Syst. Biol. 45(4):524–545, 1996

TESTING FOR PHYLOGENETIC CONFLICT AMONG MOLECULAR DATA SETS IN THE TRIBE TRITICEAE (GRAMINEAE)

ROBERTA J. MASON-GAMER AND ELIZABETH A. KELLOGG

Harvard University Herbaria, 22 Divinity Avenue, Cambridge, Massachusetts 02138, USA;
E-mail: rmason@oeb.harvard.edu (R.M.-G.), tkellogg@oeb.harvard.edu (E.A.K.)

Abstract.—Four molecular data sets are available for the diploid intersterile genera of the cereal grain tribe Triticeae, and there are numerous differences among the four published trees. All six pairwise combinations of data sets were examined using tree comparisons, the incongruence length difference test, the Wilcoxon signed ranks test, and a permutation test. We describe some advantages, disadvantages, and properties of the different comparison methods. Test results provide no evidence for significant differences in the phylogenetic signal among the three nuclear data sets, with the exception of the placement of a single taxon. The chloroplast DNA restriction site data, however, support a significantly different tree, and the differences probably reflect a separate evolutionary history of the chloroplast genome. [Gramineae; incongruence length difference test; permutation test; phylogenetic congruence; Poaceae; T-PTP test; Triticeae; Wilcoxon signed ranks test.]

The options for handling multiple phylogenetic data sets for the same taxa have been the focus of increasing discussion in recent years. The issue is not new, but it has become of practical concern with the greater availability of molecular data. Topics of discussion have included measures of dissimilarity among trees (e.g., Estabrook et al., 1985; Penny and Hendy, 1985; Day, 1986; Lapointe and Legendre, 1992) and data sets (e.g., Mickevich and Farris, 1981; Templeton, 1983; Kluge, 1989), whether or under what circumstances multiple data sets should be combined into a single analysis (e.g., Kluge, 1989; Bull et al., 1993; Eernisse and Kluge, 1993; Jones et al., 1993; Kluge and Wolf, 1993; Chevarría and Carpenter, 1994; Chippindale and Wiens, 1994; Huelsenbeck et al., 1994), and whether trees from separate analyses should be used to generate consensus trees (e.g., Mickevich and Farris, 1981; Miyamoto, 1985; Barrett et al., 1991, 1993; de Queiroz, 1993; Nelson, 1993). Options for analyzing multiple data sets have been reviewed (Swofford, 1991; de Queiroz et al., 1995; Miyamoto and Fitch, 1995; Huelsenbeck et al., 1996). The decision to keep data sets separate generally reflects a hypothesis that (1) different evolutionary processes are acting on different data sets or portions of data sets or (2) different data sets reflect different phylogenetic histories.

Recent studies of the grass tribe Triticeae provide a unique opportunity to investigate these issues. This economically important group, which includes wheat, barley, rye, and several significant forage crops, has been the focus of numerous evolutionary and phylogenetic investigations using morphological, cytogenetic, and molecular data. The present study focuses on comparisons among four molecular data sets, including sequences from the spacers of two separate nuclear 5S ribosomal DNA (rDNA) loci (Kellogg and Appels, 1995), sequences of the internal transcribed spacer (ITS) of the nuclear ribosomal repeat (Hsiao et al., 1995), and chloroplast DNA (cpDNA) restriction sites (Mason-Gamer and Kellogg, 1995). We found that the gene trees disagree extensively in the intergeneric relationships they suggest. Because all of the data sets sample throughout the tribe and include only diploid representatives of intersterile genera, we are confident that the differences are not merely a result of poor sampling or inclusion of obviously reticulate taxa. To determine whether the disagreements reflect differences among the phylogenetic histories of the four DNA segments, we performed several tests of congruence among the data sets. Biological and evolutionary interpretations of the differences and similarities

among the data sets have been presented in detail elsewhere (Kellogg et al., 1996). Here, we focus on the details of the comparisons among the data sets.

We examined six pairwise data set combinations using four methods. First, tree-based comparisons were performed by inspection of pairwise topologies and assessment of bootstrap support of both trees by both data sets. This approach reveals differences between trees and whether data sets show underlying support for alternative trees. The remaining three methods statistically test the null hypothesis that the data sets are congruent. The incongruence length difference (ILD) test (Farris et al., 1994) addresses whether two data sets are merely arbitrary subdivisions of what should be considered one large data set. Wilcoxon signed ranks (WSR) tests (Siegel, 1956; Templeton, 1983) and Compare-2 permutation tests (Swofford, 1995) use constrained parsimony analyses to examine the increase in number of steps required by a data set on an alternative tree topology. We chose these four comparison methods from among many possibilities because they provide statistical assessment of hypotheses of conflict and are easily implemented. We report on some advantages, disadvantages, and properties of these methods and draw conclusions about congruence and conflict among the data sets. We are convinced that some of the differences among the individual topologies reflect differences among the phylogenetic histories of the genes or genomes.

METHODS

Data Sets

The 5S rDNA genes are in tandem arrays at two loci on different chromosomes in most of the Triticeae. The loci can be distinguished from one another by the length and sequence of the intergenic spacers (Gerlach and Dyer, 1980; Kellogg and Appels, 1995). There appears to be little or no interlocus recombination; the spacer sequences remain distinct even in polyploids (Scoles et al., 1988; Dvořák et al., 1989). Therefore, the two loci appear to provide independent phylogenetic estimates. The complex pattern of evolution of the 5S rDNA arrays has been described in detail (Kellogg and Appels, 1995); there are strikingly high levels of variation within arrays and among arrays within species. This variation does not greatly affect the phylogenetic utility of the spacer sequences, however, because spacers from within species or genera are nearly always more closely related to each other than to those from other species or genera.

The ITS sequences also are found in tandem arrays at two separate loci on different chromosomes throughout most of the Triticeae (Flavell and Smith, 1974; Flavell and O'Dell, 1976; Miller et al., 1983; Gill and Appels, 1988), but unlike the 5S rDNA loci, the ITS loci are homogenized by recombination. In contrast to the 5S rDNA spacer sequences, no intraspecific ITS sequence variation has been detected, and levels of intrageneric polymorphism are generally low (Hsiao et al., 1995).

The chloroplast genome is clonally and maternally inherited in most angiosperms, including the few grasses examined so far except *Secale*, in which biparental inheritance has been observed (reviewed by Harris and Ingram, 1991). Chloroplast DNA restriction site data provide strong support for some intergeneric relationships in the Triticeae (Mason-Gamer and Kellogg, 1995). In some cases, species within genera can be distinguished clearly, but levels of intraspecific variation are low.

Comparisons were performed on each of the six pairwise combinations of data sets. Within each pair, the two data sets were reduced to match exactly to the level of species (Table 1). In many cases, matched species represent the same individual plant. In three cases, "matched" genera are represented by different species, because to drop them would have required dropping the genus from that pair altogether. Pairing the different species is reasonable because all of the available data support the monophyly of these genera. The number of taxa in the data set pairs differs considerably because different data sets are

TABLE 1. Triticeae taxa in individual and pairwise combinations of data sets.[a] Taxa in only a single data set are not shown.

Taxon	Abbreviation	Individual data sets				Pairwise combinations					
		5S-S	5S-L	ITS	cp	5S-S × 5S-L	5S-S × ITS	5S-L × ITS	ITS × cp	5S-S × cp	5S-L × cp
Aegilops bicornis	*Aeg bic*	X			X					X	
A. comosa	*Aeg com*	X			X					X	
A. longissima	*Aeg lon*	X			X					X	
A. searsii	*Aeg sears*	X	X		X	X				X	X
A. sharonensis	*Aeg shar*	X	X		X	X				X	X
A. speltoides	*Aeg spelt*		X	X	X			X	X		X
A. tauschii	*Aeg tau*	X	X	X	X	X	X	X	X	X	X
A. umbellulata	*Aeg umb*	X	X		X	X				X	X
A. uniaristata	*Aeg uniar*	X			X					X	
Agropyron cristatum	*Agro cris*		X	X	X			X	X		X
A. c. ssp. *puberulum*	*Agro pub*			X	X				X		
A. mongolicum	*Agro mon*			X	X				X		
Australopyrum pectinatum	*Aust pect*	X		X			X				
A. retrofractum	*Aust ret*		X	X	X			X	X		X
A. velutinum	*Aust vel*	X	X		X	X				X	X
Bromus inermis	*Bromus*	X							(X)		
B. tectorum	*Bromus*			X	X				(X)		
Critesion bogdanii	*Crites*		X					(X)			(X)
C. brevisubulatum	*Crites*				X						(X)
C. californicum	*Crites cal*			X	X				X		
C. violoaceum	*Crites viol*			X	X			(X)	X		
Crithopsis delileana	*Crith del*	X		X	X		X		X	X	
Dasypyrum villosum	*Das vil*		X	X	X			X	X		X
Eremopyrum bonaepartis 2X	*Eremo bon*			X	X				X		
Henrardia persica	*Hen pers*	X		X	X		X		X	X	
Heteranthelium piliferum	*Het pil*			X	X				X		
Hordeum bulbosum	*Hord bulb*		X	X				X			
Peridictyon sanctum	*Per sanc*			X	X				X		
Psathyrostachys fragilis	*Psa frag*			X	X				X		
P. juncea	*Psa junc*	X		X	X		X		X	X	
Pseudoroegneria libanotica	*Pse lib*	X		X	X		X		X	X	
P. spicata	*Pse spic*	X	X	X	X	X	X	X	X	X	X
Secale cereale	*Sec cer*	X	X	X	X	X	X	X	X	X	X
S. montanum	*Sec mon*	X	X	X	X	X	X	X	X	X	X
S. sylvestre	*Sec syl*	X	X			X					
S. vavilovii	*Sec vav*	X	X			X					
Taeniatherum caput-medusae	*Tae cm*	X		X	X		X		X	X	
Thinopyrum bessarabicum	*Thi bes*	X		X	X		X		X	X	
T. elongatum	*Thi elon*	X	X	X	X	X	X	X	X	X	X
T. scirpeum	*Thi scir*	X			X					X	
Triticum monococcum	*Trit mon*	X	X	X	X	X	X	X	X	X	X

[a] 5S-S = 5S rDNA short spacers; 5S-L = 5S rDNA long spacers; ITS = internal transcribed spacers; cp = chloroplast DNA. Parentheses indicate cases where one genus was represented by different species in the pairwise comparisons.

missing different taxa. In particular, one or the other of the 5S rDNA loci is absent in some taxa (Kellogg and Appels, 1995). For each pair of data sets, a combined data set was created. Within a pair, one data set or its trees are hereinafter referred to as "rival" to the other member of the pair. The paired data sets are available on the *Systematic Biology* Worldwide Web site (http://www.utexas.edu/depts/systbiol/) and have been submitted to TreeBASE (Sanderson et al., 1994; http://phylogeny.harvard.edu/treebase/).

