



Profile Parsimony (PP): an analysis under Implied Weights (IW)

J. Salvador Arias & Daniel Rafael Miranda-Esquivel*

Laboratorio de Sistemática y Biogeografía. Grupo de Estudios en Biodiversidad, Escuela de Biología, Universidad Industrial de Santander, A.A. 678 Bucaramanga, Colombia

Accepted 29 October 2003

Abstract

Faith and Trueman [Syst. Biol. (2001) 331] recently proposed “Profile Parsimony” (PP) as a new form of phylogenetic analysis, but it is equivalent to the Implied Weights (IW) method when certain functions are used. The PP method cannot accommodate missing/inapplicable cells and/or multistate characters. Finally we demonstrate that, as defined, PP scores are not random, and there is no need to generate random trees as proposed by Faith and Trueman.

© The Willi Hennig Society 2004.

Introduction

Faith and Trueman (2001) presented what they called “a new form of parsimony” (p. 333), and coined the name “Profile Parsimony” (PP). According to their method (2001, p. 344):

Any tree hypothesis provides a profile of the observed minimum number of steps for each of k characters (in place of the total tree length for a set of characters). This profile is converted to a profile of tail probabilities, based on frequency distributions of the number of steps for different characters. To obtain these distributions, we evaluate each character on random tree topologies, counting the minimum number of steps [...]. That distribution allows us to calculate the probability of obtaining an equal or smaller number of steps [...] for a random tree [...]. The preferred tree hypothesis then has a profile of steps over k characters producing the lowest overall tail probability.¹

¹Terminology used here: $\log(n)$, is the natural logarithm; *profile*, the values of the $\log(n)$ scores of individual character-step cumulative frequencies (Faith and Trueman, 2001 also use “profile” to refer to the usual character length, reconstruction length, or character extra-steps); the character *fit* is the value of a function given the homoplasy of the character (Goloboff, 1993). In PP the fit is measured using the profile.

*Correspondence: Daniel Rafael Miranda-Esquivel, Laboratorio de Sistemática & Biogeografía. Grupo de Estudios en Biodiversidad. Escuela de Biología. Universidad Industrial de Santander. A.A. 678 Bucaramanga, Colombia.
E-mail address: dmiranda@uis.edu.co

Here, we show that the goals and justification of PP and Implied Weights (Goloboff, 1993, 1995; hereafter, IW) are alike, and that their results are equivalent under most conditions. We also demonstrate that PP and IW may give different results when the data matrix includes question marks, inapplicable cells, or multistate characters. The differences arise from the improper treatment of these entries in PP.

Objectives of IW and PP

To demonstrate that the goals of PP and IW are identical, we quote from the respective authors:

(Faith and Trueman, 2001, p. 343; our italics):

We suspect that cladistic parsimony may lead to inferential errors on some occasions when it *discounts a tree that might better explain shared features, instead [sic!] favoring one that makes a homoplasious characters a bit less homoplasious.*

(Goloboff, 1995, p. 95; our italics):

Explicit discussions of which I am aware suggest why *trees with more raw steps may be preferable* in some cases, pointing out that it *may be better to save homoplasy in a few good characters at the expense of increasing it in several poor ones.*

and (Goloboff, 1993, p. 88):

It seems more reasonable to think that adding one step to a perfectly hierarchical character worsens it more than adding one step to a character with, say, 50 extra steps.

(Faith and Trueman, 2001, p. 343; our italics):

PP provides an alternative way to base evidence on the explanation of putative synapomorphies, assigning *a better fit to trees producing a very small number of steps for some characters than to trees that reduce a large number of steps to not-quite-so large.*

So, both approaches prefer characters with relatively less homoplasy over those with relatively more homoplasy.

Methodological issues

The goal of PP is to select that tree(s) “for which its profile [fit under PP] is least probably matched or bettered if a tree were selected at random” (Faith and Trueman, 2001, p. 344). This value is calculated over the profile for all characters, and expressed as the PP score, a negative value. Here we convert this into an absolute value to facilitate a comparison with IW. Implied Weights, as implemented in Pee-Wee (Goloboff, 1998) and PAUP* (Swofford, 2001) applies a score to each character, which is then summed over all characters, and used as a measure of Goloboff-fit (hereafter, GFit, the term and acronym were introduced by Swofford, 2001). $GFit(k)$ is the value of the function $k/(k + es)$ (Goloboff, 1993, 1995) where k is a “concavity constant” (Goloboff, 1993) and where es is the number of extra steps of the character in a given tree. The objective is to find that tree which maximizes $GFit(k)$.

