[1]*División Mastozoología, Museo Argentino de Ciencias Naturales "Bernardino Rivadavia", Buenos Aires, Argentina - CONICET;*
[2]*Departamento de Ciencias Básicas, Universidad Nacional de Luján, Buenos Aires, Argentina*

Evaluating the clade size effect in alternative measures of branch support

MARÍA AMELIA CHEMISQUY[1] and FRANCISCO J. PREVOSTI[1,2]

Abstract

The clade size effect refers to a bias that causes middle-sized clades to be less supported than small or large-sized clades. This bias is present in resampling measures of support calculated under maximum likelihood and maximum parsimony and in Bayesian posterior probabilities. Previous analyses indicated that the clade size effect is worst in maximum parsimony, followed by maximum likelihood, while Bayesian inference is the least affected. Homoplasy was interpreted as the main cause of the effect. In this study, we explored the presence of the clade size effect in alternative measures of branch support under maximum parsimony: Bremer support and symmetric resampling, expressed as absolute frequencies and frequency differences. Analyses were performed using 50 molecular and morphological matrices. Symmetric resampling showed the same tendency that bootstrap and jackknife did for maximum parsimony and maximum likelihood. Few matrices showed a significant bias using Bremer support, presenting a better performance than resampling measures of support and comparable to Bayesian posterior probabilities. Our results indicate that the problem is not maximum parsimony, but resampling measures of support. We corroborated the role of homoplasy as a possible cause of the clade size effect, increasing the number of random trees during the resampling, which together with the higher chances that medium-sized clades have of being contradicted generates the bias during the perturbation of the original matrix, making it stronger in resampling measures of support.

Key words: Branch support – clade size effect – maximum parsimony – resampling – Bremer support

Introduction

Previous authors found an association between the size of a clade and its support, causing large and small clades to be more supported than medium-sized clades (Pickett and Randle 2005; Randle and Pickett 2006; Brandley et al. 2009). The effect was first described in the context of Bayesian analyses, when Pickett and Randle (2005) showed that the values of the clade prior probabilities were affected in a predictable way by the size of the clade given the number of taxa in the matrix, when uniform topological priors were stipulated. These clade prior probabilities influenced the Bayesian clade support values, translating the prior effect to the node support estimations (i.e. posterior probability of a clade).

Surprisingly, the clade size effect was also found when calculating bootstrap and/or jackknife in maximum likelihood and maximum parsimony, methods that do not depend on prior values (Pickett and Randle 2005). The authors proposed that two different mechanisms were causing the effect: clade prior probabilities in the case of Bayesian support values and noise in the data sets (caused by homoplasy) in the case of resampling support. Apart from the noise, the authors suggested that the lack of data decisiveness, caused by character conflict, was also responsible for the clade size effect (Randle and Pickett 2006). In fact, this effect was previously noted by Wenzel and Siddall (1999) in the context of the addition of noise to a matrix or the random replacement of its entries with random states. Albeit the effect reported by Wenzel and Siddall (1999) was related to a stability measurement and not to measures of branch support, it is essentially the same bias that was later called the clade size effect.

Brandley et al. (2006) questioned Pickett and Randle's position and stated that the same mechanism was responsible for the clade size effect in Bayesian support, bootstrap and jackknifing, suggesting that taxon sampling may be one of the underlying mechanisms. Finally, Brandley et al. (2009) analysed the influ-

ence of homoplasy on the clade size effect and compared the intensity of the effect in Bayesian inference, maximum likelihood and maximum parsimony. They concluded, based on simulation analyses, that homoplasy might be one of the factors causing the clade size effect by including effectively random trees in the calculation of the support values. Also, Brandley et al. (2009) found that Bayesian inference was the least affected method by the clade size effect, while maximum parsimony was the most affected one.

The aim of this study was to evaluate the presence of the clade size effect in alternative measures of group support available for maximum parsimony: symmetric resampling, expressing the group frequencies as absolute frequencies and as difference in frequencies (Goloboff et al. 2003), and Bremer support (Bremer 1994). The analyses were performed using empirical data sets, including molecular and morphological matrices, using the Clade Disparity Index (Brandley et al. 2009) and the two-dimensional Kolmogorov–Smirnov test (Garvey et al. 1998). Also, we plotted the node support values versus the clade size or the clade prior probabilities and explored the fitting of the data to the expected patterns of distribution (quadric and exponential, respectively), interpreting these results as another measure of the strength of the effect. Finally, we evaluated the possible causes of the effect, correlating the statistics with different characteristics of the data sets.

Materials and Methods

Data sets and phylogenetic analyses

For analysing the association between the size of a clade and its support, we used 50 empirical data sets, including 27 molecular matrices and 23 morphological data sets (Appendix 1). 15 of the 38 data sets analysed by Brandley et al. (2009) were included in this study. All the phylogenetic searches and node support calculations were carried out using the software TNT version 1.1 (Goloboff et al. 2008).

Heuristic searches were performed using 1000 random addition sequences (RAS) and tree bisection–reconnection (TBR). In the more

Corresponding author: María Amelia Chemisquy (amelych80@gmail.com)
Contributing author: Francisco J. Prevosti (protocyon@hotmail.com)

complex cases, searches were carried out by building 20 RAS trees, swapping each one with TBR, sectorial search (with the default parameters), and 20 iterations of tree-drifting (Goloboff 1999). The strict consensus was calculated in all the cases and was used as the reference tree for calculating the clade prior probabilities and the node support values. Clade prior probabilities were calculated using the Equation 1 published by Pickett and Randle (2005). Bootstrap and traditional jackknife were evaluated by previous authors (e.g. Pickett and Randle 2005; Brandley et al. 2009), so we chose to test the clade size effect using alternative measures of support, such as the symmetric resampling expressed as absolute frequencies (SR; Goloboff et al. 2003), and as differences of frequencies (GC; Goloboff et al. 2003), and the Bremer support (Br; Bremer 1994). Symmetric resampling is a modification of the traditional jackknife, where the probability (p) of a character of being upweighted, remaining unmodified or being deleted is the same (0.33; Goloboff et al. 2003), and it is the best choice when working with weighting schemes. The GC measures the differences in frequencies between a group and the most frequent contradictory group (Goloboff et al. 2003) and varies between -100 and 100 (contrary to other resampling methods that vary from 0 to 100). The GC can be calculated using the resampling strategy of the bootstrap, the jackknife or the symmetric resampling. Finally, the Bremer support measures the extra length needed to lose a branch in the strict consensus (Bremer 1994) and, contrary to the other methods, does not rely on perturbations to the matrix nor on a resampling scheme. Bremer support was calculated using the script *Bremer.run* (available in the web page tnt. insectmuseum.org) using the default parameters (searching with constrains for trees not lost in the one-step-longer trees, with sectorial search and 10 tree-drift cycles, repeating the search three times and showing the minimum value). SR and GC were calculated by performing 1000 pseudoreplicates saving 100 trees in each replicate, using the groups from the strict consensus tree. To avoid scaling artefacts, all the branch support values were taken to a 0–1 scale for the statistical analyses. For doing so, we divided each node support value by 100 in the case of SR; by the figure of the most supported node of each tree in the case of Bremer; and by $100 +$ the number of the lowest (most negative) supported node in the case of GC.

Evaluating the relationship between clade size and node support

We performed two different approaches to evaluate the relationship between clade size and node support. In the first one, we compared the clade prior probabilities with the branch support values using the Clade Disparity Index (CDI) proposed by Brandley et al. (2009). The CDI not only evaluates the presence of the clade size effect, but also estimates its magnitude (Brandley et al. 2009). The significance of the index was evaluated performing a randomization test where the node support values were resampled 10 000 times, obtaining new CDIs for each resampled group where the node support values were randomly distributed throughout the tree. The CDIs and its significance were calculated using the software R 2.9.2 (R Development Core Team 2004, Vienna, Austria). In the second approach, we compared the raw clade size (i.e. the number of taxa of each clade) with the branch support values using the two-dimensional Kolmogorov–Smirnov test (2DKS; Garvey et al. 1998) implemented on the software Ez2dks (distributed by J. Garvey at http://www.science.siu.edu/zoology/garvey/2dks.htm). Since the clade prior probabilities are estimated using the raw clade size, we expect to have similar results in both approaches. Moreover, because the clade prior probabilities are not used for tree searches in maximum parsimony, the examination of raw clade size becomes relevant.

Apart from the statistical evaluation of the clade size effect, we also examined the relationship between clade size (both raw values and $\log_{10}$ clade prior probabilities) and node support values using bivariate plots. We expect to find a quadric pattern of distribution (i.e. a U-shaped curve) when plotting the raw clade size values against the node support values, and an exponential pattern when plotting the clade prior probabilities (based on Pickett and Randle 2005; Brandley et al. 2009). Fittings were performed using a quantile regression, which does not require the data to be homoscedastic (Cade and Noon 2003). Curves were fitted to the graphs, and the significance of the fitting and the proportion of explained variance (R^2) were calculated using the package quantreg for the software

R. The proportion of explained variance was interpreted as another measure of the strength of the effect.

Evaluating the relationship between the clade size effect and other parameters

To analyse the hypotheses proposed by previous authors about the influence of homoplasy in the clade size effect, we tested the relationship between the CDI and the D statistic from the 2DKS test, with the amount of homoplasy measured with the Consistency Index (CI; Kluge and Farris 1969; Goloboff 1991a) using the Spearman's correlation index (Zar 1984). We also explored the correlation between the CDI and D with the following parameters: number of taxa and characters (an expression of the size of the tree and data set), number of nodes of the strict consensus tree (tree size and resolution), asymmetry of the strict consensus tree (tree shape; measured following Prevosti and Chemisquy 2009), mean support number of the strict consensus tree (a proxy of the phylogenetic signal of the data set) and cladistic decisiveness of the matrix (Goloboff 1991a) (Appendix 1). The significance of the correlations was tested using a randomization test resampling the CDI or the D values 10 000 times. The cladistic decisiveness of data is the degree to which the possible dichotomic trees differ in length (Goloboff 1991a,b), being 0 when all the possible resolved trees have the same length and increasing when the trees differ more in length (Goloboff 1991b). Cladistic decisiveness is inversely related to the homoplasy of the matrix (Goloboff 1991b) and was calculated using the equation for DD (from data decisiveness) published by Goloboff (1991a) implemented on a script designed for TNT by P. Goloboff (available from the authors of this contribution). We also analysed the effect of the type of data (i.e. molecular or morphological) on the clade size effect performing the Student's t-test with the CDI or the D values. All the statistical analyses were performed with the software R.

