THE RELATIONSHIP BETWEEN HOMOPLASY AND CONFIDENCE IN A PHYLOGENETIC TREE

MICHAEL J. SANDERSON* AND MICHAEL J. DONOGHUE†

**University of California*
Davis, California

†*Harvard University*
Cambridge, Massachusetts

INTRODUCTION

When the term "homoplasy" was coined by Lankester (1870), the evolutionary process was central:

> When identical or nearly similar forces, or environments, act on . . . parts in two organisms, which parts are exactly or nearly alike and sometimes homogenetic [homologous], the resulting correspondences called forth in the several parts in the two organisms will be nearly or exactly alike . . . I propose to call this form of agreement *homoplasis* or *homoplasy*. [p. 39]

However, the phylogenetic implications of this process were not lost on him:

> Zoology has for some time been embarrassed with the reference of all segmented Invertebrata to a common type, and the supposed homology of their segmented structures. This difficulty may, it is suggested, be possibly solved by the admission of true zooid-segmentation as being frequently due to homoplasy, and not by any means necessarily an indication of genetic affinity. [p. 43]

HOMOPLASY: The Recurrence of Similarity in Evolution, M. J. Sanderson and L. Hufford, eds.
Copyright © 1996 All rights of reproduction in any form reserved.

To many phylogeneticists homoplasy is now viewed only as "error in homology assessment." Although this restricted definition tends to ignore the profound impact that homoplasy has on the generation of organismal diversity, the emphasis is appropriate when accurate phylogeny reconstruction is the central concern. Any source of error must be characterized, and, if possible, ameliorated. Surprisingly, little attention has been paid to a subtle but basic aspect of this "error": even if homoplasy is a mistaken hypothesis of homology for an individual character, does it generally impede efforts to reconstruct phylogenies using sets of many independent characters?

Because it is easy to imagine simple cases, such as a single character, in which homoplasy leads to a mistaken idea of relationships, it is tempting to generalize and equate *overall* levels of homoplasy with confidence in a tree. Such generalizations have been common. For many years, levels of homoplasy, as encapsulated in the consistency index (Kluge and Farris 1969), were routinely reported in phylogenetic studies, often accompanied by ad hoc explanations for low values (reviewed in Sanderson and Donoghue 1989; Donoghue and Sanderson 1992). More recently, surveys of consistency indices have served as the basis for suggestions that high CI's support the superiority of one kind of data over another. For example, Givnish and Sytsma (1992) and Jansen (1995) have both suggested that chloroplast restriction site data are better than sequence data for plant phylogenetics based at least in part on relative levels of homoplasy (Jansen also considered several other variables in addition to homoplasy). Clearly, many investigators view the level of homoplasy as a valuable indicator of the quality of the phylogenetic conclusions derived based on a data set.

Another reason that some workers have assumed a close relationship between homoplasy and confidence is the close relationship between rates of evolution and confidence. Analytical (Felsenstein 1978) and simulation studies (see Archie 1996 for a review) have generally agreed that it is more difficult to estimate trees accurately when rates of evolution are high, and rates of evolution ought generally to be correlated with levels of homoplasy. However, few studies have considered the distribution of homoplasy as a factor independent of rate (although see Landrum 1993; Archie 1996), and it is possible that rates, homoplasy, and confidence are intercorrelated in a way that masks the underlying relationships among these variables.

Critical analyses of the consistency index (Sanderson and Donoghue 1989; Archie 1989; Klassen et al. 1991; Meier et al. 1991) have prompted some reevaluation of the difference between homoplasy and "informativeness" and perhaps "accuracy" (Goloboff 1991a,b). In this paper we examine such issues further by considering the relationship, if any, between levels of homoplasy in data sets and confidence in the phylogenetic tree(s) estimated from those data. We begin where our previous study of levels of homoplasy (Sanderson and Donoghue, 1989) left off—with an examination of homoplasy in relation to an independent measure of the robustness of a tree.

Some authors have examined robustness in terms of the accuracy or "nearness" of the estimated tree to the true tree in simulation studies in which the true tree is known. Under those conditions the correspondence between the estimated tree and the true tree, as measured by a consensus index, for example, is obviously a reasonable indication of accuracy. However, we propose to examine a large set of real (not simulated) phylogenetic data sets in which the true tree is unknown. We hope that the disadvantages of an unknown underlying tree are offset by the generality entailed by the diversity of evolutionary processes likely to be uncovered in any large sample of real data sets.

Because we do not know the true tree in any real data set, we focus on "confidence," rather than accuracy. The bootstrap procedure (Felsenstein 1985; Sanderson 1989, 1995; Hillis and Bull 1993) is the most widely used statistical assessment of confidence in a phylogenetic tree. It aims to provide an indication of the size of the neighborhood of trees that are similar to the true tree. Specifically, the bootstrap attempts to estimate the effects of sampling error on phylogenetic inference. It aims to predict the probability that further sampling of characters will support clades discovered in the initial analysis. It is therefore highly dependent on the number of synapomorphies of a clade, for example—the idea being that a clade in which many synapomorphies have already been discovered is probably a clade in which there are also more undiscovered synapomorphies. The bootstrap has been criticized for many reasons (reviewed in Sanderson 1995)—some philosophical, some statistical, and some empirical (e.g., Cummings et al. 1995)—but its utility, or at least the perception of its utility, remains high. Some nonstatistical measures, such as the decay index (Bremer 1988; Donoghue et al. 1992) are of no utility for comparative purposes, because their magnitudes are data set dependent.

In principle, the bootstrap is decoupled from homoplasy. A data set with just one informative character (but several uninformative ones) will automatically have no apparent homoplasy (CI = 1.0) but its bootstrap value will be very low because the probability of future sampling overturning the relationship suggested by any single character is high. Conversely, it is not difficult to construct data matrices that have fairly low CIs but have high bootstrap values, if the homoplasy is dispersed randomly around a tree and is not correlated across characters. See Sanderson and Donoghue (1989: 1789) for an example.

We use the consistency index (CI: Kluge and Farris 1969) as a measure of homoplasy. It and the retention index (RI: Farris 1989) (see Archie 1996) are the most commonly cited statistics reported in phylogenetic studies. Both measure aspects of the goodness of fit of a data set to a hierarchical tree structure. However, CI is a better measure of overall homoplasy, because RI factors in the distribution of apomorphies among taxa. RI therefore reflects something in addition to homoplasy which is related to the probability that

the data could actually be expected to exhibit a certain level of homoplasy given the distribution of character states among taxa. The difference can be seen in a simple example. If 2 of 10 taxa have the apomorphic (derived) state for a binary character, CI = 0.50 if the states are parallelisms, but RI = 0.00. If 4 of 10 taxa have the derived state, then CI = 0.25 but still RI = 0.00 when all four originations are independent. Clearly these two cases represent very different levels of homoplasy, but their retention indices are the same. Goloboff (1991b) suggests that CI provides more of an indication about homoplasy than it does about "informativeness" about relationships, implying that the two issues are logically separable.

As in our previous analysis (Sanderson and Donoghue 1989), we consider a set of phylogenetic analyses of real taxa taken from the primary literature. We use observational statistics to make inferences about the relationships between levels of homoplasy and bootstrap estimates of confidence. Each "observation" therefore represents an individual study sampled from a "population" of studies. This approach is not quite meta-analysis (Arnqvist and Wooster 1995). Meta-analysis, which is widely used to extract patterns from multiple studies, tests a hypothesis across a set of studies that is identical to a hypothesis tested (possibly by different methods) in each study. In the present analysis there is no "hypothesis" common to all studies. Instead we seek information about the relationships of variables in the population of studies. For example, is data set size correlated with robustness of the resulting trees?