Comparing raw numbers of PP scores and GFit is misleading because they are not on the same scale and PP scores have different maximum values for different characters. We calculated the fit slopes ratio for PP and IW values using a one-step difference in fit (hereafter referred as DIF) (Goloboff, 1993, p. 88). DIF is the difference in fit function values for a character with x extra steps and $x + 1$ extra steps (see Goloboff, 1993, Table 1). As DIF measures proportions, it is possible to compare PP and IW values using their DIF values, rescaled to let the first instance of homoplasy be equal to one. Our results, using the example of Faith and Trueman (2001, their Table 2) show that the behavior of IW and PP are very similar (Table 1).

PP differs from IW neither in the results obtained nor in the way the results are obtained. The PP score is a function of the number of taxa (which determines the total number of trees), and the distribution of states. For example, binary characters with more equilibrated (i.e. equal) distributions between apomorphic and plesiomorphic states are, given equal homoplasy, more “improbable”.

Table 1

Comparison of DIF values (Difference in Fit: Fit of character with x extra steps – Fit of [the same] character with $x + 1$ extra steps), between PP scores and GFit values over all possible combinations for Faith and Trueman’s example (2001, their table 2), for eight taxa. Results are rescaled to make the DIF values for a perfectly hierarchical character when compared to the same character with one extra step, equal to 1. Apo, refers to the number of apomorphic taxa, $GFit(k)$ is the fit reported by Pee-Wee (Goloboff, 1998) for a particular k -value. NA: impossible to calculate

Extra steps	DIF GFit(1)	DIF GFit(2)	DIF Apo = 2	DIF Apo = 3	DIF Apo = 4
1	1.000	1.000	1.000	1.000	1.000
2	0.333	0.500	NA	0.458	0.490
3	0.167	0.300	NA	NA	0.108

For a given character length, what Faith and Trueman (2001, p. 344) called the “profile of observed minimum number of steps for each k characters”, the number of trees is fixed, and it need not be estimated by randomization because it depends only on the number of apomorphic and plesiomorphic character states, and the number of empty cells (Goloboff, 1991). It is possible to find the exact number of trees for a given number of steps for a binary character, using Carter et al.’s (1990) theorem one (notation modified below):

$$P(a, p, s) = [(2t - 3s) \times (2t - 5)!! \times (2a - s - 1)! \times (2p - s - 1)!] / [(2t - 2s - 1)!! \times (a - s)! \times (p - s)! \times (s - 1)! \times 2^{t-2s}] \quad (1)^2$$

where $P(a,p,s)$ is the number of trees; a is the number of taxa with the apomorphic state; p is the number of taxa that have the plesiomorphic state; s is the number of steps; and t is the number of taxa that have an apomorphic or plesiomorphic state (the sum of a plus p , or the number of cells that are not empty). For a given character, independent of the number of apomorphic and plesiomorphic states, the ratio of the number of trees without homoplasy and trees with one homoplastic instance (i.e. the DIF for the first extra step) is equal to $1/(2t - 5)$. The first extra step produces a constant fit difference across all characters, a desirable property for a phylogenetic method, but one that has nothing to do with *probability* or *improbability*.

Although we were unable to find an analytical solution for larger amounts of homoplasy (a mathematical problem not under discussion here), we modified Eqn 1 to convert it into log form (Eqn 2, below and Appendix 1) to implement a simple profile probability program, written in C (“Pmaker”), that can calculate the probability profile for a binary character matrix. Note, the DIF values (Fig. 1), for more than one extra

²Double factorial is defined to natural numbers as $n!! = n \times (n - 2)!!$ with $0!! = 1!! = 1$.

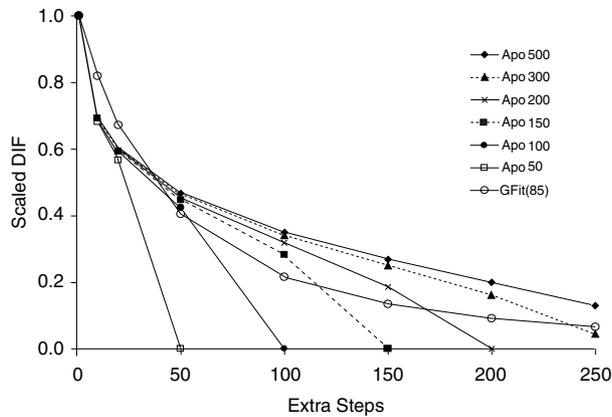


Fig. 1. Plot of scaled DIF values (see text and Table 1) for different proportions of apomorphic/pleisiomorphic taxa for a data set of 1000 terminals. *Apo*, refers to the number of apomorphic taxa. The GFit values for a concavity of 85 are included for comparison with *Apo* = 500. All the DIF values for PP are concave in the first part of the graph.

step are nearly constant across all characters. These DIF values are a function of the number of homoplastic instances rather than the *improbability* of the character, and they are smaller than the previous DIF values. DIF values are greatest for the first extra steps of a character, and they decrease with increased homoplasy. This is the behavior of a concave function of homoplasy (Farris, 1969, p. 380; Goloboff, 1993, pp. 87–88).

