THE BOTANICAL REVIEW

VOL. 59 OCTOBER–DECEMBER, 1993 NO. 4

Phylogenetic Studies of a Large Data Set. I. Bambusoideae, Andropogonodae, and Pooideae (Gramineae)

ELIZABETH A. KELLOGG

Arnold Arboretum of Harvard University
22 Divinity Avenue
Cambridge, MA 02138

LESLIE WATSON

Taxonomy Laboratory
Research School of Biological Sciences
Australian National University
Canberra, ACT 2601

Abstract

Large data sets, with several hundred terminal taxa, are becoming increasingly common in phylogenetic studies, but are proving very difficult to analyze because existing algorithms cannot explore the enormous number of trees efficiently. This article presents the results of an ongoing project to carry out phylogenetic analyses on a data base with 760 terminal taxa, the genera of the grass family (Gramineae), initially scored for more than 400 morphological and anatomical characters. The approach consists of three steps: (1) Using a small number of highly consistent characters, determine which large groups are demonstrably monophyletic and which may be polyphyletic. Treat the large monophyletic groups as single terminal taxa, and focus on the overall structure of the entire group. This results in a tree that links the large monophyletic "black boxes" with smaller basal groups; the latter can then be taken as provisional outgroups. (2) Use the outgroups defined in step 1 to analyze the cladistic structure of the big monophyletic groups, with a much larger set of characters. (3) Use a cladistically-guided sample of basal taxa from each large clade to redo the family-level analysis. Kellogg and Campbell (1987) carried out step 1 for the grass family and defined four monophyletic groups (subfamilies) that were derived from within a highly polyphyletic assemblage of genera. This article reports on step 2, analyses of three of the four monophyletic groups, the pooid clade (184 genera), the bambusoid clade (166 genera), and the Andropogonodae plus Arundinelleae (121 genera). 150–220 characters per clade, a much larger number than commonly used in morphological studies, were chosen from the comprehensive database. The initial descriptions of characters, their division into states, and their application to particular genera were the result of 20 years of work on the family by one of us (LW). Subsequent choice of characters for cladistic analysis was done by the other author (EAK) using only the pattern of variation of the character rather than the morphological descriptor, thus eliminating possible bias from a priori ideas of a character's value. Each clade was analyzed in two ways, (1) with all terminal taxa for which there were adequate data, and (2) for only mono- and ditypic taxa; the latter analysis was to minimize the effect of possibly polyphyletic genera. In all cases, the reduced data set produced groups similar to those of the entire data set. The bambusoid clade consists of several well-defined subclades corresponding approximately to previously-recognized tribes. The relationships among the subclades are not resolved by these data. The Andropogonodae is made up of two major groups, an awned group and an awnless group; the latter includes taxa previously included in the Maydeae and also genera conventionally assigned to the Rottboelliinae. The pooid clade exhibits high homoplasy and

no robust cladistic relationships. This is not likely to be caused by problems with generic circumscription, but may reflect extensive lateral gene flow (hybridization), rapid radiation followed by extensive anagenetic change, or true parallelism in morphological characters. The traditional tribes of the Pooideae are, with the exception of the Triticeae, apparently not monophyletic. Morphological cladograms are evaluated in light of data from molecular characters; while the results are generally consistent, there are too few molecular data yet to make meaningful comparisons.

Resumen

Grupos de datos grandes, con varios cientos de taxones terminales, son cada día más comunes en estudios filogenéticos. Sin embargo su análisis ha resultado ser muy difícil dado que los algoritmos existentes no pueden explorar eficientemente el enorme número de árboles. Este artículo presenta los resultados de un proyecto en progreso encaminado a realizar análisis filogenéticos en una base de datos de 760 taxones terminales, los géneros de la familia de las gramíneas, evaluados inicialmente en más de 400 caracteres morfológicos y anatómicos. El procedimiento consiste en tres pasos: (1) Usando un pequeño número de caracteres últamente consistentes, determinar para cuales grupos grandes existe fuerte evidencia de que estos sean monofiléticos y cuales puedan ser polifiléticos. Considerar los grandes grupos monofiléticos como taxones terminales unitarios, y hacer énfasis en la estructura general del grupo entero. El resultado es un árbol que une las "cajas negras" monofiléticas con grupos basales más pequeños; estos últimos pueden entonces tomarse como grupos externos provisionales. (2) Utilizar los grupos externos definidos en el paso 1 para analizar la estructura cladística de los grupos monofiléticos grandes con un número de caracteres mucho mayor. (3) Utilizar una muestra de taxones basales, inferida cladísticamente, de cada uno de los clados grandes para rehacer el análisis a nivel de familia. Kellogg y Campbell (1987) realizaron el paso 1 para la familia de las gramíneas y definieron cuatro grupos monofiléticos (subfamilias), derivados a partir de un ensamblaje de géneros últamente polifilético. Este articulo presenta un reporte del paso 2, los análisis de tres de los cuatro grupos monofiléticos: el clado pooide (184 géneros), el bambusoide (166 géneros) y juntas la supertribu Andropogonodae y la tribu Arundinelleae (121 géneros). Unos 150–220 caracteres por clado, un número mucho mayor de lo comúnmente usado en estudios morfológicos, fueron escogidos de la base de datos completa. La descripción inicial de los caracteres, su división en dos atributos y su aplicación a cada género fue el resultado de 20 años de trabajo en la familia por parte del segundo autor (LW). La subsiguiente selección de caracteres para el análisis cladístico fue llevada a cabo por la primera autora (EAK) utilizando únicamente el patrón de variación del carácter envés de su significado morfológico, eliminando así el posible sesgo proveniente de ideas a priori acerca del valor de un determinado carácter. Cada clado fue analizado de dos maneras, (1) con todos los taxones terminales para los cuales existían datos adecuados, y (2) únicamente con taxones mono- y ditípicos; este último análisis, con el fin de minimizar el efecto de géneros que puedan ser polifiléticos. En todos los casos, la muestra reducida de datos produjo grupos similares a aquellos obtenidos con el juego completo. El clado bambusoide esta compuesto por varios subclados bien definidos, correspondiendo aproximadamente a las subtribus reconocidas en el pasado. No fue posible resolver las relaciones entre los subclados con estos datos. La supertribu Andropogonodae esta compuesta por dos grandes

grupos, uno aristado y otro sin aristas; este último incluye taxones anteriormente incluidos en la tribu Maydeae, además de géneros convencionalmente asignados a la subtribu Rottboelliinae. El clado pooide demuestra alta homoplasia y relaciones cladísticas no robustas. Es muy probable que lo anterior no sea el producto de problemas de circunscripción genérica, sino que puede reflejar, bien sea, un gran flujo genético lateral (hibridización), radiación rápida seguida de cambio genético anagénico extenso, o verdadero paralelismo en caracteres morfológicos. Las tribus tradicionales de la subfamilia Pooideae, exceptuando la tribu Triticeae, aparentemente no son monofiléticas. Los cladogramas morfológicos fueron evaluados a la luz de información proveniente de caracteres moleculares; mientras los resultados son generalmente consistentes, por lo pronto existen muy pocos datos moleculares como para hacer comparaciones significativas. (Translation kindly provided by S. Madriñan.)

I. General Introduction

Understanding evolutionary process requires a detailed knowledge of evolutionary pattern. Phylogeny reconstruction is thus fundamental to all evolutionary biology. More and more methods have been developed for phylogenetic analysis over the last twenty years, in part because of the increasing availability of computing power, and they continue to be improved.

Large data sets, however, continue to create problems for phylogenetic studies. Limitations of algorithmic power and of data quality affect the sort of analyses that can be done, the resolution of the analyses, and the interpretation of the results. Most phylogenetic studies are done on groups with fewer than 50, and often fewer than 30, terminal taxa (cf. Sanderson & Donoghue, 1989). Methodological studies are of necessity done on even fewer taxa, sometimes as few as four (see for example, Lake, 1987; Li & Guoy, 1992; or any recent issue of Cladistics), on the implicit assumption that a larger analysis is simply a matter of scaling up. Frequently, however, studies with only a few taxa suffer from the "long branch problem"; this is particularly true for molecular data (cf. Allard & Miyamoto, 1992, among others). According to Swofford and Olsen (1990):

"With a large number of taxa, correctly inferring every aspect of the true topology is extremely difficult, but if we were interested in the relationships of, say, only four taxa, we would be much better off to compute a tree for 20 taxa and prune 16 of them from the tree than to compute the tree for only the four taxa. As an aside, we note that for this reason, the behavior of a method may be quite different for a four-taxon data set than for a larger one." (Swofford & Olsen, 1990, p. 497)

Furthermore, considerations of global parsimony suggest that larger analyses might be appropriate in some cases; Maddison et al. (1984) have shown that the locally most parsimonious tree should in theory be obtained by a global analysis. This is not always true in practice because the global tree is frequently only an approximation of the shortest tree, and often reflects a local, rather than global, minimum length. Thus, although large analyses may be theoretically desirable, in practice, investigators usually try to create minimally-sized data sets. Unfortunately, large groups may not always be readily reduced in size by a priori subdivision into monophyletic subgroups; in addition, sampling of large groups creates its own problems, some of which will be discussed here.

This article reports on an ongoing study of a data set with 760 terminal taxa, which

makes it the largest phylogenetic project undertaken to date (at least by a small team of workers). The number of characters (150–220) is much larger than commonly used for morphological analyses, making it comparable to a molecular data matrix in size. Characters in the database were initially chosen as accurate descriptors of specimens; characters were then selected for cladistic analysis not because of any a priori belief in their utility, but rather by their pattern of variation alone. The study group is the grass family (Gramineae), a family of perhaps 8–10,000 species and (as of 1988) 760 genera (Watson & Dallwitz, 1988). The data matrix includes all genera scored for 430 characters, including data on gross morphology, leaf epidermal and cross-sectional anatomy, and cytology. The objective of the study is to test the limits of the existing data for the family; the resolution of the trees reflects most importantly the utility of the morphological and anatomical characters commonly used in grass taxonomy, but is also influenced by the completeness of the data and the monophyly of the existing genera. By doing exploratory cladistic analyses on the entire data set, we hope to determine which characters are likely to be informative and which need additional study, which clades are robust and which need more investigation, and which taxa are insufficiently understood and/or are critical to determining a well-supported phylogeny. At the same time we hope to clarify some of the problems that are peculiar to large data sets, and suggest some possible methodological solutions.

Ideally, all 760 genera would be included as terminal taxa in a cladistic analysis and a few parsimonious trees would emerge. However, existing computer algorithms are not powerful enough to search such a data set even heuristically in a reasonable length of time. Chase et al. (1993) note that one run of their 499-taxon data set on rbcL took 200 hours on Sun Workstation, a time comparable to that of each of the numerous analyses described in this paper. For such data sets detailed evaluations of the resulting trees by any of the many resampling techniques now available [e.g., bootstrap (Felsenstein, 1985), or randomization tests (Archie, 1989; Faith & Cranston, 1991)] would be virtually impossible. The general approach to the large analysis has therefore been as follows:

1. Using a small number of highly consistent characters, determine which large groups are demonstrably monophyletic and which may be polyphyletic. Treat the monophyletic groups as black boxes, and focus on the structure of the family. This phase of the study was done by Kellogg and Campbell (1987), who published two cladograms for the Gramineae differing in the stringency of their underlying assumptions, particularly as regards the monophyly of the tribe Danthonieae (Fig. 1). They demonstrated that the four subfamilies Pooideae, Bambusoideae, Panicoideae and Chloridoideae are likely to be strictly monophyletic, although the monophyly of the chloridoids is not strongly supported. The fifth subfamily, the Arundinoideae, had long been known to be phenetically heterogeneous and appeared to be polyphyletic, but was also problematical because of uncertainty about the limits of many genera.

2. Use the outgroups thus defined to analyze the cladistic structure of the monophyletic subfamilies, with a much larger set of characters. This article reports on the analyses for the bambusoid clade, the pooid clade, and the Andropogonodae, a supertribe that includes the Maydeae and Andropogoneae and forms a large subgroup in the panicoid clade. A subsequent paper will deal with the chloridoid clade and the remainder of the panicoids.

3. Use a sample of basal taxa from each clade to reanalyze at the family level. This analysis will constitute a future publication and will also incorporate information from a redefinition of the genera of the Arundinoideae (Linder, Ellis, & Barker, in prep.).

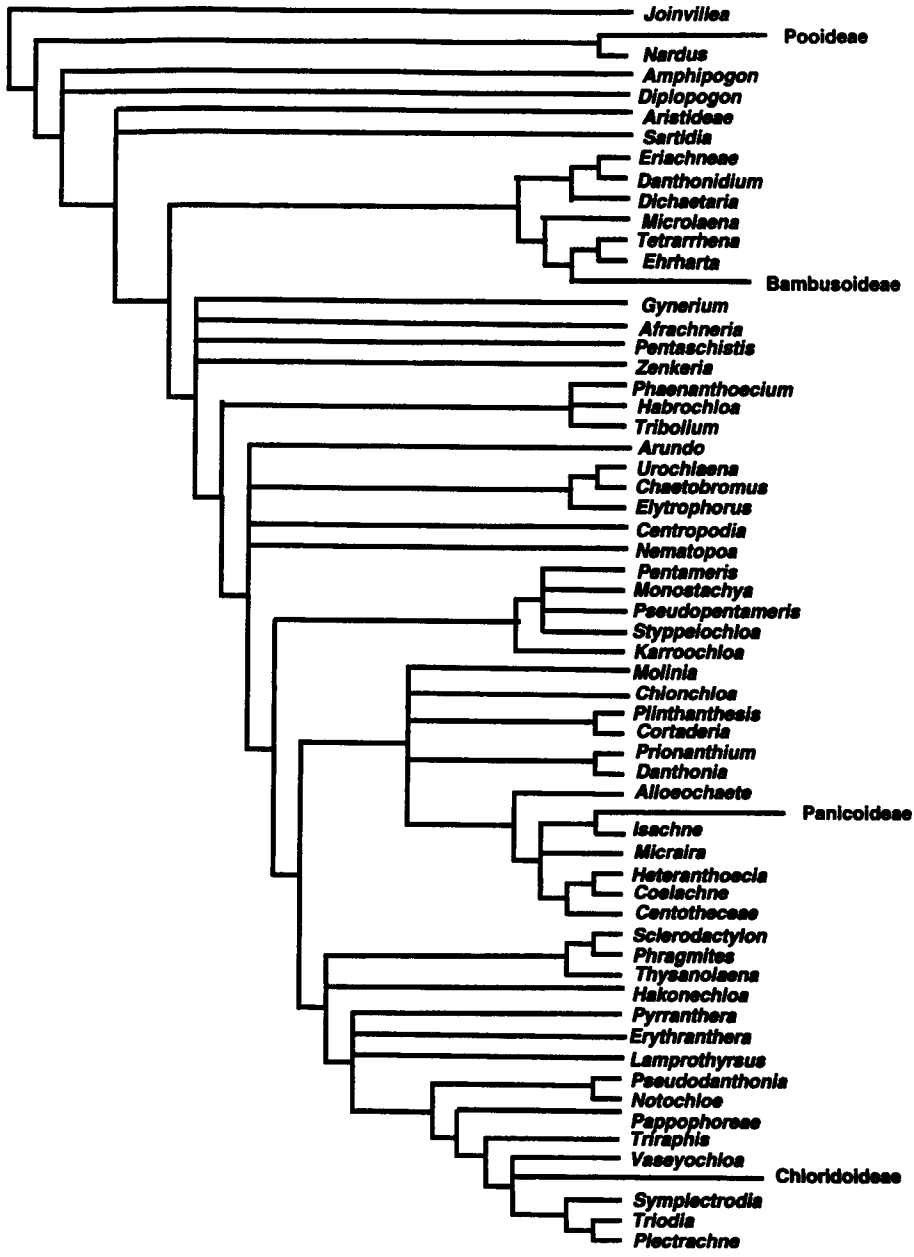

Fig. 1. One phylogenetic hypothesis for the Gramineae, redrawn from Kellogg and Campbell (1987). Tree is rooted at *Joinvillea* (Joinvilleaceae). Pooideae, Bambusoideae, Panicoideae, and Chloridoideae are monophyletic subfamilies. All other genera are commonly placed in the Arundinoideae. Tree length = 106; CI = 0.367.

Like the work of Kellogg and Campbell (1987), this article relies on information contained in an interactive data base, "Grass Genera of the World" (Watson & Dallwitz, 1988), compiled over a period of some 20 years by one of us (LW) and colleagues. The data were initially taken from the literature but have been extensively augmented by original observations. Characters and character states have been evaluated and redefined when necessary, based on continuing study of the grass genera. The information has been assembled in a data base using the programs in the DELTA package (DEscriptive Language for TAxonomy; Dallwitz, 1980). As of 1988, the data base contained 760 terminal taxa (genera) scored for 430 characters; the number of characters is now much larger. The data are easily accessible in machine-readable form, and are also comprehensive, so are easily restructured for phylogenetic analysis. There are three major limitations of the data base, which also apply to any of the molecular data bases, such as GenBank. First, it has been compiled from many published papers and includes data from at least a dozen other contributors in addition to the data produced by the Watson lab; this means that the data are generally reliable and constantly improving, but the quality is not absolutely consistent. Continual use and refinement of the data base over a period of years have mitigated this problem. Second, the data base is designed to summarize information at the generic level, and thus its quality depends on the quality of generic circumscription in the family; this is less of a problem than it might be because grass genera tend to be small relative to those in many other plant groups. Over 400 of the 760 genera have only one or two species, so for the majority of taxa, generic characters are actually characters of single species. Third, the primary objective of the data base was to create a tool for multiple purposes rather than for phylogenetic analysis per se. The data thus had to be extensively reinterpreted for cladistic analysis, as described in section IIB below.

The most accurate picture of phylogeny is obtained from analyses that include as many of the relevant taxa as possible. This was shown clearly by Donoghue et al. (1989) in a set of analyses in which data sets including fossils were compared with those excluding them—the picture of evolution was substantially different (and presumably more accurate) when all taxa, both extant and extinct, were included. We have therefore avoided restrictive sampling of large complex groups (such as the Pooideae), primarily because, as illustrated in Section IIIC, such samples tend to give artefactually "clean" cladograms and a false sense of security. Not only is consistency index (CI) heavily influenced by number of taxa (Sanderson & Donoghue, 1989), but cladogram resolution may be affected as well. Some of the added taxa will exhibit different character correlations, apparent patterns of concordance will be disrupted and the cladograms will become less resolved. We prefer an unresolved, confusing cladogram that accurately reflects the data to a clear but misleading one. Nonetheless, we have not included fossils because the available descriptions of most grass fossils have too little information available about one or more major character systems. They thus are like any of the other taxa with missing data (see below)—although they can and should in principle be included, in practice they make it impossible to analyze the data.

The unfortunate reality is that if a group does not break neatly into monophyletic units of say, 50 terminal taxa or fewer, the "tree space" becomes difficult to analyze with any precision (cf. Maddison, 1991). Obviously this is partly a function of homoplasy, missing data, and polymorphism (which is treated here as missing data; see below). With "clean" data, the maximum number of terminal taxa could be much higher, but as Sanderson and Donoghue (1989) have shown, large data sets are rarely

"clean." In fact, the more detailed the information, the more homoplasy is revealed. The crux of the problem thus remains algorithmic. Although computer programs are constantly improving, they are not yet sufficiently powerful to handle a data set of this size. Thus taxa with missing data, including fossils, aggravate the already-considerable computational problems.

The three data subsets presented here each illustrate different problems, requiring different solutions. The first, the bambusoid clade, is a data set of 42 terminal taxa (after exclusion of many poorly known genera of woody bamboos) with clear hierarchical structure. Problems of generic limits plague one clade (the woody bamboos), but the internal cladistic structure of this group appears to have little effect on the rest of the analysis. The second data set, the Andropogonodae is much larger, although still with clear hierarchical structure. Its size alone makes it harder to work with but there is comparatively little character conflict. The third data set, the pooid clade, is the most difficult, both because of its size, which is comparable to the Andropogonodae, but also because of its extensive homoplasy and minimal hierarchical structure. Tests to explore the amounts of "signal" and "noise" are complicated, and in most cases rendered impracticable, by the large size of the data set.

The problems of analysis encountered in this study are paralleled in molecular data sets, and similar solutions may be applicable. There are now well over 500 rbcL sequences (1428 bp each) available for higher plants (Chase et al., 1993). The database of the genes for the small subunit of ribosomal RNA (ca. 1600–1900 bp) contains 927 sequences (DeRijk et al., 1992), and for the large subunit of ribosomal RNA (over 2000 bp), 128 (Gutell et al., 1992). Methods for studying restriction site variation in chloroplast DNA are becoming routine in plant systematics and are producing ever-larger data sets (e.g. 60 taxa × 82 characters; Bruneau & Doyle, 1993). (The utility of these however, may be limited by the difficulty of comparing restriction site maps generated by different investigators.) Molecular data sets are thus approaching the size of the grass database. Like the morphological grass data, the sequences and maps have been produced by many investigators in different labs, and thus may be presumed to be generally reliable, but with possible inaccuracies. Also like the grass data, analyses will probably have to be done in an iterative fashion to make computations more tractable. In molecular studies, as in this morphological one, there will inevitably be a trade-off of time between double-checking data for particular taxa or characters versus using more powerful algorithms for exploring cladistic structure. These points will be amplified in the General Discussion.

In addition to illustrating methodology, the morphological phylogenies presented here will provide a comparison for molecular studies in the family, expanding on the work of Kellogg and Campbell (1987), which was the first explicitly cladistic study of the family as a whole. The phylogeny of the grass family has been the subject of much discussion, particularly since the influential work of Stebbins and Crampton (1961), in which the classical view of the family (Brown, 1810, 1814) was supplanted by an interpretation based on characters of the embryo, leaf epidermis, and leaf cross sectional anatomy, as well as on cytology. This has led to the comprehensive works of Watson and Dallwitz (1988) and Clayton and Renvoize (1986); the classification of Watson and Dallwitz (1988) is phenetic, grouping by overall similarity, whereas that of Clayton and Renvoize (1986) uses a more intuitive and non-algorithmic approach.

As molecular phylogenies are generated, it will be important to have a morphological phylogeny such as those presented here against which to compare them.

Donoghue and Sanderson (1992) pointed out that "the limiting factor [in such comparisons] is likely to be the number of solid morphological analyses" (cf. also Watson, 1971). Until this work, there have been no cladistic studies of large groups in the grass family; cladograms based on molecular data cannot be properly compared to phenograms based on morphological data. Discrepancies between molecular and morphological cladograms can be used to illuminate parallelisms in one or the other set of characters [as has been done for drosophilids (DeSalle & Grimaldi, 1991)], to understand more about the nature of character state change, and possibly to illuminate questions of past hybridization and its importance in the evolution of the family.

II. General Methods

A. TAXA

Genera were initially assumed to be monophyletic, as in the work of Kellogg and Campbell (1987), but this assumption was then tested in later analyses. The assumption seems reasonable because the family is highly split (well over 400 of the genera have only one or two species), and the fact that most genera have at least one clear autapomorphy. For each of the three groups discussed in this paper, some analyses were done using only genera with one or two species to test for the confounding effects of paraphyletic groups.

Hybridization may also lead to polyphyletic genera, and is known to create problems for the reconstruction of phylogenies. Cladistic analyses assume divergent evolution, and known or suspected hybrids cannot be accommodated (cf. Kellogg, 1989; McDade, 1990, 1992; and discussion in section IVC below). In their analyses, Kellogg and Campbell (1987) argued that hybridity was not a problem because intergeneric hybridization outside the Triticeae was rare. This was based largely on Knobloch's (1968) reports of hybridization in the family, in which few intergeneric hybrids were reported. [Note that this information is now available in the grass database (Watson & Dallwitz, 1991)]. However, checking ploidy levels in the family revealed that many genera, particularly monotypic genera, have no known diploids. This may well point to a history of hybridization. Hybrids are easily segregated into new genera in a non-cladistic classification; taxa are not required to have apomorphies as evidence of monophyly, and any group with a distinctive combination of characters can be put in its own taxon (in this case a genus). The prevalence of these putatively hybrid "genera" suggests that hybridization has been more important in the evolution of the family than Kellogg and Campbell (1987) had thought. Thus, for each group, analyses either (1) included only taxa known to contain diploids, or (2) included taxa known to contain diploids as well as those of unknown ploidy.

Taxa with data missing for more than 1/3 of the characters were initially included in an effort to determine their approximate placement. For many large analyses, however, they seriously affected the ability of the program to search for trees, as many equally parsimonious trees could be found representing alternative placements for them. Taxa with more than one third of their characters unrecorded have therefore been excluded from subsequent analyses. However, these taxa represent only 10–15% of the total number of genera, and are mostly monotypic (Appendices A1–A3). It should be noted that many of them are now much more fully described than they were when the data sets were initially compiled and could be included in future analyses. [The grass database is now available via Internet (Watson & Dallwitz, 1991)].

For many genera, the character state data encompass an appreciable amount of polymorphism; in our cladistic analyses this has been rescored throughout as though the data were missing. Nixon and Davis (1991) have pointed out that such an approach can be misleading for two reasons. First, taxa with much polymorphism (missing data) may tend to be misplaced in phylogenetic analyses, in part because the extensive "missing data" imply many more possible character state combinations than actually occur. Second, levels of homoplasy and actual length of trees will tend to be underestimated because of "hidden homoplasy" created by variation within the terminal taxon. The second problem seems to us to be one that must be recognized, but then simply accepted. All terminal taxa, whether they be species, genera, or families are somewhat variable and hence some polymorphism will be included; there is hidden homoplasy in all data sets. The first problem, however, is more worrisome. It can be mitigated in part by use of "exemplar taxa" (a single specimen as proxy for the whole genus) rather than "summary taxa" (a taxonomic unit that summarizes all character states for all species in the genus). However, as will be noted below particularly for the pooid clade, samples of taxa introduce their own biases; these will be particularly misleading if a sample species or specimen happens to be a derived member of the genus it represents, in which case its placement may still be erroneous.

B. CHARACTERS

Many of the characters in the data base were included for purposes other than cladistic analyses, so characters had to be reevaluated and in some cases recoded. The method of initial character selection was the same for all clades and made extensive use of the interactive identification program in DELTA, INTKEY (Dallwitz, 1992). The taxa to be included in a particular analysis (e.g., the bambusoid clade, below) were decided upon and then all character analyses were conducted with reference to that set of genera. Given a particular set of taxa, the command "Summary" in INTKEY produces an output similar to that shown in Fig. 2, which is a summary of the character data for characters 160 through 169 of the Bambusoideae. Each character is listed by number (rather than by descriptor); the character number is followed by either the number of states or a statement that it is a continuous variable ("Real" or "Integer"). The output then lists the number of taxa for which that character is unknown (U), inapplicable (I), or recorded, giving an indication of the extent of sampling. For example, for the 123 genera of the Bambusoideae, the state for character 162 is unknown in 103 genera, inapplicable in 18 genera, and recorded in 2 genera. Finally, INTKEY lists the distribution for the group of taxa in question (e.g., for character 160, 7 genera with character state 1 and 77 with character state 2); the number of genera may add up to more than the number scored because some genera will have both character states. "Quasi-characters"—nomenclature, subfamilial and tribal assignments, etc.—were naturally excluded, as were geographic distribution data on the grounds that biogeography should not be considered when forming groups. Characters having to do with susceptibility to pathogens were also eliminated because of difficulty scoring and interpreting homologies.

Because characters were listed by number rather than by name, the remaining characters could be evaluated "anonymously," that is, with reference only to their pattern of variation and not to their actual morphological descriptor. This eliminated bias in character selection, encouraged inclusion of characters not commonly used,

Character 160 (2 states)					
U/I/Recorded	19/23/83				
Distribution of values	1(7)	2(77)			
Character 161 (Real)					
U/I/Recorded	64/58/2				
Mean	0.75				
Standard deviation	0				
Minimum	1 (Item 237)				
Character 162 (3 states)					
U/I/Recorded	103/18/2				
Distribution of values	1(2)				
Character 163 (2 states)					
U/I/Recorded	86/28/9				
Distribution of values	1(4)	2(5)			
Character 164 (2 states)					
U/I/Recorded	102/20/1				
Distribution of values	1(1)				
Character 165 (4 states)					
U/I/Recorded	0/123/0				
Character 166 (2 states)					
U/I/Recorded	0/123/0				
Character 167 (6 states)					
U/I/Recorded	103/19/1				
Distribution of values	1(1)				
Character 168 (Integer)					
U/I/Recorded	45/21/64				
Distribution of values	0(10)	1(11)	2(5)	3(17)	4(10)
	5(26)	6(12)	7(14)	8(5)	9(12)
	10(6)	11(7)	12(4)	13(5)	14–15(3)
	16(2)	17(3)	18(2)	19–30(1)	
Mean	5.925				
Standard deviation	6.184				
Character 169 (2 states)					
U/I/Recorded	0/26/105				
Distribution of values	2(105)				

Fig. 2. Sample output using the "Summary" command of INTKEY (Dallwitz, 1992). Summarizes characters 160 through 169 for the Bambusoideae. See text for discussion.

and discouraged undue reliance on well-known "important" characters. This approach, of course, relies heavily on the quality of observations and taxonomic decisions made in constructing the grass database.