However, several problems in true meta-analyses could be problematic here. Sometimes meta-analysis is applied to studies so different from each other that the only thing they share in common is a significance level (the "apples and oranges problem"). Phylogenetic analyses, however, share a fundamentally similar methodology. They use a data matrix of taxa by characters. In our study the characters are all discrete. Continuous characters and distance date are excluded. Phylogenetic studies generally employ some optimization strategy to estimate a phylogenetic tree. In our study maximum parsimony is the only algorithm used. Moreover, because the same algorithms and software are used, precisely the same set of statistics are gathered on each study.

Another issue in meta-analysis is the "file-drawer" problem, a bias introduced because studies that fail to find significance are rarely reported in the literature (they are relegated to the file drawer). However, perhaps because of the systematics community's collective uncertainty about how to assign significance to phylogenetic trees, systematists seem unlikely to fail to publish because of lack of significance. Only one journal that publishes many phylogenetic studies, *Molecular Biology and Evolution*, encourages as part of its editorial policy the reporting of significance levels (although reviewers of other journals certainly may encourage it), and we suspect that a significant

number of systematists regards the reporting of bootstrap values as positively misguided (e.g., Wendel and Albert 1992).

MATERIALS AND METHODS

A set of 101 phylogenetic studies (Table 1) was extracted from the TreeBASE data base (WWW URL http: //phylogeny.harvard.edu/treebase). A summary of the characteristics of these data sets is provided in Table 2. They included 50 morphological, 29 RFLP, and 22 DNA sequence data sets, ranging in size from 5 to 68 taxa and from 10 to 2226 characters. All studies were on green plants.

Each data set was subjected to parsimony analysis using PAUP 3.1 (Swofford 1993). Heuristic search options consisted of simple addition sequence ("hold" set to 5) and "TBR" branch swapping with "maxtrees" set to 250. The following tree statistics were recorded for each run: consistency index (CI), retention index (RI), number of most parsimonious trees (Ntrees), and degree of resolution of one of the most parsimonious trees (arbitrarily, the first one, if more than one was found). Degree of resolution was measured by the normalized consensus fork index, which reports the fraction of possible clades that actually occur on a tree of given size (Swofford 1991). All runs were performed on the same computer.

Each data set was then bootstrapped (100 replicates) using the same set of heuristic search options. The list of clades appearing among the bootstrap replicates was saved and later examined by a UNIX shell script to summarize findings. For each complete bootstrap analysis, two statistics were extracted: (1) the maximum bootstrap level observed in the tree (B_{max}) and (2) the fraction of clades supported at the level of greater than 50% (B_{50}). Note that B_{50} is really the consensus fork index of the majority rule consensus tree (Swofford 1991). These two statistics convey different aspects of overall support for a tree. Neither of them is ideal, and other measures could be used, such as the average bootstrap value computed over all clades in the strict consensus tree (Jansen and Wee, personal communication), or Bremer's (1994) total support index. Of the two measures reported here, the B_{50} value should give a better overall summary of support but is influenced by the average degree of resolution possible with the given set of characters (i.e., low resolution in the most parsimonious trees probably will also lead to lower values of B_{50}. The B_{max} statistic, on the other hand, should be less sensitive to resolution and the distribution of character support among clades, but cannot discriminate among overall support when it is fairly high. There are many data sets for which B_{max} is 1.0.

Statistical analyses were performed using JMP version 3 (SAS Institute, Inc.). A matrix of partial correlation coefficients among all the variables is

TABLE 1 Raw Data from an Analysis of 101 Phylogenetic Data Sets[a]

Ref.	Data	No. characters	No. taxa	CI	RI	No. trees	Resolution (CFI)[b]	B_{50}	B_{max}
1	Morph	32	40	0.415	0.721	36	0.919	0.205	1.000
2	Morph	28	14	0.783	0.811	5	1.000	0.308	1.000
3	Morph	46	44	0.459	0.776	250	0.927	0.279	1.000
4	Morph	23	14	0.644	0.724	1	0.909	0.462	0.818
5	Morph	29	23	0.433	0.600	4	0.950	0.091	0.673
6	Morph	10	7	0.833	0.818	6	0.750	0.333	0.943
7	Morph	40	15	0.742	0.734	250	0.833	0.071	0.948
8	Morph	22	13	0.931	0.875	1	0.800	0.250	0.993
9	Morph	22	15	0.750	0.843	18	0.917	0.500	0.815
10	Morph	23	11	0.711	0.780	7	1.000	0.500	0.980
11	Morph	24	21	0.491	0.746	155	0.889	0.200	0.700
12	Morph	45	30	0.741	0.851	117	0.963	0.586	1.000
13	Morph	20	22	0.839	0.896	6	0.632	0.286	0.972
14	Morph	15	6	0.789	0.750	2	0.667	0.400	0.910
15	Morph	28	14	0.481	0.604	18	0.909	0.308	0.674
16	Morph	34	20	0.600	0.730	8	1.000	0.316	0.718
17	Morph	17	5	0.909	0.714	1	1.000	0.500	0.830
18	Morph	29	20	0.688	0.826	2	0.941	0.632	0.928
19	Morph	36	11	0.873	0.850	2	1.000	0.700	0.958
20	Morph	13	18	0.542	0.780	24	0.533	0.176	0.890
21	Morph	49	18	0.545	0.658	93	1.000	0.294	0.982
22	Morph	63	13	0.585	0.700	18	1.000	0.583	0.990
23	Morph	15	11	1.000	1.000	15	0.750	0.300	0.688
24	Morph	21	9	0.846	0.875	1	1.000	0.750	0.936
25	Morph	46	23	0.465	0.663	55	1.000	0.227	0.720
26	Morph	25	11	0.617	0.647	1	1.000	0.500	0.989
27	Morph	48	19	0.552	0.566	1	1.000	0.333	0.980
28	Morph	17	19	0.774	0.868	35	0.688	0.500	0.960
29	Morph	25	9	0.750	0.714	4	1.000	0.500	0.970
30	Morph	70	14	0.507	0.538	1	1.000	0.692	0.957
31	Morph	42	17	0.769	0.829	66	0.714	0.375	0.966
32	Morph	88	56	0.398	0.663	250	0.906	0.182	0.870
33	Morph	29	12	0.938	0.944	26	0.778	0.545	0.990
34	Morph	28	11	0.630	0.595	11	1.000	0.200	0.888
35	Morph	39	8	0.565	0.565	2	1.000	0.571	0.901
36	Morph	56	40	0.679	0.892	18	0.919	0.512	0.960
37	Morph	27	14	0.658	0.729	24	0.818	0.462	0.789
38	Morph	27	11	0.644	0.754	2	1.000	0.500	0.801
39	Morph	19	5	0.909	0.833	1	1.000	0.500	1.000
40	Morph	36	6	0.944	0.905	2	1.000	0.400	1.000
41	Morph	36	23	0.529	0.721	99	1.000	0.364	0.989
42	Morph	28	16	0.595	0.630	6	1.000	0.400	0.702
43	Morph	15	6	0.864	0.625	2	1.000	0.400	0.704
44	Morph	92	9	0.579	0.451	1	1.000	0.375	0.880
45	Morph	55	19	0.625	0.814	2	0.875	0.500	0.990
46	Morph	16	9	0.889	0.923	1	1.000	0.750	0.980

(continues)

