WILEY-BLACKWELL

Cladistics 26 (2010) 657–663

Cladistics

10.1111/j.1096-0031.2010.00308.x

Distinctions between optimal and expected support

Ward C. Wheeler

Division of Invertebrate Zoology, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024-5192, USA

Accepted 8 February 2010

Abstract

Several commonly used support measures are discussed and described as either optimal or expected support. This distinction is based on whether the indices are based on a function of optimal and non-optimal hypotheses, or on the statistical expectation of clades.

© The Willi Hennig Society 2010.

Background

There is a plethora of support measures used to indicate the strength of cladistic (i.e. sister-group) statements in systematics today. These include: (i) those based on statistical resampling and probabilistic reasoning, such as jackknifing (Quenouille, 1949, 1956; Tukey, 1958; Farris et al., 1996), bootstrapping (Efron, 1979; Felsenstein, 1985; Efron and Tibshirana, 1993), and Bayesian clade posterior probability (Mau et al., 1999), and (ii) those linked to relative optimality values of hypotheses such as Bremer support (Goodman et al., 1982; Bremer, 1988), likelihood ratio (Fisher, 1912), and Bayes odd ratios (Jeffreys, 1935, 1961). While resampling techniques are broadly applicable–and their numerical values comparable as percentages—those linked to specific optimality criteria are not (e.g. marginal likelihoods cannot be calculated for parsimony scores).

Statistical significance tests have also been developed either explicitly (likelihood ratio tests summarized in Felsenstein, 2004; Yang, 2006) or implicitly (bootstrap values as confidence intervals–Felsenstein, 1985; Hillis and Bull, 1993). The discussion here is not concerned with whether or not a particular level of support is "significant," but with the relationship among the types of measures.

Given this diversity of measures and an absence of a general definition of "support" separate from the specifics of a particular index–how do these values relate to one another? This discussion argues that there are two classes of support measures: those that relate to optimality criteria directly, and those that are derived from statistical expectation of clades.

Optimality and optimal support

If we restrict ourselves to discussion of the three most prevalent optimality criteria (parsimony, likelihood, and Bayesian posterior probability), phylogenetic trees (hypotheses) are adjudicated by reference to minimum cost (weighted or unweighted), likelihood (probability of the data given a stochastic evolutionary model and time vector–μt branch lengths), or posterior probability (probability of the tree given data, a set of stochastic models, time vector, and prior probabilities of trees, models, and time vectors[1]).

We can then define *optimal support* (S^{o}) of a group (g) as a function f of the optimality values of the optimal tree, A, with a particular clade or group g and that of B, the best tree without that clade (Fig. 1):

Corresponding author:
E-mail address: wheeler@amnh.org

[1]This is based on use of the maximum *a posteriori* tree (MAP) Rannala and Yang (1996) as opposed to clade posteriors (Mau et al., 1999; Wheeler and Pickett, 2008)

© The Willi Hennig Society 2010

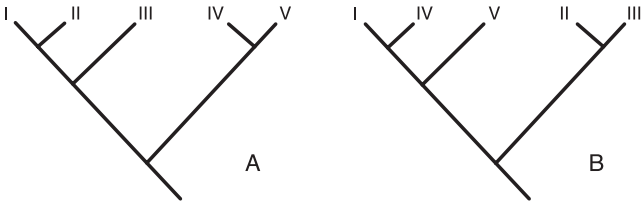

Fig. 1. Trees A (left) and B (right) with group $g = \{I, II, III\}$ present in A but not B.

$$S^o(g) = f(A, B).$$

In parsimony, f is usually the Bremer support, the difference in cost between the two hypotheses A and B:

$$S^o_{\text{parsimony}}(g) = B_{\text{cost}} - A_{\text{cost}}.$$

For likelihood, f is their likelihood ratio:

$$S^o_{\text{likelihood}}(g) = \frac{pr(D|A)}{pr(D|B)}.$$

The Bayes odds ratio contains terms for the prior probabilities of alternate trees, $pr(T)$, as well as their likelihood, resulting in an analogous ratio to that of the likelihood ratio, but between marginal integrated likelihoods.