Unweighted parsimony analyses were done for each of the 12 reduced data sets and the 6 combined data sets with PAUP 3.1.1 (Swofford, 1993) using the heuristic search with the tree bisection–reconnection branch-swapping procedure. For each data set, the set of shortest trees, the strict consensus tree, and bootstrap estimates of support for nodes on the consensus tree were obtained. All of the analyses were carried out on unweighted data. No attempts were made to improve the most-parsimonious trees by experimenting with, for example, the removal of taxa or characters. Such analyses have been carried out elsewhere (Kellogg et al., 1996). The number of potentially informative characters, number and length of most-parsimonious trees, consistency index (uninformative characters excluded), retention index, and rescaled consistency index were obtained following individual and combined analyses using PAUP 3.1.1.

Methods of Comparison

Inspection and assessment of support.—In the initial comparisons, the most-parsimonious consensus trees from the individual data sets were inspected. Paired trees were examined for conflicts involving nodes with bootstrap values >70%. Weakly supported nodes only ambiguously represent patterns within individual data sets, and therefore conflict among data sets cannot be inferred from comparisons involving weak nodes. Although our 70% cutoff is not unreasonable (Hillis and Bull, 1993), any single line drawn between strong and weak bootstrap support must be considered arbitrary.

From each data set, underlying bootstrap support for each node on the rival tree was estimated using the table of "partitions found in one or more trees and frequency of occurrence" from the PAUP 3.1.1 bootstrap output. The results indicate whether a data set supports any rival nodes that do not appear in its own most-parsimonious tree.

ILD tests.—The data sets were further compared using the incongruence length difference test (Farris et al., 1994), implemented as the combinability test in PAUP* 4.0 29d-31d (Swofford, 1995). The test compares the Mickevich and Farris (1981) index to a null distribution based on multiple randomizations (Farris et al., 1994, 1995). First, the lengths of the shortest trees are obtained for each data set and are added to give a sum of tree lengths. The data sets are then combined and randomly repartitioned into two subsets equal in size to the original data sets, and the lengths of the shortest trees and the sum of tree lengths is again determined. The random repartitioning is repeated multiple times to generate a random distribution of the sum of tree lengths. Finally, the sum of tree lengths from the original data sets is compared with the random distribution. If the probability of randomly obtaining a smaller sum of tree lengths than that of the separate data sets is low, the data sets are interpreted as incongruent. Here, 999 random repartitions were used to generate the distribution.

WSR tests.—More detailed comparisons were done using nonparametric WSR tests (Siegel, 1956) as proposed by Templeton (Templeton, 1983; see also Larson's [1994] discussion of the Templeton test). This application of the WSR test compares the number of changes required by each character on the most-parsimonious tree with the changes required by the same characters on a constraint tree. When data are reanalyzed using a constraint tree, some characters may require more steps than on the most-parsimonious tree and others may require fewer. Using PAUP 3.1.1, char-

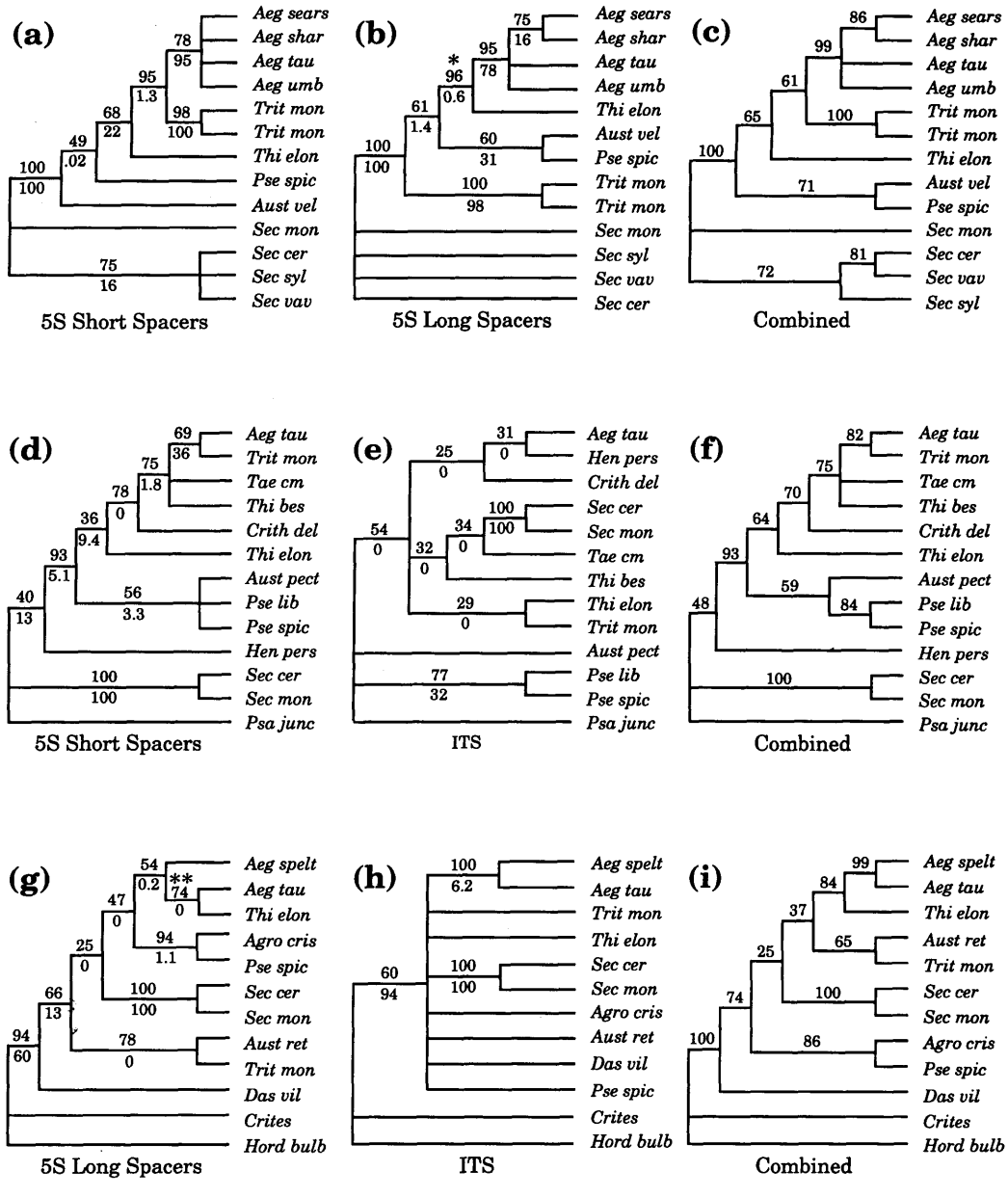

FIGURE 1. Results of individual and combined parsimony analyses for the six pairs of Triticeae data sets. Taxon abbreviations are given in Table 1. Numbers above nodes give the bootstrap support for the tree by the data used to generate the tree. Numbers below the nodes give bootstrap support from the rival paired data set. Asterisks indicate individual nodes that give a significant Wilcoxon signed ranks two-tailed test result when tested against the rival paired data set: $*P < 0.1$; $**P < 0.05$; $***P < 0.01$.

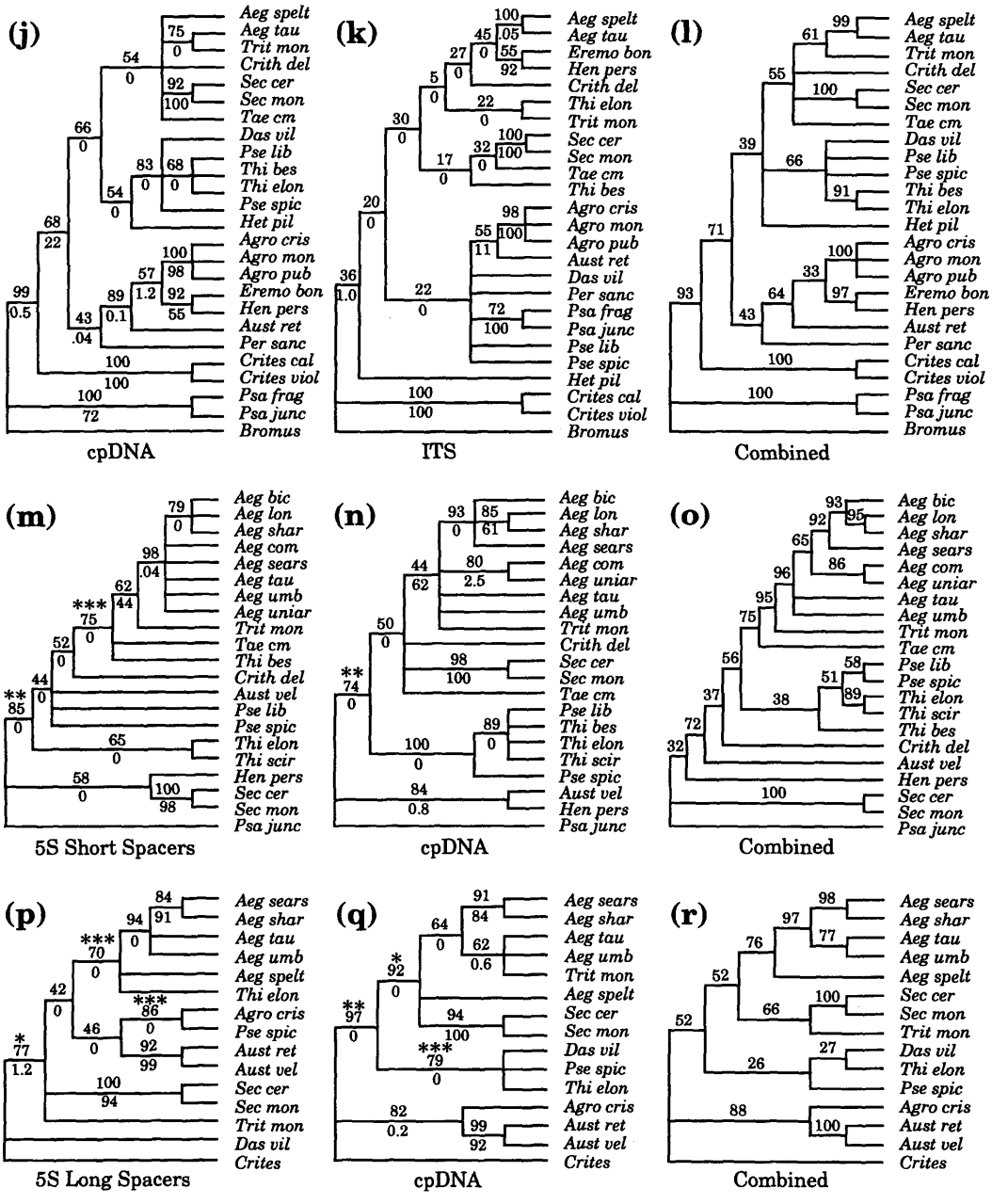

FIGURE 1. Continued.

