Short Communication

REP provides meaningful measurement of support across datasets

Taran Grant a,*, Arnold G. Kluge b

aLaboratório de Sistemática de Vertebrados, Faculdade de Biociências, Pontifícia Universidade Católica do Rio Grande do Sul, Av. Ipiranga 6681, 90619-900 Porto Alegre, RS, Brazil
b3140 Dolph Drive, Ann Arbor, MI 48103, USA

A R T I C L E  I N F O

Article history:
Received 12 August 2009
Revised 22 October 2009
Accepted 22 October 2009
Available online 27 October 2009

Keywords:
Group support
Goodman–Bremer support
Support measure
Parsimony
Relative fit difference
Optimality

A B S T R A C T

The relative optimality of the best and next-best hypotheses indicates the strength of support for the optimal hypothesis and may be calculated as either the difference or ratio of their optimality scores (Grant and Kluge, 2007). The Goodman–Bremer support measure (GB; Goodman et al., 1982; Bremer, 1988; Grant and Kluge, 2008b) calculates the support for a given clade in the most parsimonious tree as the difference between the length of the optimal tree that lacks that clade (S) and the length of the optimal tree (S), S−S. This measure allows comparison of support within a dataset, but the GB values from datasets of different sizes and tree lengths are not comparable.

Grant and Kluge (2007) derived a support measure based on the ratio of optimality scores, which simplifies to (S−S)/(G−S), where G denotes the length of the least parsimonious tree, 1 and for epistemological reasons (for review see Grant and Kluge, 2009) they named it the ratio of explanatory power, REP. As such, the REP value for a given clade is the difference in optimality scores between the optimal trees with and without that clade (GB), relative to how great that difference could be (the maximum possible GB for any clade in the tree, GBmax), GB/GBmax. That is, REP measures support as the proportional difference in optimality values and is therefore a scaled or relativized GB support measure. GB support could be scaled according to any number of factors, but this one has the advantages (among others) of maintaining the same rank order as GB within a dataset while also providing a logical basis to compare support across datasets (Grant and Kluge, 2007), thereby solving a persistent problem in the interpretation of GB.

Goloboff and Farris (2001) proposed the relative fit difference (RFD) to calculate the support for a given clade as the ratio of the evidence favorable (F) and contradictory to that clade, which simplifies to GB/F. As such, RFD scales GB according to the amount of evidence in favor of a given clade, F, which is the maximum possible GB for that clade. This scale factor varies among clades within a tree, such that clades of the same tree with the same GB value can have different scaled support values and the rank order of clades according to RFD can differ from that of GB (Grant and Kluge, 2008a). As Grant and Kluge (2007) noted when discussing the possibility of scaling GB according to branch-length, this may be suitable for some purposes, but it decouples support from optimality and makes comparison of support values across datasets meaningless. These are the practical differences between REP and RFD.

Farris and Goloboff (2008) challenged Grant and Kluge’s (2007, 2008a) claim that REP provides meaningful comparison of support across datasets and provided four examples (reproduced here as Figs. 1–4) designed to reveal the problematic behavior of REP. Below we will show that, far from demonstrating the problems that...
with REP. Farris and Goloboff’s (2008) examples actually illustrate the logically consistent behavior of REP when applied to different datasets.

Farris and Goloboff’s (2008) first example (Fig. 1) demonstrates that REP and RFD can exhibit the same behavior. Grant and Kluge (2008a) never claimed otherwise, and this example has no bearing on their finding that RFD does not necessarily vary in direct proportion with optimality. For Matrix 2B, Farris and Goloboff mistakenly reported RFD = 0.96 and REP = 0.92. For this dataset REP = \( \frac{(S_0 - S)}{(G - S)} = \frac{51 - 27}{52 - 27} = 0.96 \). If two characters contradict the clade YZ, then REP = 0.92, as does RFD.

Farris and Goloboff’s (2008) second and third examples show that datasets that differ in amount of evidence (synapomorphy) have different REP values. Fig. 2 compares matrices in which only the number of terminals that share an incongruent synapomorphy differs. In these matrices, the support for a clade measured as the difference in optimality scores (GB; squares) appears to be the same, because the lengths of the optimal tree \( S \) and optimal tree that lacks that clade \( S_0 \) are the same in each matrix. However, for support values to be comparable across datasets they must take into account how strong support could be (GBmax) for each dataset. Increasing the number of terminals with apomorphic character states increases the number of terminals with state 1, T. Grant, A.G. Kluge / Molecular Phylogenetics and Evolution 55 (2010) 340–342 341

Farris and Goloboff’s (2008) second and third examples show that datasets that differ in amount of evidence (synapomorphy) have different REP values. Fig. 2 compares matrices in which only the number of terminals that share an incongruent synapomorphy differs. In these matrices, the support for a clade measured as the difference in optimality scores (GB; squares) appears to be the same, because the lengths of the optimal tree \( S \) and optimal tree that lacks that clade \( S_0 \) are the same in each matrix. However, for support values to be comparable across datasets they must take into account how strong support could be (GBmax) for each dataset. Increasing the number of terminals with apomorphic character states increases the support for the clade YZ, as correctly indicated by the REP values, the proportional difference in optimality scores between the optimal trees with and without the clades in question is greater for the first dataset than the second. From this, it follows that the optimal solution of the first dataset is more strongly supported than that of the second dataset.