“Comparing the fit of different trees using a concave, decreasing function of homoplasy” (Goloboff, 1995, p. 91) is the most distinctive feature of IW. As the function used with PP is a concave one, we assert that PP is not a new parsimony method but a new concave function. We do not suggest here an exact mathematical equivalence between PP and Goloboff’s (1993, 1995) function, rather, we assert that the results will be identical or very similar because the same basic calculation (a concave function) is employed. The only methodological difference between the Goloboff’s, and Faith and Trueman’s versions of (implied) weighting is that Goloboff used an explicit and simple concave function, while Faith and Trueman suggested an implicit and complex one.

Empirical examples

To illustrate the similarities between PP and IW we prepared a set of programs in C to calculate the PP scores using both Faith and Trueman’s method (with the random trees, program “TreeP”) and Carter et al.’s (1990) method (program “PMaker” based on Eqn 2 below). We searched all optimal trees and several near-optimal trees using parsimony under equal weights as a criterion of optimality. We also produced trees using IW

Table 2

Faith and Trueman’s (2001) example. The trees are from their Fig. 1. The Faith and Trueman PP scores are based on the entire tree universe; our PP scores are based only supported tree topologies (see text). GFit values are calculated with Pee-Wee (Goloboff, 1998). PP values are the PP absolute values

Tree	Length	PP score			Our PP score
		FT2001	GFit(1)	GFit(2)	
a	33	23.50	108.3 (39%)	130.8 (44%)	21.850
b	35	29.10	123.2 (51%)	137.8 (52%)	25.754
c	36	31.00	126.6 (54%)	139.2 (54%)	27.804
d	36	31.00	126.6 (54%)	139.2 (54%)	27.559

and employed all concavity values available in the present version of Pee-Wee (from 1 to 6). We then submitted these to our programs to get the “best” PP tree(s).

To generate the profile for the hypothetical dataset of Faith and Trueman (2001, Table 2), we used two sets of trees, one was composed of the entire set of 10 395 dichotomous trees, as Faith and Trueman (2001, p. 345, Fig. 1) did, while the other included only those 3313 trees with supported nodes. The trees were then submitted to TreeP. Our results are presented in Table 2. Implied Weights, as implemented in Pee-Wee (Goloboff, 1998); using a concavity value of 1 or 2, the same trees (c and d) were chosen as PP. Using the complete set of trees, we plotted the scores of each tree, comparing fit and length (Fig. 2).

We conducted additional tests using real datasets, some of which were slightly modified to eliminate some multistate characters or taxa with many empty cells to facilitate the calculation of PP values. Using the “standard” PP technique (“TreeP” program) we analyzed a blackfly (Diptera: Simuliidae) dataset (Coscarón and Miranda-Esquivel, 1998; 49 characters, 13 taxa). We deleted one taxon (*Petrowygomysia chacabamba*) due to a high number of missing entries, although the matrix still has some missing entries (8.8% of matrix entries). 10 000 random trees were generated by Component (Page, 1993).

Under equal weights, NONA (Goloboff, 1998) found five trees (86 steps). Pee-Wee (Goloboff, 1998), with a concavity value of 1, found one tree, one step longer than the most parsimonious trees (87 steps, GFit(1) = 296.5 (70%)). With concavity values of from two to six, the resulting tree is identical to one of the most parsimonious trees found (86 steps, GFit(2) = 319.3 (72%)). PP also prefers one of the most parsimonious trees (86 steps, PP score = 152.814). The third best PP tree is the same as that selected by IW with concavity values equal to or greater than 2 (86 steps, PP score = 152.529; see Fig. 3 for graphical comparison). The results, for the optimal trees of the two methods are different. We attribute the differences between the results due to the behavior of the PP function when missing entries and multistate characters are present (see below under PP problems).

