A molecular phylogeny of the subfamily Arundinoideae (Poaceae)

Nigel Paul Barker

Thesis Presented for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Botany

UNIVERSITY OF CAPE TOWN

Febuary 1995

UT 580 BARK.

95/10718

Dedicated to my mother, Audrey L. Barker,
who believed in it but never lived to see
it finished, and to my wife, Janet, for
her support all through it.

ACKNOWLEDGEMENTS

I would like to extend my grateful thanks to the following people:

My supervisors, Prof. **H. Peter Linder** and Prof. **Eric H. Harley**, for their advice, interest and support throughout the course of this study.

The **Foundation for Research and Development** for financial support.

Dr **E.A. Kellogg** (Harvard University, Massachusetts) for her generosity in sharing *rpo*C2 primers, ideas, computer time, hospitality and support.

Dr **Jerold I. Davis** and Dr **Robert J. Soreng** (Cornell University, New York) for the provision of their data, ideas, hospitality, humour and encouragement.

Prof. **Ed P. Rybicki** for use of his version of PAUP.

Without the specimens provided by the following peiople, this research would have been impossible: **R. Buxton** (New Zealand), **L.G. Clark** (Iowa State University, Iowa) **H.E. Connor** (New Zealand), **S. Darbyshire** (Canada), **W. Hahn** (Smithsonian Institute, Washington DC.), **E.A. Kellogg** (Harvard University), **M. Kobayashi** (Utsunomiya University, Japan), **H.P. Linder** (University of Cape Town), **M. Lyle** (Frankfurt), **C. McDowell** (University of Cape Town), **E. Moll** (University of Queensland), **S.A. Renvoize** (Royal Botanic Gardens, Kew), **N. Snow** (Missouri Botanical Gardens, St Louis), **L. Watson** (Australia) and **J.F. Wendel** (Iowa State University).

CONTENTS

ABSTRACT

Chapter 1. INTRODUCTION: THE SUBFAMILY ARUNDINOIDEAE

Chapter 2. PHYLOGENETIC ANALYSIS OF THE rpoC2 SEQUENCE DATA

ABSTRACT

The subfamily Arundinoideae has long been considered to be an unnatural assemblage of genera, the relationships of which are obscure or unknown. Because morphological and anatomical data have, to date, been unable to elucidate relationships among these genera, sequence data from two chloroplast genes are used to elucidate relationships among 33 arundinoid genera.

Sequence data from the variable, grass-specific insert in the rpoC2 gene is used to determine the relationships among 73 grass species from all currently recognised subfamilies. Phylogenetic analysis of this sequence data required the development of specialised alignment techniques based on testing assumptions of positional homology. Results of the analyses based on these alignments suggest that the Arundinoideae is divisible into four lineages, corresponding approximately to the tribes Danthonieae, Arundineae, Aristideae and Thysanolaeneae. Several arundinoid representatives are placed in other subfamilies.

The rpoC2 sequence data was too variable to elucidate relationships at the tribal and subfamilial level. For this purpose, sequence data of the highly conserved rbcL gene was obtained from 22 taxa selected from the lineages identified by the rpoC2 study. Phylogenetic analysis of a total of 36 sequences resolved some of the relationships of the major clades, but other relationships were poorly supported.

In an attempt to improve the resolution of these major clades, the rpoC2 and rbcL data sets were combined with restriction site data. These three data sets were analysed in a variety of combinations using both data combination and tree consensus methods to assess support of the phylogenetic relationships. Despite this, the resolution of the relationships among the Arundineae, Danthonieae, Aristideae and Chloridoideae was not

resolved with any finality, although a (Arundineae (Danthonieae (Aristideae, Chloridoideae))) relationship is proposed as being most likely.

The molecular phylogeny implies that eight grass subfamilies should be recognised. Two of these, the Danthonioideae and Aristidoideae, are new and the Arundinoideae is redelimited. Furthermore, new tribes in the subfamilies Centothecoideae (Thysanolaeneae) and Chloridoideae (Centropodieae) are proposed to accommodate lineages and taxa misplaced in the subfamily Arudinoideae as previously delimited.

CHAPTER 1

INTRODUCTION:

THE SUBFAMILY ARUNDINOIDEAE

The subfamily Arundinoideae was erected by Tateoka (1957) to accomodate those taxa with one or more florets (the uppermost of which is often reduced) and a basic chromosome number of 12, 11 or 10. In addition, leaf sectional anatomy of this group was described as festucoid or bambusoid, while the epidermides were noted to be panicoid or festucoid. Tateoka thus drew heavily on the anatomical work of Prat (1936) and the cytology of Avdulov (1931), although he re-interpreted their work with respect to several genera.

The subfamily has subsequently been variously defined by numerous grass systematists, all of whom have only managed to agree on one fact; that the group as a whole is problematic, and probably polyphyletic (Campbell 1985, Clayton and Renvoize 1986, Conert 1987, Ellis 1987, Kellogg and Campbell 1987, Renvoize 1981, Watson 1990). As noted by Kellogg and Campbell (1987), there is no single character that unites the subfamily, or even a large subset of it. These studies, summarised in Table 1.1, were based on a variety of characters and methods.

It is clear that these studies have been unable to resolve the relationships within the subfamily, or even agree on the tribal level classification. It is difficult to reconcile the continued maintenance of this subfamilial classification with the repeated observations that the subfamily (and some of its constituent tribes) are probably polyphyletic. The reason for

Table 1.1. A summary of recent studies in which the relationships of the Arundinoideae are assessed.

Classification	Character source	Analytical methods	Major conclusions	Comments
Renvoize (1981)	Leaf anatomy	Phenetic (ordination)	8 tribes in Arundinoideae	Genera arbitrarily divided into "core" and outlying groups
Hilu and Wright (1982)	Morphology	Phenetic	Arundinoideae includes Arundineae and Danthonieae	Small sample size of Arundinoideae
Renvoize (1986a); Clayton and Renvoize (1986)	Morphology Leaf anatomy	Pragmatic phyletics	4 tribes in Arundinoideae	No cladistic analysis
Conert (1987)	Breeding systems		3 tribes, incl. Cortaderieae	Incomplete survey; ignored certain groups (eg. Aristideae)
Kellogg and Campbell (1987)	Morphology Leaf anatomy Embryology	Cladistic	Arundinoideae polyphyletic and basal	Limited by computer technology; some missing data
Watson (1990)	Morphology Leaf anatomy Embryology Biochemistry	Phenetic	Arundinoideae an unsatisfactory group; 11 tribes	

this can be ascribed at least in part to the lack of confidence associated with the various character sets. This may be caused by the wide range of character states, erroneous assessments of homology of these characters, incomplete data across the range of taxa and lack of material and data for many of the smaller genera.

Tribal classification

When the subfamily was erected, Tateoka (1957) identified no less than 17 tribes. This number has subsequently been reduced, largely by the transfer of tribes and genera to other subfamilies, but no consensus can be reached on the issue. Today there are three competing tribal classifications. These are discussed below, and are arbitrarily named after the main proponents: Clayton and Renvoize (1986), Conert (1987), and Watson (1990). Table 1.2 compares these classifications on the basis of the genera included in the tribes.

Clayton and Renvoize's system

Following on from studies by Renvoize (1981, 1986a), Clayton and Renvoize (1986) proposed four tribes: the Arundineae, Aristideae, Thysanolaeneae and Micraireae, the latter two tribes being monogeneric. Of the four tribes, the Arundineae remains by far the largest, and includes genera considered by others to belong in the tribe Danthonieae. The Arundineae are defined by embryo features, non-kranz leaf anatomy (including the presence of slender microhairs), and a generally simple spikelet structure. Despite these apparently unifying features, Clayton and Renvoize (1986) consider the Arundineae to be a heterogenous collection of taxa with uncertain relationships consistent with a history of decline.

The Arundineae is further divided into six facies which are not homogenous or sufficiently defined to be recognised as subtribes (Clayton and Renvoize 1986). These groups are:

Table 1.2. The genera which have been, or are, in the Subfamily Arundinoideae. This table lists the tribe in which the genera are placed, and the workers who placed them there. Three classification systems are compared; Clayton and Renvoize (1986), Conert (1987) and Watson (1990). **Note:** 1) That these are according to the most recent taxonomies available and do not reflect the earlier concepts of these workers. 2) Not all genera of the tribe Stipeae sensu Watson are included; the group is represented by the genus Stipa. 3) Those genera in **bold** are of such uncertain placings that the subfamilial concepts differ between the workers. 4) Empty cells indicate that the genus is not mentioned by that particular worker. 5) Clayton and Renvoize's concept of Rytidosperma is not followed here. Instead, the individual genera that they included in Rytidosperma are retained.

Key (subfamilies): Arund = Arundinoideae, Chlor = Chloridoideae, Panic = Panicoideae, Pooid = Pooideae. **Key (Tribes):** Arist = Aristideae, Aru = Arundineae, Aven = Aveneae, Cortad = Cortaderieae, Cyp = Cyperochloeae, Danth = Danthonieae, Erag = Eragrostideae, Eriac = Eriachneae, Lyg = Lygeae, Mic = Micrairieae, Nard = Nardeae, Stey = Steyermarkochloeae, Stip = Stipeae, Spart = Spartochloeae, Thys = Thysanolaeneae, n/a = genus was described subsequent to the classification under consideration.

Genus	Clayton and Renvoize (1986)		Conert (1987)		Watson (1990)	
	S-Fam	Tribe	S-Fam	Tribe	S-Fam	Tribe
Alloeochaete	Arund	Aru	Arund	Cortad	Arund	Danth
Amphipogon	Arund	Aru			Arund	Danth
Anisopogon	Arund	Aru			Arund	Stip
Aristida	Arund	Arist			Arund	Arist
Arundo	Arund	Aru	Arund	Aru	Arund	Aru
Arundoclaytonia	n/a		n/a		Arund	Stey
Centropodia	Arund	Aru	Arund	Danth	Arund	Danth
Chaetobromus	Arund	Aru	Arund	Danth	Arund	Danth
Chionochloa	Arund	Aru	Arund	Cortad	Arund	Danth

Genus	Clayton and Renvoize (1986)		Conert (1987)		Watson (1990)	
	S-Fam	Tribe	S-Fam	Tribe	S-Fam	Tribe
Cortaderia	Arund	Aru	Arund	Cortad	Arund	Danth
Crinipes	Arund	Aru			Arund	Danth
Cyperochloa	n/a		n/a		Arund	Cyp
Danthonia	Arund	Aru	Arund	Danth	Arund	Danth
Danthonidium	Arund	Aru			Arund	Danth
Dichaetaria	Arund	Aru			Arund	Danth
Diplopogon	Arund	Aru			Arund	Danth
Dregeochloa	Arund	Aru	Arund	Danth	Arund	Danth
Duthiea	Pooid	Aven			Arund	Danth
Elytrophorus	Arund	Aru			Arund	Danth
Eriachne	Panic	Eriac			Arund	Eriac
Erythanthera	Arund	Aru	Arund	Danth	Arund	Danth
Gynerium	Arund	Aru	Arund	Aru	Arund	Danth
Habrochloa	Chlor	Erag			Arund	Danth
Hackonechloa	Arund	Aru	Arund	Aru	Arund	Danth
Karroochloa	Arund	Aru	Arund	Danth	Arund	Danth
Lamprothyrsus	Arund	Aru	Arund	Cortad	Arund	Danth
Leptagrostis	Arund	Aru			Pooid	Avena
Lygeum	Pooid	Lyg			Arund	Lyg
Merxmuellera	Arund	Aru	Arund	Cortad	Arund	Danth
Metcalfia	Pooid	Aven			Arund	Danth
Micraira	Arund	Mic			Arund	Mic

Genus	Clayton and Renvoize (1986)		Conert (1987)		Watson (1990)	
	S-Fam	Tribe	S-Fam	Tribe	S-Fam	Tribe
Molinia / Moliniopsis	Arund	Aru	Arund	Aru	Arund	Danth
Monachather	Arund	Aru			Arund	Danth
Monostachya	Arund	Aru	Arund	Danth	Arund	Danth
Nardus	Pooid	Nard			Arund	Nard
Nematopoa	Arund	Aru			Arund	Danth
Notochloe	Arund	Aru			Arund	Danth
Notodanthonia	Arund	Aru				
Pentameris	Arund	Aru	Arund	Danth	Arund	Danth
Pentaschistis	Arund	Aru	Arund	Danth	Arund	Danth
Phaenanthoecium	Arund	Aru			Arund	Danth
Pheidochloa	Panic	Eriac			Arund	Eriac
Phragmites	Arund	Aru	Arund	Aru	Arund	Aru
Piptophyllum	Arund	Aru			Chlor	Chlor
Plinthanthesis	Arund	Aru	Arund	Danth	Arund	Danth
Poagrostis	Arund	Aru	Arund	Danth	Arund	Danth
Prionanthium	Arund	Aru			Arund	Danth
Pseudodanthonia	Pooid	Aven			Arund	Danth
Pseudopentameris	Arund	Aru	Arund	Danth	Arund	Danth
Pyrrhanthera	Arund	Aru	Arund	Danth	Arund	Danth
Rytidosperma	Arund	Aru	Arund	Danth	Arund	Danth
Sartidia	Arund	Arist			Arund	Arist
Schismus	Arund	Aru	Arund	Danth	Aru	Danth

Genus	Clayton and Renvoize (1986)		Conert (1987)		Watson (1990)	
	S-Fam	Tribe	S-Fam	Tribe	S-Fam	Tribe
Spartochloa	Arund	Aru			Arund	Spart
Steyermarkochloa	Panic	Stey			Arund	Stey
Stipa	Pooid	Stipa			Arund	Stipa
Stipagrostis	Arund	Arist			Arund	Arist
Styppeiochloa	Arund	Aru			Arund	Danth
Thysanolaena	Arund	Thys			Arund	Aru
Tribolium	Arund	Aru			Arund	Danth
Urochlaena	Arund	Aru			Arund	Danth
Zenkeria	Arund	Aru			Arund	Danth

1. The "primitive" genera, a heterogenous group united by their short glumes and multinerved lemmas with entire tips. This group consists of *Tribolium*, *Urochlaena*, *Elytrophorus*, *Prionanthium*, *Spartochloa*, *Notochloe*, *Zenkeria*, *Piptophyllum* and *Styppeiochloa*.

2. Those taxa with long glumes and bilobed lemmas with a geniculate central awn; a group thought to be an agglomeration of closely related and narrowly defined genera whose taxonomy is still controversial. Included here are *Chionochloa*, *Danthonia*, *Chaetobromus*, *Plinthanthesis*, *Pentaschistis*, *Pentameris*, *Poagrostis*, *Pseudopentameris*, *Rytidosperma*, *Phaenanthoecium*, *Alloechaete*, *Monachather*, *Schismus*, *Pyrrhanthera*, *Dregeochloa* and *Centropodia*. This and the previous group approximates to a large extent the tribe Danthonieae *sensu* Watson (1990).

3. Taxa with one-flowered spikelets with wrap-around lemmas. This character overlaps with the tribe Aristideae, and links to this group are considered to be strong. Taxa included here are *Danthonidium*, *Anisopogon*, *Diplopogon* and *Amphipogon*. The former two genera are considered to have relatives in group 2, but the latter two genera are considered to be intermediate between the Arundineae and the Aristideae. They are placed in the former tribe on the basis of their unreduced palea.

4. Four genera with three-nerved lemmas are placed in the fourth group. This trend is also found in the next two groups. These genera are *Crinipes*, *Nematopoa*, *Leptagrostis* and *Dichaetaria*.

5. This group is characterised by a dioecious breeding system. In addition to this, the megagametophyte has haustorial synergid cells, a character which also is found in group 2. Included here are *Cortaderia* and *Lamprothyrsus*.

6. This group, comprising *Arundo*, *Hakonechloa*, *Molinia*,

Phragmites and Gynerium, is united by a reedy habit, but the constituent genera are considered to be heterogeneous.

Conert's system

These tribal definitions are based on the basis of breeding systems, and Conert (1987) outlined three tribes: Arundineae, Danthonieae and Cortaderieae. The latter tribe is unique to this classification. The Cortaderieae approximates Clayton and Renvoize's group 5 and part of group 2, discussed above. Conert (1987) considers this tribe to be the youngest and most modern in the subfamily. However, he makes no mention of the Aristideae or its constituent genera.

The Cortaderieae was initially erected by Zotov (1963), who comments that the tribe occupies a position intermediate between the Arundineae and the Danthonieae. Only Cortaderia was, at that time, allocated to this tribe, and Zotov (l.c.) notes that the floral structure of this genus resembles that of Arundo, while the foliar structure resembles that of Chionochloa, which was then placed in the Danthonieae.

Prior to Zotov's work, Conert (1961) had divided the Arundineae into five subtribes; the Arundiniinae, the Cortaderiinae, the Crinipinae, the Ampelodesminae and the Moliniinae. Thus, Cortaderia and Lamprothysus were already identified as members of a separate subtribe, the Cortaderiinae, on the basis of the plants being dioecious. Zotov's recognition of this group was therefore not original. Zotov (1963) also validated the tribal name "Danthonieae", which was adopted without formal characterisation by Nevski (1937) after it was first proposed by Prat (1936).

Zotov does not cite Conert's 1961 work, and it is not clear if these two workers arrived independantly at a similar conclusion. Conert (1987) thus serves only to modify Zotov's tribal concept by placing Chionochloa, Lamprothyrsus, Merxmuellera and Allochaete into this tribe, largely on the basis of their gynodioecism or dioecism. This action invalidates the use of the earlier subtribe Cortaderiinae. In spite of the emphasis on breeding systems, Conert (1987) notes that the African and New Zealand taxa are generally hermaphrodite, while the South American groups have modified breeding systems.

The second tribe, the Danthonieae, comprises many genera, most of which are Gondwanean in distribution, with the exception of Danthonia. The third tribe, the Arundineae is considered to be very ancient. In support of this, Conert cites the small number of species in each genus in this group, and the fact that many are pandemic.

Watson's system

In 1976 Watson and Clifford placed representatives of the subfamily Arundinoideae in four somewhat informal, groups; the Aristideae, the Stipeae, the "arundinoids" and the "danthonioids". The "arundinoids" comprised Arundinella, Arundo, Garnitia, Phragmites and Thysanolaena, while the "danthonioids" comprised (among others) Chionochloa, Cortaderia, Danthonia, Molinia, Pentaschistis and Schismus. This classification was retained in the automated DELTA database of the world's grasses (Watson 1987, 1990, Watson and Dallwitz 1988, 1992, Watson, Dallwitz and Johnstone 1986).

In a recent review of the classification of the family, Watson (1990) recognises no less than 11 tribes within the subfamily Arundinoideae. These are: the Stipeae, Steyermarkochloeae, Nardeae, Lygeae, Arundineae, Danthonieae, Spartochloeae, Cyperochloeae, Micraireae, Aristideae and Eriachne. These 11 groups were recognised after what Watson (1990) describes as extensive and continuing taxonomic analyses of a large set of descriptive data, employing a variety of mainly phenetic numerical methods. The inclusion of three of these (the

Stipeae, Lygeae and Nardeae) is controversial as they are placed by other workers (eg. Clayton and Renvoize 1986) in the subfamily Pooideae. Watson feels that his classification corresponds in many areas with that of Clayton and Renvoize (1986), or at least with certain of their lineages, and considers the Arundinoideae to be an assemblage of convenience which is not amenable to anything approaching a diagnostic description and which is probably polyphyletic. The individual arundinoid tribes are coherent and useful, but their relationships to one another and to the other subfamilies remain controversial (Watson 1990).

To date, all the attempts to clarify the relationships of and within the subfamily have been based on extensive surveys of morphology and anatomy, and to a lesser extent micromolecular characters. The apparent failure of these studies to clarify relationships of the Arundinoideae suggests that these characters are unsuitable for assessing phylogenetic relationships. A novel data set is thus required; one in which the homologies of the characters can be tested, and in which the character states are comparatively uncomplicated. Such data is readily available from the DNA sequences of genes, particularly chloroplast genes.

The use of chloroplast DNA sequences in systematics

Phylogenies based on sequence data reflect the history of the sequence, not the organism. Assumptions about the mode of inheritance, evolution and mutation allow for the extrapolation of the gene trees into species trees. Plastids are maternally inherited in many plants, and in all grasses examined to date, with the exception of Secale cereale, which has been documented to have biparental inheritance (Harris and Ingram 1991). Phylogenies obtained from plastid sequences will thus reflect the maternal lineage. Should hybridisation have occurred in the history of an organism, it will not be reflected in the plastid gene phylogeny. Instead, it will only be detectable if compared to another phylogeny that is based on the nuclear genome, where the positions of the taxon of hybrid origin will be in conflict. These caveats noted, the use of gene sequences nevertheless becomes a powerful means of generating substantial quantities of phylogenetic characters.

Previous cpDNA studies in the Poaceae

There have been previous attempts to determine subfamilial relationships in the grasses by means of DNA-based studies, and the grasses were among the first plant families to be the subject of pioneering sequencing studies (Doebley et al. 1990, Hamby and Zimmer 1988). These studies were carried out to assess methodological approaches and to test the suitability of the genes used.

Past studies take the form of either gene sequencing or plastid genome restriction site studies. Details of these are listed in Table 1.3. With the exception of the studies by Cummings et al. (1994) and Davis and Soreng (1993), these studies have all been limited by under-sampling and the use of distant outgroups.