Finally, we tested the hypothesis of the relationship between the distribution of the signal of a character on the tree and the clade size effect (Brandley et al. 2006). If the hypothesis is correct, we expect a 'U' distribution in the spreading of the synapomorphies versus the clade size, where larger and smaller clades have higher numbers of synapomorphies. For doing so, we obtained the number of synapomorphies of each node of one randomly chosen most parsimonious tree. Then, we analysed whether there was an association between the number of synapomorphies and the size of each node using the 2DKS test. We also analysed the relationship between both variables in a qualitative way using bivariate plots and fitting a quadric pattern of distribution.

The statistical significance of the p value was determined, in all the analyses, using a p-plot and a sharpened Bonferroni method (García 2004). We considered each measure of group support and each statistic as a different family of tests.

Results

Relationship between priors, clade size and clade support

For most of the data sets, the CDI and the 2DKS test showed a similar pattern. The 2DKS test was significant for four data sets of the 50 in the Br, 14 data sets in the GC and 16 in the SR (Table 1). Of the 47 data sets for which clade prior probabilities were obtained, two had a significant CDI in the Br, nine had in the GC, and eight had in the SR (Table 1). The CDIs of the matrices that showed a significant relationship ranged between 0.08 and 0.4. Six of the 47 matrices had negative CDIs in the GC and the SR, but the values were never higher than -0.05, while for the Br, 16 matrices had negative CDIs reaching values of -0.14 (Table 1). SR and GC showed similar values of CDI in most of the data sets, but the relationship between the CDIs of Br and the resampling measures of support was not constant, being sometimes higher, sometimes lower and sometimes negative when GC and SR had positive values (Fig. 1).

When plotting clade support against clade size, results were ambiguous. The U-shaped pattern was not found, and some plots had dots evenly distributed throughout the graphic (even plots of data sets with high levels of significance in the 2DKS test;

Table 1. 2DKS and CDI results of Bremer support and symmetric resampling expressed as absolute frequencies (SR) and frequency differences (GC)

Study	Br				GC				SR			
	D	p	CDI	p	D	p	CDI	p	D	p	CDI	p
1 **Aliscioni et al. (2003)**	0.04010	0.55000	-0.00612	0.92570	0.03110	0.93700	0.02219	0.42890	0.02940	0.95500	0.01287	0.53260
2 **Anderson et al. (2004)**	0.09400	0.01900	0.14912	0.12440	0.12240	0.00160	0.20450	0.00040*	0.11400	0.00380	0.18769	0.00010*
3 Armbruster (2004)	0.61000	0.67000	0.11151	0.08290	0.12100	0.00020*	0.14352	0.00001*	0.10600	0.00040*	0.15708	0.00010*
4 Baker et al. (2000)	0.05500	0.66700	0.08796	0.44620	0.13000	0.06300	0.09058	0.35350	0.13000	0.07300	0.07077	0.32620
5 **Barkworth et al. (2008)**	0.10600	0.00360	0.02059	0.84930	0.11700	0.00120*	0.07523	0.03420	0.12100	0.00060*	0.07072	0.02340
6 **Barns et al. (1996)**	0.07870	0.11300	0.03705	0.88960	0.09040	0.05300	0.11442	0.00640	0.09300	0.03680	0.12092	0.00360
7 **Bell and Donoghue (2003)**	0.05500	0.90100	-0.01075	0.93910	0.05500	0.86700	0.02011	0.62850	0.05500	0.86700	0.00833	0.69120
8 **Berbee and Taylor (2001)**	0.05000	0.64100	0.02430	0.78040	0.09000	0.06100	0.07740	0.06340	0.09000	0.06380	0.07950	0.04690
9 **Whitten et al. (2007)**	0.05930	0.21800	0.10695	0.28850	0.05600	0.34100	0.05088	0.25250	0.06500	0.17000	0.03907	0.19460
10 **Brown et al. (2008)**	0.14500	0.00020*	0.22473	0.03340	0.15700	0.00020*	0.19448	0.00080*	0.15700	0.00020*	0.17727	0.00100*
11 Cabrero-Sañudo (2007)	0.05100	0.41100	-0.04678	0.73470	0.12600	0.00220*	0.09789	0.06100	0.11000	0.01100	0.10552	0.07930
12 Jaramillo (2008)	0.03900	0.52700			0.06210	0.06420			0.05930	0.08120		
13 **Chemisquy and Morrone 2010**	0.09250	0.39100	-0.05571	0.85440	0.13900	0.04720	-0.00893	0.86310	0.13900	0.04740	-0.00920	0.80900
14 **Chemisquy and Morrone in press**	0.05300	0.56500	-0.14319	0.13060	0.04400	0.83600	0.00919	0.77100	0.04300	0.86600	-0.00206	0.91870
15 **Chemisquy et al. (2010)**	0.05900	0.19700	-0.03782	0.88410	0.05660	0.34700	0.04327	0.23440	0.05660	0.34700	0.03830	0.28600
16 de Jong et al. (1996)	0.08700	0.17900	-0.09484	0.32350	0.08920	0.21700	0.01222	0.84110	0.08800	0.23100	0.00532	0.90810
17 **Des Marais et al. (2003)**	0.09260	0.37600	-0.08793	0.61360	0.06170	0.83000	-0.06611	0.31400	0.06170	0.82400	-0.03738	0.34590
18 **Dohrmann et al. (2008)**	0.05280	0.62200	0.10953	0.24590	0.03760	0.25600	0.04937	0.19110	0.06810	0.31000	0.02795	0.26290
19 **Edwards (2005)**	0.05330	0.78600	0.02243	0.86750	0.06220	0.62100	0.03412	0.32220	0.05670	0.72000	0.02587	0.33310
20 Freitas and Brown (2004)	0.03200	0.90100	0.03736	0.62420	0.07100	0.16000	0.06699	0.03940	0.07300	0.14500	0.06297	0.03960
21 **Garey et al. (1996)**	0.13200	0.00100*	0.36952	0.00001*	0.11900	0.00480	0.14107	0.00060*	0.12700	0.00280*	0.08441	0.00080*
22 Gasparini et al. (2006)	0.38000	0.49000	0.05034	0.38450	0.12720	0.00100*	0.18133	0.00200*	0.14210	0.00020*	0.15156	0.00020*
23 Gaudin (2004)	0.05500	0.56400	0.20391	0.00810	0.11200	0.01220	0.16558	0.00020*	0.10500	0.03080	0.12047	0.00040*
24 Grant et al. (2006)	0.07600	0.08300	0.10585	0.01930	0.07600	0.18500	0.06870	0.24710	0.11600	0.00600	0.14917	0.06390
25 **Ives and Taylor (2008)**	0.07810	0.88300	-0.24025	0.26130	0.12500	0.30700	-0.05775	0.22180	0.12500	0.30700	-0.05930	0.21820
26 James (2004)	0.02200	0.93600	0.03260	0.71260	0.08900	0.16200	0.15924	0.06180	0.11000	0.04700	0.14267	0.02100
27 **Jordan et al. (2003)**	0.06250	0.30300	0.03860	0.70890	0.06760	0.20700	0.00411	0.91250	0.06760	0.21000	0.00373	0.87700
28 Kaila (2004)	0.03000	0.79000			0.11700	0.00020*			0.11900	0.00020*		
29 **Kelch and Baldwin (2003)**	0.16500	0.06400	0.04703	0.87130	0.11600	0.30500	0.07847	0.14360	0.16500	0.06200	0.07044	0.04520
30 Komarek and Beutel (2007)	0.04960	0.84890	0.07526	0.41390	0.11200	0.13000	0.05362	0.43080	0.91000	0.33000	0.06142	0.18560
31 Leschen and Buckley (2007)	0.09640	0.00880	0.09124	0.13670	0.17760	0.00020*	0.23111	0.00060*	0.17600	0.00020*	0.29592	0.00050*
32 Lindgren et al. (2004)	0.08400	0.25300	-0.14390	0.10960	0.08800	0.23800	-0.00950	0.78370	0.08800	0.23800	-0.02321	0.39300
33 Lipscomb (1989)	0.11600	0.00120*	0.12263	0.14000	0.16800	0.00020*	0.21033	0.00040*	0.15800	0.00020*	0.23667	0.00060*
34 **Mori et al. (2007)**	0.05310	0.29700	-0.06270	0.76770	0.07810	0.02640	0.06474	0.01420	0.07250	0.05200	0.04940	0.02820
35 Nei et al. (2007)	0.39500	0.81600	-0.00676	0.94920	0.12800	0.00320	0.11082	0.03140	0.13700	0.00180*	0.11402	0.01580
36 O'Leary et al. (2004)	0.03300	0.98600	0.07083	0.45460	0.08100	0.32700	0.10653	0.22440	0.08100	0.33200	0.06858	0.25910
37 **O'Leary et al. (2009)**	0.05100	0.51000	-0.00817	0.93090	0.09160	0.01540	-0.00237	0.94830	0.06300	0.32300	-0.00294	0.91130
38 Pitts et al. (2006)	0.05400	0.16800	0.07192	0.29650	0.10500	0.00720	0.13886	0.01150	0.12390	0.01200	0.17287	0.01250
39 Pramuk (2006)	0.07500	0.37500	0.03037	0.89060	0.08900	0.34600	0.13601	0.09900	0.08900	0.34600	0.11732	0.08860
40 **Prevosti (2010)**	0.07100	0.70010	0.05669	0.78440	0.04400	0.99000	0.01763	0.88180	0.04400	0.98500	0.02524	0.56400
41 **Reinert et al. (2004)**	0.08700	0.00600	0.20188	0.00001*	0.17700	0.00020*	0.39988	0.00001*	0.17700	0.00020*	0.33537	0.00001*

Table 1. (continued)