TABLE 1—*Continued*

Ref.	*Data*	*No. characters*	*No. taxa*	*CI*	*RI*	*No. trees*	*Resolution (CFI)*[b]	B_{50}	B_{max}
47	Morph	23	16	0.487	0.557	4	1.000	0.133	0.560
48	Morph	35	14	0.631	0.730	2	1.000	0.846	0.960
49	Morph	60	24	0.301	0.553	12	1.000	0.174	0.995
50	Morph	24	15	0.617	0.727	7	0.917	0.429	0.941
51	Seq	118	43	0.674	0.851	8	0.650	0.452	1.000
52	Seq	671	24	0.798	0.901	21	0.857	0.609	1.000
53	Seq	82	6	0.689	0.367	1	1.000	0.400	0.663
54	Seq	17	8	0.758	0.692	1	1.000	0.571	0.959
55	Seq	547	23	0.623	0.458	1	1.000	0.318	1.000
56	Seq	86	12	0.625	0.674	1	1.000	0.727	1.000
57	Seq	358	12	0.854	0.806	3	1.000	0.727	0.980
58	Seq	88	12	0.515	0.470	4	1.000	0.545	1.000
59	Seq	174	18	0.580	0.663	2	1.000	0.411	1.000
60	Seq	631	22	0.768	0.776	250	0.947	0.524	1.000
61	Seq	138	31	0.615	0.757	48	0.821	0.433	1.000
62	Seq	718	19	0.622	0.714	6	0.875	0.722	1.000
63	Seq	1428	13	0.788	0.716	4	1.000	0.750	1.000
64	Seq	151	28	0.751	0.843	8	1.000	0.741	1.000
65	Seq	154	19	0.570	0.635	27	0.938	0.333	1.000
66	Seq	637	22	0.737	0.755	2	1.000	0.619	1.000
67	Seq	2167	33	0.559	0.434	2	1.000	0.656	1.000
68	Seq	720	41	0.550	0.729	3	1.000	0.800	1.000
69	Seq	486	9	0.923	0.942	1	1.000	0.750	1.000
70	Seq	2226	24	0.648	0.652	24	1.000	0.609	1.000
71	Seq	1428	21	0.615	0.558	31	1.000	0.400	1.000
72	Seq	270	28	0.636	0.749	3	0.960	0.111	1.000
73	Seq	652	35	0.752	0.654	250	0.906	0.529	1.000
74	RFLP	161	33	0.459	0.781	6	1.000	0.688	1.000
75	RFLP	82	68	0.562	0.892	250	0.708	0.433	1.000
76	RFLP	58	24	0.784	0.957	5	0.667	0.435	1.000
77	RFLP	65	11	0.755	0.835	4	0.875	0.300	1.000
78	RFLP	311	66	0.475	0.709	250	0.840	0.308	1.000
79	RFLP	43	12	0.872	0.935	1	1.000	0.818	1.000
80	RFLP	21	42	0.957	0.994	48	0.282	0.244	1.000
81	RFLP	104	6	0.889	0.917	1	1.000	0.600	1.000
82	RFLP	245	23	0.573	0.729	8	1.000	0.636	1.000
83	RFLP	110	27	0.894	0.845	6	0.625	0.538	1.000
84	RFLP	56	20	0.862	0.962	8	0.765	0.526	1.000
85	RFLP	45	32	0.381	0.746	6	0.724	0.290	0.970
86	RFLP	74	10	0.935	0.889	3	0.857	0.555	1.000
87	RFLP	115	16	0.846	0.868	19	0.615	0.400	1.000
88	RFLP	38	27	0.844	0.951	250	0.500	0.346	0.990
89	RFLP	29	11	0.967	0.957	1	0.500	0.400	1.000
90	RFLP	72	14	0.673	0.670	1	1.000	0.846	0.970
91	RFLP	44	33	0.740	0.949	1	0.533	0.437	1.000
92	RFLP	53	14	0.855	0.938	2	0.818	0.692	1.000

(continues)

TABLE 1—*Continued*

Ref.	Data	No. characters	No. taxa	CI	RI	No. trees	Resolution (CFI)[b]	B_{50}	B_{max}
93	RFLP	128	46	0.888	0.941	250	0.465	0.378	0.990
94	RFLP	91	45	0.561	0.834	250	0.810	0.432	0.990
95	RFLP	55	10	0.902	0.926	1	1.000	0.777	1.000
96	RFLP	194	9	0.975	0.970	1	0.667	0.500	1.000
97	RFLP	26	9	1.000	1.000	1	0.333	0.250	1.000
98	RFLP	72	38	0.459	0.824	64	0.829	0.541	1.000
99	RFLP	10	23	1.000	1.000	1	0.250	0.227	0.980
100	RFLP	30	10	0.762	0.792	1	0.857	0.333	0.900
101	RFLP	25	40	0.962	0.985	2	0.270	0.256	0.940

[a] See Appendix for references to studies.
[b] Normalized consensus fork index

provided in Table 3. These indicate the association among pairs of variables when all other variables are held constant. Significance tests for the matrix of partial correlations were not reported owing to multiple test issues (Sokal and Rohlf 1981). Instead, multiple regression was used. Principal component analysis was used to summarize variation patterns in the eight-dimensional space of the bootstrap indices plus the data set parameters (Table 4 and Fig. 1). Least squares multiple regression (Tables 5 and 6) of B_{50} and B_{max} against all the other variables simultaneously was used to examine the effect of the other variables on these measures of confidence. Finally, multiple regression was also used to repeat the analysis of Sanderson and Donoghue (1989), which examined the effect of these variables on the consistency index (Table 7). Unlike in Sanderson and Donoghue (1989) no attempt was made to factor out the effect that autapomorphies tend to inflate CI. The RI provides an index that is corrected for this (see Archie 1996), and a strong case can be made that an index of homoplasy per se, such as the CI, should in fact include autapomorphies (Gòloboff 1991b).

RESULTS

The summary statistics reveal several interesting differences among the three different kinds of data. RFLP data sets have the highest average CIs and RIs, the largest number of most parsimonious trees, and the lowest average level of resolution in any given minimal tree. Sequence data sets have the highest bootstrap support levels for the B_{50} statistic, and they are about tied with RFLP studies for the B_{max} statistic. They also have the greatest resolution. Morphological data sets have the lowest bootstrap support, but their reten-

TABLE 2 Summary Statistics Classified by Type of Data

Statistic	*Morphological*	*RFLP*	*DNA sequence*
N	50	29	22
No. of characters			
Mean	33.8	86.8	606.4
Minimum	10	10	17
Maximum	92	311	2226
No. of taxa			
Mean	16.8	25.9	21.8
Minimum	5	6	6
Maximum	56	68	43
Consistency index (CI)			
Mean	0.67	0.77	0.68
Minimum	0.30	0.38	0.51
Maximum	1.00	1.00	0.92
Retention index (RI)			
Mean	0.74	0.88	0.69
Minimum	0.45	0.67	0.36
Maximum	1.0	1.0	0.94
No. of minimal trees			
Mean	33.5	49.9	30.5
Minimum	1	1	1
Maximum	250[a]	250[a]	250[a]
Resolution (CFI)[b]			
Mean	0.92	0.72	0.95
Minimum	0.53	0.25	0.65
Maximum	1.00	1.00	1.00
B_{50}			
Mean	0.41	0.48	0.55
Minimum	0.07	0.22	0.11
Maximum	0.85	0.84	0.80
B_{max}			
Mean	0.90	0.99	0.98
Minimum	0.56	0.90	0.66
Maximum	1.00	1.00	1.00

[a] Maximum set by MAXTREES option during PAUP runs.
[b] Normalized consensus fork index (see Swofford 1991).

tion indices are higher than those for sequence data sets. These findings hint at complex and subtle differences among different classes of data, which sample size may not have allowed us to detect in earlier work (Sanderson

TABLE 3 Partial Correlation Coefficients for Variables Surveyed

Variable	CI	No. characters	No. taxa	No. trees	B_{max}	B_{50}	RI	Resolution (CFI)
CI		0.308	−0.679	0.187	−0.007	0.170	0.619	−0.367
NChars			0.325	−0.082	0.195	0.197	−0.463	0.046
Ntaxa				0.546	0.116	0.077	0.361	−0.394
Ntrees					0.033	−0.234	−0.039	0.206
B_{max}						0.303	0.202	−0.076
B_{50}							0.202	0.529
RI								−0.238
Resolution								

and Donoghue 1989). Significance of these differences is tested in regression analysis (see below).

