Cladistic Characters and Cladogram Stability

Jerrold I. Davis; Michael W. Frohlich; Robert J. Soreng

Systematic Botany, Vol. 18, No. 2. (Apr. - Jun., 1993), pp. 188-196.

Cladistic Characters and Cladogram Stability

Jerrold I Davis

L. H. Bailey Hortorium, 462 Mann Library Building, Cornell University,
Ithaca, New York 14853

Michael W. Frohlich

Department of Biology, Union College,
Schenectady, New York 12308

Robert J. Soreng

L. H. Bailey Hortorium, 462 Mann Library Building, Cornell University,
Ithaca, New York 14853

ABSTRACT. Phylogenetic stability is the tendency for monophyletic groups that are resolved by an analysis to continue to be resolved when either the data or the analytical method is altered. Bootstrapping and similar procedures assess the stability of monophyletic groups, but provide little information to indicate which characters are crucial in the support of particular groups. We present a method that does so. Sequential Character Removal consists of a series of cladistic analyses, each conducted with a different subset of a complete data set; the subsets are obtained by the sequential and exhaustive removal of characters, individually and in all combinations of two or more. A character or combination of characters is critical to the resolution of a monophyletic group when the removal of that character or set of characters results in the loss of resolution of the group. Thus, for any monophyletic group, Sequential Character Removal generates a list of characters and character combinations that are critical to the resolution of the group, in the context of an otherwise complete data set. An analysis of a real data set, representing chloroplast DNA restriction site variation among 27 species of Poaceae, demonstrates that characters critical to the recognition of a monophyletic group need not be synapomorphies of that group, and that a clade that is resolved by a complete data set, and fails to be resolved when a particular set of characters is removed, may reappear when additional characters are removed.

Systematists have long been interested in the degree to which the results of their analyses are altered by changes in data or analytical method. The tendency for results to remain constant has been termed *stability* (Mickevich 1980; Rohlf 1984; Rohlf and Sokal 1980), while the alternative tendency, for results to change, has been called *sensitivity* (Neff 1987). Indeed, authors have expressed opposing opinions on which attribute—stability or sensitivity—is preferable in an analytical method (Neff 1987; Rohlf 1984).

Mickevich (1980, p. 163, also 1978) used the term "stability" to denote the degree of "congruence between classifications of the same organisms derived from different data." Rohlf and Sokal (1980) extended the usage of this term to include the degree of congruence between classifications produced when taxa, methods of character scoring, or methods of data analysis are varied. We use the term "stability" in a general sense, to refer to the tendency for results of an analysis to remain unaltered when either the data or the analytical technique is varied.

Although it would be desirable to determine the likelihood that the phylogenetic structure detected by an analysis is an accurate representation of descent relationships, phylogenetic stability is not a measure of congruence with "true" relationships; any assessment of stability is specific to a particular set of data and analytical methods.

The most widely applied method for assessing phylogenetic stability in the face of changes in data is the bootstrap (Felsenstein 1985, 1988; Sanderson 1989). This technique involves a series of "replicate" cladistic analyses conducted with random samples of characters, drawn with replacement, from a complete data set. Bootstrap analysis generates a frequency for each monophyletic group detected in any of the replicate analyses. This numeric score, often depicted as a percentage associated with each

monophyletic grouping in a majority-rule consensus tree, represents the frequency with which each such group is resolved among the replicate analyses conducted. Other approaches to the evaluation of phylogenetic stability are the jack-knife (Mueller and Ayala 1982), and analysis of the resilience of cladistic structure to the relaxation of parsimony (e.g., Bremer 1988).

All of these methods focus on the stability of phylogenetic structure *per se*, that is, on the monophyletic groups that are resolved, while the importance of individual characters in defining these groups tends to fall into the background. Character transformations can be mapped onto the consensus trees that are generated by bootstrap analysis, but the placement of characters on consensus trees, in general, often violates the strictures of parsimony, and therefore can be misleading (Maddison 1989; Miyamoto 1985). Thus, the investigation of phylogenetic stability has tended to become isolated from the analysis of character support for individual clades.