Study	Br				GC				SR			
	D	p	CDI	p	D	p	CDI	p	D	p	CDI	p
42 Rokas et al. (2003)	0.08330	0.80100	0.08721	0.71290	0.08330	0.82600	-0.00671	0.85340	0.08330	0.82600	0.00041	1.00000
43 Shivonen (2005)	0.06010	0.09840	-0.13293	0.60920	0.15400	0.00020*	0.20632	0.00680	0.15400	0.00020*	0.23967	0.00340
44 Sikes et al. (2008)	0.07140	0.29100	0.20559	0.11800	0.13300	0.00280	0.09006	0.01550	0.08860	0.11700	0.07167	0.02390
45 Swenson and Anderberg (2005)	0.10200	0.00900	0.16128	0.11810	0.14400	0.00020*	0.24243	0.00030*	0.14300	0.00040*	0.33163	0.00001*
46 Wanntorp et al. (2006)	0.03800	0.91500	-0.01269	0.97940	0.06230	0.85400	0.01544	0.92090	0.06230	0.87600	0.01171	0.91580
47 Weisrock et al. (2006)	0.08240	0.01040	0.09702	0.21110	0.08120	0.00900	0.03895	0.22070	0.08120	0.00860	0.02867	0.16550
48 Wills et al. (1998)	0.13780	0.01040	0.32072	0.01080	0.13780	0.09320	0.33998	0.02990	0.13780	0.09320	0.24140	0.03320
49 Wu et al. (2001)	0.06500	0.21000	-0.02711	0.72630	0.10900	0.00860	0.14775	0.00570	0.11700	0.00200*	0.15215	0.00820
50 Yen et al. (2005)	0.06900	0.00100*			0.10600	0.00020*			0.10600	0.00020*		
Median			0.03860				0.07739				0.07072	

*Significant p values after Bonferroni correction: 0.0125 (D, Br); 0.00113 (CDI, Br); 0.00227 (D, GC); 0.00207 (CDI, GC); 0.00208 (D, SR); 0.0022 (CDI SR).

p expresses the statistical significance of each test.

Molecular data sets are in bold.

Fig. 2). When fitting a quadric function, we found that the R^2 was never higher than 0.46 (Table 3), meaning that the fitting of this model was not very good. Although the plots with higher R^2 showed a pattern that could be seen as a U, both small and large clades (mainly the small ones) had high and low values of node support (i.e. a larger variance than the middle-sized clades; Fig. 2). There was a difference between measures of node support in the results of the regressions; Br showed much lower R^2 values than SR and GC, having only 12 data sets with a R^2 higher than 0.1 (versus 32 matrices for the SR and GC; Table 3). Results were similar when plotting clade support against clade prior probabilities (Fig. 2). In this case, an exponential function was fitted to the graphs, being the highest R^2 0.48. Similar to what happened with the clade size, GC and SR had three times more matrices with an R^2 over 0.1 than Br (26 versus 7; Table 3).

Differences between molecular and morphological matrices

Molecular and morphological matrices had a similar performance with the Br, having both the same number of significant data sets in the 2DKS test and the CDI test (Table 1). The CDI values were also similar between both kinds of data sets (Fig. 3a), and the t-test was not significant for the 2DKS test ($t = -1.47$, p = 0.1546) nor for the CDI ($t = -0.8676$, p = 0.3903).

Conversely, the clade size effect was significantly greater for morphological matrices than molecular matrices (Table 1, Fig. 3b,c). The Student's t-test showed significant differences in the effect between both kinds of data sets for the 2DKS test and the CDI in the GC ($t = -36.217$, p = 0.0007066 and $t = -38.547$, p = 0.000494) and only for the CDI in the SR ($t = -42.41$, p = 0.00018 and $t = -1.9797$, p = 0.05944 for the 2DKS test).

Correlation with other parameters of the matrix

The CDI values and the D parameter based on the Br did not show any significant correlation with the parameters tested (Table 2). The CDI values and the D parameter based on GC and SR showed a significant negative correlation with the Consistency Index, the number of characters, the mean support and the data decisiveness (Table 2). The CDI values calculated using the SR also had a significant and positive correlation with the number of taxa. The asymmetry and the number of nodes of the strict consensus tree never had a significant correlation.

Synapomorphies and clade size

The number of synapomorphies of each node did not show a significant pattern with the clade size, and only one matrix was significant under the 2DKS test (matrix Reinert et al. 2004; Table S1). When analysing the plots of the synapomorphies versus the clade size, results were inconclusive. The plots did not show a clear U pattern (see examples in Fig. 4), and when regressing a quadric function, none of the matrices had a significant regression, and the R^2 was never higher than 0.25 (Table S1).

Discussion

Comparison between measures of support

Our results clearly show that the clade size effect has a stronger influence on resampling measures of support. Only 5% of the matrices had a significant CDI between Bremer support and the clade prior probabilities, while almost 20% were significant when analysing the symmetric resampling (both absolute frequencies
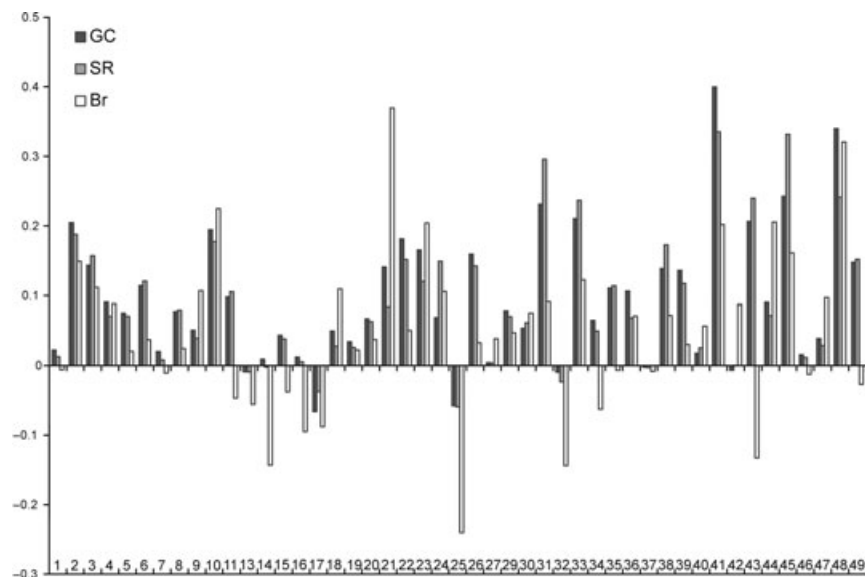
Figure 1. Clade Disparity Indexes (CDIs) of the 47 empirical data sets described in Table 1. Br, Bremer support; SR, symmetric resampling expressed in absolute frequencies; CG, SR expressed in frequency differences. The numbers of the *x*-axis correspond to the number that each matrix has in the Table 1

and frequency differences). Maximum observed CDI values were quite elevated, reaching almost 0.4 in the GC and Br and 0.33 in the SR. However, for each branch support measure, only six matrices had CDI values over 0.2, and the median was of ~0.07 for SR and GC and 0.039 for Br. This means that although the clade size effect affects the performance of the clade support measures, the strength of the effect (measured both with the figures and with the number of statistically significant CDIs) is not really high. It is noteworthy that while only six matrices had a negative CDI for the GC and the SR (and never lower than −0.066), 14 matrices had a negative CDI calculated with the Bremer support, even reaching −0.24. These results imply that in Br, there is a tendency to have the inverse effect (i.e. having higher values of support in the middle-sized clades). Only one of these matrices with negative CDIs was significant with the randomization test. Results with the 2DKS show the same tendency, but the number of matrices with a significant relationship between clade support and its size was larger (8% for BR, 28% for GC and 32% for SR). The larger number of matrices that were statistically significant in the 2DKS could be due to the test itself and not to the use of the node size or the clade prior probabilities, because the analysis of the node support versus the clade prior probabilities with the 2DKS showed similar results than the node size analyses (data not shown).

We found that the number of matrices with significant p values is lower (20% versus 31%) than the number found by Brandley et al. (2009). However, if we employ Bonferroni correction in the same way as Brandley et al. (2009) did, we have the same percentage of statistically significant matrices. Nonetheless, the number of matrices with CDI values over 0.2 is similar in both analyses (13% versus 16%) as well as the maximum value (0.32 in Brandley et al.'s work). Moreover, the Student's *t*-test was not significant between the results presented here and the results from Brandley et al. (2009) ($t = 0.5691$, $p = 0.571$ with SR; $t = 0.7332$, $p = 0.4657$ with GC). It is important to mention that Brandley et al. (2009) calculated the branch supports using PAUP, which uses a majority rule consensus tree instead of a strict consensus tree for estimating the branch support values (Davis et al. 1998; Grandcolas 2004; Simmons and Freudenstein 2011). Although the use of the majority rule consensus tree over-

estimates the branch support and could be one of the causes of the clade size effect (Simmons and Freudenstein 2011), we obtained similar results calculating the branch supports using the strict consensus tree (at least for resampling measures of support), implying that there are other factors causing this effect.

When contrasting our results with Brandley et al.'s (2009) results for maximum likelihood (ML), a similar pattern to the one described above was observed, because our results had a lower number of matrices with significant p values (20% versus 37%; again, the numbers become similar when correcting our results differently), and the maximum CDI remained similar to the one we obtained with the absolute frequencies of the symmetric resampling support (0.33 versus 0.31). For the ML results, the percentage of matrices with CDI values over 0.2 was only 8% (versus 13% for the MP results); however, this difference was not significant when comparing the CDI values of both analyses with a Student's *t*-test ($t = 0.6469$, $p = 0.5195$ with SR and $t = 0.8252$, $p = 0.4117$ with GC).

The comparison with the Bayesian posterior probabilities (PP) showed different results, because only 16% of the matrices had significant p values, and most importantly, the highest CDI only reached 0.092. These results confirm the hypothesis proposed by Brandley et al. (2009) that Bayesian posterior probabilities are less affected by the clade size effect than the MP resampling measures of support, because the difference in the CDI values between PP, SR and GC was highly significant when compared with a Student's *t*-test ($t = 4.6888$, $p = 0.000012$ and $t = 4.8914$, $p = 0.0000056$, respectively).