The first sequencing study undertaken on the grasses used nuclear ribosomal RNA sequence data from three subfamilies (Hamby and Zimmer 1988). In part a methodological exercise, this study sequenced over 1600 bases from each of the 10 species. This yielded 85 phylogenetically informative sites. Parsimony analysis of this data (using the very distant outgroup Psilotum) indicated a (bambusoid (pooid, panicoid)) relationship, although the Bambusoideae were shown to be paraphyletic. In spite of the low number of informative sites obtained from the data, the authors were optimistic that the gene and method could be used to successfully elucidate phylogenetic relationships.

Table 1.3. A synthesis of the DNA-based studies carried out on the grass family. Unless otherwise indicated, all studies are based on chloroplast DNA. **Key**: RSM = Restriction Site Mapping; rbcL = large subunit of ribulose-1,5-bisphosphate carboxylase; rpoC2 = C" subunit of RNA polymerase; rps4 = ribosomal protein subunit 4. The figures in the column of sample size refer to the number of grass taxa only; outgroups are not counted.

Study	Sample size	Subfamilies	Results	Comments
Nuclear rRNA (Hamby and Zimmer 1988)	9	Bambusoideae, Panicoideae and Pooideae	Bambusoideae not monophyletic (based on 2 samples)	Insufficient sampling across subfamilies; distant outgroup
rbcL sequence (Doebley et al. 1990)	10	Bambusoideae Panicoideae Pooideae	Unable to confirm basal lineage (analysis dependant)	Insufficient sampling across subfamilies; distant outgroup
rpoC2 sequence (Cummings et al. 1994)	13	All (they state five; include Centothecoideae in Arundinoideae)	Gene too variable to resolve relationships between subfamilies	Distant outgroup; computerised alignments
rps4 sequence (Nardot et al. 1994)	28	Four; Centothecoideae and Arundinoideae excluded	Gene useful at subfamily and tribal level	Need additional samples from excluded subfamilies.
RSM (Davis and Soreng 1993)	31	All (six)	Unable to confirm basal lineage; PACC clade monophyletic	Arundinoideae not well sampled

In the first chloroplast DNA sequencing on the grasses, Doebley et al. (1990) used sequence data from the large subunit of ribulose-1,5-bisphosphate carboxylase (rbcL) to 1) determine if the gene evolved in a clock-like manner in the grasses, 2) to determine and statistically test phylogenies, 3) to test tree building methods and 4) to obtain novel insights into the evolution of the grass family. This pioneering study (carried out prior to the popularisation of rbcL in molecular systematics) found that the gene had evolved in a clock-like manner in the grasses, but the conflict between results of the different tree building methods and the insufficient information from the sequence detracted from its usefulness as the gene of choice for systematics studies at this level. The use of these sequences has subsequently been avoided.

Recent DNA based studies have centered on elucidating generic or infrageneric relationships by means of restriction site studies, such as those on Puccinellia (Choo et al. 1994) and the subfamily Pooideae (Soreng et al. 1990). The exception to these smaller studies is the work of Davis and Soreng (1993), which used plastid restriction site mapping data to examine the relationships among the grass subfamilies. Although this study includes representatives of all the subfamilies, 24 of the 31 genera included in the study are from the Pooideae and Bambusoideae. While attempting to elucidate relationships among the higher lineages of the grasses, this study thus concentrates on examining the relationships of the pooid and bambusoid lineages. The remaining seven taxa (from four subfamilies) in the study form a monophyletic clade Davis and Soreng (1993) termed the PACC clade (Panicoideae - Arundinoideae - Centothecoideae - Chloridoideae), an acronym adopted here.

Only recently (following the simplification of the gene amplification sequencing methodologies) have other genes been sequenced in an attempt to find genes of suitable variability to assess subfamilial relationships within the grasses. Cummings et al. (1994) used sequence data from the grass specific insert in the plastid rpoC2 (RNA polymerase C2) gene to show relationships among 13 representatives from all the subfamilies, while Nardot et al. (1994) used sequence data from the plastid rps4 (ribosomal protein subunit 4) gene to elucidate relationships among 28 representatives of four subfamiles.

The aim of this thesis

The aim of this thesis is to use data from chloroplast DNA gene sequences to elucidate relationships among the genera and tribal lineages of the subfamily Arundinoideae. Sequence data from two chloroplast genes will be used to infer phylogenetic relationships. The variable grass-specific region within the rpoC2 gene, first used by Cummings et al. (1994) is used to show relationships between genera and tribes, and the more conserved rbcL gene is used to determine tribal and subfamilial relationships of the major lineages in the family.

Because the rpoC2 and rbcL sequence differ in degree of variability, stability, structure and alignment, this thesis presents results based on each of these sequences as separate chapters. The rpoC2 study is presented first, as results of this study are used to determine sampling strategy of the rbcL study, which appears as a subsequent chapter. The combination and analysis of these and other data sets is discussed in separate chapter. In all these chapters, details of the biology (morphology, anatomy etc) of the taxa sampled are not discussed. These are included in the last chapter, concerning the taxonomic implications of the various molecular phylogenies.

CHAPTER 2

PHYLOGENETIC ANALYSIS OF THE rpoC2 SEQUENCE DATA

INTRODUCTION

Recently Cummings *et al*. (1994) described a grass-specific, variable region inserted in the plastid *rpo*C2 gene, which codes for the B" subunit of the plastid RNA polymerase (Igloi *et al*. 1990). The insert has been conclusively shown to be absent in other members of the Poales (Cummings *et al*. 1994).

This insert comprises a number of repeat motifs, each generally coding for seven amino acids. These repeats (termed heptameric repeats by Igloi *et al*. 1990) comprise most of the insert, which ranges in size from 210 bp (*Anomochloa*, Bambusoideae) to 459 bp (*Streptochaeta*, Bambusoideae). The majority of grasses sequenced to date possess an insert of 417 or 438 bases.

Two types of sequence variation have been observed in this region; base substitutions (i.e. mutations) and insertion - deletion events (indels). Cummings *et al*. (1994) postulate that the major repeat structure of the insert arose as a result of slipped strand mispairing (SSM) events, as has been described in plastid genome sequences of other plant taxa (Blasko *et al*. 1988, Wolfson *et al*. 1991). Slipped strand mispairing has been postulated to be of major significance in the evolution of genes and genomes (Levinson and Gutman 1987). In addition to these plant examples, at least two other genes are known to contain repetitive elements: Djian and Green (1989) found that the (nuclear) gene coding for involucrin in hominids contains a series of 10-codon (30 base pair) repeats which are always based on one of two possible templates, and in plants, at least some of the nuclear coded LEA (Late Embryo Accumulating) genes contain a variable number of repeats 11 amino acids (33 base pairs) in length (Curry and Walker-Simmons 1993, Dure et al. 1989).

In their study, Cummings et al. (1994) designed oligonucleotide primers for use in PCR (Polymerase Chain Reaction) and sequenced 13 grass species from the six major subfamilies recognised by Clayton and Renvoize (1986). Results from this study, which was the first sequencing study to include all six subfamilies, indicated that, while the grass-specific insert was highly variable, the data was too homoplasious to determine relationships at the subfamily level. It was subsequently suggested that the region may serve as a useful tool in studies below the subfamilial level (Cummings et al. 1994), hence its application to systematic problems in the subfamily Arundinoideae. The use of this sequence for phylogenetic reconstruction is, however, associated with two problems; the choice of an outgroup and the alignment of the very variable sequences. The former issue is addressed below, while alignment methodologies are discussed later.

Choice of outgroup

The immediate sister family to the Poaceae, the Joinvilleaceae (Campbell and Kellogg 1987; Doyle et al. 1992; Kellogg and Linder in press) lacks the rpoC2 insert (Cummings et al. 1994). Thus only the relatively short regions flanking the insert, which are in common with the outgroup taxon, can be used to root the tree. Despite this drawback, analyses were carried out using Joinvillea as the outgroup. The rooting of the trees obtained from these analyses is thus based on very few characters, and may therefore be incorrect. The validity of the root of the tree may, however, be confirmed in the rbcL topologies presented in the next chapter.

Aims of the rpoC2 sequencing study

The rpoC2 sequence data are used to address the following issues:

1) What are the phylogenetic relationships among the genera and tribes of the subfamily Arundinoideae?

The relationships of 48 representatives of 33 "arundinoid" genera are assessed using a number of phylogenetic methods. The relationships of the major groups of these taxa with other subfamilies are also examined.

2) Is the rpoC2 insert suited for use in phylogenetic studies?

The link between a gene tree, based on the relationships between gene sequences, and a phylogeny of organisms rests on the assumptions that the sequences under comparison are orthologous, and that they are adequate representations of the variability of the taxonomic units under comparison. Excessive variation associated with gene sorting may result in a discrepancy between the gene and organism phylogeny. This issue thus questions the validity of using one sequence from a single specimen to represent a species, genus or higher category. While this assumption is generally accepted (although seldom tested) in studies using comparatively well understood and / or conserved genes (such as rbcL), the variability of rpoC2 sequence has not been investigated.

Testing variability (and thus suitability) of any gene used in a systematic study involves the sequencing of more than one sample of each taxon. This needs to be done in a manner that maximises the chances of finding such variability (a "worst case" scenario). This implies that a species containing as much variation as possible should be used for this purpose. Such species include those that are geographically widespread or morphologically variable.

In this study rpoC2 sequence variability is examined using two species of Phragmites, one of which is pandemic. Sequence data from P. australis (=P. communis), was obtained from four specimens; one each from southern Africa, Canada, Australia and Japan. An additional single sequence from P. karka (restricted to the orient and Australasia) is used to indicate sequence variability between species. If sequence variability is substantial within the pandemic species, then sequences taken from specimens from different continents would be expected to vary markedly and thus call into question the systematic value of the rpoC2 sequence. However, if the sequences are conserved across such broad geographic distributions, then the rpoC2 insert may be viewed as suitable for molecular systematic studies that use single specimens as exemplars.

As the hierarchical unit in the study on the Arundinoideae is generally the genus, the variability within this unit should be assessed by sampling more than one species from each genus. However, as the existing taxonomy of some genera is known or thought to be dubious, results suggesting that these "genera" are poly- or paraphyletic cannot be used to show demonstrate the unsuitability of the gene for systematic purposes. In the study using rpoC2 sequence data, several genera are represented by more than one species.

3) Can the slipped strand mispairing evolutionary events be tracked along a lineage, thus providing an indication of the sequence of slipped strand mispairing events during the evolution of the insert?

Given a phylogeny of a well sampled lineage based on nucleotide data, then the evolution of the insert by means of slipped strand mispairing (as represented by gaps in the aligned sequences) can be examined by plotting their distribution on the most parsimonious cladogram. This is done using the danthonioid lineage, for which there are 32 sequences available.

4) Can any further insight be obtained as to how the insert has originated and evolved in the grasses?

MATERIALS AND MOLECULAR METHODS

DNA extraction, amplification and sequencing

A total of 67 grass specimens were sequenced in this study, details of which appear in Table 2.1. Note that the generic concepts used in this study, and shown in this table, do not conform to those proposed by Clayton and Renvoize (1986). Particular areas of difference include the retention of the genera Clayton and Renvoize (1986) place into *Rytidosperma* and the retention of *Moliniopsis* (as opposed to its inclusion in *Molinia*, proposed by Clayton and Renvoize 1986)

Leaf material was usually dried in the field by means of silica gel, as described by Chase and Hills (1991). Fresh material was used when plants were available locally. Total DNA was extracted using the CTAB method of Doyle and Doyle (1987). The insert in the *rpo*C2 gene sequence was then amplified by means of the PCR method, using primers specific to regions flanking the grass-specific insert. These primers (and the corresponding primer sequence information) were obtained from E.A. Kellogg (Harvard University, Cambridge, MA). Additional quantities were subsequently manufactured locally by the Nucleotide Synthesis Unit of the Department of Biochemistry, University of Cape Town.

Primer *rpo*C2-1: 5′-CggAATTCTTTTACgTAgAAATACTA-3′
Primer *rpo*C2-2: 5′-CggTCgACTTgTTCCTCgATgCTCAA-3′

Although these primers worked for almost all the grass species sequenced in this study, a second set of primers was designed using the published rice and maize *rpo*C2 sequences (Igloi *et al.* 1990). These primers, called "*rpo*C2-U" (upstream) and "*rpo*C2-D" (downstream), flank the priming sites of the *rpo*C2-1 and *rpo*C2-2 primers, as illustrated in Figure 2.1.

Primer rpoC2-U: 5'-TTTgATTTgTTAgCgAAAAAgCg-3'
Primer rpoC2-D: 5'-ATCCACTCCAATAATACTATTgTC-3'

This second set of primers was used to amplify and sequence the outgroup *Joinvillea*, as well as the few grasses where sequence variation at one of the priming sites (usually the *rpo*C2-2 priming site) was sufficient to cause PCR failure when the original primers were used. The primers that were used in the amplification reactions were also used to sequence the amplified product.

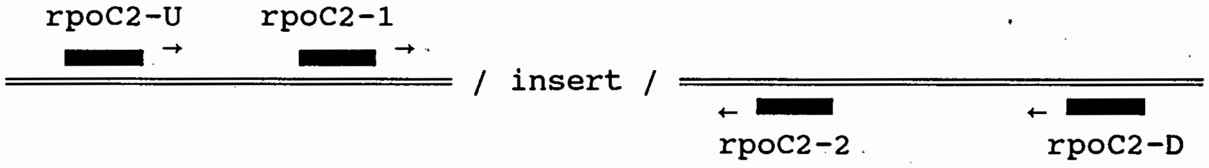


Figure 2.1. The layout of the primers used in the detection of the *rpo*C2 insert.

The amplification products were electrophoresed in a 1% agarose gel. The region of the gel containing the DNA was then excised using a scalpel, and the DNA purified by means of the QIAEX (QIAGEN) DNA purification kit. The purified fragments were sequenced using the Sequenase sequencing kit version 2 (United States Biochemicals), the manufacturer's protocol being modified according to Casanova *et al*. (1990) and Bachmann *et al*. (1990). Sequencing reactions were labelled with ^{35}S dCTP, and the products electrophoresed in a 6% denaturing polyacrylamide gel using a Hoeffer Poker Face II sequencing system. Gels were fixed for 15 minutes in methanol: acetic acid: glycerol: water (10:10:2:78), dried under vacuum, and detected autoradiographically.

Table 2.1. Species which were sequenced for the *rpo*C2 study, arranged according to subfamily and tribe *sensu* Clayton and Renvoize (1986). Note that the genera placed under *Rytidosperma* by Clayton and Renvoize (1986) are retained here, and their inclusion in *Rytidosperma* indicated in parentheses. Where possible, garden numbers are provided for those specimens taken from cultivation. DNA of the outgroup, *Joinvillea* (Joinvilleaceae) was kindly provided by W. Hahn (Smithsonian Institution, Washington).

Subfamily Tribe Species	Source of material and accession number
Bambusoideae	
Anomochloeae	
Anomochloa marantoidea Brongn.	DNA provided by L. Clark, Iowa State University (Clark 1299)
Bambusoideae	
Streptochaeteae	
Streptochaeta angustifolia T.S. Soderstrom	DNA provided by L. Clark, Iowa State University (Clark 1304)
Bambusoideae	
Bambuseae	
Bambusa vulgaris Schrad.	Cultivated; Durban Botanic Gardens
Bambusoideae	
Olyreae	
Olyra latifolia L.	Ngoye Forest, Kwazulu-Natal (Linder 5742)
Pooideae	
Stipeae	
Stipa dregeana Steud. var. dregeana	Constantia, Cape Town, South Africa McDowell s.n.

Centothecoideae	
Centotheceae	
Chasmanthium latifolium (Michx.) Yates	Cultivated; Missouri Botanical Gardens, St. Louis, USA (Snow, 5944)
Arundinoideae	
Aristideae	
Aristida congesta Roem. and Schult. subsp. barbicollis (Trin. and Rupr.) De Winter	Pretoria Botanic Gardens, South Africa (Barker 1130)
Stipagrostis zeyheri (Nees) De Winter subsp. zeyheri	Cape Point Nature Reserve, South Africa (Barker 1133)
Arundinoideae	
Arundineae	
Amphipogon strictus R. Br.	Kings Tableland, New South Wales, Australia (Linder 5634)
Anisopogon avenaceus R. Br.	Tiantara Falls, New South Wales, Australia (Linder 5590)
Arundo donax L.	University of Cape Town campus, South Africa (Barker 1131)
Centropodia glauca (Nees) T.A. Cope	Alexander Bay, Northern Cape, South Africa (Linder 5410)
Chaetobromus dregeanus Nees	Leliefontein, Kamiesberg, South Africa (Barker 978)
Chionochloa macra Zotov	Mt Cheeseman, Cantebury, New Zealand (CHR 475278)
Cortaderia fulvida (Buchan.) Zotov	Lincoln, New Zealand (in cult.; Garden No. 5088)
Cortaderia selloana (Schult.) Asch. & Graeb.	Roodepoort, Johannesburg (Robinson, s.n.)
Danthonia pallida R. Br.	Black Mtn, Canberra, Australia (Linder 5564)
Danthonia secundiflora J. & C. Presl	South America (Lyle 1617)
Danthonia spicata Roem. and Schult.	Mt Desert Island, Maine, USA. (Kellogg, s.n.)

Dregeochloa pumilla (Nees) Conert	Alexander Bay, Northern Cape, South Africa (Linder 5408)
Erythranthera pumila (Kirk) Zotov (=Rytidosperma)	Mt. Somers, New Zealand (Linder 5747)
Gynerium sagittatum (Aubl.) P. Beauv.	Kew Gardens (originally from Peru; Garden No. 1991-1276 Kall)
Hakonechloa macra (Munro) Makino	Utsunomiya University campus, Japan (Kobayashi s.n.)
Karroochloa purpurea (L.f.) Conert and Tuerpe (=Rytidosperma)	Botterkloof, SW Cape, South Africa (Linder 5360)
Lamprothyrsus peruvianus Hitchc.	Lincoln, New Zealand (in cult. Garden No. 11154)
Merxmuellera arundinacea (Berg.) Conert	Groot Swartberg, South Africa (Barker 1017)
Merxmuellera cincta (Nees) Conert	Cape Town, South Africa (Barker 1160)
Merxmuellera davyi (C.E. Hubb.) Conert	Mt. Mulanje, Malawi (Barker 942)
Merxmuellera disticha (Nees) Conert	Niew Bethesda, E. Cape, South Africa (Barker 1002)
Merxmuellera dura (Stapf) Conert	Nieuwoudtville, S.W. Cape, South Africa (Barker 983)
Merxmuellera guillarmodae (Conert)	Drakensberg mountains, South Africa (Barker 1009)
Merxmuellera macowanii (Stapf) Conert	Drakensberg mountains, South Africa (Barker 1008)
Merxmuellera rangei (Pilg.) Conert	Aus, Namibia (Barker 960)
Merxmuellera rufa (Nees) Conert	Bainskloof, SW Cape, South Africa (Barker 1149)
Merxmuellera setacea N.P. Barker	Ceres, S.W. Cape, South Africa (Barker 987)
Merxmuellera stricta (Schrad.) Conert	Table Mountain, Cape Town, South Africa (Barker 1159)

Moliniopsis japonica (Hack.) Hayata	Utsunomiya University campus, Japan (Kobayashi 1253)
Monachather paradoxus Steud.	Eulo, Queensland, Australia (Moll 1)
Notochloe microdon Domin	Wentworth Falls, New South Wales, Australia (Linder 5633)
Pentameris thuarii Beauv.	Tradouw's Pass, Swellendam, South Africa (Linder 5456)
Pentaschistis aspera (Thunb.) Stapf	Cape Town, South Africa (Barker 1164)
Pentaschistis curvifolia (Schrad.) Stapf	Cape Town, South Africa (Barker 1165)
Phragmites australis (Cav.) Steud.	Black River, Cape Town, South Africa (Barker 1132)
Phragmites australis (Cav.) Steud.	Utsunomiya University campus, Japan (Kobayashi 1256)
Phragmites australis (Cav.) Steud.	Canberra, Australia (in cult., Watson s.n.)
Phragmites australis (Cav.) Steud.	Canada (Darbyshire s.n.)
Phragmites karka Trin. ex Steud.	Utsunomiya University campus, Japan (Kobayashi 1254)
Plinthanthesis paradoxa (R. Br.) S.T. Blake	Wollongong, NSW, Australia (Linder 5638)
Prionanthium ecklonii (Nees) Stapf	Clanwilliam, SW Cape, South Africa (Linder 5402)
Pseudopentameris macrantha (Schrad.) Conert	De Hoop, S.W. Cape, South Africa (Linder 5470)
Rytidosperma laeve (Vickery) Connor & Edgar	Canberra, Australia (Watson s.n.)
Rytidosperma nudiflorum (Morris) Connor & Edgar	Cradle Mt., Tasmania (Linder 5693)
Schismus barbartus (Loefl. ex L.) Thell.	Botterkloof, SW Cape, South Africa (Linder 5359)
Spartochloa scirpoidea (Steud.) C.E. Hubbard	Australia (Craven 8892)
Tribolium uniolae (L. f.) Renvoize	Cape Town, South Africa (Barker 1163)
Urochlaena pusilla Nees	Clanwilliam, SW Cape, South Africa (Linder 5402)

Arundinoideae Micraireae Micraira lazeridesii Clark, Wendel & Craven (ined)	DNA provided by L. Clark, Iowa State University (Clark 1157)
Arundinoideae Thysanolaeneae Thysanolaena maxima Kuntze	Royal Botanic Gardens, Kew (originally from Barbados; Garden No. 1979-3225 Warr)
Chloridoideae Cynodonteae Tragus berteronianus Schult.	Pretoria Botanic Gardens, South Africa (Barker 1128)
Chloridoideae Eragrostideae Eragrostis capensis (Thunb.) Trin.	Kenilworth Race Course, Cape Town, South Africa (Barker 1135)
Chloridoideae Pappophoreae Enneapogon scaber Lehm. var. scaber	Leeuw Gamka, SW Cape, South Africa (Barker 1023)

Panicoideae	
Andropogoneae	
Hyparrhenia hirta (L.) Stapf	Kirstenbosch Botanic Gardens, Cape Town, South Africa (Barker 1134)
Saccharum officinarum L.	Johannesburg, South Africa (in private cultivation)
Panicoideae	
Paniceae	
Brachiaria serrata (Thunb.) Stapf	Pretoria Botanic Gardens, South Africa (Barker 1124)
Rhynchelytrum (Willd.) C.E. Hubb. subsp. repens	Pretoria Botanic Gardens, South Africa (Barker 1129)
Panicum maximum Jacq.	Pretoria Botanic Gardens, South Africa (Barker 1125)
Panicoideae	
Arundinelleae	
Tristachya biseriata Stapf	Pretoria Botanic Gardens, National Botanical Institute, South Africa (Barker 1126)

Sequence manipulation and alignment

The sequences were entered into the program DAPSA (DNA And Protein Sequence Analysis) written by Prof. Eric H. Harley (Dept. Chemical Pathology, University of Cape Town). Previously published rpoC2 sequences (Cummings et al. 1994) were obtained from M. Cummings, or GENBANK (the Sorghum sequence).