$$S^o_{\text{Bayes}}(g) = \frac{pr(A) \cdot pr(D|A)}{pr(B) \cdot pr(D|B)}$$

Furthermore, we can show close links between these functions via the use of the No-Common-Mechanism model (Tuffley and Steel, 1997). If we consider a set of n characters, each with r states, and l_i the number of changes for the ith character on tree T. The non-additive Fitch (1970) parsimony cost of that tree is $T_{\text{cost}} = \sum_i^n l_i$. The likelihood of T is (Tuffley and Steel, 1997):

$$T_{\text{likelihood}} = \prod_i^n r^{-(l_i+1)}.$$

If we take the logarithm of the likelihood value we get:

$$\begin{aligned} \log\ T_{\text{likelihood}} &= \log \prod_i^n r^{-(l_i+1)} \\ &= \sum_i^n -(l_i + 1) \log\ r \\ &= -T_{\text{cost}} \cdot \log\ r - n \cdot \log\ r. \end{aligned}$$

The logarithm of the likelihood ratio for group g given A and B above will then be:

$$\begin{aligned} \log\ S^o_{\text{likelihood}}(g) &= \log\ A_{\text{likelihood}} - \log\ B_{\text{likelihood}} \\ &= (B_{\text{cost}} \cdot \log\ r - n\ \log\ r) \\ &\quad - (A_{\text{cost}} \cdot \log\ r - n\ \log\ r) \\ &= (B_{\text{cost}} - A_{\text{cost}}) \cdot \log\ r \\ &= S^o parsimony\ \cdot \log\ r \end{aligned}$$

(Wheeler, 2006). Given flat priors for trees, the Bayesian posterior odds ratio reduces to the same value.

Resampling, clade posteriors, and expected support

Resampling support measures are not calculated from optimal tree values, hence are not measures of optimal support. Even though the determination of tree quality in resampled data occurs (as it must) via a specific optimality criterion, the actual tree optimality values are not used further, only the frequency of reconstructed groups. Likewise, clade posteriors (Mau et al., 1999; Huelsenbeck and Ronquist, 2003) are not functions of the posterior probability of the MAP tree (and may conflict; Wheeler and Pickett, 2008). If these values are not measuring optimal support, what are they measuring?

Consider a probability distribution h on trees in tree sample space (set of all potential trees) Ω. Let

$$m(g|T) = \begin{cases} 0 & \text{if } g \notin T \\ 1 & \text{otherwise} \end{cases}$$

be the occurrence ($m()$) of group g in tree T. Then the expectation (mean) of the occurrence of group g over all trees is the product of m and h integrated over tree space, Ω:

$$E[m(g|T)] = \sum_{T \in \Omega} m(g|T) \cdot h(T)$$

This expression yields the average support of a group over the tree space, and can be referred to as expected support, $S^e(g) = E[m(g|T)]$.

Bayesian posterior probabilities of trees offer $h(T)$ explicitly (based on observations and prior distributions of character change, trees, and edge lengths). Hence, the clade posteriors as in Mau et al. (1999) can be considered expected support. Resampling methods generate pseudo-replicate data sets of the same (bootstrap) or reduced (jackknife) size. Typically, a single tree (however collapsed) is determined as best for that data set. This process will also generate the $h(T)$ distribution of equation 3. Jackknife and bootstrap methods therefore, also measure expected support.

An example

As an illustration of the behaviour of optimal and expected supports, consider the arthropod anatomical data of Giribet et al. (2001). These data consist of 303 characters for 54 taxa. Here, all characters are treated as non-additive (seven were additive in the original analysis). For the likelihood and Bayesian analyses, No-Common-Mechanism (NCM; Tuffley and Steel, 1997) was employed with a flat prior distribution (for trees $T, p(T_i) = p(T_j); \forall i, j \in \Omega_T$)) on tree topologies. Flat clade priors are impossible (Steel and Pickett, 2006), hence clade posteriors (in the form of Bayesian expected support) will have the well known clade size effect (Pickett and Randle, 2005).