Partial correlation analysis reveals the association of pairs of these variables. The highest partial correlations are found between CI and number of taxa (−0.67), between CI and RI (0.61), between number of taxa and number of minimal trees (0.55), and between degree of resolution and B_{50} (0.53). A more synthetic view of the relationships among the variables is provided in a principal components analysis (Table 4 and Fig. 1). Plotted in the space of the first three principal components are the data points and vectors representing the directions in this space of the the original variables, measured as deviations from the mean. For example, the axis for CI points from the population mean to a point one standard deviation away in the positive

TABLE 4 Component Loadings for First Five Principal Components of Variables Surveyed

	Component 1	Component 2	Component 3	Component 4	Component 5
Eigen value	2.3226	2.0250	1.6027	0.7282	0.5634
Percentage	29.0327	25.3130	20.0342	9.1026	7.0422
CumPercent	29.0327	54.3457	74.3799	83.4825	90.5247
Eigenvectors					
CI	0.5193	−0.3057	−0.0439	−0.2011	0.3998
NChars	−0.1771	−0.0519	0.5507	−0.7408	0.2084
Ntaxa	0.0330	0.5864	0.3025	0.0634	−0.1324
Ntrees	0.0322	0.5502	0.1682	0.2401	0.6369
B_{max}	0.2805	−0.0132	0.5663	0.1742	−0.5196
B_{50}	0.0306	−0.4299	0.4804	0.3742	0.1955
RI	0.6066	−0.0077	0.0114	0.1619	0.1113
Resolution	−0.4990	−0.2685	0.1514	0.3904	0.2296

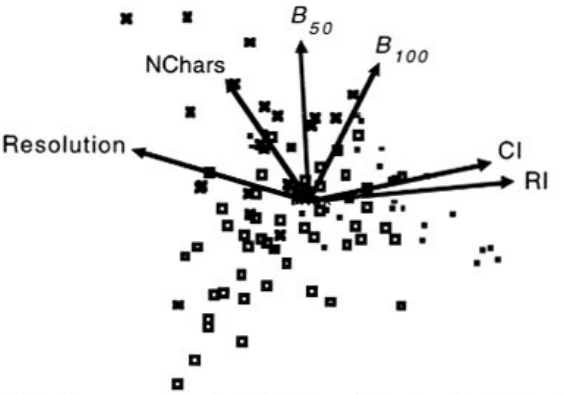

FIGURE 1 Principal component analysis of eight variables listed in Tables 3 and 4. Open squares are morphological data sets; crosses are DNA sequence data sets; closed squares are RFLP data sets. Plot was originally a three-dimensional plot of the first three principal components (Table 4), standardized so that each axis has the same standard deviation. The arrows represent vectors corresponding to a deviation of one standard deviation above the population mean for that variable, all others held constant. The graph is a projection of the original three-dimensional graph into two dimensions. Six of the eight variables have strong components in this plane; the other two, number of taxa and number of trees, are more or less perpendicular to the plane of the paper. Correlation among variables corresponds roughly to similarity in direction of vectors.

direction, all other variables being held constant. Variables with arrows in more or less the same direction are highly correlated. CI and RI are highly correlated, and this graph has the two measures of bootstrap support at roughly right angles to the measures of homoplasy, indicating poor association. Interestingly, degree of resolution is almost 180° away from the two measures of homoplasy, indicating that studies with high CIs or RIs tend to

TABLE 5 Multiple Regression Analysis[a] of B_{50}

Source	*No. params.*	*dF*	*Sum of squares*	*F ratio*	*Prob>F*
Data type	2	2	0.2057	5.3576	0.0063
CI	1	1	0.0293	1.5288	0.2194
NChars	1	1	0.0478	2.4931	0.1178
Ntaxa	1	1	0.0024	0.1265	0.7229
Ntrees	1	1	0.0765	3.9857	0.0488
RI	1	1	0.1385	7.2172	0.0086
Resolution	1	1	0.8396	43.7233	0.0000

[a] For 101 observations: $r^2 = 0.4900$. Whole model ANOVA F ratio = 11.0495 ($P < 0.0001$).

TABLE 6 Multiple Regression Analysis[a] of B_{max}

Source	*No. params.*	*dF*	*Sum of squares*	*F ratio*	*Prob>F*
Data type	2	2	784.9901	5.3164	0.0065
CI	1	1	2.9939	0.0406	0.8408
NChars	1	1	118.0663	1.5992	0.2092
Ntaxa	1	1	24.7098	0.3347	0.5643
Ntrees	1	1	0.3578	0.0048	0.9446
RI	1	1	554.8130	7.5151	0.0074
Resolution	1	1	121.9801	1.6522	0.2019

[a]For 101 observations: $r^2 = 0.32366$. Whole model ANOVA F ratio = 5.5034 ($P < 0.0001$).

have low resolution. The figure clearly shows that this effect is most pronounced for RFLP studies, and that those same studies with high CIs tend to have just average bootstrap support. The differences in CI and bootstrap support noted earlier between types of data can also be observed in this plot.

Multiple regression analysis (Tables 5 and 6) provides estimates of the significance of components in these associations. Both B_{max} and B_{50} depend strongly on the type of data and the retention index. Neither of them depends on CI, the number of characters, or the number of taxa. In addition, B_{50} depends on both the number of most parsimonious trees and the degree of resolution, whereas B_{max} does not. Although overall significance of both models is high, the amount of the variation explained is only 50% in the case of B_{50} and considerably less than that in the case of B_{max}. Clearly, other factors in these data sets must explain the remaining variation in bootstrap support.

TABLE 7 Parameter Estimates in Multiple Regression Analysis of CI on Variables in Survey[a]

Term	*Estimate*	*Standard error*	*t ratio*	*Prob>|t|*
Intercept	0.4223	0.1065	3.97	0.0001
Data[RFLP–seq]	−0.0123	0.0176	−0.70	0.4849
Data[morph–seq]	−0.0263	0.0159	−1.65	0.1017
NChars[a]	0.00007	0.00008	0.95	0.3460
Ntaxa	−0.0081	0.00093	−8.80	0.0000
Ntrees	0.00027	0.00015	1.80	0.0757
RI	0.7980	0.08351	9.56	0.0000
Resolution	−0.2087	0.06524	−3.20	0.0019

[a]For 101 observations: $r^2 = 0.7581$.

[b]Four sequence records excluded: these had over 1000 characters in their matrices, and when included lead to a marginally significant effect due to characters.

Multiple regression analysis of CI on various factors other than bootstrap support (Table 7) reconfirms the main findings of Sanderson and Donoghue (1989). CI depends markedly on the number of taxa ($P < 0.001$), but not on the number of characters (but see footnote to Table 7), the number of most parsimonious trees, or the type of data. It also depends on RI and on the degree of resolution; neither factor was included in our previous analysis. Oddly enough, the direction of the dependence of CI on degree of resolution is negative, meaning that homoplasy is correlated with level of resolution. Although one might predict that highly resolved cladograms would show low homoplasy, in fact, the reverse is true; the pattern is swayed by many studies that are very poorly resolved and therefore have high CIs. We believe that this is partly explained by the fact that these studies have fewer "effective" taxa than it appears, owing to large polytomies The polytomies arise because many of the taxa are actually indistinguishable on the basis of the characters in the matrix. The retention index is less sensitive to this phenomenon (see Table 3).

