an analogy to determining exchange rates among unknown currencies.

## Probability-based methods on morphological data

For probability-based phylogenetic methods applied to morphological data, the Mk model is still one of the most commonly used (Lewis 2001). It derives from the Jukes-Canter model (JC69), the simplest model for DNA evolution, in which each pairwise nucleotide substitution rate is assumed to be equal (Jukes & Cantor 1969). The Mk model was proposed as an extension of the JC69 model, and the latter can be regarded as a specified Mk model when k = 4 (Lewis 2001). In the Mk model, all transitional rates among different character states are equal (Lewis 2001). Considering that morphological data often omit constant characters, Lewis (2001: 918) termed Mkv the model that corrects the acquisition bias. For morphological characters, evolutionary heterogeneity becomes a more serious issue compared to the situation in DNA sequences. Although previous studies (Clarke & Middleton 2008, Harrison & Larsson 2014) suggested that when efforts are made to adjust for evolutionary heterogeneity, such as using more sophisticated

data partitioning schemes or adding the gamma parameter, the performance of statistical phylogenetic methods often gets improved, some other research claimed that when evolutionary heterogeneity is high, likelihood-based methods do not perform well, even when the gamma parameter is added to the model (Kolaczkowski & Thornton 2004). For DNA sequence data, the JC69 model is considered to be significantly oversimplified, and more sophisticated models are found to be more congruent with empirical cases (Tavaré 1986, Felsenstein 2004). Correspondingly, in empirical studies using morphological data, which are more diversified than the molecular sequence data, assumptions in the Mk model, or other models, are always violated. These violations include, but are not limited to, the lack of justification for interdependence among characters, restriction on evolutionary rates across different characters, the lack of correction for acquisition bias against parsimonious non-informative characters, and difficulty in application to continuous traits (Ronquist et al. 2016, Parins-Fukuchi 2017, 2018). Despite the previous debate between probability-based and parsimony-based methods on inferring phylogenies using phenotypic data (Wright & Hillis 2014, Puttick et al. 2017, Goloboff et al. 2018a, 2018b, 2018c, O'Reilly et al. 2018a, 2018b, Puttick et al. 2019), because of the serious deficiencies of the Mk model, there is little ground to claim that probability-based methods should be used as the default in phylogenetic analysis on phenotypic datasets (contra Puttick et al. 2017). Parsimony, as a traditional philosophical thinking framework that has been applied in various fields for more than two thousand years, remains a powerful tool for inferring phylogenies, particularly for morphological data. The key is that the parsimony phylogenetic methods need to be applied correctly, as further elaborated in the following sections, starting with reviews of the properties of morphological characters.

#### Overview of properties of morphological characters

Morphological characters are often divided into qualitative and quantitative, or discrete and continuous (Stevens 1991, Thiele 1993, Wiens 2001). These two kinds of distinctions, however, represent different levels of data classification. When classified from the perspective of property, firstly, data can be divided into qualitative and quantitative. In contrast to quantitative data, qualitative data "deals with meanings. Meanings are mediated mainly through language and action" (Dey 2003: 11). As determined by its own property, qualitative data are used far more frequently in social sciences than in natural sciences. Regarding phenotypic data used in phylogenetic analysis for organisms, some of them, such as cultural and behavioral data, may belong to the qualitative category. For morphological characters, however, all are fundamentally quantitative rather than qualitative (Thiele 1993, Wiens 2001). Quantitative data are often further divided into continuous and discrete (Thiele 1993). Qualitative data have no such distinction. For morphological characters, most, if not all, of them should fall into the category of continuous variables (Thiele 1993, Wiens 2001).

Despite sharing the same property, in practice, morphological characters are expressed differently depending on the specific situations, and the different expressions are often interchangeable (Stevens 1991, Thiele 1993, Wiens 2001). It is conventional to treat some characters, such as shape, color, position, as "discrete" compared to some others, such as size. For example, even though color is a continuous numeric variable when examined in detail (i.e., as expressed in appearance parameters), it is often expressed as discrete variables, using qualitative terms. For traditional metric characters, which usually take up a substantial portion of the characters being used in phenotypic phylogenetic studies, they may be coded differently under various situations, as further elaborated below.