Alignment methods used in this study

Owing to the controversial nature of alignment methods as well as the variability of the data obtained from the rpoC2 insert, as broad a spectrum as possible of alignment methods is used to obtain alignments of the data. There are three possible approaches that can adopted:

1) Computerised alignment methods. Several programs are available, including DAPSA, CLUSTAL V and MALIGN.

2) Intuitive alignment methods, including manual alignment methods, or alignments done "by eye", a precedent for which has already been set (see Table 2.2 and discussion below).

3) Develop a method based on testing character homology.

Computerised alignment methods

The links between "relationships" and alignment are exploited by a number of computerised clustering methods which construct a distance based Fitch-Margoliash tree prior to alignment, and then align the sequences beginning with those that are, according to the tree, most closely related. This alignment procedure is often based on an algorithm developed by Needleman and Wunsch (1970). This algorithm adds insertions (gaps) according to a weighting or penalty scheme. This approach has subsequently been used in various formats (for example Hein, 1989a, 1989b, Higgins and Sharp 1988). This method, however, suffers from four weaknesses.

1) The initial step is based on a phenetic tree construction algorithm which uses a concept of (overall) similarity; a phenetic concept. Thus, in terms of hennigian phylogenetics, the distance-based alignment procedures are flawed. The use of a parsimony-based method of tree construction in place of a phenetic one is, however, not possible, as a tree constructed using parsimony relies, a priori, on correct (positional) homology assessments, which implies a knowledge of phylogeny.

2) The sampling strategy (sample size, taxon density etc.) will affect the initial relationships displayed in the distance-based dendrogram. More taxa may result in different alignments, not because of modifications to any homologies, but because the phenetic algorithms (such as UPGMA) are sensitive to such factors as "outliers" or excessive numbers of very similar sequences (Sneath and Sokal 1973).

3) These methods are dependant on the order in which the taxa are considered for alignment. While this order may be dependant on the initial calculation of phenetic similarity, some methods require the taxa to be entered in a sequence approximating phylogenetic relatedness. However, this may not be known, and the elucidation of a phylogenetic sequence is, after all, the very aim of the study in the first place.

4) It is questionable whether it will effectively work on very divergent sequence data. This is evidenced to some extent by the "throw away" attitude toward "unalignable" data, and the final dependence on alignments "by eye" that is carried out after the initial computerised alignments are obtained.

Wheeler and Gladstein (1993) point out that all the computerised alignment procedures utilising the above approach fail to realise that alignment is merely a way to examine putative homology statements. They further state that there is

> "... no better way to test homology than with parsimony. Only an alignment procedure footed firmly in cladistic parsimony can rigorously test alternative putative homology schemes." (Page 5).

Their alignment program (MALIGN) generates alignments using parsimony in the sense that the

> "...minimum number of steps or changes required by an alignment is of course the most parsimonious branching diagram for these sequences." (Page 3).

They further state that

> "The best alignment is that which yields the most parsimonious cladogram." (Page 3).

While the efforts of Wheeler and Gladstein in introducing an alignment logic rooted in cladistic principles are to be applauded, the principles of their method (or at least their explanation of them) may be questioned. The issue as to whether parsimony itself is a test of homology is debatable. According to Patterson (1982 - see later discussions), there are only three tests of homology; similarity, conjunction and congruence. Parsimony as a criterion of cladogram generation is not a test of homology. Homology may, however, be tested on a most parsimonious cladogram. This may in fact be what is implied by their statement, but it is not at all clear.

The alignment method of Wheeler and Gladstein (1993) inserts gaps in such a way that cost is minimised, and it is left to the user to choose a cost value for gaps (similar to the gap penalty in CLUSTAL). But apart from the cost of adding a gap, it needs to be borne in mind that a gap is not real - it is a reflection of a past mutational event, which can be phylogenetically informative. Gaps should thus be added in a manner consistent with a parsimonious phylogeny. In order to accommodate and utilise this, Wheeler and Gladstein (1993) consider gaps to be a fifth state (after A, C, T & G). The cost of transformation between a gap and these states is determined by the user. However, in light of the evolutionary importance of gaps in a sequence it seems to be somewhat misleading to associate a "cost" or "penalty" with their use.

In using computerised alignment methods, three programs were considered: DAPSA, CLUSTAL V and MALIGN.

DAPSA

DAPSA (DNA and Protein Sequence Analysis), written by Prof. Eric H. Harley (Department of Chemical Pathology, University of Cape Town), aligns sequences in a pairwise manner. In doing so, the user is able to select a "stringency" value which determines the degree of similarity required for the recognition of two regions of sequence as identical, or "homologous".

In its present format, DAPSA unfortunately suffers from two weakness that rule it out for use in this study:

1) It requires that the sequences be entered in the order of most to least closely related, although this requirement is not absolute. This requirement implies an a priori knowledge of the relationships of the taxa, which are not known.

2) It is unable to operate within the confines of the conserved flanking regions, ultimately homologised by the PCR primers. This means that, instead of keeping the beginning and end regions aligned while inserting gaps in the middle of the region, the program starts by inserting gaps as it comes across the regions it identifies as requiring them, and the ends of the sequences (which are more conserved) can end up being completely unaligned. In so doing, the individual sequences end up a variety of lengths. For these reasons, particularly the latter, this program is not used in this study to align sequences; merely to manipulate them manually.

CLUSTAL V

This program, written by Higgins and Sharp (1988), is based on the Needleman-Wunsch algorithm, and can carry out multiple alignments. This program was used by Cummings et al. (1994) in their initial analysis of 13 rpoC2 sequences. It allows the user to set a penalty for gaps (insertions / deletion events) and transversions. While frequently used in smaller studies, the PC-based version of this program can only accommodate 30 taxa. As sequence data for more than double that number of taxa are used here, these sequences were aligned using CLUSTAL on a SUN workstation by Dr. E.A. Kellogg (Harvard University, Boston). The gap weighting used by Cummings et al. (1994) was initially used (a value of 35), but values of 25 and 45 were also employed.

MALIGN

MALIGN (Wheeler and Gladstein 1993, discussed above) is the only alignment package available that aligns sequences in a method based on parsimony. Unfortunately, it is very memory-intensive, the amount of memory required being dependant on the length and number of sequences to be aligned. When attempted, no more than 10 sequences could be aligned using this program, and it therefore could not be used in any large analyses of the rpoC2 data set.

Intuitive alignment methods

Intuitive methods vary in degree of subjectivity, with the more objective methods employing rules to govern the insertion of gaps (for example Golenberg et al. 1993, discussed above). In the intuitive method developed specifically for the rpoC2 sequence data (the successive stringency alignment method), rules governing where gaps may be inserted have been drawn up, and the logic governing the recognition of regions which need re-alignment is rooted in the cladistic principle of homology testing by congruence (character consistency). In this way the homology of the repeat units and nucleotides is tested.

In this study, alignment was done using the more conserved amino acid phenotype, a method considered by Li and Graur (1991) to be more reliable than using DNA sequence data. In this process, gaps to account for similarities of the repeats were placed using three rules:

1) Insertions corresponding to complete heptameric repeats were used when ever possible.

2) Heptameric repeats are considered to start with the tyrosine code ("Y", codon = TAT or TAC)

3) Complete heptameric repeats were not unnecessarily interrupted with insertions.

Once the amino acid sequences had been aligned in this manner, the DNA sequences were then aligned to reflect the amino acid alignment.

These rules are based on assumptions about the evolutionary dynamics of the gene. They are artificial in the sense that they do not necessarily reflect biological processes; the inserts do not have to begin with TAT, and repeats may in fact be interrupted. However, without these rules, no consistent alignment would be possible. The alignment is thus objective in the sense that it is rule governed, but the rules themselves are subjective.

Alignments based on tests of homology

Phylogenetic analysis of DNA sequences involves their alignment in such a way that homologous nucleotide base positions within homologous genes (orthologues) are compared (Mindell 1991). Alignment is thus the first and most crucial analytical step in any molecular phylogenetic study, and is essentially a problem of assessing positional homology (Miyamoto and Cracraft 1991; Swofford and Olsen 1990). Many molecular systematic studies rely on the conservative nature of the genes used to ensure that alignment procedures are unproblematic. Such sequences have been used in systematic studies across a broad spectrum of organisms. However, excessively variable genes, or variable regions within genes, result in sequence alignment problems, and are thus not favoured for use in systematics.

The relationship between alignment and phylogeny is one akin to the "chicken and egg" scenario. Alignment processes (on orthologous genes for systematic purposes, or paralogous genes for gene evolution studies) cannot be independent of phylogeny, as the very acceptance of a notion of orthology or paralogy implies an understanding of ancestry. Molecular systematic studies utilising the PCR method inherently have the concept of homology as their foundation; primers are used to recognise and amplify "specific" genes.

A cladogram derived from aligned sequence data may carry with it a bias based on the interpretations of "phylogeny" / similarity made during the alignment process. The inability to divorce the alignment procedure from the assessments of phylogeny or similarity may result in "incorrect" trees (i.e. those not reflecting true ancestor - descendant relationships, or "reality"). This would occur if the initial assumptions of organismal phylogeny or relatedness (used, for example, to determine the order in which sequences are aligned) were incorrect. However, there are no known instances where this has obviously occurred. Indeed, it would be difficult to prove that a result is erroneous purely because of an a priori bias; methodological problems such as sample size and taxon density would obscure this source of error.

Very often, alignment issues are either ignored, or rules are made up as needed. This is apparent in a survey of alignment methodologies discussed in recent botanical molecular systematic papers. This survey, summarised in Table 2.2, reveals four approaches to sequence alignment:

1) Alignment is not mentioned or discussed at all. In many of these cases the gene involved is highly conserved, and it is likely that no alignment problems are encountered.

2) Alignment methods are mentioned or discussed, but no alignment problems are found.

3) Alignment methods are discussed, alignment problems are present, and data are excluded because of uncertainties in the alignment process or results. These instances generally occur in studies using more variable genes, such as nuclear ribosomal ITS (Internal Transcribed Spacer) sequences. However, alignment problems in both this and the next category may also be exacerbated by insufficient sampling.

4) Alignment methods are discussed, problems are detected, and attempts are made to examine alternative solutions without sacrificing data. In these instances, the most frequently used method was to code the indels as binary characters and add them to the nucleotide data set, which was first analysed without the indel characters. Some of these studies did this instead of including nucleotide data from the indel regions, while others retained the nucleotide data from these regions with "absent" data suitably coded. When the indel characters are added to data matrices obtained from the latter method, the indel regions would receive additional weighting in a parsimony analyses.

Table 2.2. A survey of recent studies presenting phylogenetic analyses based on sequence data. The studies are categorised according to the approach taken toward sequence alignment, and where relevant the approaches adopted toward resolving alignment problems are summarised. Unless otherwise stated, all genes are from the chloroplast.

Approach to alignment	Reference	Gene	Alignment method	Solution to problems
1) Not discussed	Les et al. (1991)	rbcL		
	Brunsfeld et al. (1994)	rbcL		
	Soltis et al. (1993)	rbcL + 5'region		
	Soltis et al. (1990)	rbcL		
	Bremer et al. (1994)	rbcL		
	Swensen and Mullin (1994)	rbcL		
2) Discussed; no problems	Nadot et al. (1994)	rps4	CLUSTAL	
	Williams et al. (1994)	rbcL	Manually (on amino acid sequences)	
	Conti et al. (1993)	rbcL	Manually, using a reference sequence	
	Clark et al. (1993)	rbcL	Manually	

Approach to alignment	Reference	Gene	Alignment method	Solution to problems
3) Discussed; problems result in data exclusion	Olmstead <u>et al</u>. (1992)	<u>rbc</u>L	Manually by comparison to reference sequence	Excluded positions of uncertain homology
	Baldwin (1992)	nDNA (ITS)	Manually by sequential pairwise comparison	Excluded sites of ambiguous alignment.
	Wilcox <u>et al</u>. (1993)	nDNA (18S rRNA)	By eye using program ESEE	Hypervariable regions excluded
	Suh <u>et al</u>. (1993)	nDNA (ITS)	By eye using LINEUP	Indel regions excluded, but were superimposed on tree
	Baldwin (1993)	nDNA (ITS)	Manually by sequential pairwise comparison	Excluded gaps
	Savard <u>et al</u>. (1993)	Various	ITS sequences aligned using a computer	Gaps (in ITS sequences) were excluded
	Gatesy <u>et al</u>. (1993)	mtDNA (12S & 16S rDNA)	Used MALIGN to generate alignments with varying gap penalties	Removed alignment-ambiguous regions
	Rettig <u>et al</u>. (1992)	<u>rbc</u>L	Not discussed	Excluded variable regions
	Giannasi <u>et al</u>. (1992)	<u>rbc</u>L	Not discussed	Excluded variable regions

Approach to alignment	Reference	Gene	Alignment method	Solution to problems
4) Discussed; problems addressed	Golenberg _et al_. (1993)	Noncoding cpDNA	Manually following preset criteria	Analysed indels separately from nucleotide data
	Bult and Zimmer (1993)	nRNA (18S and 26S)	Various programs; by eye	Indels treated as binary characters
	Steele and Vilgalys (1994)	_matK_	Manually	Indel was coded as a binary character in data set
	Wojciechowski _et al_. (1993)	nDNA (ITS)	Manually by sequential pairwise comparisons	Alignment variable regions excluded; various coding methods for remaining indels in some analyses
	Waters _et al_. (1992)	nRNA (18S and 26S)	Computerised; adjusted by eye	Variable regions excluded, but indels coded as binary characters
	Johnson and Soltis (1994)	_matK_	Manually using amino acids	Indels coded as binary characters and added to nucleotide data set

The latter two categories are pertinent to the discussion concerning alignment of the variable rpoC2 sequences. Studies in both categories are based on more variable sequences, a response resulting from the increasing application of sequencing methods to more hierarchically limited studies (inter and intra generic studies) as opposed to "big picture" problems addressed using more conserved genes. In using variable sequences, one of two actions follow the occurrence of alignment problems: discard data or investigate the options. The former is clearly an unsatisfactory solution, for the reasons discussed below, while the latter has been limited to studying the effect of adding binary data coding for insertion / deletion (indel) events to the nucleotide data sets.

Swofford and Olsen (1990) advise that

> "When regions of the sequences are so divergent that a reasonable alignment cannot be obtained by manual methods using a sequence editor ('by eye'), those regions should probably be eliminated from the analysis." (Page 417).

This procedure appears to have been followed in many studies where alignment problems have been encountered. However, data exclusion on the basis of excess variability has been criticised for a number of reasons. Gatesy et al. (1993) consider data exclusion to be an extreme form of character weighting - variable regions being given a weight of zero. Furthermore, the decision to exclude data is often made subjectively. If data has to be excluded, Gatesy et al. (1993) propose the use of an objective, repeatable method to select the relevant nucleotide positions. They offer a solution based on the consistency of alignment of nucleotide positions over a variety of alignment parameters, and only then exclude alignment-variable (also termed alignment ambiguous) regions from further phylogenetic analysis.

In a study using variable sequence data, Golenberg et al. (1993) addressed alignment problems by applying a set of predefined criteria in the alignment process. This approach is favoured by Hillis (1994) who maintains that inferences of positional homology can only be robust if alignments are rule-driven.

Penny et al. (1990), in a discussion on problems associated with sequence data, appear not to consider data exclusion to be a result of alignment problems, and it does not feature in the section concerning alignment problems in their discussion. They do, however, view data exclusion as a property of the various methods of tree reconstruction. Within this context, the discarding of information such as insertions and deletions is seen as a methodological problem resulting in incomplete use of information. However, elsewhere they state that it is necessary to consider trees and alignments simultaneously (Penny et al. 1990 p. 23). This apparent confusion concerning the role of alignments, tree building and data exclusion summarises the difficulties inherent in using variable molecular data.

It is thus easy to see why sequences (or regions within them) which need numerous insertion or deletion events to align (i.e. to explain positional homologies) are not ideal for use in phylogeny generation; the positioning of the gaps is problematic and (if coded for) many gap-characters will be homoplasious, but without these gaps, the nucleotide data will be homoplasious. As outlined by Siebert (1992), character conflict (i.e. homoplasies) can be dealt with in one of four ways:

1) Re-evaluate the characters.
2) Weight the characters.
3) Subject the characters to a transformation series analysis.
4) Use a consensus tree.

However, for sequence data some of these options are invalid or inappropriate. Assuming the sequences in question are accurate, characters (= nucleotides) cannot be re-assessed. They are always one of A, C, T or G. As such, no direction can be placed on these states, and so transformation series are inappropriate. The weighting of nucleotide data appears inappropriate, unless it takes the form of (for example) weighting tranversions over transitions, or the first and second bases of codons over the third base. The use of a consensus tree is acceptable, but will result in a loss of resolution.

None of the above proposals can thus contribute to the improved resolution of the cladogram, unless character weighting is invoked. What is not explicitly mentioned in Siebert's (1992) proposals (which are not specific for any one kind of data) but which might be viewed as a special case of the first proposal, is that character re-evaluation can be carried out <u>in the context of sequence alignments</u>.

Characters (nucleotides) should be viewed as "low level hypotheses" (Neff 1986). The accuracy of any cladogram, and that of these hypotheses is affected by the accuracy of the alignment. By the process of reciprocal illumination (Hennig 1966), the initial alignment hypothesis based on crude similarity can be refined by looking for homoplasy on the initial cladogram, and then refining the alignments. In this manner, these low level hypotheses can be effectively tested.

The transformation of "bad" sequences (very variable and of questionable value for use in molecular systematics) into "good" (well resolved) phylogenies is best made possible by the implementation of an alignment method based on the principle of hypothesis testing. The hypothesis to be tested is one of positional homology. In the instance of <u>rpo</u>C2 sequence data, the concept of positional homology is not restricted to just the nucleotide level. The unusual nature of the sequence (heptameric repeats) and the mutational events that produced it (slipped strand mispairing) imply that the homology of these repeats as a whole also needs to be examined and tested. The repeats are thus also low-level hypotheses.

Patterson (1982) proposed three tests of homology: similarity, conjunction and congruence.

Similarity: As a test for homology, the test of similarity is impossible to fail. As noted by Stevens (1984), without similarity, homology would not be dreamed of. In the rpoC2 gene, all the repeats comprise a general pattern (phenotype) dictated by functional requirements of the gene (Igloi et al. 1990) and are thus similar. Homology assessment of repeats by means of similarity is further complicated by the possibility of two independent origins for similar or identical repeats in a certain position, or by point mutations in the repeat sequences, which can result in the (false) impression that two orthologous heptameric repeats are paralagous. Furthermore, the fact that two consecutive repeats (within a sequence) may be identical as a result of slipped strand mispairing duplication events (or two events based on the same template) makes the choice of which of the two copies should be orthologous impossible.

Conjunction: The above scenario of two identical repeats at (either or both) amino acid and DNA level within a single sequence means that the conjunction test is failed. The implications of this are discussed below.

Congruence: In this test, homologies specify groups that are rendered monophyletic by them (Patterson 1982, 1988). As stated by Patterson (1988), a true homology will circumscribe a group that is congruent with those specified by other homologies. This equates to what Wagner (1986) calls the "criterion of coincidence", or testing by consistency.

De Pinna (1991) points out that all similarities are initially deemed homologous, until non-homology is indicated by a pattern detecting procedure. In this context, De Pinna recognises two interdependent and complementary stages in homology propositions. The first, or primary homology, is conjectural and based on similarity, while the secondary homology is the outcome of the pattern detecting analysis, and tested by congruence (a test based on additional criteria). Parsimony is shown by De Pinna (1991) to be the best method to maximise the number of statements of secondary homology.