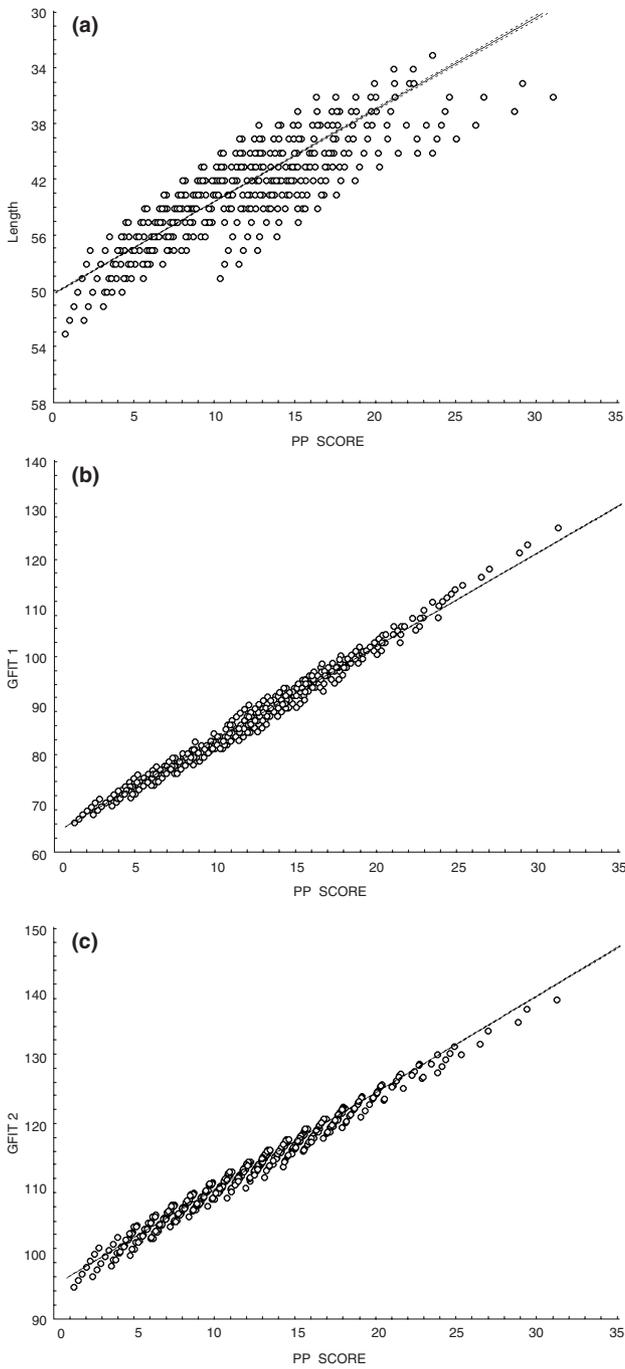


Fig. 2. Plots of PP scores against lengths, GFit(1) and GFit(2), for eight taxa over the entire tree universe, using the matrix from Faith and Trueman's (2001) Table 2. (a) PP score vs. Length ($r^2 = 0.7506$). (b) PP scores vs. GFit(1) values ($r^2 = 0.9860$). (c) PP scores vs. GFit(2) values ($r^2 = 0.9795$). The r^2 values are given only for reference and are not intended to be statistical statements.

The second data set is a Pentatomomorpha (Heteroptera) matrix (Henry, 1997; 58 characters, 34 taxa) modified to contain only binary characters. Eleven multistate characters were re-coded by merging several states in two