A similar result led Brandley et al. (2009) to conclude that MP bootstrapping has a poor performance, and it is highly affected by clade size, encouraging researchers to prefer Bayesian posterior probabilities. They also stated that the increased magnitude of the effect is related to the lack of capacity of MP to deal appropriately with homoplasy. However, our results obtained with Bremer support are as good as Bayesian PP results, having fewer matrices with significant p values (5% for Br versus 16% for PP), and non-significant differences on the CDI values ($t = 1.7181$, $p = 0.09036$). Moreover, the same tendency to have the reverse effect in 26% of the matrices reported by Brandley et al. (2009) is similar to our results from Bremer
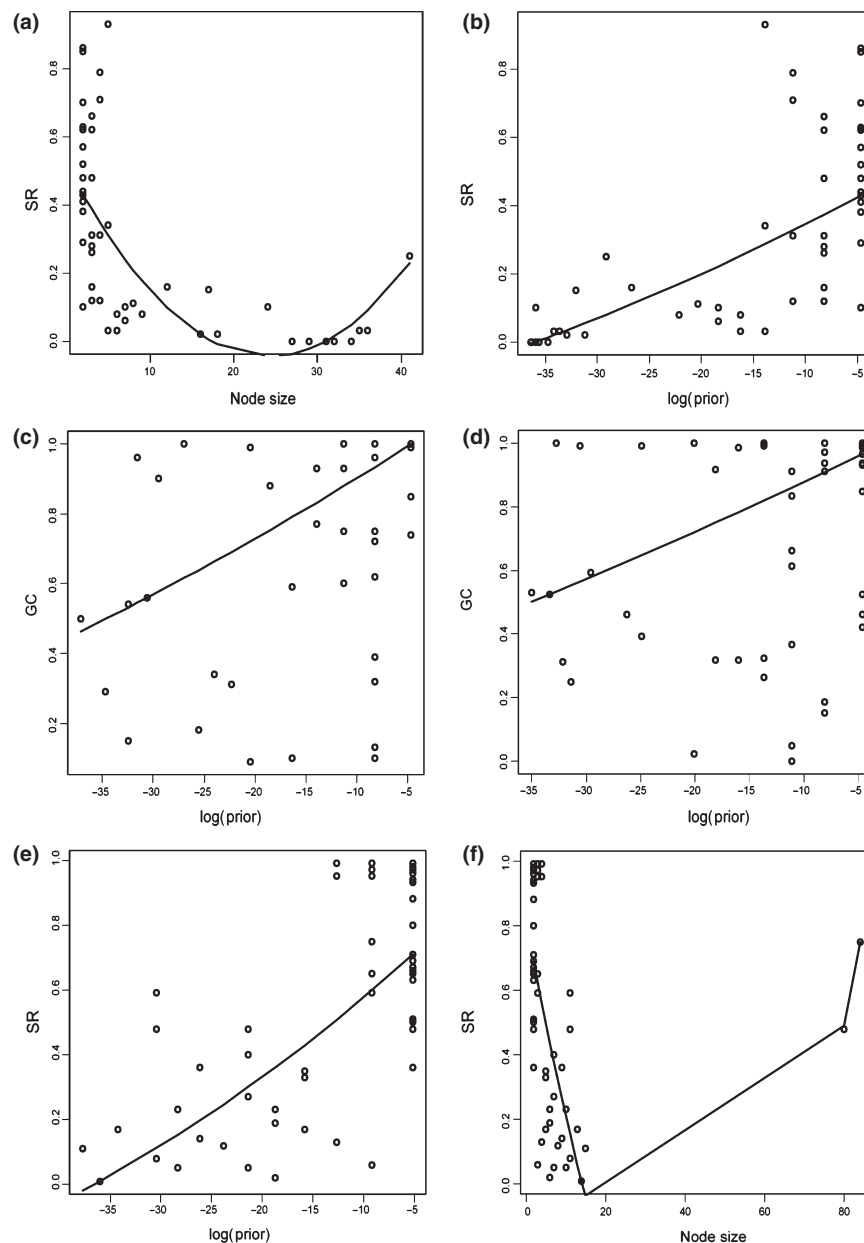
Figure 2. Distribution of measures of nodal support plotted against prior clade probability or node size for the following data sets: (a) and (b) Leschen and Buckley (2007), (c) Garey et al. (1996), (d) Berbee and Taylor (2001), (e) and (f) Lipscomb (1989). Lines indicate exponential or quadric fitting estimated using quantile regression. Br, Bremer support; SR, symmetric resampling expressed in absolute frequencies; CG, SR expressed in frequency differences. Node support values were rescaled to a 0–1 scale. Plots were chosen to show different examples of distribution and fitting. See Discussion for more details

support. In conclusion, the poor performance of MP bootstrap has nothing to do with the method used to infer the phylogeny, but with the method used to obtain the branch support values (i.e. a resampling method).

Possible causes of the clade size effect

Our results indicate that homoplasy, the number of characters, the mean support and the data decisiveness might be responsible for the clade size effect in the resampling measures of support. This is interesting because these four factors are connected. For example, an increase in the number of characters (and taxa) leads to higher levels of homoplasy (i.e. there is an inverse relationship

between the number of characters and taxa with the CI, because higher values of CI imply less homoplasy; see Archie 1989, 1996; Sanderson and Donoghue 1996), while more homoplasy is related to lower values of support (i.e. there is a positive correlation between CI and mean support; e.g. Sanderson and Donoghue 1996; Prevosti and Chemisquy 2010). Although decisiveness is not a function of the amount of homoplasy on a data set and it is not directly related to the CI (Goloboff 1991a), the same author reported that when the CI is higher than 0.5, there is a positive relationship between CI and DD. With this data set, we found a positive correlation between the data decisiveness and the CI (Spearman's $r = 0.65$, $p < 0.00001$), something that could be interpreted as a relationship between both parameters.
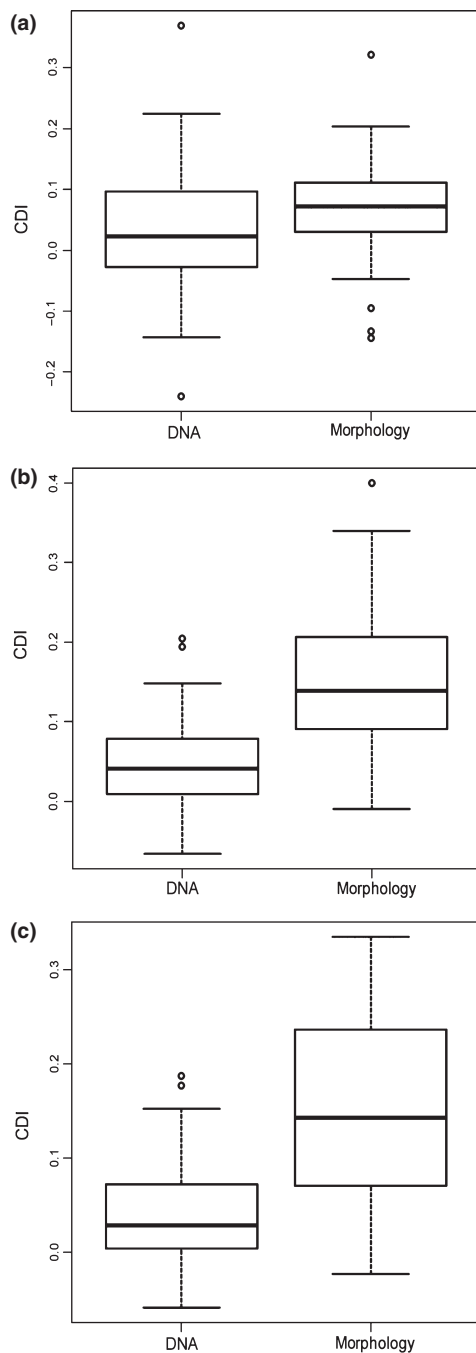
Figure 3. Boxplot of the CDIs discriminated by type of data (molecular or morphological). (a) Bremer support; (b) symmetric resampling expressed in absolute frequencies; (c) symmetric resampling expressed in frequency differences

Consequently, more homoplasy probably leads to less mean support and data decisiveness. The correlation of these variables with the CDI agrees with previous interpretations about the causal role of the homoplasy in the size clade effect (Pickett and Randle 2005; Brandley et al. 2009).

The relationship between homoplasy and the clade size effect can also be seen in the differences between molecular and morphological data sets. It is widely known that morphological data sets have higher levels of homoplasy than molecular matrices (e.g. Sanderson and Donoghue 1996; Givnish and Sytsma 1997a, b) and that is also true for the matrices included in the present study (Fig. 4a; Student's t-test $t = 5.664$, p = 0.000001). This

difference in the levels of homoplasy is translated in higher CDIs for the morphological matrices and more morphological data sets with a significant relationship between nodes or clade prior probabilities and clade support (Fig. 4b–d; Table 1).

The question that arises is why homoplasy is particularly problematic for resampling measures of support? The answer lies in the way in which each measure of branch support is calculated. Posterior probabilities are calculated using the trees sampled in the MCMC chains (Huelsenbeck and Ronquist 2005), while Bremer support is based on suboptimal trees (Bremer 1994), but none of them alter the data matrix during support calculation. Conversely, resampling techniques modify the matrix in different ways to explore the stability of the nodes obtained in the original optimal trees (Felsenstein 1985; Farris et al. 1996). The modification of the matrix during its resampling may cause the appearance of a secondary signal and noise (Brochu 1999; Wenzel and Siddall 1999) that modifies the resampled tree, and in this context, the homoplasy could originate new contradictory clades that are not present in optimal and suboptimal trees obtained from the non-modified matrix. The number of these contradictory groups probably increases with the homoplasy, which is expected due to the negative relationship between mean support and homoplasy (Archie 1989, 1996; Sanderson and Donoghue 1996; Prevosti and Chemisquy 2010; this study).