DISCUSSION

If it is agreed that the consistency index is the best measure of homoplasy per se; then these results suggest that homoplasy and confidence are decoupled. CI has no influence on either measure of bootstrap support for a tree. Table 1 includes numerous studies in which bootstrap support is high but CI is low and vice versa, which contradicts the conventional notion that the level of homoplasy has a substantial bearing on confidence in the associated phylogenetic hypothesis. This suggests that regardless of the amount of homoplasy observed in a data set, it has often proven possible to reconstruct trees with a requisite level of robustness. That this is true independent of the number of taxa in a data set should be some encouragement to investigators studying large and highly homoplastic clades.

This result apparently contradicts other findings that the level of homoplasy correlates directly with accuracy (not confidence) in the tree. For example, Archie (1996) reports results from simulation studies that show that accuracy of the tree estimate does indeed improve with lower levels of homoplasy. Although condidence and accuracy need not be the same thing, they should not be completely uncoupled. We suspect that the discrepancy arises from the different contexts in which homoplasy is being examined in a simulation versus a "meta-analysis." In a simulation, all results are conditional on a given model of evolution. Conditional on that model, the accuracy of tree estimation probably depends on the rate of evolution and thereby the level of homoplasy. In a survey such as ours the results are conditional not on a given evolutionary model but on a given survey sampling scheme.

Nothing about our sample survey methodology can guarantee that there is anything approaching a common model of evolution underlying all 101 data sets. The factors that do impinge on the issue of confidence in this survey emerge despite the vagaries of sample design and bias. The effect of the level of homoplasy may simply be too weak in the face of ordinary systematic practice (as mimicked by our survey design) to exert much influence on bootstrap values. Recall that even with seven independent variables we can explain at best only half the variation in bootstrap levels.

One result that was not entirely unexpected was the finding that some variables were influenced significantly by the type of data. Our earlier study had found no effect of type of data on level of homoplasy (CI), but the sample size for the molecular studies was small. The present, much larger analysis, also finds no effect on CI but the type of data does evidently affect bootstrap confidence levels. Sequence data are the "best" and morphological data the "worst" according to the bootstrap tests. The point of the present paper is not to examine the relative reliability of molecular versus morphological data, but we feel compelled to point out that our results can be viewed from at least two perspectives. One perspective would view our finding that molecular data sets have higher bootstrap values as confirmation of the view that molecular data are better than morphological data. However, morphological data are "nearly" as reliable as molecular data, and are usually considerably less expensive to obtain. This leads inevitably into consideration of the cost/benefit ratio measured in something like dollars per bootstrap percentage point.

Moreover, it is also clear that the overall resolution of individual maximum parsimony trees is worst for one class of molecular data, RFLP studies. This is a curiosity that remains to be explained adequately. Perusal of the trees obtained from chloroplast RFLP studies suggests that they are qualitatively different from trees obtained in either morphological or DNA sequence studies. They regularly contain one or more large polytomies, sometimes entailing 20–50% of the taxa. Morphological and sequence studies often have unresolved regions but they are rarely so unresolved as many RFLP studies. Whether this reflects an intrinsic bias in character "choice" in RFLP studies relative to other data or real differences in underlying evolutionary process is unknown. Even though individual trees from RFLP studies are the least resolved of the three kinds of data sets, it is important to note that the average resolution of the bootstrap majority rule trees (i.e, as measured by B_{50}) is still higher than those for morphological data sets.

This apparent contradiction might be explained as follows. Lack of resolution in individual most parsimonious trees (the factor surveyed as "Resolution" in this study) is probably due to either too few synapomorphies across the board or a very uneven distribution of synapomorphies among clades. Either would cause some taxa to be scored identically. Low resolution of the bootstrap majority rule tree can be due to this factor, too, which is why level of resolution and B_{50} are highly correlated (Table 3), but bootstrap

resolution also reflects conflicting signals in the data owing to real character conflict. Thus, the bootstrap support values for RFLP studies may stem from lack of resolution due to the distribution of synapomorphies, whereas the somewhat lower bootstrap support values for morphological studies may just reflect real noise obscuring the phylogenetic signal.

In sum, we conclude that the conventional view that confidence is directly related to the level of homoplasy in a data set is not supported by available data from phylogenetic studies. The consistency index is a measure of homoplasy, not robustness. The implication of this is that homoplasy and confidence are two separate issues—not completely independent perhaps, especially in carefully controlled simulation experiments—but effectively so in real data sets.

ACKNOWLEDGMENTS

Thanks to T. Givnish and K. Sytsma, R. Jansen for providing us information on their studies of patterns of homoplasy in surveys of real data sets, and the Harvard Phylogeny Discussion Group for comments.

APPENDIX: SOURCES FOR SURVEY OF PHYLOGENETIC DATA SETS

Morphological Data

1. Anderberg, A. A. 1991. Taxonomy and phylogeny of the tribe Inuleae (Asteraceae). *Plant Systematics and Evolution* **176**:75–123.
2. Anderberg, A. A. 1994. Phylogeny of the Empetraceae, with special emphasis on character evolution in the genus *Empetrum*. *Systematic Botany* **19**:35–46.
3. Anderberg, A. A., and K. Bremer. 1991. Parsimony analysis and cladistic reclassification of the *Relhania* generic group (Asteraceae–Gnaphalieae). *Annals of the Missouri Botanical Garden* **78**:1061–1072.
4. Anderberg, A. A., and P. Eldenas. 1991. A cladistic analysis of *Anigozanthos* and *Macropidia* (Haemodoraceae). *Australian Systematic Botany* **4**:655–664.
5. An-Ming, L. 1990. A preliminary cladistic study of the families of the superorder Lamiiflorae. *Biological Journal of the Linnean Society* **103**: 39–57.
6. Ayers, T. J. 1990. Systematics of *Heterotoma* (Campanulaceae) and the evolution of nectar spurs in the New World Lobelioideae. *Systematic Botany* **15**:296–327.

7. Barfod, A. S. 1991. A monographic study of the subfamily Phytelephantoideae (Arecaceae). *Opera Botanica* **105**:1–73.
8. Berry, P. E. 1989. A systematic revision of *Fuchsia* sect. *Quelusia* (Onagraceae). *Annals of the Missouri Botanical Garden* **76**:532–584.
9. Boufford, D. E., J. V. Crisci, H. Tobe, and P. C. Hoch. 1990. A cladistic analysis of *Circaea* (Onagraceae). *Cladistics* **6**:171–182.
10. Bruyns, P., and H. P. Linder. 1991. A revision of *Microloma* R. Br. (Asclepiadaceae–Asclepiadeae). *Botanische Jahrbücher für Systematik, Pflanzengeschichte und Pflanzengeographie* **112**:453–527.
11. Campbell, C. S. 1986. Phylogenetic reconstructions and two new varieties in the *Andropogon virginicus* complex (Poaceae: Andropogoneae). *Systematic Botany* **11**:280–292.
12. Caputo, P., and S. Cozzolino. 1994. A cladistic analysis of Dipsacaceae (Dipsacales). *Plant Systematics and Evolution* **189**:41–61.
13. Carr, B. L., J. V. Crisci, and P. C. Hoch. 1990. A cladistic analysis of the genus *Gaura* (Onagraceae). *Systematic Botany* **15**:454–461.
14. Cox, P. A. 1990. Pollination and the evolution of breeding systems in Pandanaceae. *Annals of the Missouri Botanical Garden* **77**:816–840.
15. Cox, P. B., and L. E. Urbatsch. 1990. A phylogenetic analysis of the coneflower genera (Asteraceae: Heliantheae). *Systematic Botany* **15**: 394–402.
16. Crane, P. R. 1985. Phylogenetic analysis of seed plants and the origin of angiosperms. *Annals of the Missouri Botanical Garden* **72**:716–793.
17. Crisci, J. V., and P. E. Berry. 1990. A phylogenetic reevaluation of the Old World species of *Fuchsia*. *Annals of the Missouri Botanical Garden* **77**:517–522.
18. Crisp, M. D. 1991. Contributions towards a revision of *Daviesia* Smith (Fabaceae: Mirbelieae). II. The *D. latifolia* group. *Australian Systematic Botany* **4**:229–298.
19. Cruden, R. W. 1991. A revision of *Isidrogalvia* (Liliaceae): Recognition for Ruiz and Pavon's genus. *Systematic Botany* **16**:270–282.
20. Curry, K. J., L. M. McDowell, W. S. Judd, and W. L. Stern. 1991. Osmophores, floral features, and systematics of Stanhopea (Orchidaceae). *American Journal of Botany* **78**:610–623.
21. Doyle, J. A., and M. J. Donoghue. 1992. Fossils and seed plant phylogeny reanalyzed. *Brittonia* **44**:89–106.
22. Eriksson, R. 1994. Phylogeny of Cyclanthaceae. *Plant Systematics and Evolution* **190**:31–47.
23. Goldblatt, P., and J. E. Henrich. 1991. *Calydorea* Herbert (Iridaceae–Tigridieae): Notes on this New World genus and reduction to synonymy of *Salpingostylis*, *Cardiostigma*, *Itysa*, and *Catila*. *Annals of the Missouri Botanical Garden* **78**:504–511.