Homology concepts in the context of the rpoC2 heptameric repeat sequences

The repeat structure within the rpoC2 sequence suggests a series of paralogous sub-sequences; each heptameric repeat is paralogous to its template, and ultimately a paralogue of the original source sequence. It must be noted that these are not true paralogues as defined by Fitch (1970) and used elsewhere (Patterson 1988, Mindell 1991, Li and Graur 1991, Hillis 1994, Nelson 1994, Williams 1992). Fitch's definition considers only whole duplicated genes to be paralogues. Thus in comparing the plastid rpoC2 genes from two taxa the gene is orthologous, but contains an insert consisting of repeats, some of which will be "homologous" and some "paralogous". To avoid calling these repeats "paralogues" and "orthologues" (which should be retained in their original, whole-gene, context), the terms "endo-paralogues" and "endo-orthologues" are used here instead. As with whole-gene orthologues, once two repeats are considered to be endo-orthologues (i.e. the positional homology of the repeats is accepted), the positional homology of the individual nucleotide positions within the repeats is (in the absences of gaps) assumed.

The repeat structure of the insert both helps and hinders the alignment process. Although the gaps corresponding to indel events will be a particular size, the probability that SSM events can occur anywhere in the sequence means that the assessment of homology of the repeats is not always easy, but can be done using the test of congruence (testing by consistency) to determine positional homologies.

In the alignment method developed and used here, character consistency is used to assist in aligning the heptameric repeats. In the context of De Pinna's (1991) homology concepts, the initial alignment, obtained using rules defined below, sets up a series of primary homologies; statements about the positional homologies of the repeats. Once a parsimonious cladogram has been obtained, statements of secondary homology can be tested. The aim of the alignment method is thus to align the repeat units in such a way that homologous repeat units (those having arisen from a common ancestor) are recognised. Once the repeats are correctly aligned, the nucleotides in the sequence will also be correctly aligned.

Within the rpoC2 sequences, there are three levels, or "degrees of coarseness", at which testing by congruence can be applied.

1. The lowest (finest) level is that of the nucleotide sequence data. This data set is the one from which the cladograms are (in this study) always determined, and it is nucleotide sequence data that is the desired end product of any sequencing study. However, the cladogram is only as correct as the positional homologies deduced by the alignment process.

2. The amino acid sequence data is the middle level at which congruence can be applied. Because of the multiple codon usage for many amino acids in the genetic code, variability at the third base in a codon is likely to be homoplasious, particularly when distantly related taxa are being studied. The use of the amino acid sequence data thus excludes this

source of noise, and also reflects any functional constraints that the sequence may be under.

3. The highest (coarsest) level at which congruence may be applied is at the indel level. The nature of the SSM mechanism of gene evolution implies that homoplasious indel data is likely to be rare; the independent origin of two identical repeats in an identical position is unlikely. Their subsequent independent loss is, however, possible.

As an indication of character congruence, character consistency is examined at each of the three levels mentioned above by using either the basic units of each level (nucleotides, amino acids and repeats), or using only nucleotides and examining them within the context of the relevant units. The "value" of phylogenetically informative characters obtained from an alignment is quantifiable by means of two cladistic information measures; the character consistency index (cCI) and character retention index (CRI). It would be expected that mis-aligned (i.e. non-homologous) repeats would have characters with either low CCI values or high CRI values, or both. The reasons for this are as follows:

1) Characters (nucleotide sites) that have state/s that are unique to a particular clade, but which are misaligned, will have fortuitously high CRI values. This would indicate that, while the characters are phylogenetically informative, they may not have been derived from a positional homologue, and that the alignment needs to be modified such that positional homologies are recognised.

2) Characters (nucleotide sites) with states which show reversals or convergences and / or which are saturated will have low CCI values. These homoplasies are likely to be reduced with alignment modifications, which may then result in higher values.

These two indices may be combined, and Farris (1988) proposed the use of the rescaled consistency index (RCI). The RCI of a character is the cCI x cRI, and can achieve a value of zero (this would occur should the cRI be zero). Farris (1988) utilises the RCI in his successive weighting approach to character weighting, and the RCI of each character (i.e. its "weight") is an indication of its performance or compatibility in the cladogram. In other words, characters (or their states) that are consistent with the tree topology derived from the data are to be retained at the expense of those that are not consistent; the homoplasies.

The preservation of these consistent (clade-specific) characters is done through alignment modifications in which insertions are made such that the relevant repeats from which the characters are derived are isolated. However, as will become evident in the discussion, the isolation of these repeats by re-alignment may mean that the character (nucleotide site) becomes invariant and is lost from the data set of informative sites. For this reason, a second data set was constructed in which these (potentially) lost data were replaced by a character coding for the presence of complete repeats and gaps.

These alignment solutions depend on the strictness of the homology concept that was applied in the recognition of the heptameric units as endo-orthologues. In other words, the amount and numerical value of what was considered to be a low cCI value, high cRI or low RCI value, as well as the distribution of these characters within the repeat units dictates whether or not a repeat is considered to be orthologous.

In this way, new primary homologies (De Pinna 1991) are erected (by re-aligning the sequences) and need to be tested again before they can become statements of secondary homology. Owing to the re-iterative nature of the method, it has been called the Successive Stringency Alignment (SSA) method.

The Successive Stringency Alignment method

As a starting point, the alignment obtained using the intuitive method was used to obtain phylogenetically informative sites. From these data, a cladogram was obtained using the m*bb* options of HENNIG86 (Farris 1988). In all analyses the number of trees found was limited by computer memory. The strict consensus tree was then calculated from the trees obtained. CLADOS (Nixon 1992) was then used to step through the data set, and the distribution of the states of each character on the consensus tree was examined. The consistency and retention index of each character was noted (cCI and cRI respectively), and those characters that had low cCI values (<= 33) and high cRI (>= 90) values were marked on a printout of the aligned sequences. In addition, the rescaled consistency index (RCI; Farris 1988, 1989) of each character was obtained by means of a single successive weighting iteration (xsteps w in HENNIG86) based on all available trees derived from the m*bb* routines in HENNIG86. Those characters with an RCI of zero were also marked on the printout. Characters obtained from all the three levels of coarseness (discussed above) could thus be assessed for their consistency and cladistic informativeness.

Those heptameric repeats which contained several characters with high cRI or low cCI were noted. These repeats were considered to be potentially non-homologous, and the alignment subsequently modified to accommodate them as separate entities. The realignment of the these repeats affects the cCI and cRI values of the original characters, and may introduce new characters with good cCI and cRI values. Thus the re-alignment contributes to increasing the ensemble CI and RI.

In this way, the sequence alignment is modified by maximising phylogenetic information (i.e. increasing the ensemble ri) and reducing noise (homoplasy) by maximising the ensemble ci. This is done here in a number of steps of increasing alignment "stringency". In each of these steps, a few (no more than four) heptameric repeats are recognised as non-homologous, and are re-aligned by means of inserting gaps in the region where they were originally aligned, and moving them to a position such that they form a block on their own. The insertion and moving of repeat blocks was done using DAPSA.

The process of character examination is reiterated using the informative sites from the new alignment, regions (repeats) of high homoplasy identified, and the alignment further expanded. This results in an increasingly stringent alignment as character homologies (i.e. repeat homologies examined on the basis of constituent nucleotide data) are more carefully assessed. The allocation of numbers to the alignment stringencies refers merely to the order in which alignments were produced, and is not correlated in any way to numerical factors such as the number of additional gaps or informative sites.

Phylogenetic analysis.

Two approaches were used in the analysis of sequence data: parsimony-based methods and distance-based methods. Parsimony-based methods utilise phylogenetically informative substitutions at variable sites in the aligned sequences, and produce a most parsimonious tree from these data. Distance-based methods use an estimate of sequence similarity based on the entire sequence (not just the variable sites) and a tree is constructed according to one (or more of) a number of principles and algorithms. Both these approaches to phylogeny reconstruction were utilised. DAPSA was used to output files of a suitable format for the phylogenetic analyses.

Parsimony methods:

Parsimony analysis was carried out on the computer-derived (CLUSTAL) alignments, the intuitive alignment and alignments at subsequent stringencies, as well as on the data sets to which characters coding for gaps had been added.

In performing a parsimony analysis, particularly on large data sets, one is confronted by three problems:

1) finding the shortest tree
2) finding all the shortest trees in an island of most-parsimonious trees
3) finding alternative islands of most-parsimonious trees

For the purposes of the routine analyses (i.e. addressing the first two problems listed above), the program HENNIG86 (Farris 1988) was used. Platnick (1989) tested a number of phylogenetic analysis programs and found HENNIG86 to be the most effective and efficient tool for systematic studies. However, subsequent to this comparison, more recent versions of the various software packages have appeared, and this distinction may no longer be valid.

In order to obtain as many parsimonious trees as possible in the routine analyses, the m* and bb* options of HENNIG86 were used. The m* option constructs several trees (each by a single pass, adding the terminals is different sequences) and then applies branch swopping to each of these trees. The bb* option, executed after the m* option, uses all available memory space (if required) to store the shortest trees found after extended branch swopping of the set of trees obtained from the analysis (Farris 1988).

In order to address the third problem (searching for as many islands of most-parsimonious trees as possible), different input sequences must be used to provide as many different starting points as possible in the search for shortest trees. Although the HENNIG86 m* option does this in a limited manner, the use of many more random input sequences has been recommended in order to increase the likelihood of finding all islands of most-parsimonious trees. Examples where this search strategy has been employed indicate that there are a variety of methods available for searching for alternative islands of most-parsimonious trees. Olmstead and Sweere (1994) use PAUP (Phylogenetic Analysis Using Parsimony, Swofford 1993) to carry out 100 random entry replicates using the Tree Bisection-Reconnection (TBR) search option in conjunction with MULPARS. These parameters allow for exhaustive searching of each potential island. However, the use of such exhaustive procedures was facilitated by the small size of their data set, which comprised only 18 terminals.

In a study with a larger sample size (45 terminals), Clark _et al_. (submitted) searched for islands of most-parsimonious trees using 100 random addition replicates, but restricted the TBR branch swopping to a set of input trees that had been obtained from a heuristic search and Nearest-Neighbour Interchange (NNI) branch swopping procedure; a procedure that Maddison (1991) considers to be the least rigorous of the available branch swopping routines in PAUP.

The size of the data set (the number of terminals) affects both the possibility of finding different islands of most parsimonious trees and the comprehensiveness of the analytical procedures to detect these islands. In an exhaustive empirical study, Maddison (1991) found that data sets which had more than 22 terminals or data matrices which had an RI of less than 0.67 were candidates for producing multiple islands of most-parsimonious trees. As the _rpo_C2 data set is much larger (74 terminals) and the ensemble RI values obtained from the analyses using HENNIG86 are generally lower than 0.50, the data sets are likely to produce multiple islands of most-parsimonious trees. The large number of terminals in the _rpo_C2 data set implies that all but the most narrow of searches will be prohibitively time consuming. The search option chosen is described below. As even this search method took over 36 hours to run to completion, these analyses were restricted to those data sets on which subsequent taxonomic discussions are based.

Method used to search for islands of most-parsimonious trees:

PAUP version 3.1.1 (Swofford 1993) was used to perform a heuristic search and branch swopping on 500 replicates of randomly inputed datasets. This search was done using the heuristic search algorithm HSEARCH with the addition sequence (ADDSEQ) to be used in the stepwise addition procedure set to RANDOM. In accordance with the requirement of this option, the number of replicates (NREPS) was set to 500. The tree bisection-reconnection (TBR) branch-swopping algorithm was used (SWAP = TBR). In order to save on search time, the NOMULPARS option was set and the number of trees (NCHUCK) greater than or equal to a specified length (CHUCKLEN; obtained from the initial m*bb* analysis) was set at 1.

Using these options trees meeting the length requirements and representing islands of most parsimonious trees obtained from different random order inputs were retained (Swofford and Begle 1993). These trees were then used as the starting trees for a second round of analyses using the bb* option of HENNIG86. The strict consensus tree (sensu Page 1989) of all trees found by this procedure was obtained using the "nelsen" command in HENNIG86, and this tree is used as the basis for subsequent taxonomic discussion.

Bootstrap method:

Bootstrap support values (Felsenstein 1985) and Bremer support values (Bremer 1988) of the nucleotide data set were obtained using RNA version 1.1 (Farris 1994). One thousand replicates were done, and a topology showing values at all nodes was obtained (cutting-point = 0).

Distance methods:

Distance-based phylogenetic analyses was performed on the computer-derived alignments, the intuitive alignment and alignments at subsequent stringencies. MEGA version 1.01 (Kumar et al. 1993, licensed to E.H. Harley) was used to produce trees by means of the Neighbour Joining algorithm (Saitou and Nei 1987). This algorithm has been found by means of simulations to be the most efficient method in recovering correct topologies (Nei 1991) and empirical studies have shown that the method gives "reasonable trees" Kumar et al. 1993; quotation marks my own).

Distance between a pair of DNA sequences is expressed as a measure of nucleotide substitutions between them. There are several methods for calculating distance, depending on the pattern of substitutions. In this study the Jukes and Cantor correction (Jukes and Cantor 1969) was used in the calculation of distances between sequences. This correction assumes that the rate of substitution is the same for all of the four nucleotides (A, T, G and C).

The Neighbour Joining method requires a pairwise comparison of the sequences. However, owing to the alignment procedure, sequences pairs may not share all regions of the sequence, as regions (generally corresponding to heptameric repeats) may have been inserted or lost and be represented by gaps. The use of the pairwise deletion option in MEGA allows for comparisons across regions of sequence common to both taxa being compared. The alternative option excludes all those regions which are not common to all taxa in the analysis, an option which would (because of the nature of the various alignments, and the use of an outgroup without the insert) effectively limit the comparison to the short flanking regions on either side of the insert. MEGA was also used to obtain bootstrap figures from 1000 replicates for the Neighbour Joining trees.

RESULTS AND DISCUSSION

Phylogenetic relationships of the genera of the Arundinoideae

The perceived phylogenetic relationships of the taxa sampled may be dependant on the phylogenetic methods used to obtain a cladogram, as well as the method used to produce the sequence alignment. The phylogenetic relationships derived from the different alignment methods are examined and discussed. The intuitive alignment, and the subsequent successive stringency alignments, are discussed first, followed by the results of the computerised alignments.

The phylogenetic statistics of the cladistic analyses of these alignments are listed in Table 2.3. Not unexpectedly, the number of phylogenetically informative sites decreases with increasing alignment stringency. This reflects a loss of homoplasic characters as a result of the recognition of unique heptameric repeats. That these repeats are indeed homoplasic is demonstrated by the 109-step reduction in tree length accompanying the loss of just 21 informative sites between alignment stringencies 1 and 6. This is also evidenced by the increase, albeit slight, in the ensemble consistency and retention indices with increasing alignment stringency.

Strict consensus trees from the cladistic analyses of each alignment stringency are presented and discussed below. In all cases, the number of equal length trees that was found using HENNIG86 was limited by the computer memory capacity. These trees are compared to trees obtained from an expanded data set where characters coding for gaps have been added, as well as to trees obtained using the Neighbour Joining method from the same alignment. The different alignments and associated phylogenies are discussed separately below. To avoid breaking the flow of the text, the figures accompanying each alignment are placed together at the end of the discussion of the relevant alignment.

Table 2.3. Details of the data sets and cladograms derived from increasing alignment stringencies. The first row in each stringency is for the nucleotide data only, the second row includes data coding for gaps. The length column refers to the complete length of the aligned block of sequences, the repeats re-aligned column refers to the number of repeat sized insertions added in each successive stringency. The number of characters refers to the number of phylogenetically informative sites derived from each alignment, including characters for gaps, where relevant. Tree length, ensemble consistency indices (ci) and ensemble retention indices (ri) are also listed.

Stringency	Length (bp)	Repeats re-aligned	No. characters	Tree length	ci	ri
1	819		206	613	45	75
			364	689	44	75
2	855	2	203	606	45	75
			233	688	44	75
3	939	4	197	578	45	75
			232	670	44	75
4	1023	4	190	551	45	76
			229	652	44	76
5	1065	3	189	525	47	77
			229	633	45	77
6	1128	3	185	504	47	74
			228	619	45	76

```
Joinvill    FAYCeDPRYRK--NSSGILK-----------------------  -------  -----------
                                                                 2
Olyra       ...........DKKGSGIVKFRYRTLeDe-------------- |-------| YRTReeDSeDe
Bambusa     ..........KDKKGSGIVKFRYRTLeDe-------------- |YRTLeDe| YRTReeDSeDe
Oryza       VAYFDDPRYKKDKKGSGIVKFRYRTLeDe-------------- |-------| YRTReKDSeNe
Ehrharta    VAYFDDPRYRKDKKGSGIVKFRYRTLeDe-------------- |YRTReDe| YRTReeDSeDe

Bromus      FAYFDDPRYRKDKKGSGIVKFRYRTLeee--------------  -------  YRTReeDSeee
Briza       FAYFDDPRYRKDKKGSGIVKFRYRTLeee--------------  -------  YRTReeDLeYe
Phleum      FAYFDDPRYRKDKKGSGIVKFRYRTLeee--------------  -------  YRTReeDSeDe
Nardus      FAYFDDPRYRKeKKGSGVVKFRYRTLeDe--------------  -------  YKIReADLeNe
Lygeum      FAYFDDPRYRKDKKGSGVVKFRYRALeDe--------------  -------  YKTReADSeNe
Stipa       ...........DKKGSGIVKFRYRTLeDe--------------  -------  YRTReeDSeDe
Anisopog    ....DDPRYRKDKKGSGIVKFRYRTLeNe--------------  -------  YRTReeDSeDe
Zea         FAYFDDPRYRKDKRGSGIVKFRYRTLeeeYRTQeee-------  YRTReee  YRTReeDSeDe
Hyparrhe    ....DDPRYRKDKRGSGIVKFRYRTLeeeYRTQeeeYRTReee  YRTReee  YRTReeDSeDe
Sorghum     FAYFDDPRYRKDKRGSGIVKFRYRTLeee--------------  YRTReee  YRTReeDSeDe
Saccharu    ..........KDKRGSGIVKFRYRTLeeeYRTQeee-------  YRTReee  YRTReeDSeDe
Tristach    .........RKDKRGSGIVKFRYRTLeeeYRTQeeeYRTReee  YRTReee  YRTReeDSeDe
Penniset    FAYFDDPRYRKDKRGSGIVKFRYRTLeee--------------  YRTReee  YRTReeDSeDe
Rhynchel    ..........KDKRGSGIVKFRYRTLeeeYRTLeee-------  YRTReee  YRTReeDSeDe
Brachyar    ..........KGKRGSGIVKFRDRTLeee--------------  YRTLeee  YRTReeDSeee
Panicum     ..........KDKRGSGIVKFRYRTLeee--------------  -------  YRTReeDSeee
Lophathe    FAYFDDPRYRKDKKGSGIVKFRYRTLeee--------------  YRTReee  YRTReeDAeDe
Chasmant    ...........DKRGSGIVKFRYRTLeee--------------  YRTReee  YRTReeDSeDe
Micraira    ..........KDKRGSGIVKFRYRTLeee--------------  YRTReee  YRTReeDSeDe
Monachat    ..........KDKRGSGIVKFRYRTLeee--------------  YRTReee  YRTReeDSeDe
Thysanol    ....DDPRYRKDKRGSGIIKFRYRTLeee--------------  YRTReee  YRTReeDSeDe
Spartoch    ....DDPRYRKDKRGSGIIKFRYRTLeee--------------  YRTReee  YRTReeDSeDe
Stipagro    ......PRYRKDKRGSGIVKFRYRTLeee--------------  YRTReee  YRTReeDSeee
Aristida    ........YRKDKKGSGIVKFRYRTLeee--------------  YRTReee  YRTReeDLeDe
Gynerium    ....DDPRYRKDKRSSGIVKFRYRTLeee--------------  YRTReee  YRTReeDSeDe
Arundo      ...FDDPRYRKDKRGSGIVKFRYRTLeee--------------  YRTReee  YRTReeDSeDe
Mx range    ......PRYRKDKRGSGIVKFRYRTLeee--------------  YRTReee  YRTLeeDSeee
Centropo    ...FDDPRYRKDKRGSGIVKFRYRTLeee--------------  YRTReee  YRTLeeDSeDe
Spartina    FAYFDDPRYRKDKRGSGIVKFRYRTLeee--------------  YRTReee  YRTLeeDSeDQ
Enneapog    ..........KDKRGSGIVKFRYRTLeee--------------  YRTLeee  YRTLeeDSeDe
Eragrost    .......RYRKDKRGSGIVKFRYRTLeee--------------  YRTLeeG  YRTLeeDSeDe
Tragus      ........YRKDKRGSGIVKFRYRTLeee--------------  YRTReee  YRTLeeDSeDe
Hakonech    FAYFDDPRYRKDKRGSGIVKFRYRTLeee--------------  YRTReee  YRTReeDSeNe
Phragmit    FPYFDDPRYRKDKRGSGIVKFRYRTLeee--------------  YRTReee  YRTReeDSeNe
Moliniop    ....DDPRYRKNKRGSGIVKFRYRTLeee--------------  YRTReee  YRTReeDSeNe
Dregeoch    .....DPRYRKDKRGSGIVKFRYRTLeee--------------  YRNReee  YRTReeDSeDG
Amphipog    ......PRYRKDKRGSGIVKFRYRTLeee--------------  -------  YRTReeDSeee
Mx davyi    ..........KDKRGSGIVKFRYRTLeee--------------  -------  YRTReeDSeee
Mx macow    ...FDDPRYRKDKRGSGIVKFRYRTLeee--------------  -------  YRTReeDSeee
Mx rufa     ....DDRPYRKDKRGSGIVKFRYRTLeee--------------  -------  YRTReeDSeee
Mx arund    ..........KDKRGSGIVKFRYRTLeee--------------  -------  YRTReeDSeee
Mx cinct    ..........KDKRGSGIVKFRYRTLeee--------------  -------  YRTReeDSeee
Mx setac    ..........KDKRGSGIVKFRYGTLeee--------------  -------  YRTReeDSeee
Psch asp    ........YRKDKKGSGIVKFRYRTLeee--------------  -------  YRTReeeSeQe
Psch cur    ....DD....KDKKGSGIVKFRYRTLeee--------------  -------  YRTReeeSeQe
Prionant    ...FDDPRYRKDKKGSGIVKFRYRALeee--------------  -------  YRTLeeeSeQe
Pentamer    ..........KDKRGSGIVKFRYRTLeee--------------  -------  YRTReeDLeee
Chionoch    ...FDDPRYRKDKRGSGIVKFRYRTLeee--------------  -------  YRTReeDSeee
Notochlo    ......PRYRKDKRGSGIVKFRYRTLeee--------------  -------  YRTReeDSeee
Cort sel    ....DDPRHRKDKRGSGIVKFRYRTLeee--------------  -------  YRTReeDSeee
Cort ful    ..........KDKRGSGIVKFRYRTLeee--------------  -------  YRTReeDSeee
Plinthan    ........YKKDKRGSGIVKFRYRTLeee--------------  -------  YRTReeDSeee
Lamproth    ...........DKRGSGIVKFRYRTLeee--------------  -------  YRTReeDSeee
Danth se    FAYFDDPRYRKDKRGSGIVKFRYRTLeee--------------  -------  YRTReeDSeee
Danth sp    ..YFDDPRYRKDKRGSGIVKFRYRTLeee--------------  -------  YRTReeDSeee
Chaetobr    .....DPRYRKDKRGSGIVKFRYRTLeee--------------  -------  YRTReeDSeee
Pseudope    .....DPRYRKDKRGSGIVKFRYRTLeee--------------  -------  YRTReeDSeee
Mx stric    ........YRKDKRGSGIVKFRYRTLeee--------------  -------  YRTReeDSeee
Mx dura     ..........KDKRGSGIVKFRYRTLeee--------------  -------  YRTReeDSeee
Mx disti    ........YRKDKRGSGIVKFRYRTLeee--------------  -------  YRTLeeDSeee
Mx guill    ..........KDKRGSGIVKFRYRTLeee--------------  -------  YRTReeDSeee
Urochlae    ....DDPRYRKDKRGSGIVKFRYRTLeee--------------  -------  YRTReeDLeee
Triboliu    ...FDDPRYRKDKRGSGIVKFRYRTLeee--------------  -------  YRTReeDLeee
Danth pa    ...FDDPRYRKDKRGSGIVKFRYRTLeee--------------  -------  YRTReeDSeee
Ryt laev    ...FDDPRYRKDKRGSGIVKFRYRTLeee--------------  -------  YRTReeDSeee
Erythant    ..........KDKRSSGIVKFRYRTLeee--------------  -------  YRTReeDSeee
Ryt nudi    ...........DKRGSGIVKFRYRTLeee--------------  -------  YRTReeDSeee
Schismus    ....DDPRYKKDKRGSGIVKFRYRTLeee--------------  -------  YRTReeDSeee
Karrooch    FAYFDDPRYRKDKRGSGIVKFRYRTLeee--------------  -------  YRTReeDSeee
```