## Character state delimitation issues for metric data

For morphological characters, especially metric characters, when character states are expressed discretely, or even qualitatively, it induces subjective interpretation on the raw data, and this practice potentially loses and/or alters the original information (Wiens 2001). How to delimit character states attracted extensive interest from past research (Mickevich & Johnson 1976, Colless 1980, Almeida & Bisby 1984, Thorpe 1984, Archie 1985, Baum 1988, Goldman 1988, Thiele & Ladiges 1988, Chappill 1989, Thiele 1993, Strait et al. 1996, Wiens 2001). A number of character discretization methods, such as those various versions of gap-coding methods (Mickevich & Johnson 1976, Archie 1985, Chappill 1989), bear the problem that terminals whose means are (not) significantly different are coded with the same (different) character states (Farris 1990, Goloboff et al. 2006). Differing from these ones, the method proposed by Wiens (2001) preserves the relative quantitative distances among different means of the terminals.

Constrained by the limitation on the number of distinct states, Wiens' method, however, is not applicable if the number of taxa which bear a unique character state exceeds 32 (or 64) for a 32 (or 64) bits system (Wiens 2001) when using PAUP\* (Swofford 2001), the default software for the minimum step method at that time. The TNT program (Goloboff et al. 2008), differing from PAUP\*, can incorporate a much larger range of character states (values between 0 and 65, up to three decimals) (Goloboff et al. 2006). As a result, metric characters can be coded as such when using the program TNT, and the entry can be either a single value, usually the mean, or a range (Goloboff et al. 2006). Goloboff and colleagues explicitly suggested using range entry such as the 95% confidence interval, because "that the ranges of two terminal taxa will overlap (thus producing a step count of 0) whenever their means are not statistically different" (Goloboff et al. 2006: 591). However, doing so will bring in additional problems, because testing for overlapping of confidence intervals is not a logical basis for assessing statistical significance of the difference between two means. Two means may very well be significantly different from one another at  $\alpha = 0.05$  with overlapping 95% confidence intervals (Austin & Hux 2002).

As determined by the complicated nature of morphological characters, character state delimitation remains an issue awaiting to be solved. On the one hand, it is advantageous to delimit metric characters as finely as possible and to keep the modification as little as possible. From this perspective, using characters "coded as such" was notable progress compared to prior delimitation methods. On the other hand, how to cope with the boundaries and overlapping issues remains an open question. The most natural way is to keep the original statistical distributions. Consequently, most of the characters are expected to be polymorphic characters with non-flat weights assigned to each presenting state, but that would significantly increase the computing burden and thus may not be ideal. Alternatively, one may opt to apply Wiens' method (2001) using the TNT program, or simply code the character states "as such" using only the means, but these actions would erase valuable information. More sophisticated solutions require further studies.

Finely delimiting character states inevitably increases the step number changing range of a character. The among-character scaling issue, thus, is the twin issue of character state delimitation. This is another issue that has not yet been solved, but the present paper (also in Zhang 2016) provides a solution, as further discussed below.

# Scaling evolutionary step changing ranges among characters: accuracy vs. resolution

Some proponents of the minimum step method argued against the practice of scaling for the reason that more "informative" characters may deserve higher weights (Farris 1990, Goloboff 1993). To solve this puzzle, it is necessary to discuss the resolution and accuracy of characters' phylogenetic information, and the logical relationship between characters' resolution and accuracy. Given any data matrix, for two random characters, a and b, if character a has a higher step number changing range, it means that among different topologies, finer difference of the phylogenetic expression of character a, compared to that of character b, can be distinguished. Therefore, the resolution of the phylogenetic information carried by character a is higher. However, this is not equivalent to the claim that character a is more informative. Character a is more informative if, and only if, its phylogenetic

information is more accurate. On the other hand, it is unclear whether the phylogenetic information of character *a* is more accurate compared to that of character *b*, because the accuracy of character's phylogenetic information is based on the test of congruence (Hennig 1950, 1966), that is, its acceptance by other characters. Therefore, it is illogical to up-weight a character *a priori* just based on the condition that the resolution of its phylogenetic information is higher, as performed in the minimum step method. Below, Hennig's position on this issue is reviewed and the corrected parsimony principle is presented.