With the exception of the determination of the clade posteriors, all analyses were performed using POY version 4.1.1 Varón et al. (2008, 2010)). Bremer support values were determined by examining a pool of 100 TBR neighborhoods after random addition Wagner builds with command line: `build(100) swap(all, visited: "tmp") report("bremer.pdf", graphsupports:bremer:"tmp")`. Resampling supports were determined from 250 replicates in each case with command line: `calculate support(jackknife: (resample:250), build(), swap(tbr, trees:5)) report("jack.pdf", graphsupports:jackknife:consensus)` for jackknife (e^{-1} deletion) and: `calculate_support(bootstrap: 250, build(), swap(tbr, trees:5)) report ("boot.pdf", graphsupports:bootstrap:consensus)` for bootstrap runs. Clade posterior probabilities were calculated as in Wheeler and Pickett (2008) using MrBayes (Huelsenbeck and Ronquist, 2003) with options: `lset parsmodel=yes mcmc ngen=10 000 000 samplefreq=500 nchains=4` resulting in two runs of four chains each with 10 million generations sampling every 500th visited topology.

Results are summarized in Figs 2–4. Parsimony runs are shown in Fig. 2 (98 most parsimonious trees at length 596), with broad but not complete agreement among the Bremer, jackknife, and bootstrap supported