Figure 2.2. The amino acid sequences of the <u>rpo</u>C2 insert and flanking regions from the the 74 taxa included in the phylogenetic study, arranged approximately according to relationships suggested by the results of the phylogenetic analyses. Blocks depict groups of heptameric repeats that were considered to be non-homologous at the various levels of alignment stringency (indicated by the numbers at the top right hand corners of the blocks). For ease of visual comparison, the code for Glutamine (E) has been entered in lower case. Gaps are represented by hyphens (-) and unknown amino acids (unsequenced regions) shown by a period (.). The first eight letters of the taxon names only are provided. Cort = <u>Cortaderia</u>, Danth = <u>Danthonia</u>, Mx = <u>Merxmuellera</u>, Psch = <u>Pentaschistis</u>, Ryt = <u>Rytidosperma</u>.

```
Joinvill   -------   -------------  --- ------  -------  -------------- -------
                  3                                                            2
Olyra      YGSPeNG   YR--TR--e--Ke  YYe TLeDe-  YGVLeDe  -------------- YeTLeDe
Bambusa    YGSPeNK   YR--TR--e--De  Y-e TLeDe-  YRTLeDe  -------------- YeTLeDe
Oryza      YGSPeNe   YR--TR--e--ee  C-K TLeDe-  YRTReee  -------------- YeTLeDe
Ehrharta   YGSPeNe   YR--TR--e--ee  Y-e TLeDK-  YRTLeDe  -------------- YeTLeDe
                  5
Bromus     YeTLeDG   YR--TR--e--De  Y-e ------  -------  -------------- -------
Briza      YeTLeDe   YR--NR--e--De  Y-e ------  -------  -------------- -------
Phleum     YeTLeDe   YR--TR--e--De  Y-e ------  -------  -------------- -------
                                              3
Nardus     YGILeDe   -------------  Y-e TLeeDK  -------  -------------- -------
Lygeum     YeTLeDe   YR--TR--e--De  Y-e ILeeDK  -------  -------------- -------
Stipa      YeTLeDe   YR--TR--e--De  Y-e TLeeDK  -------  -------------- -------
Anisopog   YeTLeDe   YR--TR--e--De  Y-e TLeeDK  -------  -------------- -------
                                   5
Zea        YeSPeNK   YR--TR--eGeGe  Y-K ILeDe-  YRTLeDe  -------------- -------
Hyparrhe   YeSPeNK   YR--TR--eGeGe  Y-e ILeDe-  YRTLeDe  -------------- -------
Sorghum    YeSPeNK   YR--TR--eGeGe  Y-e ILeDe-  YRTLeDe  -------------- -------
Saccharu   YeSPeNK   YR--TR--eGeGe  Y-e ILeDe-  YRTLeDe  -------------- -------
Tristach   YeSPeNK   YR--TR--eGeGe  Y-e ILeDe-  -------  -------------- -------
Penniset   YeSPeNe   YR--TR--eGeGK  Y-e ILeDK-  YRTLeDe  -------------- -------
Rhynchel   YeSPeNe   YR--AR--eGeGK  Y-e ILeDK-  YSTLeDe  -------------- -------
Brachyar   YeSPeNe   YK--TR--eGeGK  Y-e ILeDK-  -------  -------------- -------
Panicum    YeSPeNG   YK--TR--eGeGe  Y-e ILeDK-  YRTLeDe  -------------- -------
Lophathe   YeSPeNe   YR--IR--e--Ge  Y-e TLeDe-  -------  -------------- -------
Chasmant   YeSPeNe   YR--TR--e--Ge  Y-e TLeDe-  YRTLeDe  -------------- -------
Thysanol   YeSPeNK   YR--TR--e--Ge  Y-e TLeDe-  YRTLeDe  -------------- -------
Spartoch   YeSPeNK   YR--TR--e--Ge  Y-e TLeDe-  YRTLeDe  -------------- -------

Micraira   YeSPeNe   YR--TR--e--DG  Y-K TLeDe-  YRTLeNe  -------------- -------
Monachat   YeSPeNe   YR--TR--e--De  Y-e TLeDe-  YRTLQDe  -------------- -------
Stipagro   YeIPeNe   YR--TL--e--De  --- ------  -------  -------------- -------
Aristida   YeSPeKe   YG--VR--e--De  Y-e TLeDe-  -------  -------------- -------
Gynerium   YeIPeNe   YR--TL--e--De  Y-e TLeDe-  YRTLeDe  -------------- -------
Arundo     YeSPeNe   YR--TL--e--De  Y-e TLeDe-  -------  -------------- -------
Mx range   YeSPeNe   YR--TR--D--DK  Y-e TLeDe-  YRTLeDe  -------------- -------
Centropo   YeSPeNe   YR--TR--e--De  Y-e TLeDe-  YRTLeDe  -------------- -------
Spartina   YeSPeK-   ----IR--e--De  Y-e TLeDe-  YRALeDe  -------------- -------
Enneapog   YeSPeNe   YR--TR--e--De  Y-e TLeDe-  YRTLeDe  -------------- -------
Eragrost   YeSLeNe   YR--TR--e--De  Y-e TLeDe-  YSTLeDe  -------------- -------
Tragus     YeSPeNe   YR--TL--e--De  Y-e TLeNe-  YRTLeDe  -------------- -------
                  5
Hakonech   YeTLeDe   YR--TL--e--De  Y-e TLeDe-  -------  -------------- -------
Phragmit   YKTLeee   YR--TL--e--De  Y-e TLeee-  YRTLeDe  -------------- -------
Moliniop   YeTLeDe   YR--TL--e--De  Y-e TLeDe-  YRTLeDe  -------------- -------
Dregeoch   YeTLeDe   YR--TLALe--De  --- ------  -------  -------------- -------

Amphipog   YeSPeNe   YK--TR--e--De  Y-e TLeDe-  YRALeDK  -------------- -------
                  6
Mx davyi   YeSLeNe   YK--TL--e--DG  Y-e TLeDe-  YRTLeDe  -------------- YeTReDe
Mx macow   YeSLeNe   YK--TL--e--DG  Y-e TLeDe-  YRTLeDe  -------------- YeTReDe
Mx rufa    YeSLeNe   YR--TR--e--De  Y-e TLeDe-  YRTLeDe  -------------- YeTReDe
Mx arund   YeSLeNe   YR--TR--e--De  Y-e TLeDe-  YRTLeDe  -------------- YeTReDe
Mx cinct   YeSLeNe   YR--TR--e--De  Y-e TLeDe-  YRTLeDe  -------------- -------
Mx setac   YeSLeNe   YR--TR--e--De  Y-e TLeDe-  YRTQeDe  -------------- YeTReDe

Psch asp   YeSLeNe   YG--PG--e--De  Y-e ALeDe-  YGAReDe  -------------- -------
Psch cur   YeSLeNe   YG--PR--e--De  Y-e AIeDe-  YGPReDe  -------------- -------
Prionant   YeSLeNe   YG--PR--e--De  Y-G ALeDe-  YGALeDe  -------------- -------
Pentamer   YeSLeDe   YG--TR--e--De  Y-e ALeDe-  YeTReGe  -------------- -------

Chionoch   YeSLeDe   YK--TR--e--De  Y-e TLeDe-  YRTLeDe  -------------- YeTReDe
                                                        6
Notochlo   YeSLeDe   YK--TR--e--ee  Y-e SLeNe-  YKTReDe  -------------- YeTReDe
Cort sel   YeSLeDe   YK--TR--e--ee  Y-e SLeNe-  YKTReDe  -------------- YeTReDe
Cort ful   YeSLeDe   YK--TR--e--ee  Y-e SLeNe-  YKTReDe  -------------- YeTReDe
Plinthan   YeSLeDe   YK--TR--e--ee  Y-e SLeNe-  YKTReDe  -------------- YeTReDe
Lamproth   YeSLeDe   YK--TR--e--ee  Y-e SLeNe-  YKTReDe  -------------- -------
Danth se   YeSLeDe   YKNKTR--e--ee  Y-e SLeNe-  YKTReDe  -------------- YeTReDe
Danth sp   YeSLeDe   YK--TR--e--ee  Y-e SLeNe-  YKTReDe  -------------- YeTReDe
Chaetobr   YeNLeD-   ----AR--e--ee  Y-e NLeNe-  YKTReDe  -------------- YeTReDe
Pseudope   YeSLeD-   ----AR--e--ee  Y-e SLeNe-  YKTReGe  -------------- YeTReDe
Mx stric   YeSLeDe   YK--PR--e--Ne  --- ------  YKTReDe  -------------- -------
Mx dura    YeSLeDe   YK--TR--e--Ne  --- ------  -------  -------------- YeTReDe
Mx disti   YeSPeDe   YK--TR--e--ee  Y-e SLeDe-  YKTReDe  -------------- YeTReDe
Mx guill   YeSLeDe   YK--TR--e--ee  Y-e SLeNe-  YKTReDe  -------------- YeTReDe
Urochlae   YeSLeDe   YK--TR--e--ee  Y-e SLeNe-  YKTReDe  -------------- YeTReDe
Triboliu   YeSLeDe   YK--TR--e--ee  Y-e SLeNe-  YKTReDe  -------------- YeTReDe
Danth pa   YeSLeNe   YK--TR--Q--De  --- ------  -------  YeTReDeYGILeDe YeTReDe
Ryt laev   YeSLeDe   YK--TR--e--ee  Y-e SLeNe-  YKTReDe  -------------- YeTReDe
Erythant   YeSLeDe   YK--TR--e--ee  Y-e SLeNe-  YKTReDe  -------------- YeTReDe
Ryt nudi   YeSLeDe   YK--TR--e--ee  Y-e SLeNe-  YKTRQDe  -------------- YeTReDe
Schismus   YeSLeDe   YK--TR--e--ee  Y-e SLeNe-  YKTReDe  -------------- YeIReDe
Karrooch   -------   -------------  Y-e SLeNe-  YKTReDe  -------------- YeTReDe
```

Figure 2.2 cont.

```
Joinvill    ----------------- -------------- ------- -------------- -------
                                                                           3
Olyra       ----------YGILeGe YeTLeDe------- YGILeDe -------------- YRTReeD
Bambusa     ----------YGILeDe YeTLeDe------- ------- -------------- YRTReeD
Oryza       ----------YGIPeNe YeTLeDe------- YGILeDe -------------- YRTReee
Ehrharta    ----------YGILeDe YeTLeDe------- YGIPeNe -------------- YRTReeD
                                                    3
Bromus      -------TSKYGIRDDG YeTLeDe------- YGSPeNe -------YGNPeNe YRTLeKD
Briza       ----------------- YeTLeDe------- YGSLeNe -------------- YRTFeKD
Phleum      ----------------- YeTLeDe------- YGSPeNe -------YGNPeNe YRTLeKD
Nardus      ----------YRILeDe YeTLeee------- YGSPeNe -------------- YRTLeKN
Lygeum      ----------YeILeDe YeTLeDe------- YGSPeNe -------------- YRTLeKN
Stipa       ----------YGILeDe YeTLeDe------- YGSPeNe -------YGSPeNe YRILeKD
Anisopog    ----------YGILeDe YeTLeDe------- YGSSeKe -------------- YRTLeKD

Zea         ----------------- YeTLeDe------- YGILeDe -------------- YRTLeKD
Hyparrhe    ----------------- YeTLeDe------- YGILeDe -------------- YRTLeKD
Sorghum     ----------------- YeTLeDe------- YGILeDe -------------- YRTLeKD
Saccharu    ----------------- YeTLeDe------- YGILeDe -------------- YRTLeKD
Tristach    ----------------- -------------- YGILeNe -------------- YRTLeKD
Penniset    ----------------- YeTLeDe------- YGILeDe -------------- YeTLeKD
Rhynchel    ----------------- YeTLeDe------- YGILeDe -------------- YRTLeKD
Brachyar    ----------------- -------------- ------- -------------- -------
Panicum     ----------------- YeTLeDe------- YGILeDe -------------- YRTLeKD
Lophathe    ----------------- -------------- YGILeDe -------------- YRTLeKD
Chasmant    ----------------- YeTLeDe------- YGILeDe -------------- YRTLeKD
Micraira    ----------------- YeTLeDe------- YGILeDe -------------- YRTPWKD
Monachat    ----------------- YeTLeDe------- YGILeDe -------------- YRTLeKD
Thysanol    ----------------- YeTLKDe------- YGILeDe -------------- YRTLeKD
Spartoch    ----------------- YeTLeDe------- YGILeDe -------------- YRTLeKD
Stipagro    ----------YGILeDe YeTLeDe------- YGILeDe YRILeDe------- YRTLeKD
Aristida    ----------------- YeTLeDe------- YGMLeDe YeTLeDe------- YRTLeKD
Gynerium    ----------------- YeTLeDe------- YGILeDe -------------- YRTLeKD
Arundo      ----------------- -------------- YGILeDe -------------- YRTLeKD
Mx range    ----------------- YeTLeDe------- YGILeDe -------------- YRTLeKD
Centropo    ----------------- YeTLeDe------- YGILeDe -------------- YRTLeKD
Spartina    ----------------- YeTLeDe------- YGILeDe -------------- YeALeKD
Enneapog    ----------------- YeTLeDe------- YGILeAK -------------- YRTLeKD
Eragrost    ----------------- YeILeDe------- YGILeAK -------------- YRTLeKD
Tragus      ----------------- YeTLeDe------- YGILeDe -------------- YRTLeKD
Hakonech    ----------------- -------------- YGILeDe -------------- YRTLeKD
Phragmit    ----------------- YeTLeDe------- YGILeDe -------------- YRTLeKD
Moliniop    ----------------- YeTLeDe------- YGILeDe -------------- YRTLeKD
Dregeoch    ----------------- YeTLeDe------- YGTLeDe -------------- YRTIeKD
Amphipog    YeILGDe---------- -------------- YGILeDe -------------- YRTLeKD
Mx davyi    ----------YGILeDe YeTLeDe------- YGILeDe -------------- YRTLeKD
Mx macow    ----------YGILeDe YeTLeDe------- YGILeDe -------------- YRTLeKD
Mx rufa     ----------YGILeDe YeTLeDe------- YGILeDe -------------- YRTLeKD
Mx arund    ----------YGILeDe YeTLeDe------- YGILeDK -------------- YRTLeKD
Mx cinct    ----------------- YeTLeDe------- ------- -------------- YGTLeKD
Mx setac    ----------YGILeDe YeTLeDe------- YGILeNe -------------- YRTReKD
                             4
Psch asp    ----------YGSLeDe -------------- YGILeDe -------------- YRTLeKD
Psch cur    ----------YGALeDe YeTLeDe------- YGILeDe -------------- YRTLeKD
Prionant    ----------YGALeDe YeTLeDe------- YGILeDe -------------- YRTLeKD
Pentamer    ----------YeALeDe YeTLeDe------- YGILeDe -------------- YRTLeKD

Chionoch    ----------YGILeDe YeTLeDe------- YGILeDe -------------- YRTLeKD
Notochlo    ----------YGILeDe YeTLeDK------- YGILeDe -------------- YRTLeKD
Cort sel    ----------YGILeDe YeTLeDe------- YGILeDe -------------- YRTLeKD
Cort ful    ----------YGILeDe YeTLeDK------- YGILeDK -------------- YRTLeKD
Plinthan    ----------YGILeDe YeTLeDK------- YGILeDe -------------- YRTLeKD
Lamproth    ----------------- YeTLeDe------- YGILeDe -------------- YRTLeKD
Danth se    ----------YGILeDe YeTLeDK------- YGILeDe -------------- YRTLeKD
Danth sp    ----------YGILeDe YeTLeDK------- YGILeDe -------------- YRTLeKD
Chaetobr    ----------YGTLeee YeTLeDe------- YGILeDe -------------- YRTLeKD
Pseudope    ----------YGILeee YeTLeDe------- YGILeDe -------------- YRTLeKD
Mx stric    ----------YeILeDe YeTLeDe------- YGILeDe -------------- YRTLeKD
Mx dura     ----------YGILeDe YeTLeDeYeTLeDe YGILeDe -------------- YRTLeKD
Mx disti    ----------YGILeDe YeILeDeYeILeDe YGILeDe -------------- YRTLeKD
Mx guill    ----------YGILeDe YeTLeDeYeILeDe YGILeDe -------------- YRTLeKD
Urochlae    ----------YGILeDe YeTLeDeYeILeDe YGILeDe -------------- YRTLeKD
Triboliu    ----------YGILeDe YeTLeDeYeILeDe YGILeDe -------------- YRTLeKD
Danth pa    ----------YGILeDe YeTLeDeYeILeDe YGILeDe -------------- YRTLeKD
Ryt laev    ----------YGILeDe YeTLeDeYeILeDe YGILeDe -------------- YRTLeKD
Erythant    ----------YGILeDe YeTLeDeYeILeDe YGILeDe -------------- YRTLeKD
Ryt nudi    ----------YGILeDe YeTLeDeYeILeDe YGILeDe -------------- YRTLeKD
Schismus    ----------YGILKDe YeTLeDeYeILeDe YGILeee -------------- YKTLeKD
Karrooch    ----------YGILeDe YeTLeDeYeILeDe YGILeDe -------------- YRTLeKD
```