Characters that did not vary within the clade were excluded, as were those that were scored in fewer than 10% of the taxa. For the characters shown in Fig. 2, #161, #162, #163, #164 and #167, were eliminated from further consideration because they are too poorly sampled. Characters #165 and #166 are inapplicable to the genera in this group. Character #169 was eliminated because it is invariant. A character state appearing only once within a clade was eliminated if it was also unique within the family. If the unique state also appeared in potential outgroups, then it could be either (a) an autapomorphic reversal or (b) the character state of the basal taxon. In each case, the taxon with the unique character state was determined. If it was potentially a basal member [based on data from Kellogg & Campbell (1987)], then the character was used for the analysis. If, however, the taxon with the character state was clearly nested within a larger clade (e.g. three spikelets per node in *Hordeum*; see below), then the character was eliminated from the analysis as being phylogenetically uninformative.

There are relatively few quantitative characters in the database (32 out of 430, or about 7%). For each quantitative character (listed as "real" or "integer" in INTKEY; e.g. characters #160 and #168 in Fig. 2), the values for each genus were retrieved and all values were graphed to check for clear breaks in the distribution. In most cases, variation was continuous, and the character was eliminated as recommended by Pimentel and Riggins (1987), Cranston and Humphries (1988), and Stevens (1991). Because initial character selection was "anonymous," some continuous characters were not detected at this stage (e.g. culms greater than 3 m tall or not); many of these were later excluded on the basis of high homoplasy in initial cladograms.

Most of the binary characters were included in the initial data analyses; character #160 of Fig. 2, for example, was included. Some were excluded, however, on the grounds that an overwhelming majority of the genera (well over 75%) had both states. This is obviously a conservative criterion; even 25% overlap might be justification for excluding the character. However, it is not uncommon for a character to be highly homoplasious in one part of a cladogram and very consistent and phylogenetically informative in another part. To check for this possibility, such characters were included at least in initial analyses.

Multistate characters (10–15% of each data set) were evaluated, and if necessary recoded according to the degree of co-occurrence of states. Thus, if all the genera with character state A also had members with state B, but there were few that had both B and C, the character was recoded as a binary character where state 1 was equivalent to A plus B, and state 2 equivalent to C. As with the binary characters, character states were combined only when a large majority of genera were polymorphic, which led to the inclusion of some characters in which the states are probably not useful for distinguishing groups. However, because of the possibility that characters are invariant in one region of the cladogram even if variable in another, they were left in the initial round of analyses. This method of recoding multistates led to some difference in coding between clades. For example, the nature of the ligule is coded as a binary character in the bambusoid and pooid clades (A–ligule an unfringed membrane vs. B–ligule a fringed membrane or a fringe of hairs), but has three states in the andropogonoid group (A–an unfringed membrane; B–a fringed membrane; C–a fringe of hairs).

All multistate characters were treated as unordered. Although Mickevich (1982) felt that unordered characters were "the equivalent of indifference" i.e. uninformative about relationship, Hauser and Presch (1991) have shown that unordered characters have no consistent effect on tree length or resolution. Depending on the constellation of other binary characters in the data set, trees might be longer, shorter or the same length, and have more, less or the same amount of homoplasy, if multistates were ordered vs. unordered. Because we have no a priori information on transformation series in the multistate characters, we chose to leave them unordered for these analyses.

Characters to be included, with any necessary changes in character states, were converted from DELTA format to PAUP format by the TOPAU facility of the DELTA system (Dallwitz, 1992). This eliminated errors in transcription of data. Polymorphisms were consistently treated as missing data. All data matrices as employed for cladistic analyses are available from the first author on request, either as hard copy or on a Macintosh diskette. The current version of the original generic descriptions is obtainable from the second author, as a package comprising an interactive data set using the program INTKEY, plus screen readable versions of the full printed descriptions. The original data in DELTA format are also available.

After initial heuristic searches, some characters were eliminated from the analyses because of high levels of homoplasy. These are noted in discussion of the results for each clade. This approach is an easily interpreted form of a posteriori character weighting (Neff, 1986), and is based on the assumption that "unreliable characters will each vary from the phylogeny in [their] own random way, and chances are very slight that a series of random variables will by accident form a pseudo-hierarchic pattern of variation" (Carpenter, 1988; Farris, 1969; Ladiges et al., 1989). This is the rationale behind successive approximations character weighting (Farris, 1969), in which characters are successively down-weighted as a function of their unit consistencies on initial trees. In this study, character weighting was only done once, rather than successively. This does make a strong assumption (as pointed out by Farris, 1969 and 1983) that all similarities in that character are irrelevant to estimating phylogeny. However, given the size of these analyses, it seemed reasonable to overlook some possible local resolutions in favor of a clearer overall hierarchy. Exclusion of an entire set of homoplasious characters resulted in a redistribution of the homoplasy on the tree such that some of them actually changed fewer times on the tree in subsequent analyses (their consistency index improved). These few characters were then reintroduced into the analyses.

C. PROGRAMS

This study has used maximum parsimony algorithms exclusively. Maximum likelihood methods are computationally unable to handle data sets of this size (J. Felsenstein, pers. comm.). These data have already been extensively explored using phenetic algorithms (e.g., Watson et al., 1985, and unpubl.); reference to phenetic results is made below as counterpoint to the cladistic results. More recently developed phenetic methods such as neighbor-joining (Saitou & Nei, 1987) may prove useful but were not available during most of this study.

All cladograms published here were generated by PAUP 3.0s (Swofford, 1989) on an Apple Macintosh II or by a ß version for UNIX operating systems. However, other analyses described herein were conducted over a period of six years on three continents at six institutions on nine different computers, using, at various times, PHYSYS

(Mickevich & Farris, 1984), PAUP 2.2 (Swofford, 1984), PAUP 2.4 (Swofford, 1985), and Hennig86 version 1.5 (Farris, 1988). Initially, information was copied by hand from the microfiche descriptions of the grass data into a data matrix. Later, data were extracted using the program ONLINE (Pankhurst, 1986). Since 1988, all work has been done using the DELTA program INTKEY (Dallwitz, 1992; see below). As will become apparent, the project has been limited primarily by computer software; the new generation programs (PAUP 3.0 and Hennig86) permit much more detailed exploration of large data sets. The effectiveness of the search depends in part on how many trees are saved; with the newer programs analyses can now be done saving hundreds or thousands, rather than tens, of trees. In general, PAUP 3.0 has proved to be preferable to Hennig86 for exploratory analyses because of the ease of manipulating both input and output. The ability of PAUP 3.0 to interrupt a run without losing accumulated trees, has also been important for these analyses, most of which tend to be measured in days or weeks rather than hours, which limits possibilities for data exploration.

Unless otherwise noted, trees were constructed by a heuristic search, which explores many trees but does not guarantee that the trees found will in fact be the shortest for the data set. Exhaustive search methods, which do find the shortest possible trees, become computationally prohibitive if there are more than 11 taxa (Swofford, 1993); the branch-and-bound algorithm, also guaranteed to find the shortest trees, is somewhat more efficient, but is still usually good only for data sets of less ca. 30 taxa.

PAUP provides many options for exploring trees in a heuristic search. For these analyses, branch swapping was by tree bisection and recombination, and character optimization used accelerated transformations (ACCTRAN). The addition sequence was initially random, and was replicated 100–2000 times (the exact number differing with the data set; see below) with generally two trees saved at each replication. The shortest trees obtained were then used for more extensive exploration with the maximum number of trees saved (MAXTREES) set at a level that would allow the analysis to be completed in two to seven days (48–168 hours).

The program also generates many statistics for evaluating and comparing trees. The two reported here are the consistency index (CI), which divides the minimum number of changes of characters on a tree by the actual number of changes (Farris, 1989; Kluge & Farris, 1969), and the retention index (RI), which corrects for the actual distribution of characters states in the data matrix by subtracting both the minimum number of changes and the actual number of changes from the maximum number possible (Farris, 1989; see also Swofford, 1993). Both indices are measures of homoplasy, and can be applied to entire trees or to individual characters. When the statistics are used to describe trees, we include only phylogenetically informative characters.

III. The Analyses

A. THE BAMBUSOID CLADE

1. Introduction

The bambusoid clade, as defined by Kellogg and Campbell (1987), is largely equivalent to the Bambusoideae sensu lato as defined by Soderstrom and Ellis (1987). It includes the "bambusoid core" of the Bambusodeae and Olyrodeae, but also the Puelieae, Oryzeae, Streptogyneae, Phareae, Guaduelleae, and Zizanieae. The Ehrharteae were found to be basal by Kellogg and Campbell (1987), and the outgroups

for the entire clade represented by the arundinoid genera *Eriachne, Dichaetaria, Pheidochloa,* and *Danthonidium.* Taxa included are listed in Appendix A1.

Brachyelytrum, Phaenosperma, and *Diarrhena* were included in early analyses of the clade as were the members of the Centotheceae, and all members of the Bambuseae (woody bamboos) listed by Watson and Dallwitz (1988). In these early analyses, Brachyelytrum was found to have a branch length of 13 steps, and it generally appeared basal to the Centotheceae. It had a much shorter branch length when placed in the pooid clade and so was eliminated from further consideration in the bamboos. The Centotheceae always formed a monophyletic group supported by five to seven synapomorphies; the clade was generally basal to the remainder of the bambusoid clade, in line with the contention of Kellogg and Campbell (1987) that there was little cladistic support for its inclusion in the Bambusoideae. Molecular data (Cummings et al., 1993; Davis & Soreng, 1993; see below) also indicate that the Centotheceae are unrelated to the bamboos. The Centotheceae were therefore removed from subsequent analyses of the bambusoid clade.

In initial analyses all genera of the woody bamboos (Bambuseae) were included and invariably formed a monophyletic group. There is little agreement, however, on generic limits in the tribe, with treatments by Clayton and Renvoize (1986), Soderstrom and Ellis (1987), and Watson and Dallwitz (1988) differing considerably among each other (see Fig. 5). This is compounded by the very fragmentary information on many genera of woody bamboos. Fortunately, a reassessment of the generic limits in the group is currently underway (S. Dransfield, pers. comm.), but for the moment generic monophyly cannot be assumed in the woody bamboos, even as a first approximation, so any apparent relationships within the tribe are probabaly meaningless. We have used only seven genera of Bambuseae as place holders for the rest of the tribe; these were chosen for completeness of data and generally low levels of variability rather than presumed phylogenetic significance. Early analyses did use a summary taxon "bambuseae," in which all characters that varied were coded as missing. This resulted in unstable placement of the taxon.

Excluding the Centotheceae and most of the woody bamboos reduced the data matrix from its original 166 terminal taxa to a much more manageable 42. The characters used are listed in Appendix B1. Phylogenetic analysis was done using the heuristic search algorithm of PAUP 3.0 with 100 random addition sequences. In addition to finding consensus trees, a bootstrap analysis was also done with 100 replicates. *Phaenosperma* and *Diarrhena* were variously included and excluded as discussed below. Trees were rooted at *Danthonidium* (Arundinoideae), following Kellogg and Campbell (1987).

2. Results and Discussion

One hundred heuristic searches with random addition sequences found 10 equal length trees; branch-swapping on these found 32 trees, each 371 steps long (excluding *Diarrhena* and *Phaenosperma*) with a CI of 0.358 and RI of 0.592. The major clades are similar in all trees, but they vary in their relationship to each other; two equal length trees are shown in Figs. 3 and 4, and the strict consensus tree in Fig. 5. Tribal designations of Clayton and Renvoize (1986), Soderstrom and Ellis (1987), and Watson and Dallwitz (1988), which are all quite similar, are compared to the consensus tree in Fig. 5. In all trees, the Olyreae, Phareae, Buergersiochloeae, and Phyllorhachidae form a clade, as do the Oryzeae, and the two clades are sister taxa. The

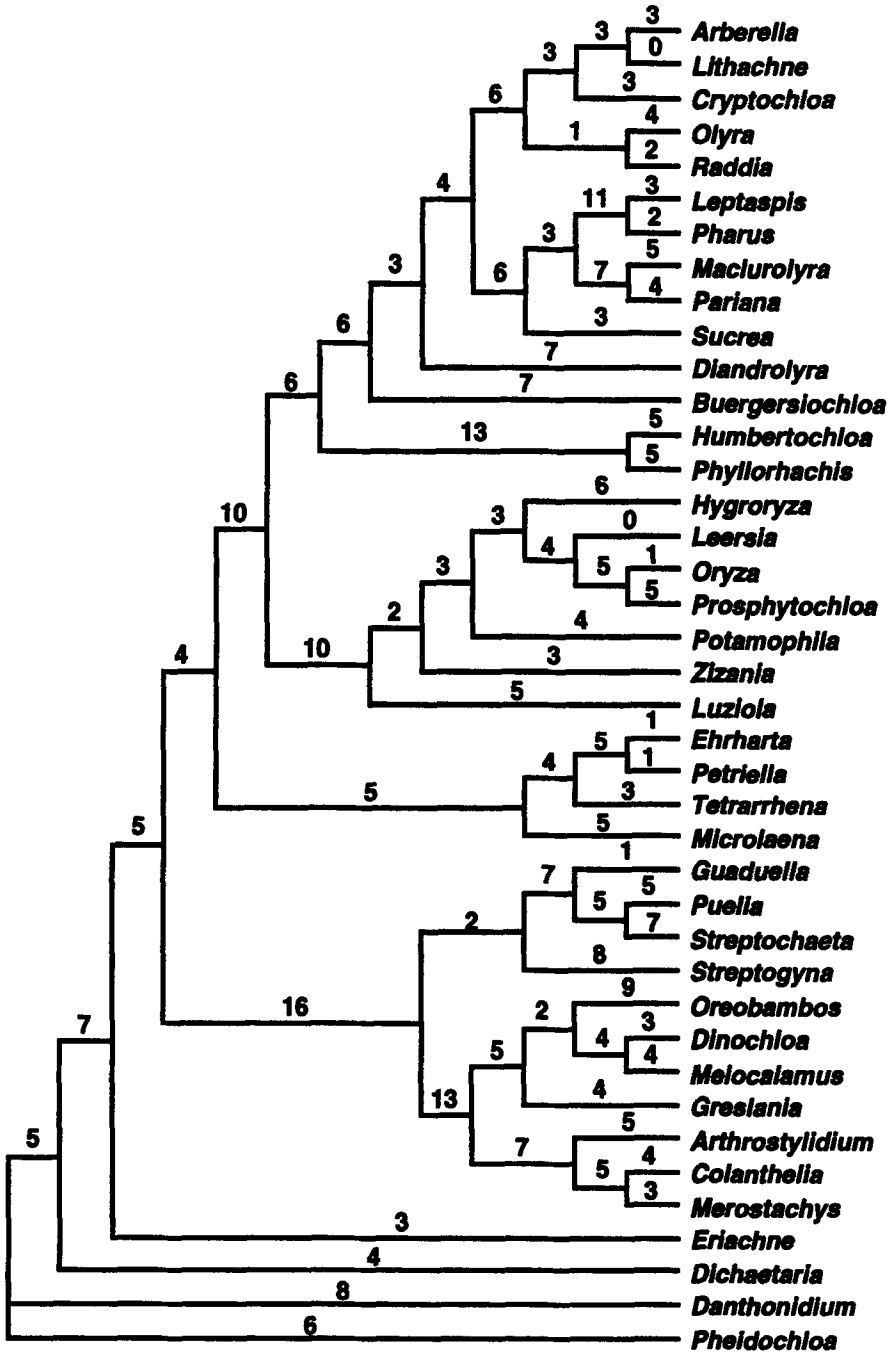

Fig. 3. One phylogenetic hypothesis for the bambusoid clade. Numbers above branches indicate branch lengths with ACCTRAN character optimization. Tree length = 371; CI = 0.358; RI = 0.592.

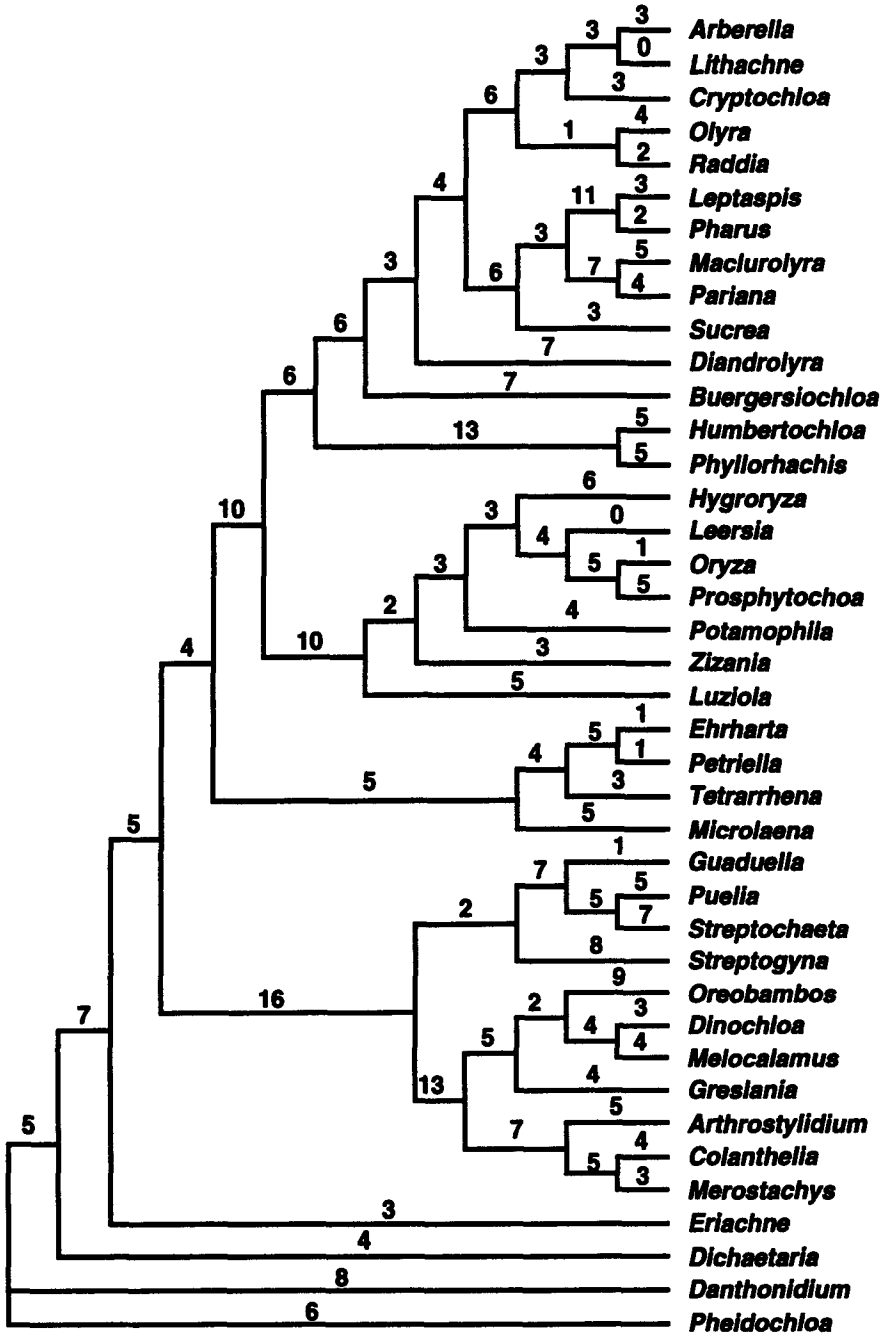

Fig. 4. A second phylogenetic hypothesis for the bambusoid clade, differing from that in Fig. 3 in the monophyly of the Ehrharteae. Numbers above branches indicate branch lengths with ACCTRAN character optimization. Tree length = 371; CI = 0.358; RI = 0.592.

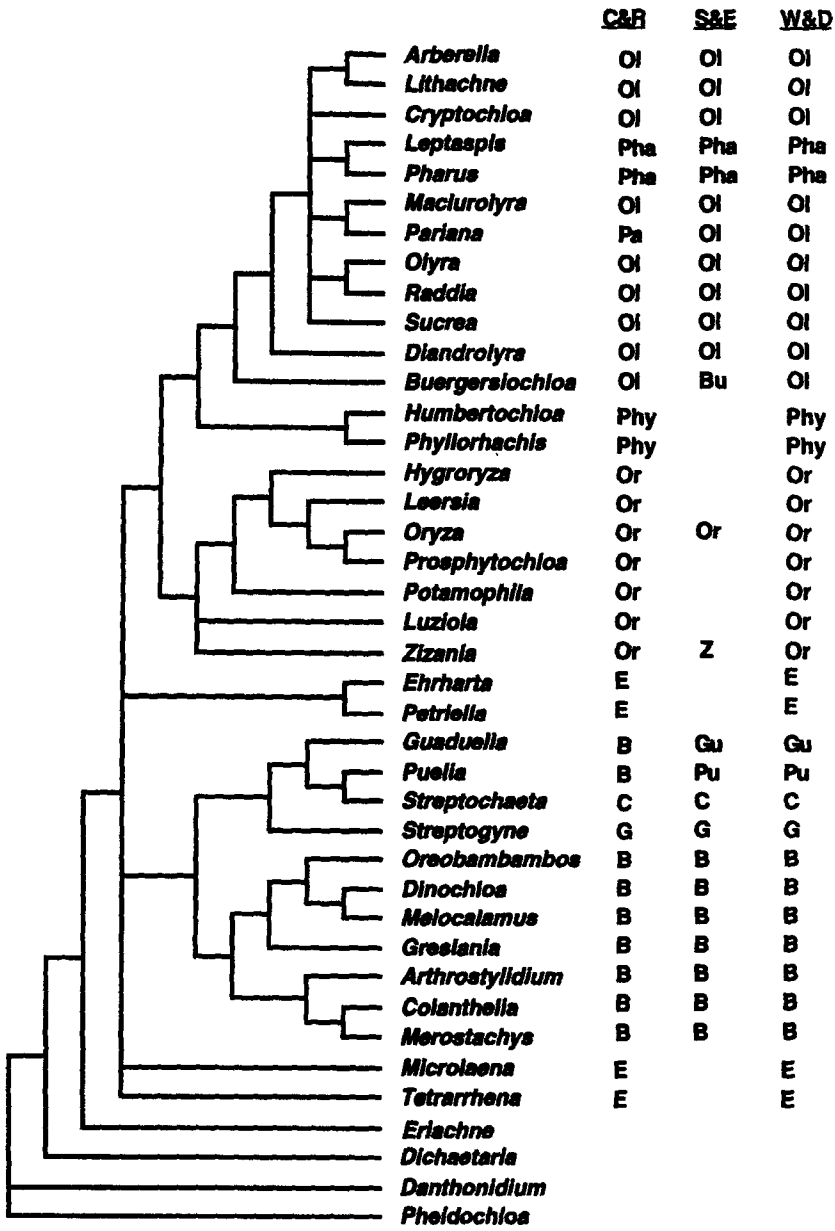

Fig. 5. Strict consensus of 32 equally parsimonious trees for the bambusoid clade. Statistics as in Figs. 3 and 4. Recent classifications of the group are compared in the columns at the right. C&R = Clayton and Renvoize (1986); S&E = Soderstrom and Ellis (1987); W&D = Watson and Dallwitz (1988). Tribal abbreviations: Ol = Olyreae; Pha = Phareae; Pa = Parianeae; Bu = Buergersiochloeae; Phy = Phyllorhachidae; Or = Oryzeae; Z = Zizanieae; E = Ehrharteae; B = Bambuseae; Gu = Guaduelleae; Pu = Puelieae; C = Streptochaeteae; G = Streptogyneae. Note that Clayton and Renvoize (1986) include all the Ehrharteae in the genus *Ehrharta*.

herbaceous genera *Streptochaeta, Streptogyna, Puelia,* and *Guaduella* form a clade (the "Guaduella group") that is the sister taxon to the woody bamboos. The Ehrharteae may be paraphyletic and basal to the clade including the Guadella group and the woody bamboos (Fig. 3) or monophyletic and basal to the Olyreae plus Oryzeae (Fig. 4). The trees vary in whether *Zizania* or *Luziola* is basal in the Oryzeae, and in the position of *Cryptochloa*—whether sister taxon to *Arberella/Lithachne,* as shown here, sister taxon to the Phareae/Parianeae group or sister taxon to just the Phareae.

The higher-level classifications of Clayton and Renvoize (1986) and Soderstrom and Ellis (1987) are not consistent with the phylogenetic relationship of the tribes. Soderstrom and Ellis (1987) define a "bambusoid core" including tribes Bambuseae, Anomochloeae (excluded here because of missing data), Buergersiochloeae, Olyreae and Streptochaeteae. This group is clearly polyphyletic. Their supertribes, Olyrodae and Bambusodae, are therefore also polyphyletic. The Bambuseae of Clayton and Renvoize (1986) includes not only the woody bamboos, but also two herbaceous genera, *Guaduella* and *Puelia.* This group is also polyphyletic, in that it excludes *Streptochaeta* and *Streptogyna* which they place in tribes of their own.

The Oryzeae are consistently monophyletic only if *Zizania* is included. Soderstrom and Ellis (1987) placed *Zizania* in its own tribe, a classification consistent with some but not all of the cladograms. Clayton and Renvoize (1986) and Watson and Dallwitz (1988) place *Zizania* in the Oryzeae, a classification that is cladistically defensible. The Ehrharteae may or may not be monophyletic depending on the position of *Tetrarrhena* and *Microlaena.* The questionable monophyly of the Ehrharteae is responsible for the lack of resolution at the base of the consensus tree (Fig. 5). The genus *Ehrharta* itself has been carefully examined by Ellis (1987a, b), Gibbs-Russell (1987), and Gibbs-Russell and Ellis (1987, 1988), but *Tetrarrhena* and *Microlaena* have received much less attention. It will be critical for future analyses to evaluate the nature of the many differences in morphological characters among the three genera.

Diarrhena and *Phaenosperma,* if included, appear basal to the "*Guaduella* group" plus woody bamboos with branch lengths of 7 and 10 respectively. The two genera thus share no more clear apomorphies with the bambusoids than they do with the pooid clade. There is no suggestion of any linkage to the Phareae as suggested by Clayton and Renvoize (1986).

The 14 characters used in Soderstrom and Ellis (1987) are the ones most heavily relied upon by other authors; character distributions are based on optimizations on the tree in Fig. 3. Hilum form (#123) is linear or variably short and long-linear in all included taxa, thus not providing any phylogenetic information at this level. A short punctiform hilum is found in the Centotheceae and is one of the characters that excludes them from the bambusoid clade. One embryo character, presence of an epiblast, is invariant for all taxa in this cladogram. Two other embryo characters (#130 and #132, embryo with a scutellar tail and with overlapping leaf margins) each change only twice on the tree. The loss of the scutellar tail is a synapomorphy for the Oryzeae excluding *Zizania,* but it is regained in *Oryza.* The short mesocotyl internode (#131) is a synapomorphy for the entire ingroup (i.e., it changes on the internode above *Eriachne*), although it reverses in *Microlaena.* Embryonic leaf margins are overlapping in most taxa on the cladogram, but simply meet in *Pheidochloa* (one of the basal outgroups), and in *Potamophila*; the character is thus phylogenetically uninformative and has a retention index of 0.

The embryo is consistently small in all taxa in this analysis. The number of lodicules

(#104) changes from two to three twice, once at the base of the Olyreae s.l. and once at the base of the Bambuseae s.l. As noted by Kellogg and Campbell (1987), this character appears to be a reversal rather than the primitive condition, consistent with the hypothesis that the third lodicule is in fact developmentally, structurally, and evolutionarily different from the other two (Clifford, 1987, and references therein). The first seedling leaf (#134) lacks a well-developed lamina in the Olyreae plus Oryzeae and also in *Streptochaeta,* thus giving two changes on the tree. Microhairs on the abaxial epidermis (#135) are lost three times independently, in the Phareae (which invert their leaf blades during development), in *Puelia,* and in *Streptogyna.* Photosynthetic pathway is not included in this data set, although *Eriachne* and *Pheidochloa* are known to be C_4 and all other taxa C_3 (Hattersley, 1987).