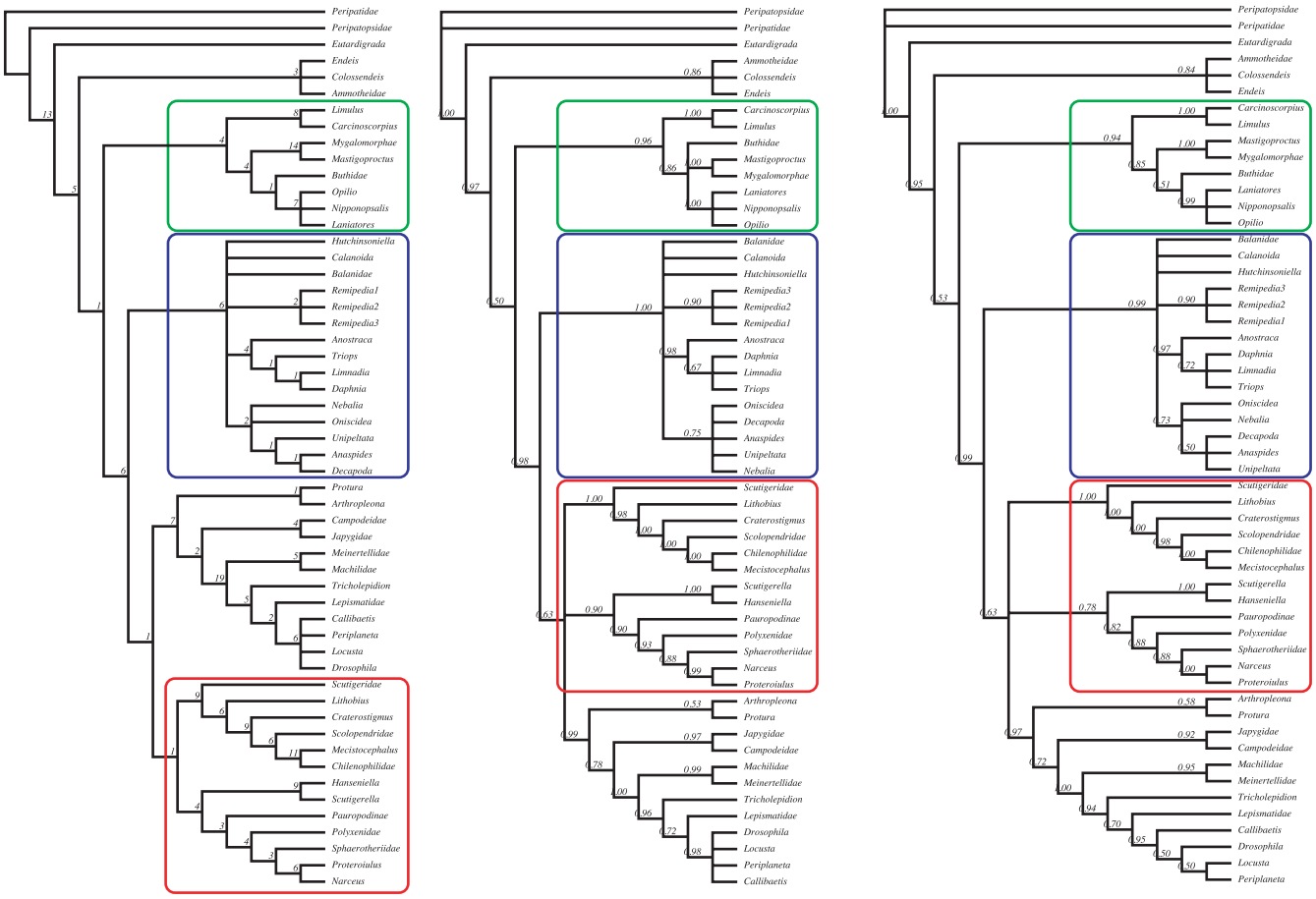

Fig. 2. Parsimony support analyses with: Bremer (support > 0) left, Jackknife (> 0.50) centre, and Bootstrap (> 0.50) right. Myriapod taxa are outlined in red, crustacean in blue, and euchelicerate in green.