#### Hennig's auxiliary principle, character conflict, and the corrected parsimony approach

Synapomorphy is "always reason for suspecting kinship, and that their origin by convergence should not be assumed a priori" (Hennig 1966: 121). This criterion is called the auxiliary principle by Hennig (1950, 1966). The auxiliary principle is congruent with the parsimony principle when only a single character is examined, because when synapomorphies are used to group taxa, the tree(s) is/are the most parsimonious tree(s) with fewest evolutionary steps. When multiple characters are present, however, Hennig (1966: 121) stated the following: "(t)hus the question of whether kinship relations based on a single character or a single presumed transformation series of characters correspond to the actual phylogenetic relationships of the species is tested by means of other series of characters: by trying to bring the relationships indicated by the several series of characters into congruence. In the final analysis this is again the method of 'checking, correcting, and rechecking'."

It is clear that Hennig's method is distinct from the minimum step method (Felsenstein, 1983). Hennig never proposed any generic mathematic approach to bring characters into congruence, which is not likely to be found. To mathematically express the problem, let us use E to represent the expression of phylogenetic signal by any given character, and define that Eranges from 0 to 1: when E = 0, none of the phylogenetic signal of that particular character gets expressed in the tree that takes into account all characters; when E = 1, all of the phylogenetic signal of the character gets expressed in the tree. Therefore, in a data matrix that contains *n* characters, to find the most congruent signal among all characters becomes the question of maximizing  $\sum_{i=0}^{n} W_i E_i$  (i = 1, 2, 3, ..., n), where  $E_i$  denotes the expression of the phylogenetic information carried by character *i*, and  $W_i$  denotes the weight of character *i*.

The value of E should reflect how well a character's phylogenetic signal is expressed, and it needs to be standardized across all characters. Accordingly, the retention index r (Farris 1989) is adopted here to quantify E. The value of rranges from 0 to 1, and it denotes "the fraction of apparent synapomorphy in the character that is retained as synapomorphy on the tree" (Farris 1989: 418). Certainly, to maximize the total congruent phylogenetic expression is consistent with minimizing *ad hoc* hypotheses of homoplasy, which can be expressed as minimizing  $\sum_{i=1}^{n} W_i (1-E_i)$  (i = 1, 2, 3, ..., n). Without loss of generality, for the purpose of simplicity, here temporarily assuming  $W_i = 1$  (i = 1, 2, ...,*n*), then, it is to minimize  $\sum_{i=1}^{n} (s_i - m_i)/(g_i - m_i)$ , where for character  $i, g_i$  represents the maximum possible number of steps,  $m_i$  represents the minimum possible number of steps, and s, represents the actual number of steps in a given tree.

In contrast, in the minimum step method, it is to minimize  $\sum_{i=1}^{n} (s_i - m_i)$ , equating to minimizing  $\sum_{i=1}^{n} (g_i - m_i)(s_i - m_i)/(g_i - m_i)$ , which is  $\sum_{i=1}^{n} (g_i - m_i)(1 - E_i)$ . Hence, in the minimum step method, the step number changing range,  $g_i - m_i$ , becomes a coefficient of character weight. Accordingly, when steps are simply added across all characters, characters with higher step number changing ranges get overweighted, while those with lower step number changing ranges get underweighted.

To distinguish it from the minimum step method, the method presented here is referred to as the corrected parsimony approach (Zhang 2016). Comparisons between the corrected parsimony approach and the minimum step method on both hypothetical and empirical cases, as well as TNT scripts, were presented in Zhang (2016). As for how to quantify characters' weights, it is another critical issue, further discussed below.

## Scaling, weighting *a posteriori*, and weighting *a priori*

Scaling, or standardization, is a separate issue from weighting. Scaling sets up the basis for character weighting. Without properly scaling evolutionary step number changing ranges among different characters, the minimum step method represents a unjustified character weighting scheme a priori, which is a problem that cannot be circumvented by approaches invented for weighting a posteriori, such as successive weighting (Farris 1969), implied weighting (Goloboff 1993), etc. These approaches weight characters based on extra step numbers. However, without justified weighting a priori, it is meaningless to discuss characters' extra step numbers (Neff 1986, Bryant 1989). In other words, because the minimum step method weights characters a priori in an unjustifiable way, any additional efforts for a posteiori weighting cannot fix the problem.