Farris and Goloboff’s (2008) second and third examples show that datasets that differ in amount of evidence (synapomorphy) have different REP values. Fig. 2 compares matrices in which only the number of terminals that share an incongruent synapomorphy differs. In these matrices, the support for a clade measured as the difference in optimality scores (GB; squares) appears to be the same, because the lengths of the optimal tree \( S \) and optimal tree that lacks that clade \( S_0 \) are the same in each matrix. However, for support values to be comparable across datasets they must take into account how strong support could be (GBmax) for each dataset. Increasing the number of terminals with apomorphic character states increases the support for the clade YZ, as correctly indicated by the REP values, the proportional difference in optimality scores between the optimal trees with and without the clades in question is greater for the first dataset than the second. From this, it follows that the optimal solution of the first dataset is more strongly supported than that of the second dataset.
the length of the least parsimonious tree \((G)\), which, in turn affects \(GB_{\text{max}}\). Consequently, the same GB value is proportionately smaller because \(GB_{\text{max}}\) is greater. Unlike other parsimony support measures, REP measures the support for a clade as a proportion of the maximum possible support and therefore detects this proportional difference.

Fig. 3 compares matrices composed of different numbers of terminals with different numbers of congruent synapomorphies. If support is measured as the difference in optimality values (GB), then it appears that the clades of the larger matrices are more strongly supported than those of the smaller matrices. However, if support is measured as the ratio of optimality values (REP), it becomes apparent that the clades of the larger matrices are proportionately less strongly supported than those of the smaller matrices. That is, by scaling GB by how strong support could be for each dataset (\(GB_{\text{max}}\)), REP reveals that the increases in support do not keep pace with the increases in the size of the datasets.

Farris and Goloboff’s (2008:1067) statement that “such behavior seems thoroughly unsuitable for a measure of support, since the state frequencies have no apparent logical connection with support” is misleading. Their examples do not address the effect of altering state frequencies per se. Instead, they compare matrices composed of different amounts of evidence (i.e., synapomorphy), which alters the maximum support possible (\(GB_{\text{max}}\)) and therefore has an obvious logical connection with support. As correctly reported by REP, clades with identical GB values may have proportionately more or less support, depending on the amount of evidence in their respective datasets.

Farris and Goloboff’s (2008) final attempt to show the illogic of REP provides yet another example of REP’s logically consistent behavior. In this example, one 20-taxon dataset includes eight characters that favor a 2-taxon clade and seven that favor a contradictory 2-taxon clade, giving REP = 0.125, while another consists of 15 characters that all favor the same clade composed of half of the terminals, giving REP = 0.111. According to Farris and Goloboff, this is problematic because the first dataset includes about as much contradictory evidence as favorable evidence, whereas the second dataset has no contradictory evidence at all. The only reason Farris and Goloboff find this to be problematic is that they apply their epistemologically unjustified definition of support as the ratio of favorable and contradictory evidence. As is correctly indicated by the REP values, the proportional difference in optimality scores between the optimal trees with and without the clades in question is greater for the first dataset than the second. From this, it follows that the optimal solution of the first dataset is more strongly supported than that of the second dataset.

Although Farris and Goloboff (2008:1068) exclaimed with derision that REP and GB support are negatively correlated in their third and fourth examples (Figs. 3 and 4), this illustrates precisely why Grant and Kluge (2007) proposed using REP instead of GB to compare results from different datasets. The values of REP and GB are perfectly correlated within a dataset (unlike RFD); differences only arise when different datasets are compared. GB correctly measures the support within each dataset, but it is incapable of detecting relative differences. As such, even though GB values are greater in the larger matrices in Fig. 3, they are proportionately smaller. REP correctly reports this proportional difference, which is why it provides meaningful measurement of support across datasets.

**Acknowledgments**

We are grateful to Ward Wheeler and an anonymous reviewer for their helpful criticisms and suggestions. Any errors that remain are ours alone. This research was supported by CNPq Proc. 305473/2008-5 and FAPESP Proc. 2008/50928-1.

**References**