Arm and fusoid cells (#161 and #162) are characters on which Kellogg and Campbell (1987) relied extensively in their argumentation for the structure of the Bambusoideae, but they do not appear to be particularly strong characters in this analysis. Arm cells are synapomorphic for the Oryzeae/Olyreae clade, but are then lost in *Leptaspis*; they are also synapomorphic for the woody bamboos and are subsequently lost in *Colanthelia.* Fusoid cells are gained in the Olyreae excluding *Buergersiochloa*; they are gained independently in the Oryzeae and then lost in *Luziola* and in the *Leersia/Oryza/Prosphytochloa* clade. They are also gained in the Bambuseae s.l. Midrib structure (#166) was initially eliminated as being too variable, but with later alterations in the data set the character was reintroduced. Midribs with supernumerary bundles arise in the Bambuseae s.l., in the Oryzeae and in the Phareae, but are subsequently lost in *Hygroryza* and *Colanthelia* plus *Merostachys. Luziola* and the genera of the Olyreae have an arc of bundles in the midrib. There may be some error associated with scoring this character; the extra bundles in some taxa appear only in the proximal part of the blade, which is not customarily examined. Oryzoid (vertically elongated) silica bodies (#148) are more consistent with other characters than are the other types of silica bodies. Even so, the character changes three times on this cladogram, being lost in the *Leptaspis/Pharus* clade and on the internode above *Eriachne* and then regained at the base of the Oryzeae/Olyreae.

Base chromosome number and level of ploidy were not included as characters in this analysis, but the distribution of polyploid taxa is striking. Diploid chromosome numbers are common throughout the bambusoid clade except among the woody bamboos where tetraploids and hexaploids predominate (but see Pohl & Clark, 1992). It is possible that all woody bamboos are descended from a polyploid ancestor and that the infrequent reports of diploid numbers reflect secondary reductions in chromosome number.

To correct for the possibility that genera might be polyphyletic, one analysis was done on only those genera with one or two species. As with the full analysis, the number of woody bamboo genera was reduced to a small number of "place-holders." Six trees were found, with length 188, CI= 0.532. The strict consensus and one of the six trees are shown in Fig. 6. The overall topology is similar to that of the larger analysis, although the Olyreae are monophyletic only if *Buergersiochloa* is included. The Phyllorachidae are monophyletic, as are the woody bamboos. *Streptogyna* and *Streptochaeta* are sister taxa to the woody bamboos, and the Oryzeae are basal and paraphyletic.

The cladistic structure favored by Kellogg and Campbell (1987)—woody bamboos nested within the bambusoid core, with the Oryzeae basal—is more apparent in the analysis of monotypic taxa than it is in the larger analysis. Although the Bambuseae

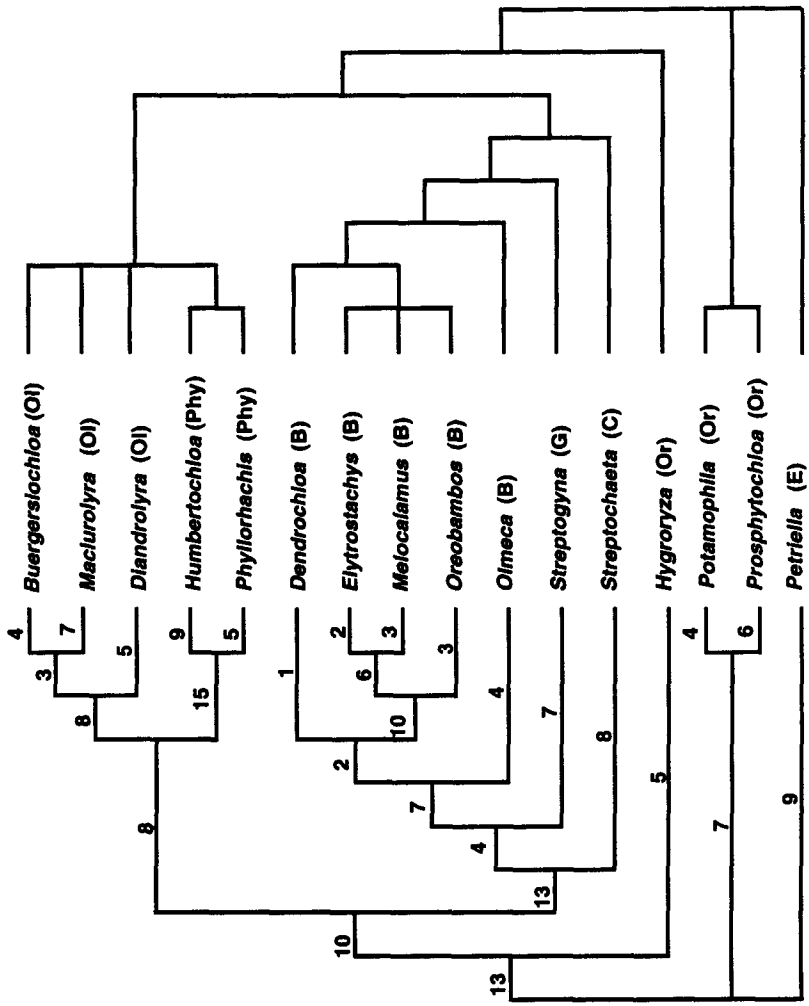

Fig. 6. Bambusoid clade, including only genera with one or two species. Left hand cladogram: one phylogenetic hypothesis for the bambusoid clade, using only genera with one or two species. Numbers above branches indicate branch lengths with ACCTRAN character optimization. Right hand cladogram: strict consensus of six trees, all with length = 188, CI = 0.532; RI = 0.579.

s.l., the Oryzeae, the Olyreae/Phyllorachidae, and the Ehrharteae all are monophyletic groups, the relationship among them is not stable. In any future family-level analyses, several representatives of each clade will need to be included. Finally, as noted above, the generic limits in the woody bamboos are not well-enough defined for even a preliminary hypothesis of relationship. Because of the economic importance of this group, this should be a high priority for future research.

The bamboos are often thought to be among the most primitive of the grasses, and indeed Soderstrom (1981) suggested that *Streptochaeta, Streptogyna* and *Pharus* may be most similar to the ancestral graminoid stock. Kellogg and Campbell (1987) found little cladistic support for this view, however, and hence the trees presented here are rooted with reference to several arundinoid genera. If we assume that Kellogg and Campbell (1987) were wrong, and the tree in Fig. 3 is rooted at *Pharus,* then the *Guaduella/Puelia/ Streptochaeta/ Streptogyna* group is among the basal grasses, and Soderstrom (1981) was correct; in this case, the other herbaceous bamboos are then not basal. Conversely, the tree may be rooted near the *Guaduella/Puelia/ Streptochaeta/Streptogyna* group, in which case the woody bamboos are also near the base. In any case, the trees presented here reinforce the conclusion of Kellogg and Campbell (1987) that the Bambusoideae are unlikely to be both basal and monophyletic.

B. ANDROPOGONODAE

1. Introduction

The panicoid clade, as defined by Kellogg and Campbell (1987), includes all the Panicoideae, plus *Alloechaete, Micraira, Isachne, Heteranthoecia, Coelachne* and the Centotheceae. This comprises 256 genera, and so still needs to be broken down into smaller subgroups for analysis. The supertribe Andropogonodae is almost certainly monophyletic and can legitimately be analyzed separately from the rest of the panicoids. The monophyly of the tribe is based on several characters: Nearly all members have spikelets in pairs, one sessile and one pedicellate. The tribe consistently exhibits the NADP-ME subtype of the C_4 photosynthetic pathway, with leaf cross sectional anatomy characterized by a single, rather than double, sheath of cells around the vascular tissue (absence of a mestome sheath). All members examined have agranal chloroplasts in the bundle sheath, inferred to be associated with a reduction in the proteins associated with photosystem II. Because of their anatomical similarities, all members of the Andropogonodae are expected to have a photosynthetic pathway similar to maize, an assertion borne out by several studies (Hattersley, 1987, and references therein).

Previous work on the supertribe Andropogonodae suggested that it might be easily divided into smaller groups. Clayton and Renvoize (1986) recognize a single tribe with 11 subtribes, and Watson and Dallwitz (1988) divide it (as the supertribe Andropogonanae) into two tribes, the largest of these (the Andropogoneae) with two subtribes, comprising the awned and the awnless species (see Fig. 8). The analyses presented here, therefore, explored whether these subdivisions were strictly monophyletic and determined a small set of taxa that could justifiably be used in future analyses of the entire subfamily Panicoideae. We followed the suggestion of Clayton and Renvoize (1986) that the Andropogoneae were likely to be closely related to the Arundinelleae and that two tribes might have been a basal offshoot of the remainder of the Panicoideae, having diverged approximately at the same time as the Neurachneae.

We therefore analyzed the Andropogoneae using Arundinelleae and Neurachneae as outgroups (Appendix A2) and rooted the trees at *Thyridolepis.* This rooting will be tested in future analyses.

Characters are listed in Appendix B2. Data were analyzed in PAUP 3.0 ß for Unix systems using 2000 random addition sequences to search for multiple islands of equally parsimonious trees. Only a single island was found, with trees of length 662. Over 5000 trees were found at this length.

2. Results and Discussion

One of the 5000 equally parsimonious trees is shown in Fig. 7 (CI = 0.249; RI = 0.570), and the strict consensus in Fig. 8. The tribal and subtribal designations of Clayton and Renvoize (1986) and Watson and Dallwitz (1988) are compared to the consensus tree. Even with the large number of equal-length trees, the consensus tree is well-resolved. The largest of the three main groups corresponds (with a couple of exceptions noted below) to the Andropogonineae of Watson and Dallwitz (1988). The second group is primarily made up of the Rottboelliinae plus the Maydeae [sensu Watson & Dallwitz (1988)], and the third group, the Arundinelleae. Several taxa, however, appear out of place. *Garnotia,* a member of the Arundinelleae, in this tree appears in the midst of the Andropogoneae, as sister to *Leptosaccharum* and *Oxyrhachis.* This small clade is part of an equally heterogeneous larger clade including *Elionurus, Eriochrysis, Imperata,* and *Vetiveria.* Clayton and Renvoize (1986) unite *Leptosaccharum* and *Eriochrysis* on the basis of their rufous panicles. *Elionurus* is usually placed among the Rottboelliinae because of its lack of awns, as is *Oxyrhachis.* In the tree shown in Fig. 7, this clade is supported by only three characters, leaves mostly basal (#9), palea apically notched (#127), and costal short cells predominantly paired (#187). These three characters change 7, 5 and 11 times on the tree, respectively. This group is therefore somewhat suspect. Clayton and Renvoize (1986) point out, however, that *Oxyrhachis* is a genus of uncertain affinity and this may reflect its distance from the remainder of the Rottboelliinae. Its inflorescence is of the "rat-tail" type that has appeared in a number of grass genera that are otherwise unrelated; studies of inflorescence development in *Oxyrhachis* and other genera of similar aspect (e.g. *Pholiurus* and *Henrardia,* both Pooideae) might illuminate the frequent convergent evolution of this inflorescence type. The anomalous placement of all the members of this clade would indicate the need for reexamination of all relevant characters. *Oxyrhachis* is linked to *Leptosaccharum* and *Garnotia* by having spikelets mainly solitary (#54), costal regions with crescentic silica bodies (#175), and the leaf blade with prominent adaxial ribs (#206). The character consistency indices are 0.29, 1.0 and 0.2, respectively. Note that this is one of the few cases where silica bodies show a consistent pattern of distribution and are therefore phylogenetically informative.

Included within the Rottboelliinae are three members of the *Ischaeminae, Digastrium, Ischaemum* and *Thelepogon. Digastrium* and *Ischaemum* always appear as sister taxa, supporting Clayton and Renvoize's (1986) combination of the two. In other analyses of this data set, the three genera form part of a larger clade that includes *Phacelurus, Vossia,* and the *Loxodera/Lasiurus/Urelytrum* group, and the larger clade is then the sister taxon to the Arundinelleae. The placement of these eight genera at the base of the Rottboelliinae is thus not highly robust.

The Arundinelleae are monophyletic and nearly basal to the rest of the Andropogoneae. In the cladogram shown here their immediate sister taxon is *Polytrias,* a

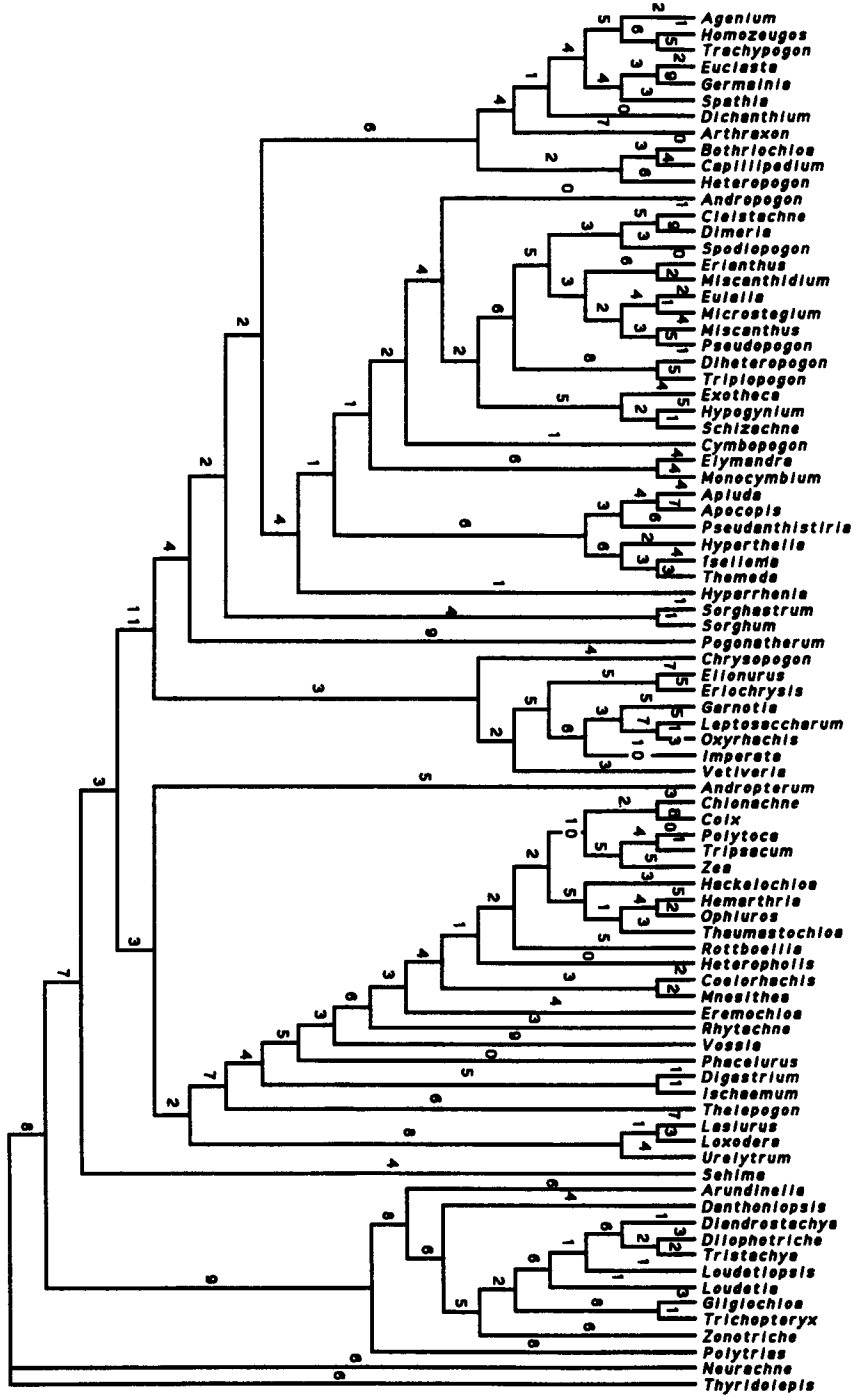

Fig. 7. One phylogenetic hypothesis for the genera of the Andropogonodae plus Arundinelleae and Neurachneae. Numbers on left of branches indicate branch lengths with ACCTRAN character optimization. Tree length = 662; CI = 0.249; RI = 0.570.

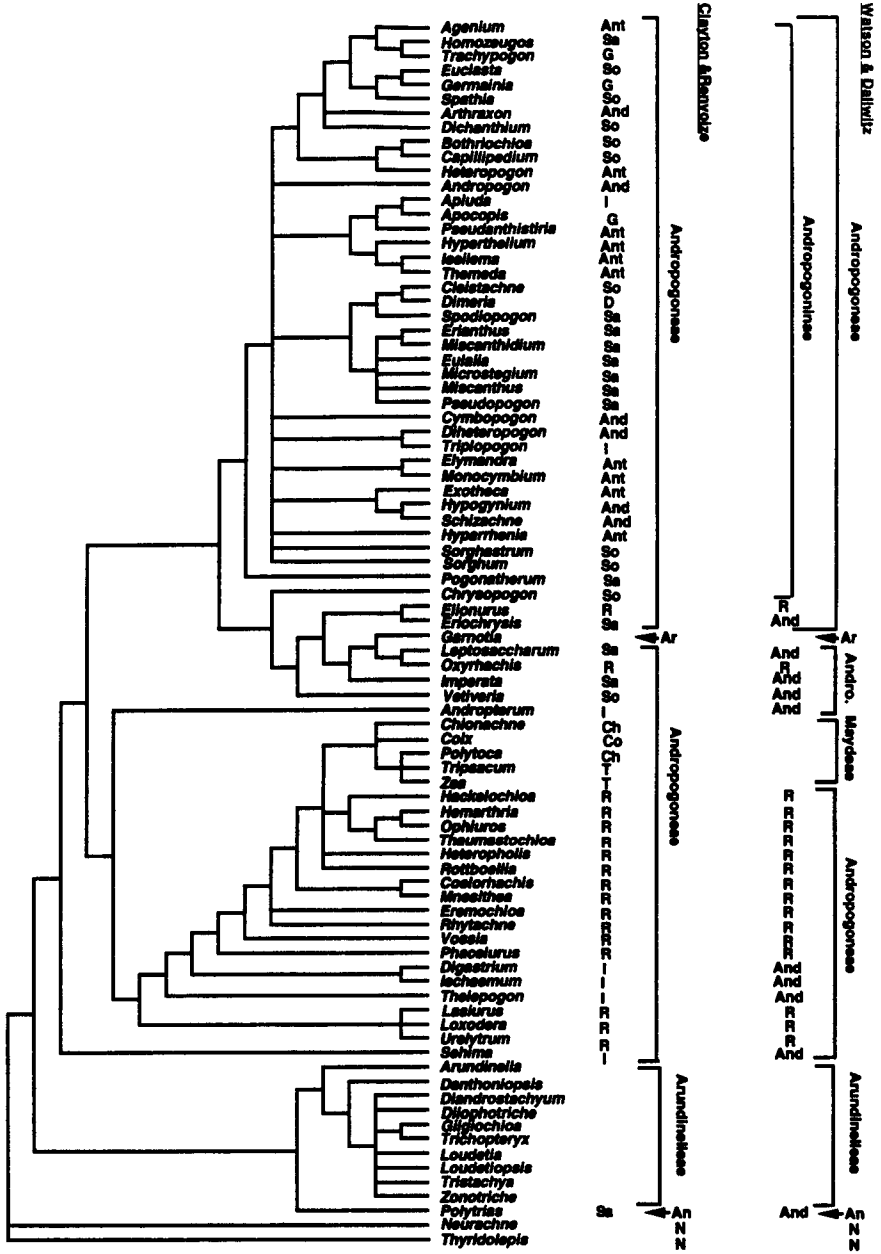

Fig. 8. Andropogonodae, Arundinelleae, and Neurachneae. Strict consensus of 5000 trees; statistics as in Fig. 7. Tribal and subtribal classifications of Clayton and Renvoize (1986) and Watson and Dallwitz (1988) are compared above. Tribal and subtribal abbreviations are as follows: Ar = Arundinelleae; An = Andropogoneae; N = Neurachneae; Ant = Anthistiriinae; Sa = Saccharinae; G = Germainiinae; So = Sorghinae; And = Andropogoninae; I = Ischaeminae; D = Dimerinae; R = Rottboelliinae; Ch = Chionachninae; Co = Coicinae; T = Tripsacinae.

member of the Saccharinae, but in other analyses *Polytrias* is placed more conventionally with *Pseudopogonatherum.*

In no analysis was *Dichanthium* linked with *Bothriochloa* and *Capillipedium,* despite the fact that they are all interfertile at the tetraploid level (DeWet & Harlan, 1970; see Section IVC below, for discussion of hybrids and cladograms). *Dichanthium* remains in a separate clade even in the consensus tree, although it is basal in that clade, and in the tree shown in Fig. 7 it has a branch length of zero. The clade that includes *Dichanthium* is supported by having culm nodes hairy (#5) and an inflorescence of spikelike main branches (#30), the ultimate branches clustered (#41), and the long cells markedly different in shape costally and intercostally (#163), all characters for which *Dichanthium* is variable. Of the characters supporting its placement as shown in Fig. 7, the only constant one is a digitate or subdigitate inflorescence (#34); the inflorescence in *Capillipedium* is not digitate, and that of *Bothriochloa* is variable.

Phenograms show that the subtribes of Clayton and Renvoize (1986), other than their Germainiinae and Andropogoninae, are phenetically cohesive at low levels, though with numerous exceptions, e.g. *Agenium, Eremopogon, Sehima, Pogonatherum,* and *Cleistachne* are "misplaced" (data not shown). However, in the cladistic analyses presented here the subtribes are not strictly monophyletic entities. The members of the Saccharinae are cladistically distant from each other, although the core "*Saccharum* group" is monophyletic, including *Erianthus, Miscanthidium, Eulalia, Microstegium, Miscanthus,* and *Pseudopogonatherum,* the latter put in *Eulalia* by Clayton and Renvoize (1986). *Erianthus* is sometimes included in *Saccharum,* but is kept separate here as representative of the diploid members of the genus. Members of the Ischaeminae are widely scattered over the phylogeny. In particular the position of *Andropterum* deserves further analysis. It is depicted here as basal to the Rottboelliinae sensu latissimo, but in the cladogram of mono- and dispecific taxa it is basal to the entire supertribe.

The genus *Dimeria,* put in its own subtribe because of having two fertile florets, rather than one fertile and the other reduced, consistently appears as the sister genus of *Cleistachne,* itself interpreted by Clayton and Renvoize (1986) as being part of the Sorghinae. The genus *Germainia* is placed with *Trachypogon* in the Germainiinae by Clayton and Renvoize (1986) on the basis of a shared reduction in the sessile spikelet. In both cladistic and phenetic analyses, however, *Germainia* is consistently the sister genus to *Euclasta* (Sorghinae) and *Trachypogon* the sister taxon to *Homozeugos* (Saccharinae). There is no evidence of a relationship with the central "*Saccharum* group," as suggested by Clayton and Renvoize (1986).

The Maydeae, in the sense of Watson and Dallwitz (1988) and this cladogram, is supported by 10 characters: leaf blades rolled in bud (#18), plants monoecious (#23), plants without hermaphrodite florets (#24), male and female spikelets in different inflorescences (#26), spikelets not all embedded in the rachis (#37), rachis "articles" without a basal callus-knob (#47), sessile spikelets pistillate only (#62), lodicules absent (#134), styles fused (#142), and vascular bundles in the culms in three or more rings (#219). The three subtribes recognized by Clayton and Renvoize (1986), Tripsacinae, Coicinae, and Chionachninae, (1986) are seemingly not monophyletic. *Zea* plus *Tripsacum* are included in the Tripsacinae, which by itself is either monophyletic or paraphyletic (see also Kellogg & Birchler, 1993). If it is recognized, however, the Chionachninae, including *Polytoca* and *Chionachne,* becomes para- or polyphyletic; in these two genera the pedicel is interpreted as being completely fused to the internode

it subtends (Clayton, 1981). *Coix,* the sole genus in the Coicinae is also part of this broadly circumscribed Maydeae. "Pedicel fused to internode" (#61), changes only once on the tree on the internode below *Heteropholis* (subtribe Rottboelliinae).

The close relationship of *Zea* and *Tripsacum* is well known, and there is a wealth of genetic, cytogenetic, and molecular data on the two genera (review in Kellogg & Birchler, 1993). The Maydeae are clearly derived from within a larger monophyletic group, corresponding to the awnless Andropogoneae; many of the members of this group are placed by Clayton and Renvoize (1986) in the Rottboelliinae. If the Maydeae are formally recognized, however, the Rottboelliinae becomes paraphyletic. The "core Rottboelliinae," (*Rhytachne* on up) is supported by six characters in the tree shown here, some of them partially redundant: inflorescence a single raceme (#30) and not digitate (#34), spikelet-bearing axes spike-like (#40) and solitary (#41), lower glume convex on the back (#93), and the palea of the incomplete florets reduced or vestigial (#99).

In many taxa of the subtribe, the pedicel of the pedicellate spikelet is interpreted as being more or less fused to the rachis. There have been no developmental studies on this character, however. As scored here, *Coelorachis* and *Mnesithea* are interpreted as non-fused, whereas some taxa (e.g., *Polytoca*) are thought to exhibit complete fusion of the pedicel and rachis; this character needs to be examined in more detail.

As noted by Jain (1970), the taxonomic history of the group is confused, with species frequently being transferred between genera. A case in point is the genus *Manisuris,* which features in many of the discussions of the origin of maize and *Tripsacum.* A. S. Hitchcock (1935, 1950) recognized five species in the genus, four American and one Asian; because his publication was and is the only comprehensive treatment of North American grasses it has been widely followed. However, Jain (1970) confined *Manisuris* to a set of nine Asian species, so the American species are now properly placed in the genus *Coelorachis (C. cylindrica* (Michx.) Nash, *C. rugosa* (Nutt.) Nash, *C. tesselata* (Steud.) Nash, and *C. tuberculosa* Nash), along with a number of Old World species. Then in 1981, Clayton restricted *Manisuris* to its type species (native to India) and placed Jain's other species in the genus *Glyphochloa.* Neither *Manisuris* s.s. nor *Glyphochloa* is included in the cladograms shown here because both had extensive missing data in one or more character systems; the species commonly called *Manisuris cylindrica* in the American literature is included in *Coelorachis.*

Both Clayton (1966, 1970, and 1973) and Veldkamp et al. (1986) have noted that many of the generic limits are poorly defined, and may represent arbitrary divisions of a continuum (although the potential continuity was not evaluated in a phylogenetic context). In a phenetic study of *Coelorachis, Rhytachne, Robynsiochloa,* and *Chasmopodium* (which are not sister taxa in the cladograms presented here), Clayton (1970) found two clear groups not corresponding to previously delimited genera. deKoning et al. (1983) presented two alternative cladograms depicting relationships of *Heteropholis* and *Thaumastochloa*; both show *Thaumastochloa* as monophyletic, but *Heteropholis* may be either para- or polyphyletic. In the more comprehensive cladogram in Fig. 8, the two genera are not sister taxa. Veldkamp et al. (1986) claim that most of the genera in the Rottboelliinae could justifiably be included in *Mnesithea,* and they do in fact sink *Coelorachis* accordingly. The foregoing suggests that some of the genera in our analyses may not be strictly monophyletic; however, the tree structure is clearly hierarchical, indicating that generic-level polyphyly is not common enough to obscure a basically divergent evolutionary pattern.

The general topology of the large cladogram is reflected in the cladograms of the mono- and ditypic genera. The strict consensus tree and one of the seven equally parsimonious trees are shown in Fig. 9. The subtribes of Clayton and Renvoize (1986) remain non-monophyletic, and *Andropterum* is the sister taxon of all other taxa. *Zea* is part of a monophyletic group that includes *Polytoca,* and of a larger group that includes *Hackelochloa* of the Rottboelliinae. *Thelepogon, Vossia* and *Digastrium* form a clade distinct from all other taxa. *Oxyrhachis* is never linked to other awnless Andropogoneae, suggesting that the loss of awns occurred independently.

C. THE POOID CLADE

1. Introduction

The pooid clade includes the Pooideae sensu Watson and Dallwitz (1988), plus *Nardus, Lygeum,* the Stipeae, *Brachyelytrum, Diarrhena,* and *Phaenosperma* (full list in Appendix A3; note that the latter three were also included in analyses of the bambusoid clade because of ambiguities in their position). Kellogg and Campbell (1987) concluded that this clade is monophyletic, based on having a "pooid embryo" with an epiblast, no scutellar cleft, and a short mesocotyl internode (although some species in the Triticeae appear to have secondarily lost the epiblast). *Nardus* is basal in this group; the more restricted pooid clade (i.e., excluding *Nardus*) is united by lack of microhairs and non-vascularized lodicules. (*Nardus* lacks lodicules, and the character is thus not informative at that level.) *Euthryptochloa* was initially included because of its apparent similarity to *Phaenosperma,* but was deleted because of extensive missing data; it is somewhat similar to *Cyathopus* and may be related. The starting character list included 158 characters (see annotated character list, Appendix B3), 25 multistate and the remainder binary. Trees were rooted using *Nardus* as the outgroup, based on the conclusions of Kellogg and Campbell (1987).