Apparently, the corrected parsimony method does not prevent characters from being weighted a priori, as long as the weighting is justified. It is not expected that all characters deserve the same weights, and thus characters are supposed to be carefully weighted, with explicitly presented reasons (Poe & Wiens 2000). Although in parsimony phylogenetic analysis characters are assumed to be independent from each other, in real cases this assumption is almost never satisfied. Instead, many, if not most, characters are correlated. Character redundancy is an extreme example of character correlation. In such cases, the redundant characters should be given relatively lower weights, as practiced in some empirical studies (e.g., Strait & Grine 2004). In most cases, however, it is not easy to quantify character correlation. Background knowledge is an important source for detecting character correlation. Especially, studies in developmental biology have provided substantial evidence for inferring character correlation (Kangas et al. 2004, Kavanagh et al. 2007, Harjunmaa et al. 2014). For example, mammalian dental characters are common sources for phylogenetic analysis, but in most cases their weights are not justified for developmental correlation. Experimental studies on mice suggested that many mammalian dental traits that

were regarded as stepwise transitional stages were developmentally linked rather than independent (Kangas *et al.* 2004, Harjunmaa *et al.* 2014). These studies provided a biological basis to justify weighting schemes not only among different characters, but also among different between-state transitions of a specified character.

Character weighting a priori is a complicated topic that is beyond the scope of this paper, but it is the place where Hennig's suggestion of "checking, correcting, and rechecking" comes into play. That means, it is unlikely to find a universal weighting a priori approach with clearly defined numerical formula. The corrected parsimony approach presented here should be viewed as a backbone scheme in parsimony phylogenetics. Character weights should be estimated, if possible, by researchers based on explicit biological reasonings (Poe & Wiens 2000), as originally suggested by Hennig (1950, 1966). Weighting a posteriori, based on how characters covary with each other, is certainly useful, but weighting *a priori*, with valid biological reasons, is more crucial. After careful weighting a priori is executed, weighting a posteriori can step in, and further refine the analysis.

#### Analogy with exchange rates among different currencies

The difference between the minimum step method and the corrected parsimony method can be more intuitively understood via an analogy below. The key in parsimony phylogenetic analysis is to choose the best option among various tradeoffs, which are inevitably created by conflicting or partially conflicting characters. Exchanges of phylogenetic information among different characters can be intuitively viewed as an analogy to the exchanges among different currencies. As shown in Fig. 1A, the mid-market exchange rates among different currencies are generally uncontroversial, as determined by the real-time market.

Characters in phylogenetic analyses are like different currencies. The exchange rates among them, however, are largely unknown, as it is usually less clear which characters provide more reliable phylogenetic information and which provide less, except for extreme cases, such as character



**Fig. 1.** The analogy with exchange rates among different currencies. — **A**: The mid-market exchange rates among Japanese Yen, Euro and US Dollar as of 16 September 2024. — **B**: Assigning random numbers to the exchange rates among unknown currencies, analogous to the minimum step method. — **C**: First assigning flat exchange rates to the unknown currencies, and then trying to figure out more accurate rates using information from the market, if available, analogous to the corrected parsimony method. The bill images are from Wikimedia Commons: €5 bill by Robert Kalina, redesign by Reinhold Gerstetter, ECB decisions ECB/2003/4 and ECB/2003/5; ¥1000 bill from Nippon Ginko, Government of Japan, public domain; \$5 bill from United States Government, public domain.

redundancy, poor documentation, etc. Hence, the task is similar to figuring out exchange rates among unknown currencies. What the minimum step method does is to assign an arbitrary set of exchange rates, as determined by the step count changing range of each character, to the currencies (Fig. 1B). This does not automatically result in an absurd phylogeny, because a set of arbitrarily chosen numbers may happen to be very close to the true numbers, though it is more likely not the case. An implicit "advantage" of the minimum step method, that has been utilized implicitly or explicitly in many past empirical studies, is that a slight modification of the character state delimitation scheme may significantly change the set of the arbitrary numbers, until reaching the "expected" phylogeny, but such an "advantage" is

hardly justifiable. In contrast, under the corrected parsimony approach, the starting point is that the exchange rates among different currencies are flat because there is no basis to justify a specific set of exchange rates. The next step is to study the market, in order to reach a more credible set of exchange rates (Fig. 1C). In biology, to study the "market" of the characters requires a thorough investigation of how valuable each character is, compared to each other, in its capability of revealing the phylogenetic relationships of the terminals.