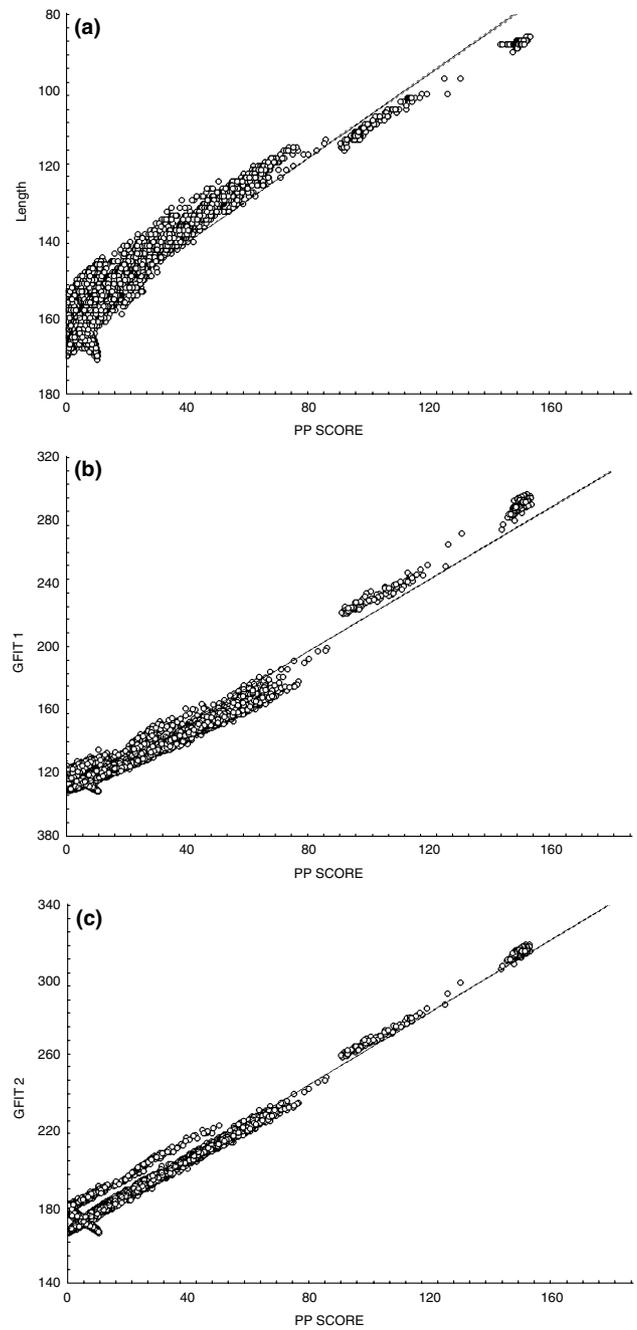


Fig. 3. Plots of PP scores against length, GFit(1) and GFit(2), for 16000 trees (including the best trees and some suboptimal) for Coscarón and Miranda-Esquivel's (1998) dataset (see text for details). (a) PP scores vs. Lengths ($r^2 = 0.9239$). (b) PP scores vs. GFit(1) values ($r^2 = 0.9712$). (c) PP scores vs. GFit(2) values ($r^2 = 0.9847$). The r^2 values are given only for reference and are not intended to be statistical statements.

states (Appendix 2). We used Carter et al.'s (1990) formula to calculate PP-score values, using PMaker.

With NONA, 12 trees were found (153 steps, CI = 0.37, RI = 0.68). Under IW, with a concavity

value of six, the only tree found was one of the most parsimonious (153 steps, $G\text{Fit}(6) = 466.9$ (56%)). This tree was also selected by PP (PP score = 372.509). Note that the Henry (1997) data matrix is nearly complete, with few empty cells (0.6% of matrix entries).

PP problems

The values obtained with PP are a function of the relative number of apomorphic and plesiomorphic states across taxa for a particular character. There are two ways in which the proportion of apomorphic/plesiomorphic could be distorted, the first is a result of missing entries and the second a result of multistate characters.

Using the log-form of Eqn 1 and Goloboff's (1991) modification to accommodate missing entries and by dividing by the total number of trees to obtain the log of difference in probability ($\log \Delta\text{PP}$), Carter et al.'s (1990) theorem 1 may be reduced to:

$$\begin{aligned} \log(2t - 3s) + \sum [n = 1; a - 1](\log(2a - s - n)) \\ + \sum [n = 1; p - 1](\log(2p - s - n)) \\ - \sum [n = 0; t - s - 1](\log(2t - 2s - 1 - 2n)) \\ - \sum [n = 0; s - 1](\log(s - n)) - (t - 2s) \log 2 \quad (2) \end{aligned}$$

Note that the number of missing data points does not change the form of the equation (see Appendix 1). The number of filled cells (cells with apomorphic or plesiomorphic character states) is lower in a dataset with missing entries than in the same dataset without missing entries, so the frequencies of the character with missing cells are artificially increased. The number of trees (and therefore its frequency) with a given character length increases when missing values are present, because the total number of trees remains the same (see Goloboff, 1991), but the number of trees with non-missing cells is lower than the number of total trees. As a result, character DIF values decrease in comparisons between trees with low homoplasy and increase when trees with higher homoplasy are compared. The following comparison clearly shows how DIF values must decrease for the first extra steps when missing values are added, where t is equal to taxa, and x is equal to the number of missing cells: $(1/(2t - 5)) < (1/(2(t - x) - 5))$.

Missing and/or inapplicable entries are common and unavoidable in most phylogenetic studies. The reduced PP value that results when they are present is not a result of a loss of information but an artifact of the missing cells. For example, consider the matrix in Table 3, where parsimony under equal weights and Implied Weights for any concavity value gives two trees: (A (B (C ((D E) (F G)))))) and (A (B (C ((D F)(E G))))). PP prefers the

Table 3

A matrix with two missing cells (see text for details)

	1	2	3	4	5	6	7	8	9	10	11	12
A	0	0	0	0	0	0	0	0	0	0	0	?
B	1	0	0	0	0	0	0	0	?	0	0	0
C	1	1	0	0	0	0	0	0	0	0	0	0
D	1	1	1	1	1	1	0	0	1	1	0	0
E	1	1	1	1	1	1	0	0	0	0	1	1
F	1	1	1	1	0	0	1	1	1	1	0	0
G	1	1	1	1	0	0	1	1	0	0	1	1

first tree because there are four non-missing characters (5–8, see Table 3) that support the grouping, whereas the alternative solution is also supported by four characters (9–12), but two of them have empty cells (the question marks in characters 9 and 12). Characters 9 and 12 have the largest numbers of trees with non-homoplastic reconstructions. They have higher numbers of trees than, for example, characters 5 or 7. Faith and Trueman (2001) did not discuss or refer to this kind of problem with PP.