In initial runs of the matrix, *Lygeum* appeared at the base with *Nardus.* It was subsequently deleted on the basis of: (1) the discovery of chloridoid microhairs on the leaf epidermis (Watson, unpublished data) and (2) a tetraploid chromosome number, indicating possible hybrid origin. Note, however, that molecular studies do not indicate hybrid origin of *Lygeum* (Cummings et al., forthcoming; see below).

Early analyses indicated few robust groups. The Triticeae were monophyletic, but not the Triticodae (Triticeae plus Bromeae plus Brachypodieae; Macfarlane & Watson, 1987, as Triticanae); within the Triticeae, *Lophopyrum* and *Thinopyrum* were consistently sister taxa, as were *Aegilops* (s.l.) and *Henrardia,* and *Hordeum* plus *Taeniatherum*). The Triticeae were thus treated as a single terminal taxon in the analyses illustrated, with character states being a summary of all character states in the group (i.e., not assuming any in-group structure). The one exception to this is the clade of *Hordeum* plus *Taeniatherum,* which never appeared to be basal; thus a character state occurring only in one of them was interpreted as derived in the tribe and hence not affecting the character state at the basal node.

After several initial analyses, *Diarrhena* and *Phaenosperma* were excluded from the clade, primarily for computational reasons. As might be expected from their unusual combination of characters they had long branch lengths (8 and 10 respectively) and their position was highly unstable. Leaving them in the analysis increased the already-lengthy computational time, since many of the equally parsimonious trees simply represented differing placements for the two genera. Their inclusion or

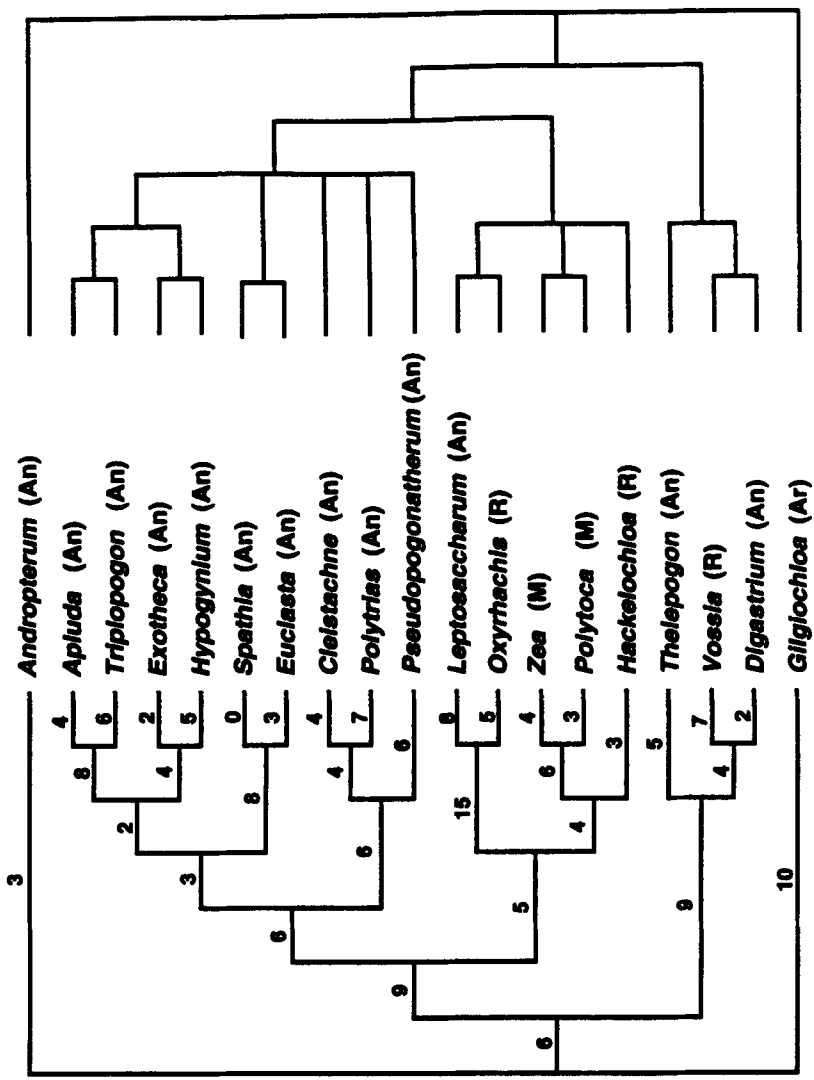

Fig. 9. Andropogonodae and Arundinelleae, including only genera with one or two species. Left hand cladogram: one phylogenetic hypothesis. Numbers above branches indicate branch lengths with ACCTRAN character optimization. Right hand cladogram: strict consensus of seven trees; all with tree length = 186; CI = 0.462; RI = 0.487.

exclusion did not affect the relationships of other taxa. (There is also the possibility that with n = 17 and 20 for *Diarrhena* and n = 12 for *Phaenosperma,* the two are allopolyploids.) Excluding the two genera from this analysis does not necessarily imply that they belong elsewhere—merely that the data are insufficient to permit their reliable placement in the pooid clade and thus their inclusion is uninformative.

Characters included were only those with CI's greater than 0.14 (see Appendix B3), i.e. those that changed fewer than 7 times on the tree. The maximum number of trees saved in this analysis was 500, but there are undoubtedly many more. This data set was extensively explored, by excluding taxa and characters, combining taxa in different ways, and recoding some characters. More than a dozen analyses, each lasting from 7 to 15 days, were undertaken on the full data set. Other subanalyses were used to examine smaller parts of the data. The tree presented here was reached by doing 100 heuristic searches with a random addition sequence, which yielded only a single island (Maddison, 1991). The shortest tree (316 steps) was then explored more fully and 500 trees were found that were two steps shorter.

2. Results and Discussion

The most striking aspect of this data set was the complete lack of higher-level structure. The tree shown in Fig. 10 is one of 500 equally parsimonious trees (CI = 0.266; RI = 0.576) of length 319. The strict consensus tree is shown in Fig. 11 and is typical of those found for all permutations of the data. The tree has no consistent hierarchical structure, although some small monophyletic groups seem to be well-supported.

The trees are, in general, dominated by inflorescence characters and thus contain some groups that are likely (on the basis of other evidence) to be artificial. The basal taxa consistently share inflorescence characteristics with *Nardus,* having narrow spicate inflorescences that are often one-sided. Thus *Narduroides,* the Triticeae, and *Lolium* are placed here, as are some members of the Hainardieae. In some other permutations of this data set, *Rhizocephalus* appears at or near the base of the cladogram, along with *Mibora,* with its one-sided racemes; these latter two genera share with *Nardus* the character state of fused styles, interpreted here as plesiomorphic. The position of these taxa at the base may well be spurious, however, reflecting superficial resemblance to *Nardus.*

The subfamilial classifications of Clayton and Renvoize (1986) and Watson and Dallwitz (1988) are compared to the consensus tree in Fig. 11. Although both classifications rcognize the tribes Aveneae, Poeae, Stipeae, and Meliceae, the genera included are somewhat different, as indicated by the arrows.

Clayton and Renvoize (1986) recognize the tribe Hainardieae, including *Narduroides, Pholiurus, Parapholis, Scribneria,* and *Hainardia,* whereas Watson and Dallwitz (1988) place these in the Poeae. This tribe is not monophyletic here, but *Parapholis* and *Hainardia* are sister taxa in all trees, based on having costal short-cells predominantly paired (#144); *Parapholis* and *Hainardia* are part of a larger group that includes *Gaudinia,* supported by having spikelet-bearing axes that disarticulate (#30). (Characters #16, leaf blades folded in the bud, and #117, having a loose coleoptile, are also interpreted as supporting this group, but are in fact only recorded for *Parapholis.*) It should be noted that, although various members of the Hainardieae plus *Gaudinia* often formed a clade in other analyses of this data set, this clade was never supported by the same characters. The synapomorphy suggested by Clayton and Renvoize (1986) for the tribe, collateral glumes, was not used in these analyses;

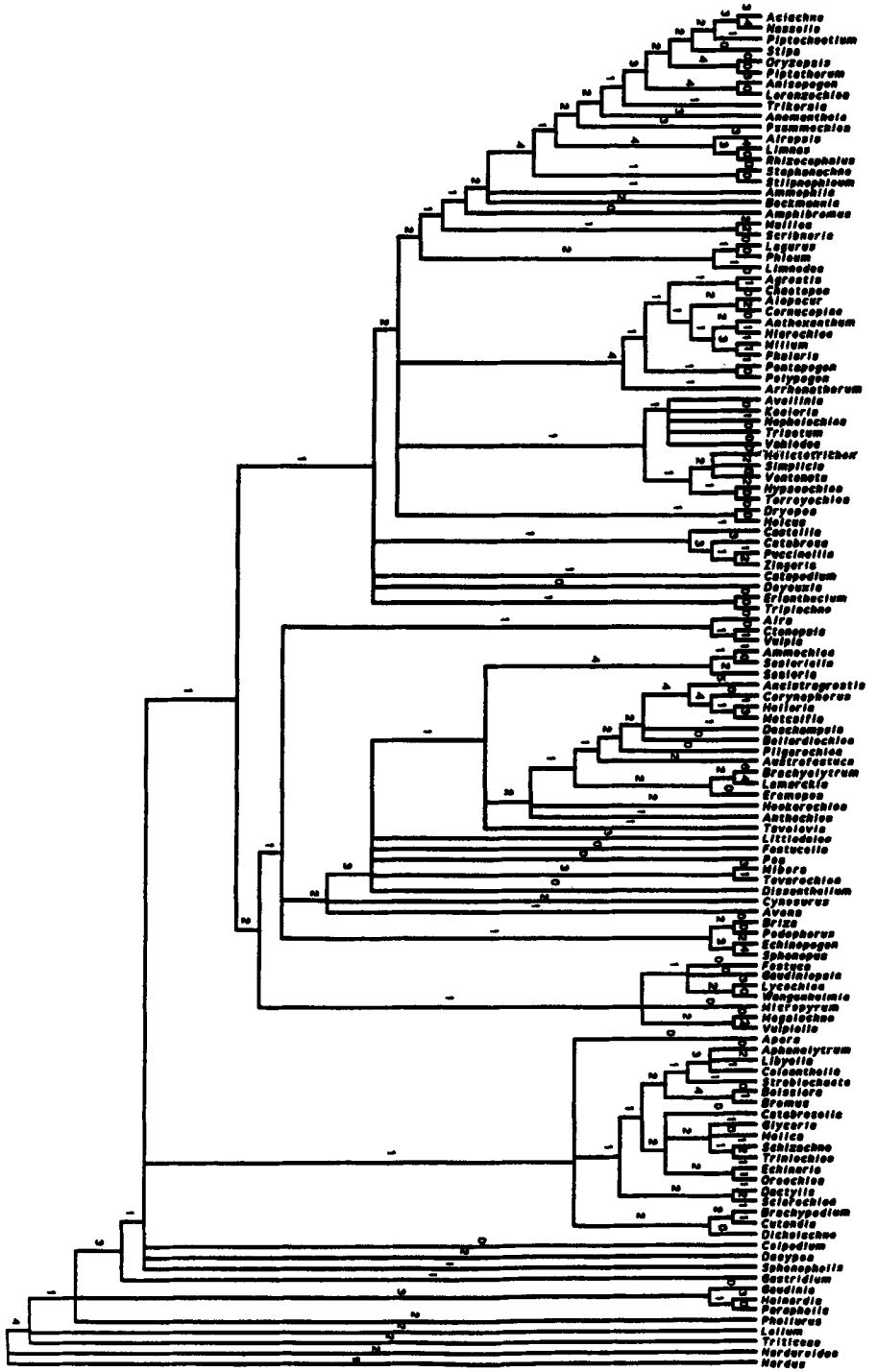

Fig. 10. One phylogenetic hypothesis for the pooid clade. Numbers on left of branches indicate branch lengths with ACCTRAN character optimization. Tree length = 319; CI = 0.266; RI = 0.576.

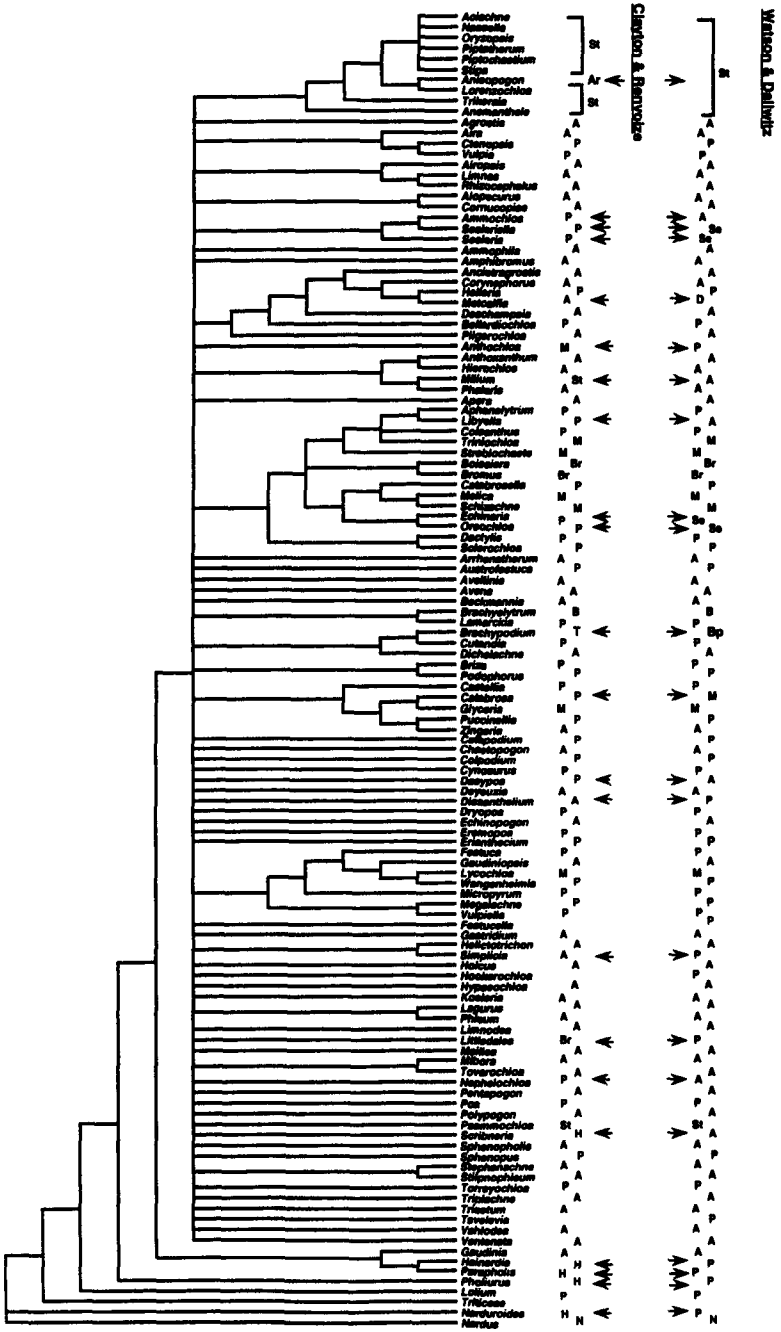

Fig. 11. Pooid clade, strict consensus of 500 trees; statistics as in Fig. 10. Tribal classifications of Clayton and Renvoize (1986) and Watson and Dallwitz (1988) are compared above; arrows indicate points of disagreement. Tribal abbreviations as follows: St - Stipeae; Ar = Arundineae (in Arundinoideae); A = Aveneae; P = Poeae; Se = Seslerieae; D = Danthonieae (in Arundinoideae); M = Meliceae; Br = Bromeae; Bp = Brachypodieae; H = Hainardieae; N = Nardeae.

however, it is clearly a correlate of having sessile spikelets partially embedded in the rachis; its inclusion would have the effect of weighting the inflorescence characters still more than they already are.

Kellogg and Campbell (1987) suggested that the Stipeae were basal and paraphyletic in the pooid clade. None of the analyses performed since have supported that conclusion. As in the trees shown here, the Stipeae are consistently monophyletic and not basal. Moving the Stipeae to the base of the cladogram using MacClade (Maddison & Maddison, 1992) adds 9 steps to the total tree length. The group (including *Psammochloa,* which was placed here in 498 out of the 500 trees) is supported by two consistent characters: (1) lodicules 3 (#91). This character reverses in *Nassella* and in *Anemanthele*; it also appears independently in *Metcalfia*; (2) guard cells of stomata flush with or overlapping interstomatals (#138). This is the plesiomorphic condition, shared with *Nardus,* and is one of the reasons that the Stipeae might have been expected to be the sister taxon to all other Pooideae. However, as noted above, the basal taxa are determined more by inflorescence characters than by anatomical ones. This presumably reflects number of characters as well as a high level of homoplasy throughout the data such that any "signal" is being swamped out by "noise." The character of sunken guard cells is known to be a very consistent character in trees showing subfamilial relationships, but it is nearly invisible in this analysis. The presence of three lodicules in many members of the tribe is also supposed to indicate primitiveness and/or a link with the Bambusoideae (Dahlgren et al., 1985). Because the outgroup, *Nardus,* lacks lodicules the character is unpolarized in this analysis, but other characters do not pull the Stipeae to the base of the tree. The lack of parallel-sided subsidiaries is plesiomorphic in the family as a whole, but appears here as a reversal.

The "core Stipeae" are *Aciachne, Nassella, Piptochaetium, Stipa, Oryzopsis* and *Piptatherum,* a robust group supported in these clades by 2 characters: lemmas becoming indurated when dry (#68) and paleas indurated (#88), characters that appear in parallel in the Phalarideae (q.v.). Character 88 reverses in *Aciachne* and *Nassella,* and 68 also appears in *Hypseochloa* and the Phalarideae.

Phalaris, Hierochloë and *Anthoxanthum* are generally considered to be closely related and are often placed in a tribe of their own, the Phalarideae (Hitchcock, 1935, 1950), traditionally defined by having both proximal and distal incomplete florets. The tribe plus the genus *Milium* is monophyletic in all 500 trees, with the synapomorphy of lemmas becoming indurated when mature, a character that also appears in *Hypseochloa* and in the Stipeae. *Milium* is linked to *Phalaris* by an indurated palea, a loose coleoptile, and a flat leaf blade. This argues against Clayton and Renvoize's (1986) interpretation of *Milium* as "little more than an awnless version of *Oryzopsis.*" The chromosome numbers of *Milium* are based on $x = 4$, 5, 7, and 9 (Petrova, 1975), and the character is thus scored as missing; it might, however, indicate that the genus belongs elsewhere. *Milium* does not have proximal and distal incomplete florets; rather it has only one floret per spikelet with no incomplete florets at all. It is scored as missing data for this character and the algorithm assigns it the same value as *Hierochloë, Anthoxanthum* and *Phalaris.* This may represent an artefact of the coding method. *Anthoxanthum* and *Hierochloë* are joined by having aromatic shoots and only two stamens per hermaphrodite floret. Schouten and Veldkamp (1985) have suggested that *Anthoxanthum* and *Hierochloë* should be combined, which Figs. 10 and 11 show to be cladistically defensible, but not strictly necessary.

The clade of *Limnas* and *Rhizocephalus* appeared in one other preliminary analysis

(tree not shown), but the cladogram presented here is the only one in which *Airopsis* is added to the clade. The characters supporting it are #89 (palea one-keeled), and #104 (fruit compressed). These characters change 9 and 14 times on the tree respectively, indicating that this group is poorly supported despite its appearance in the consensus tree. In many other analyses of this data set, *Limnas* and *Rhizocephalus* appear near the base of the cladogram.

Stephanachne and *Stilpnophleum* form a robust group in many analyses. Here they are linked by having leaf blade sclerenchyma all associated with vascular bundles (#156); they also share character #40 (rachilla apically prolonged), interpreted as a synapomorphy in some analyses. *Phleum* and *Lagurus* appeared as sister taxa in many analyses, with the synapmorphy of awned glumes (#56).

The sister group relationship of *Dactylis* plus *Sclerochloa* is one of the more robust conclusions to emerge from this analysis. The pair are linked by folded leaf vernation (#16) and secund spikelets (#34). They also appear together in the analysis of monotypic taxa, and in many trees in the smaller analysis of monotypic taxa with known diploids. Studies of chloroplast DNA (Soreng et al., 1990, discussed below) find that the two taxa have very different plastid genomes.

Brachyelytrum, placed here instead of the bambusoid clade (see above), has a branch length of 5 steps; it often appeared in or near the base of the Stipeae, although in Fig. 10 it is the sister taxon of *Lamarckia* and *Eremopoa.* A detailed discussion of the problematical character combinations of *Brachyelytrum* is provided by Campbell et al. (1986); although they suggest provisionally placing the genus in the Bambusoideae, they concede that the evidence is inconclusive. The analyses presented here are likewise ambiguous.

The Meliceae in its usual sense is not monophyletic in this analysis. *Melica, Glyceria, Triniochloa,* and *Schizachne* are sister taxa, on the basis of having lodicules that are joined, at least basally (#92), and are distally fleshy (#93). However, *Catabrosa* is linked to *Puccinellia* and *Zingeria,* supported by characters #125 (papillae present on epidermal cells), #146 (leaf blade flat in cross section), and #151 (bulliform cells absent). *Streblochaete* appears with the clade of *Aphanelytrum/Libyella/Coleanthus.*

The Seslerieae, recognized by Watson and Dallwitz (1988) but not by Clayton and Renvoize (1986), are also not monophyletic. *Echinaria* and *Oreochloa* are linked on the basis of non-compressed fruits (#104) and fused styles (#99); fused styles also have arisen independently elsewhere in the tree to link *Sesleria* and *Sesleriella* with *Ammochloa.* The latter group is also supported by having separate staminate and pistillate spikelets on the same plant (#19), spikelets associated with bract-like involucres (#32), and spikelets subsessile (#36).

Bromus and *Boissiera* always appear as sister taxa in these analyses but do not form a clade with either *Brachypodium* or the rest of the Triticeae.

Some characters in the annotated character list in Appendix C1 merit particular attention here. Character #49, length of glumes relative to the adjacent lemmas, consistently changed more that 20 times on the tree and so was excluded from the analyses. This is the character defining the traditional Aveneae, which never, under any circumstances, appears as monophyletic. Likewise #64, flowers one vs. more than one per spikelet, was excluded because of a very low CI, indicating little support for the traditional Agrostideae, which was based on the former character state.

Character #97, ovary hairy or not, changes 12 times on the tree and is thus excluded,

although it is sometimes thought to support the monophyly of the Triticodae. Yet another putatively higher-level character, disarticulation above or below the glumes (#38), was excluded because it changed more than 10 times on the tree.

The various characters related to silica bodies are generally uncorrelated with all other characters, a point already noted by Kellogg and Campbell (1987). This is particularly true of #128, the "pooid type" silica body (horizontally elongated-sinuous or elongated-crenate) which had to be excluded from these analyses because of its large number of changes on the cladograms. Nonetheless, some types of silica bodies may be useful locally.

The lack of hierarchy in the consensus tree in Fig. 11 may need several explanations. First, it may simply reflect the size of the analysis and the inability to conduct an efficient search; this seems unlikely, however, given that the data matrix for the Andropogonodae is of similar size (see above). Second, it may indicate problems with generic circumscription; if a significant proportion of the genera are actually polyphyletic, a cladistic analysis will fail to find hierarchical structure. Third, it may be an accurate description of the pattern of variation, indicating either extensive lateral gene flow (hybridization) or a rapid burst of morphological change. To address the second possibility, we analyzed a data set of only taxa with one or two species, in order to minimize homoplasy caused by improper generic circumscription. The analysis included 70 terminal taxa and found 800 trees (MAXTREES setting) of 259 steps, CI = 0.309, RI = 0.517; the consensus tree is shown in Fig. 12. The overall structure (or lack of it) is the same, indicating that it is not an artefact of poor generic circumscription. The Stipeae and the Triticeae (excluding *Australopyrum*) are monophyletic.

There remains the possibility that some monotypic genera are in fact allopolyploids, erroneously included simply because their chromosome number is unknown. We therefore analyzed only those taxa with known diploids, which produced 1000 equally parsimonious trees of length 133, CI = 0.42 and RI = 0.48; the strict consensus is shown on the left side of Fig. 13. The general structure of the tree is still unresolved. We thus conclude that the high homoplasy and lack of robust groups does not reflect multiple mistakes in generic circumscription. Possible allopolyploid taxa have been removed by restricting the analysis to those with known diploid members, and possible paraphyletic genera have been removed by analyzing only single species. The pattern is therefore not a taxonomic artefact, but rather reflects (1) extensive lateral gene flow, or (2) a rapid burst of morphological change, or (3) true parallelism in the morphological data, of the sort described by Kellogg (1990). We return to this problem below in the discussion of molecular data.

This was the only one of the pooid data sets small enough to generate an Adams consensus tree in a reasonable time; this tree is shown on the right side of Fig. 13. The Adams consensus tree places terminal taxa at the lowest node common to all trees, which thus helps pinpoint "unstable" taxa by putting them in basal polytomies. The topology generated may not be congruent with any one of the equally parsimonious trees. The multiple basal polytomies, show that at least nine of the 34 monotypic genera analyzed (ca. 25%) are unstable in their placement.

Several analyses were done with fewer than 30 genera (trees not shown). In these analyses, many fewer trees were found and consensus trees were clearly hierarchical. In light of the foregoing discussion, relationships suggested by these smaller trees are very likely to be sampling artefacts, the result of leaving out so many genera that existing character conflicts are not discovered. This means that a tree based on other

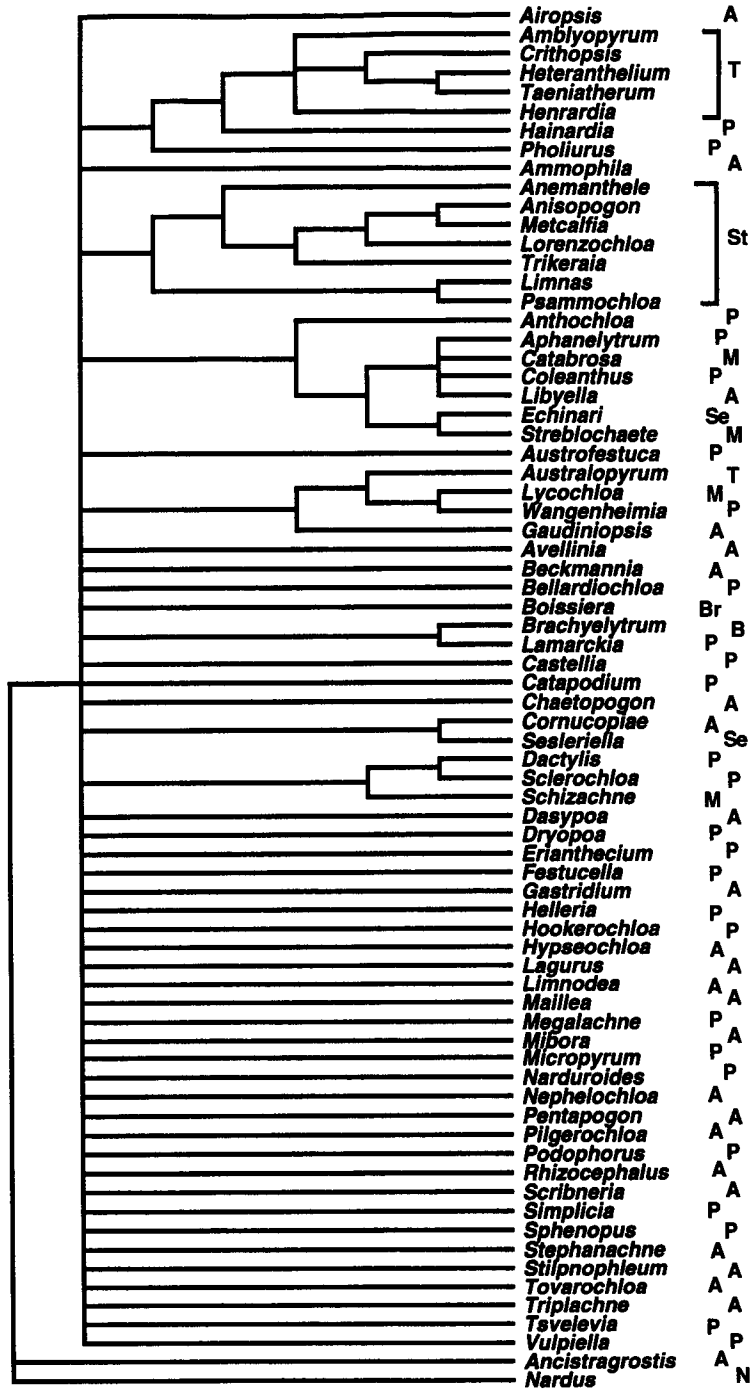

Fig. 12. Pooid clade, genera with only one or two species. Strict consensus of 800 trees, all of length = 259; CI = 0.309; RI = 0.517. Tribal designations following Watson and Dallwitz (1988); abbreviations as in Figure 11.