24. Graham, L. E., C. F. Delwiche, and B. D. Mishler. 1991. Phylogenetic connections between the 'green algae' and the 'bryophytes'. *Advances in Bryology* **4**:213–244.
25. Gustafsson, M. H. G., and K. Bremer. 1995. Morphology and phylogenetic interrelationships of the Asteraceae, Calyceraceae, Campanulaceae, Goodeniaceae, and related families (Asterales). *American Journal of Botany* **82**:250–265.
26. Hammel, B. E., and G. J. Wilder. 1989. *Dianthoveus*: A new genus of Cyclanthaceae. *Annals of the Missouri Botanical Garden* **76**:112–123.
27. Hickey, L. J., and D. W. Taylor. 1991. The leaf architecture of *Ticodendron* and the application of foliar characters in discerning its relationships. *Annals of the Missouri Botanical Garden* **78**:105–130.
28. Hoch, P. C., J. V. Crisci, H. Tobe, and P. E. Berry. 1993. A cladistic analysis of the plant family Onagraceae. *Systematic Botany* **18**:31–47.
29. Huck, R. B., W. S. Judd, W. M. Whitten, J. D. Skean, Jr. , R. P. Wunderlin, and K. R. Delaney. 1989. A new *Dicerandra* (Labiatae) from the Lake Wales ridge of Florida, with a cladistic analysis and discussion of endemism. *Systematic Botany* **14**:197–213.
30. Hufford, L. D., and P. R. Crane. 1989. A preliminary phylogenetic analysis of the 'lower' Hamamelidae. *In* P. R. Crane and S. Blackmore (Eds.), *Evolution, systematics, and fossil history of the Hamamelidae*, pp. 175–192. Clarendon Press, Oxford.
31. Judd, W. S. 1989. Taxonomic studies in the Miconieae (Melastomataceae). III. Cladistic analysis of axillary-flowered taxa. *Annals of the Missouri Botanical Garden* **76**:476–495.
32. Karis, P. O. 1989. Systematics of the genus *Metalasia* (Asteraceae–Gnaphalieae). *Opera Botanica* **99**:1–150.
33. Kenrick, P., and P. R. Crane. 1991. Water-conducting cells in early fossil land plants: Implications for the early evolution of tracheophytes. *Botanical Gazette* **152**:335–356.
34. Knapp, S. 1991. A cladistic analysis of the *Solanum* sessile species group (Section *Geminata pro parte*: Solanaceae). *Biological Journal of the Linnean Society* **106**:73–89.
35. Kress, W. J. 1990. The phylogeny and classification of the Zingiberales. *Annals of the Missouri Botanical Garden* **77**:698–721.
36. Kurzweil, H., H. P. Linder, and P. Chesselet. 1991. The phylogeny and evolution of the *Pterygodium–Corycium* complex (Coryciinae, Orchidaceae). *Plant Systematics and Evolution* **175**:161–223.
37. Ladiges, P. Y., and C. J. Humphries. 1983. A cladistic study of *Arillastrum*, *Angophora*, and *Eucalyptus* (Myrtaceae). *Botanical Journal of the Linnean Society* **87**:105–134.
38. Lavin, M. 1990. The genus *Sphinctospermum* (Leguminosae): Taxonomy and tribal relationships as inferred from a cladistic analysis of traditional data. *Systematic Botany* **15**:544–559.

39. Luteyn, J. L. 1991. A synopsis of the genus *Didonica* (Ericaceae: Vaccinieae) with two new species. *Systematic Botany* **16**:587–597.
40. Manning, J. C., and P. Goldblatt. 1991. Systematic and phylogenetic significance of the seed coat in the shrubby African Iridaceae, *Nivenia*, *Klattia*, and *Witsenia*. *Botanical Journal of the Linnean Society* **107**: 387–404.
41. Mishler, B. D. 1990. Reproductive biology and species distinctions in the moss genus *Tortula*, as represented in Mexico. *Systematic Botany* **15**:86–97.
42. Olmstead, R. 1989. Phylogeny, phenotypic evolution, and biogeography of the *Scutellaria angustifolia* complex (Lamiaceae): Inference from morphological and molecular data. *Systematic Botany* **14**:320–338.
43. Schutte, A. L., and B.-E. Van Wyk. 1990. Taxonomic relationships in the genus *Dichilus* (Fabaceae–Crotalarieae). *South African Journal of Botany* **56**:244–256.
44. Schwarzwalder, R. N., and D. L. Dilcher. 1991. Systematic placement of the Platanaceae in the Hamamelidae. *Annals of the Missouri Botanical Garden* **78**:962–969.
45. Simpson, M. G. 1990. Phylogeny and classification of the Haemodoraceae. *Annals of the Missouri Botanical Garden* **77**:722–784.
46. Smith, G. F., and B.-E. Van Wyk. 1991. Generic relationships in the Alooideae (Asphodelaceae). *Taxon* **40**:557–581.
47. Taylor, D. W., and L. J. Hickey. 1992. Phylogenetic evidence for the herbaceous origin or angiosperms. *Plant Systematics and Evolution* **180**:137–156.
48. Tucker, S. C., and A. W. Douglas. 1993. Utility of ontogenetic and conventional characters in determining phylogenetic relationships of Saururaceae and Piperaceae (Piperales). *Systematic Botany* **18**:614–641.
49. Varadarajan, G. S., and A. J. Gilmartin. 1988. Phylogenetic relationships of groups of genera within the subfamily Pitcairnioideae (Bromeliaceae). *Systematic Botany* **13**:283–293.
50. Wilken, D., and R. L. Hartman. 1991. A revision of the *Ipomopsis spicata* complex (Polemoniaceae). *Systematic Botany* **16**:143–161.

Sequence Data

51. Baldwin, B. G., and R. H. Robichaux. 1995. Historical biogeography and ecology af the Hawaiian silversword alliance (Asteraceae): New molecular phylogenetic perspectives. *In* W. L. Wagner and V. A. Funk (Eds.), *Hawaiian biogeography. Evolution on a hot spot archipelago*, pp. 259–287. Smithsonian Institution Press, Washington, DC.