#### Conclusion

For morphological data, using parsimony meth-

ods to infer phylogenetic relationships is still essential. This paper revisits the properties of morphological characters, character state delimitation approaches, and the accompanying among-character scaling issue. Given the necessity of standardizing evolutionary step number changing ranges among different characters, the corrected parsimony method is restated here following my earlier work (Zhang 2016). I argue that the corrected parsimony approach should replace the minimum step algorithm and be viewed as the backbone framework in parsimony phylogenetic methods. Distinctions between the minimum step method and the corrected parsimony approach are intuitively presented through an analogy to exchange rates among unknown currencies. The minimum step method assigns arbitrary weights to characters, just like using unjustified arbitrary exchange rates among unknown currencies. In contrast, the corrected parsimony approach assigns flat weights to characters by default, and then differentiates the weights based on valid biological reasons, if applicable. It is like assigning flat exchange rates to unknown currencies, and then tuning the weights by studying the market. The principle of the corrected parsimony approach, unlike the minimum step method, agrees with the original messages delivered by Hennig, who had no intention to solve character conflicting issues by arbitrarily weighting characters. This corrected parsimony principle can also be applied in fields beyond biology.

#### Acknowledgements

First and foremost, I express my gratitude to the late Professor Richard C. Fox (1933-2022). I have had the fortune of receiving his mentoring on numerous occasions, and those are the most memorable moments of my whole academic career. During his six-decade-long academic career, Professor Fox always stood with solid evidence and reasonings and never stepped back, regardless of those countless ungrounded critiques and negative comments, often from so-called authorities. I have seen the true spirit of a real scientist from Professor Fox. The inspiration and academic mentoring I have received from Professor Fox is greater than that I have received from anyone else, even collectively. In addition, for the corrected parsimony approach, I thank many people including but not limited to (in alphabetic order) Bill Ausich, Chris Brochu, Joe Felsenstein, John Freudenstein, Pablo Goloboff, Debbie Guatelli-Sternberg, Tony Harper, John Hunter, Jukka Jernvall, Charles Marshall, Jin Meng,

Xiaoming Wang, John Wenzel, and Greg Wilson, for discussions, comments, and critiques. In particular, I want to express my special thanks to John Wiens, whose past studies on this issue and beyond built the foundation for the work presented here. Finally, I thank Jukka Jernvall for inviting me to contribute to the special issue in honor of the "paleontologist's paleontologist", Professor Björn Kurtén, and I thank the editor Oscar Wilson and two anonymous reviewers (especially reviewer 2) for insightful and constructive comments, which significantly improved the paper. It is like winning a lottery to have you fellows on board.

#### References

- Almeida, M. & Bisby, F. 1984: A simple method for establishing taxonomic characters from measurement data. — *Taxon* 33: 405–409.
- Archie, J. W. 1985: Methods for coding variable morphological features for numerical taxonomic analysis. — Systematic Zoology 34: 326–345.
- Austin, P. C. & Hux, J. E. 2002: A brief note on overlapping confidence intervals. — *Journal of Vascular Surgery* 36: 194–195.
- Baum, B. R. 1988: A simple procedure for establishing discrete characters from measurement data, applicable to cladistics. — *Taxon* 37: 63–70.
- Bryant, H. N. 1989: An evaluation of cladistic and character analyses as hypothetico-deductive procedures, and the consequences for character weighting. — *Systematic Zoology* 38: 214–227.
- Chappill, J. A. 1989: Quantitative characters in phylogenetic analysis. — *Cladistics* 5: 217–234.
- Clarke, J. A. & Middleton, K. M. 2008: Mosaicism, modules, and the evolution of birds: results from a Bayesian approach to the study of morphological evolution using discrete character data. — *Systematic Biology* 57: 185–201.
- Cochrane, E. E. 2015: Phylogenetic analysis of Polynesian ritual architecture suggests extensive cultural sharing and innovation. — *The Journal of the Polynesian Soci*ety 124: 7–46, https://doi.org/10.15286/jps.124.1.7-46.
- Colless, D. 1980: Congruence between morphometric and allozyme data for Menidia species: a reappraisal. — Systematic Zoology 29: 288–299.
- Darwin, C. 1859: On the origin of species by means of natural selection. — John Murray, London.
- Dey, I. 2003: Qualitative data analysis: a user friendly guide for social scientists. — Routledge, London and New York.
- Farris, J. S. 1969: A successive approximations approach to character weighting. — Systematic Biology 18: 374–385.
- Farris, J. S. 1989: The retention index and the rescaled consistency index. — *Cladistics* 5: 417–419.
- Farris, J. S. 1990: Phenetics in camouflage. *Cladistics* 6: 91–100.
- Felsenstein, J. 1981: Evolutionary trees from gene frequencies and quantitative characters: finding maximum likelihood estimates. — *Evolution* 35: 1229–1242.