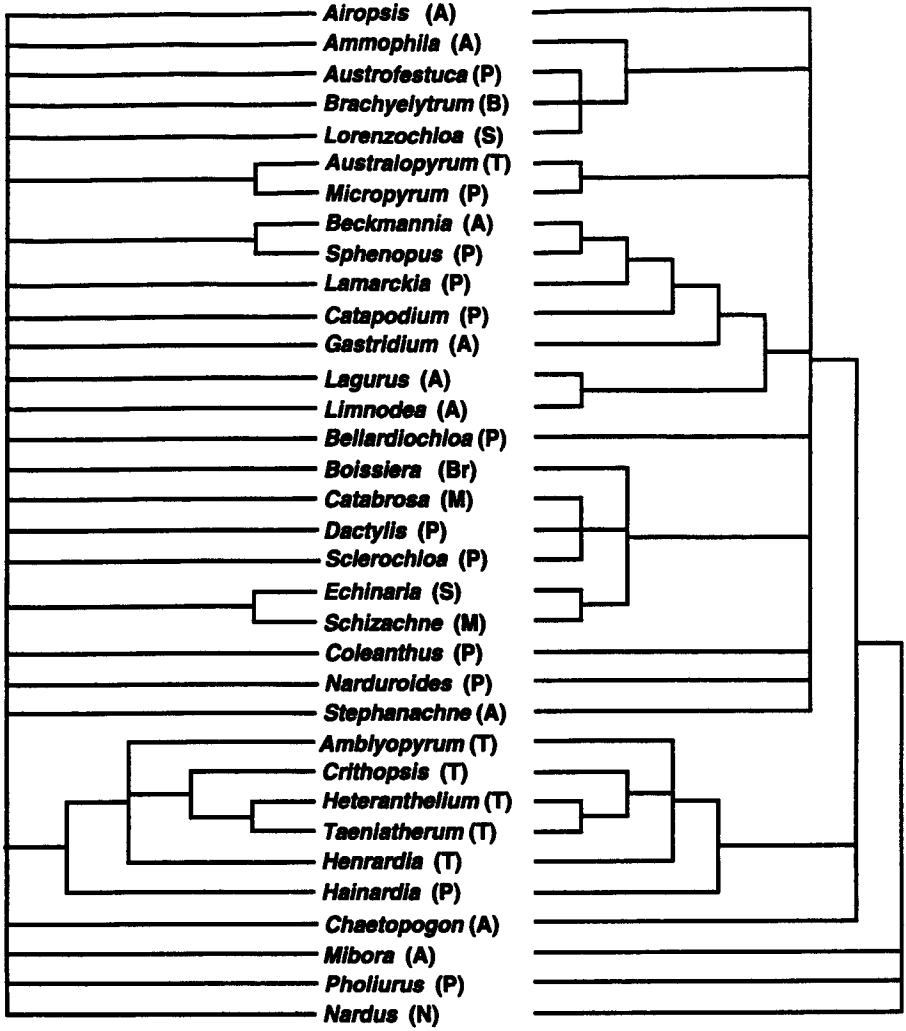

Fig. 13. Pooid clade, genera with only one or two species and at least one documented diploid chromosome count. Left hand cladogram: strict consensus of 1000 trees, all of length = 133; CI = 0.42; RI = 0.48. Right hand cladogram: Adams consensus of same 1000 trees. Tribal designations following Watson and Dallwitz (1988); abbreviations as in Figure 11.

characters (e.g. molecular data) will need to include more than 30 taxa if it is to provide an adequate comparison to the larger trees shown here.

The trees in Figs. 12 and 13 support recognition of the Stipeae and the Triticeae, excluding *Australopyrum,* as monophyletic tribes. Three of the Seslerieae form a monophyletic group, viz. *Echinaria, Oreochloa,* and *Sesleria.* No other groups in the consensus tree are stable and appeared consistently in other analyses of this data set. They are thus not discussed in detail. There is no support at all for the Aveneae, Poeae, or Agrostideae, and therefore no defense for their continued use, unless it be purely for ease of key construction.

IV. General Discussion

A. GENERAL IMPLICATIONS FOR LARGE DATA SETS

As data sets become larger, they will become harder to analyze. The data presented here illustrate an iterative approach to the problem, whereby a large data set is divided into putatively monophyletic groups and analysed. A set of functional outgroups is then determined for each monophyletic group, which is then analysed to determine monophyletic subgroups. Representatives of the monophyletic subgroups will then be used for a second full-scale analysis aimed at elucidating the overall structure of the whole study group.

Such an iterative approach is necessitated by the tension between computation time and sampling artefacts. If a sample of taxa is used, the amount of homoplasy will be underestimated, sometimes severely so, and erroneous relationships may be produced (analogous to extinction). If taxa are combined as summary taxa, they will have more variation and hence more missing data and their cladistic position may be misleading (Nixon & Davis, 1991; Platnick et al., 1991). However, if all taxa are included, the number of possible trees becomes so large that efficient exploration of the data set becomes impossible, and the shortest tree becomes a mere phantom. Here we have explored in some detail the largest data sets that can be conveniently handled by PAUP 3.0 on a Macintosh—ca. 100 terminal taxa. These will then be used to guide sampling for a larger family-level analysis. The samples will be chosen from the basal members of the major clades; hence for the bamboos, *Zizania, Buergersiochloa, Phyllorhachis, Pharus, Guaduella* and *Diandrolyra* might be used in a revised family-level analysis, along with the putative basal group *Ehrharta.* For the Andropogonodae, the choice would be *Polytrias, Sehima, Andropterum, Lasiurus, Chrysopogon* and *Pogonatherum,* along with *Arundinella* (representing the Arundinelleae) and the Neurachneae. The lack of hierarchy in the non-triticoid pooids makes the choice more arbitrary, but would probably include *Stipa, Bromus, Melica, Lolium, Pholiurus,* and *Narduroides,* as well as *Nardus.*

Data quality (including questions of character state definition) and extent of missing data remain problems with the analyses presented here, as with any analysis. Not all characters and character systems can be examined with equal care or across identical sets of taxa. In a study that relies on data from multiple sources, the quality of the data is certain to be inconsistent and cannot always be conveniently rechecked. There is no easy solution to this problem, other than care in interpreting results. A critical analysis should be able to pinpoint particular characters or taxa for which reexamination will be most fruitful. Missing data also contributes to analytical problems. Extensive missing data tends to slow the tree search process and so exacerbates the

difficulties caused by large amounts of homoplasy. Our general observation has been that, whatever the limitations in speed of available algorithms, the ultimate limiting factor is thus the quality of the data base.

This study has produced several general conclusions.

1. Very large data sets can be analyzed by an iterative approach, in which large monophyletic groups are defined and their interrelationships tested; these interrelationships are then used to structure subanalyses of the monophyletic groups. The groups are then divided up to reintroduce into a global analysis. The central point here is that all decisions are made on the basis of characters in the one data base. If (for example) the data set to be analyzed is made up of DNA sequences, then the first assessment of monophyletic clusters needs to be made on the basis the sequence data, not on external (e.g. morphological) criteria. If the latter is done, then the two data sets are in effect being analyzed together and cannot be then interpreted as independent estimates of phylogeny.

2. Small samples of taxa can result in artefactual structure; hence larger (more global) analyses are preferable. Large analyses, though, strain current computer algorithms, and limit the ability to explore data structure. This problem is exacerbated by missing data.

3. Although the analyses presented here are based on morphological data, the general conclusions are equally applicable to molecular data. Data quality may become a particularly difficult problem with molecular sequence data. It is often very difficult to check such data without redoing an entire sequence, perhaps more difficult than, for example, re-examining herbarium specimens or anatomical slides.

4. Large data sets have problems of their own. Many methods developed for smaller data sets (for example, any test of the statistical support for a given cladogram) simply cannot be applied in a reasonable amount of time. Some of these problems will be solved by more powerful computer algorithms, but in many cases approximations may always be unavoidable. The costs, in terms of statistical power or tree resolution, will need to be assessed. There are many phylogenetic problems that are not computationally convenient; we need to develop ways to attack them.

B. WHAT NEXT FOR MORPHOLOGICAL DATA?

For these three data sets, there are two possible future approaches. First, they can be analyzed and reanalyzed with ever-more-sophisticated programs, and second, the data can be checked and refined to improve both the extent and the accuracy of the character coding. Both are necessary. Unless computer programs improve so that data analyses of this size can be run in a very few hours, the ability to explore the data will continue to be limited. Very long runs may be justified, but only if the investigator has confidence in the details of the data matrix. Time may be best spent in gaining a more detailed understanding of the characters and the organisms.

Character delimitation is always the critical aspect of a morphological study. The difficulties are aggravated when the number of taxa is large and similarities need to be assessed across a broad range of forms. When the database is so large that a single investigator can no longer check each character for each taxon him- or herself, then the problem can easily outweigh any algorithmic limitations. The matrix is rarely as powerful as the analytical methods.

Ideally, one would work with species descriptions, to minimize problems with

interpretation and scoring of polymorphism. However, this almost certainly means that data matrices will become even more unwieldy than they already are, and/or computer searches would have to be confined to inadequate samples. Many of the problems could be minimized by careful attention to comparative morphology, the interpretation of characters and their subsequent scoring for analysis. It will also be important to have more definitive information on the hybrid origins of particular taxa (see Section IVC below).

The following groups are in particular need of attention:

1. In the bambusoid clade, the Ehrharteae, particularly *Tetrarrhena* and *Microlaena* and their relationship to *Ehrharta* itself influence the basal structure of the clade and need more careful evaluation. Generic limits of the woody bamboos are, as noted above, in dire need of attention and should be given high priority.

2. In the Andropogonodae, the positions of *Andropterum, Oxyrhachis,* and *Garnotia* are anomalous; their morphology needs careful study. In particular, developmental studies on the inflorescence in the Rottboelliinae and outlying genera should clarify the assessment of homologous inflorescence types. Generic limits in the *Saccharum* group are unclear, as are generic limits in the Rottboelliinae.

3. Morphological data may continue to be uninformative for phylogenetic relationships in the Pooideae, at least at the level of the subfamily. Clearly any future morphological work will have to address questions of homology and character state division in a much more sophisticated way than has been possible here.

C. THE DETECTION OF HYBRIDIZATION

From the foregoing pages, it should be clear that potential hybridization is a continuing problem for phylogenetic analyses, and uncertainty about its frequency complicates interpretation of cladograms. Hybridization has been postulated to be a major force in the evolution of the angiosperms (for example, Stebbins (1950, p. 252) has suggested that most plant genera that create problems for classification are of hybrid origin). Because it introduces a reticulate pattern into an evolutionary tree, it is not easily accommodated in an analytical method that assumes strict divergence (Cronquist, 1987; Funk, 1981; Kellogg, 1989). Both theoretical and empirical studies show (Funk, 1981, 1985; Humphries, 1983; McDade, 1990, 1992) that hybrids do not create any predictable pattern in a cladogram and thus cannot be readily detected. The morphology of hybrids may be intermediate (McDade, 1990; Wagner, 1983), virtually indistinguishable from one parent (Bennett, 1984), or completely unlike either parent (Rieseberg et al., 1990). Allopolyploid genera can be considered as likely hybrids, and can be removed from an analysis, as we have done here, but there remain two other possible sources of confusion. The first is hybridization at the diploid level, and the second, hybridization that produces polyploids similar enough to one parent that they are included in the same taxon, but increase its variability with respect to one or more characters. These "cryptic hybrids" may not be detected by cladistic analysis of any sort of characters. If hybrids are detectable at all, it will require molecular data that allow individual genomes to be tracked. If recombination is common, however, then even molecular data will be unable to recover phylogeny. The best that can be hoped for in this case will be to pinpoint where the phylogeny becomes ambiguous and unrecoverable.

Clearly the pattern of character conflict in the pooid clade could be caused by hybridization. Reference to Funk's (1985) guidelines for detecting possible hybrids

and their parents shows that they apply to many pooid taxa. Taxa that are defined solely by character conflict, taxa with reversals and taxa whose position is different in equally parsimonious cladograms may be hybrids; taxa without autapomorphies may be parents. The clear structure of the Andropogonodae and the bambusoid trees implies that hybridization is less rampant, but still could be invoked as a testable hypothesis for the origin of some "unstable" taxa. As demonstrated by McDade (1992), inclusion of a few hybrids does not necessarily create unresolved trees. Recall that *Dichanthium, Bothriochloa* and *Capillipedium* are known to hybridize (DeWet & Harlan, 1970), yet do not create lack of resolution in the cladogram (Section IIB2, above, and Figs. 7 and 8); they also do not form a monophyletic group.

D. COMPARISONS WITH MOLECULAR DATA

There are as yet few generic-level molecular phylogenies to compare with the morphological cladograms. There is no published molecular phylogeny of bambusoids. Friar and Kochert (1991) have described species-specific probes for *Phyllostachys.* Kanno and Hirai (1992) have studied chloroplast DNA variation in *Oryza,* and Wang et al. (1992) analyzed RFLP variation in the nuclear genome. None of these studies, however, has addressed generic level relationships within the bambusoid group. Duvall et al. (1993) used chloroplast restriction site mutations to infer relationships among North American Oryzeae; their results are congruent with the morphological cladograms, demonstrating a monophyletic Oryzeae (relative to *Olyra*), and a somewhat ambiguous relationship between *Zizania* and *Luziola.*

Data on *rbcL* sequences place *Oryza* as the sister taxon to all other grasses (Chase et al., 1993). Ribosomal RNA sequences place *Oryza* at the base of a panicoid clade and *Arundinaria* (a bamboo) sister to the rest of family (Hamby & Zimmer, 1988, 1992). The *rbcL* and the rRNA data are thus at odds with each other and with the morphological data, which places bamboos and rices together. However, sequences of a portion of *rpo*C2, the chloroplast gene for the ß" subunit of RNA polymerase II, place *Oryza* with *Ehrharta,* in accord with morphological cladograms (Cummings et al., forthcoming). Chloroplast restriction site data suggest that the Bambusoideae are in fact polyphyletic, with the woody bamboos being sister taxon to the Centotheceae, and with the Oryzeae part of a clade with *Nardus, Pharus* and *Brachyelytrum* (Davis & Soreng, 1993). Of the four data sets (*rbc*L, rRNA, *rpo*C2, and chloroplast restriction sites) only the latter two sample all five subfamilies and are thus less likely to show artefactual groups. Nonetheless, the number of taxa sampled in each is miniscule compared with the morphological data set and comparisons need to be made cautiously.

There is likewise little data bearing on the Andropogoneae — *rbc*L sequences are available for *Zea* and *Sorghum,* and the two are sister taxa in the study of Doebley et al. (1990). Duvall and Doebley (1990) studied restriction site variation in the chloroplast genome of *Sorghum,* and found that genus is either para- or polyphyletic; furthermore, the chloroplast of the genus *Cleistachne* is similar enough to that of *Sorghum* that the two genera might be part of a single lineage. The potential polyphyly of *Sorghum* may in part explain its position as part of a multichotomy in the consensus tree in Fig. 8; note, however, that it remains morphologically quite distinct from *Cleistachne* which appears as the sister taxon of *Dimeria.* [The distant relationship of *Cleistachne* and *Sorghum* is not an artefact of the cladistic algorithm; the two are also

widely separated in phenograms (data not shown).] Davis and Soreng (1993), also using plastid restriction site variation, showed that *Miscanthus* (Andropogoninae) and *Zea* (Maydeae) form a strongly supported clade. Ribosomal RNA shows that *Zea* and *Tripsacum* are sister taxa, as indicated by all morphological and cytological data available, and *Saccharum* and *Sorghum* are also sister taxa (Andropogoninae). The four taxa together form a monophyletic andropogonoid group. The higher-level structure of the Andropogonodae is thus supported by the rRNA data; however, the cladistic structure may be "clear" only because there are no data for most of the "difficult" genera. In the *rbc*L cladogram of Doebley et al. (1990), *Neurachne* appears basal to the entire subfamily and may thus be more distantly related to the Andropogonodae than would appear from Figs. 7 and 8.

More molecular data are available for the Pooideae; this is fortunate because the morphological data have produced such ambiguous results, as shown in the trees in Figs. 11–13. There are three possible reasons for the lack of resolution of these trees: (1) A rapid burst of evolution producing a pattern of short basal branches followed by longer terminal branches; (2) Extensive intergeneric gene flow; (3) True parallelism in morphological characters, rendering them inadequate for phylogenetic inference. These three possibilities can, in principle, be differentiated by comparing the morphological cladogram with cladograms using chloroplast (cp) DNA and a nuclear gene (biparental). If the pattern is produced by a rapid radiation followed by a long period of anagenetic change, cpDNA and nuclear cladograms should be as unresolved as the morphological one. If the pattern is due to extensive gene flow then a cpDNA cladogram may be hierarchical but incongruent with that of a nuclear gene. If the morphological data are simply inherently homoplasious, then cpDNA and nuclear DNA should give similar clear well-supported clades with little homoplasy.

Interpretation of the cladistic pattern (or lack of it) is thus potentially resolvable by recourse to comparable molecular data sets. However, as will be seen in the following discussion, they must be on a comparable set of taxa. A small set of genera will give apparently well-supported trees in which low homoplasy reflects inadequate sampling of the variation in the larger group. This was shown by Kellogg (1992) in studies of six pooid grasses; a cladogram based on morphological characters was only slightly different in topology and consistency index from one based on chloroplast DNA.

The use of *Nardus* to root the pooid cladogram is supported by the data of Cummings et al. (forthcoming), in which sequences of *rpo*C2 link *Nardus* and *Lygeum* to a clade made up of *Bromus, Briza* and *Phleum,* but is contradicted by Davis and Soreng (1993) who found that *Nardus* is unrelated to the Pooideae. Molecular data within the pooid clade are provided by Soreng et al. (1990), based on cpDNA restriction site variation in 28 species. The Meliceae are the basal clade, followed by the Stipeae, although reversing the positions of the two groups requires a tree only one step longer. [The tree presented by Davis and Soreng (1993), on a more restricted set of pooid species, places the Stipeae as the sister group to all other pooids.] The two tribes are monophyletic. The Triticodae are monophyletic and are the sister taxon to all remaining pooids in the tree of Soreng et al. (1990), but are nested within the pooids in the tree of Davis and Soreng (1993). *Arctagrostis* (2n=28, 56, 62; Bowden, 1960) is likely to be hybrid with a *Poa*-like maternal parent. *Leucopoa sclerophylla* (=*Hesperochloa*; 2n=56) is likewise potentially of hybrid origin with a species of *Festuca* as the maternal parent. If polyploid (polyphyletic) genera are to be avoided, and if the hybrid origin of these taxa is supported by future data, then they should be returned to their

respective parental genera [as was done for *Leucopoa* by Clayton and Renvoize (1986)]. The morphological tree places *Dactylis* unequivocally with *Sclerochloa,* yet the chloroplast tree keeps the two genera firmly in separate clades. In the chloroplast tree, *Sclerochloa* is placed with other species (*Puccinellia distans, Catabrosa aquatica*) which it resembles in having somewhat blunt lemmas with nearly parallel nervation, but this character is not even hinted at in *Dactylis*. Either the lemma character is a parallelism and the morphological cladogram reflects the organismal phylogeny, or the joined sheath margins, lack of silica in the short cells, the nodular leaf cross section, and the plicate leaf vernation shared by the two taxa are convergences and the chloroplast group reflects the organismal phylogeny. A similar point can be made about the chloroplast group called "Aveneae" by Soreng et al. (1990). In the chloroplast cladogram it is the most strongly supported of the groups in the core Pooideae, but is not remotely similar to any group (however unstable) found in the morphological analysis.

Imposing the structure of the chloroplast cladogram on the morphological data results in a cladogram that is 30 steps longer than the most parsimonious morphological tree for those taxa (203 vs. 173 steps). Again, until more cpDNA data from more taxa are collected, it is impossible to say if the chloroplast phylogeny can be interpreted as an organismic phylogeny. All groups in the chloroplast cladogram are somewhat suspect until the sample size is enlarged; on the other hand, many groups in the morphological cladogram are so susceptible to small permutations of the data matrix that they are equally questionable. At present, there is no clear way to choose between the cladograms.

The above observations could be generalized to any studies involving molecular data. All current molecular data sets that address generic relationships in the grasses include only a very small sample of the taxa, much smaller than have been included in the morphological analyses. Molecular data may be as prone to convergence and parallelism as morphological data (Sanderson & Donoghue, 1989; Donoghue & Sanderson, 1992); hence comparisons between morphological and molecular data sets will need to be done on comparable sets and numbers of taxa to be reliably interpreted. We see that the sample of taxa studied does influence the topology of the cladogram, as would be expected from any data with appreciable homoplasy. Thus a small sample of taxa from any group, using any set of data, might well give artefactual groups.

This brings us back to the central point of this article. We conclude that large data sets are frequently preferable to small ones, because they are likely to give a more accurate picture of character state change and hence of phylogenetic relationships. However, as illustrated here, large data sets have problems of their own. Assessing and maintaining accuracy of data is difficult. Construction and comparison of trees is laborious and can be severely limited by available hardware and software. Randomization and resampling methods are virtually impossible. We hope that as more systematists work with large data sets, methods will be developed to address these problems. This article suggests an iterative approach that involves first, principled, character-directed sampling of the entire study group, second, complete analyses of sub-groups, and third, cladistically-informed sampling and re-analysis of the entire group. Ultimately, the reliability of this approach must be tested by comparison with independent data sets, but such independent (presumably molecular) data sets will need to be equally large.

Acknowledgments

We would like to thank M. J. Dallwitz for help in modifying DELTA for the purposes of this investigation; the computing staff at CSIRO and RSBS for their patience; J. Beach, M. Cummings, D. Davison, D. Maddison for help and access to computing facilities; P. Cranston and J. Davis for running the data through Hennig86; J. Carpenter, P.F. Stevens, and J. Stewart for use of their personal computers; J. Bruhl, J. Carpenter, M. Henwood, J. Lenz, and many others for helpful discussion; N. Gift, M. Hershkovitz, H. P. Linder, P. F. Stevens and an anonymous reviewer for comments on the manuscript. EAK thanks the Australian National University for fellowships as a Visiting Scholar in 1988 and 1992; an unrestricted grant from the Stevens Foundation throughout the study; and the Arnold Arboretum and NSF Grant #9106581 for support during the completion of the manuscript.

Literature Cited

Allard, M. W., & M. M. Miyamoto. 1992. Perspective: testing phylogenetic approaches with empirical data, as illustrated with the parsimony method. Molec. Biol. Evol. **9:** 778–786.

Archie, J. W. 1989. A randomization test for phylogenetic information in systematic data. Syst. Zool. **38:** 239–252.

Bennett, M. D. 1984. Pages 469–502 *in* J. P. Gustafson (ed.), Gene manipulation in plant improvement. Plenum, New York.

Bowden, W. M. 1960. Chromosome numbers and taxonomic notes on northern grasses. Canad. J. Bot. **38:** 117–131.

Brown, R. 1810. Prodromus Florae Novae Hollandieae et Insulae Van-Diemen. Volume 1. J. Johnson and Co., London.

———. 1814. General remarks, geographical and systematical, on the botany of Terra Australis, Appendix 3. Pages 533–613 *in* M. Flinders, A voyage to Terra Australis; undertaken for the purpose of completing the discovery of that vast country, and prosecuted in the years 1801, 1802, and 1803. W. Bulmer & Company, London.

Bruneau, A., & J. J. Doyle. 1993. Cladistic analysis of chloroplast DNA restriction site characters in *Erythrina* (Leguminosae: Phaseolae). Syst. Bot. **18:** 229–241.

Campbell, C. S., P. E. Garwood, & L. P. Specht. 1986. Bambusoid affinities of the north temperate genus *Brachyelytrum* (Gramineae). Bull. Torrey Bot. Club. **113:** 135–141.

Carpenter, J. M. 1988. Choosing among multiple equally parsimonious cladograms. Cladistics **4:** 291–296.

Chase, M. W., D. E. Soltis, R. G. Olmstead, D. Morgan, D. H. Les, B. D. Mishler, M. R. Duvall, R. A. Price, H. G. Hills, Y.-L. Qiu, K. A. Kron, J. H. Rettig, E. Conti, J. D. Palmer, J. R. Manhart, K. J. Sytsma, H. J. Michaels, W. J. Kress, K. G. Karol, W. D. Clark, M. Hedrén, B. S. Gaut, R. K. Jansen, K.-J. Kim, C. F. Wimpee, J. F. Smith, G. R. Furnier, S. H. Strauss, Q.-Y. Xiang, G. M. Plunkett, P. S. Soltis, S. M. Swensen, S. E. Williams, P. A. Gadek, C. J. Quinn, L. E. Eguiarte, E. Golenberg, G. H. Learn, Jr., S. W. Graham, S. C. H. Barrett, S. Dayanandan, & V. A. Albert. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbc*L. Ann. Missouri Bot. Gard. **80:** 528–580.

Clayton, W. D. 1966. Studies in the Gramineae. 9. Andropogoneae. Kew Bull. **20:** 257–274.

———. 1970. Studies in the Gramineae. 21. *Coelorhachis* and *Rhytachne*: a study in numerical taxonomy. Kew Bull. **24:** 309–314.

———. 1973. Studies in the Gramineae. 33. The awnless genera of Andropogoneae. Kew Bull. **28:** 49–58.

——— **& S. A. Renvoize.** 1986. Genera graminum. Grasses of the world. Her Majesty's Stationery Office, London.

Clifford, H. T. 1987. Spikelet and floral morphology. Pages 21–30 *in* T. R. Soderstrom, K. W. Hilu, C. S. Campbell, & M. E. Barkworth (eds.), Grass systematics and evolution. Smithsonian Institution Press, Washington, D.C.

Cranston, P. S., & C. J. Humphries. 1988. Cladistics and computers: A chironomid conundrum? Cladistics **4:** 72–92.

Cronquist, A. 1987. A botanical critique of cladism. Bot. Rev. **53:** 1–52.

Cummings, M. P., L. M. King, & E. A. Kellogg. forthcoming. Slipped-strand mispairing in a plastid gene: *rpo*C2 in grasses (Poaceae). Molec. Biol. Evol. **11.** (In press.)

Dahlgren, R. M. T., H. T. Clifford, & P. F. Yeo. 1985. The families of the monocotyledons. Springer-Verlag, Heidelberg, Berlin, New York.

Dallwitz, M. J. 1980. A general system for coding taxonomic descriptions. Taxon **29:** 41–46.

———. 1992. DELTA and INTKEY. *In* R. Fortuner (ed.) Proceedings of workshop on artificial intelligence and modern computer methods for systematic studies in biology. University of California, Davis.

Davis, J. I., & R. J. Soreng. 1993. Phylogenetic structure in the grass family (Poaceae), as determined from chloroplast DNA restriction site variation. Amer. J. Bot. (In press.)

deKoning, R., M. S. M. Sosef, & J. F. Veldkamp. 1983. A revision of *Heteropholis* and *Thaumastochloa* (Gramineae). Gard. Bull. Straits Settlem. **36:** 137–162.

De Rijk, P., J.-M. Neefs, Y. Van de Peer, & R. DeWachter. 1992. Compilation of small ribosomal subunit RNA sequences. Nucl. Acids Res. **20** (Suppl): 2075–2089.

DeSalle, R., & D. Grimaldi. 1991. Morphological and molecular systematics of the Drosophilidae. Annual Rev. Ecol. Syst. **22:** 447–475.

DeWet, J. M. J., & J. R. Harlan. 1970. Apomixis, polyploidy, and speciation in *Dichanthium.* Evolution **24:** 270–277.

Doebley, J., M. Durbin, E. M. Golenberg, M. T. Clegg, & D. P. Ma. 1990. Evolutionary analysis of the large subunit of carboxylase (*rbc*L) nucleotide sequence among the grasses (Gramineae). Evolution **44:** 1097–1108.

Donoghue, M. J., J. A. Doyle, J. Gauthier, A. G. Kluge, & T. Rowe. 1989. The importance of fossils in phylogeny reconstruction. Annual Rev. Ecol. Syst. **20:** 431–460.