Figure 2.2. cont.

```
Joinvill   -------------------------------------------------------------
Olyra      ----SeDe-YRSPeNeYRTReDe--------------------------------------
Bambusa    ----SeDe-YGSPeNKYRTReDe----------------YGTLDeDSeDeYGSPeeGSeDe
Oryza      ----SeDe-YGSPeNKYRPReDK----------------YGTLeeDSeDe-----------
Ehrharta   ----SeAe-YGSPeNeYRTReDe----------------YGTLeeDSeee-------SeDe
Bromus     ----Seee-YGSPeSKYRTQeDe----------------YGTLeeDSeDeYGSPGeSAeDK
Briza      ----LeDe-YGSPeSKYRTReGe----------------YGTLeeDSeDeYGNLGeSSeDK
Phleum     ----SeDe-YGSPeSKYRTReGK----------------YGTLeeDSeDeYGSPGeSSeDK
Nardus     ----SeDeeYGSPeSeYRTReDe----------------YGTLeeDSeDeYGSAGGSSeDe
Lygeum     ----SeDeeYGNPeSeYRTReNe----------------YGTLeeDSeDeYGNPGeGSeDe
Stipa      ----SeDe-YGSPeSQYRTReDe----------------YGTLeeDSeDeYGSPGeGSeDK
Anisopog   ----SeDe-YGSPeSeYRTReDe----------------YGTLeeDSeYeYGSLGeDSeDK
Zea        ----Seee-YGSLeNKYRTReGe-------GeYeI-------LeeeSeeeYGSSeDGSeKe
Hyparrhe   ----Seee-YGSLeNKYGTReGe-------GeYeI-------LeeDSeeeYGSSeDGSeKe
Sorghum    ----Seee-YGSLeNKYRTReGe-------GeYeI-------LeeDSeeeYGSSeDGSeKe
Saccharu   ----Seee-YGSLeNKYRTReGe-------GeYeI-------LeeDSeeeYGSSeDGSeKe
Tristach   ----Seee-YGSLeNKYRTReGe-------GeYeI-------LeeDSeeeYGSSeDGSeKe
Penniset   ----Seee-YGSPeNeYKTReGe-------GeYeILeDKYRTLeeDSeeeYGNSeDGSeKe
Rhynchel   ----Seee-YGSPeNeYKTReGe-------GeYeILeDKYRTLeeDSeee-----------
Brachyar   ---------------------------------------YRTLeeDSeeeYGSSeDGSeKe
Panicum    SeKDSeee-YGSPeNGYKTReGe-------GeYeILeDRYRTLeeDSeeeYGSSKDGSeKe
Lophathe   ----Seee-YGSPeNeYKTReDe----------------YGTLeeDSeeeYRSPeDGSeNe
Chasmant   ----Seee-YGSPeNeYKTReDe----------------YGTLeeDSeeeYGSPeDGSeNe
Micraira   ----SeeQ-YGSPeNeYNTReDe----------------YGTLeeDSeeeYGSPeDGSeNe
Monachat   ----Seee-YGSPeNeYKTReDe----------------YGTLGeDSeeeYGSPeDGSeNe
Thysanol   ----Seee-YGSPeNeYKTReDe----------------YGTLeeDSeeeYGSPeNDSeNe
Spartoch   ----Seee-YGSPeNeYKTReDe----------------YGTLeeDSeeeYRSPeDGSeNe
Stipagro   ----SQeD-YeNPeNeYKTRKDe----------------YGTLeeDSeeeYGSPeDGSeNe
Aristida   ----SQee-YRSPeNeYKTRKNe----------------YGTLeeDPeeeYGSPeDGSeNe
Gynerium   ----Seee-YGSPeNeYKTReDe----------------YGTLeeDSeeeYGSPeDGSeNe
Arundo     ----Seee-YRSPeNeYKTReDe----------------YGTLeKDSeeeYGSPeDGSeNe
Mx range   ----Seee-YGSPeNeYKTReDe----------------YGTLeeDSeeGYGSReDGSeNe
Centropo   ----Seee-YGSPeNeYKTReDe----------------YeTLeeDSeKGYGSReDGSeNe
Spartina   ----Seee-YGNPeNeYKTReDe----------------YGTLee---eeYGSReDGSeNe
Enneapog   ----Seee-SGNPeNKYKSReDe----------------YGTLeeDSeeeYGSHeDGSeNe
Eragrost   ----Seee-SGNPeNeYKSReDe----------------YGTLeeDSeeeYGSReDGSeNe
Tragus     ----Seee-YGNPeNKYKTReDQ----------------YGTLKe---eeYGSReDGSeNe
Hakonech   ----Seee-YGSPeNeYKTReDe----------------YGTLeeDSeeeYGSPeDGSeNe
Phragmit   ----Seee-YGSPeNeYKTReDe----------------YGTLeeDSeeeYGSPeDGSeNe
Moliniop   ----Seee-YGSPeNeYKTReDe----------------YGTLeeDSeeeYGSPeDGSeNK
Dregeoch   ----Seee-YGSLeNeYKTReDe----------------YRTLeeDSeeeYGNPeDGSeNe
Amphipog   ----Seee-YGSPeDeYKTReDe----------------YGTLeeDSeeeYGSPeDGSeNe
Mx davyi   ----Seee-YGSPeNeYKTReDe----------------YGTLeeDSeeeYGNReDGSeNe
Mx macow   ----Seee-YGSPeNeYKTReDe----------------YGTLeeDSeeeYGNReDGSeNe
Mx rufa    ----Seee-YGSPeNeYKTReDe----------------YGILeeDLeeeYGSReDGSeNe
Mx arund   ----Seee-YGSPeNeYKTReDe----------------YGILeeDLeeeYGSReDGSeNe
Mx cinct   ----SeeQ-YGSPeNeYKTReDe----------------YGILeeDLeeeYGSReDGSeNe
Mx setac   ----Seee-YGSPeNeYKTReDe----------------YGILeeDLeeeYGSReDGSeNe
Psch asp   ----Seee-YGSPeNeYKTReDe----------------YGILeeDLeeeYGSHeDGSeNe
Psch cur   ----Seee-YGSPeNeYKTReDe----------------YGILeeDLeeeYGSHeDGSeNe
Prionant   ----Seee-YGSPeNeYKTReDe----------------YGILeeDLeeeYGSHeDGSeNe
Pentamer   ----Seee-YGSPeNeYKTReDe----------------YGILeeDLeeeYGSReDGSeNe
Chionoch   ----Seee-YGSPeNQYKTReDe----------------YGILeeDLeeeYGSReDGSeNe
Notochlo   ----Seee-YGSPeNQYKTReDe----------------YGILeeDLeeeYGSReDGSeNe
Cort sel   ----Seee-YGSPKNQYKTReDe----------------YGILeeDLeeeYGSReDGSeNe
Cort ful   ----Seee-YGSPeNQYKTReDe----------------YGILeeDLeeeYGSReDGSeNe
Plinthan   ----Seee-YGSPeNQYKTReDe----------------YGILeeDLeeeYGSReDGSeNe
Lamproth   ----Seee-YGSPKNQYKTReDe----------------YGILeeDLeeeYGSReDGSeNe
Danth se   ----Seee-YGSPeNQYKTReDe----------------YGILeeDLeeeYGSReDGSeNe
Danth sp   ----Seee-YGSPeNQYKTReDe----------------YGILeeDLeeeYGNReDGSeNe
Chaetobr   ----Seee-YGSPeNeYKTReDeYeTReDe--------------------YGSReDGSANR
Pseudope   ----Seee-YGSPeNeYKTReDeYeTReDe--------------------YGSReeGSANe
Mx stric   ----Seee-YGSPeNQYKTReDe----------------YGILeeDLeeeYGSReDSSeKe
Mx dura    ----Seee-YGS--NQYKTReDe----------------YGILeeDLeeeYGSReDGSeNe
Mx disti   ----Seee-YGSPeNQYKTReDe----------------YGILeeDLeeeYGSReDGSeNe
Mx guill   ----Seee-YGSPeNQYKTQeDe----------------YGILeeDLeeeYGSReDGSeNe
Urochlae   ----Seee-YGSPeNQYKTQeDe----------------YGILeeNLeeeYGSReDGSeNe
Triboliu   ----Seee-YGSPeNQYKTQeDe----------------YGILeeNLeeeYGSReDGSeNe
Danth pa   ----Seee-YGSPeNQYKIQeDe----------------YGILeeDLeeeYGSReDGSeNe
Ryt laev   ----Seee-YGSPeNQYKIQeDe----------------YGILeeDLeeeYGSReDGSeNe
Erythant   ----SeKe-YGSPeNQYKIQeDe----------------YGILeeDLeeeYGSReDGSeNe
Ryt nudi   ----Seee-YGSPeNQYKIQeDe----------------YGILeeDLeeeYGSReDGSeNe
Schismus   ----Seee-YGSPeNQYKTQeDe----------------YGILeKDLeeKYGNReDGSeNe
Karrooch   ----Seee-YGSPeNQYKTQeDe----------------YGILeeDLeEEYGSPEDGSeNe
```

Figure 2.2. cont.

```
Joinvill    -------------------YGTLeeDSIIKKeGL
Olyra       YGTLeeDSee----DSeDeYGSPeeDSI......
Bambusa     YGTLeeDSee----DSeDeYGSPeeDSIL.....
Oryza       HGTLeeDSeeDSeeDSeDeYGNPeeDSVLKKGVL
Ehrharta    YGTLeeDSee----DSeDeYGSPQeDSILKKeGL
Bromus      YGTLeeDSee----DSeDeYeSPeeDSILKKeGL
Briza       YRTLeeDLee----DSeDeYDSPeeDSILKKeGL
Phleum      YGTLeeDLee----DSeDeYeSPeeDSILKKeGL
Nardus      YGTLeeDSee----DSeDeYGSSeeDSVLKKeGL
Lygeum      YGTLeeDSee----DSeDeYGSPeeDSLLKKeGL
Stipa       YGTLeeDSee----DSeDeYeSPeeDSI......
Anisopog    YGTLeeDSee----DSeDeYeSPeeDSILKKe..
Zea         YGTLeeDSeeDSeeDSeDeYGSPeeDSILKKeGF
Hyparrhe    YGTLeeDSeeDSeeDSeDeYGSPeeDSILKK...
Sorghum     YGTLeeDSeeDSeeDSeDeYGSPeeDSILKKeGF
Saccharu    YGTLeeDSeeDSeeDSeDeYGSPeeDSILKK...
Tristach    YGTLeeDSeeDSeeDSeDeYGSPeeDS.......
Penniset    YGTLeeDSee----DSeDeYGSPeeDSILKReS.
Rhynchel    YGTLeeDSee----DSeDeYGSPee.........
Brachyar    YGTLeeDSee----DSeDeYGSPeeDSILKK...
Panicum     YGILeeDSee----DSeDeYGGPee.........
Lophathe    YGTLeeDSee----DSeDeYGSLeeDSILKKeGF
Chasmant    YGTLeeDSee----DSeDeYGSPe....LKK...
Micraira    YGTLeeDSee----DSeDeYGSPeeDSMLKKeG.
Monachat    YGTLeeDSee----DSeDeYGSPee.........
Thysanol    YGTLeeDSee----DSeDeYGNPee....KK...
Spartoch    YGTLeeDSee----DSeDeYGSPeeDSILK....
Stipagro    YGTLeeNSee----ePeDeYGSPee.........
Aristida    YGTLeeYSee----DSeDeYGRPeeDSI......
Gynerium    YGTLeeDSee----DSeDeYGSPeeDSILKK...
Arundo      YGTLeeDSee----DSeDeYGSPeeDSILKKeG.
Mx range    YGALeeDSee----DSeDeYGSSeeDS.......
Centropo    YGALeeDSee----DSeDeYGSPeeDSIFKK...
Spartina    YGALeeDSee----DSeDeYGSPeeDSILKKeGF
Enneapog    YGALeeDSee----DSeDeYGSSeeDSILKK...
Eragrost    YGALeeDSee----DSeDeYGSPeeDSILKKeG.
Tragus      YGALeeDSee----DSeDeYGSSeeDSILKK...
Hakonech    YGTLeeDSee----DSeDeYGSPeeDSILKKeGF
Phragmit    YGTLeeDSee----DSeDeYGSPeeDSILKKeGF
Moliniop    YGTLeeDSee----DSeDeYGSPeeDSILKKeG.
Dregeoch    YGTLeeDSee----DSeDeYGSPeeDSILKK...
Amphipog    YGTLeeDSee----DSeDeYGSPeeDSILK....
Mx davyi    YGTLeeDSee----DSeDeYGSPe..........
Mx macow    YGTLeeDSee----DSeDeYGSPeeDSILKKGG.
Mx rufa     YGTLeeDSee----DSeDeYGSPeeDSILKK...
Mx arund    YGTLeeDSee----DSeDGYGSP...........
Mx cinct    YGTLeeDSee----DSeDeYGSPee.........
Mx setac    YGTLeeDSee----DSeDeYGSPeeDSILKK...
Psch asp    YGTLeeDSee----DSeDeYGSPeeGSILKKeA.
Psch cur    YGTLeeDSee----DSeDeYGSPeeGPILKK...
Prionant    YGTLeeDSee----DSeNeYGSPeeGSILKKDG.
Pentamer    YGTLeeDSee----DSeDeYGSPe..........
Chionoch    YGTLeeDSee----DSeDeYGNPeeDSILKKeG.
Notochlo    YGTLeeDSee----DSeDeYGSPee.........
Cort sel    YGTLeeDSee----DSeDeYGSPeeDSLLKKeG.
Cort ful    YAILeeDSeK----DSeDeYGSPeeDSLLKKe..
Plinthan    YGILeeDSee----DSeDeYGSPeeD........
Lamproth    YGTLeeDSee----DSeDeYGSPeeDSLLKKe..
Danth se    YGTLeeDSee----DSeDeYGSPeeDSLLKKe..
Danth sp    YGTLeeDSee----DSeDeYGSPeeDSLLKKeG.
Chaetobr    YGTLeeDSee----DSeDeYGSPeeDSLLKK...
Pseudope    YGTLeeDSee----DSeDeYGSPee.SLLKKeG.
Mx stric    YGTLeeDSeeDSeeDSeDeYGSPeeDSL......
Mx dura     YGTLeeDSeeDSeeDSeDeYGSPeeDSLKK....
Mx disti    YGTLeeNSeeDSeeDSeeeYGSPeG.........
Mx guill    YGALeeDSeeDSeeDSeDeYGSPe..........
Urochlae    YGALeeDSeeDSeeDSeDeYGSPeeDSLFKK...
Triboliu    YGALeeDSeeDSeeDSeDeYGSPeeDSLFKKeG.
Danth pa    YGALeeDSeeDSeeDSeDeYGSPeeDSLLKKe..
Ryt laev    YGALeeDSeeDSeeDSeDeYGSPeeDSLLKKe..
Erythant    YGALeeDSeeDSeeDSeDeYGSPeeDSLLK....
Ryt nudi    YGALeeDSeeDSeeDSeDeYGSPeeDSLLK....
Schismus    YGALeeDSeeDSeeDSeDeYGSPeeDSLLKKeG.
Karrooch    YGALeeDSeeDSeeDSeDeYGSPeeDSLLKK...
```

Figure 2.2. cont.

Phylogenies from the intuitive alignment (stringency 1)

The alignment produced using the rule-governed method described above is shown in Figure 2.2. In this figure, the amino acid sequences of the taxa are shown, as these were the basis of the rule-governed alignment method. This alignment is the initial stringency level from which the alignments at successive stringencies are derived. Blocks of the heptameric repeats that were re-aligned in the subsequent stringencies are marked in this figure. In this alignment, the majority of the insertions serve only to align major blocks of similar repeats. The positional homology of the repeats was not critically evaluated.

Informative sites only:

This data set was the only one which appeared to contain multiple islands of most-parsimonious trees, or at least suffer from the memory limitations imposed upon HENNIG86's m*bb* analysis; the consensus tree obtained from the m*bb* analysis was better resolved than that found following the 500 replicate random order input search. The latter, less resolved tree is presented and discussed below.

The consensus tree from the random order input search is not well resolved. There is a basal polychotomy that comprises individual genera (Arundo, Gynerium, Amphipogon, Lophatherum and Chasmanthium) and other more resolved clades corresponding to the major lineages of the grasses (Figure 2.3). The bambusoid and pooid subfamilies taxa are shown to be sisters. Panicoideae and Chloridoideae are both well resolved, as is a lineage corresponding approximately to the Danthonieae sensu Watson (1990). Several two-taxon clades are also part of the basal polychotomy. These correspond approximately to the tribes Thysanolaeneae, Micraireae and Aristideae sensu Clayton & Renvoize (1986) and Arundineae sensu Watson (1990). The previously monotypic Micraireae and Thysanolaeneae include other arundinoid taxa; Micraira is shown to be sister to Monachather, while Thysanolaeana is sister to Spartochloa.

The bootstrap analysis produced a different topology (Figure 2.4) from the consensus tree discussed above. As in the consensus tree, the major clades are well resolved, although bootstrap support for these is not always strong. The pooid clade has the highest bootstrap support (86.8%; Bremer support value = 6). The PACC clade is not resolved, as the poorly supported (45.7%) bambusoid clade is shown to be sister to the Panicoideae (33.6%; Bremer support value = 3). The danthonioid clade is retained, although the relationships of its constituent taxa are different from those shown in the consensus tree. Furthermore, Amphipogon is shown to be the basal-most taxon in this lineage. Bootstrap support for this clade (i.e. including Amphipogon) is 60.5% (Bremer support value = 5). Bootstrap support for the danthonioid clade excluding Amphipogon is weak (41.6%). The arundinoid clade (reeds) includes the centothecoid Lophatherum, and support for this clade is weak (18.9%; Bremer support value = 3).

Of the smaller tribes of the Arundinoideae, the Aristideae (Aristida, Stipagrostis) are reasonably supported (61.5%; Bremer support value = 5), while representatives of the other monotypic tribes are shown to be associated with other taxa, as shown in the consensus tree discussed above. The (Micraira, Monachather) clade is weakly supported (42.0%; Bremer support value = 1), while the (Thysanolaena, Spartochloa) clade is well supported (92.5%; Bremer support value = 8) and shown to be related to Gynerium (39.0% bootstrap support).

The basal position in the topology is occupied by Brachiaria, a panicoid genus which has a short rpoC2 insert. The position of this taxon next to Joinvillea is probably caused by the substantial number of "?" characters from the insert region that both these taxa would have. This relationship is retained in the RNA-produced trees from all the alignment stringencies.

Informative sites and gaps:

The addition of characters coding for gaps improves the resolution of the tree substantially (Figure 2.5). The PACC clade is now resolved, although the relationships of the constituent major lineages is not fully resolved. The topology of the consensus trees from the analysis of this expanded data set resolves a paraphyletic Arundineae sensu Watson (1990) and Aristideae as the sister clade to the rest of the PACC clades, which form a four-clade polytomy. Lineages in this polytomy include a panicoid clade, now shown to include centothecoid taxa (Chasmanthium), Thysanolaeneae and two arundinoid taxa, Gynerium and Spartochloa. The latter taxon appears as sister to Thysanolaena, a position it retains irrespective of method of cladistic analysis or alignment. Another improvement to resolution of the PACC clade is the inclusion of Amphipogon as the basal taxon in the danthonioid clade; a relationship also shown in the neighbour joining tree.

Neighbour Joining method:

In this analysis, the pooid clade is shown to be the basal lineage in the grasses and (unlike the trees discussed above) not sister to Bambusoideae (Figure 2.6). Bootstrap support for the pooid clade is moderate (66%), while that for the bambusoid clade is weak (41%). The PACC clade is retrieved in this analysis, and receives a bootstrap support value of 52%. The basal taxon in the PACC clade is Lophatherum, after which the poorly supported (Panicoideae, Centothecoideae) clade is derived. The relationships among the remaining major lineages in the PACC clade, (Aristideae and Arundineae sensu Watson (1990)), Chloridoideae, Micraireae and Danthonieae sensu Watson (1990) are unresolved. The topology within the danthonioid clade is somewhat different from that obtained from the parsimony analyses, and the clade as whole (excluding Amphipogon) recieves 46% bootstrap support.