It would be useful to have a method for assessing relationships between individual characters and the stability of particular monophyletic groups. In this paper we describe a method in which individual characters and groups of characters are removed sequentially from a data set in exhaustive combinations, and cladistic analyses are then conducted on the resulting incomplete data sets. For a data set that includes n characters, the analysis generates separate cladistic analyses for all possible incomplete data sets with $n - 1$ characters, $n - 2$ characters, and so on. This method superficially resembles the statistical jackknife, but its intended purpose is not to generate quantitative indices of stability; rather, it is proposed as a means for indentifying specific characters and character combinations that are crucial in the support of particular monophyletic groups, in the context of an otherwise complete data set. We apply this procedure to a real data set, with the goal of identifying the characters and character combinations that support a controversial monophyletic assemblage that has emerged in our phylogenetic studies of grasses. Among our results is the observation that a monophyletic group resolved in an analysis may fail to be resolved when characters that are *not* synapomorphies of that group are removed from the analysis. This observation emphasizes the com-

plexity of character interactions in phylogenetic analysis, and indicates that "support" for a monophyletic group derives from a complete data set, not simply from the synapomorphies of that group.

MATERIALS AND METHODS

Our example is based on a previously published data set (Soreng et al. 1990) involving chloroplast DNA restriction site variation in the grass family (Poaceae). All characters represent the presence or absence of a restriction site and thus are binary. For the purpose of clarity the data set was reduced in size by the removal of: 1) cladistically uninformative characters (either invariant or present as one state or the other in only one taxon); 2) taxa for which there is a missing value for any character; and 3) all but a single taxon from groups that are uniform for the characters under study. These adjustments resulted in a set of 27 terminal taxa scored for 74 characters (Appendix 1). Methods employed in generating the data, and accession information for the plants that exhibit these plastome types, are provided by Soreng et al. (1990).

A baseline analysis using the complete data set was conducted initially to generate a comprehensive set of most-parsimonious cladograms. The remainder of the analysis then focused on character support for a particular monophyletic group (see below) that occurred in all most-parsimonious cladograms. We analyzed the stability of this group by conducting analyses with subsets of the data obtained by removing individual characters (i.e., for $n = 74$ characters, 74 different analyses were conducted, each with $n - 1 = 73$ characters) and all combinations of two characters (i.e., for $n = 74$ characters, removal of characters 1 and 2, 1 and 3, . . . 1 and 74; 2 and 3, 2 and 4, . . . 2 and 74; . . . 73 and 74, for a total of $73 + 72 + 71 + . . . + 1 = 2701$ analyses, each involving a different combination of 72 characters). Additional analyses involving selected three-character removals also were conducted (see below).

Cladistic analyses were conducted using HENNIG86 (Farris 1988); character distributions on cladograms were examined and figures were generated using CLADOS (Nixon 1991). The baseline analysis with the complete data set, and the 74 analyses using 73 characters each, were conducted with the HENNIG86 command

ie* (implicit enumeration of all possible trees, with all available "tree space" in memory used to store results). Analyses involving the removal of two or more characters were conducted with the command mhennig* (construction of several trees, each by a single pass of the data, followed by branch-swapping), then bb* (extensive branch-swapping applied to trees obtained from previous step, use of all available "tree space" to store results); results of particular interest were verified with ie*. In all cases, the recognition of a monophyletic group was determined by its occurrence in a strict consensus tree (using the command nelsen) of all most-parsimonious trees detected.