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

Duvall, M. R., & J. F. Doebley. 1990. Restriction site variation in the chloroplast genome of *Sorghum* (Poaceae). Syst. Bot. 15: 472–480.

———, **M. T. Clegg, M. W. Chase, W. D. Clark, J. W. Kress, E. A. Zimmer, H. G. Hills, L. E. Eguiarte, J. F. Smith, B. S. Gaut, & G. H. Learn, Jr.** 1993. Rapid radiation of ancestral monocotyledons into seven primary lineages indicated by analysis of the DNA sequence of a plastid locus obtained from one-hundred four species. Ann. Missouri Bot. Gard. **80:** 607–619.

———, **P. M. Peterson, E. E. Terrell, & A. H. Christensen.** 1993. Phylogeny of North American oryzoid grasses as construed from maps of plastid DNA restriction sites. Amer. J. Bot. **80:** 83–88.

Ellis, R. P. 1987a. Leaf anatomy of the genus *Ehrharta* (Poaceae) in southern Africa: the Setacea group. Bothalia **17:** 75–89.

———. 1987b. Leaf anatomy of the genus *Ehrharta* (Poaceae) in southern Africa: the Villosa group. Bothalia **17:** 195–204.

Faith, D. P., & 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. 1969. A successive approximations approach to character weighting. Syst. Zool. **18:** 374–385.

———. 1983. The logical basis of phylogenetic analysis. Pages 7–36 *in* N. I. Platnick & V. A. Funk (eds.), Advances in Cladistics: Proceedings of the Second Meeting of the Willi Hennig Society. Columbia University Press, New York.

———. 1988. Hennig86, version 1.5. Published by the author, Port Jefferson Station, N. Y.

———. 1989. The retention index and the rescaled consistency index. Cladistics **5:** 417–419.

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

Friar, E., & G. Kochert. 1991. Bamboo germplasm screening with nuclear restriction fragment length polymorphisms. Theor. Appl. Genet. **82:** 697–703.

Funk, V. A. 1981. Special concerns in estimating plant phylogenies. Pages 73–86 *in* V. A. Funk & D. R. Brooks (eds.), Advances in Cladistics: Proceedings of the First Meeting of the Willi Hennig Society. New York Botanical Garden, New York.

———. 1985. Phylogenetic patterns and hybridization. Ann. Missouri Bot. Gard. **72:** 681–715.

Gibbs-Russell, G. E. 1987. Taxonomy of the genus *Ehrharta* (Poaceae) in southern Africa: the Setacea group. Bothalia **17:** 67–73.

——— **& R. P. Ellis.** 1987. Species groups in the genus *Ehrharta* (Poaceae) in southern Africa. Bothalia **17:** 51–65.

——— & ———. 1988. Taxonomy and leaf anatomy of the genus *Ehrharta* (Poaceae) in southern Africa: the Dura group. Bothalia **18:** 165–171.

Gutell, R. R., M. N. Schnare & M. W. Gray. 1992. A compilation of large subunit (23S- and 23S-like) ribosomal RNA structures. Nucl. Acids Res. **20**(Suppl.): 2095- 2109.

Hamby, R. K. & E. A. Zimmer. 1988. Ribosomal RNA sequences for inferring phylogeny within the grass family (Poaceae). Pl. Syst. Evol. **160:** 29–37.

——— & ———. 1992. Ribosomal RNA as a phylogenetic tool in plant systematics. Pages 50–91 *in* P. S. Soltis, D. E. Soltis, & J. J. Doyle (eds.), Molecular systematics of plants. Chapman and Hall, New York.

Hattersley, P. W. 1987. Variations in photosynthetic pathway. Pages 49–64 *in* T. R. Soderstrom, K. W. Hilu, C. S. Campbell, & M. E. Barkworth (eds.), Grass systematics and evolution. Smithsonian Institution Press, Washington, D.C.

Hauser, D. L. & W. Presch. 1991. The effect of ordered characters on phylogenetic reconstruction. Cladistics 7: 243–265.

Humphries, C. J. 1983. Primary data in hybrid analysis. Pages 89–103 *in* N. I. Platnick & V. A. Funk (eds.), Advances in Cladistics: Proceedings of the Second Meeting of the Willi Hennig Society. Columbia University Press, New York.

Jain, S. K. 1970. The genus *Manisuris* L. (Poaceae) in India. Bull. Bot. Surv. India **12:** 6–17.

Johnston, C. R. & L. Watson. 1976. Microhairs: a universal characteristic of non- festucoid grass genera? Phytomorphology 26: 297–301.

Kanno, A. & A. Hirai. 1992. Comparative studies of the structure of chloroplast DNA from four species of *Oryza*: cloning and physical maps. Theor. Appl. Genet. **83:** 791–798.

Kellogg, E. A. 1989. Comments on genomic genera in the Triticeae. Amer. J. Bot. **76:** 796–805.

———. 1990. Ontogenetic studies of florets in *Poa* (Gramineae): allometry and heterochrony. Evolution **44:** 1978–1989.

———. 1992. Tools for studying the chloroplast genome in the Triticeae (Gramineae): an *Eco* RI map, a diagnostic deletion, and support for *Bromus* as an outgroup. Amer. J. Bot. **79:** 186–197.

——— **& C. S. Campbell.** 1987. Phylogenetic analyses of the Gramineae. Pages 310–322 *in* T. R. Soderstrom, K. W. Hilu, C. S. Campbell, & M. E. Barkworth (eds.), Grass systematics and evolution. Smithsonian Institution Press, Washington, D.C.

Kluge, A. G., & J. S. Farris. 1969. Quantitative phyletics and the evolution of the anurans. Syst. Zool. **18:** 1–32.

Knobloch, I. W. 1968. A checklist of crosses in the Gramineae. East Lansing, Michigan.

Ladiges, P. Y., M. R. Newnham & C. J. Humphries. 1989. Systematics and biogeography of the Australian "green ash" eucalypts (*Monocalyptus*). Cladistics **5:** 345–364.

Lake, J. A. 1987. A rate-independent technique for analysis of nucleic acid sequences: evolutionary parsimony. Molec. Biol. Evol. **4:** 167–191.

Li, Wen-Hsiung & M. Gouy. 1992. Statistical methods for testing molecular phylogenies. Pages 249–277 *in* M. M. Miyamoto & J. Cracraft (eds.), Phylogenetic analysis of DNA sequences. Oxford University Press, New York and Oxford.

Maddison, D. R. 1991. The discovery and importance of multiple islands of most-parsimonious trees. Syst. Zool. 40: 315–328.

Maddison, W. P., M. J. Donoghue & D. R. Maddison. 1984. Outgroup analysis and parsimony. Syst. Zool. **33:** 83–104.

——— **& D. R. Maddison.** (1992) MacClade. Analysis of phylogeny and character evolution. Version 3. Sinauer Associates, Sunderland, Massachusetts.

McDade, L. A. 1990. Hybrids and phylogenetic systematics I. Patterns of character expression in hybrids and their implications for cladistic analysis. Evolution **44:** 1685–1700.

———. 1992. Hybrids and phylogenetic systematics II. The impact of hybrids on cladistic analysis. Evolution **46:** 1329–1346.

Mickevich, M. F. 1982. Transformation series analysis. Syst. Zool. **31:** 461–478.

——— **& J. S. Farris.** 1984. PHYSYS (phylogenetic analysis system). Documentation for VAX/VMS installation at Harvard University, Cambridge.

Neff, N. A. 1986. A rational basis for a priori character weighting. Syst. Zool. **35:** 110–123.

Nixon, K. C. & J. I. Davis. 1991. Polymorphic taxa, missing values and cladistic analysis. Cladistics **7:** 233–241.

Pankhurst, R. J. 1986. A package of computer programs for handling taxonomic databases. Computer Applic. Biosci. **2:** 33–39.

Petrova, O. A. 1975. On the main chromosome numbers in the genus *Milium*. Bot. Žurn. **60:** 393–394.

Pimentel, R. A. & R. Riggins. 1987. The nature of cladistic data. Cladistics 3: 201–209.

Platnick, N. I., C. E. Griswold, & J. A. Coddington. 1991. On missing entries in cladistic analysis. Cladistics **7:** 337–343.

Pohl, R. W. & L. G. Clark. 1992. New chromosome counts for *Chusquea* and *Aulonemia* (Poaceae: Bambusoideae). Amer. J. Bot. **79:** 478–480.

Reeder, J. R. 1957. The embryo in grass systematics. Amer. J. Bot. **44:** 756–769.

Renvoize, S. A. 1985. A survey of leaf-blade anatomy in grasses. VI. Stipeae. Kew Bulletin **38:** 469–478.

Rieseberg, L. H., R. Carter, & S. Zona. 1990. Molecular tests of the hypothesized hybrid origin of two diploid *Helianthus* species (Asteraceae). Evolution **44:** 1498–1511.

Saitou, N. & M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425.

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

Scholz, H. 1982. Über Mikro- und Makrohaare einiger *Piptatherum* und *Stipa*-Arten (Stipeae, Gramineae). Willdenowia **12:** 235–240.

Schouten, Y. & J. F. Veldkamp. 1985. A revision of *Anthoxanthum* including *Hierochloë* (Gramineae) in Malesia and Thailand. Blumea **30:** 319–351.

Soderstrom, T. R. 1981. Some evolutionary trends in the Bambusoideae (Poaceae). Ann. Missouri Bot. Gard. **68:** 15–47.

——— **& R. P. Ellis.** 1987. The position of bamboo genera and allies in a system of grass classification. Pages 225–238 *in* T. R. Soderstrom, K. W. Hilu, C. S. Campbell, & M. E. Barkworth (eds.), Grass systematics and evolution. Smithsonian Institution Press, Washington, D.C.

Soreng, R. J., J. I. Davis & J. J. Doyle. 1990. A phylogenetic analysis of chloroplast DNA restriction site variation in Poaceae subfam. Pooideae. Pl. Syst. Evol. 172: 83–97.

Stebbins, G. L. 1950. Variation and evolution in plants. Columbia University Press, New York.

——— **& B. Crampton.** 1961. A suggested revision of the grass genera of temperate North America. Recent advances in botany, **1:** 133–145. University of Toronto Press.

Stevens, P. F. 1991. Character states, morphological variation, and phylogenetic analysis: a review. Syst. Bot. **16:** 553–583.

Swofford, D. L. 1984. PAUP 2.2. Phylogenetic analysis using parsimony. Illinois Natural History Survey, Champaign, Illinois.

———. 1985. PAUP 2.4. Illinois Natural History Survey, Champaign, Illinois.

———. 1989. PAUP 3.0s. Illinois Natural History Survey, Champaign, Illinois.
———. 1992. PAUP 3.0 ß for UNIX. Illinois Natural History Survey, Champaign, Illinois.
———. 1993. PAUP: Phylogenetic analysis using parsimony, Version 3.1. Illinois Natural History Survey, Champaign, Illinois.
——— **& G. J. Olsen.** 1990. Phylogeny reconstruction. Pages 411–501 *in* D. M. Hillis & C. Moritz (eds.), Molecular systematics. Sinauer Associates, Sunderland, Massachusetts.
Tateoka, R. 1962. Starch grains of endosperm in grass systematics. Bot. Mag. Tokyo **75:** 377–383.
Veldkamp, J. F., R. deKoning & M. S. M. Sosef. 1986. Generic delimitation of *Rottboellia* and related genera (Gramineae). Blumea **31:** 281–307.
Wagner, W. H., Jr. 1983. Reticulistics: The recognition of hybrids and their role in cladistics and classification. Pages 63–79 *in* N. I. Platnick and V. A. Funk (eds.), Advances in Cladistics: Proceedings of the Second Meeting of the Willi Hennig Society. Columbia University Press, New York.
Wang, Z. Y., G. & S. D. Tanksley. 1992. Polymorphism and phylogenetic relationships among species in the genus *Oryza* as determined by analysis of nuclear RFLPs. Theor. Appl. Genet. **83:** 565–581.
Watson, L. 1971. Basic taxonomic data: the need for organisation over presentation and accumulation. Taxon **20:** 131–136.
———, **H. T. Clifford & M. J. Dallwitz.** 1985. The classification of Poaceae: subfamilies and supertribes. Austral. J. Bot. 33: 433–484.
——— **& M. J. Dallwitz.** 1988. Grass genera of the world: illustrations of characters, descriptions, classification, interactive identification, information retrieval. Australian National University, Canberra.
——— & ———. 1991. Grass genera of the world: an INTKEY package for automated identification and information retrieval of data including synonyms, morphology, anatomy, physiology, cytology, classification, pathogens, world and local distribution, and references. 2nd edition. Flora Online **22.**
Zhang, Guang-zhu. 1987. Studies on the chromosome numbers of some bamboo species with clump rhizomes. Pages 175–178 *in* A. N. Rao, G. Dhanarajan, & C. B. Sastry (eds.), Recent research on bamboos. The Chinese Academy of Forestry, People's Republic of China and International Development Research Centre, Canada.

Appendix A1

TAXA INCLUDED–BAMBUSOID CLADE

Genera with more than 1/3 missing data or with only polyploid taxa known were excluded from cladistic analyses. "Diploid" indicates those taxa in which a diploid count has been recorded. "Polyploid" indicates those taxa in which counts have been made, but none is diploid. Note that many taxa are of unkown ploidy and are not marked in either column. Some chromosome numbers taken from Zhang (1987). * = genera of woody bamboos, not included in all analyses. See text for discussion.

	Species one or two	Polypoid	Diploid	Missing data
Acidosasa*	x			x
Actinocladum*	x			
Alvimia*				
Anomochloa	x			x
Apoclada*				
Arberella			x	
Arthrostylidium*	x			
Arundinaria*		x		

Appendix A1
Continued

	Species one or two	Polypoid	Diploid	Missing data
Arthroostachys*				x
Atractantha*	x			
Aulonemia*				
Bambusa*		x		
Brachyelytrum	x		x	
Bromuniola	x			
Buergersiochloa	x			
Calderonella	x			
Centotheca			x	
Cephalostachyum*		x		
Chasmanthium			x	
Chevalierella	x			
Chikusiochloa			x	x
Chimonobambusa*		x		
Chusquea*			x	
Colanthelia*				
Cryptochloa			x	
Danthonidium				
Decaryochloa*	x			x
Dendrocalamus*		x		
Dendrochloa*	x			
Diandrolyra	x			
Diarrhena	x		x	
Dichaetaria				
Dinochloa*				
Ehrharta			x	
Elytrostachys*	x			
Eriachne				
Euthryptochloa	x			x
Fargesia*	x			x
Gigantochloa*		x		
Glaziophyton*	x			x
Gouldochloa	x		x	
Greslania*				
Guaduella				
Hickelia*	x			x
Hitchcockella*	x			x
Humbertochloa	x			
Hydrochloa	x			x
Hygroryza	x		x	
Indocalamus*		x		
Indosasa*				x
Leersia			x	
Leptaspis			x	
Lithachne			x	
Lophatherum	x	x		
Luziola				
Maclurolyra	x		x	
Maltebrunia				x
Megastachya	x	x		
Melocalamus*	x			

Appendix A1
Continued

	Species one or two	Polypoid	Diploid	Missing data
Melocanna*		x		
Merostachys*				
Metasasa*	x x			
Microlaena				
Mniochloa	x			x
Myriocladus*				
Nastus*				
Neohouzeana*				x
Neurolepis*		x		
Ochlandra*		x		
Olmeca*	x			
Olyra			x	
Oreobambos*	x			
Orthoclada	x		x	
Oryza			x	
Otatea*	x			x
Oxytenanthera*	x	x		
Pariana			x	
Perrierbambus*	x			x
Petriella	x			
Phaenosperma	x		x	
Pharus			x	
Pheidochloa				
Phyllorhachis	x		x	
Phyllostachys*			x	
Piresia			x	x
Pohlidium	x		x	
Porteresia	x	x		
Potamophila	x			
Prosphytochloa	x		x	
Pseudocoix*	x			
Pseudosasa*		x		
Pseudostachyum*	x			
Puelia			x	
Racemobambos*				
Raddia			x	
Raddiella				x
Rehia	x			x
Reitzia	x			x
Rhipidocladum*				x
Rhynchoryza	x			x
Sasa*		x		
Schizostachyum*				
Scrotochloa	x			x
Semiarundinaria*		x		
Shibataea*		x		
Sinobambusa*		x		
Steyermarkochloa	x			
Streptochaeta	x		x	
Streptogyna	x		x	
Sucrea			x	

Appendix A1
Continued

	Species one or two	Polypoid	Diploid	Missing data
Suddia	x			
Swallenochloa*				x
Teinostachyum*				
Tetrarrhena				
Thamnocalamus*		x		
Thyrsostachys*	x			
Yushania*	x			x
Zeugites				
Zizania			x	
Zizaniopsis			x	x

Appendix A2

TAXA INCLUDED– ANDROPOGONODAE PLUS ARUNDINELLEAE AND NEURACHNEAE

Genera with more than 1/3 missing data or with only polyploid taxa known were excluded from cladistic analyses. "Diploid" indicates those taxa in which a diploid count has been recorded. "Polyploid" indicates those taxa in which counts have been made, but none is diploid. Note that many taxa are of unkown ploidy and are not marked in either column.

	Mono- or ditypic	Polyploid	Diploid	Missing data
Agenium				
Anadelphia	x			x
Andropogon			x	
Andropterum	x			
Apluda	x		x	
Apocopis				
Arthraxon			x	
Arundinella				
Asthenochloa	x			x
Bhidea	x			x
Bossia				
Bothriochloa			x	
Capillipedium			x	
Chandrasekharania	x			x
Chasmopodium	x			x
Chrysopogon				
Chumsriella	x			x
Cleistachne	x			
Coelorhachis			x	
Cymbopogon				
Danthoniopsis			x	
Diandrostachya				
Dichanthium			x	
Digastrium	x			
Diheteropogon			x	
Dilophotriche				

Appendix A2
Continued

	Mono- or ditypic	Polyploid	Diploid	Missing data
Dimeria				
Dybowskia	x			x
Eccoilopus				x
Elionurus			x	
Elymandra				
Eremochloa			x	
Eremopogon		x		
Erianthus				
Eriochrysis				
Euclasta	x			
Eulalia				
Eulaliopsis	x			x
Exotheca	x			
Garnotia				
Germainia				
Gilgiochloa	x			
Glyphochloa				x
Hackelochloa	x		x	
Hemarthria			x	
Hemisorghum	x			x
Heteropholis				
Heteropogon			x	
Homozeugos				
Hyparrhenia			x	
Hyperthelia				
Hypogynium	x			
Imperata				
Isalus				x
Ischaemum				
Ischnochloa	x			x
Iseilema			x	
Jansenella	x			x
Jardinea				x
Kerriochloa	x			x
Lasiorrachis				x
Lasiurus			x	
Lepargochloa	x			x
Leptosaccharum	x			
Lophopogon	x			x
Loudetia				
Loudetiopsis				
Loxodera				
Manisuris	x		x	x
Microstegium				
Miscanthidium			x	
Miscanthus				
Mnesithea			x	
Monium				x
Monocymbium			x	
Narenga	x	x		
Neurachne			x	

Appendix A2
Continued

	Mono- or ditypic	Polypoid	Diploid	Missing data
Ophiuros				
Oxyrhachis	x			
Parahyparrhenia				x
Paraneurachne	x	x		
Phacelurus			x	
Pleiadelphia	x			x
Pogonachne	x			x
Pogonatherum				
Polliniopsis	x			x
Polytoca	x			
Polytrias	x			
Pseudanthistiria				
Pseudodichanthium	x			x
Pseudopogonatherum	x			
Pseudosorghum	x			x
Pseudovossia	x			x
Ratzeburgia	x			x
Rhytachne				
Robynsiochloa	x			x
Rottboellia			x	
Saccharum		x		
Schizachyrium				
Sclerachne	x			x
Sclerostachya	x			x
Sehima				
Sorghum				
Sorghastrum			x	
Spathia	x			
Spodiopogon				
Thaumastochloa				
Thelepogon	x			
Themeda				
Thyridolepis			x	
Thyrsia				x
Trachypogon				
Trichopteryx			x	
Trilobachne	x			x
Triplopogon	x			
Tristachya				
Urelytrum				
Vetiveria				
Vossia	x			
Ystia	x			x
Zea				
Zonotriche				

Appendix A3

TAXA INCLUDED–POOID CLADE

Genera with more than 1/3 missing data or with only polyploid taxa known were excluded from cladistic analyses. "Diploid" indicates those taxa in which a diploid count has been recorded. "Polyploid" indicates those taxa in which counts have been made, but none is diploid. Note that many taxa are of unkown ploidy and are not marked in either column.

Genus	Mono- or ditypic	Polyploid	Diploid	Missing data
Aciachne				
Aegilops			x	
Agropyron			x	
Agropyropsis	x			x
Agrostis			x	
Aira			x	
Airopsis	x		x	
Alopecurus			x	
Ambylopyrum	x		x	
Ammochloa				
Ammmophila	x		x	
Ampelodesmos	x	x		
Amphibromus				
Ancistragrostis	x			
Anemanthele	x			
Aniselytron	x	x		
Anisopogon	x			
Anthochloa	x			
Anthoxanthum			x	
Antinoria	x			x
Apera			x	
Aphanelytrum	x			
Arctagrostis		x		
Arctophila	x	x		
Arrhenatherum			x	
Australopyrum	x		x	
Austrofestuca	x		x	
Avellinia	x			
Avena			x	
Beckmannia	x		x	
Bellardiochloa	x		x	
Boissiera	x		x	
Brachyelytrum	x		x	
Brachypodium			x	
Briza			x	
Bromus			x	
Brylkinia	x	x		
Calamagrostis		x		
Calosteca	x	x		
Castellia	x			
Catabrosa	x		x	
Catabrosella				
Catapodium	x		x	
Chaetopogon	x		x	
Cinna		x		
Cockaynea	x	x		

Appendix A3
Continued

Genus	Mono- or ditypic	Polyploid	Diploid	Missing data
Coleanthus	x		x	
Colpodium				
Cornucopiae	x			
Corynephorus			x	
Crithopsis	x		x	x
Ctenopsis			x	
Cutandia			x	
Cyathopus	x			x
Cynosurus			x	
Dactylis	x		x	
Danthoniastrum	x			x
Dasypoa	x			
Dasypyrum			x	
Deschampsia			x	
Desmazeria			x	
Deyeuxia				
Diarrhena				
Dichelachne				
Dielsiochloa	x			x
Dissanthelium				
Dryopoa	x			
Dupontia	x	x		
Duthiea			x	x
Echinaria	x		x	
Echinopoa				
Eremopoa				
Eremopyrum	x			
Erianthecium	x			
Euthryptochloa	x			x
Festuca			x	
Festucella	x			
Festucopsis	x		x	x
Gastridium	x		x	
Gaudinia			x	
Gaudiniopsis	x			
Glyceria			x	
Gymnachne	x	x		
Hainardia	x		x	
Helictotrichon			x	
Helleria	x			
Henrardia	x		x	
Heteranthelium	x		x	
Hierochloe			x	
Holcus			x	
Hookerochloa	x			
Hordelymus	x			
Hordeum			x	
Hyalopoa		x		
Hypseochloa	x			
Koeleria			x	
Lagurus	x		x	

Appendix A3
Continued

Genus	Mono- or ditypic	Polyploid	Diploid	Missing data
Lamarckia	x		x	
Leptagrostis	x			x
Leucopoa		x		
Libyella	x			
Limnas	x			
Limnodea	x		x	
Lindbergella	x			x
Littledalea				
Loliolum	x			x
Lolium			x	
Lombardochloa	x			x
Lophopyrum			x	
Lorenzochloa	x		x	
Lycochloa	x	x		
Lygeum	x	x		
Maillea	x			
Megalachne	x			
Metcalfia	x			
Melica			x	
Mibora	x		x	
Microbriza	x			x
Micropyropsis	x			x
Micropyrum			x	
Milium			x	
Narduroides	x		x	
Nardus	x		x	
Nassella				
Nephelochloa	x			
Oreochloa			x	
Ortachne	x	x		
Oryzopsis			x	
Parafestuca	x	x		
Parapholis			x	
Pascopyrum	x	x		
Pentapogon	x			
Peyritschia	x	x		
Periballia			x	x
Phaenosperma	x		x	
Phalaris			x	
Phippsia		x		
Phleum			x	
Pholiurus	x		x	
Pilgerochloa	x			
Piptatherum				
Piptochaetium	x			
Pleuropogon		x		
Poa				
Podophorus	x			
Poidium	x			x
Polypogon			x	
Psammochloa	x			

Appendix A3
Continued

Genus	Mono- or ditypic	Polyploid	Diploid	Missing data
Psathyrostachys			x	
Pseudarrenatherum	x		x	x
Pseudobromus		x		
Pseudophleum	x			x
Pseudoroegneria	x			
Psilathera	x			x
Psilurus	x	x		
Puccinellia			x	
Rhizocephalus	x			
Rhomboelytrum	x			x
Schizachne	x		x	
Sclerochloa	x		x	
Scolochloa	x	x		
Scribneria	x			
Secale			x	
Sesleria			x	
Sesleriella	x			
Simplicia	x			
Sinochasea	x			x
Sphenopholis			x	
Sphenopus	x		x	
Stephanachne	x		x	
Stilphophleum	x			
Stipa				
Streblochaete	x			
Taeniatherum	x		x	
Thinopyrum			x	
Torreyochloa				
Tovarochloa	x			
Trikeraia	x			
Triniochloa				
Triplachne	x			
Trisetum	x			
Triticum		x		
Tsvelevia	x			
Vahlodea			x	
Ventenata			x	
Vulpia			x	
Vulpiella	x			
Wangenheimia	x			
Zingeria			x	

Appendix B1

ANNOTATED CHARACTER LIST—BAMBUSOID CLADE

Numbers in parentheses refer to consistency index and retention index, respectively, in analyses of full data set (Fig. 3). u = uninformative for final set of included taxa. Inapplicable characters scored as missing. PAUP records CI even for excluded characters, but these are not included in length calculations. Many characters are illustrated in Watson and Dallwitz (1988). Unless otherwise noted, "lemmas" refers to lemmas of florets with fertile pistils (= "female-fertile lemmas" of Watson and Dallwitz (1988; 1991)). Notes on synapomorphies are based on the distribution of characters on tree shown in Fig. 3.

1. Habit. A–long-rhizomatous, or long-stoloniferous; B–Caespitose or decumbent. (0.25; 0.57)
2. Flowering culms. A–leafless; B–leafy. (0.5; 0.0)
3. Height of mature plants. A–tall plants, to 3 m or more; B–never reaching 3 m in height. (0.25; 0.5) Although this character shows little homoplasy in this analysis, its inclusion in future work may not be justified because of its fundamentally continuous nature.
4. Culms. A–woody and persistent; B–herbaceous. (0.2; 0.6)
5. Culms. A–scandent; B–not scandent. (0.5; 0.5)
6. Culms. A–branching above; B–unbranched above. (0.2; 0.6)
7. Primary branches per mid-culm node. A–1; B–2 or more. (u)
8. Culm nodes. A–hairy; B–glabrous. (u)
9. Culm sheaths; A–persistent; B–deciduous in their entirety. (u)
10. Culm internodes. A–solid; B–hollow. (0.5; 0.83)
11. Bambusoid habit. A–unicaespitose; B–pluricaespitose. (1.0; 1.0) A synapomorphy for *Arthrostylidium*/*Colanthelia*/*Merostachys*. This character will be most useful for an ingroup analysis of the woody bamboos.
12. Rhizomes. A–pachymorph; B–leptomorph. (0.5; 0.0) This character is not really relevant at this level; it will be more useful for ingroup analysis of the woody bamboos.
13. Young shoots; A–extravaginal; B–intravaginal. (0.33; 0.33)
14. Leaves. A–mostly basal; B–not basally aggregated. (0.5; 0.0)
15. Auricles. A–present; B–absent. (0.2; 0.43)
16. Auricular setae. A–present; B–absent. (0.25; 0.5)
17. Leaf blades. A–broad; B–narrow. (0.33; 0.75)
18. Leaf blades. A–cordate; B–sagittate; C–not cordate, not sagittate. (0.67; 0.0)
19. Leaf blades. A–flat or folded; B–rolled. (u)
20. Leaf blades. A–pseudopetiolate; B–not pseudopetiolate. (0.2; 0.69)
21. Leaf venation. A–pinnate; B–palmate; C–neither pinnate nor palmate. (1.0; 1.0) Pinnate leaf venation is a synapomorphy for the Phareae.
22. Transverse veins in leaf blades. A–readily visible; B–not readily visible. (0.17; 0.67) Excluded because of low CI.
23. Leaf blades. A–disarticulating from the sheaths; B–not disarticulating. (0.25; 0.7)
24. Adaxial ligule. A–present; B–absent, at least from upper leaves. (u)
25. Adaxial ligule. A–an unfringed membrane; B–a fringed membrane or a fringe of hairs. (0.33; 0.75)
26. Adaxial ligule. A–truncate; B–not truncate. (0.25; 0.0) An arbitrary division of a continuum; the low retention index is not surprising.
27. Abaxial ligule. A–present; B–absent. (1.0; 1.0) An artefact of sampling; present in *Streptogyna* and *Dichaetaria* so scored as present in all outgroups, the "*Guaduella* group" and the woody bamboos.
28. Plants. A–monoecious with all fertile spikelets unisexual; B–bisexual, with bisexual spikelets; C–dioecious. (0.5; 0.93)
29. Hermaphrodite florets. A–present; B–absent. (0.33; 0.87)
30. Spikelets. A–of at least 2 sexually distinct forms on the same plant; B–all alike in sexuality. (0.33; 0.88)
31. Male and female-fertile spikelets. A–in different inflorescences; B–on different branches of same inflorescence, or segregated in different parts of same inflorescence branch; C–mixed in inflorescence. (0.33; 0.33)
32. Inflorescence. A–determinate; B–indeterminate. (0.33; 0.5)
33. Pseudospikelets. A–present; B–absent. (0.5; 0.67) Some authors have suggested that pseudospikelets are primitive (e.g. Soderstrom, 1981), but these analyses support the contention of Clayton and Renvoize (1986) that they are in fact derived.