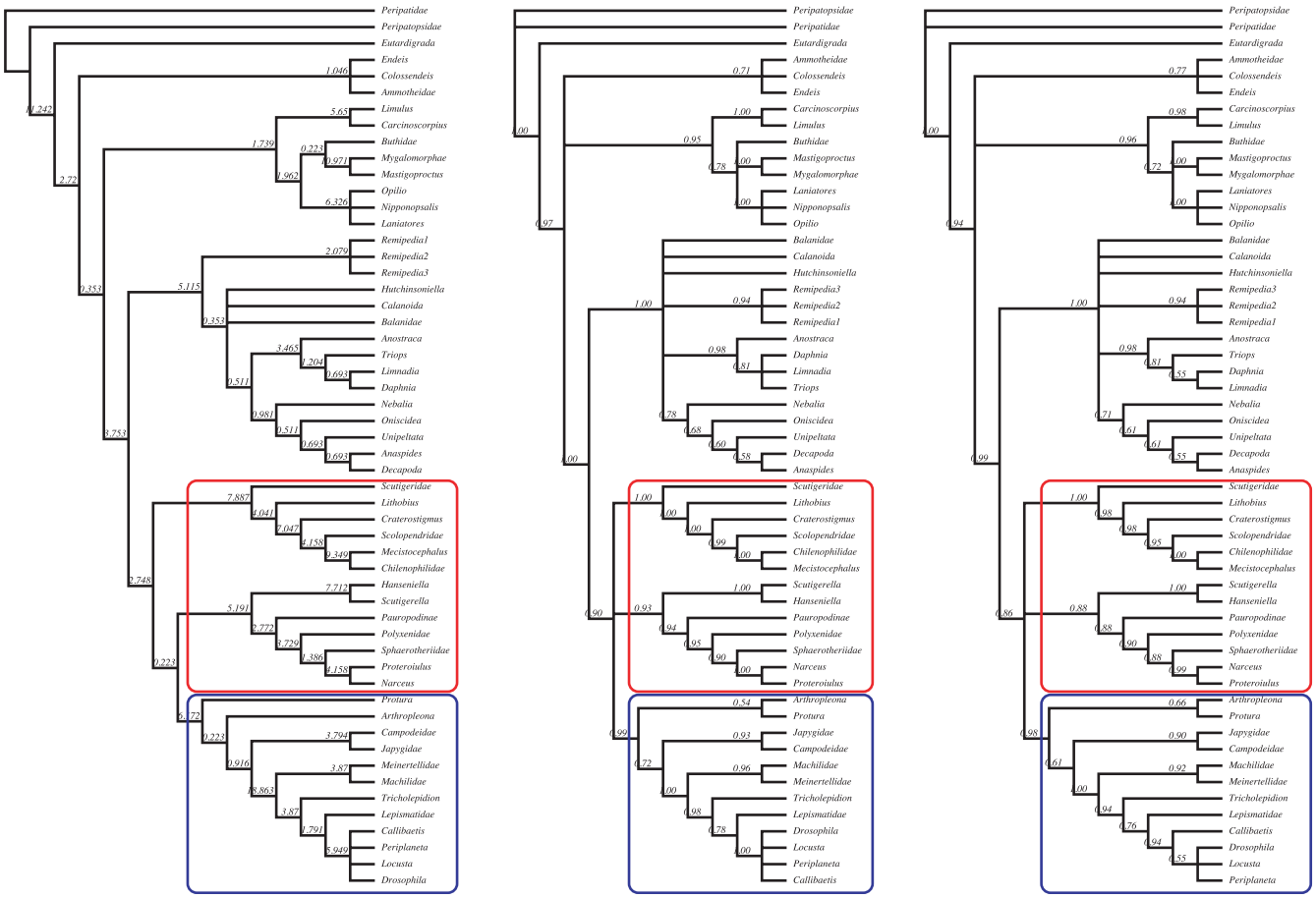

Fig. 3. Likelihood (NCM) support analyses with: −log likelihood ratios (support > 0) left, Jackknife (> 0.50) centre, and Bootstrap (> 0.50) right. Myriapod taxa are outlined in red and hexapod in blue.

clades. In general, a greater number of nodes show Bremer support (e.g. Myriapoda) and the jackknife (e.g. Euchelicerata and Crustacea) the least.

Likelihood results are shown in Fig. 3 (53 most likely trees at log lik −773.817). As with parsimony, likelihood ratios support more groups than resampling, with the jackknife most conservative. Again, Myriapoda is an example, in this case a supported paraphyletic arrangement with respect to Hexapoda (as opposed to the monophyly in the parsimony analysis).

Bayesian support results are shown in Fig. 4. Since the priors of trees are equal, Bayes factors reduce to likelihood ratios (although integrated) and are repeated from Fig. 3 on the MAP tree. One of the interesting differences among these support measures, and between the clade-based expected Bayesian support (Fig. 4, right) and all the others is the upholding of chelicerate monophyly. Optimal support under all three optimality criteria upheld chelicerate paraphyly (with pycnogonids–*Endeis*, *Collosendeis*, and Ammotheidae basal) as did expected parsimony support. Both instances of expected likelihood support (jackknife and bootstrap) were agnostic with support slightly lower than 0.5. The expected Bayesian (via MrBayes) stood alone with 0.6 posterior probability of monophyly.

The purpose of this demonstration was not to determine general aspects of support behavior, but to illustrate the applicability of optimal and expected support in these three analytical frameworks. A second motivation was to demonstrate that these measures can support alternate phylogenetic scenarios. This is not a defect in any particular measure, but a result of the fact that optimal and expected support measure different aspects of the data.

Support and hypothesis testing

Given the extensive discussion of the ideas of Popper with respect to support (e.g. Grant and Kluge, 2007, 2008; Farris, 2008), it is fitting and proper to integrate these ideas with Popper's formalisms of explanatory power (Popper, 1959), corroboration (Popper, 1959), severity of test (Popper, 1963).