RESULTS AND DISCUSSION

The initial analysis of the complete 74-character data set, with *Oryza sativa* L. as the outgroup, yielded 10 most-parsimonious cladograms of 130 steps and consistency index 0.57; one of these cladograms is illustrated in Figure 1. The trichotomy at the base separates *Oryza* (Oryzoideae) from two monophyletic groups: *Danthonia* with *Eragrostis* (representing subfams. Arundinoideae and Chloridoideae); and a second group that includes all other taxa in the analysis and corresponds in membership to a broadly defined subfam. Pooideae. Among the monophyletic groups present within Pooideae in all most-parsimonious cladograms are assemblages corresponding in membership to Clayton and Renvoize's (1986) tribes Meliceae and Stipeae, and Macfarlane's (1987) supertribes Triticodae and Poodae (minus Meliceae) (Appendix 1, Fig. 1). Also present in all most-parsimonious cladograms are the following combinations of these four groups: Poodae + Triticodae; Poodae + Triticodae + Stipeae; and Poodae + Triticodae + Stipeae + Meliceae. These results are in agreement with those presented previously on the basis of a broader sampling of species and restriction sites (Soreng et al. 1990).

One aspect of these results of particular interest is the consistent placement of Meliceae as the sister group of Stipeae + Triticodae + Poodae (Fig. 1). Previous assessment of these taxa, some involving rigorous cladistic methodology (e.g. Davis and Soreng, unpubl. data; Kellogg and Campbell 1987), suggest that Meliceae are more closely related than Stipeae to

Triticodae + Poodae, i.e., that the positions of Meliceae and Stipeae in the present analysis should be exchanged. Hence, we focused our assessment of stability on character support for the Stipeae + Triticodae + Poodae group. This assemblage is united in Figure 1 and throughout the set of most-parsimonious cladograms by characters 37 and 72; with alternative character optimizations, additional characters sometimes appear on this internode as well.

Analyses run with 73 characters (i.e., with one character removed) indicate that there is no single character that, when removed, prevents the resolution of Stipeae + Triticodae + Poodae as a monophyletic group in the consensus tree.

Of the 2701 character-pairs tested, four characters, when removed in five different pairwise combinations, result in the loss of the Stipeae + Triticodae + Poodae group. These four characters, each with corresponding character number, restriction enzyme, and map location in Soreng et al. (1990), are: 9 (11, Hpa I, 56), 25 (29, Kpn I, 91), 37 (48, Pst I, 94), and 72 (92, Sma I, 131). Two of these characters (37 and 72), as noted above, are synapomorphies of Stipeae + Triticodae + Poodae (Fig. 1), and it might have been predicted that their removal from the data set would result in the loss of this clade. The other two characters (9 and 25) are synapomorphies of the less inclusive clade, Triticodae + Poodae. All four of these characters are homoplasious in all 10 most-parsimonious cladograms generated from the complete data set, and different character optimization schemes placed some of their state transformations in different locations, but gains of restriction sites 37 and 72 always are synapomorphies of Stipeae + Triticodae + Poodae, and transformations of sites 9 and 25 never are synapomorphies of this clade.