Medium-sized clades have larger probabilities of being contradicted. Take as an example a fully pectinate tree of $n = 10$ taxa (Fig. 5a); the odds of breaking a clade of T taxa could be given by the Equation

$$\frac{\left[\prod_{i=2}^{n} 2i - 3\right] - \left[\prod_{i=2}^{T} 2i - 3\right]\left[\prod_{i=T+1}^{n} 2i - 2T - 1\right]}{\left[\prod_{i=2}^{n} 2i - 3\right]} \quad (1)$$

which represents the number of possible rearrangements of the whole tree minus the rearrangements that do not compromise the monophyly of the clade i of n taxa (inside or outside the clade i). When plotting the chances of breaking a clade calculated using Equation 1, versus its size, we obtained a pattern inverse from the clade size effect pattern (Fig. 5b). The presence of more homoplasy could generate random contradictory clades, which will impact harder on medium-sized clades due to the higher probabilities given by Equation 1. If all the trees obtained from the resampled matrices (during the branch support calculation) were random trees, a perfect fit to the clade size effect would be obtained (such as Fig 1, Pickett and Randle 2005). This is difficult to obtain with real data sets, because the matrices retain some phylogenetic signal after the resampling, making the effect softer. However, as Brandley et al. (2009) suggested, when obtaining resampled measures of support, the presence of homoplasy during the random modification of the matrices could generate some amount of random or 'near random' trees, causing a significant (but not very strong) clade size effect. This is the case of the data sets analysed in this contribution.

A similar explanation was proposed by Goloboff and Pol (2005) to account for inconsistent estimations of Bayesian posterior probabilities and incorrect estimations of monophyly provided by Bayesian analyses. The authors also mentioned that something similar could be affecting bootstrap values estimated using PAUP* because of the implementation of the bootstrapping and jackknifing in that software. However, the authors did not deal with the clade size effect, nor with problems associated with branch support estimated using other softwares.

Wenzel and Siddall (1999) found a pattern similar to the clade size effect between clade size and a measure of stability (mojo values) when including random signal (e.g. random characters

Table 2. Results from the correlation test of CDI and D against different parameters

	nTax		nChar		CI		Imbalance		Mean support		nNodes		DD	
	R	p	R	p	R	p	R	p	R	p	R	p	R	p
CDI Br	−0.0368	0.0822	−0.1087	0.0473	−0.2442	0.0113	−0.0638	0.0646	−0.0787	0.0608	−0.0002	0.1	−0.2083	0.016
CDI GC	0.1941	0.0185	**−0.4076**	**0.0006**	**−0.5745**	**0.00001**	−0.0643	0.0679	**−0.5606**	**0.00001**	0.1193	0.0173	**−0.4492**	**0.0003**
CDI SR	**0.2685**	**0.0059**	**−0.4399**	**0.0001**	**−0.6152**	**0.00001**	−0.0844	0.0559	**−0.5617**	**0.00001**	0.2521	0.0074	**−0.4459**	**0.00001**
D Br	−0.1281	0.0371	−0.0387	0.0772	−0.0933	0.0501	−0.0717	0.0613	−0.0757	0.0601	−0.1051	0.048	−0.1359	0.0366
D GC	−0.0105	0.0944	**−0.4326**	**0.00001**	**−0.3645**	**0.0007**	−0.0427	0.0773	**−0.4824**	**0.00001**	0.0173	0.0901	**−0.2621**	**0.007**
D SR	0.0034	0.0977	**−0.4943**	**0.00001**	**−0.3943**	**0.0005**	−0.0849	0.0545	**−0.541**	**0.00001**	−0.0482	0.0752	**−0.3602**	**0.0016**

nTax, number of taxa; nChar, number of characters; CI, Consistency Index; nNodes, number of nodes of the strict consensus tree; DD, data decisiveness; CDI, Clade Disparity Index; D, statistical from the 2DKS; Br, Bremer support; SR, symmetric resampling expressed in absolute frequencies; CG, SR expressed in frequency differences. Significant results are in bold.

and/or states) in a data set. Wenzel and Siddall's results agree with our interpretation of the causes of the clade size effect, because similar to resampling measures of support, what they did was perturb matrices, although they did it using random signal.

This problem is not expected, at least not in the same degree, with the Bayesian posterior probabilities and Bremer support, because they do not depend on randomly modified matrices. Brandley et al. (2009) proposed something similar when mentioning that homoplasy could cause the inclusion of randomly resolved trees during the resampling, and that could be one of the reasons of the clade size effect. However, we must mention that some contriving examples could be generated to show that posterior probabilities are also susceptible to the clade size effect (Goloboff and Pol 2005; Pickett and Randle 2005).

Another hypothesized cause for the clade size effect was the taxon sampling strategy, of which the outgroup taxa (i.e. the larger nodes) are distant from the ingroup, while the smaller nodes group specimens of the same species (Brandley et al. 2006). This could concentrate the character signal in the terminal and in the most basal nodes. If this was the case, we would have expected to find the synapomorphies more concentrated in the larger and smaller clades, but our analyses did not find any support for this statement. Moreover, this particular taxon sampling is more common in phylogenies of living taxa, especially in molecular studies. The presence of a stronger CDI bias in morphological matrices does not agree with this hypothesis either. Related to this, one of the reviewers pointed out that in phylogenetic analyses that include fossils, the incompleteness of the fossil record could be causing smaller nodes to be more supported than mid-sized clades due to the accumulation of derivate states (see also Sidor and Hopson 1998). Although this need to be tested with a larger sample of matrices, the five paleontological data sets included in this analysis did not describe a different pattern from other morphological matrices with only extant taxa (see matrices Gaudin 2004, Gasparini et al. 2006, Leschen and Buckley 2007, O'Leary et al. 2004 and Wills et al. 1998 in Table 1).

Does the clade size effect really affect our daily analyses?

Brandley et al. (2009) stated that the poor performance of MP bootstrapping should concern the researchers, while researchers using likelihood-based methods should not worry. As said before, the results presented here using resampling measures of support are not different from the results obtained by Brandley et al. (2009) using ML bootstrapping. But going further on the analyses and despite what the statistical analyses might say to us, the actual impact of the clade size effect may not be so dramatic.

When plotting the clade support versus the clade size (raw number or the log_{10} of the clade prior probabilities), it becomes evident that the effect is not as strong as the CDI or the 2DKS test might suggest. Take as an example the data set of Leschen and Buckley (2007), which has a relatively high CDI with the SR (0.3, p = 0.0005) and a significant relationship between SR and clade size measured with the 2DKS ($D = 0.176$; p = 0.0002). Although the bivariate plot of the clade size versus the SR had a significant quadric fitting, the R^2 was only 0.42 (Table 3). A graphical analysis of the plot showed that the smallest nodes have a wide range of support values, and while the largest node has a higher branch support value than the middle-sized ones, the branch support value of this node is low (less than 40%; Fig. 2a). Although the bivariate plot of the clade prior probabilities versus the SR showed a tendency to have highest support values on the nodes with the highest priors, there is a wide dispersion of the data on that part of the graph (Fig. 2b), and again, the R^2 was low (0.43; Table 3). Another good exam-
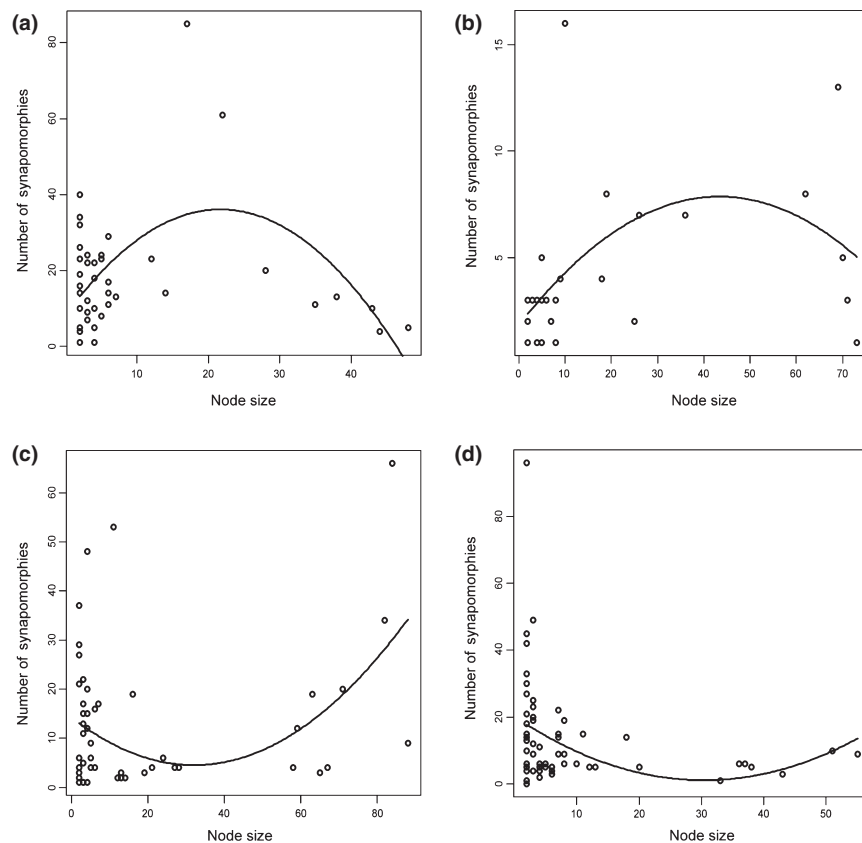
Figure 4. Distribution of the number of synapomorphies plotted against node size for the following data sets: (a) Dohrmann et al. (2008), (b) de Jong et al. (1996), (c) Chemisquy and Morrone (2012), (d) Brown et al. (2008)
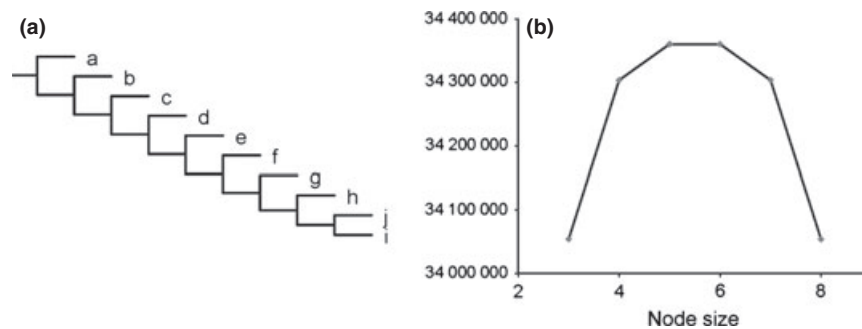

Figure 5. Chances of breaking a clade according to Equation 1. (a) Model tree, (b) possibility of breaking a clade plotted against node size

ple is the data set of Garvey et al. (1998) that has a significant CDI of 0.14 with the GC, and the plot of the GC versus the clade size is not much different from the plot of Berbee and Taylor (2001) that has a low, not significant CDI (0.08, p = 0.05; Fig. 2c and d). In this case, none of the plots showed a significant fitting (Table 3).