Multistate characters present a similar situation because they, like characters with missing data, alter the proportion of apomorphic to plesiomorphic states. Faith and Trueman (2001) suggested re-coding multistate characters: “if an initial character is recoded to produce many new characters, then those individual characters are evaluated as a unit” (p. 344). This procedure will almost certainly be impractical for real datasets, especially those with unordered multistates. Goloboff explicitly avoids the problem by developing a function that is not dependent on the number of character states (Goloboff, 1993, p. 89).

In contrast with IW, PP implicitly weights characters with missing cells and multistate characters prior to the analysis. We do not know how this form of weighting “results in reduced inferential errors” (Faith and Trueman, 2001, p. 345).

From our analyses, we found another possible flaw in the PP method. Many random trees are arbitrarily resolved polytomous trees (Coddington and Scharff, 1994). These trees are considered independently by PP, but they are really the same. Indeed, counting them as “independent events” is clearly an error. One of the tacit assumptions of PP is that all dichotomous trees are independent but in fact there is tree duplication in randomly generated trees and in the entire tree universe. These duplicated trees rely on the same character state distributions. The total number of unsupported trees is impossible to know without an exhaustive analysis of the dataset, so the effect of their inclusion is unpredictable.³

³This criticism may apply to many methods that imply or assume equal distribution of dichotomous trees (as Bayesian inference; Huelshbeck et al., 2002, p. 678) or a random character distribution based on dichotomous trees (as Strong Evidence; Salisbury, 1999, p. 145).

Faith and Trueman rationalized PP under a Popperian epistemological framework of corroboration. While we do not wish to enter into a philosophical discussion of this topic, we can point out that IW received strong criticism under precisely this framework (Kluge, 1997). The concave function used by Faith and Trueman (2001) does not exclude PP from Kluge's (1997) criticism, because the use of homoplasy down-weighting functions undermines the severity of test advocated by Popper. Moreover, Farris (for example, 1995, 2000, Farris et al., 2001) criticized several statistical interpretations of the Popperian corroboration, and these criticisms would be equally applicable to Faith and Trueman's proposal.

The most common criticism raised against IW is the use of the k -value. Indeed, this value is still being evaluated (Goloboff, 1993, 1995), and some have asserted that it is "subjective" or "arbitrary" (e.g. Turner and Zandee, 1995). We argue that IW, in contrast to PP, has more flexibility because the user can select a concavity value to down-weight homoplastic characters (Goloboff, 1993). "Random" trees fix PP values beforehand and there are no constants to choose. That PP values are calculated in this way does not make them any more "objective". Not all possible trees have support, because it is a question satisfied by the characters themselves and not with the length distribution. We cannot imagine a justification for the preferred tree (h) being "the one having a profile least probably matched or bettered under *not-h*, given that selecting trees at random approximates looking for trees other than h " (Faith and Trueman, 2001, p. 344). If Implied Weights can be criticized as arbitrary the same can be said for PP. Perhaps PP values can be considered "more" arbitrary because a concavity function is being used, but the user is not aware of it.

A strong empirical criticism against down-weighting homoplastic characters was raised by Källersjö et al. (1999). They found that highly homoplastic characters may be more informative than less homoplastic ones. Their conclusion was that the use of concave functions of homoplasy (as IW of Goloboff, 1993; successive weighting of Farris, 1969) may not be well advised (Källersjö et al., 1999, p. 93). Although there are responses to Källersjö et al.'s criticism (Fontal-Cazalla et al., 2002, p. 175; also Farris, 2001), PP also down-weights homoplastic characters with a concave function and therefore it shares the same criticism.

Aside from all these problems PP is clearly an inefficient method. Tree searches are becoming increasingly efficient (Farris, 1988; Goloboff, 1998), and algorithms for optimal or partial solutions are continuing to be developed (Goloboff, 1996; Gladstein, 1997; Ronquist, 1998; Goloboff, 1999; Moilanen, 1999; Nixon, 1999; Goloboff and Farris, 2001; Moilanen, 2001). Using these technologies it seems reasonable to

conclude that searching for optimal trees with PP could be as fast as a search with Pee-Wee, but PP searches require two cycles to complete their evaluation. The initial examination of randomly generated trees that is required to calculate the probability profile is very slow: the length of each of the random trees must be calculated (without the benefit of powerful shortcuts). A way to avoid the first cycle of the PP method is to use a deterministic approach, an exact value for the number of reconstructions, given the number of steps of the reconstruction used, for binary characters, by employing Carter et al.'s (1990) theorem one. This approach was used by Goloboff (1991) in his data decisiveness program. Faith and Trueman (2001) seem fixed on the random trees approach (p. 344, our italics): "we evaluate each character *on random tree topologies*... allowing estimation of improbability of the observed character fit *over random trees*". Indeed in the header of their Table 2 they remark "[t]he number of trees is typically estimated from a *random sample of trees*" (our italics), but as the number of taxa increases, it is difficult to get non or low-homoplastic reconstructions for the majority of the characters, and as a result arbitrary values could be assigned to these characters. Faith and Trueman (2001) did not consider the use of an exact number based on a tree distribution equation, and the exact strategy offered by Faith and Trueman (2001), the counting of "all trees", is only possible using very small data sets.