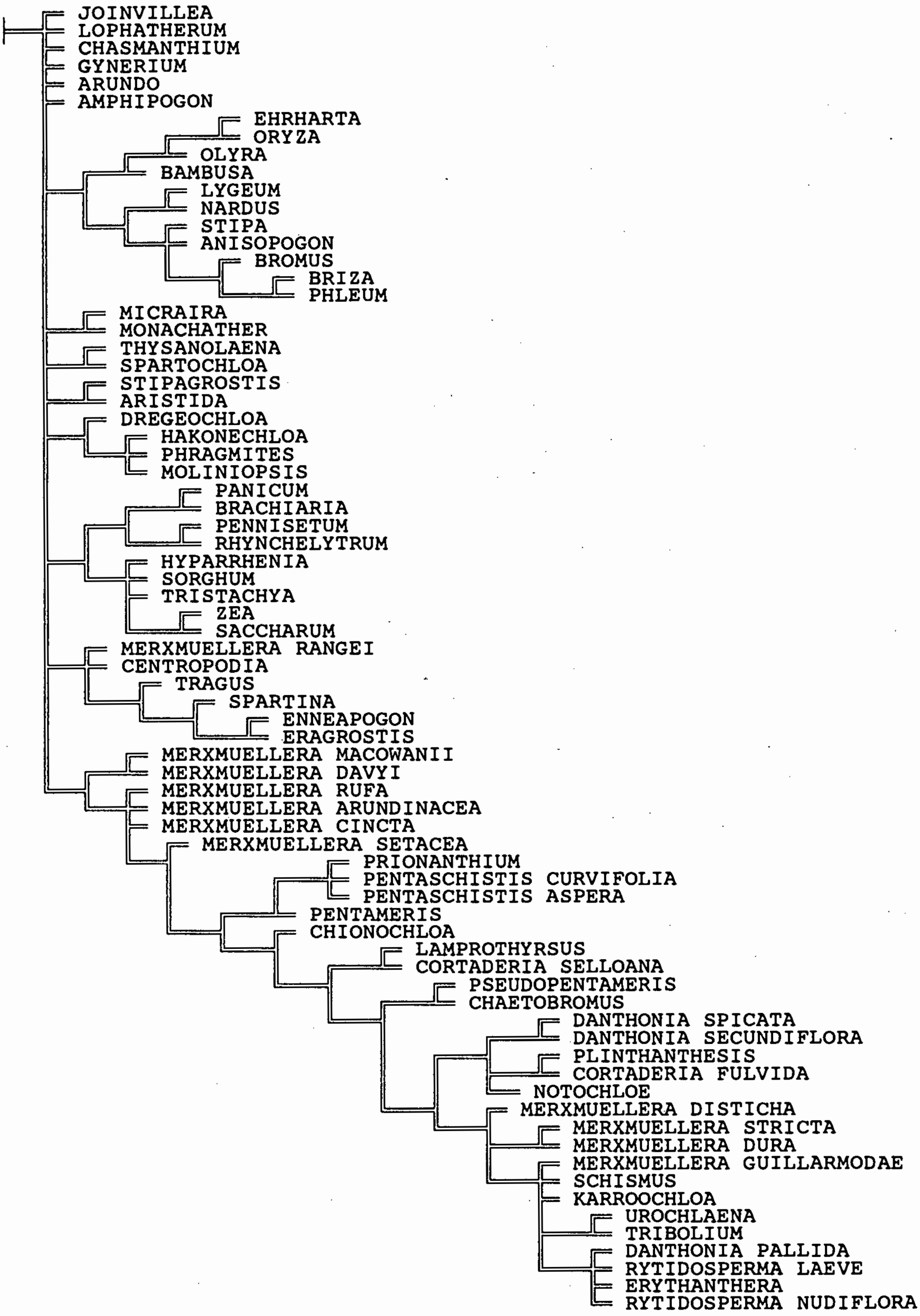


Figure 2.3. Consensus tree of the 1239 trees (memory overflow) obtained from a bb* analysis of 54 trees obtained from a random input order analysis carried out using PAUP. This analysis is based on the sequences aligned at stringency 1. Length = 613, ci = 45, ri = 75.

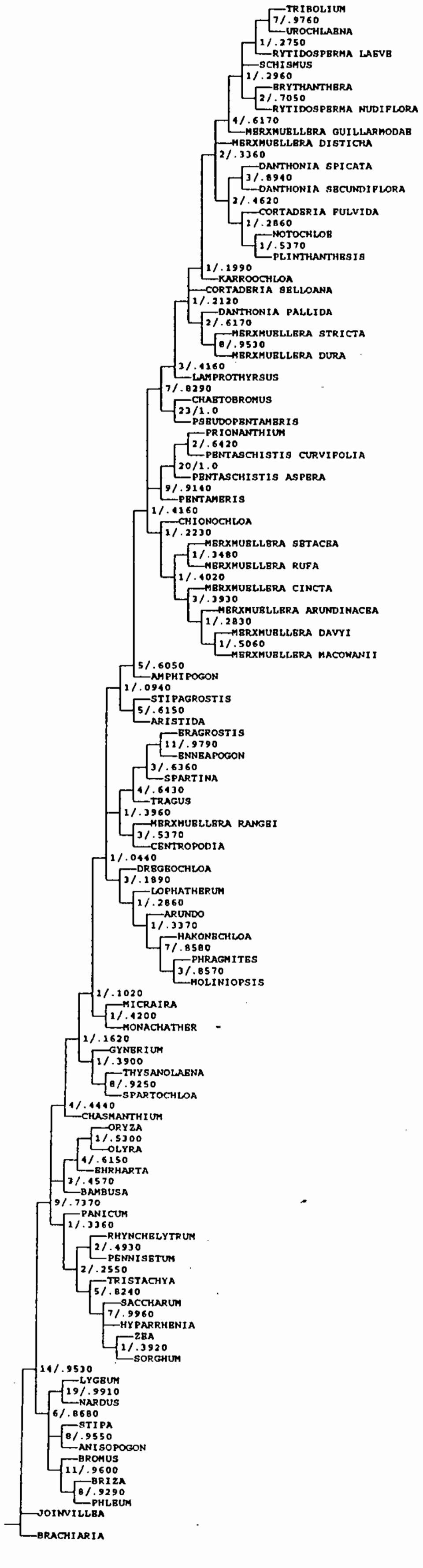


Figure 2.4. Topology showing Bremer support (left of "/") and bootstrap support values produced from 1000 replicates of the nucleotide data aligned at stringency 1.

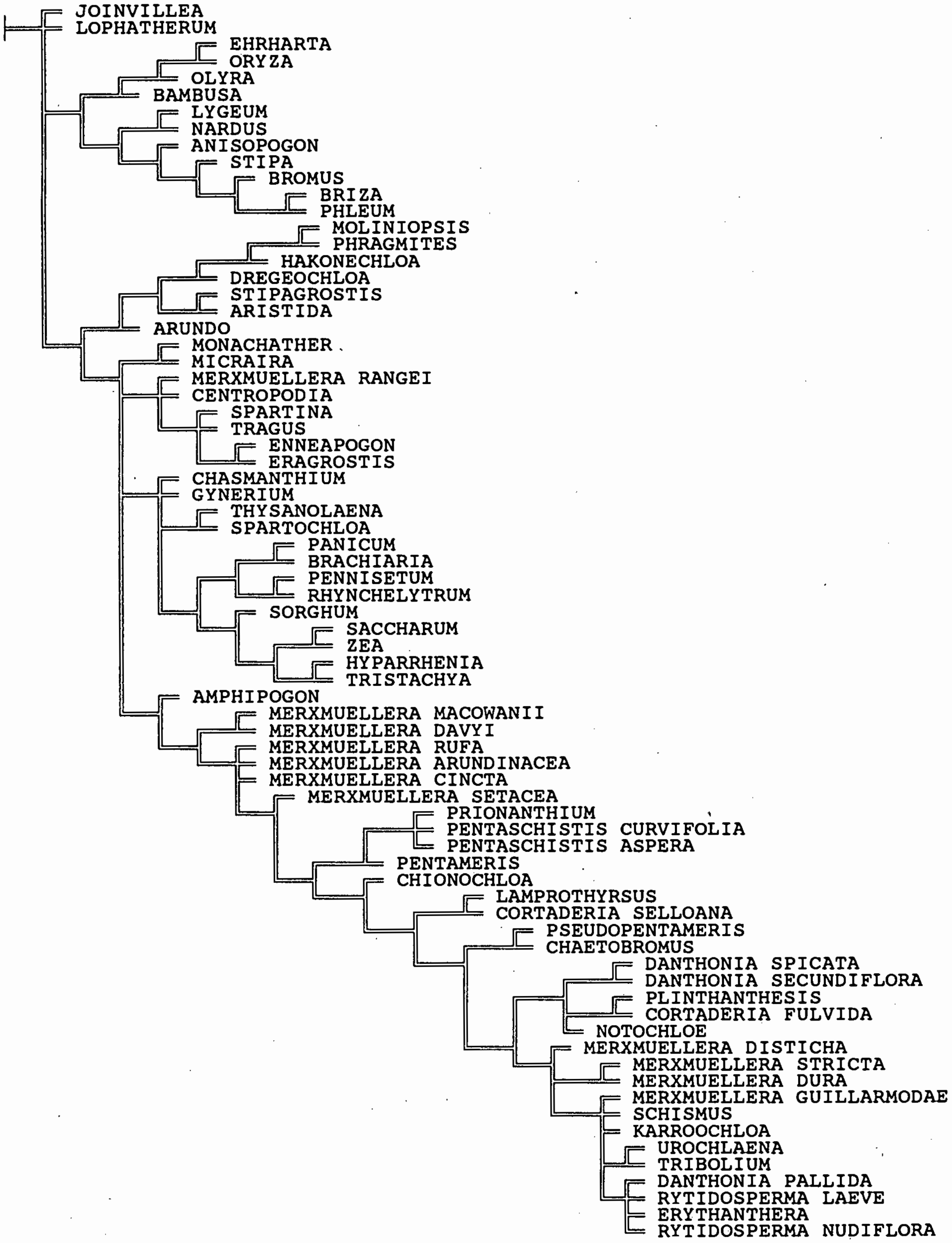


Figure 2.5. Consensus tree of 1292 (memory overflow) equally parsimonious trees obtained from a bb* analysis of 67 trees obtained from a random input order analysis carried out using PAUP. This analysis is based on the sequences aligned at stringency 1, with characters added for gaps. Length = 689, ci = 44, ri = 75.

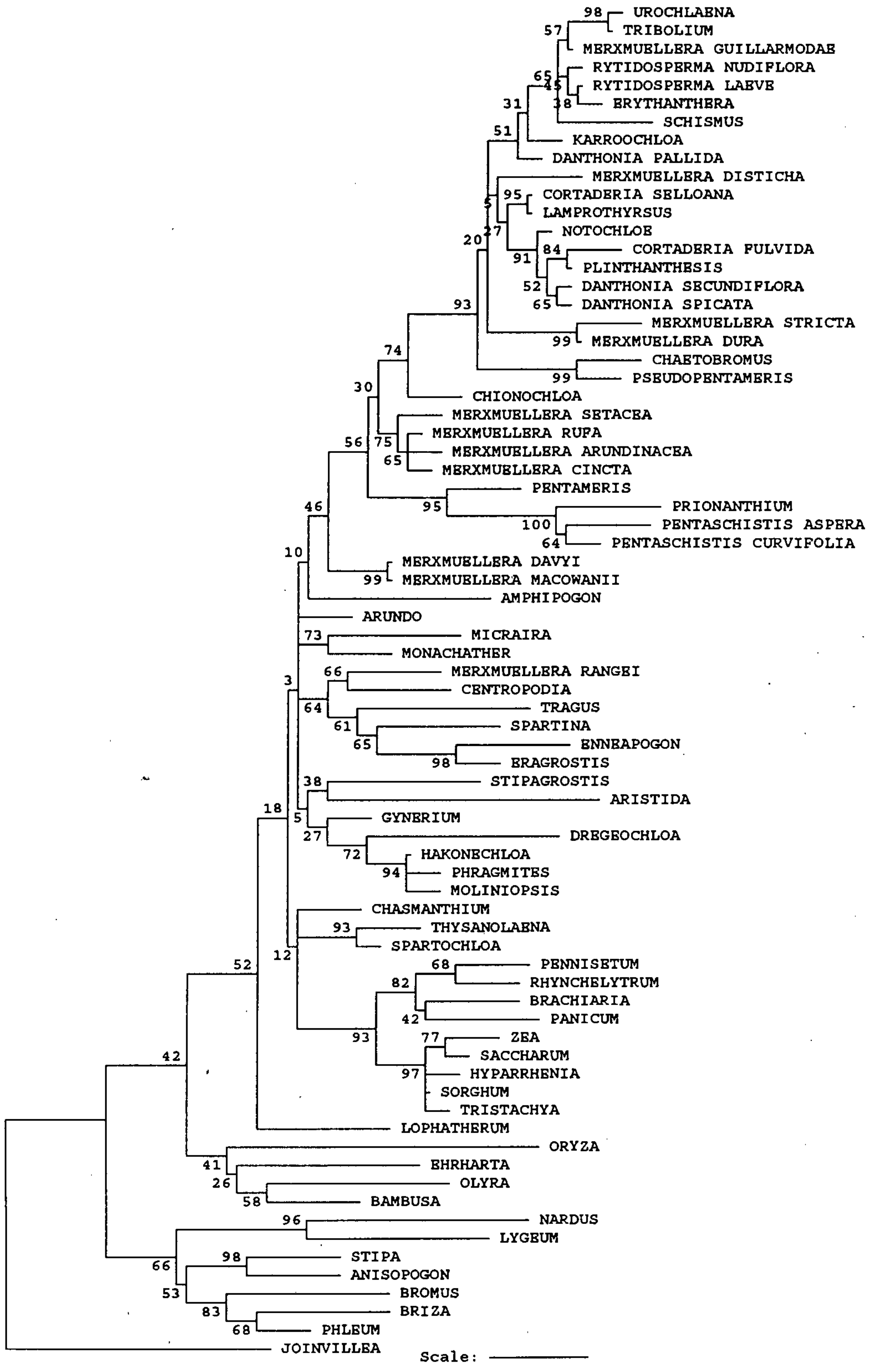


Figure 2.6. The Neighbour Joining tree derived from the nucleotide sequence data at alignment stringency 1, using the Jukes and Cantor correction. Bootstrap support values from 1000 replicates are provided. Scale bar is equal to a distance of 0.02127.

Phylogenies from alignment at stringency 2

The first iteration of the alignment process examined the data for repeats that were homoplasic. As discussed above, this is the third and coarsest level at which congruence may be applied. In examining the data set at this level, whole repeats were treated as single characters. In doing this, repeats which appear to be superficially similar but which, on the basis of the initial topology have had independent origins, are identified and re-aligned. In this way, two repeats (both affecting the representatives of the Bambusoideae) were identified as homoplasic, and the alignment corrected accordingly.

<u>Informative sites only:</u>

The alignment changes resulted in the loss of three informative sites, and a decrease in tree length of seven steps - an average of just over two steps per character. As there are four possible states, the affected sites were thus not saturated.

The topology of the consensus tree derived from the new alignment (Figure 2.7) is less resolved than that from the initial alignment. The PACC clade is lost, although the same subfamilial and tribal clades are retained. Other differences include the presence of a monophyletic lineage comprising (Arundineae, Danthonieae), and the resolution of <u>Amphipogon</u> as the basal taxon to the (<u>Monachather</u>, <u>Micraira</u>) clade.

The topology of the tree derived from the bootstrap analysis at this stringency level (Figure 2.8) differs only slightly from that obtained from the initial stringency. These differences include an unresolved Bambusoideae and re-arrangements within the danthonioid lineage. The bootstrap support for the latter lineage drops markedly (from 60.5% to 49.0%). Support for the other arundinoid lineages remains relatively unchanged at this stringency.

Informative sites and gaps:

The consensus tree from this data set is identical to that obtained from the previous alignment with gap data added, and is thus not shown. The length of this tree is only one step shorter; 688 steps (ci = 44, ri = 75).

Neighbour Joining method:

The Neighbour Joining tree from this alignment is interesting in that there are no "singleton" taxa in unusual places (Figure 2.9) (except Lophatherum which is shown as the basal taxon of the PACC clade).

The pooids, with bootstrap support of 64%, are the basal lineage. The monophyletic PACC clade (including Lophatherum) receives 52 % bootstrap support. The basal dichotomy within the PACC clade shows the (Panicoideae, Centothecoideae, Thysanolaeneae) group (abbreviated here to the PCT clade) to be sister to the chloridoid and remaining arundinoid clades. the PCT clade is, however, very poorly supported (10% bootstrap support). The chloridoid and remaining arundinoid clades form a polychotomy. Of these clades, the large danthonioid lineage (excluding Amphipogon) receives 50% bootstrap support.

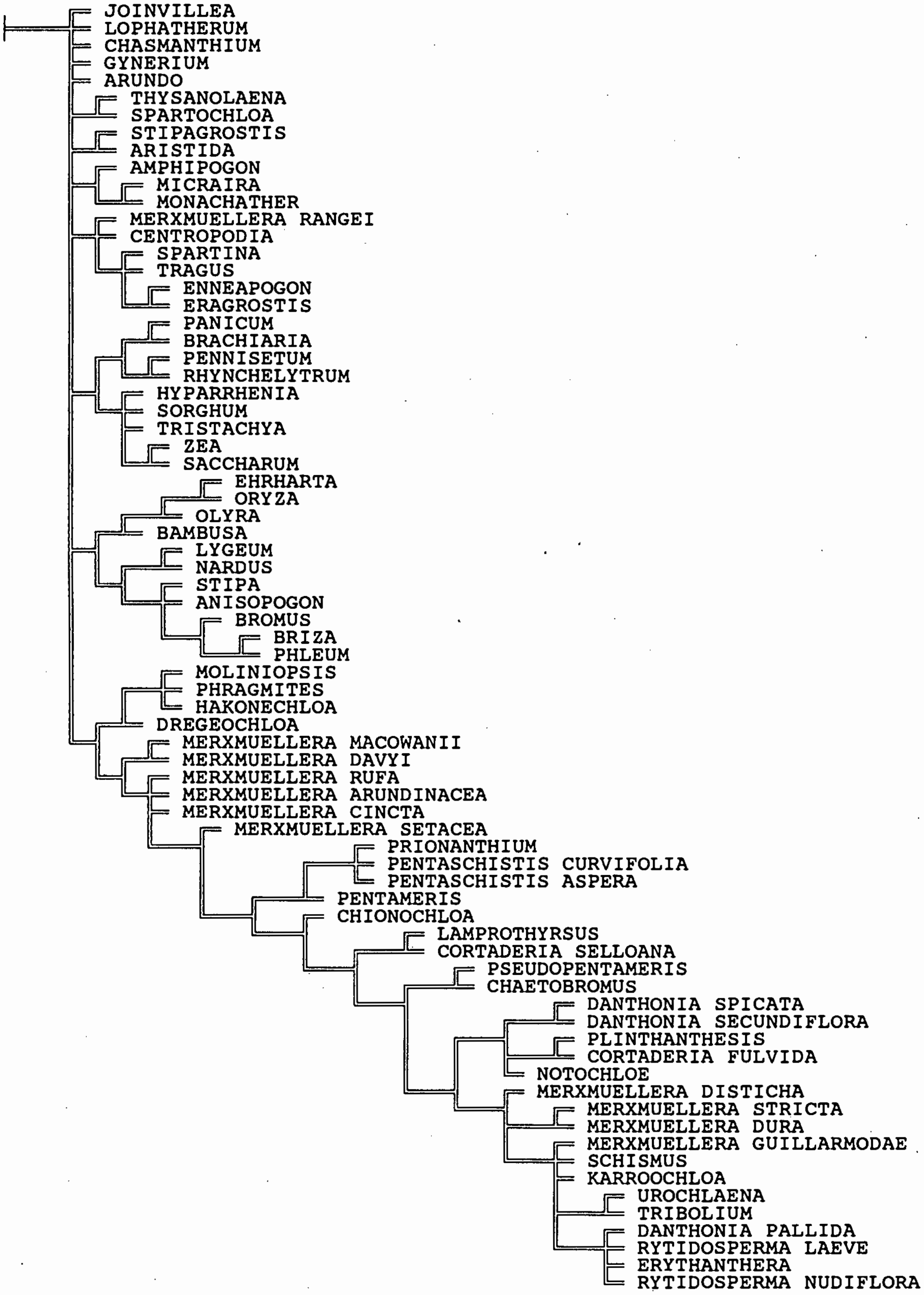


Figure 2.7. Consensus tree of the 1220 trees (memory overflow) obtained from a bb* analysis of 73 trees obtained from a random input order analysis carried out using PAUP. This analysis is based on the sequences aligned at stringency 2. Length = 606, ci = 45, ri = 75.

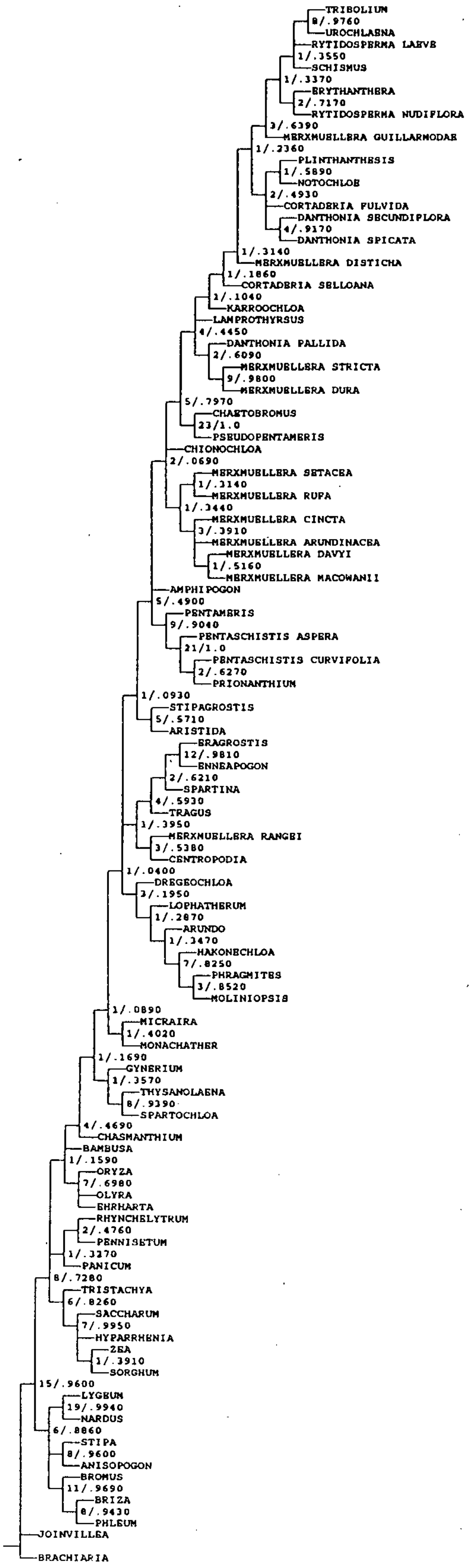


Figure 2.8. Topology showing the Bremer support (left of "/") and bootstrap support values produced from 1000 replicates of the nucleotide data aligned at stringency 2.