Of the six possible pairwise combinations among these four characters, only five (9 & 25, 9 & 37, 25 & 37, 25 & 72, and 37 & 72), when removed from the data set, cause the resolution of Stipeae + Triticodae + Poodae as a monophyletic group to be lost. The combination without this effect is character 9 with character 72. Of the five pairwise character combinations that must be present in this data set for the monophyly of Stipeae + Triticodae + Poodae to be resolved, one (character 9 with 25) is a pair of characters that *are not synapomorphies* of this group.
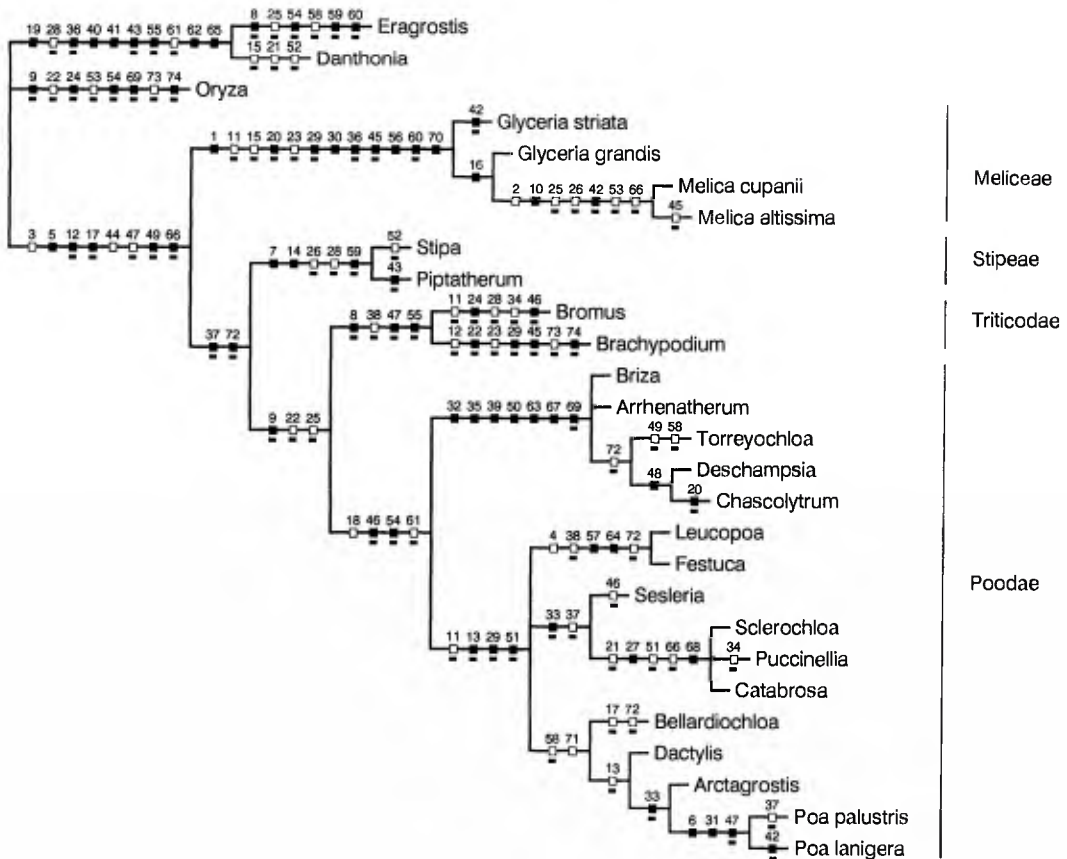
FIG. 1.   One of 10 most-parsimonious cladograms depicting variation among 27 species of Poaceae in 74 informative chloroplast DNA restriction site characters (cf. Appendix 1). Length = 130 steps, consistency index = 0.57. Solid bars signify site gains, open bars site losses; additional solid bar below a character transformation indicates that the character is homoplasious.

The final test of the stability of the Stipeae + Triticodae + Poodae clade was based upon the five 72-character subsets of the data that did not resolve this group. Beginning with each of these five subsets of the data, all possible additional single-character removals were conducted, which resulted in the testing of all three-character combinations that included the five critical two-character combinations.

For two of the five character-pairs tested in this manner (9 and 25; 9 and 37), the removal of a third character sometimes resulted in the reappearance of a monophyletic Stipeae + Triticodae + Poodae (i.e., the clade is *not* resolved when only characters 9 and 25 are removed, but it *is* resolved when the data set lacks these two and any one of the following: 56, 60, or 73; it is also resolved when the data set lacks char-

acters 9 and 37 and any one of the following: 3, 5, 6, 19, 20, 21, 24, 27, 32, 33, 34, 36, 40, 41, 42, 44, 45, 50, 51, 52, 56, 60, 61, 63, 64, 65, 67, 69, 70, 71, 73, 75, 76). The "revival" of this clade was verified for all of these three-character combinations with the HENNIG86 command ie*. For the remaining three-character pairs tested (25 and 37; 25 and 72; and 37 and 72), the removal of a third character never resulted in the reappearance of a monophyletic Stipeae + Triticodae + Poodae.