Final remarks

PP and IW have the same aim and are, essentially, variations of the same approach: a tree search using a concave function of homoplasy. For large data sets, at least as presented by Faith and Trueman (2001), PP is computationally unworkable, even with the most powerful computer available. It is preferable to use computer time to improve the quest for optimal trees, rather than to evaluate random (and likely useless) trees. Even if the computer time were not a constraint, real data sets typically include multistate characters and inapplicable/unknown entries, and the method suggested by Faith and Trueman is surely misguided in these instances. This is a practical problem that can be overcome using Implied Weights. Faith and Trueman suggested that a PP-like approach could be used to combine the results of different data sets. If the reader is interested in this application or any other use of PP, we suggest that Pee-Wee be employed.

Paraphrasing the final words of Faith and Trueman (2001, p. 348), and given the superiority of IW, we suggest PP be synonymized with "PIWE*", highlighting that this is not a new form of parsimony, but one (albeit ill-defined) of the many functions of homoplasy available for use under the implied weights approach.

Acknowledgements

We thank Paula Posadas, Jonathan Coddington and Mike Sharkey for their suggestions and help with the English grammar in early drafts. Three anonymous referees provided several comments, which are also appreciated. Arnold Kluge provided editorial suggestions. This research was supported in part by the grant 1102-05-13563 Colciencias (Consejo Nacional de Ciencia y Técnica, Colombia). The programs and datasets used are available at Laboratorio de Sistemática y Biogeografía UIS website <<http://ciencias.uis.edu.co/labsist/intro.html>>.

References

- Carter, M., Hendy, M., Penny, D., Székely, L.A., Wormald, N.C., 1990. On the distribution of lengths of evolutionary trees. *SIAM J. Disc. Math.* 3, 38–47.
- Coddington, J., Scharff, N., 1994. Problems with zero-length branches. *Cladistics*, 10, 415–423.
- Coscarón, S., Miranda-Esquivel, D.R., 1998. *Petrovomyia*, a new neotropical genus of Prosimuliini (Diptera: Simuliidae): *Gigantodax* s. lat. split into two genera. *Ent. Scand.* 29, 161–167.
- Faith, D.P., Trueman, J.W.H., 2001. Towards an inclusive philosophy for phylogenetic inference. *Syst. Biol.* 50, 331–350.
- Farris, J.S., 1969. A successive approximations approach to character weighting. *Syst. Zool.* 18, 374–385.
- Farris, J.S., 1988. Hennig86, Version 1.5. Port Jefferson, New York.
- Farris, J.S., 1995. Conjectures and refutations. *Cladistics*, 11, 105–118.
- Farris, J.S., 2000. Corroboration versus ‘strongest evidence’. *Cladistics*, 16, 385–393.
- Farris, J.S., 2001. Support weighting. *Cladistics*, 17, 389–394.
- Farris, J.S., Kluge, A.G., Carpenter, J.M., 2001. Popper and likelihood versus ‘Popper*’. *Syst. Biol.* 50, 438–444.
- Fontal-Cazalla, F.M., Buffington, M.L., Nordlander, G., Liljeblad, J., Ros-Farré, P., Nieves-Aldrey, J.L., Pujade-Villar, J., Ronquist, F., 2002. Phylogeny of Eucolliinae (Hymenoptera: Cynipoidea: Figitidae). *Cladistics*, 18, 154–199.
- Gladstein, D.S., 1997. Efficient incremental character optimization. *Cladistics*, 13, 21–26.
- Goloboff, P.A., 1991. Homoplasy and the choice among cladograms. *Cladistics*, 7, 215–232.
- Goloboff, P.A., 1993. Estimating character weights during tree search. *Cladistics*, 9, 83–91.
- Goloboff, P.A., 1995. Parsimony and weighting: a reply to Turner and Zandee. *Cladistics*, 11, 91–104.
- Goloboff, P.A., 1996. Methods for faster parsimony analysis. *Cladistics*, 12, 199–220.
- Goloboff, P.A., 1998. Pee-Wee/NONA, Versions 3.0/2.0. Instituto Miguel Lillo, Tucumán, Argentina.
- Goloboff, P.A., 1999. Analyzing large data sets in reasonable times: solutions for composite optima. *Cladistics*, 15, 415–428.
- Goloboff, P.A., Farris, J.S., 2001. Methods for quick consensus estimation. *Cladistics*, 17, S26–S34.
- Henry, T.J., 1997. Phylogenetic analysis of family groups within the infraorder Pentatomomorpha (Hemiptera: Heteroptera), with emphasis on the Lygaeoidea. *Ann. Entomol. Soc. Am.* 90, 275–301.
- Huelsbeck, J.P., Larget, B., Miller, R.E., Ronquist, F., 2002. Potential applications and pitfalls of Bayesian inference of phylogeny. *Syst. Biol.* 51, 673–688.
- Källersjö, M., Albert, V.A., Farris, J.S., 1999. Homoplasy increases phylogenetic structure. *Cladistics*, 15, 91–93.
- Kluge, A.G., 1997. Sophisticated falsification and research cycles: consequences for differential character weighting in phylogenetic systematics. *Zool. Scripta*, 26, 349–360.
- Moilanen, A., 1999. Searching for most parsimonious trees with simulated evolutionary optimization. *Cladistics*, 15, 39–50.
- Moilanen, A., 2001. Simulated evolutionary optimization and local search: introduction and application to tree search. *Cladistics*, 17, S12–S25.
- Nixon, K.C., 1999. The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics*, 15, 407–414.
- Page, R.D.M., 1993. COMPONENT, Version 2.0. Natural History Museum, London. Available at <http://taxonomy.zoology.gla.ac.uk>.
- Ronquist, F., 1998. Fast Fitch-parsimony algorithms for large data sets. *Cladistics*, 14, 387–400.
- Salisbury, B.A., 1999. Strongest evidence: maximum apparent phylogenetic signal as a new cladistic optimality criterion. *Cladistics*, 15, 137–149.
- Swofford, D.L., 2001. PAUP*: Phylogenetic Analysis Using Parsimony, Version 4. Sinauer Associates, Sunderland, MA.
- Turner, H., Zandee, R., 1995. The behavior of Goloboff tree fitness measure F. *Cladistics*, 11, 57–72.