34. Inflorescence. A–of spike-like main branches; B–a false spike, with clusters of spikelets on reduced axes; C–a single raceme. (0.5; 0.0)
35. Inflorescence. A–open; B–contracted. (0.25; 0.0)
36. Capillary inflorescence branchlets. A–present; B–absent. (0.33; 0.0)
37. Inflorescence axes. A–ending in spikelets; B–not ending in spikelets. (0.33; 0.0)
38. Rachides. A–hollowed, or flattened, or winged; B–neither flattened nor hollowed, not winged. (0.5; 0.5)
39. Inflorescence. A–spatheate; B–espatheate. (0.17; 0.54) Excluded because of low CI.
40. Inflorescence. A–a complex of partial inflorescences and intervening foliar organs; B–not comprising partial inflorescence and foliar organs. (0.25; 0.57)
41. Spikelet-bearing axes. A–very much reduced; B–spikes; C–racemes, or spike-like, or paniculate; D–capitate. (0.75; 0.0)
42. Spikelet-bearing axes. A–solitary; B–paired; C–clustered. (u)
43. Spikelet-bearing axes. A–disarticulating; B–persistent. (0.33; 0.33)
44. Spikelets. A–associated with bractiform involucres; B–unaccompanied by bractiform involucres, not associated with setiform vestigial branches. (0.5; 0.0)
45. Spikelets. A–solitary; B–in pairs or in triplets. (1.0; 1.0) A synapomorphy for the clade including the Phareae plus *Pariana* and *Maclurolyra*.
46. Spikelets. A–secund; B–not secund. (0.33; 0.0)
47. Spikelets; A–all sessile; B–subsessile; C–pedicellate. (0.5; 0.0)
48. Long-and-short combinations of spikelets. A–present; B–absent. (0.5; 0.75)
49. Spikelets. A–in pedicellate/sessile combinations; B–unequally pedicellate in each combination. (u)
50. Shorter spikelets. A–hermaphrodite; B–female only. (u)
51. Longer spikelets. A–hermaphrodite; B–male only. (u)
52. Disarticulation. A–above glumes; B–below glumes. (0.12; 0.12) Excluded because of low CI.
53. Spikelets. A–not disarticulating between florets; B–disarticulating between florets. (1.0; 1.0) A synapomorphy for the "*Guadella*" clade plus the woody bamboos.
54. Spikelets. A–with conventional internode spacings; B–with a distinctly elongated rachilla internode above glumes; C–with distinctly elongated rachilla internodes between florets. (0.5; 0.5)
55. Rachilla. A–prolonged apically; B–not apically prolonged. (0.2; 0.6) Excluded because of low CI.
56. Rachilla. A–hairy; B–hairless. (u)
57. Callus. A–absent; B–short; C–long. (0.5; 0.5)
58. Hairy callus. A–present; B–absent. (0.33; 0.6)
59. Glumes. A–present; B–absent. (1.0; 1.0) A synapomorphy for the Oryzeae s.l.
60. Glumes. A–minute; B–relatively large. (0.5; 0.67)
61. Glumes. A–very unequal; B–more or less equal. (0.17; 0.5) Excluded because of low CI.
62. Glumes. A–markedly < spikelets; B–= or > spikelets. (0.5; 0.0)
63. Glumes. A–decidedly adjacent lemmas; B–> adjacent lemmas. (0.25; 0.73)
64. Glumes. A–joined; B–free. (1.0; 1.0) This may be a synapomorphy for the Oryzeae plus *Zizania*; glumes are fused where present, but are lacking in most species in the clade.
65. Glumes. A–hairy; B–hairless. (0.5; 0.0)
66. Glumes. A–pointed; B–not pointed. (0.33; 0.71)
67. Glumes. A–awned; B–awnless. (0.25; 0.0)
68. Glumes. A–carinate; B–not carinate. (0.25; 0.0)
69. Glumes. A–very dissimilar; B–similar. (1.0; 1.0) A synapomorphy for the Phyllorhachidae.
70. Lower glume. A–< 1/2 length of lowest lemma; B–> 1/2 length of lowest lemma. (u)
71. Incomplete florets. A–absent; B–present. (0.12; 0.56) Excluded because of low CI.
72. Incomplete florets. A–proximal to female-fertile florets; B–distal to or both distal and proximal to female-fertile florets. (0.5; 0.88)
73. Distal florets. A–merely underdeveloped; B–clearly specialised and modified in form. (u)
74. Proximal incomplete florets. A–present; B–absent. (0.13; 0.46) Excluded because of low CI.
75. Proximal incomplete florets. A–up to 2; B–2 or more. (0.5; 0.75) Excluded because dependent on #74, which is excluded.
76. Proximal incomplete florets. A–paleate; B–epaleate. (1.0; 1.0) Excluded because dependent on #74, which is excluded.
77. Proximal incomplete florets. A–male; B–sterile. (1.0; 1.0) Excluded because dependent on #74, which is excluded.
78. Proximal lemmas. A–awned; B–awnless. (0.5; 0.0) Excluded because dependent on #74, which is excluded.

79. Proximal lemmas. A–up to 13-nerved; B–26-nerved or more. (u) Excluded because dependent on #74, which is excluded.
80. Proximal lemmas. A–< female-fertile lemmas; B–= or > female-fertile lemmas. (0.33; 0.5) Excluded because dependent on #74, which is excluded.
81. Proximal lemmas. A–less firm than female-fertile lemmas; B–similar in texture to female-fertile lemmas; C–decidedly firmer than female-fertile lemmas. (0.67; 0.0) Excluded because dependent on #74, which is excluded.
82. Proximal lemmas. A–becoming indurated; B–not becoming indurated. (u) Excluded because dependent on #74, which is excluded.
83. Female-fertile florets. A–1; B–2 or more. (0.2; 0.33)
84. Lemmas. A–convolute; B–not convolute. (u)
85. Lemmas. A–less firm or similar in texture to glumes; B–decidedly firmer than glumes. (0.67; 0.75)
86. Lemmas. A–becoming indurated; B–not becoming indurated. (0.33; 0.82)
87. Lemmas. A–entire; B–incised. (0.5; 0.0)
88. Lemmas. A–pointed; B–blunt. (0.20; 0.43)
89. Lemmas. A–awnless to mucronate; B–awned. (0.2; 0.33)
90. Awns. A–hairless; B–hairy; C–long-plumose. (u)
91. Lemmas. A–hairy; B–hairless. (0.09; 0.23) Excluded because of low CI.
92. Lemmas. A–glabrous; B–scabrous. (u)
93. Lemmas. A–carinate; B–not carinate. (0.33; 0.71)
94. Germination flap. A–present; B–absent. (1.0; 1.0) A synapomorphy for *Arberella*/*Lithachne*/ *Olyra*/*Raddia*.
95. Lemma veins. A–confluent towards tip; B–non-confluent apically. (u)
96. Palea. A–relatively long; B–conspicuous but relatively short to very reduced. (0.5; 0.0)
97. Palea. A–convolute; B–not convolute. (u)
98. Palea. A–entire; B–apically notched to deeply bifid. (0.25; 0.4)
99. Palea. A–thinner than lemma; B–similar in texture to lemma. (1.0; 1.0) A synapomorphy for *Ehrharta* plus *Petriella*.
100. Palea. A–indurated; B–not indurated. (0.33; 0.71)
101. Palea. A–1-nerved; B–2-nerved; C–with several nerves. (0.25; 0.63)
102. Palea. A–one-keeled; B–2-keeled; C–keel-less. (0.5; 0.87)
103. Palea keels. A–winged; B–wingless. (u)
104. Lodicules. A–2; B–3. (0.5; 0.92) See text for discussion.
105. Lodicules. A–joined; B–free. (u)
106. Lodicules. A–fleshy; B–membranous. (0.25; 0.4)
107. Lodicules. A–ciliate; B–glabrous. (1.0; 1.0) A synapomorphy for the clade including the woody bamboos, the "*Guadella* group" and the Ehrharteae.
108. Lodicules. A–toothed; B–not toothed. (0.50; 0.83)
109. Lodicules. A–heavily vascularized; B–not or scarcely vascularized. (0.5; 0.0) Note that the circumscription of this character is different from that used by Kellogg and Campbell (1987).
110. Stamen number. A–2 or 3; B–4 to 6. (0.43; 0.64) This is equivalent to 1 whorl vs. 2 whorls of stamens, each state then including reductions within that whorl. Numbers larger than 6 are interpreted as proliferations of stamens with an ancestral 6-stamened plant. The lower numbers (1 or 2 and 4 or 5) occur only in scattered genera or species and appear to be parallel losses.
111. Stamens. A–with free filaments; B–monodelphous or diadelphous. (1.0; 1.0) A synapomorphy for *Puelia* plus *Streptochaeta*.
112. Anthers. A–penicillate; B–not penicillate. (u)
113. Anther connective. A–apically prolonged; B–not apically prolonged. (0.33; 0.33)
114. Ovary. A–glabrous; B–hairy. (0.33; 0.0)
115. Apical appendage of ovary. A–present, conspicuous; B–absent. (0.25; 0.57)
116. Ovary appendage. A–long, stiff and tapering; B–broadly conical, fleshy. (u)
117. Styles. A–fused; B–free to their bases. (0.33; 0.83)
118. Stigmas. A–1; B–2; C–3. (0.5; 0.5)
119. Stigmas. A–white or brown; B–red. (u)
120. Fruit. A–small; B–medium sized; C–large. (0.4; 0.4) Although this character shows low homoplasy in this analysis, its inclusion in future work may not be justified because of its fundamentally continuous nature and apparently arbitrary division into states.
121. Fruit. A–longitudinally grooved; B–not grooved. (0.25; 0.25)
122. Fruit. A–compressed laterally; B–compressed dorsiventrally; C–not noticeably compressed; D- trigonous. (0.25; 0.4)

123. Hilum. A–short. B–long-linear. (u) See text for discussion.
124. Pericarp. A–thin; B–thick and hard; C–fleshy. (u)
125. Pericarp. A–loose; B–fused. (0.25; 0.25)
126. Embryo. A–large (1/3 length of caryopsis); B–small. (u)
127. Seed. A–endospermic; B–not endospermic. (1.0; 1.0) A synapomorphy for *Dinochloa* and *Melocanna*. This character may be more useful for an ingroup analysis of the Bambuseae.
128. Endosperm. A–liquid in mature fruit; B–hard. (u)
129. Endosperm starch grains. A–simple; B–compound. (u) Tateoka (1962) has questioned the value of this character in systematic studies.
130. Scutellar tail. A–present; B–absent. (0.5; 0.5) This character and the following two were shown by Reeder (1957) to be highly significant systematically, a conclusion confirmed by all our analyses to date. See text for discussion.
131. Embryonic mesocotyl internode. A–elongated; B–short. (0.5; 0.5) See text for discussion.
132. Embryonic leaf margins. A–meeting; B–overlapping. (0.5; 0.0) See text for discussion.
133. Seedling mesocotyl. A–short; B–long. (0.5; 0.5)
134. Lamina of first seedling leaf. A–well-developed; B–absent. (0.5; 0.8)
135. Microhairs on abaxial leaf surface. A–present; B–absent. (0.33; 0.33) See text for discussion.
136. Microhairs. A–ostensibly one-celled; B–clearly two-celled. (1.0; 1.0) A possible synapomorphy for the Oryzeae, but relatively poorly sampled (7 of 42 taxa).
137. Intercostal zones. A–of typical long-cells; B–with many atypical long-cells. (0.33; 0.0)
138. Long-cell shape. A–similar costally and intercostally; B–markedly different costally and intercostally. (0.25; 0.67)
139. Long-cell wall thickness. A–similar costally and intercostally; B–differing markedly costally and intercostally. (0.33; 0.0)
140. Papillae. A–present; B–absent. (0.14; 0.54) Excluded because of low CI.
141. Papillae. A–present on subsidiaries; B–absent from subsidiaries. (0.5; 0.5) Excluded because dependent on #140, which is excluded.
142. Intercostal papillae. A–over-arching stomata; B–not over-arching stomata. (0.33; 0.33)
143. "Pooid-type" silica bodies. A–present; B–absent. (u)
144. "Panicoid-type" silica bodies. A–present; B–absent. (0.14; 0.65) Excluded because of low CI.
145. Tall-and-narrow silica bodies. A–present; B–absent. (0.33; 0.33)
146. Saddle-shaped silica bodies. A–present; B–absent. (0.12; 0.56) Excluded because of low CI.
147. Crescentic silica bodies. A–present; B–absent. (0.5; 0.0)
148. Oryzoid silica bodies. A–present; B–absent. (0.33; 0.83)
149. Round to oval silica bodies. A–present; B–absent. (u)
150. Elongated, smooth silica bodies. A–present; B–absent. (u)
151. Guard-cells. A–overlapped by interstomatals; B–overlapping to flush with interstomatals. (0.25; 0.4)
152. Subsidiaries. A–triangular; B–not triangular. (0.25; 0.67) Character state A includes dome-shaped subsidiaries, which so often co-occur with triangular ones that the two states cannot be reliably separated.
153. Subsidiaries. A–parallel-sided; B–not parallel-sided. (0.33; 0.33)
154. Subsidiaries. A–triangular and parallel-sided on same leaf; B–not parallel-sided and triangular on same leaf. (u)
155. Intercostal short-cells. A–common; B–absent. (0.25; 0.5)
156. Intercostal short-cells. A–in cork/silica-cell pairs; B–not paired. (0.2; 0.0)
157. Intercostal short-cells. A–silicified; B–not silicified. (u)
158. Costal short-cells. A–conspicuously in long rows; B–predominantly paired; C–not distinctly grouped into long rows nor predominantly paired. (0.4; 0.57)
159. Adaxial palisade in mesophyll. A–present; B–absent. (0.25; 0.5)
160. Columns of colourless cells in mesophyll. A–present; B–absent. (u)
161. Arm cells. A–present; B–absent. (0.25; 0.63) See text for discussion.
162. Fusoid cells. A–present; B–absent. (0.2; 0.67) See text for discussion.
163. Fusoid cells. A–an integral part of PBS; B–external to PBS. (u) A synapomorphy for the Centotheceae, subsequently excluded from the bambusoid clade.
164. Leaf blade. A–with distinct, prominent adaxial ribs; B–"nodular" in section; C–adaxially flat. (0.22; 0.12)
165. Midrib. A–conspicuous; B–not readily distinguishable from other main veins. (0.2; 0.2)
166. Midrib. A–with one bundle only; B–with an arc of bundles; C–with complex vascularization. (0.25; 0.63)

167. Midrib, adaxial colourless tissue. A–extensive; B–lacking. (0.17; 0.37)
168. Adaxial bulliform cells. A–in simple fan-shaped groups; B–not in simple fans. (0.5; 0.5)
169. Adaxial bulliform cells. A–combined with colourless cells to form deeply-penetrating fan-shaped groups; B–without deeply-penetrating fans of bulliforms-plus- colourless cells. (u)
170. Combined girders. A–forming "figures"; B–nowhere forming 'figures.' (0.5; 0.0)
171. Sclerenchyma. A–all associated with vascular bundles; B–not all bundle-associated. (u)
172. Chromosome base number, $x =$. A–7; B–9 to 10; C–11 to 12. (u)

Appendix B2

ANNOTATED CHARACTER LIST–ANDROPOGONODAE

Numbers in parentheses refer to consistency index and retention index, respectively, in analyses of full data set (Fig. 7). u = uninformative for final set of included taxa. Inapplicable characters scored as missing. PAUP records CI even for excluded characters, but these are not included in length calculations. Many characters are illustrated in Watson and Dallwitz (1988). Unless otherwise noted, "lemmas" refers to lemmas of florets with fertile pistils (= "female-fertile lemmas" of Watson and Dallwitz (1988; 1991)). Notes on synapomorphies are based on the distribution of characters on tree shown in Fig. 7.

1. Longevity. A–annual; B–perennial. (0.14;0.14)
2. Habit. A–long-rhizomatous, or long-stoloniferous; B–caespitose or decumbent.(0.33; 0.0)
3. Culms. A–woody and persistent; B–herbaceous. (u)
4. Culms. A–branching; B–unbranched. (0.33;0.86)
5. Nodes. A–hairy; B–glabrous. (0.2;0.5)
6. Culm internodes. A–solid; B–hollow. (0.14;0.40)
7. Young shoots. A–extravaginal; B–intravaginal. (u)
8. Shoots. A–aromatic; B–not aromatic. (0.5; 0.0)
9. Leaves. A–mostly basal; B–not basally aggregated. (0.14; 0.40)
10. Auricles. A–present; B–absent. (0.5; 0.0)
11. Auricular setae. A–present; B–absent. (u)
12. Leaf blades. A–broad; B–narrow. (0.17; 0.38)
13. Leaf blades. A–cordate or sagittate; B–not cordate, not sagittate. (0.5; 0.0)
14. Leaf blades. A–setaceous; B–not setaceous. (1.0; 1.0) A synapomorphy for *Leptosaccharum* plus *Oxyrhachis*.
15. Leaf blades. A–pseudopetiolate; B–not pseudopetiolate. (0.25; 0.0)
16. Transverse veins on leaf blades. A–readily visible; B–not readily visible. (u)
17. Leaf blades. A–disarticulating from sheaths; B–not disarticulating. (0.33; 0.33)
18. Leaf blades. A–rolled in bud; B–once-folded in bud, or folded like a fan in bud. (0.33; 0.0)
19. Adaxial ligule. A–an unfringed membrane; B–a fringed membrane; C–a fringe of hairs; D- a rim of minute papillae. (0.15; 0.61)
20. Adaxial ligule. A–truncate; B–not truncate. (0.33; 0.33)
21. Adaxial ligule. A–up to 2 mm long; B–2 mm long or more. (u)
22. Abaxial ligule. A–present; B–absent. (u)
23. Plants. A–monoecious; B–bisexual, with bisexual spikelets; C–dioecious. (0.33; 0.67)
24. Plants. A–with hermaphrodite florets; B–without hermaphrodite florets. (0.25; 0.57)
25. Spikelets. A–of sexually distinct forms on same plant; B–all alike in sexuality. (0.17; 0.81)
26. Male and female-fertile spikelets. A–in different inflorescences; B–on different branches of same inflorescence; C–segregated, in different parts of same inflorescence branch; D- mixed in inflorescence. (1.0; 1.0) Character exhibits three changes, all in the Maydeae s.l.
27. Spikelets. A–overtly heteromorphic; B–externally homomorphic. (0.14; 0.57)
28. Spikelets. A–in both homogamous and heterogamous combinations; B–all in heterogamous combinations. (0.14; 0.57)
29. Hidden cleistogenes. A–present; B–absent. (u)
30. Inflorescence. A–of spike-like main branches; B–a false spike, with clusters of spikelets on reduced axes; C–a single raceme; D- paniculate. (0.22; 0.63)
31. Inflorescence. A–deciduous in its entirety as a "tumbleweed"; B–not deciduous. (u)

32. Inflorescence. A–open; B–contracted. (0.25; 0.4)
33. Capillary branchlets in inflorescence. A–present; B–absent. (0.17; 0.0)
34. Inflorescence. A–digitate or subdigitate; B–not digitate. (0.17; 0.62)
35. Inflorescence axes. A–ending in spikelets; B–not ending in spikelets. (0.5; 0.0)
36. Rachides. A–hollowed; B–flattened or winged; C–neither flattened nor hollowed, not winged. (0.33; 0.72)
37. Spikelets. A–all partially embedded in rachis; B–not all embedded. (0.14; 0.46)
38. Inflorescence. A–spatheate; B–espatheate. (0.14; 0.79)
39. Inflorescence. A–a complex of "partial inflorescences" and intervening foliar organs; B–not comprising "partial inflorescences" and foliar organs. (0.17; 0.78)
40. Spikelet-bearing axes. A–very much reduced; B–spikes; C–"racemes," or paniculate, or capitate; D-spike-like. (0.38; 0.72)
41. Spikelet-bearing axes. A–solitary; B–clustered. (0.25; 0.67)
42. Rachides of spikelet-bearing axes. A–slender; B–substantial. (0.12; 0.71)
43. Spikelet-bearing axes. A–disarticulating; B–persistent. (0.11; 0.56)
44. Spikelet-bearing axes. A–falling entire; B–disarticulating at joints. (0.33; 0.0)
45. Longitudinal, translucent furrow on pedicels and internodes of rachis. A–present; B–absent. (0.5; 0.5)
46. "Articles." A–linear; B–not linear. (0.17; 0.78)
47. "Articles." A–with a basal callus-knob; B–without a basal callus-knob. (1.0; 1.0) A synapomorphy at the node below *Phacelurus*.
48. "Articles." A–appendaged; B–not appendaged. (0.33; 0.5)
49. "Articles." A–disarticulating transversely; B–disarticulating obliquely. (0.07; 0.32) Excluded because of low CI.
50. "Articles." A–densely long-hairy; B–somewhat hairy; C–glabrous. (0.4; 0.75)
51. Spikelets. A–associated with bractiform involucres; B–unaccompanied by bractiform involucres, not associated with setiform vestigial branches; C–with "involucres" of "bristles"; D- subtended by solitary "bristles." (0.5; 0.5)
52. "Bristles." A–spiny, markedly coalescent basally; B–relatively slender, not spiny. (u)
53. "Bristles." A–persisting on axis; B–deciduous with spikelets. (u)
54. Spikelets. A–solitary; B–in pairs; C–in triplets. (0.29; 0.54)
55. Spikelets. A–secund; B–not secund. (0.08; 0.42)
56. Spikelets. A–biseriate; B–distichous, or not two-ranked. (u)
57. Pedicellate spikelets. A–spikelets all sessile; B–spikelets subsessile; C–having pedicellate spikelets. (0.5; 0.5)
58. Pedicel apices. A–oblique to discoid; B–cupuliform. (u)
59. Spikelets. A–consistently in long-and-short combinations; B–not in distinct long-and- short combinations. (0.33; 0.78)
60. Spikelets. A–in pedicellate/sessile combinations; B–unequally pedicellate in each combination. (0.2; 0.2)
61. Pedicels of pedicellate spikelets. A–discernible, but fused with rachis; B–free of rachis. (1.0; 1.0) A synapomorphy for the Maydeae s. l. plus *Hackelochloa*, *Hemarthria*, *Ophiuros*, *Thaumastochloa*, *Heteropholis*, and *Rottboellia*.
62. Shorter spikelets. A–hermaphrodite; B–female-only; C–male-only, or sterile. (0.5; 0.0)
63. Longer spikelets. A–hermaphrodite; B–female-only; C–male-only, or sterile. (0.22; 0.56)
64. Spikelets. A–abaxial; B–adaxial. (u)
65. Spikelets. A–compressed laterally; B–not noticeably compressed; C–compressed dorsiventrally. (0.22; 0.46)
66. Disarticulation of spikelets. A–above glumes; B–below glumes. (1.0; 1.0) A synapomorphy for the Arundinelleae.
67. Disarticulation of spikelets. A–not between florets; B–between florets. (1.0; 1.0) A synapomorphy for the Arundinelleae plus *Polytrias*.
68. Spikelets. A–with conventional internode spacings; B–with a distinctly elongated rachilla internode between glumes; C–with a distinctly elongated rachilla internode above glumes. (u)
69. *Ichnanthus*-type stipe. A-present; B–absent. (u)
70. Stipe beneath upper floret. A–filiform; B–not filiform. (u)
71. Stipe beneath upper floret. A–straight and swollen; B–curved, not swollen. (u)
72. Apically prolonged rachilla. A–present; B–absent. (u)
73. Rachilla. A–hairy; B–hairless. (u)

74. Callus. A–absent; B–short; C–long. (0.17; 0.17)
75. Callus. A–pointed; B–blunt. (0.17; 0.38)
76. Hairy callus. A–present; B–absent. (0.17; 0.71)
77. Glumes. A–present; B–absent. (u)
78. Glumes. A–one per spikelet; B–two to several. (0.5; 0.0)
79. Glumes. A–minute; B–relatively large. (u)
80. Glumes. A–very unequal; B–more or less equal. (0.12; 0.46)
81. Glumes. A–markedly < spikelets; B–about = spikelets; C–> spikelets. (0.5; 0.0)
82. Glumes. A–decidedly < adjacent lemmas; B–> adjacent lemmas. (u)
83. Glumes. A–dorsiventral to rachis; B–lateral to rachis. (u)
84. Glumes. A–hairy; B–hairless. (0.25; 0.67)
85. Distinct hair tufts on glumes. A–present; B–absent. (u)
86. Glumes. A–pointed; B–not pointed. (0.33; 0.71)
87. Glumes. A–awned; B–awnless. (0.17; 0.17)
88. Glumes. A–carinate; B–not carinate. (u)
89. Keel of glumes. A–conspicuously winged; B–not winged. (u)
90. Glumes. A–very dissimilar; B–similar. (0.17; 0.44)
91. Lower glume. A–<lowest lemma; B–>lowest lemma. (1.0; 1.0) A synapomorphy for the Andropogonodae.
92. Lower glume. A–two-keeled; B–not two-keeled. (0.06; 0.37) Excluded because of low CI.
93. Back of lower glume. A–convex; B–flattened; C–sulcate. (0.18; 0.31)
94. Lower glume. A–conspicuously pitted; B–not pitted. (u)
95. Upper glume. A–distinctly saccate; B–not saccate. (u)
96. Spikelets. A–with hermaphrodite florets only; B–with incomplete florets. (0.5; 0.0)
97. Incomplete florets. A–proximal to female-fertile florets; B–distal to female-fertile florets; C–both distal and proximal to female-fertile florets. (u)
98. Proximal incomplete florets. A–paleate; B–epaleate. (0.25; 0.87)
99. Palea of proximal incomplete florets. A–fully developed; B–reduced. (0.25; 0.4)
100. Palea of proximal incomplete florets. A–becoming conspicuously hardened and enlarged laterally; B–not becoming conspicuously hardened and enlarged laterally. (u)
101. Proximal incomplete florets. A–male; B–sterile. (0.17; 0.76)
102. Proximal lemmas. A–awned; B–awnless. (u)
103. Proximal lemmas. A–<female-fertile lemmas; B–= female-fertile lemmas. (0.25; 0.0)
104. Proximal lemmas. A–less firm than female-fertile lemmas; B–similar in texture to female-fertile lemmas; C–decidedly firmer than female-fertile lemmas. (0.5; 0.5)
105. Proximal lemmas. A–becoming indurated; B–not becoming indurated. (u)
106. Female-fertile florets. A–1; B–2 or more. (u)
107. Lemmas. A–less firm than glumes; B–similar in texture to glumes; C–decidedly firmer than glumes. (0.5; 0.0)
108. Lemmas. A–smooth; B–rugose. (u)
109. Lemmas. A–becoming indurated; B–not becoming indurated. (u)
110. Lemmas. A–entire; B–incised. (0.14; 0.79)
111. Lemmas. A–pointed; B–blunt. (0.5; 0.5)
112. Lemmas. A–deeply cleft; B–not deeply cleft. (0.17; 0.5)
113. Lemmas. A–crested at tip; B–not crested. (u)
114. Lemmas. A–awnless to mucronate; B–awned. (0.2; 0.85)
115. Awns. A–from sinus; B–apical. (0.33; 0.80)
116. Awns. A–non-geniculate; B–geniculate. (u)
117. Awns. A–hairless; B–hairy; C–long-plumose. (0.14; 0.40)
118. Awns. A–deciduous; B–persistent. (0.5; 0.0)
119. Lemmas. A–hairy; B–hairless. (0.14; 0.4)
120. Hairs. A–in tufts; B–not in tufts. (1.0; 1.0) A synapomorphy for *Gilgiochloa* plus *Trichopteryx*.
121. Hairs. A–in transverse rows; B–not in transverse rows. (0.5; 0.0)
122. Lemmas. A–carinate; B–not carinate. (0.5; 0.0)
123. Lemmas. A–having margins lying flat and exposed on palea; B–having margins tucked in onto palea. (1.0; 1.0) A synapomorphy for all Arundinelleae except *Arundinella*.
124. Lemmas. A–with a clear germination flap; B–without a germination flap. (0.33; 0.33)
125. Palea. A–present; B–absent. (0.17; 0.50)
126. Palea. A–relatively long; B–conspicuous but relatively short; C–very reduced. (0.2; 0.58)