acters that differ in the number of steps they require on the most-parsimonious tree versus the constraint tree can be identified by comparing the "tree steps" columns of the "character diagnostics" output from both searches. For each character, the difference in number of steps required is ranked according to its magnitude regardless of the direction of the change. The original positive or negative signs are then

applied to the rank values, and the positive and negative ranks are summed separately. If, according to a table of significance values for the WSR test, the sum of the ranks of characters requiring more steps on the rival tree is significantly greater than the sum of ranks of characters requiring fewer, the data are assumed to disagree significantly with the constraint topology. Like Larson (1994), we followed the recommendation of Felsenstein (1985) and used two-tailed probability values, which give conservative estimates of significance values. For each pair of data sets, both sets were tested using numerous constraint trees based on trees from the rival data set. Constraints included (1) the rival strict consensus tree, (2) the 70% majority rule bootstrap rival tree, representing the combination of all nodes with moderate to strong support, (3) sets of trees with single resolved nodes, each of which includes one of the nodes from the 70% bootstrap majority rule rival tree, and (4) the strict consensus topology from the combined analysis of the two data sets. In this way, we tested whether a data set conflicts with the well-supported features of the rival tree and then tested which nodes were the sources of conflict.

Compare-2 tests.—The last method of comparison was the Compare-2 test (Swofford, 1995), which is related to the topology-dependent cladistic permutation tail probability (T-PTP) test described by Faith (1991). The "compare-2" option in PAUP* 4.0 29d-31d (Swofford, 1995) was used. In this test as applied here, the number of steps required by a data set on its most-parsimonious tree is determined, followed by the number of steps required by the same data on a test constraint tree. The data are then permuted by individually randomizing each character's states among the taxa. The randomized data are analyzed, constrained first by the most-parsimonious tree and then by the test constraint tree. Multiple data set permutations are created, and constrained analyses are repeated for each to create a random distribution of the difference in steps required on the most-parsimonious tree versus the test constraint tree. If the difference in steps required by the real data set on the two topologies is significantly greater than that required by randomized data, then the data might be assumed to disagree with the test tree. In our pairwise tests, the test constraint tree was the strict consensus from the rival data set, and the difference in the number of steps required by the real data on the rival tree versus its own most-parsimonious tree was compared with a distribution generated from 999 randomizations of the data set.

TABLE 2. Statistical information[a] for Triticeae trees in Figure 1.

Data set	No. informative characters	No. trees	Tree length	CI	RI	RCI
5S Short	90	39	128	0.883	0.925	0.817
5S Long	88	12	121	0.843	0.890	0.750
Combined	178	3	255	0.843	0.893	0.753
5S Short	111	9	188	0.782	0.701	0.548
ITS	40	4	69	0.667	0.662	0.441
Combined	151	2	262	0.737	0.663	0.489
5S Long	126	1	220	0.773	0.600	0.464
ITS	56	61	98	0.653	0.630	0.412
Combined	182	1	326	0.718	0.576	0.413
cpDNA	74	19	109	0.679	0.843	0.572
ITS	77	9	168	0.583	0.657	0.383
Combined	151	15	293	0.587	0.717	0.421
5S Short	141	12,510	278	0.719	0.713	0.513
cpDNA	33	13	44	0.750	0.888	0.666
Combined	174	3	341	0.683	0.708	0.484
5S Long	126	18	219	0.763	0.677	0.516
cpDNA	33	6	39	0.846	0.919	0.778
Combined	159	1	279	0.717	0.664	0.476

[a] CI = consistency index; RI = retention index; RCI = rescaled consistency index.

RESULTS

Phylogenetic Analyses and Tree Comparisons

A summary of the individual and combined tree topologies, along with the bootstrap support for each node from both paired data sets, is shown in Figure 1. Tree statistics, obtained from PAUP 3.1.1, are shown in Table 2.

5S short spacers and 5S long spacers.—These trees differ mainly in the placement

of a single taxon, *Triticum monococcum* (Figs. 1a, 1b). Although the 5S short spacer data place this species as sister to the *Aegilops* clade with bootstrap support of 95%, the 5S long spacers instead place *Thinopyrum elongatum* as sister to *Aegilops* with 96% bootstrap support. The 5S long spacers show only 1.3% bootstrap support for the *Triticum / Aegilops* clade, and similarly the 5S short spacers show only 0.6% bootstrap support for the *Thinopyrum / Aegilops* clade (Figs. 1a, 1b). Clearly, the data sets do not show strong underlying support for the conflicting rival nodes. The combined analysis reflects the 5S short spacer tree with regard to the placement of *Triticum monococcum*, although the bootstrap support is considerably lower (Fig. 1c). Another but more poorly supported difference between the trees is the relative placements of *Australopyrum velutinum* and *Pseudoroegneria spicata*. The 5S long spacers suggest they are sister taxa, with 60% bootstrap support (Fig. 1b, below node). However, the 5S short spacers also exhibit weak underlying support for the clade, with 31% bootstrap support (Fig. 1b), and the two data sets together provide moderate (71%) support for the clade in the combined tree (Fig. 1c).

5S short spacers and ITS.—Although the 5S short spacer topology has little in common with the ITS tree (Figs. 1d, 1e), the ITS nodes are not well supported; therefore, there are no indisputable differences between the trees. The ITS data provide very weak support for the nodes on the 5S short spacer tree (Fig. 1d), and the 5S short spacer data provide no bootstrap support for any of the intergeneric nodes on the ITS tree (Fig. 1e). The combined ITS and 5S short spacer tree (Fig. 1f) reflects the short spacer topology, with similar bootstrap support for nearly all nodes. Two nodes show increased bootstrap support in the combined analysis: from 69% to 82% for the *Aegilops tauschii / Triticum monococcum* node and from 36% to 64% for the clade of *Aegilops tauschii, Triticum monococcum, Taeniatherum caput-medusae, Thinopyrum bessarabicum, Crithopsis delileana,* and *Thinopyrum elongatum*. In both cases, the ITS data alone exhibit weak underlying support for these nodes (Fig. 1d, below nodes).

5S long spacers and ITS.—This tree comparison (Figs. 1g, 1h) is limited by the lack of resolution of the ITS strict consensus (Fig. 1h). The ITS data provide between 0% and 13% bootstrap support for intergeneric nodes on the 5S long spacer tree (Fig. 1g, below nodes). The combined topology is well resolved, with several moderately to well-supported nodes (Fig. 1i). Although the ITS data alone provide little resolution, their influence in the combined analysis is easily seen. The combined tree differs from the 5S long spacer tree in several respects, particularly in the reversed positions of the *Australopyrum retrofractum / Triticum monococcum* clade with the *Agropyron cristatum / Pseudoroegneria spicata* clade. In addition, support for two nodes that appear in the 5S long spacer tree and not in the ITS tree increases in the combined analysis. Support for the basal position of *Dasypyrum villosum, Critesion,* and *Hordeum bulbosum* increases modestly, from 66% to 74%; the ITS data alone show 13% bootstrap support for this node (Fig. 1g, below node). Support for *Aegilops speltoides, A. tauschii,* and *Thinopyrum elongatum* increases from 54% to 84%. This result is surprising, given the low level of ITS support for the node (0.2%; Fig. 1g, below node).

Chloroplast genome and ITS.—These topologies differ in many details (Figs. 1j, 1k); none of the well-supported intergeneric clades in the cpDNA tree are reflected in the ITS tree, except the *Eremopyrum bonaepartis / Henrardia persica* clade. Although the ITS data weakly support (22% bootstrap) the chloroplast node that places *Psathyrostachys* and *Critesion* at the base of the tree, their support for the other apparently conflicting intergeneric chloroplast nodes ranges from 0% to 1.2% (Fig. 1j, below nodes). The cpDNA data give weak (11%) support to the ITS *Agropyron / Australopyrum* node but only very weak support (0–1%) to the other intergeneric ITS nodes (Fig. 1k, below nodes). The combined topology (Fig. 1l) is similar to the

cpDNA tree, with the nodes showing similar to much lower bootstrap support.

5S short spacers and chloroplast genome.—Differences between these trees involve several well-supported nodes from both data sets (Figs. 1m, 1n). Both data sets moderately support the *Triticum / Aegilops* clade (5S short: 62%; cpDNA: 44%). The cpDNA data show <1% support for the other intergeneric 5S short spacer nodes (Fig. 1m, below nodes), just as the 5S short spacer data show <1% bootstrap support for the other cpDNA intergeneric nodes (Fig. 1n, below nodes). The combined topology (Fig. 1o) shares features of both of the individual trees and has some unique nodes. The *Aegilops / Triticum* clade is more strongly supported in the combined tree than in either individual tree, and the relationships within it are more resolved than in either individual tree.

5S long spacers and chloroplast genome.—These trees differ in all intergeneric groupings (Figs. 1p, 1q). Each tree has six resolved intergeneric nodes; neither data set supports any rival intergeneric node at a level >1.2% (Figs. 1p, 1q, below nodes). The combined tree (Fig. 1r) is intermediate in its placement of *Triticum monococcum* but otherwise more closely resembles the cpDNA tree in spite of the smaller size of the cpDNA data set (33 vs. 126 potentially informative characters).

Incongruence Length Difference Tests

In five of the six pairs of data sets, the summed lengths of the separate trees is significantly lower than when the combined data set is partitioned randomly (Fig. 2), which suggests that patterns of character state variation in the predefined data sets differ significantly. The one exception is the 5S short spacer/ITS pair, in which 10% of the random partitions yielded a sum of tree lengths lower than that of the 5S short spacer/ITS partition (Fig. 2b).

Wilcoxon Signed Ranks Tests

The results of the WSR tests are summarized in Table 3. Significant tests of individual nodes are also indicated in Figure 1 with asterisks above the relevant nodes.