Appendix 1. Mathematics

Carter et al.’s (1990) theorem 1 (here Eqn 1), was extended by Goloboff (1991) to include missing data. Goloboff (1991) noted that missing entries do not add steps, but increase the number of trees. So (1) becomes

$$P(a, p, s) \times (\Pi[n = t; \text{taxa} - 1](2n - 1)) \quad (\text{A1})$$

Applying log transformation on (A1) to get only sums, we get:

$$\begin{aligned} \log(2t - 3s) + \sum[n = 2; t - 1](\log(2t - 1 - 2n)) \\ + \sum[n = 1; a - 1](\log(2a - s - n)) \\ + \sum[n = 1; p - 1](\log(2p - s - n)) \\ + \sum[n = t; \text{taxa} - 1](\log(2n - 1)) \\ - \sum[n = 0; t - s - 1](\log(2t - 2s - 1 - 2n)) \\ - \sum[n = 0; s - 1](\log(s - n)) - (t - 2s) \log 2. \quad (\text{A2}) \end{aligned}$$

(A2) is the log(number of trees for s steps), so it is the log of the difference in the number of trees with s or fewer steps and $s + 1$ or fewer steps. Dividing (A1) by the total number of trees for a given taxon number, we get the ΔPP . Then $\log(\Delta\text{PP})$ is equal to (A2) by canceling the underlined factors (i.e. the total number of trees).

Appendix 2. Modified characters in Henry (1997)

The following characters in Henry’s (1997) matrix were re-scored and transformed into binary variables. These changes were required to analyze a real matrix with PP and do not represent our dissatisfaction with Henry’s (1997) matrix, nor do they challenge Henry’s (1997) results.

(3) *Sutures between ocelli and compound eyes*. States 1 and 2 were merged. (0) absent; (1) present.

(8) *Antennal segment I*. States 0 and 1 were merged. (0) short; (1) very long.

(20) *Veins on hemelytral membrane*. States 0 and 1 were merged. (0) numerous veins (more than 5) and closed cells; (1) only 3–4 simple veins.

(23) *Position of spiracles*. States 1–4 merged. (0) all ventral; (1) at least II dorsal.

(28) *Fusion of abdominal sterna*. Separate segments (state 0) scored as unknown, states 2 and 3 were merged. (0) sterna 2–4 fused; (1) sterna 2–6 (2–7) fused.

(34) *Type of ovipositor*. States 0 and 1 were merged. (0) lanceolate; (1) plate-like.

(35) *Sternite VII*. States 0–2 merged. (0) split by ovipositor; (1) not split by ovipositor.

(47) *Abdominal trichobothria*. States 1 and 2 were merged. (0) absent; (1) present.

(48) *Position of trichobothria in abdominal segment IV*. Trichobothrial absence (state 0) scored as unknown. (0) lateral; (1) medioventral.

(52) *Number of salivary gland lobes*. States 0 and 1 were merged. (0) 1–3; (1) 4.

(56) *Dorsal abdominal scent gland opening*. States 1 and 2 were merged. (0) single; (1) double.