Having determined that four characters were crucial to the resolution of Stipeae + Triticodae + Poodae, we examined other attributes of these characters. Two of these restriction sites (25 and 37) occur within the largest of the three inversions that are found in the large single-copy (LSC) region of the chloroplast genome of

grasses, both of them near endpoints of the two smaller inversions (Hiratsuka et al. 1989; Quigley and Weil 1985; Soreng et al. 1990). This region has been identified previously as one of five in the grass chloroplast genome that exhibit elevated levels of restriction site variation (Soreng et al. 1990). The observation that these regions exhibited unusually high levels of variation was initially made on the basis of numbers of variable restriction sites, not on the basis of homoplasy levels of these sites.

Of the two remaining critical restriction sites, one (character 9) also occurs in the LSC region. Although not in the area of the inversions, this site occurs in another of the five previously identified regions of elevated restriction site variation. The fourth character (number 72) is in the small single-copy region, and is not in an area of noted restriction site variation.

Three of the four restriction sites that are most critical in support of a monophyletic Stipeae + Triticodae + Poodae, thus, are located in previously identified highly variable regions of the chloroplast genome; if mutation rates are unusually high in these regions, homoplasious sites may be particularly abundant. Having made this observation, we note that the restriction sites in these highly variable regions support other groupings elsewhere in the cladogram, often in congruence with sites from less variable regions. Programmatic removal from the data of all sites in highly variable regions (or a priori downweighting), therefore, might yield any number of additional alterations in the results. We emphasize, then, that the enumeration of these critical sites is not intended to suggest that their veracity as phylogenetic indicators be discounted. Rather, our goal is to demonstrate that Sequential Character Removal allows for the identification of those characters that play substantial roles in the support of particular monophyletic groups. In a data set comprising morphological characters, it might be discovered that a particular clade is supported predominantly by seed characters, pollination-related characters, or characters with some other common feature suggesting nonindependence (cf. Farris 1983). Attentiveness to the roles of individual characters in the recognition of particular clades might identify those that warrant reexamination with particular care (Frohlich 1987). Whether or not the focus provided by this procedure ultimately results in alterations in char-

acter scoring, or in the decision to gather additional data, the investigator is made aware of the particular characters and possibly character types that support the recognition of particular monophyletic groups.

Two general observations emerge from our analysis. First, we have demonstrated with a real data set that character support for the monophyly of a group of taxa need not derive entirely from that group's synapomorphies. In a data set with no inconsistency, support for monophyletic groups is indeed equivalent to each group's synapomorphies. However, when there is inconsistency among characters, as in the present case, the groups that are resolved are the products of complex interactions among characters; a character transformation in one location may influence perceived relationships among taxa in a different part of the cladogram (Neff 1987; Nixon and Davis 1991; Rohlf 1984). Thus, when there is homoplasy in a data set, character support for individual monophyletic groups is not adequately summarized by the simple enumeration of each group's synapomorphies.

Our second general observation, also a consequence of character interaction, is that a clade that is resolved by a complete data set, but is not resolved when one or more characters are removed, may again appear when additional characters are removed. Cladistic structure results from a delicate interplay of idiosyncratic and often mutually inconsistent hypotheses of character homology. Systematists should distinguish between average effects of severe and randomized perturbations (as in bootstrap analysis) and threshhold effects of minimal defined perturbations (as in Sequential Character Removal).

Many practitioners routinely examine the results of analyses conducted with subsets of their data, and some, perhaps, have conducted exercises much like the one we have described. Sequential Character Removal can proceed through the removal of sets of three, four, or more characters. Each successive round (until half of the characters have been removed) is more computationally demanding than the previous round, for it involves the testing of a greater number of character combinations. The number of cladistic analyses that must be conducted in each round is determined by the number of characters, whereas the time consumed

by each analysis within a round is determined principally by the number of taxa. Therefore, the depth to which any particular analysis can be taken is dependent upon both of these factors. However, it may frequently be the case, as in the present example, that the first few rounds of character removal identify a set of pivotal characters whose importance might not be evident from character optimizations on most-parsimonious cladograms derived from the complete data set.