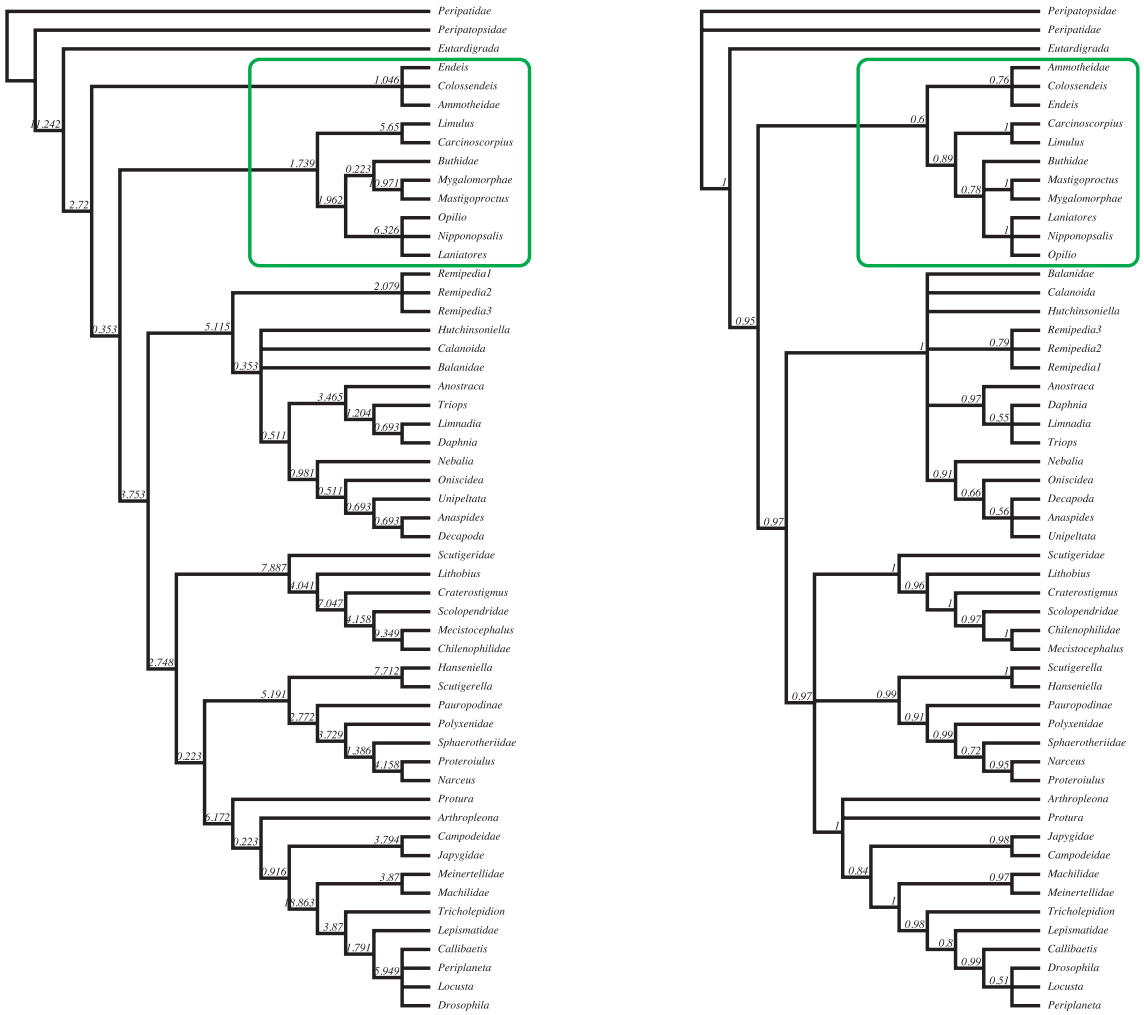

Fig. 4. Bayesian support analyses with MAP: −log odds ratios (support > 0) left and clade posterior probabilities as reported by MrBayes (> 0.50) right. Chelicerate taxa are outlined in green.