127. Palea. A–entire; B–apically notched to deeply bifid. (0.2; 0.6)
128. Palea. A–awnless, without apical setae; B–with apical setae. (u)
129. Palea. A–thinner than lemma; B–similar in texture to lemma. (1.0; 1.0) A synapomorphy for the Arundinelleae except *Arundinella* and *Danthoniopsis*.
130. Palea. A–indurated; B–not indurated. (u)
131. Palea nerves. A–1; B–0. (0.14; 0.78)
132. Palea. A–2-keeled; B–keel-less. (0.5; 0.89)
133. Palea keels. A–winged; B–wingless. (0.5; 0.0)
134. Lodicules. A–present; B–absent. (0.14; 0.40)
135. Lodicules. A–joined; B–free. (u)
136. Lodicules. A–fleshy; B–membranous. (u)
137. Lodicules. A–ciliate; B–glabrous. (0.5; 0.5)
138. Lodicules. A–toothed; B–not toothed. (u)
139. Lodicules. A–heavily vascularized; B–not or scarcely vascularized. (u)
140. Stamens. A–0 to 3; B–6. (u)
141. Ovary. A–glabrous; B–hairy. (u)
142. Styles. A–fused; B–free to their bases. (0.12; 0.36)
143. Stigmas. A–white; B–red pigmented; C–brown. (0.4; 0.0)
144. Fruit. A–longitudinally grooved; B–not grooved. (0.2; 0.5)
145. Fruit. A–compressed laterally; B–compressed dorsiventrally; C–not noticeably compressed. (0.5; 0.5)
146. Hilum. A–short; B–long-linear. (1.0; 1.0) A synapomorphy for the Arundinelleae except for *Arundinella*.
147. Embryo. A–large; B–small. (u)
148. Embryo. A–waisted; B–not waisted. (u)
149. Starch grains in endosperm. A–simple; B–compound. (0.17; 0.44)
150. Epiblast. A–present; B–absent. (u)
151. Scutellar tail. A–present; B–absent. (u)
152. Mesocotyl internode. A–elongated; B–negligible. (1.0; 1.0) A synapomorphy for the Andropogonodae and the Arundinelleae.
153. Embryonic leaf margins. A–meeting; B–overlapping. (u)
154. Seedling mesocotyl. A–short; B–long. (u)
155. First seedling lamina. A–broad; B–narrow. (u)
156. First seedling lamina. A–erect; B–supine. (0.33; 0.0)
157. First seedling lamina. A–up to 12 veined; B–13 to 20 veined; C–21 veined or more. (u)
158. Microhairs. A–present; B–absent. (0.5; 0.5)
159. Microhairs. A–panicoid-type; B–chloridoid-type. (u)
160. Microhairs. A–up to 14.9 microns wide at septum; B–15 microns wide at septum or more. (u)
161. Costal/intercostal zonation. A–conspicuous; B–lacking. (u)
162. Intercostal long-cells. A–typical; B–many atypical; C–absent. (0.5; 0.0)
163. Long-cells, costal vs. intercostal. A–similar in shape; B–markedly different in shape. (0.11; 0.33)
164. Long-cell walls, costal vs. intercostal. A–of similar thickness; B–differing markedly in thickness. (0.25; 0.0)
165. Mid-intercostal long-cells. A–more or less rectangular; B–more or less fusiform. (u)
166. Mid-intercostal long-cell walls. A–markedly sinuous; B–straight or only gently undulating. (0.5; 0.0)
167. Papillae. A–present; B–absent. (0.11; 0.70)
168. Papillae on subsidiaries. A–present; B–absent. (0.33; 0.33)
169. Intercostal papillae. A–over-arching stomata; B–not over-arching stomata. (0.25; 0.57)
170. Intercostal papillae. A–one per cell; B–several per cell. (0.2; 0.64)
171. "Pooid-type" silica bodies. A–present; B–absent. (0.5; 0.0)
172. "Panicoid-type" silica bodies. A–present; B–absent. (0.33; 0.33)
173. Tall-and-narrow silica bodies. A–present; B–absent. (0.5; 0.5)
174. Saddle-shaped silica bodies. A–present; B–absent. (u)
175. Crescentic silica bodies. A–present; B–absent. (1.0; 1.0) A synapomorphy linking *Garnotia* with *Leptosaccharum* and *Oxyrhachis*.
176. Sharp-pointed silica bodies. A–present; B–absent. (0.25; 0.40)
177. Round to oval silica bodies. A–present; B–absent. (0.5; 0.0)
178. Elongated, smooth silica bodies. A–present; B–absent. (u)

179. Abaxial stomata. A–absent or very rare; B–common. (1.0; 1.0) A synapomorphy for *Leptosaccharum* and *Oxyrhachis*.
180. Guard cells. A–overlapped by interstomatals; B–overlapping to flush with interstomatals. (0.25; 0.0)
181. Triangular subsidiaries. A–absent; B–present. (0.25; 0.25)
182. Parallel-sided subsidiaries. A–absent; B–present.(u)
183. Subsidiaries. A–a mixture of triangular and parallel-sided; B–not parallel-sided and triangular on same leaf. (u)
184. Intercostal short-cells. A–common; B–absent or very rare. (0.06; 0.5) Excluded because of low CI.
185. Intercostal short-cells. A–in cork/silica-cell pairs; B–not paired. (0.14; 0.45)
186. Intercostal short-cells. A–silicified; B–not silicified. (0.2; 0.64)
187. Costal short-cells. A–conspicuously in long rows; B–predominantly paired; C–neither A nor B. (0.18; 0.31)
188. Maximum cells-distant count. A–1; B–2 or more. (u)
189. Anatomical organization. A–conventional; B–unconventional. (0.5; 0.0)
190. Organization of PCR tissue. A–*Alloteropsis* type; B–*Arundinella* type. (u)
191. Biochemical type. A–PCK, or NAD-ME; B–NADP-ME. (u)
192. Leaf blade xylem mestome sheath. A–present; B–absent . (0.33; 0.0)
193. PCR sheath outlines. A–uneven; B–even. (0.1; 0.1)
194. PCR sheath extensions. A–present; B–absent. (0.5; 0.67)
195. Maximum number of extension cells. A–1; B–2 or more. (0.5; 0.0)
196. Grana in PCR cell chloroplasts. A–well-developed; B–reduced. (u)
197. PCR cell chloroplasts. A–centrifugal/peripheral; B–centripetal. (u)
198. Chlorenchyma in mesophyll. A–radiate; B–non-radiate. (0.5; 0.5)
199. Adaxial palisade. A–present; B–absent. (u)
200. Mesophyll. A–*Isachne*-type; B–not *Isachne*-type. (u)
201. "Circular cells" in mesophyll. A–present; B–absent. (u)
202. Columns of colourless cells in mesophyll. A–present; B–absent. (0.17; 0.44)
203. Arm cells. A–present; B–absent. (u)
204. Fusoid cells. A–present; B–absent. (u)
205. Fusoid cells. A–an integral part of PBS; B–external to PBS. (u)
206. Leaf blade. A–with distinct, prominent adaxial ribs; B–"nodular" in section; C–adaxially flat. (0.2; 0.33)
207. Ribs of leaf blade. A–more or less constant in size; B–very irregular in size. (0.2; 0.2)
208. Midrib. A–conspicuous; B–not readily distinguishable from other main veins. (0.17; 0.64)
209. Midrib. A–with one bundle only; B–with a conventional arc of bundles; C–with complex vascularization. (0.2; 0.5)
210. Adaxial colourless tissue in midrib. A–extensive; B–absent. (0.06; 0.36) Excluded because of low CI.
211. Bulliform cells. A–present in discrete, regular adaxial groups; B–not in discrete, regular groups. (0.08; 0.4) Excluded because of low CI.
212. Bulliform cells. A–in simple fan-shaped groups; B–not in simple fans. (0.04; 0.18) Excluded because of low CI.
213. Bulliforms combined with colourless cells forming deeply-penetrating fan-shaped groups. A - present; B–absent. (0.17; 0.44)
214. Bulliforms and associated colourless cells forming arches over small vascular bundles. A - present; B–absent. (0.11; 0.47)
215. Vascular bundles. A–many unaccompanied by sclerenchyma; B–all accompanied by sclerenchyma. (0.08; 0.29) Excluded because of low CI.
216. Vascular bundles combining both adaxial and abaxial girders of sclerenchyma. A–present ; B–absent. (0.25; 0.0)
217. Combined girders. A–forming "figures"; B–nowhere forming "figures." (0.07; 0.07) Excluded because of low CI.
218. Sclerenchyma. A–all associated with vascular bundles; B–not all bundle-associated. (0.25; 0.25)
219. Culm internode bundles. A–in one or two rings; B–in three or more rings; C–scattered. (0.4; 0.4)
220. Chromosome base number, x = . A–7; B–9 to 10; C–11 to 12. (0.67; 0.0)

Appendix B3

ANNOTATED CHARACTER LIST—POOID CLADE

Numbers in parentheses refer to consistency index and retention index, respectively, in analyses of full data set (Fig. 10). u = uninformative for final set of included taxa. Inapplicable characters scored as missing. PAUP records CI even for excluded characters, but these are not included in length calculations. Many characters are illustrated in Watson and Dallwitz (1988). Unless otherwise noted, "lemmas" refers to lemmas of florets with fertile pistils (= "female-fertile lemmas" of Watson and Dallwitz (1988; 1991)). Notes on synapomorphies are based on the distribution of characters on tree shown in Fig. 10.

1. Longevity of plants. A-annual or biennial; B–perennial. (0.03;0.31) Excluded because of low CI.
2. Habit. A–long-rhizomatous or long-stoloniferous; B–caespitose or decumbent. (0.14; 0.0) Excluded because of high variability within genera.
3. Culms. A–branched above; B–unbranched. (u)
4. Culms. A–tuberous at base; B–not tuberous at base. (u)
5. Culm nodes. A–hairy; B–not hairy. (0.33; 0.0)
6. Culm internodes. A–solid or spongy; B–hollow. (0.33; 0.0) This character has been suggested as helpful in distinguishing subfamilies, but the low RI indicates that it is not useful below that level.
7. Young vegetative shoots. A–extravaginal; B–intravaginal. (0.5; 0.67) Excluded because poorly recorded.
8. Fresh shoots. A–aromatic when crushed; B–not aromatic. (1.0; 1.0) A synapomorphy for *Anthoxanthum* plus *Hierochloë*.
9. Leaves. A–mostly basal; B–not basally aggregated. (0.25; 0.79)
10. Auricles. A–present; B–absent. (0.33; 0.0)
11. Sheath margins. A–joined to at least 1/4 their length; B–free. (0.33; 0.87)
12. Leaf blades. A–setaceous; B–not setaceous. (0.2; 0.5)
13. Leaf blades. A–pseudopetiolate; B–not pseudopetiolate. (u)
14. Transverse veins on leaf blades. A–readily visible; B–not readily visible. (u)
15. Leaf blades. A–ultimately disarticulating from sheaths; B–not disarticulating. (u)
16. Leaf blades. A–rolled in bud; B–once-folded in bud; C–folded like a fan in bud. (0.17; 0.38)
17. Adaxial ligule. A–an unfringed membrane; B–a fringed membrane or a fringe of hairs. (u)
18. Adaxial ligule. A–truncate; B–not truncate (acute, obtuse, or rounded). (0.04; 0.25) This character represents an arbitrary division of a continuum; hence the low consistency index is not surprising. Excluded because of low CI.
19. Spikelets. A–of at least two sexually distinct forms on the same plant; B–alike in sexuality. (0.25; 0.4)
20. Plants. A–outbreeding; B–inbreeding. (0.17; 0.44) Excluded because of insufficient sampling.
21. Inflorescence. A–a single spike; B–a single raceme; C–paniculate. (0.67; 0.86)
22. Inflorescence. A–open; B–contracted. (0.04; 0.24) Excluded because of low CI.
23. Inflorescence. A–capitate to elongate-symmetrical, spike-like; B–more or less irregular. (u)
24. Divaricate branchlets in inflorescence. A–present; B–absent. (0.5; 0.0)
25. Capillary branchlets in inflorescence; A–present; B–absent. (0.1; 0.0) Excluded because of low CI.
26. Inflorescence axes. A–ending in spikelets; B–often not ending in spikelets. (0.5; 0.0)
27. Rachides. A–hollowed; B–flattened; C–neither flattened nor hollowed, not winged. (0.33; 0.75) Excluded because widely inapplicable.
28. Spikelets. A–all more or less partially embedded in the rachis; B–not all embedded. (0.5; 0.86)
29. Inflorescence. A–spatheate; B–espatheate. (u)
30. Spikelet-bearing axes. A–disarticulating; B–persistent. (0.25; 0.4)
31. Spikelet-bearing axes. A–falling entire; B–disarticulating at joints. (0.5; 0.0) Excluded because of insufficient sampling.
32. Spikelets. A–associated with bractiform involucres; B–all unaccompanied by bractiform involucres; C–with "involucres" of "bristles" at least some of them subtended by solitary "bristles." (0.5; 0.67)
33. Spikelets. A–mainly solitary; B–consistently in pairs or triplets. (u)
34. Spikelets. A–secund; B–not secund. (0.10; 0.31)
35. Spikelets. A–biseriate; B–distichous; C–not two-ranked. (0.67; 0.0) Excluded because of insufficient sampling; widely inapplicable.

36. Spikelets. A–all sessile; B–subsessile; C–some pedicellate. (0.67; 0.83)
37. Female-fertile spikelets. A–laterally compressed; B–not noticeably compressed; C–dorsally, ventrally, or dorsiventrally compressed. (0.18; 0.5)
38. Disarticulation. A–above glumes; B–below glumes; C–not disarticulating. (0.08; 0.21) Excluded because of low CI.
39. Female-fertile spikelets. A–not disarticulating between florets; B–disarticulating between florets. (0.2; 0.2)
40. Rachilla. A–prolonged apically; B–not prolonged apically. (0.11; 0.72)
41. Rachilla. A–hairy; B–hairless. (0.5; 0.94)
42. Callus. A–absent; B–short; C–long. (u)
43. Callus. A–pointed; B–blunt. (0.2; 0.2)
44. Hairy callus. A–present; B–absent. (0.04; 0.34) Excluded because of low CI.
45. Glumes. A–present; B–absent. (u)
46. Glumes. A–one per spikelet; B–two. (0.5; 0.0)
47. Glumes. A–minute; B–relatively large. (0.25; 0.0)
48. Glumes. A–very unequal in length; B–more or less equal. (0.04; 0.27) Excluded because of low CI. The underlying character is quantitative and continuous.
49. Glumes. A–<adjacent lemmas; B–>adjacent lemmas. (0.03; 0.36) Underlying character is quantitative and continuous. Excluded because of low CI. This has been used as the diagnostic character of the traditional Aveneae. As shown here it does not delimit any large monophyletic grup.
50. Glumes. A–joined, at least basally; B–free. (0.5; 0.0)
51. Glumes. A–conspicuously ventricose basally; B–not ventricose. (0.5; 0.0)
52. Glumes of sessile to subsessile spikelets. A–dorsiventral to rachis; B–lateral to rachis. (0.17; 0.0) Excluded because widely inapplicable. The character has been important in recognizing that *Lolium* is unrelated to the Triticeae.
53. Glumes. A–hairy; B–hairless. (0.33; 0.33)
54. Glumes. A–pointed; B–not pointed. (0.08; 0.08) Excluded because of low CI.
55. Glumes. A–subulate; B–not subulate. (u)
56. Glumes. A–awned; B–awnless. (0.2; 0.33)
57. Glumes. A–one-keeled; B–not one-keeled. (0.04; 0.47)
58. Glumes. A–very dissimilar in form or texture; B–more or less similar. (0.25; 0.25) Excluded. Very dissimilar glumes may be produced in different ways, and are not comparable.
59. Incomplete florets. A–absent; B–present. (0.04; 0.52) Excluded because of low CI.
60. Incomplete, male or sterile florets. A–proximal to female-fertile florets; B–distal to female-fertile florets. (1.0; 1.0) A synapomorphy for the Phalarideae.
61. Proximal incomplete florets. A–1; B–2. (u)
62. Proximal incomplete florets. A–paleate; B–epaleate. (u)
63. Lemmas of proximal incomplete florets. A–awned; B–awnless. (u)
64. Number of female-fertile florets per spikelet. A–1; B–2. (0.05; 0.62) Excluded because of low CI. This is the character traditionally used to recognize the tribe Agrostideae.
65. Lemmas. A–convolute; B–not convolute. (0.5; 0.5)
66. Lemmas. A–saccate; B–not saccate. (1.0; 1.0) A synapomorphy for *Nassella* and *Piptochaetium.*
67. Lemmas. A–less firm than the firmer of the glumes; B–similar in texture to glumes; C–decidedly firmer than glumes. (0.06; 0.42) Excluded because of low CI. The underlying character is quantitative and continuous.
68. Lemmas. A–becoming indurated when mature and dry; B–not becoming indurated. (0.33; 0.78)
69. Apex of lemmas. A–entire; B–incised. (0.04; 0.46) Excluded because of low CI.
70. Apex of lemmas. A–pointed; B–blunt. (0.33; 0.82)
71. Lemmas. A–awnless to mucronate; B–awned. (0.04; 0.33) Excluded because of low CI.
72. Awns. A–median; B–median and lateral; C–lateral only. (0.14; 0.0) Excluded because of low CI, but also dependent on #71, which is excluded.
73. Median awns. A–different in form from laterals; B–similar in form to laterals. (0.5; 0.5) Excluded because dependent on #72, which is excluded.
74. Awns of lemmas. A–from sinus; B–dorsal; C–apical. (0.09; 0.26) Excluded because of low CI, but also because dependent on #71, which is excluded.
75. Awns of dorsally-awned lemmas. A–from near top; B–from well down back. (0.09; 0.23) Excluded because of low CI, but also because dependent on #74, which is excluded.
76. Awns of lemmas. A–straight or curved; B–geniculate. (0.04; 0.26) Excluded because of low CI, but also because dependent on #71, which is excluded.

77. Awns. A–<<body of lemma; B–ca. = body of lemma; C–>>body of lemma. (0.22; 0.36) Excluded because dependent on #71, which is excluded.
78. Awns. A–entered by one vein; B–entered by several veins. (0.17; 0.44) Excluded because dependent on #71, which is excluded.
79. Awns. A–deciduous; B–persistent. (0.25; 0.0) Excluded because dependent on #71, which is excluded.
80. Lemmas. A–conspicuously hairy; B–hairless. (0.05; 0.24) Excluded because of low CI.
81. Lemmas. A–with 1 median keel; B–rounded, flat, or with 2 or more keels. (0.04; 0.08) Excluded because of low CI.
82. Veins of lemmas. A–confluent towards tip; B–not confluent apically. (1.0; 1.0) Excluded because of inadequate sampling.
83. Palea. A–relatively long; B–conspicuous but relatively short. (0.07; 0.0) Excluded because of low CI.
84. Palea. A–gaping; B–tightly clasped by lemma. (0.5; 0.5) Excluded because of inadequate sampling.
85. Palea. A–entire; B–apically notched. (0.14; 0.33)
86. Palea. A–awnless, without apical setae; B–with apical setae; C–awned. (0.5; 0.0)
87. Palea. A–thinner than lemma; B–similar in texture to lemma. (0.2; 0.56)
88. Palea. A–indurated; B–not indurated. (0.33; 0.5)
89. Palea keels. A–one; B–two; C–absent. (0.22; 0.68)
90. Palea keels. A–winged; B–wingless. (u)
91. Lodicules. A–2; B–3. (0.25; 0.5) Lack of lodicules is scored as missing data, because it could be derived from a plant with either two or three lodicules.
92. Lodicules. A–joined at least basally; B–free. (0.5; 0.75)
93. Lodicules. A–distally fleshy; B–distally membranous. (0.33; 0.6)
94. Lodicules. A–ciliate or hairy; B–glabrous. (0.17; 0.17)
95. Lodicules. A–toothed; B–not toothed. (0.05; 0.47) Excluded because of low CI.
96. Stamen number. A–1; B–2; C–3. (0.29; 0.0)
97. Ovary apex. A–glabrous; B–hairy. (0.07; 0.07) Excluded because of low CI. This is sometimes used to support the monophyly of the Triticodae.
98. Apical appendage on ovary. A–present; B–absent. (0.5; 0.5)
99. Styles. A–fused, at least basally; B–free to base. (0.14; 0.45)
100. Stigmas. A–1; B–2–4. (u)
101. Fruit. A–adhering to lemma and/or palea; B–free from both lemma and palea. (0.08; 0.14) Excluded because of low CI.
102. Fruit. A–small, <4 mm long when mature; B–medium-sized (4–10mm long) to large (more than 10 mm long). (0.07; 0.07) Excluded because of low CI. An artificial division of a continuous quantitative character.
103. Fruit. A–longitudinally grooved; B–not grooved. (0.06; 0.52) Excluded because of low CI.
104. Fruit. A–compressed laterally; B–compressed dorsally, ventrally, or dorsiventrally; C–not noticeably compressed. (0.21; 0.58)
105. Fruit hairs. A–confined to a terminal tuft; B–on body of fruit. (0.5; 0.5)
106. Hilum. A–short, punctiform or shortly elliptical, <1/2 length of fruit; B–long-linear, >1/2 as long as fruit. (0.04; 0.32) Excluded because of low CI. An arbitrary division of a continuum. This character is quite consistent, however, in other parts of the family.
107. Pericarp. A–loose, free; B–fused. (0.33; 0.0)
108. Embryo. A–large, at least 1/3 as long as fruit; B–small, <1/3 as long as fruit. (u)
109. Embryo. A–waisted in surface view; B–not waisted. (0.5; 0.0)
110. Endosperm. A–liquid in mature fruit; B–hard. (0.08; 0.2) Excluded because of low CI.
111. Endosperm. A–with lipid; B–without lipid. (0.14; 0.74) This character is unique to the pooids, being unknown elsewhere in the family.
112. Starch grains in endosperm. A–simple; B–compound. (0.17; 0.29)
113. Epiblast. A–present; B–absent. (u)
114. Scutellar tail.. A–present; B–absent. (u)
115. Embryonic leaf margins. A–meeting; B–overlapping. (0.5; 0.0)
116. Seedling mesocotyl. A–short; B–long. (0.25; 0.4)
117. Coleoptile. A–loose; B–tight. (0.25; 0.25)
118. First seedling leaf. A-erect; B–curved; C–supine. (0.5; 0.0)
119. Microhairs. A–present; B–absent. (u) The one-celled structures found in some *Stipa* species (Johnston

& Watson, 1976; Scholz, 1982; Renvoize, 1985) are here assumed not to be homologous with microhairs. *Anisopogon* & *Metcalfia* lack microhairs. Thus this character is phylogenetically uninformative in this data set.

120. Abaxial leaf blade epidermis with costal/intercostal zonation. A–conspicuous; B–lacking. (0.09; 0.23) Excluded because of low CI.
121. Abaxial leaf blade epidermis long-cells. A–similar in shape costally and intercostally; B–markedly different in shape costally and intercostally. (0.07; 0.19) Excluded because of low CI.
122. Walls of long-cells of abaxial leaf blade epidermis. A–of similar thickness costally and intercostally; B–differing markedly in wall thickness costally and intercostally. (0.08; 0.25) Excluded because of low CI.
123. Mid-intercostal long-cells. A–more or less rectangular; B–more or less fusiform. (0.04; 0.29) Excluded because of low CI.
124. Walls of mid-intercostal long-cells. A–sinuous; B–straight or gently undulating. (0.04; 0.52) Excluded because of low CI.
125. Papillae. A–present in abaxial leaf blade epidermis; B–absent. (0.25; 0.4)
126. Intercostal papillae of abaxial leaf blade epidermis. A–frequently over-arching stomata; B–not over-arching stomata. (0.5; 0.5)
127. Crown cells. A–present; B–absent. (0.5; 0.5)
128. "Pooid-type" silica bodies. A–present; B–absent. (0.04; 0.35) Excluded because of low CI.
129. "Panicoid" silica bodies. A–present; B–absent. (0.2; 0.69)
130. Tall-and-narrow silica bodies. A–present; B–absent. (0.14; 0.5)
131. "Chloridoid-type" silica bodies. A–present; B–absent. (u)
132. Crescentic silica bodies. A–present; B–absent. (0.11; 0.11)
133. "Oryzoid-type" silica bodies. A–present; B–absent. (u)
134. Sharp-pointed or acutely-angled silica bodies. A–present; B–absent. (0.5; 0.0)
135. Round, oval or potato-shaped silica bodies. A–present; B–absent. (0.03; 0.28) Excluded because of low CI.
136. Horizontally elongated, smooth silica bodies. A–present; B–absent. (0.04; 0.37) Excluded because of low CI.
137. Stomata on abaxial leaf blade epidermis. A–absent or very rare; B–common. (0.04; 0.24) Excluded because of low CI.
138. Guard cells. A–overlapped by interstomatals, i.e., sunken in epidermis; B–overlapping interstomatals. (0.33; 0.6)
139. Triangular subsidiaries. A–absent; B–common. (0.33; 0.82)
140. Parallel-sided subsidiaries. A–absent; B–common. (0.2; 0.71)
141. Intercostal short-cells. A–common; B–absent or very rare. (0.04; 0.47) Excluded because of low CI.
142. Intercostal short-cells. A–in cork/silica cell pairs; B–not paired. (0.08; 0.56) Excluded because of low CI.
143. Intercostal short-cells. A–silicified; B–not silicified. (0.1; 0.57) Excluded because of low CI.
144. Costal short-cells. A–conspicuously in long rows of five or more cells; B–predominantly paired; C–not distinctly grouped into long rows or paired. (0.15; 0.64)
145. Leaf blade mesophyll. A–with radiate chlorenchyma; B–without radiate chlorenchyma. (0.5; 0.5)
146. Leaf blade. A–with distinct, prominent adaxial ribs only; B–nodular in section; C–adaxially more or less flat. (0.18; 0.61)
147. Leaf blade ribs. A–more or less constant in size; B–very irregular in size. (0.25; 0.73)
148. Midrib of leaf blade. A–conspicuous, prominent in outline, with distinctive sclerenchyma; B–not readily distinguishable from other main veins. (0.05; 0.44) Excluded because of low CI.
149. Midrib of leaf blade. A–with one bundle only; B–with a conventional arc of bundles; C–with complex vascularization. (u)
150. Midrib and/or middle part of leaf blade. A–with extensive colourless tissue adaxially; B–without conspicuous colourless tissue adaxially. (0.5; 0.0)
151. Adaxial groups of bulliform cells in leaf blade. A–present in discrete, regular groups; B–absent, or in irregular groups, or constituting most of the epidermis. (0.17; 0.54)
152. Simple fan-shaped groups of bulliform cells. A–present; B–absent. (0.05; 0.17) Excluded because of low CI.
153. Smallest vascular bundles. A–unaccompanied by sclerenchyma; B–all vascular bundles accompanied by sclerenchyma. (0.25; 0.40)

154. Vascular bundles combining both adaxial and abaxial girders of sclerenchyma. A–at least some present, if only the midrib; B–absent. (0.04; 0.19) Excluded because of low CI.
155. Adaxial and abaxial sclerenchyma girders in one or more bundles of the leaf blade. A–forming "figures" ("anchors," I's, or T's); B–nowhere forming "figures." (0.05; 0.4) Excluded because of low CI.
156. Leaf blade sclerenchyma. A–all associated with vascular bundles, apart from any marginal fibres; B–not all obviously bundle-associated. (0.2; 0.69)
157. Culm internode bundles. A–in 1 or 2 rings; B–in 3 or more rings; C–scattered. (u)
158. Chromosome base number, x=. A–7; B–9 or 10; C–11 or 12. (0.4; 0.7)