52. Baum, D. A., K. J. Sytsma, and P. C. Hoch 1994. A phylogenetic analysis of *Epilobium* (Onagraceae) based on nuclear ribosomal DNA sequences. *Systematic Botany* **19**:363–388.
53. Bremer, K. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution* **42**:795–803.
54. Bult, E. A., and E. A. Zimmer. 1993. Nuclear ribosomal RNA sequences for inferring tribal relationships within Onagraceae. *Systematic Botany* **18**:48–63.
55. Campbell, C. S., M. J. Donoghue, B. G. Baldwin, and M. F. Wojciechowski. 1995. Phylogenetic relationships in Maloideae (Rosaceae): Evidence from sequences of internal transcribed spacers of nuclear ribosomal DNA and its congruence with morphology. *American Journal of Botany* **82**:903–918.
56. Conti, E., A. Fischbach, and K. J. Sytsma. 1993. Tribal relationships in Onagraceae: Implications from *rbc*L sequence data. *Annals of the Missouri Botanical Garden* **80**:672–685.
57. Cullings, K. W., and T. D. Bruns. 1992. Phylogenetic origin of the Monotropoideae inferred from partial 28S ribsomal RNA gene sequences. *Canadian Journal of Botany* **70**:1703–1708.
58. Donoghue, M. J., R. G. Olmstead, J. F. Smith, and J. D. Palmer. 1992. Phylogenetic relationships of Dipsacales based on *rbc*L sequences. *Annals of the Missouri Botanical Garden* **79**:333–345.
59. Doyle, J. A., M. J. Donoghue, and E. A. Zimmer. 1994. Integration of morphological and ribosomal RNA data on the origin of angiosperms. *Annals of the Missouri Botanical Garden* **81**:419–450. [ribosomal RNA matrix, Table 3]
60. Eriksson, T., and M. J. Donoghue. Phylogenetic analysis of *Sambucus* and *Adoxa* (Adoxaceae) based on nuclear ribosomal ITS sequences and morphological characters. Submitted for publication.
61. Johnson, L. A., and D. E. Soltis. 1994. *mat*K DNA sequences and phylogenetic reconstruction in Saxifragaceae *s. str. Systematic Botany* **19**:143–156.
62. Kim, K.-J., and R. K. Jansen. 1994. Comparisons of phylogenetic hypotheses among different data sets in dwarf dandelions (*Krigia*, Asteraceae): Additional information from internal transcribed spacer sequences of nuclear ribosomal DNA. *Plant Systematics and Evolution* **190**:157–185.
63. Les, D., D. K. Garvin, and C. F. Wimpee. 1991. Molecular evolutionary history of ancient aquatic angiosperms. *Proceedings of the National Academy of Science USA* **88**:10119–10123.
64. Manen, J.-F., A. Natali, and F. Ehrendorfer. 1994. Phylogeny of Rubiaceae–Rubieae inferred from the sequence of a cpDNA intergene region. *Plant Systematics and Evolution* **190**:195–211.

65. Martin, P. G., and J. M. Dowd. 1991. A comparison of 18S ribosomal RNA and rubisco large subunit sequences for studying angiosperm phylogeny. *Journal of Molecular Evolution* **33**:274–282.
66. Nickrent, D. L., K. P. Schuette, and E. M. Starr. 1994. A molecular phylogeny of *Arceuthobium* (Viscaceae) based on nuclear ribosomal DNA internal transcribed spacer sequences. *American Journal of Botany* **81**:1149–1160.
67. Olmstead, R. G., and P. A. Reeves. 1995. Evidence for the polyphyly of the Schrophulariaceae based on chloroplast *rbc*L and *ndh*F sequences. *Annals of the Missouri Botanical Garden* **82**:176–193.
68. Sanderson, M. J., and M. F. Wojciechowski. 1996. Diversification rates in a temperate legume clade: Are there "so many species" of *Astragalus* (Fabaceae)? *American Journal of Botany*, in press.
69. Sang, T., D. J. Crawford, T. F. Stuessy, and M. S. O. 1995. ITS sequences and the phylogeny of the genus *Robinsonia* (Asteraceae). *Systematic Botany* **20**:55–64.
70. Scotland, R. W., J. A. Sweere, P. A. Reeves, and R. G. Olmstead. 1995. Higher-level systematics of Acanthaceae determined by chloroplast DNA sequences. *American Journal of Botany* **82**:266–275. [24 *ndh*F sequences, Fig. 4]
71. Scotland, R. W., J. A. Sweere, P. A. Reeves, and R. G. Olmstead. 1995. Higher-level systematics of Acanthaceae determined by chloroplast DNA sequences. *American Journal of Botany* **82**:266–275. [21 *rbc*L sequences, Fig. 7]
72. Steele, K. P., and R. Vilgalys. 1994. Phylogenetic analyses of Polemoniaceae using nucleotide sequences of the plastid gene *mat*K. *Systematic Botany* **19**:126–142.
73. Wojciechowski, M. F., and M. J. Sanderson. In preparation.

RFLP Data

74. Bremer, B., and R. K. Jansen. 1991. Comparative restriction site mapping of chloroplast DNA implies new phylogenetic relationships within Rubiaceae. *American Journal of Botany* **78**:198–213.
75. Bruneau, A., and J. J. Doyle. 1993. Cladistic analysis of chloroplast DNA restriction site characters in *Erythrina* (Leguminosae: Phaseoleae). *Systematic Botany* **18**:229–247.
76. Conant, D. S., D. B. Stein, A. E. C. Valinski, and P. Sudarsanam. 1994. Phylogenetic implications of chloroplast DNA variation in the Cyatheaceae. I. *Systematic Botany* **19**:60–72.
77. Crisci, J. V., E. A. Zimmer, P. C. Hoch, G. B. Johnson, C. Mudd, and N. S. Pan. 1990. Phylogenetic implications of ribosomal DNA restriction site variation in the plant family Onagraceae. *Annals of the Missouri Botanical Garden* **77**:523–538.

78. Doyle, J. J., and J. L. Doyle. 1993. Chloroplast DNA phylogeny of the papilionoid legume tribe Phaseoleae. *Systematic Botany* **18**:309–327.
79. Freudenstein, J. V., and J. J. Doyle. 1994. Character transformation and relationships in *Corallorhiza* (Orchidaceae: Epidendroideae). I. Plastid DNA. *American Journal of Botany* **81**:1449–1457.
80. Freudenstein, J. V., and J. J. Doyle. 1994. Plastid DNA, morphological variation, and the phylogenetic species concept: The *Corallorhiza maculata* (Orchidaceae) complex. *Systematic Botany* **19**:273–290.
81. Johansson, J. T., and R. K. Jansen. 1991. Chloroplast DNA variation among five species of Ranunculaceae: Structure, sequence divergence, and phylogenetic relationships. *Plant Systematics and Evolution* **178**: 9–25.
82. Kadereit, J. W., and K. J. Sytsma. 1992. Disassembling *Papaver*: A restriction site analysis of chloroplast DNA. *Nordic Journal of Botany* **12**:205–217.
83. Larson, S. R., and J. Doebley. 1994. Restriction site variation in the chloroplast genome of *Tripsacum* (Poaceae): Phylogeny and rates of sequence evolution. *Systematic Botany* **19**:21–34.
84. Lavin, M., S. Mathews, and C. Hughes. 1991. Chloroplast DNA variation in *Gliricidia sepium* (Leguminosae): Intraspecific phylogeny and tokogeny. *American Journal of Botany* **78**:1576–1585.
85. Manos, P. S., K. C. Nixon, and J. J. Doyle. 1993. Cladistic analysis of restriction site variation within the chloroplast DNA inverted repeat region of selected Hamamelididae. *Systematic Botany* **18**:551–562.
86. Morgan, D. R. 1993. A molecular systematic study and taxonomic revision of *Psilactis* (Asteraceae: Astereae). *Systematic Botany* **18**: 290–308.
87. Mummenhoff, K., and M. Koch. 1994. Chloroplast DNA restriction site variation and phylogenetic relationships in the genus *Thlaspi sensu lato* (Brassicaceae). *Systematic Botany* **19**:73–88.
88. Pennington, R. T. 1995. Cladistic analysis of chloroplast DNA restriction site characters in *Andira* (Leguminosae: Dalbergieae). *American Journal of Botany* **82**:526–534.
89. Peterson, G., and J. F. Doebley. 1993. Chloroplast DNA variation in the genus *Secale* (Poaceae). *Plant Systematics and Evolution* **187**:115–125.
90. Philbrick, C. T., and R. K. Jansen. 1991. Phylogenetic studies of North American *Callitriche* (Callitrichaceae) using chloroplast DNA restriction fragment analysis. *Systematic Botany* **16**:478–491.
91. Pillay, M., and K. W. Hilu. 1995. Chloroplast-DNA restriction site analysis in the genus *Bromus* (Poaceae). *American Journal of Botany* **82**:239–249.