- Felsenstein, J. 1983: Parsimony in systematics: biological and statistical issues. — Annual Review of Ecology and Systematics 14: 313–333.
- Felsenstein, J. 2004: Inferring phylogenies. Sinauer Associates, Sunderland.
- Goldman, N. 1988: Methods for discrete coding of morphological characters for numerical analysis. — *Cladistics* 4: 59–71.
- Goloboff, P. A. 1993: Estimating character weights during tree search. — *Cladistics* 9: 83–91.
- Goloboff, P. A., Mattoni, C. I. & Quinteros, A. S. 2006: Continuous characters analyzed as such. — *Cladistics* 22: 589–601.
- Goloboff, P. A., Farris, J. S. & Nixon, K. C. 2008: TNT, a free program for phylogenetic analysis. — *Cladistics* 24: 774–786.
- Goloboff, P. A., Torres, A. & Arias, J. S. 2018a: Weighted parsimony outperforms other methods of phylogenetic inference under models appropriate for morphology. — *Cladistics* 34: 407–437.
- Goloboff, P. A., Torres Galvis, A. & Arias, J. S. 2018b: Parsimony and model-based phylogenetic methods for morphological data: comments on O'Reilly et al. — *Palaeontology* 61: 625–630.
- Goloboff, P. A., Pittman, M., Pol, D. & Xu, X. 2018c: Morphological data sets fit a common mechanism much more poorly than DNA sequences and call into question the Mkv model. — *Systematic Biology* 68: 494–504.
- Harjunmaa, E., Seidel, K., Häkkinen, T., Renvoisé, E., Corfe, I. J., Kallonen, A., Zhang, Z.-Q., Evans, A. R., Mikkola, M. L. & Salazar-Ciudad, I. 2014: Replaying evolutionary transitions from the dental fossil record. — *Nature* 512: 44–48.
- Harrison, L. B. & Larsson, H. C. 2014: Among-character rate variation distributions in phylogenetic analysis of discrete morphological characters. — *Systematic Biology* 64: 307–324.
- Hennig, W. 1950: Grundzuge einer Theorie der phylogenetischen Systematik. — Deutscher Zentralverlag, Berlin.
- Hennig, W. 1966: *Phylogenetic systematics*. University of Illinois Press, Urbana and Chicago.
- Jukes, T. H. & Cantor, C. R. 1969: Evolution of protein molecules. — In: Munro, H. N. (ed.), *Mammalian protein metabolism*: 21–132, Academic Press, New York.
- Kangas, A. T., Evans, A. R., Thesleff, I. & Jernvall, J. 2004: Nonindependence of mammalian dental characters. — *Nature* 432: 211–214.
- Kavanagh, K. D., Evans, A. R. & Jernvall, J. 2007: Predicting evolutionary patterns of mammalian teeth from development. — *Nature* 449: 427–432.
- Kolaczkowski, B. & Thornton, J. W. 2004: Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. — *Nature* 431: 980–984.
- Le Bomin, S., Lecointre, G. & Heyer, E. 2016: The evolution of musical diversity: the key role of vertical transmission. — *PLoS ONE* 11(3), e0151570, https://doi. org/10.1371/journal.pone.0151570.
- Lewis, P. O. 2001: A likelihood approach to estimating phylogeny from discrete morphological character data. — *Systematic Biology* 50: 913–925.