In recent years, several methods have been presented for the evaluation of information content and related attributes of cladistic data sets (e.g., Archie 1989a, 1989b, 1989c; Faith 1991; Faith and Cranston 1991; Felsenstein 1985, 1988; Goloboff 1991a, 1991b; Huelsenbeck 1991; Penny and Hendy 1986; Sanderson 1989). Some of these methods generate indices of relative support for particular groupings of taxa. Others involve statistical tests for departure of data sets from random structure, and some authors speak of *confidence* in monophyletic groups supported by data sets for which the null hypothesis of randomness is rejected. We would note that departure from randomness by a cladistic data set, by virtue of congruence among characters, may take many forms, including the presence of two or more groups of characters that support conflicting phylogenetic relationships (Farris 1991). Furthermore, character congruence might be caused by any number of possible factors other than common descent (e.g., mutational biases, concerted evolution, parallel selection for particular character syndromes). Thus it is one thing to demonstrate that a data set is nonrandom in structure, and quite another to argue that the structure in the data is an accurate reflection of patterns of shared ancestry. We agree with Albert (1991) and others that "confidence" tests of phylogenetic hypotheses are inappropriate. A bootstrap probability, as an index of stability, is an assessment of support within a particular data set for a particular cladistic grouping; biases in the data unrelated to common descent may result in high bootstrap frequencies for groupings that are not reflective of phylogenetic history (Nei 1991). It is consistent with this position that we have presented Sequential Character Removal as a means for assessing the influence of particular characters—alone and in combination—on the stability of particular clades, not for determining whether the clades

are accurate depictions of descent relationships. Analyses of characters that support particular groupings, with attention to possible common causes, may help to generate hypotheses of common cause other than shared ancestry. Thus, Sequential Character Removal is an avenue for exploration of potential bias in a cladistic data set; as such, its use is antithetical to the assumption that departure from randomness implies that phylogenetic groupings are historically true. In the present study we have observed that three of the characters that support the phylogenetic grouping of Stipeae + Triticodae + Poodae occur in highly variable regions of the chloroplast genome, two of them near the endpoints of inversions. This suggests that molecular characters, like those of morphology, should be regarded as individual putative homologues, and that congruence among these characters may be attributable to parallel responses to common influences.

Attentiveness to the assumptions and methodology of phylogenetic analysis has highlighted the need for objective interpretation of data within a rigorous analytical framework, but may have had the unfortunate effect of diminishing the focus on the basic homology assessments that underlie empirical studies. We believe that the continued development of methods for the identification of critical characters and character combinations, and the phylogenetic groupings they support, will help to maintain attention on the preeminent role of character analysis in phylogenetics.

LITERATURE CITED

ALBERT, V. A. 1991. Misapplication of the bootstrap in phylogenetic reconstruction. American Journal of Botany 78 (suppl to 6): 162–163 (abstract).
ARCHIE, J. W. 1989a. Phylogenies of plant families: A demonstration of phylogenetic randomness in DNA sequence data derived from proteins. Evolution 43: 1796–1800.
———. 1989b. A randomization test for phylogenetic information in systematic data. Systematic Zoology 38: 239–252.
———. 1989c. Homoplasy excess ratios: New indi-

ces for measuring levels of homoplasy in phylogenetic systematics and a critique of the consistency index. Systematic Zoology 38: 253–269.

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

CLAYTON, W. D. and S. A. RENVOIZE. 1986. *Genera graminum*. Kew Bulletin Additional Series 13. London: HMSO.

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

——— and P. S. CRANSTON. 1991. Could a cladogram this short have arisen by chance alone?: On permutation tests for cladistic structure. Cladistics 7: 1–28.

FARRIS, J. S. 1983. The logical basis of phylogenetic analysis. Pp. 7–36 in *Advances in cladistics*, vol. 2, eds. N. I. Platnick and V. A. Funk. New York: Columbia Univ. Press.

———. 1988. HENNIG86, ver 1.5. Port Jefferson Station, New York: Publ. by the author.

———. 1991. Excess homoplasy ratios. Cladistics 7:81–91.

FELSENSTEIN, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39: 783–791.

———. 1988. Phylogenies from molecular sequences: Inference and reliability. Annual Review of Genetics 22: 521–565.