The implications of the results discussed above are clear. Although the clade size effect exists and it has a moderate influence on the resampling measures of support, the actual outcome of the effect is not as strong as expected according to the significance of statistical analyses. The bivariate plots of the clade support versus its size clearly show that small-sized clades have a wide range of support values, and the same is also true for other sizes of clades. This is also evident by the lack of fitting of the graph to a quadric or exponential function, where the best fits had a proportion of explained variance of 0.46 and 0.48, respectively, for the matrix of Lipscomb et al. (1989) (Fig. 2 e,f;

Table 3). We believe that we can still rely on the results of our maximum parsimony analyses and on the node support values obtained for our cladograms. But, if researchers are concerned by the clade size effect, they can use the Bremer support or the Bayesian posterior probabilities, which are almost not affected by it. In the case of concern of the bias in a particular node, it is possible to compare the values from resampling measures of support and Bremer support or Bayesian posterior probabilities. Another possibility is to analyse the number of uncontradicted synapomorphies of the node, at least in the case of maximum parsimony.

Conclusions

We corroborated the presence of the clade size effect in a small proportion of the matrices analysed (less than 25%). When branch support was calculated using Bremer support, the bias

Table 3. Quadric or exponential fitting of the branch support values to the node size or the prior clade probabilities respectively estimated using quantile regression

	GC		SR		Bremer	
Study	R^2 exponential	R^2 quadric	R^2 exponential	R^2 quadric	R^2 exponential	R^2 quadric
Aliscioni et al. (2003)	0.0071	0.0116	0.0078	0.0116	0.0096	0.0068
Anderson et al. (2004)	**0.2384**	**0.379**	**0.2397**	**0.379**	0.036	0.007
Armbruster (2004)	0.2117	0.2924	0.1585	0.2924	0.0036	0.0798
Baker et al. (2000)	0.1693	0.267	0.1891	0.267	0.0081	0.0034
Barkworth et al. (2008)	0.1252	0.111	0.1228	0.111	0.0231	0.0086
Barns et al. (1996)	0.0095	0.0142	0.0177	0.0142	0.0529	0.0066
Bell and Donoghue (2003)	0.0362	0.0409	0.0293	0.041	0.0151	0.0164
Berbee and Taylor (2001)	0.0581	0.071	0.0852	0.071	0.0054	0.0006
Whitten et al. (2007)	0.0365	0.0322	0.0412	0.0322	0.0443	0.0334
Brown et al. (2008)	**0.3576**	**0.3098**	**0.3105**	**0.3098**	0.1402	0.1143
Cabrero-Sañudo (2007)	**0.4118**	**0.4009**	**0.3984**	**0.401**	0.0336	0.0364
Jaramillo et al. (2008)		0.0364		0.0364		0.0014
Chemisquy and Morrone (2010)	0.0025	0.002	0.0034	0.0006	0.0977	0.125
Chemisquy and Morrone (2012)	0.0195	0.0024	0.0101	0.0036	0.0009	0.0498
Chemisquy et al. (2010)	0.0008	0.0108	0.0008	0.0044	0.0133	0.0053
de Jong et al. (1996)	0.1004	0.1086	0.1051	0.1255	0.1068	0.1067
Des Marais et al. (2003)	0.0613	0.1059	0.0628	0.0647	0.0506	0.049
Dohrmann et al. (2008)	0.0229	0.1602	0.0105	0.1614	0.0285	0.0396
Edwards et al. (2005)	0.2126	0.2699	0.1658	0.2062	0.0492	0.2383
Freitas and Brown (2004)	0.0429	0.0529	0.0581	0.0715	0.0282	0.0058
Garey et al. (1996)	0.184	0.1622	0.1837	0.1617	0.1637	0.1253
Gasparini et al. (2006)	**0.2925**	**0.2895**	**0.2486**	0.2281	0.0276	0.0034
Gaudin (2004)	0.1645	0.2361	0.1915	0.2681	0.000001	0.0226
Grant et al. (2006)	**0.1151**	**0.2301**	**0.2006**	**0.3351**	0.0344	0.0338
Ilves and Taylor (2008)	0.1583	0.0047	0.1892	0.000002	0.0193	0.199
James (2004)	0.0161	0.0843	0.0334	0.1509	0.000001	0.0073
Jordan et al. (2003)	0.013	0.0126	0.0167	0.0172	0.0137	0.0034
Kaila (2004)		**0.2698**		**0.3494**		0.003
Kelch and Baldwin (2003)	0.1224	0.1651	0.1937	0.2399	0.1142	0.122
Komarek and Beutel (2007)	0.0321	0.0383	0.0449	0.0398	0.000001	0.00001
Leschen and Buckley (2007)	**0.4520**	**0.4433**	**0.4299**	**0.4226**	0.0024	0.00001
Lindgren et al. (2004)	0.0308	0.0281	0.0006	0.00007	0.0661	0.0727
Lipscomb (1989)	**0.4705**	**0.4355**	**0.4838**	**0.456**	0.0151	0.0545
Mori et al. (2007)	**0.2266**	**0.2474**	0.2073	0.2246	**0.0184**	**0.0198**
Nihei and Barros de Carvalhlo (2007)	0.2015	0.1858	0.2036	0.1765	0.0001	0.0101
O'Leary et al. (2004)	0.0723	0.1065	0.0711	0.1381	0.0494	0.1401
O'Leary et al. (2009)	0.0054	0.0462	0.0069	0.0486	0.0384	0.0828
Pitts et al. (2006)	**0.2889**	**0.2986**	**0.3457**	**0.3514**	0.0006	0.061
Pramuk (2006)	0.2882	0.2817	0.3691	0.3757	**0.0428**	0.0767
Prevosti (2010)	0.0071	0.1514	0.0071	0.1417	0.0033	0.044
Reinert et al. (2004)	**0.2061**	**0.3256**	**0.3049**	**0.4278**	**0.0279**	0.0479
Rokas et al. (2003)	0.0786	0.4092	0.0241	0.4343	0.0241	0.0193
Shivonen (2005)	**0.2731**	**0.2866**	**0.3058**	**0.2997**	0.0131	0.1719
Sikes et al. (2008)	0.2067	0.1964	0.2345	0.234	0.1281	0.0183
Swenson and Anderberg (2005)	**0.3204**	**0.29**	**0.4224**	**0.39**	**0.1438**	0.12
Wanntorp et al. (2006)	0.0046	0.0124	0.0048	0.0329	0.0191	0.1343
Weisrock et al. (2006)	0.0626	0.0641	0.0778	0.082	0.0761	0.0583
Wills et al. (1998)	0.2552	0.2255	0.2614	0.2383	0.0021	0.089
Wu et al. (2001)	0.1877	0.1669	0.1669	0.1507	0.0285	0.0374
Yen et al. (2005)		**0.1933**		**0.2542**		0.0269

Br, Bremer support; SR, symmetric resampling expressed in absolute frequencies; CG, SR expressed in frequency differences.
Numbers in bold indicate statistically significant fittings.

was lower (fewer matrices with a statistical significant bias and lower values of CDI).

Our results using symmetric resampling are similar to the results previously found by Pickett and Randle (2005) and Brandley et al. (2009) using bootstrap and jackknife under maximum parsimony and maximum likelihood. The bias was significantly stronger for symmetric resampling than for Bayesian posterior probabilities, while Bremer support presented a better performance (although not statistically different from Bayesian posterior probabilities).

Contrary to what was proposed by Brandley et al. (2009), the problem is not on the method used for phylogenetic reconstruc- tion (i.e. maximum parsimony), but on the resampling measures of branch support. Homoplasy is in part responsible for the bias, because more homoplasy increases the number of random or near random trees obtained during the resampling of the data sets. If we add the higher chance that medium-sized clades have of being contradicted, the clade size effect appears, affecting medium-sized clades during the perturbation of the original matrix for the calculation of the branch support. This explains why resampling measures of support have considerably higher values of CDI, as well as more matrices with statistically significant CDIs than Bremer support and Bayesian posterior probabilities.

Finally, qualitative analyses showed that although the effect does exist, it is not as strong as expected by the results of the statistical analyses. The plotting of the branch support values versus the node size (and clade prior probabilities) revealed that the fit of the data to the expected pattern was not good, implying that although the statistical analyses show that the effect exists, the intensity of the effect is low (or moderate, in a few cases). Consequently, we can still rely on the results of our branch support analyses, even the resampling ones.

Acknowledgements

References

Aliscioni SS, Giussani LM, Zuloaga FO, Kellogg EA (2003) A molecular phylogeny of *Panicum* (Poaceae: Paniceae): Tests of monophyly and phylogenetic placement within the Panicoideae. Am J Bot **90**:796–821.

Anderson FE, Córdoba AJ, Thollesson M (2004) Bilaterian phylogeny based on analyses of a region of the sodium–potassium atpase β–subunit gene. J Mol Evol **58**:252–268.

Archie JW (1989) Homoplasy excess ratios: new indices for measuring levels of homoplasy in phylogenetic systematics and a critique of the consistency index. Syst Zool **38**:253–269.

Archie JW (1996) Measures of homoplasy. In: Sanderson MJ, Hufford L (eds), Homoplasy: The Recurrence of Similarity in Evolution. Academic Press, New York, pp 153–206.

Armbruster JW (2004) Phylogenetic relationships of the suckermouth armoured catfishes (Loricariidae) with emphasis on the Hypostominae and the Ancistrinae. Zool J Linn Soc **141**:1–80.

Baker WJ, Dransfield J, Hedderson TA (2000) Phylogeny, character evolution, and a new classification of the Calamoid Palms. Syst Bot **25**:297–322.

Barkworth ME, Arriaga MO, Smith JF, Jacobs SWL, Valdés–Reyna J, Bushman BS (2008) Molecules and morphology in South American Stipeae (Poaceae). Syst Bot **33**:719–731.

Barns SM, Delwiche SF, Palmer JD, Pace NR (1996) Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. Proc Natl Acad Sci **93**:9188–9193.

Bell CD, Donoghue MJ (2003) Phylogeny and biogeography of Morinaceae (Dipsacales) based on nuclear and chloroplast DNA sequences. Org Divers Evol **3**:227–237.