5S short spacers and 5S long spacers.—There are no instances of very strong conflict between either data set and its rival trees, but there are several borderline results. The 5S short spacer data do not significantly conflict with the 5S long spacer consensus tree ($0.099 < P < 0.108$; an increase of 13 steps and a concurrent decrease of 5 steps) and show only weak conflict with the combination of stronger 5S long spacer nodes ($0.055 < P < 0.078$; increase 7, decrease 1). The 5S long spacer data conflict with the 5S short spacer consensus tree ($0.22 < P < 0.048$; increase 12, decrease 3) but not significantly with the well-supported nodes alone (increase 8, decrease 3). Neither data set rejects the combined topology. The only well-supported difference between the tree topologies themselves involves the sister taxon to *Aegilops*. In the corresponding WSR tests, the 5S short spacer data weakly conflict with the *Aegilops / Thinopyrum* node from the 5S long spacer tree ($0.55 < P < 0.78$; increase 7, decrease 1), but the 5S long spacer data do not significantly conflict with the *Triticum / Aegilops* node from the 5S short spacer tree (increase 8, decrease 3).

5S short spacers and ITS.—The 5S short spacer data strongly reject the ITS consensus tree (increase 44, decrease 1), but this result is misleading in terms of character conflict between the data sets because the nodes of the ITS tree show very low bootstrap support. The use of possibly spurious nodes as constraints on another data set is a questionable test for conflict between data sets. The failure of the 5S short spacer data to reject the ITS strong nodes (increase 0, decrease 0) is trivial because there are no well-supported intergeneric ITS nodes. The ITS data do not significantly reject any nodes or combinations of nodes from the 5S short spacer topology, not because all ITS characters require the same number of changes on the ITS and the 5S short spacer trees but because characters that require more steps are partially balanced by those that require fewer. For example, when the ITS data are analyzed with the *Henrardia / Secale / Psathyrostachys*

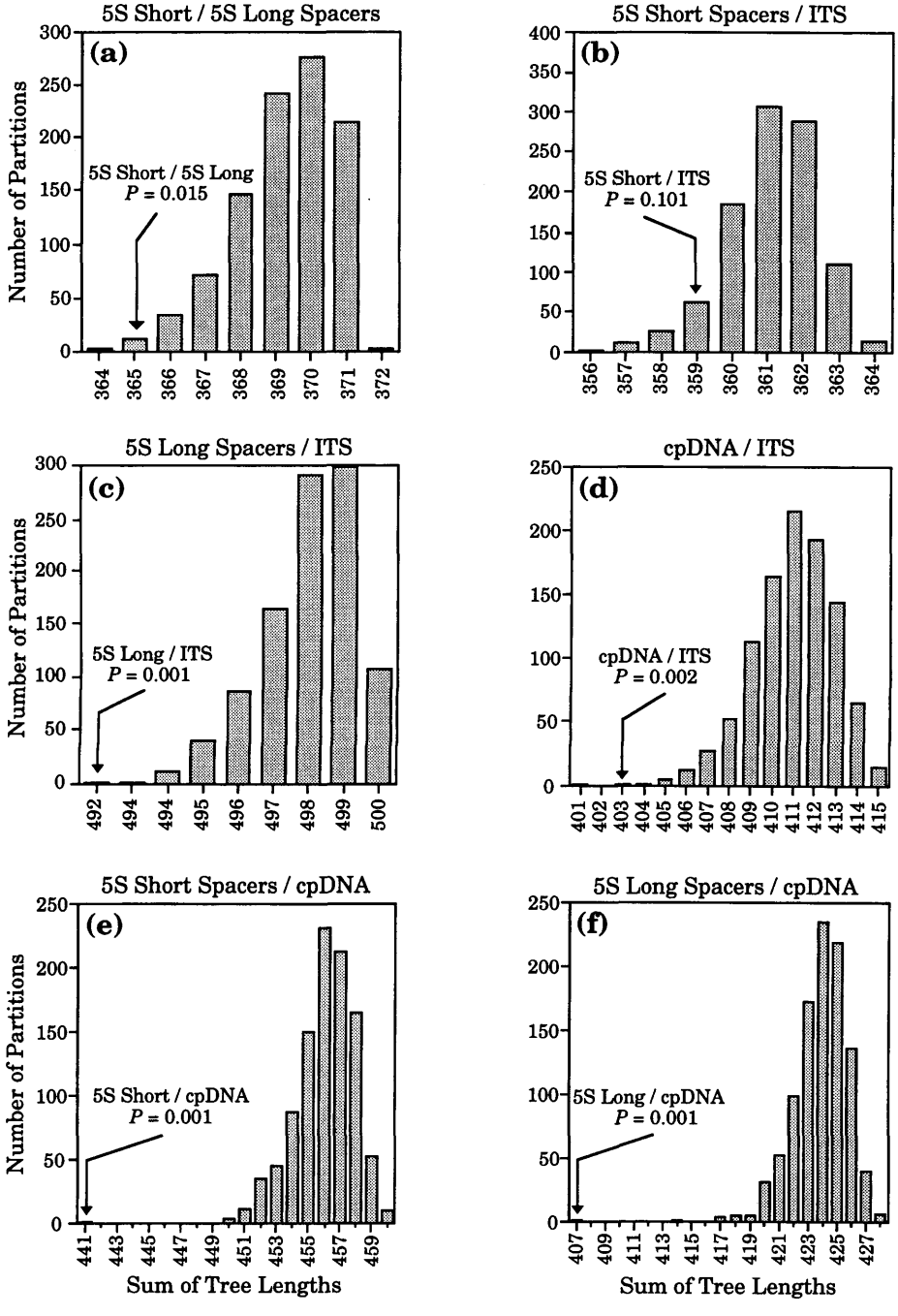

FIGURE 2. Results of the incongruence length difference tests. The arrows indicate the summed length of trees for the actual Triticeae data sets. Vertical gray bars show the random distribution of summed tree lengths generated with 999 random repartitions of the combined data set.

TABLE 3. Summary of Wilcoxon signed ranks test results for all Triticeae data set combinations.

Data set	Constraint	No. steps[a] Gain	Loss	Net	P
5S Short/5S Long					
5S Short	5S Long consensus tree	13	5	8	$0.099 < P < 0.108$
	5S Long strong nodes	7	1	6	$0.055 < P < 0.078$
	Combined tree	3	2	1	>0.1
	Aeg/Thi[b]	7	1	6	$0.055 < P < 0.078$
5S Long	5S Short consensus tree	12	3	9	$0.022 < P < 0.048$
	5S Short strong nodes	8	3	5	>0.1
	Combined tree	8	3	5	>0.1
	Aeg/Trit	8	3	5	>0.1
5S Short/ITS					
5S Short	ITS consensus tree	44	1	43	<0.01
	ITS strong nodes	0	0	0	>0.1
	Combined tree	0	0	0	>0.1
ITS	5S Short consensus tree	11	6	5	>0.1
	5S Short strong nodes	10	6	4	>0.1
	Combined tree	11	6	5	>0.1
	Aeg/Trit/Tae/Thi bes	6	4	2	>0.1
	Aeg/Trit/Tae/Thi/Crith	7	4	3	>0.1
	Hen/Sec/Psa	12	9	3	>0.1
5S Long/ITS					
5S Long	ITS consensus tree	1	0	1	>0.1
	ITS strong nodes	1	0	1	>0.1
	Combined tree	3	1	2	>0.1
	Aeg/Trit	13	3	10	$0.02 < P < 0.05$
ITS	5S Long consensus tree	19	2	17	<0.01
	5S Long strong nodes	13	2	11	<0.01
	Combined tree	9	3	6	>0.1
	Aeg/Thi	10	1	9	0.02
	Agro/Pse	4	2	2	>0.1
	Aust/Trit	7	3	4	>0.1
cpDNA/ITS					
cpDNA	ITS consensus tree	46	2	44	<0.01
	ITS strong nodes	1	0	1	>0.1
	Combined tree	2	1	1	>0.1
ITS	cpDNA consensus tree	27	6	21	<0.01
	cpDNA strong nodes	23	6	17	$0.01 < P < 0.025$
	Combined tree	22	9	13	$0.05 < P < 0.1$
	Agro/Eremo/Hen/Aust	9	5	4	>0.1
	Das/Pse/Thi	7	2	5	>0.1
	Eremo/Hen	0	0	0	>0.1
	Psa/Bromus	4	0	4	>0.1
5S Short/cpDNA					
5S Short	cpDNA consensus tree	26	4	22	<0.01
	cpDNA strong nodes	26	4	22	<0.01
	Combined tree	15	7	8	>0.1
	Aust/Hen/Psa	16	5	11	$0.048 < P < 0.021$
	Pse lib/Thi	13	5	8	$0.099 < P < 0.108$
	Pse lib/Thi/Pse spic	9	6	3	>0.1
	Aust/Hen	10	7	3	>0.1
	Three nodes above	27	4	23	<0.01
cpDNA	5S Short consensus tree	29	0	29	<0.01
	5S Short strong nodes	15	0	15	<0.01
	Combined tree	11	0	11	<0.01
	Aeg/Trit/Tae/Thi bes	9	0	9	<0.01
	Aust/Hen/Psa	10	1	9	0.039

TABLE 3. Continued.

Data set	Constraint	No. steps[a] Gain	Loss	Net	P
5S Long/cpDNA					
5S Long	cp consensus tree	32	4	28	<0.01
	cp strong nodes	21	2	19	<0.01
	Combined tree	21	2	19	<0.01
	Agro/Aust/Crites	12	3	9	$0.022 < P < 0.048$
	Aeg/Sec/Trit	7	1	6	$0.055 < P < 0.078$
	Das/Pse/Thi	20	5	15	<0.01
	Agro/Aust	2	0	2	>0.1
	Trit/Aeg	9	4	5	>0.1
cpDNA	5S Long consenus tree	26	1	25	<0.01
	5S Long strong nodes	23	3	20	<0.01
	Combined tree	2	0	2	>0.1
	Agro/Pse	14	2	12	0.008
	Aeg/Thi	11	0	11	<0.01
	Crites/Das	10	3	7	0.094

[a] The total gain, total loss, and net gain of steps required by all characters given the indicated constraint tree relative to the steps required by the same set of characters on the most-parsimonious unconstrained tree.
[b] See Table 1 for taxon abbreviations.

node as a constraint (Fig. 1d), they require an increase of 12 steps along with a concurrent decrease of 9. The combined topology, which is more similar to the 5S short spacer tree, is not significantly rejected by either the 5S short spacer or the ITS data.