FROHLICH, M. W. 1987. Common-is-primitive: A partial validation by tree counting. Systematic Botany 12: 217–237.

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

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

HIRATSUKA, J., H. SHIMADA, R. WHITTIER, T. ISHIBASHI, M. SAKANOTO, M. MORI, C. KONDO, Y. HONJI, S.-R. SUN, B.-Y. MENG, Y.-Q. LI, A. KANNO, Y. NISHIZAWA, A. HIRAI, K. SHINOZAKI, and M. SUGIURA. 1989. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: Intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Molecular and General Genetics 217: 185–194.

HUELSENBECK, J. P. 1991. Tree-length distribution skewness: An indicator of phylogenetic information. Systematic Zoology 40: 257–270.

KELLOGG, E. A. and C. S. CAMPBELL. 1987. Phylogenetic analyses of the Gramineae. Pp. 310–322 in *Grass systematics and evolution*, eds. T. R. Soderstrom, K. W. Hilu, C. S. Campbell, and M. E. Barkworth. Washington, D.C.: Smithsonian Inst.

MACFARLANE, T. D. 1987. Poaceae subfamily Pooideae. Pp. 265–276 in *Grass systematics and evolution*, eds. T.R. Soderstrom, K. W. Hilu, C. S. Campbell, and M. E. Barkworth. Washington, D.C.: Smithsonian Inst.

MADDISON, W. 1989. Reconstructing character evolution on polytomous cladograms. Cladistics 5: 365–377.

MICKEVICH, M. F. 1978. Taxonomic congruence. Systematic Zoology 27: 143–158.

———. 1980. Taxonomic congruence: Rohlf and Sokal's misunderstanding. Systematic Zoology 29: 162–176.

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

MUELLER, L. D. and F. J. AYALA. 1982. Estimation and interpretation of genetic distance in empirical studies. Genetical Research 40: 127–137.

NEFF, N. A. 1987. An analysis of the sensitivity of minimum length (Wagner) tree topology to changes in data. Systematic Zoology 36: 227–236.

NEI, M. 1991. Relative efficiencies of different tree-making methods for molecular data. Pp. 90–128 in *Phylogenetic analysis of DNA sequences*, eds. M. M. Miyamoto and J. Cracraft. New York: Oxford Univ. Press.

NIXON, K. C. 1991. CLADOS, ver. 1.0. Trumansburg, New York: Publ. by the author.

——— and J. I DAVIS. 1991. Polymorphic taxa, missing values and cladistic analysis. Cladistics 7: 233–241.

PENNY, D. and M. HENDY. 1986. Estimating the reliabilty of evolutionary trees. Molecular Biology and Evolution 3: 403–417.

QUIGLEY, F. and J. H. WEIL. 1985. Organization and sequence of five transfer RNA genes and of an unidentified reading frame in the wheat chloroplast genome: Evidence for gene rearrangements during the evolution of chloroplast genomes. Current Genetics 9: 495–503.

ROHLF, F. J. 1984. A note on minimum length trees. Systematic Zoology 33: 341–343.

——— and R. R. SOKAL. 1980. Comment on taxonomic congruence. Systematic Zoology 29: 97–101.

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

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.

APPENDIX 1. Presence/absence of 74 variant chloroplast DNA restriction sites among 27 grass species. Species numbered as in Soreng et al. (1990); sites renumbered consecutively following removal of uninformative and other sites (see text).