Berbee ML, Taylor JW (2001) Fungal molecular evolution: gene trees and geologic time. In: McLaughlin DJ, Mc–Laughlin, Lemke PA (eds), Mycota–Systematics and Evolution VII. Springer–Verlag, New York, pp. 229–245.

Brandley MC, Leaché AD, Warren DL, Mcguire JA (2006) Are unequal priors problematic for bayesian phylogenetics? Syst Biol **55**:138–146.

Brandley MC, Warren DL, Leaché AD, Mcguire JA (2009) Homoplasy and clade support. Syst Biol **58**:184–198.

Bremer K (1994) Branch support and tree stability. Cladistics **10**:295–304.

Brochu CA (1999) Taxon sampling and reverse successive weighting. Syst Biol **48**:808–813.

Brown GK, Murphy DJ, Miller JT, Ladiges PY (2008) Acacia s.s. and its relationship among tropical legumes, tribe Ingeae (Leguminosae: Mimosoideae). Syst Bot **33**:739–751.

Cabrero–Sañudo FJ (2007) The phylogeny of Iberian Aphodiini species (Coleoptera, Scarabaeoidea, Scarabaeidae, Aphodiinae) based on morphology. Syst Entomol **32**:156–175.

Cade BS, Noon BR (2003) A gentle introduction to quantile regression for ecologists. Front Ecol Environ **1**:412–420.

Chemisquy MA, Morrone O (2010) Phylogenetic analysis of the subtribe Chloraeinae (Orchidaceae): a preliminary approach based on three chloroplast markers. Aust Syst Bot **23**:38–46.

Chemisquy MA, Morrone O (2012) Molecular phylogeny of *Gavilea* (Chloraeinae: Orchidaceae) using plastid and nuclear markers. Mol Phylogenet Evol **62**:889–897.

Chemisquy MA, Giussani LM, Scataglini MA, Kellogg EA, Morrone O (2010) Phylogenetic studies favour the unification of *Pennisetum*, *Cenchrus* and *Odontelytrum* (Poaceae): a combined nuclear, plastid and morphological analysis, and nomenclatural combinations in *Cenchrus*. Ann Bot **106**:107–130.

Davis JI, Simmons MP, Stevenson DW, Wendel JF (1998) Data decisiveness, data quality, and incongruence in phylogenetic analysis: an example from the Monocotyledons using mitochondrial *atpA* sequences. Syst Biol **47**:282–310.

de Jong R, Vane–Wright RI, Ackery PR (1996) The higher classification of butterflies (Lepidoptera): problems and prospects. Entomol Scand **27**:65–101.

Des Marais DL, Smith AR, Britton DM, Pryer KM (2003) Phylogenetic relationships and evolution of extant horsetails, *Equisetum*, based on chloroplast DNA sequence data (*rbcL* and *trnL-F*). Int J Plant Sci **164**:737–751.

Dohrmann M, Janussen D, Reitner J, Collins AG, Wörheide G (2008) Phylogeny and evolution of glass sponges (Porifera, Hexactinellida). Syst Biol **57**:388–405.

Edwards EJ, Nyffeler R, Donoghue MJ (2005) Basal cactus phylogeny: implications of *Pereskia* (Cactaceae) paraphyly for the transition to the cactus life form. Am J Bot **92**:1177–1188.

Farris JS, Albert VA, Källersjö M, Lipscomb D, Kluge AG (1996) Parsimony jackknifing outperforms neighbor-joining. Cladistics **12**:1199–1201.

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39**:783–791.

Freitas AVL, Brown KS Jr (2004) Phylogeny of the Nymphalidae (Lepidoptera). Syst Biol **53**:363–383.

García LV (2004) Escaping the Bonferroni iron claw in ecological studies. Oikos **105**:657–663.

Garey JR, Near TJ, Nonnemacher MR, Nadler SA (1996) Molecular evidence for Acanthocephala as a subtaxon of Rotifera. J Mol Evol **43**:287–292.

Garvey JE, Marschall EA, Wright RA (1998) From star charts to stoneflies: detecting relationships in continuous bivariate data. Ecology **79**:442–447.

Gasparini Z, Pol D, Spalletti LA (2006) An unusual marine Crocodyliform from the Jurassic–Cretaceous boundary of Patagonia. Science **311**:70–73.

Gaudin TJ (2004) Phylogenetic relationships among sloths (Mammalia, Xenarthra, Tardigrada): the craniodental evidence. Zool J Linn Soc **140**:255–305.

Givnish TJ, Sytsma KJ (1997a) Consistency, characters, and the likelihood of correct phylogenetic inference. Mol Phylog Evol **7**:320–330.

Givnish TJ, Sytsma KJ (1997b) Homoplasy in molecular vs. morphological data: the likelihood of correct phylogenetic inference. In: Givnish TJ, Sytsma KJ (eds), Molecular evolution and adaptive radiation. Cambridge University Press, New York, pp 55–101.

Goloboff PA (1991a) Homoplasy and the choice among cladograms. Cladistics **7**:215–232.Goloboff PA (1991b) Random data, homoplasy and information. Cladistics **7**:395–406.

Goloboff PA (1999) Analyzing large data sets in reasonable times: solutions for composite optima. Cladistics **15**:415–428.

Goloboff PA, Pol D (2005) Parsimony and Bayesian phylogenetics. In: Albert V (ed.), Parsimony, Phylogeny, and Genomics. Oxford University Press Inc., New York, pp 148–217.

Goloboff PA, Farris JS, Källersjö M, Oxelman B, Ramírez MJ, Szumik CA (2003) Improvements to resampling measures of group support. Cladistics **19**:324–332.

Goloboff PA, Farris JS, Nixon KC (2008) TNT, a free program for phylogenetic analysis. Cladistics **24**:774–786.

Grandcolas P (2004) Abstracts of the 23rd Annual Meeting of the Willi Hennig Society. "Phylogenetics and Evolutionary Biology". Cladistics **20**:583–608.

Grant T, Frost DR, Caldwell JP, Gagliardo R, Haddad CFB, Kok PJR, Means DB, Noonan BP, Schargel WE, Wheeler W (2006) Phylogenetic systematics of dart–poison frogs and their relatives (Amphibia, Athesphatanura, Dendrobatidae). Bull Am Mus Nat Hist 299:1–262.

Huelsenbeck JP, Ronquist F (2005) Bayesian analysis of molecular evolution using MrBayes. In: Nielsen R (ed.), Statistical Methods in Molecular Evolution. Springer, New York, pp 2–52.

Ilves KL, Taylor EB (2008) Evolutionary and biogeographical patterns within the smelt genus *Hypomesus* in the North Pacific Ocean. J Biogeogr 35:48–64.

James HF (2004) The osteology and phylogeny of the Hawaiian finch radiation (Fringillidae: Drepanidini), including extinct taxa. Zool J Linn Soc 141:207–255.

Jaramillo MA, Callejas R, Davidson C, Smith JF, Stevens AC, Tepe EJ (2008) A phylogeny of the tropical genus *Piper* using its and the chloroplast intron *psbJ–petA*. Syst Bot 33:647–660.

Jordan S, Simon C, Polhemus D (2003) Molecular systematics and adaptive radiation of Hawaii's endemic damselfly genus *Megalagrion* (Odonata: Coenagrionidae). Syst Biol 52:89–109.

Kaila L (2004) Phylogeny of the superfamily Gelechioidea (Lepidoptera: Ditrysia): an exemplar approach. Cladistics 20:303–340.

Kelch DG, Baldwin BG (2003) Phylogeny and ecological radiation of New World thistles (Cirsium, Cardueae – Compositae) based on ITS and ETS rDNA sequence data. Mol Ecol 12:141–151.

Kluge AG, Farris JS (1969) Quantitative phyletics and the evolution of anurans. Syst Zool 18:1–32.

Komarek A, Beutel RG (2007) Phylogenetic analysis of Anacaenini (Coleoptera: Hydrophilidae: Hydrophilinae) based on morphological characters of adults. Syst Entomol 32:205–226.

Leschen RAB, Buckley TR (2007) Multistate characters and diet shifts: evolution of Erotylidae (Coleoptera). Syst Biol 56:97–112.

Lindgren AR, Giribet G, Nishiguchi MK (2004) A combined approach to the phylogeny of Cephalopoda (Mollusca). Cladistics 20:454–486.

Lipscomb DL (1989) The relationships among the eukaryotes. In: Fernholm B, Bremer K, Jornvall H (eds), The Hierarchy of Life. Elsevier Science Publishers B.V. (Biomedical Division), Amsterdam, pp 161–178.

Mori SA, Tsou CH, Wu CC, Cronholm B, Anderberg AA (2007) Evolution of Lecythidaceae with an emphasis on the circumscription of neotropical genera: information from combined *ndhF* and *trnL–F* sequence data. Am J Bot 94:289–301.

Nihei SS, De Carvalho CJB (2007) Phylogeny and classification of Muscini (Diptera, Muscidae). Zool J Linn Soc 149:493–532.

O'Leary MA, Allard M, Novacek MJ, Meng J, Gatesy J (2004) Building the Mammalian sector of the Tree of Life: combining different data and a discussion of divergence times for Placental Mammals. In: Cracraft J, Donoghue MJ (eds), Assembling the Tree of Life. Oxford University Press, New York, pp 490–516.

O'Leary N, Yuan YW, Chemisquy MA, Olmstead RG (2009) Reassignment of species of paraphyletic *Junellia s.* l. to the new genus *Mulguraea* (Verbenaceae) and new circumscription of genus *Junellia*: molecular and morphological congruence. Syst Bot 34:777–786.

Pickett KM, Randle CP (2005) Strange Bayes indeed: uniform topological priors imply non-uniform clade priors. Mol Phylogenet Evol 34:203–211.

Pitts JP, Wasbauer MS, Von Dohlen CD (2006) Preliminary morphological analysis of relationships between the spider wasp subfamilies (Hymenoptera: Pompilidae): revisiting an old problem. Zool Scr 35:63–84.

Pramuk JB (2006) Phylogeny of South American Bufo (Anura: Bufonidae) inferred from combined evidence. Zool J Linn Soc 146:407–452.