5S long spacers and ITS.—Constraints based on the ITS consensus tree or the combination of strong nodes have almost no resolution. The single intergeneric node is also present on the rival 5S long spacer tree. The ITS-based constraints, therefore, are not rejected by the 5S long spacer data. The ITS data strongly reject the 5S long spacer consensus tree (increase 19, decrease 2) and the combination of strong nodes (increase 13, decrease 2). The tests of individual nodes reflect the different placements of *Triticum monococcum* and *Thinopyrum elongatum* relative to the *Aegilops* species (Figs. 1g, 1h). The 5S long spacer data reject the *Aegilops/Triticum* node (increase 13, decrease 3), and the ITS data reject the *Thinopyrum/Aegilops* node (increase 10, decrease 1). The ITS data do not reject the two additional well-supported 5S long spacer nodes. Neither data set rejects the combined topology.

Chloroplast genome and ITS.—The cpDNA data strongly reject the ITS consensus tree (increase 46, decrease 2). However, as with the ITS and 5S short spacer comparison, many of the ITS nodes are not well supported by the ITS data themselves, and their rejection by the cpDNA data may not result from real conflict between the data sets. The ITS data strongly reject the cpDNA consensus tree (increase 27, decrease 6) and the combination of strong nodes (increase 23, decrease 6). They do not reject any of the four strong cpDNA nodes individually. As above, there are no well-supported intergeneric ITS nodes to test. The combined tree, which is more similar to the cpDNA tree, is not rejected by the cpDNA data (increase 2, decrease 1) and is weakly rejected by the ITS data ($0.05 < P < 0.1$; increase 22, decrease 9).

5S short spacers and chloroplast genome.—The 5S short spacer data strongly reject the cpDNA consensus tree (increase 26, decrease 4) and the combination of strong nodes (increase 26, decrease 4). Likewise, the cpDNA data strongly reject the 5S short spacer consensus tree (increase 29, decrease 0) and the combination of strong nodes (increase 15, decrease 0). The 5S short spacer data reject one cpDNA node, *Australopyrum/Henrardia/Psathyrostachys* (Fig. 1n; increase 16, decrease 5). Although these data fail to reject any of the other

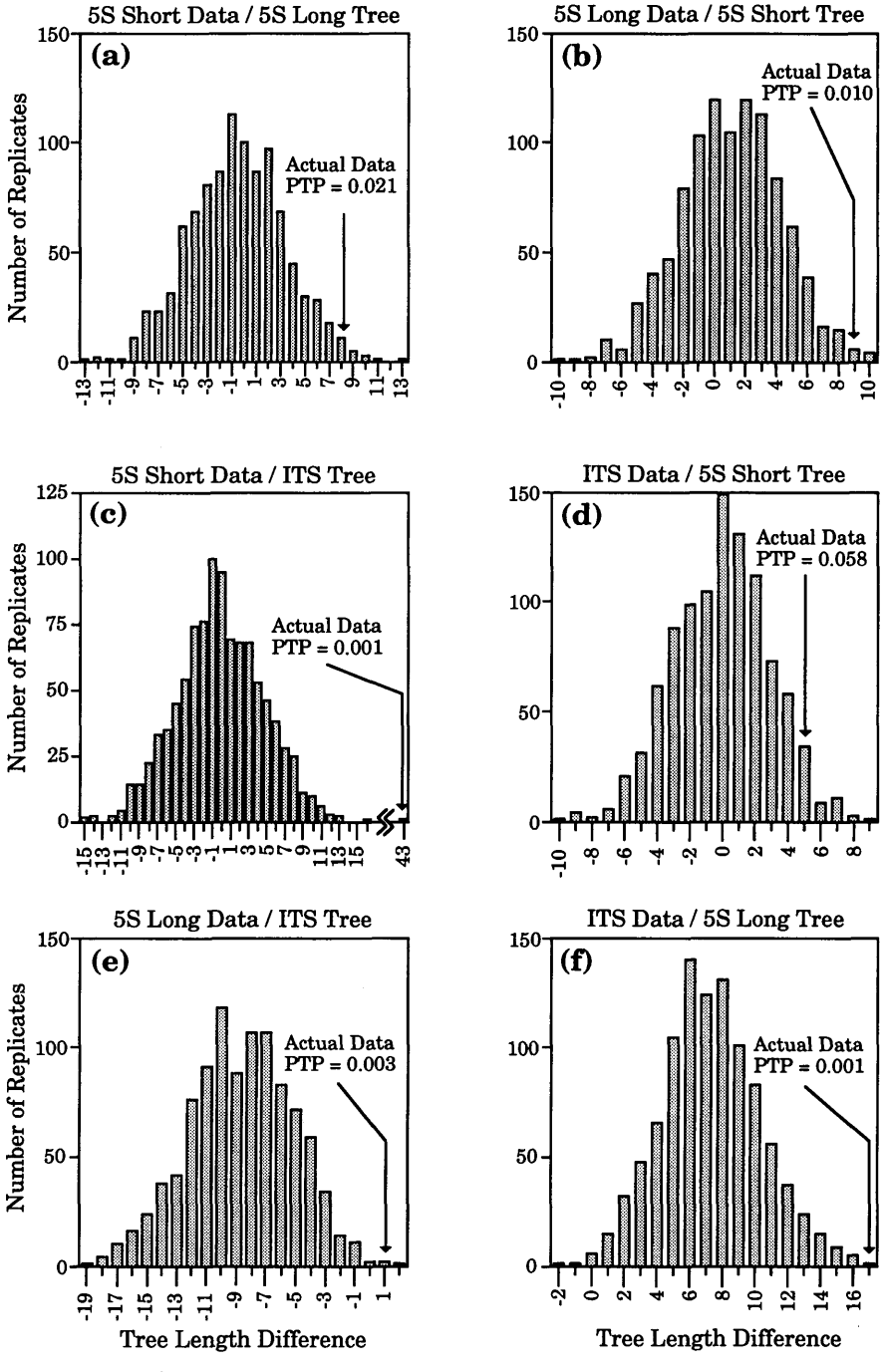

FIGURE 3. Results of the Compare-2 tests. The arrow indicates the increase in number of steps required by the real Triticeae data on the alternative strict consensus tree. The bars show the random distribution of the number of steps required on the most-parsimonious tree versus the rival strict consensus tree generated with 999 permutations of the data set.

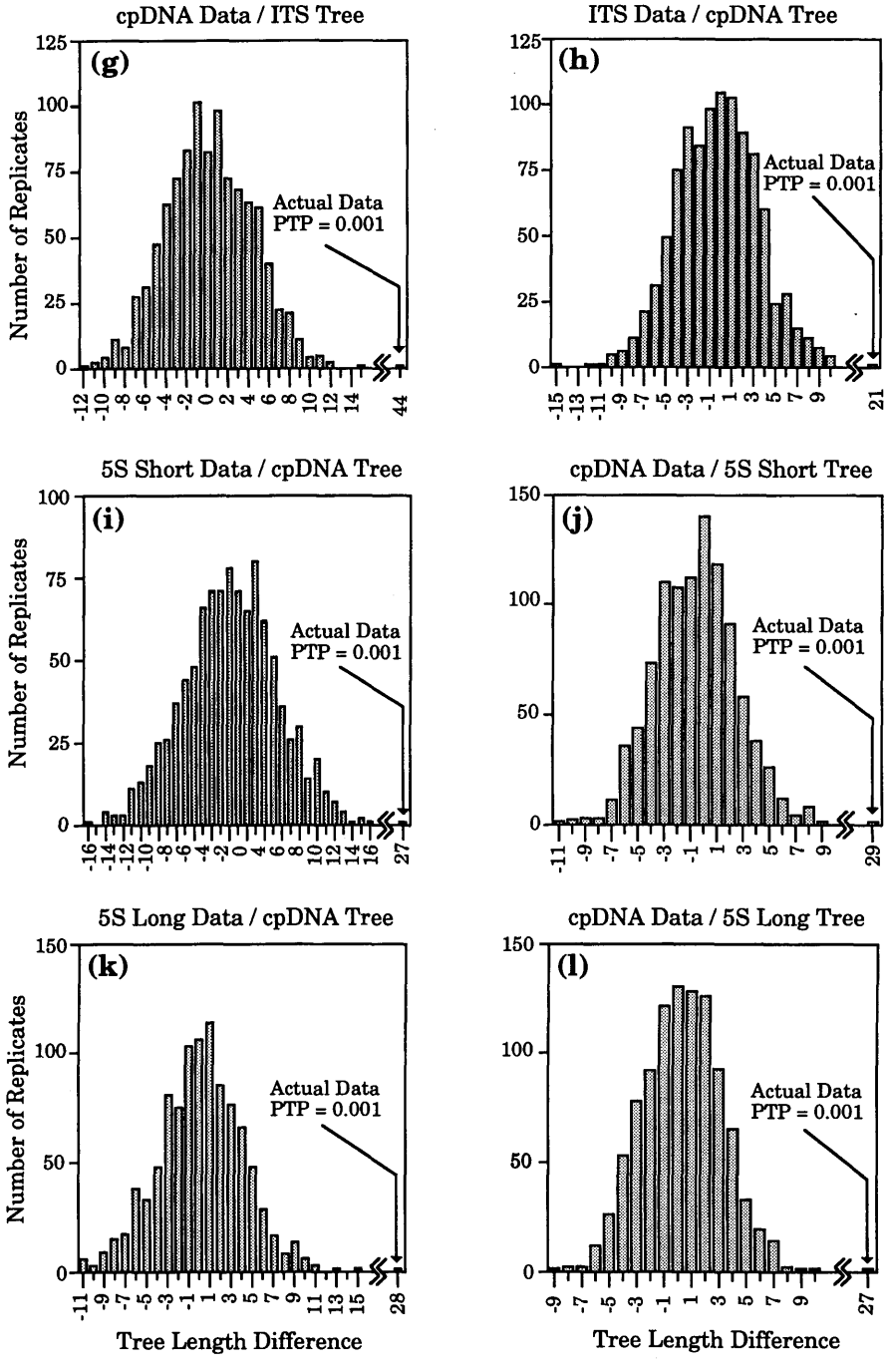

FIGURE 3. Continued.

three well-supported cpDNA nodes individually, they do reject them in combination (increase 27, decrease 4). The cpDNA data reject both of the well-supported intergeneric nodes in the 5S short spacer tree (Fig. 1m). The combined tree, which shares features of both individual trees, is significantly rejected by the cpDNA data (increase 11, decrease 0) but not by the 5S short spacer data, which would require a greater increase in steps (15) but also a greater concurrent decrease (7).