	Character numbers							
	1 1234567890	2 1234567890	3 1234567890	4 1234567890	5 1234567890	6 1234567890	7 1234567890	1234
Subfam. Oryzoideae								
3b. *Oryza sativa* L.	0111000010	1000100100	1011110100	0001000100	0001001000	0101000100	1000000010	1001
Subfam. Chloridoideae								
1. *Eragrostis* sp.	0111000100	1000100110	·1110010000	0001010101	1011001000	0111100011	0100100000	1010
Subfam. Arundinoideae								
2. *Danthonia californica* (L.) Beauv.	0111000000	1000000110	0110110000	0001010101	1011001000	0010100100	0100100000	1010
Subfam. Pooideae								
Tribe Meliceae								
19. *Glyceria grandis* S. Watson	1101100000	0100011101	1100110111	0001010100	0000100010	0110010101	1000010001	1010
20. *Glyceria striata* (Lam.) A. Hitchc.	1101100000	0100001101	1100110111	0001010100	0100100010	0110010101	1000010001	1010
21. *Melica altissima* L.	1001100001	0100011101	1100000111	0001010100	0100000010	0100010101	1000000001	1010
22. *Melica cupanii* Guss.	1001100001	0100011101	1100000111	0001010100	0100100010	0100010101	1000000001	1010
Tribe Stipeae								
40. *Piptatherum miliaceum* (L.) Cosson	0101101000	1101101100	1110100000	0001001100	0010000010	0110000110	1000010000	1110
42. *Stipa barbata* E. Desv.	0101101000	1101101100	1110100000	0001001100	0000000010	0010000110	1000010000	1110
Supertribe Triticodae								
17. *Brachypodium pinnatum* (L.) Beauv.	0101100110	1000101100	1100010110	0001001000	0000101010	0110100100	1000010000	1101
18. *Bromus inermis* Leysser	0101100110	0100101100	1011010000	0000001000	0000011010	0110100100	1000010000	1110

APPENDIX 1. Continued.

	Character numbers							
	1 1234567890	2 1234567890	3 1234567890	4 1234567890	5 1234567890	6 1234567890	7 1234567890	1234
Supertribe Poodae								
Tribe Aveneae								
7. *Arrhenatherum elatius* (L.) Beauv. ex J. Presl	0101100010	1100101000	1010010100	0101101110	0000010011	0111000100	0010011010	1110
9. *Deschampsia cespitosa* Beauv.	0101100010	1100101000	1010010100	0101101110	0000010111	0111000100	0010011010	1010
12. *Chascolytrum erectum* (Lam.) Desv.	0101100010	1100101001	1010010100	0101101110	0000010111	0111000100	0010011010	1010
11. *Briza minor* L.	0101100010	1100101000	.1010010100	0101101110	0000010011	0111000100	0010011010	1110
15. *Torreyochloa erecta* (A. Hitchc.) Church	0101100010	1100101000	1010010100	0101101110	0000010001	0111000000	0010011010	1010
Tribe Poeae								
25. *Arctagrostis poaeoides* Nash ex Britton & Rydb.	0101100010	0100101000	1010010110	0011001100	0000010010	1111000000	0000010000	0110
26. *Bellardiochloa violacea* (Bellardi) Chiov.	0101100010	.	1010010110	0001001100	0000010010	1111000000	0000010000	0010
27. *Catabrosa aquatica* (L.) Beauv.	0101100010	0110100000	1010010110	0011000100	0000010010	0111000100	0000000100	1110
28. *Dactylis glomerata* L.	0101100010	0110101000	0010011110	0001001100	0000010010	1111000000	0000010000	0110
29. *Festuca arundinacea* Schreber	0100100010	0100101000	1010010110	0001001000	0000010010	1111001100	0001010000	1010
32. *Leucopoa sclerophylla* (Boiss. & Hohen.) Krecz. & Bobrov	0100100010	0110101000	1010010110	0001001000	0000010010	1111001100	0001010000	1010
34. *Poa langera* Nees	0101110010	0110101000	1010010110	1011001100	0100011010	1111000000	0001010000	0110
35. *Poa palustris* L.	0101110010	0100101000	1010010110	1011000100	0000011010	1111000000	0000010000	0110
36. *Puccinellia distans* (L.) Parl.	0101100010	0110101000	0010011110	0010000100	0000010010	0111000100	0000000100	1110
38. *Sclerochloa dura* (L.) Beauv.	0101100010	0110101000	0010011110	0011000100	0000010010	0111000100	0000000100	1110
39. *Sesleria coerulea* (L.) Ard.	0101100010	0110101000	1010010110	0011000100	0000000010	1111000100	0000010000	1110