92. Potter, D., and J. J. Doyle. 1994. Phylogeny and systematics of *Sphenostylis* and *Nesphostylis* (Leguminosae: Phaseoleae) based on morphological and chloroplast DNA data. *Systematic Botany* **19**:389–406.
93. Soreng, R. J. 1990. Chloroplast-DNA phylogenetics and biogeography in a reticulating group: Study in *Poa* (Poaceae). *American Journal of Botany* 77:1383–1400.
94. Soreng, R. J., J. I. Davis, and J. J. Doyle. 1990. A phylogenetic analysis of chloroplast DNA restriction site variation in Poaceae subfam. Pooideae. *Plant Systematics and Evolution* **172**:83–97.
95. Sytsma, K. J., and J. F. Smith. 1988. DNA and morphology: Comparisons in the Onagraceae. *Annals of the Missouri Botanical Garden* 75: 1217–1237.
96. Systma, K. J., J. F. Smith, and P. E. Berry. 1991. The use of chloroplast DNA to assess biogeography and evolution of morphology, breeding systems, and flavonoids in *Fuschia* sect. *Skinnera* (Onagraceae). *Systematic Botany* **16**:257–269.
97. Talbert, L. E., G. M. Magyar, M. Lavin, T. K. Blake, and S. L. Moylan. 1991. Molecular evidence for the origin of the S-derived genomes of polyploid *Triticum* species. *American Journal of Botany* **78**:340–349.
98. Vaillancourt, R. E., and N. F. Weeden. 1993. Chloroplast DNA phylogeny of Old World *Vigna* (Leguminosae). *Systematic Botany* **18**:642–651.
99. Whittemore, A. T., and B. A. Schaal. 1991. Interspecific gene flow in sympatric oaks. *Proceedings of the National Academy of Sciences USA* 88:2540–2544.
100. Wolfe, A. D., and W. J. Elisens. 1994. Nuclear ribosomal DNA restriction-site variation in *Penstemon* section *Peltanthera* (Scrophulariaceae): An evaluation of diploid hybrid speciation and evidence for introgression. *American Journal of Botany* **81**:1627–1635.
101. Wolf, P. G., P. S. Soltis, and D. E. Soltis. 1993. Phylogenetic significance of chloroplast restriction site variation in the *Ipomopsis aggregata* complex and related species (Polemoniaceae). *Systematic Botany* **18**: 652–662.

REFERENCES

Arnqvist, G., and D. Wooster. 1995. Meta-analysis: Synthesizing research findings in ecology and evolution. Trends in Ecology and Evolution 10:236–240.

Archie, J. W. 1989. Homoplasy excess ratios: New indices for measuring levels of homoplasy in phylogenetic systematics and a critique of the consistency index. Systematic Zoology 38: 235–269.

Archie, J. W. 1996. Measures of homoplasy. Pp. 153–188 *in* M. J. Sanderson and L. Hufford, eds. Homoplasy: The recurrence of similarity in evolution. Academic Press, San Diego.

Bremer, K. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. Evolution 42:795–803.

Bremer, K. 1994. Branch support and tree stability. Cladistics 10:295–304.

Cummings, M. P., S. P. Otto, and J. Wakeley. 1995. Sampling properties of DNA sequence data in phylogenetic analysis. Molecular Biology and Evolution 12:814–822.

Donoghue, M. J., R. G. Olmstead, J. F. Smith, and J. D. Palmer. 1992. Phylogenetic relationships of Dipsacales based on *rbc*L sequences. Annals of the Missouri Botanical Garden 79: 333–345.

Donoghue, M. J., and M. J. Sanderson. 1992. The suitability of molecular and morphological evidence in reconstructing plant phylogeny. Pp. 340–368 *in* P. S. Soltis, D. E. Soltis, and J. J. Doyle, eds. Molecular systematics in plants. Chapman and Hall, New York.

Farris, J. S. 1989. The retention index and rescaled consistency index. Cladistics 5:417–419.

Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Systematic Zoology 27:401–410.

Felsenstein, J. 1985. Phylogenies and the comparative method. American Naturalist 125:1–15.

Givnish, T. J., and K. J. Sytsma. 1992. Chloroplast DNA restriction site data yield phylogenies with less homoplasy than analyses based on morphology or DNA sequences. American Journal of Botany 79:145 [abstract]

Goloboff, P. 1991a. Homoplasy and the choice among cladograms. Cladistics 7:215–232.

Goloboff, P. 1991b. Random data, homoplasy and information. Cladistics 7:395–406.

Hillis, D. M., and J. J. Bull. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Systematic Biology 42:182–192.

Jansen, R. K. 1995. DNA sequence data isn't really better than cpDNA restriction site data for studies of plant phylogeny at the intrafamilial level. American Journal of Botany 82:137 [abstract]

Klassen, G. J., R. D. Mooi, and A. Locke. 1991. Consistency indices and random data. Systematic Zoology 40:446–457.

Kluge, A., and J. S. Farris. 1969. Quantitative phyletics and the evolution of anurans. Systematic Zoology 18:1–32.

Landrum, L. 1993. Factors influencing th accuracy of the parsimony criterion and a method for estimating true tree length. Systematic Botany 18:516–524.

Lankester, E. R. 1870. On the use of the term homology in modern zoology. Annals and Magazine of Natural History 6:34–43.

Meier, R., P. Kores, and S. Darwin. 1991. Homoplasy slope ratio: A better measurement of observed homoplasy in cladistic analyses. Systematic Zoology 40:74–88.

Sanderson, M. J. 1989. Confidence limits on phylogenies: The bootstrap revisited. Cladistics 5: 113–129.

Sanderson, M. J. 1995. Objections to bootstrapping phylogenies: A critique. Systematic Biology 44:299–320.

Sanderson, M. J., and M. J. Donoghue. 1989. Patterns of variation in levels of homoplasy. Evolution 43:1781–1795.

Sokal, R. R., and F. J. Rohlf. 1981. Biometry, 2nd ed. W. H. Freeman, New York.

Swofford, D. 1993. PAUP: Phylogenetic analysis using parsimony. Illinois Natural History Survey, Champaign, Illinois.

Swofford, D. L. 1991. When are phylogeny estimates from molecular and morphological data incongruent? Pp. 295–333 *in* M. M. Miyamoto and J. Cracraft, eds. Phylogenetic analysis of DNA sequences. Oxford Univ. Press, New York.

Wendel, J. F., and V. A. Albert. 1992. Phylogenetics of the cotton genus (*Gossypium*): Character-state weighted parsimony of chloroplast-DNA restriction site data and its systeamtic and biogeographic implications. Systematic Botany 17:115–143.