5S long spacers and chloroplast genome.—The 5S long spacer data reject the cpDNA consensus (increase 32, decrease 4) and the combination of strong cpDNA nodes (increase 21, decrease 2). Similarly, the cpDNA data reject the 5S long spacer consensus (increase 26, decrease 1) and the combination of strong nodes (increase 23, decrease 3). The 5S long spacer data reject three of the four moderately to well-supported cpDNA nodes (one of them only weakly) (Fig. 1q), and the cpDNA data reject all three of the moderately to well-supported nodes from the 5S long spacer tree (one of them weakly) (Fig. 1p). The combined tree, which more closely resembles the cpDNA tree, is strongly rejected by the 5S long spacer data (increase 21, decrease 2) but not by the cpDNA data (increase 2, decrease 0).

Summary.—The WSR tests indicate that (1) the 5S short spacer and 5S long spacer data differ with respect to the placement of *Triticum monococcum*; (2) the ITS data do not significantly conflict with the 5S short spacer tree; (3) the ITS data conflict with the placement of *Triticum monococcum* by the 5S long spacer data and with the overall cpDNA topology; and (4) both of the 5S rDNA spacer data sets conflict with the cpDNA data and vice versa, whether they are compared with overall topologies or with individual nodes.

Compare-2 Tests

In most cases, the increase in steps required by a data set on the rival consensus tree is significantly greater than that required by randomized data (Fig. 3). In the one borderline exception, the ITS data require five additional steps when constrained by the 5S short spacer tree, significant only at the 5.8% level (Fig. 3d). Some of the significant results shown in Figure 3 must be interpreted with caution because the relative amounts of resolution and strength of support of the two trees used for the Compare-2 test can have major effects on the results.

DISCUSSION

Tree Comparisons

Comparisons among tree topologies alone cannot be used to show conclusively that patterns of character state variation within data sets conflict. In this study, however, hypotheses of conflict based on inspection of trees were largely supported by subsequent statistical comparisons of data sets. Our initial tree-based hypotheses of incongruence were based only on moderately to well-supported nodes (≥70% bootstrap), and data sets often show little or no bootstrap support for apparently conflicting nodes on rival trees. However, as Barrett et al. (1991), Chippindale and Wiens (1994), and Olmstead and Sweere (1994) have demonstrated, data sets that separately yield apparently conflicting trees may provide strong support for relationships not seen in the individual trees, possibly reflecting underlying congruent signal. If the congruent underlying signal is interpreted as a reflection of the actual phylogenetic history, then the data sets do not really conflict. In our analyses, many of the clades in the combined analyses reflect one or both of the individual trees, but bootstrap support is often decreased. There are several exceptions, however, and at least one of these is of taxonomic interest. The monophyly of the included *Thinopyrum* species has been questioned, and the decision whether to recognize two separate genera has been the focus of much debate (Jauhar, 1988, 1990a, 1990b; Wang, 1989, 1992; Wang and Hsiao, 1989). Our ITS analysis does not place the species together (Fig. 1k) (although the weighted analysis of Hsiao et al. [1995] groups them weakly with 50% bootstrap support). The

cpDNA data place them together but are unable to distinguish them from *Pseudoroegneria libanotica* (Fig. 1j). The combined tree places the two species alone in a well-supported (91% bootstrap) clade (Fig. 1l).

Where data sets yield very different trees (5S short spacers/cpDNA and 5S long spacers/cpDNA), the combined tree more closely reflects one or the other of the individual data sets. Although the combined short spacer/cpDNA topology (Fig. 1o) has elements from both trees, the cpDNA data reject the combined tree in the WSR test (increase 11 steps, decrease 0) but the 5S data do not (increase 15, decrease 7). The combined 5S long spacer/cpDNA tree (Fig. 1r) closely resembles the cpDNA tree (Fig. 1q) in spite of the larger size of the 5S long spacer data set (126 vs. 33 potentially informative characters). The combined topology is not rejected by the cpDNA data (increase 2, decrease 0) but is strongly rejected by the 5S long data (increase 21, decrease 2).

Inspecting data sets to find support for individual nodes on rival trees has been more useful for our purposes than the approach suggested by Rodrigo et al. (1993). When data sets yield trees that are significantly different, these authors suggested generating trees from large numbers of bootstrap resamplings of both data sets and looking for overlap among the two sets of trees. If there is no overlap, potential problematic taxa are individually pruned and parsimony analyses are rerun, and the trees are again compared to see if they are significantly different. (Rodrigo et al. provided a method for determining whether trees are significantly different than would be expected by sampling error alone.) Pruning is repeated with each potential problem taxon individually, until the resulting trees are no longer significantly different and the problem taxa are identified. This method is used to indicate whether differences between trees are due to sampling error rather than to different phylogenetic histories and can be used to identify taxa that cause conflict.

The test proposed by Rodrigo et al. (1993) seems most applicable to situations in which there are one or a few identifiable problem taxa for removal. For data sets with numerous well-supported differences (e.g., 5S long spacers and cpDNA), the repeated cycle of comparing trees, bootstrapping, pruning, and reanalyzing may become impractical. It is difficult to guess which taxa are good candidates for removal when trees share no intergeneric nodes. In addition, saving all bootstrap trees can be problematic when the number of trees is large. When we attempted the test with the 5S long and 5S short spacer data (the data sets with the fewest taxa) using 1,000 bootstrap replicates, as recommended by Lutzoni and Vilgalys (1995), the 5S short data yielded an unwieldy 16.7-MB file of bootstrap trees. We instead chose to examine each data set for bootstrap support for each node of a rival tree to identify whether apparently conflicting nodes were indeed unsupported by the rival data set.

Incongruence within Data Sets

Conflict within a data set may be indicated by WSR test results when a data set shows many characters that require fewer steps on a constraint tree in addition to those that require more. For example, although the ITS trees differ in topology from rival nuclear DNA trees (Figs. 1d, 1e; Figs. 1g, 1h), the ITS data reject few individual nodes from the rival trees, not because of an overall agreement between the ITS data and the nodes in question but because although some ITS characters require more steps on rival nodes than they do on their own shortest trees, other characters require fewer steps. For example, when the ITS data are separately constrained by the three apparently conflicting 5S short spacer nodes, they require net increases of only two or three steps (Table 3). These net increases in fact result from a total increase of 6 steps with a concurrent decrease of 4, an increase of 7 and a decrease of 4, or an increase of 12 and a decrease of 9 (Table 3). Likewise, all but one of the cpDNA nodes reveal ITS characters that require fewer steps over the apparently conflicting node. Only a single

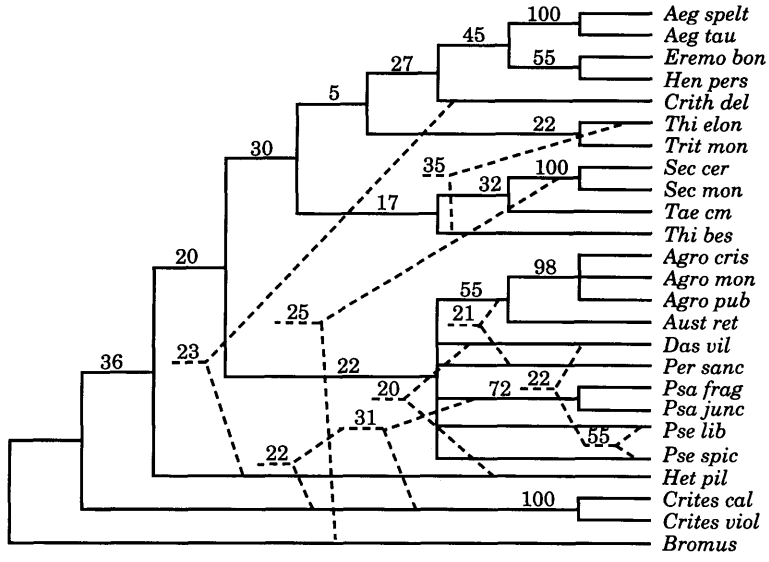

FIGURE 4. ITS strict consensus tree. This ITS data set includes the Triticeae taxa used in the ITS/cpDNA comparison. Taxon abbreviations are given in Table 1. Dashed lines indicate some of the nodes supported by the data but not found in the strict consensus tree. Numbers above all nodes indicate bootstrap support.

node, from the 5S long spacer tree, is rejected by the ITS data (Fig. 1g, asterisks).

The low bootstrap values on the ITS tree suggest internal incongruence as well. Many nodes that do not appear on the consensus tree have bootstrap support similar to those that do. Figure 4 shows the ITS strict consensus tree from the ITS/cpDNA pair. The dashed lines show groups with similar or higher bootstrap support than many of those on the consensus tree. However, few of these underlying nodes correspond to the well-supported nodes in the other three data sets (see also numbers below nodes in Figs. 1d, 1g, 1j; ITS data generally provide little or no support for nodes on rival trees). The potential placement of *Secale* at the base of the ITS tree, however, does correspond to its placement by the 5S short spacers (Figs. 1d, 1m). In addition, the monophyly of *Thinopyrum bessarabicum* and *T. elongatum* is more strongly supported (35% bootstrap) than the separate placement of these species on the strict consensus tree (17% and 22%, respectively).

Global versus Local Comparisons

The combined use of global and local methods of comparison has proved informative here. The ILD test is a useful starting point for comparisons. It determines whether or not two separate data sets are, in effect, arbitrary subdivisions of what should be considered one large data set. When the ILD test suggests incongruence, the Compare-2 test can be used as an additional global assessment of the agreement between a data set and a rival tree. The WSR test, using a full consensus tree or combination of all strong nodes as a constraint, also serves as a global indicator of agreement between a data set and a rival tree. The WSR method has a desirable combination of features: (1) the degree to which data agree or disagree with a topology can be examined in terms of individual characters, (2) the constraint topologies can be fully resolved trees, combinations of nodes, or individual nodes, so that disagreement between a data set and an alternative topology can be pinpointed to specific regions of the topology, and (3) the test shows both the in-

crease and the concurrent decrease required by all characters. Characters that decrease in the number of steps they require on an alternative topology may indicate underlying phylogenetic signal, possibly not reflected in the shortest trees.

Given evidence of conflict, it is informative to determine exactly which taxa are involved. One approach is to remove suspected problem taxa and rerun a global analysis. Because the only well-supported difference between the 5S long and 5S short spacer trees was the placement of *Triticum monococcum*, it was removed and the ILD, Compare-2, and WSR tests were rerun. None of the tests indicated even weakly significant conflict with *T. monococcum* removed (results not shown). In cases with numerous conflicting taxa, the WSR test was used for separately testing each individual well-supported conflicting node and was therefore a versatile test, useful for identifying both global and localized conflict.