Severity of test (S and S′) and explanatory power (E and E′) are both described by the same alternate formalisms[2]

$$S(e|h,b) = E(h|e,b) = \frac{pr(e|h,b) - pr(e|b)}{pr(e|h,b) + pr(e|b)}$$

$$S'(e|h,b) = E'(h|e,b) = \frac{pr(e|h,b)}{pr(e|b)}$$

based on probabilities pr of evidence e (= data), hypothesis h (in this case, a tree), and background knowledge b including all untested assumptions. These are interpreted either as the discriminating ability of the evidence in light of a hypothesis and background knowledge, for the former, or the ability of the hypothesis to account for the data in the latter.

Corroboration shares the numerator of S and E with an alternate normalizing factor in the denominator,

$$C(h,e,b) = \frac{pr(e|h,b) - pr(e|b)}{pr(e|h,b) - pr(eh|b) + pr(e|b)}$$

hence is directly proportional to these other measures.

Clearly, these formulations fall into the category of optimal support measures, with severity of test and explanatory power as functions of likelihoods with A and B from above: $pr(e|hb) = pr(D|A)$ and $pr(e,b) = pr(D|B)$ (as Popper, 1959; himself points out) conditioned upon background knowledge, including model assumptions. Optimal support does not correspond exactly to the notion of objective support (Grant and Kluge, 2007, 2008), since the latter relies on an

[2]The S and E used here for severity of test and explanatory power are not to be confused with the S and E used earlier to signify support and expectation. Changing either, I fear, would be more confusing than overusing standard, if non-unique, terms.

additional requirement based on an alternate formulation of E, namely the minimum number of character transformations required by a tree (= equally weighted parsimony score) (Kluge and Grant, 2006; Grant and Kluge, 2008).

Conclusions

The purpose of this discussion is to make more precise the inter-relationships among several commonly used indices of support. In short, if we consider the distribution of tree cost over the universe of possible trees, optimal support values measure differences among trees at the extreme tail of the distribution, ideally, the absolute extreme values. Expected support, on the other hand, looks to the central tendencies of the distribution, its centre of mass, its mean. These measures describe group support only for the data set at hand, and any conclusions drawn are specific to those observations.

The example presented here demonstrates that optimal and expected support flow naturally from optimality criteria and can be applied across different analytical paradigms. These two classes of support measures target alternate aspects of the data, hence may disagree on those clades that are "supported" and those that are not.

If one envisions characters as samples from a universe of potential observations, expected support offers predictive statements about the occurrence of groups in future, unobserved data. If one views characters as historically unique entities, such distributional statements are without much meaning. Optimal support measures are concerned only with the relative optimality values of alternate hypotheses, hence are linked directly to the criteria that governed hypothesis choice initially, placing them at the centre of optimality-based hypothesis testing.

Acknowledgements

I would like to thank Megan Cevasco, Louise Crowley, Steven Farris, Pablo Goloboff, Taran Grant, Isabella Kappner, Kurt Pickett, Chris Randle and Andrés Varón for discussion of these ideas and comments on this manuscript. Research was supported by the US Army Research Laboratory and the US Army Research Office (W911NF-05-1-0271) as well as NSF-ITR grant "Building the tree of life: A national resource for phyloinformatics and computational phylogenetics" (NSF EF 03-31495), "An Integrated approach to the origin and diversification of protostomes" (NSF DEB 05-31677), and "Assembling the tree of life: phylogeny of spiders" (NSF EAR 02-28699).