Prevosti FJ (2010) Phylogeny of the large extinct South American Canids (Mammalia, Carnivora, Canidae) using a "total evidence" approach. Cladistics 26:456–481.

Prevosti FJ, Chemisquy MA (2009) The impact of missing data on real morphological phylogenies: influence of the number and distribution of missing entries. Cladistics 25:1–14.

Randle CP, Pickett KM (2006) Are nonuniform clade priors important in Bayesian Phylogenetic Analysis? A response to Brandley et al. Syst Biol 55:147–151.

Reinert JF, Harbach RE, Kitching IJ (2004) Phylogeny and classification of Aedini (Diptera: Culicidae), based on morphological characters of all life stages. Zool J Linn Soc 142:289–368.

Rokas A, Melika G, Abe Y, Nieves–Aldrey JL, Cook JM, Stone GN (2003) Lifecycle closure, lineage sorting, and hybridization revealed in a phylogenetic analysis of European oak gallwasps (Hymenoptera: Cynipidae: Cynipini) using mitochondrial sequence data. Mol Phylogenet Evol 26:36–45.

Sanderson MJ, Donoghue MJ (1996) The relationship between homoplasy and confidence in phylogenetic trees. In: Sanderson MJ, Hufford L (eds), Homoplasy: The Recurrence of Similarity in Evolution. Academic Press, New York, pp 67–89.

Sidor CA, Hopson JA (1998) Ghost lineages and "mammalness": assessing the temporal pattern of character acquisition in the Synapsida. Paleobiology 24:254–273.

Sihvonen P (2005) Phylogeny and classification of the Scopulini moths (Lepidoptera: Geometridae, Sterrhinae). Zool J Linn Soc 143:473–530.

Sikes DS, Vamosi DM, Trumbo ST, Ricketts M, Venables C (2008) Molecular systematics and biogeography of *Nicrophorus* in part – the investigator species group (Coleoptera: Silphidae) using mixture model MCMC. Mol Phylogenet Evol 48:646–666.

Simmons MP, Freudenstein JV (2011) Spurious 99% bootstrap and jackknife support for unsupported clades. Mol Phylogenet Evol 61:177–191.

Swenson U, Anderberg AA (2005) Phylogeny, character evolution, and classification of Sapotaceae (Ericales). Cladistics 21:101–130.

Wanntorp L, Kocyan A, van Donkelaar R, Renner SS (2006) Towards a monophyletic *Hoya* (Marsdenieae, Apocynaceae): inferences from the chloroplast *trnL* region and the *rbcL–atpB* spacer. Syst Bot 31:586–596.

Weisrock DW, Papenfuss TJ, Macey JR, Litvinchuk SN, Polymeni R, Ugurtas IH, Zhao E, Jowkar H, Larson A (2006) A molecular assessment of phylogenetic relationships and lineage accumulation rates within the family Salamandridae (Amphibia, Caudata). Mol Phylogenet Evol 41:368–383.

Wenzel JW, Siddall ME (1999) Noise. Cladistics 64:51–64.

Whitten WM, Blanco MA, Williams NH, Koehler S, Carnevali G, Singer RB, Endara L, Neubig KM (2007) Molecular phylogenetics of *Maxillaria* and related genera (Orchidaceae: Cymbidieae) based on combined molecular data sets. Am J Bot 94:1860–1889.

Wills MA, Briggs DEG, Fortey RA, Wilkinson M, Sneath PHA (1998) Arthropod relationships. In: Edgecombe GD (ed.), An Arthropod Phylogeny Based on Fossil and Recent Taxa. Columbia University Press, New York, pp 33–105.

Wu SH, Hibbett DS, Binder M (2001) Phylogenetic analyses of *Aleurodiscus* s.l. and allied genera. Mycologia 93:720–731.

Yen SH, Robinson GS, Quicke DLJ (2005) The phylogenetic relationships of Chalcosiinae (Lepidoptera, Zygaenoidea, Zygaenidae). Zool J Linn Soc 143:161–341.

Zar JH (1984) Biostatistical Analysis. Prentice-Hall, Englewood Cliffs, NJ.

Appendix 1. Characteristics of the data sets used in the analyses

Study		nTax	nChar	CI	Imbalance	Br	Mean support		nNodes	DD
							GC	SR		
1	Aliscioni et al. (2003)	123	2062	0.2453	0.557	3.77	74.3	79.52	90	0.749
2	Anderson et al. (2004)	63	2943	0.2129	0.6	14.35	55.73	59.93	57	0.293
3	Armbruster (2004)	128	215	0.1954	0.547	3.78	56	62.22	107	0.707
4	Baker et al. (2000)	31	66	0.4551	0.722	1.95	42.42	51.05	19	0.651
5	Barkworth et al. (2008)	103	2479	0.569	0.653	1.366	56.19	66.95	62	0.8
6	Barns et al. (1996)	64	1620	0.3458	0.681	14.54	55.9	63.72	50	0.677
7	Bell and Donoghue (2003)	23	3025	0.7068	0.8	20.4	85.25	90.4	20	0.648
8	Berbee and Taylor (2001)	52	1531	0.3584	0.708	8.94	54.21	69.58	48	0.437
9	Whitten et al. (2007)	88	7214	0.5576	0.554	6.47	69.99	77.25	73	0.638
10	Brown et al. (2008)	64	1369	0.412	0.639	8.3	48.79	62.33	61	0.595
11	Cabrero-Sañudo (2007)	100	94	0.5178	0.583	2.58	43.78	54.61	46	0.805
12	Jaramillo et al. (2008)	181	1715	0.4135	0.43	10.51	73.28	81.23	108	0.73
13	Chemisquy and Morrone (2010)	28	1762	0.7298	0.526	8.77	67.33	82.95	20	0.666
14	Chemisquy and Morrone (2012)	92	824	0.5842	0.643	8.96	80.11	85	53	0.854
15	Chemisquy et al. (2010)	131	818	0.6146	0.449	4.26	60.04	69.59	69	0.815
16	de Jong et al. (1996)	75	103	0.2692	0.7	2.34	38.65	52.27	29	0.667
17	Des Marais et al. (2003)	22	2342	0.7818	0.789	10.37	76.58	84.05	19	0.738
18	Dohrmann et al. (2008)	51	3435	0.5631	0.533	9.35	80.39	86.41	46	0.76
19	Edwards et al. (2005)	38	6150	0.7946	0.552	12.35	85.32	88.19	31	0.774
20	Freitas and Brown (2004)	96	234	0.2242	0.531	3.28	65.13	72.51	61	0.647
21	Garey et al. (1996)	55	1934	0.3365	0.659	18.51	71.38	88.36	47	0.457
22	Gasparini et al. (2006)	59	257	0.3664	0.643	1.34	45.04	53.8	56	0.66
23	Gaudin (2004)	46	201	0.3261	0.775	3.49	45.12	58	41	0.578
24	Grant et al. (2006)	122	174	0.2032	0.604	1.4	22.49	38.1	47	0.523
25	Ilves and Taylor (2008)	15	2732	0.9282	0.571	13.56	93	93.44	9	0.925
26	James (2004)	89	84	0.276	0.5	1.38	40.74	49.06	34	0.69
27	Jordan et al. (2003)	68	2326	0.514	0.585	14.98	82.66	87.52	57	0.731
28	Kaila (2004)	156	193	0.1591	0.632	3.33	43.07	55.16	101	0.635
29	Kelch and Baldwin (2003)	52	1303	0.6736	0.125	7.67	87.67	90.33	12	0.632
30	Komarek and Beutel (2007)	56	81	0.3011	0.7	3.17	42.39	52.87	23	0.624
31	Leschen and Buckley (2007)	54	121	0.2078	0.686	1.57	14.75	29.92	51	0.436
32	Lindgren and Buckley (2004)	78	101	0.574	0.4286	1.93	72.18	75.14	28	0.899
33	Lipscomb (1989)	86	138	0.7565	0.488	1.92	41.36	52.24	50	0.732
34	Mori (2007)	111	3221	0.7011	0.493	6.57	48.68	82.72	82	0.831
35	Nihei and Barros de Carvalho (2007)	88	112	0.227	0.469	2.91	51.18	61.36	44	0.683
36	O'Leary et al. (2004)	68	184	0.2148	0.583	2.97	49.9	60.37	30	0.507
37	O'Leary et al. (2009)	70	1164	0.4875	0.738	6.69	56.87	6.64	62	0.761
38	Pitts et al. (2006)	84	77	0.3081	0.585	1.7	37.03	48.39	56	0.696
39	Pramuk (2006)	64	83	0.2348	0.667	1.85	62.5	69.9	20	0.609
40	Prevosti (2010)	25	1249	0.4177	0.824	18	74.55	83	16	0.476
41	Reinert et al. (2004)	123	172	0.1214	0.43	1.85	69.9	68.5	85	0.583
42	Rokas et al. (2003)	31	433	0.68	0	3.28	85.43	87.28	7	0.818
43	Shivonen (2005)	92	141	0.2664	0.571	3.49	45.12	58	74	0.554
44	Sikes et al. (2008)	50	2129	0.4868	0.625	17.93	74.81	79.26	42	0.778
45	Swenson and Anderberg (2005)	99	78	0.1793	0.46	2.43	18.98	30.52	53	0.558
46	Wanntorp et al. (2006)	44	1787	0.7341	0.429	1.94	48	56.44	18	0.643
47	Weisrock et al. (2006)	96	2765	0.2198	0.598	30.52	78.98	85.31	88	0.588

Appendix 1. (continued)

Study	nTax	nChar	CI	Imbalance	Mean support			nNodes	DD
					Br	GC	SR		
48 Wills et al. (1998)	64	97	0.3065	0.417	1.8	48.47	56.6	15	0.619
49 Wu et al. (2001)	71	997	0.5261	0.563	2.08	43.6	52.11	52	0.684
50 Yen et al. (2005)	207	414	0.2234	0.601	4.51	57.58	68.9	186	0.704

nTax, number of taxa; nChar, number of characters; CI, Consistency Index; nNodes, number of nodes of the strict consensus; DD, data decisiveness; Br, Bremer support; SR, symmetric resampling expressed in absolute frequencies; CG, SR expressed in frequency differences.
Molecular data sets are in bold.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Quadric function regressions of the number of synapomorphies on the node size. None of the matrices was statistically significant after the Bonferroni adjustment.