Tests Requiring Constraint Trees

Both the Compare-2 test and the WSR test require that data sets be analyzed using topological constraints, and the choice of the constraint tree will affect the outcome of the test. Because our analyses use unrooted constraint trees (Swofford and Begle, 1993), they do not test for conflict with clades but with nodes or subtrees (Trueman, 1995). Our data set comparisons illustrate two potentially misleading results from constraint-based tests.

The first misleading result is an inflated number of steps required by a data set on an alternative constraint topology with many poorly supported nodes, which will lead to a dubious conclusion of significant conflict between data sets in both the WSR and the Compare-2 tests. For example, when the 5S short spacer data are constrained by the ITS strict consensus tree (Fig. 1e), they require a net increase of 43 steps (increase 44, decrease 1), which is a highly significant result in both the Compare-2 (Fig. 3c) and WSR (Table 3) tests. All of the intergeneric nodes on the ITS constraint tree have low bootstrap support and are only ambiguously supported by the ITS data (e.g., Fig. 4). Therefore, constraining an alternative data set with those nodes does not lead to a reasonable comparison of the data sets themselves. Collapsing the poorly supported nodes gives a less well-resolved constraint tree and a more conservative representation of the patterns of character state variation within the data set. The more well-supported the constraint tree, the more accurately comparisons among trees represent comparisons among data sets. The WSR test can be successfully carried out on a collapsed tree or on a tree with any level of resolution, but the results of the Compare-2 test are dependent on the level of resolution of the constraint trees.

If the constraint trees used in the Compare-2 test differ in their level of resolution, the random distribution is affected, which can lead to another potentially misleading result. Consider a different ITS strict consensus tree (Fig. 1h), which has little resolution. In the Compare-2 test as applied here, the 5S long spacer data were constrained first by the ITS tree and then by their own shortest tree (Fig. 1g) and required only one additional step on the ITS tree. This result is not surprising given the low level of resolution of the ITS tree. However, the test still gives a significant Compare-2 result (0.003; Fig. 3e) because the difference in resolution between the 5S long spacer and ITS trees shifts the distribution downward (Fig. 3e), and a length difference of only one step is now significantly large. The randomized data nearly always require fewer steps on the ITS tree because it has less structure. As expected, in the reverse test, where the randomized ITS data are constrained by the 5S long spacer tree and by the ITS strict consensus tree, the tree length differences are nearly all positive (Fig. 3f). The length difference for the unpermuted data (17 steps) is still significant in spite of the positive shift in the random distribution. For the other five data set pairs, the consensus trees are similar in their level of resolution, and the distributions are not shifted (Figs. 3a–3d, 3g–3l).

Our T-PTP tests for monophyly (Faith, 1991) exhibited a similar phenomenon (data not shown). Swofford et al. (1996) suggested that the random distribution in this test does not correspond with the null hypothesis and therefore the test is invalid. Faith and Trueman (1996) argued that this is not the case and that the correct interpretation of the test depends on an accurate definition of the null hypothesis. In this test for monophyly or nonmonophyly of a particular group (Faith, 1991), the number of steps required by a data set is determined for the shortest tree not compatible with that node and then for the shortest tree that is compatible with the node. The difference in the number of steps is then determined. The procedure is repeated multiple times for randomized data, and a random distribution of the difference in length of the shortest incompatible tree versus the shortest compatible tree is generated. Random data commonly give mostly, or sometimes entirely, negative tree length differences because the shortest possible tree for randomized data is more likely to be incompatible with a predefined, single-node constraint tree. Of all possible tree arrangements, far fewer are compatible with a specified node than are incompatible with it. Therefore, random data will usually require more steps on the defined constraint tree than they will on a tree that is merely incompatible with that constraint tree. The resulting largely negative distribution makes it easy to obtain a significantly high positive value and difficult to obtain a significantly low negative value. In 19 of our 26 T-PTP tests (data not shown), a requirement for zero additional steps to break up a monophyletic group would be interpreted as significant support for the monophyly of the group because zero is greater than at least 95% of the values in the random distributions. Faith and Trueman (1996) maintained that "significant support for monophyly" should be equated with "failing to falsify a hypothesis of monophyly" rather than "rejecting a null hypothesis of nonmonophyly." The correct interpretation of significant T-PTP results and, in fact, whether the test is valid at all is a topic of current debate (Faith and Trueman, 1996; Swofford et al., 1996).

Weighted Analyses

Differences among trees could result from analyses based on incorrect assumptions about underlying evolutionary processes for one or both data sets (e.g., Bull et al., 1993). These differences could be accommodated by differential character weighting within a combined analysis (e.g., Chippindale and Wiens, 1994). All of the tests for incongruence and the conclusions presented here are based on unweighted cladistic parsimony analyses. Although we can never conclusively rule out the possibility that observed differences among the data sets result from differences among evolutionary processes acting on them, we have explored some different weighting strategies for the ITS and cpDNA data sets. For the full ITS data set, Hsiao et al. (1995) weighted transversions: transitions: gaps 3:1:1. This approach yields moderately higher bootstrap support, but the trees do not resemble the cpDNA or 5S rDNA spacer trees. We increased the tranversion: transition ratio to 5:1 and 10:1 and found increased conflict between the ITS and 5S rDNA trees (results not shown). For the full cpDNA data set, Mason-Gamer and Kellogg (1995) used successive weighting (Farris, 1969) and weighted gains versus losses (Albert et al., 1992) 1:1.1, 1:1.3, 1:1.5, and 1:2. These cpDNA trees are nearly identical to the unweighted trees (Mason-Gamer and Kellogg, 1995). In addition, trees from preliminary analyses of sequence data from the chloroplast gene *rpo*A share numerous features with the cpDNA restriction site tree (G. Petersen and O. Seberg, in press), suggesting that the cpDNA restriction site topology is not an artifact of an inappropriate weighting strategy or of different evolutionary processes acting on restriction site versus sequence data.

Although the unweighted analyses used here may not give the best possible estimate of the phylogeny from each individual data set, there is little to suggest that

the well-supported differences between, especially, the cpDNA and nuclear gene trees would be diminished with reasonable weighting schemes. The WSR tests suggest that few characters within the cpDNA data set support rival gene trees. When the cpDNA data are analyzed using constraint trees or individual conflicting nodes from the 5S long spacer, 5S short spacer, or ITS data sets, very few cpDNA characters require fewer steps on the constraint tree but many require more. Therefore, there are few or no characters that are candidates for upweighting, even if such an a posteriori weighting strategy were acceptable. Some data sets show considerably more ambivalence for rival trees, in that the characters requiring more changes are partially offset by those requiring fewer. Each of the four consensus trees, however, is rejected by at least two of the three rival data sets.

CONCLUSIONS

Although different tests exhibit different strengths and weaknesses, together they provide a basis for drawing conclusions about congruence and conflict among these data sets. These results, combined with numerous other analyses, have allowed Kellogg et al. (1996) to propose a detailed narrative history of the Triticeae.

Tree comparisons and the ILD test are useful for initial comparisons and will indicate whether more detailed analyses are needed. The WSR test is also useful for global comparisons but must be applied carefully, because weakly supported, possibly spurious nodes in the rival constraint tree can lead to a misleading rejection of the tree. The Compare-2 test shows similar behavior and is further affected by the level of resolution of the rival constraint tree. The independence of the null distribution from the hypothesis being tested led Swofford et al. (1996) to conclude, based on theoretical considerations, that the similar T-PTP test is invalid. After pairwise tree comparisons and global tests have been used to identify potentially conflicting individual nodes, the WSR test can be used to examine each of those nodes individually and will provide information about each character on each constraint tree.

All of the methods support three main conclusions: (1) the 5S short spacer and ITS data are not strongly incongruent, (2) the 5S short spacer and 5S long spacer data sets are significantly incongruent, but the difference can be explained by the conflicting placement of a single taxon, *Triticum monococcum*, and (3) the cpDNA data set reflects a history substantially different from that of any of the nuclear data sets. Given that the data sets support conflicting phylogenetic relationships, how can they be used to shed light on the phylogeny of the Triticeae? Based on extensive analyses of the complete data sets in addition to the results presented here, Kellogg et al. (1996) concluded that after the few problem taxa have been identified the nuclear data sets can be combined to give a single tree. An early introgression event might explain the discrepancy in the placement of *Triticum monococcum*. More numerous events must be postulated to explain the many reticulations required to reconcile the cpDNA and nuclear trees. The cause of the widespread discrepancy between the nuclear and the chloroplast genomes is not known. With existing information, it is impossible to distinguish early widespread hybridization and/or lineage sorting from rare instances of contemporary gene flow. Whatever the cause of the disagreement, however, if the chloroplast and nuclear DNA data do indeed reflect different evolutionary histories, combining them into a single analysis would not contribute to our understanding of the phylogeny and evolution of the Triticeae.

ACKNOWLEDGMENTS

We thank the members of the systematics group at the Harvard Herbaria for comments on analyses of multiple data sets and David Cannatella, Sarah Mathews, Richard Olmstead, Peter Stevens, Jonathan Wendel, and two anonymous reviewers for many insightful comments on the manuscript. Daniel Faith kindly allowed his in-press manuscript (Faith and Trueman, 1996 [this issue]) to be distributed to members of our phylogeny discussion group. David Swofford generously provided a prepress copy of his paper (Swofford et al., 1996 [this issue]) discussing the T-PTP test and allowed us to use and publish results

from test versions of PAUP* 4.0. The work was supported by NSF grant DEB 9106581 to E.A.K.

REFERENCES

ALBERT, V. A., B. D. MISHLER, AND M. W. CHASE. 1992. Character-state weighting for restriction site data in phylogenetic reconstruction, with an example from chloroplast DNA. Pages 369–401 *in* Molecular systematics of plants (P. S. Soltis, D. E. Soltis, and J. Doyle, eds.). Chapman and Hall, New York.

BARRETT, M., M. J. DONOGHUE, AND E. SOBER. 1991. Against consensus. Syst. Zool. 40:486–493.

BARRETT, M., M. J. DONOGHUE, AND E. SOBER. 1993. Crusade? A reply to Nelson. Syst. Biol. 42:216–217.

BULL, J. J., J. P. HUELSENBECK, C. W. CUNNINGHAM, D. L. SWOFFORD, AND P. J. WADDELL. 1993. Partitioning and combining data in phylogenetic analysis. Syst. Biol. 42:384–397.