References

Bremer, K., 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. Evolution 42, 795–803.
Efron, B., 1979. Bootstrap methods: another look at the jackknife. Ann. Stat. 7, 1–26.
Efron, B., Tibshirana, R., 1993. An Introduction to the Bootstrap. New York, Chapman and Hall.
Farris, J.S., 2008. Parsimony and explanatory power. Cladistics 24, 825–847.
Farris, J.S., Albert, V.A., Källersjö, M., Lipscomb, D., Kluge, A.G., 1996. Parsimony jackknifing outperforms neighbor-joining. Cladistics 12, 99–124.
Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.
Felsenstein, J., 2004. Inferring Phylogenies. Sunderland, MA, Sinauer Associates.
Fisher, R.A., 1912. On an absolute criterion for fitting frequency curves. Mess. Math. 41, 155–160.
Fitch, W.M., 1970. Distinguishing homologous from analogous proteins. Syst. Zool. 19, 99–113.
Giribet, G., Edgecombe, G.D., Wheeler, W.C., 2001. Arthropod phylogeny based on eight molecular loci and morphology. Nature 413, 157–161.
Goodman, M., Olson, C.B., Beeber, J.E., Czelusniak, J., 1982. New perspectives in the molecular biological analysis of mammalian phylogeny. Acta Zoologica Fennica 169, 19–35.
Grant, T., Kluge, A.G., 2007. Ratio of explanatory power (rep): a new measure of group support. Mol. Phylogenet. Evol. 44, 483–487.
Grant, T., Kluge, A.G., 2008. Clade support measures and their adequacy. Cladistics 24, 1051–1064.
Hillis, D.M., Bull, J.T., 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42, 182–192.
Huelsenbeck, J.P., Ronquist, F. 2003. MrBayes: Bayesian inference of phylogeny, 3.0 edn. Program and documentation available at: http://morphbank.uuse/mrbayes/.
Jeffreys, H., 1935. Some tests of significance, treated by the theory of probability. Proc. Cambridge Phil. Soc. 31, 203–222.
Jeffreys, H., 1961. Theory of Probability, 3rd edn. Oxford, Oxford University Press.
Kluge, A.G., Grant, T., 2006. From conviction to anti-superfluity: old and new justifications for parsimony in phylogenetic inference. Cladistics 22, 276–288.
Mau, B., Newton, M.A., Larget, B., 1999. Bayesian phylogenetic inference via Markov Chain Monte Carlo methods. Biometrics 55, 1–12.
Pickett, K.M., Randle, C.P., 2005. Strange Bayes indeed: uniform topological priors. Mol. Phylogenet. Evol. 34, 203–211.
Popper, K., 1959. The Logic of Scientific Discovery. London, Routledge.
Popper, K., 1963. Conjectures and Refutations, 2002 edn. London, Routledge.
Quenouille, M.H., 1949. Approximate tests of correlation in time-series. J. R. Stat. Soc. B 11, 68–84.
Quenouille, M.H., 1956. Notes on bias estimation. Biometrika 43, 353–360.
Rannala, B., Yang, Z., 1996. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. J. Mol. Evol. 43, 304–311.
Steel, M., Pickett, K.M., 2006. On the impossibility of uniform priors on clades. Mol. Phylogenet. Evol. 39, 585–586.
Tuffley, C., Steel, M., 1997. Links between maximum likelihood and maximum parsimony under a simple model of site substitution. Bull. Math. Biol. 59, 581–607.
Tukey, J.W., 1958. Bias and confidence in not-quite large samples (abstract). Ann. Math. Stat. 29, 614.

Varón, A., Vinh, L.S., Bomash, I., Wheeler, W.C. 2008. *POY 4.0*. American Museum of Natural History. http://research.amnh.org/scicomp/projects/poy.php.

Varón, A., Vinh, L.S., Wheeler, W.C., 2010. POY version 4: phylogenetic analysis using dynamic homologies. Cladistics 26, 72–85.

Wheeler, W.C., 2006. Dynamic homology and the likelihood criterion. Cladistics 22, 157–170.

Wheeler, W.C., Pickett, K.M., 2008. Topology-Bayes versus clade-bayes in phylogenetic analysis. Mol. Biol. Evol. 25, 447–453.

Yang, Z., 2006. Computational Molecular Evolution. UK, Oxford University Press.