- Mickevich, M. & Johnson, M. S. 1976: Congruence between morphological and allozyme data in evolutionary inference and character evolution. — *Systematic Zoology* 25: 260–270.
- Minocher, R., Duda, P. & Jaeggi, A. V. 2019: Explaining marriage patterns in a globally representative sample through socio-ecology and population history: a Bayesian phylogenetic analysis using a new supertree. — *Evolution and Human Behavior* 40: 176–187.
- Müller, J. & Reisz, R. R. 2006: The phylogeny of early eureptiles: comparing parsimony and Bayesian approaches in the investigation of a basal fossil clade. — *Systematic Biology* 55: 503–511.
- Neff, N. A. 1986: A rational basis for a priori character weighting. — Systematic Biology 35: 110–123.
- O'Reilly, J. E., Puttick, M. N., Pisani, D. & Donoghue, P. C. 2018a: Empirical realism of simulated data is more important than the model used to generate it: a reply to Goloboff et al. — *Palaeontology* 61: 631–635.
- O'Reilly, J. E., Puttick, M. N., Pisani, D. & Donoghue, P. C. 2018b: Probabilistic methods surpass parsimony when assessing clade support in phylogenetic analyses of discrete morphological data. — *Palaeontology* 61: 105–118.
- Parins-Fukuchi, C. 2017: Use of continuous traits can improve morphological phylogenetics. — *Systematic Biology* 67: 328–339.
- Parins-Fukuchi, C. 2018: Bayesian placement of fossils on phylogenies using quantitative morphometric data. — *Evolution* 72: 1801–1814.
- Poe, S. & Wiens, J. J. 2000: Character selection and the methodology of morphological phylogenetics. — In: Wiens, J. J. (ed.), *Phylogenetic analysis of morphological data*: 20–36. Smithsonian Institution Press, Washington D.C.
- Puttick, M. N., O'Reilly, J. E., Pisani, D. & Donoghue, P. C. 2019: Probabilistic methods outperform parsimony in the phylogenetic analysis of data simulated without a probabilistic model. — *Palaeontology* 62: 1–17, https:// doi.org/10.1111/pala.12388.
- Puttick, M. N., O'Reilly, J. E., Tanner, A. R., Fleming, J. F., Clark, J., Holloway, L., Lozano-Fernandez, J., Parry, L. A., Tarver, J. E. & Pisani, D. 2017: Uncertain-tree: discriminating among competing approaches to the phylogenetic analysis of phenotype data. — *Proceedings of the Royal Society B* 284, 20162290. https://doi. org/10.1098/rspb.2016.2290.
- Ronquist, F., Lartillot, N. & Phillips, M. J. 2016: Closing the gap between rocks and clocks using total-evidence dating. — *Philosophical Transactions of the Royal Society B* 371, 20150136, https://doi.org/10.1098/ rstb.2015.0136.
- Sober, E. 2015: Ockham's razors Cambridge University Press, Cambridge.
- Stevens, P. 1991: Character states, morphological variation, and phylogenetic analysis: a review. — *Systematic Botany* 16: 553–583.
- Strait, D. S. & Grine, F. E. 2004: Inferring hominoid and early hominid phylogeny using craniodental characters: the role of fossil taxa. — *Journal of Human Evolution*

47: 399-452.

- Strait, D. S., Moniz, M. A. & Strait, P. T. 1996: Finite mixture coding: a new approach to coding continuous characters. — *Systematic Biology* 45: 67–78.
- Swofford, D. L. 2001: PAUP\*: phylogenetic analysis using parsimony (\*and other methods) 4.0. — Sinauer Associates, Sunderland.
- Tavaré, S. 1986: Some probabilistic and statistical problems in the analysis of DNA sequences. — *Lectures on Mathematics in the Life Sciences* 17: 57–86.
- Thiele, K. 1993: The holy grail of the perfect character: the cladistic treatment of morphometric data. — *Cladistics* 9: 275–304.
- Thiele, K. & Ladiges, P. Y. 1988: A cladistic analysis of Angophora Cav. (Myrtaceae). — *Cladistics* 4: 23–42.
- Thorpe, R. 1984: Coding morphometric characters for constructing distance Wagner networks. — *Evolution* 38: 244–255.
- Toussaint, G. 2003: Classification and phylogenetic anal-

ysis of African ternary rhythm timelines. — *ISA-MA-BRIDGES Conference Proceedings*, https://archive.bridgesmathart.org/2003/bridges2003-25.pdf.

- Wheeler, Q., Assis, L. & Rieppel, O. 2013: Phylogenetics: Heed the father of cladistics. — *Nature* 496: 295–296.
- Wiens, J. J. 2001: Character analysis in morphological phylogenetics: problems and solutions. — *Systematic Biol*ogy 50: 689–699.
- Wright, A. M. & Hillis, D. M. 2014: Bayesian analysis using a simple likelihood model outperforms parsimony for estimation of phylogeny from discrete morphological data. — *PLoS ONE* 9(10), e109210, https://doi. org/10.1371/journal.pone.0109210.
- Yang, Z. & Rannala, B. 2012: Molecular phylogenetics: principles and practice. — *Nature Reviews Genetics* 13: 303–314.
- Zhang, Y. 2016: A corrected parsimony criterion for reconstructing phylogenies. — arXiv: 1610.03217, https://doi. org/10.48550/arXiv.1610.03217.