CHEVARRÍA, G., AND J. M. CARPENTER. 1994. "Total evidence" and the evolution of highly social bees. Cladistics 10:229–258.

CHIPPINDALE, P. T., AND J. J. WIENS. 1994. Weighting, partitioning, and combining characters in phylogenetic analysis. Syst. Biol. 43:278–287.

DAY, W. H. E. 1986. Analysis of quartet dissimilarity measures between undirected phylogenetic trees. Syst. Zool. 35:325–333.

DE QUEIROZ, A. 1993. For consensus (sometimes). Syst. Biol. 42:368–372.

DE QUEIROZ, A., M. J. DONOGHUE, AND J. KIM. 1995. Separate versus combined analysis of phylogenetic evidence. Annu. Rev. Ecol. Syst. 26:657–682.

DVOŘÁK, J., H.-B. ZHANG, R. S. KOTA, AND M. LASSNER. 1989. Organization and evolution of the 5S ribosomal RNA gene family in wheat and related species. Genome 32:1003–1016.

EERNISSE, D. J., AND A. J. KLUGE. 1993. Taxonomic congruence versus total evidence, and amniote phylogeny inferred from fossils, molecules, and morphology. Mol. Biol. Evol. 10:1170–1195.

ESTABROOK, G. F., F. R. MCMORRIS, AND C. A. MEACHAM. 1985. Comparison of undirected phylogenetic trees based on subtrees of four evolutionary units. Syst. Zool. 34:193–200.

FAITH, D. P. 1991. Cladistic permutation tests for monophyly and nonmonophyly. Syst. Zool. 40:366–375.

FAITH, D. P., AND J. W. H. TRUEMAN. 1996. When the topology-dependent permutation test (T-PTP) for monophyly returns significant support for monophyly, should that be equated with (a) rejecting a null hypothesis of nonmonophyly, (b) rejecting a null hypothesis of "no structure," (c) failing to falsify a hypothesis of monophyly, or (d) none of the above? Syst. Biol. 45:580–586.

FARRIS, J. S. 1969. A successive approximations approach to character weighting. Syst. Zool. 18:374–385.

FARRIS, J. S., M. KÄLLERSJÖ, A. G. KLUGE, AND C. BULT. 1994. Testing significance of incongruence. Cladistics 10:315–319.

FARRIS, J. S., M. KÄLLERSJÖ, A. G. KLUGE, AND C. BULT. 1995. Constructing a significance test for incongruence. Syst. Biol. 44:570–572.

FELSENSTEIN, J. 1985. Confidence limits on phylogenies with a molecular clock. Syst. Zool. 34:152–161.

FLAVELL, R. B., AND M. O'DELL. 1976. Ribosomal RNA genes on homoeologous chromosomes of groups 5 and 6 in hexaploid wheat. Heredity 37: 377–385.

FLAVELL, R. B., AND D. B. SMITH. 1974. Variation in nucleolar organiser rRNA gene multiplicity in wheat and rye. Chromosoma 47:327–334.

GERLACH, W. L., AND T. A. DYER. 1980. Sequence organization of the repeating units in the nucleus of wheat which contain 5S RNA genes. Nucleic Acids Res. 8:4851–4865.

GILL, B. S., AND R. APPELS. 1988. Relationships between *Nor*-loci from different Triticeae species. Plant Syst. Evol. 160:77–89.

HARRIS, S. A., AND R. INGRAM. 1991. Chloroplast DNA and biosystematics: The effects of intraspecific diversity and plastid transmission. Taxon 40:393–412.

HILLIS, D. M., AND J. J. BULL. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42:182–192.

HSIAO, C., N. J. CHATTERTON, K. H. ASAY, AND K. B. JENSEN. 1995. Phylogenetic relationships of the monogenomic species of the wheat tribe, Triticeae (Poaceae), inferred from nuclear rDNA (internal transcribed spacer) sequences. Genome 38:211–223.

HUELSENBECK, J. P., J. J. BULL, AND C. W. CUNNINGHAM. 1996. Combining data in phylogenetic analysis. Trends Ecol. Evol. 11:152–158.

HUELSENBECK, J. P., D. L. SWOFFORD, C. W. CUNNINGHAM, J. J. BULL, AND P. J. WADDELL. 1994. Is character weighting a panacea for the problem of data heterogeneity in phylogenetic analysis? Syst. Biol. 43:288–291.

JAUHAR, P. P. 1988. A reassessment of genome relationships between *Thinopyrum bessarabicum* and *T. elongatum* of the Triticeae. Genome 30:903–914.

JAUHAR, P. P. 1990a. Dilemma of genome relationship in the diploid species *Thinopyrum bessarabicum* and *Thinopyrum elongatum* (Triticeae: Poaceae). Genome 33:944–946.

JAUHAR, P. P. 1990b. Multidisciplinary approach to genome analysis in the diploid species, *Thinopyrum bessarabicum* and *Th. elongatum* (*Lophopyrum elongatum*), of the Triticeae. Theor. Appl. Genet. 80:523–536.

JONES, T. R., A. G. KLUGE, AND A. J. WOLF. 1993. When theories and methodologies clash: A phylogenetic reanalysis of the North American ambystomatid salamanders (Caudata: Ambystomatidae). Syst. Biol. 42:92–102.

KELLOGG, E. A., AND R. APPELS. 1995. Intraspecific and interspecific variation in 5S RNA genes are decoupled in diploid wheat relatives. Genetics 140: 325–343.

KELLOGG, E. A., R. APPELS, AND R. J. MASON-GAMER. 1996. When genes tell different stories: The diploid

genera of the Triticeae (Gramineae). Syst. Bot. 21 (in press).

KLUGE, A. G. 1989. A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae: Serpentes). Syst. Zool. 38:7–25.

KLUGE, A. G., AND A. T. WOLF. 1993. Cladistics: What's in a word? Cladistics 9:183–199.

LAPOINTE, F.-J., AND P. LEGENDRE. 1992. A statistical framework to test the consensus among additive trees (cladograms). Syst. Biol. 41:158–171.

LARSON, A. 1994. The comparison of morphological and molecular data in phylogenetic systematics. Pages 371–390 *in* Molecular ecology and evolution: Approaches and applications (B. Schierwater, B. Streit, G. P. Wagner, and R. DeSalle, eds.). Birkhäuser Verlag, Basel, Switzerland.

LUTZONI, F. M., AND R. VILGALYS. 1995. Integration of morphological and molecular data sets in estimating fungal phylogenies. Can. J. Bot. Suppl. 1: S649–S659.

MASON-GAMER, R. J., AND E. A. KELLOGG. 1995. Chloroplast DNA analysis of the monogenomic Triticeae: Phylogenetic implications and genome-specific markers. Pages 301–325 *in* Methods of genome analysis in plants (P. P. Jauhar, ed.). CRC Press, Boca Raton, Florida.

MICKEVICH, M. F., AND J. S. FARRIS. 1981. The implications of congruence in *Menidia*. Syst. Zool. 30:351–370.

MILLER, T. E., J. HUTCHINSON, AND S. M. READER. 1983. The identification of the nucleolus organiser chromosomes of diploid wheat. Theor. Appl. Genet. 65:145–147.

MIYAMOTO, M. M. 1985. Consensus cladograms and general classifications. Cladistics 1:186–189.

MIYAMOTO, M. M., AND W. M. FITCH. 1995. Testing species phylogenies and phylogenetic methods with congruence. Syst. Biol. 44:64–76.

NELSON, G. 1993. Why crusade against consensus? A reply to Barrett, Donoghue, and Sober. Syst. Biol. 42: 215–216.

OLMSTEAD, R. G., AND J. A. SWEERE. 1994. Combining data in phylogenetic systematics: An empirical approach using three molecular data sets in the Solanaceae. Syst. Biol. 43:467–481.

PENNY, D., AND M. D. HENDY. 1985. The use of tree comparison metrics. Syst. Zool. 34:75–82.

PETERSON, G., AND O. SEBERG. In press. Phylogenetic analysis of the Triticeae (Poaceae) based on *rpo*A sequence data. Mol. Phylogenet. Evol.

RODRIGO, A. G., M. KELLY-BORGES, P. R. BERGQUIST, AND P. L. BERGQUIST. 1993. A randomisation test of the null hypothesis that two cladograms are sample estimates of a parametric phylogenetic tree. N.Z. J. Bot. 31:257–268.

SANDERSON, M. J., M. J. DONOGHUE, W. PIEL, AND T. ERIKSSON. 1994. TreeBASE: A prototype database of phylogenetic analyses and an interactive tool for browsing the phylogeny of life. Am. J. Bot. Abstr. 81:183. [Internet address: http://phylogeny.harvard.edu/treebase/]

SCOLES, G. J., B. S. GILL, Z.-Y. XIN, B. C. CLARKE, C. L. MCINTYRE, C. CHAPMAN, AND R. APPELS. 1988. Frequent duplication and deletion events in the 5S RNA genes and the associated spacer regions of the Triticeae. Plant Syst. Evol. 160:105–122.

SIEGEL, S. 1956. Nonparametric statistics for the behavioral sciences. McGraw-Hill, New York.

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

SWOFFORD, D. L. 1993. PAUP: Phylogenetic analysis using parsimony, version 3.1.1. Illinois Natural History Survey, Champaign.

SWOFFORD, D. L. 1995. PAUP*: Phylogenetic analysis using parsimony, beta test version 4.0 29d-31d. Sinauer, Sunderland, Massachusetts.

SWOFFORD, D. L., AND D. P. BEGLE. 1993. PAUP: Phylogenetic analysis using parsimony. User's manual. Illinois Natural History Survey, Champaign.

SWOFFORD, D. L., J. L. THORNE, J. FELSENSTEIN, AND B. M. WIEGMANN. 1996. The topology-dependent permutation test for monophyly does not test for monophyly. Syst. Biol. 45:575–579.

TEMPLETON, A. R. 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. Evolution 37:221–244.

TRUEMAN, J. W. H. 1995. Sponges, plants, and T-PTP. Syst. Biol. 44:429–434.

WANG, R. R.-C. 1989. An assessment of genome analysis based on chromosome pairing in hybrids of perennial Triticeae. Genome 32:179–189.

WANG, R. R.-C. 1992. Genome relationships in the perennial Triticeae based on diploid hybrids and beyond. Hereditas 116:133–136.

WANG, R. R.-C., AND C. HSIAO. 1989. Genome relationships between *Thinopyrum bessarabicum* and *T. elongatum*: Revisited. Genome 32:802–809.

Received 17 October 1995; accepted 15 August 1996
Associate Editor: Richard Olmstead