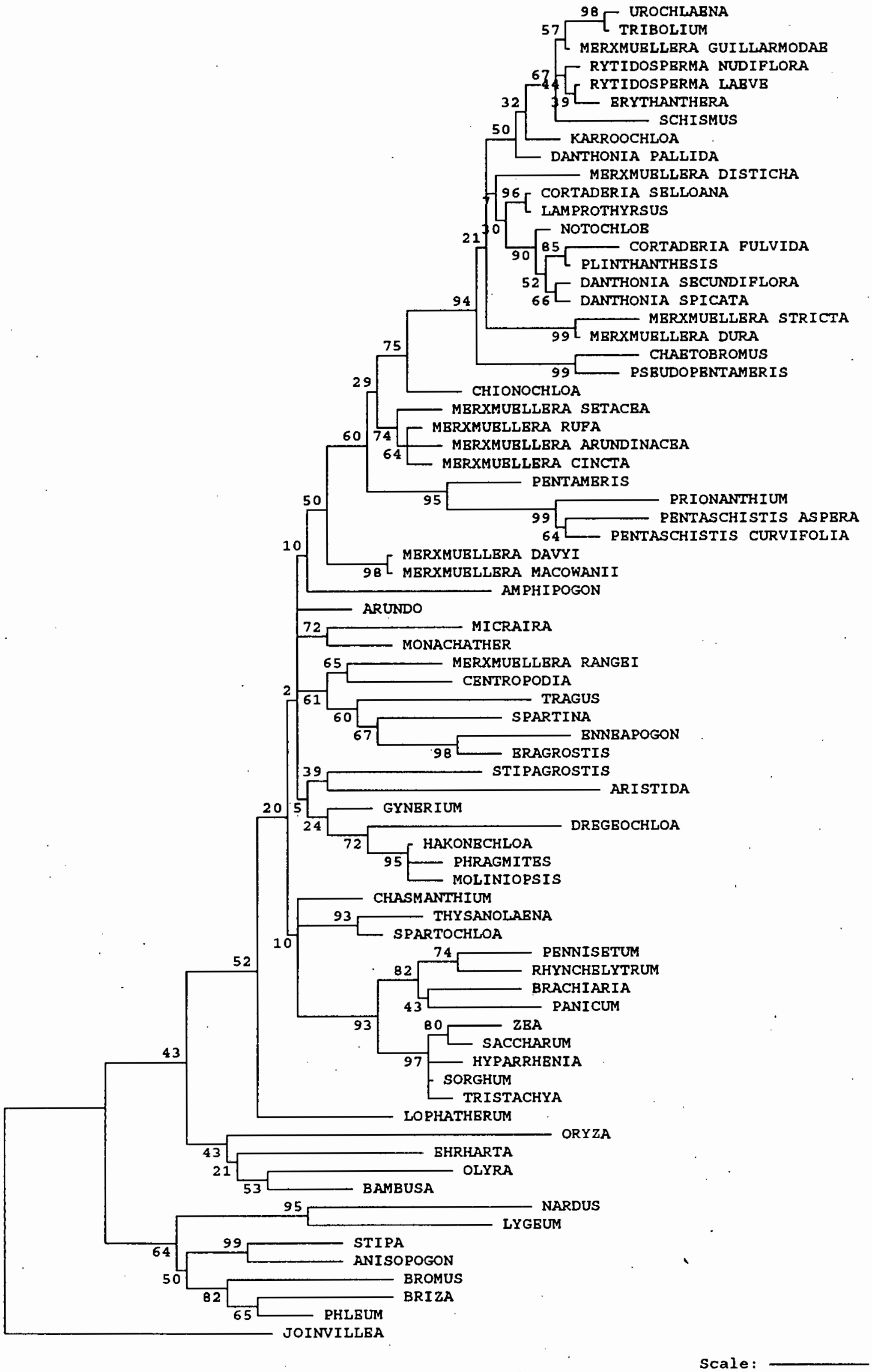


Figure 2.9. The Neighbour Joining tree derived from the nucleotide sequence data at alignment stringency 2, using the Jukes and Cantor correction. Bootstrap support values from 1000 replicate are provided. Scale bar is approximately equal to a distance of 0.02124

Phylogenies from alignment at stringency 3

In this alignment, four repeats were recognised as being misaligned on the basis of characters from those regions having low cCI values. These changes affected the Bambusoideae and Pooideae (two repeats each).

Informative sites only:

These changes resulted in a loss of six characters and a reduction in tree length of 28 steps. This equates to an average of over four steps per character, indicating that the sites from these regions were saturated and thus responsible for the homoplasy present in the previous data set. Despite this substantial reduction in tree length, these changes do not affect the ensemble ci or ri values. The consensus tree (Figure 2.10) has, however, lost some resolution at the level of the major clades; the sister group relationship between the danthonioid and arundinoid lineages is no longer supported, and these two groups are once again shown as part of the basal polytomy.

The topology of the bootstrap tree (Figure 2.11) is once again different from both the consensus tree and the other bootstrap trees. At this stringency, the PACC clade is resolved and the Bambusoideae are polyphyletic. Support for the PACC clade is strong (71.7%; Bremer support value = 6), while support for the lineages of the subfamily Arundinoideae remains weak.

Informative sites and gaps:

In the consensus tree from this analysis, the resolution of relationships among the major clades is lost, and the basal fully compensate for the loss of homoplasic, saturated sites. The resolution of the tree is improved if the weighting of all the gap characters is doubled prior to the analysis (tree not shown). Thus the loss of six apparently homoplasic characters also resulted in a loss of cladistic information for which the inclusion of four additional gap characters was unable to compensate.

Neighbour Joining method:

Unlike the Neighbour Joining tree from the previous stringency, Lophatherum once again appears in the anomalous position of the basal-most taxon in the grasses (Figure 2.13). Furthermore, the pooids (with a low bootstrap value of 21%) are nested within the bambusoids, rendering them paraphyletic. The PACC clade is retained (with 23% bootstrap support), and the relationships of the genera within the major clades approximate that obtained from other analyses, although the relationships among the larger clades is different, the tree being pectinate rather than symmetricly dichotomous. The relationships in this analysis appear as (panicoids (arundinoids (chloridoids, danthonioids))).

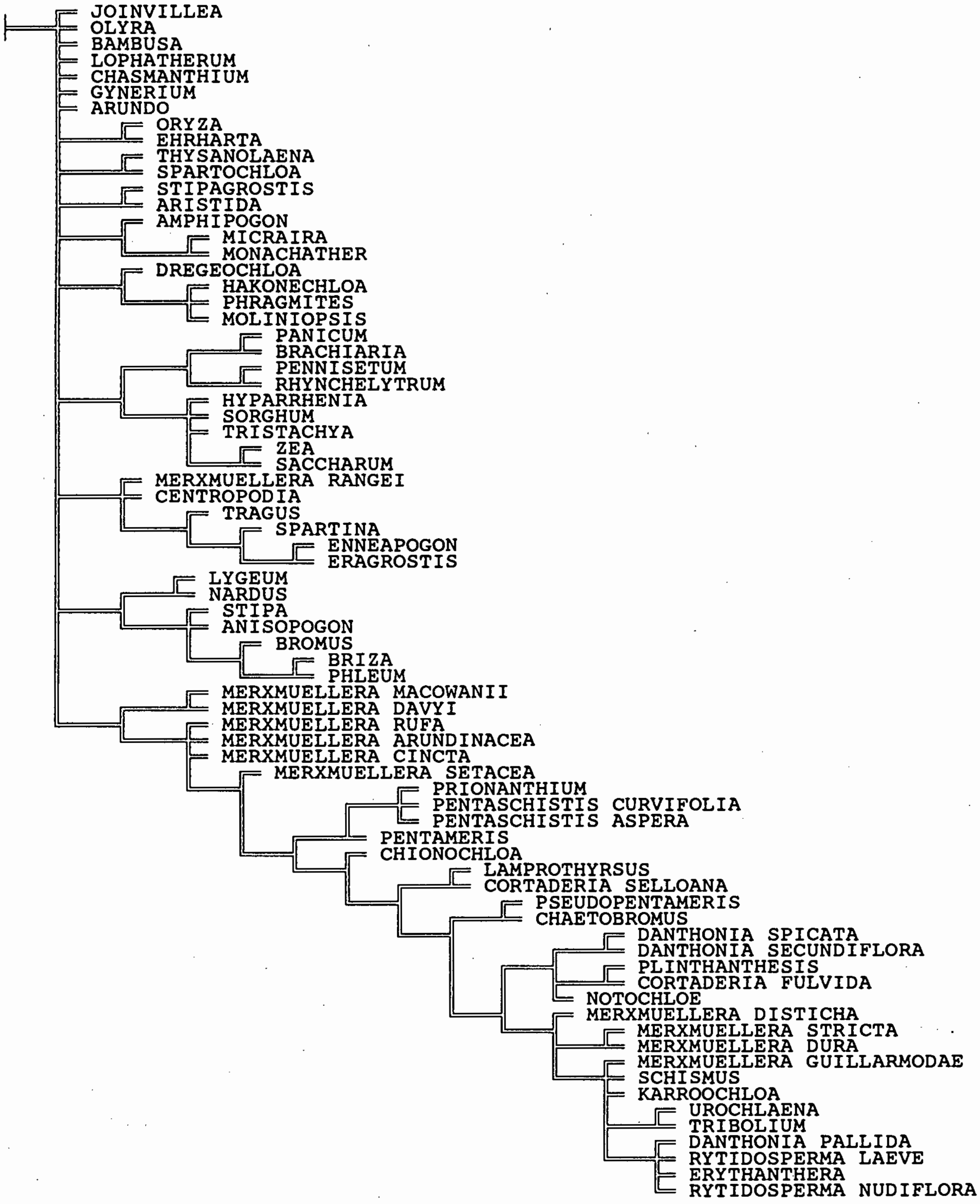


Figure 2.10. Consensus tree of 1289 trees (memory overflow) derived from an m*bb* analysis of the nucleotide sequence data at alignment stringency 3. Length = 578, ci = 45, ri = 75.

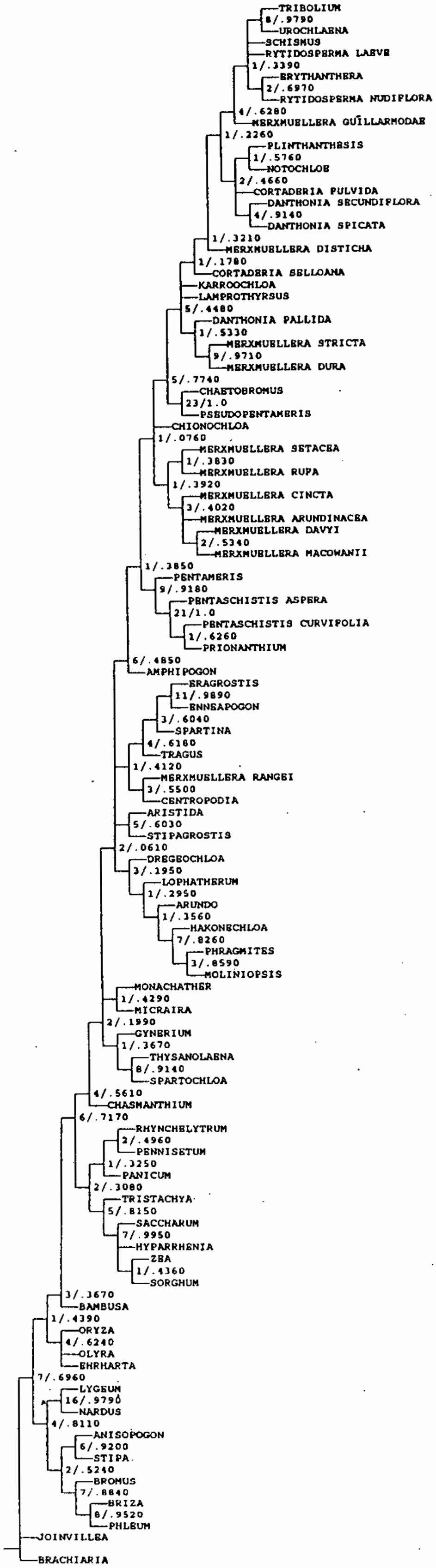


Figure 2.11. Topology showing Bremer support (left of "/") and bootstrap support values produced from 1000 replicates of the nucleotide data aligned at stringency 3.

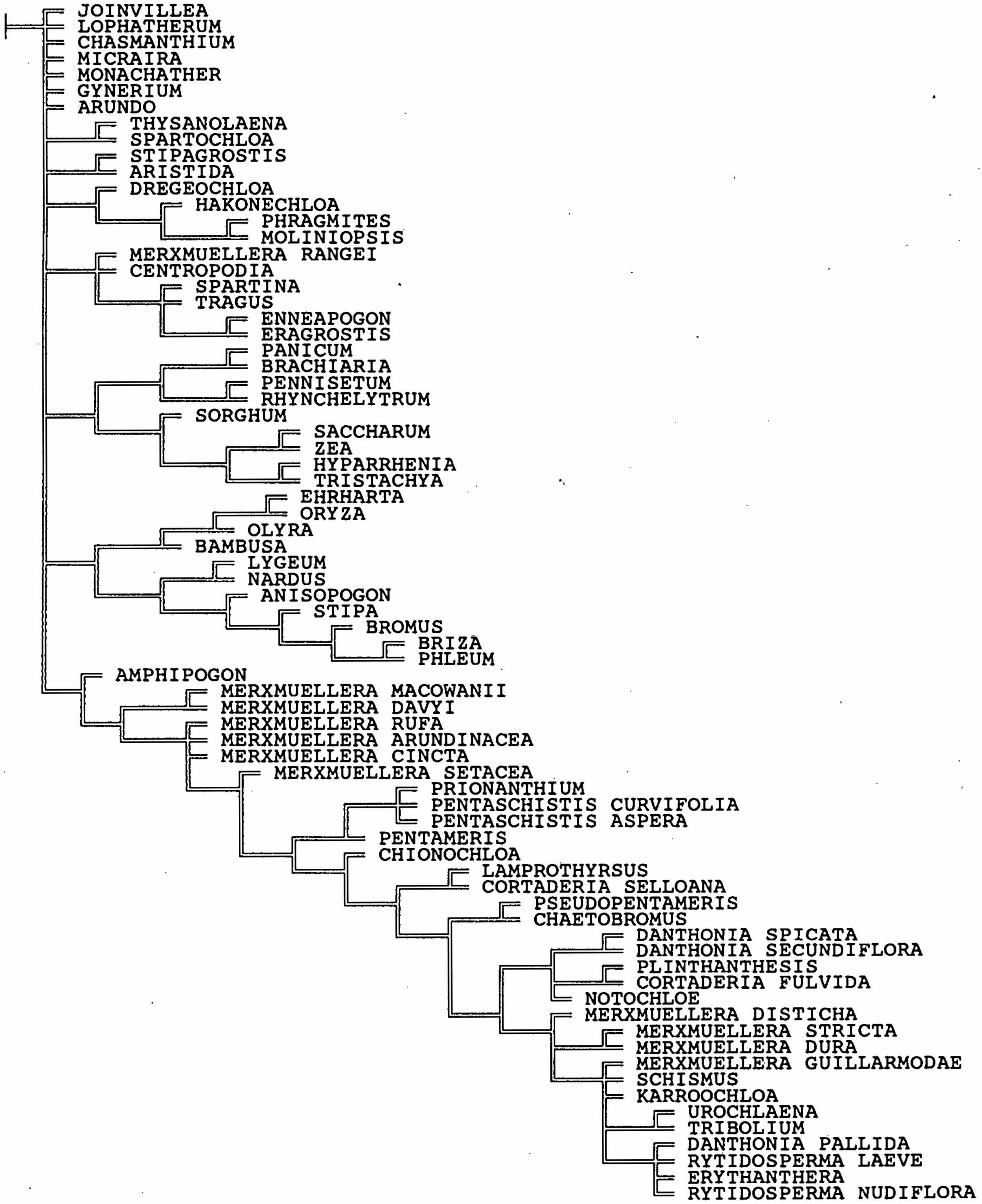


Figure 2.12. Consensus tree of 1290 (memory overflow) equally parsimonious trees obtained from a m*bb* analysis of the nucleotide sequence data at alignment stringency 3, with characters added for gaps. Length = 670, ci = 44, ri = 75.

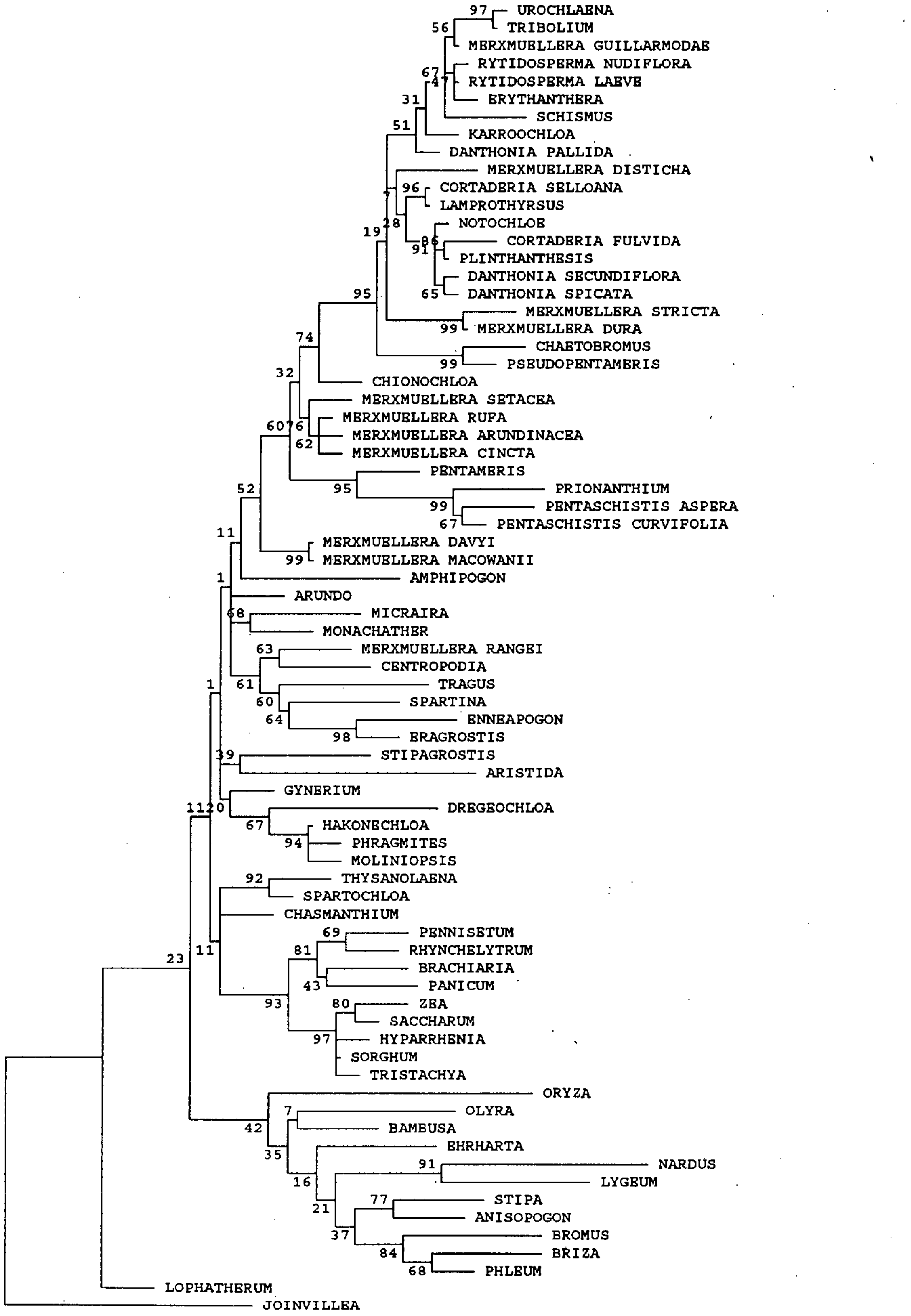


Figure 2.13. The Neighbour Joining tree derived from the nucleotide sequence data at alignment stringency 3, using the Jukes and Cantor correction. Bootstrap support values from 1000 replicates are provided. Scale bar is approximately equal to a distance of 0.02389.

Phylogenies from alignment at stringency 4

Changes made in this alignment affected only the Pentameris - Pentaschistis - Prionanthium group, and involved re-aligning four repeats.

Informative sites only:

These changes reduced the number of characters by seven, and the tree length by 21 steps - an average of three steps per lost character. These losses increase the ensemble ri by one percent. The ensemble ci remained unchanged.

The resolution of the consensus tree from the analysis of this alignment is further reduced (Figure 2.14). The main lineages remain, but some of the smaller two-taxon clades present in the tree from the previous alignment are now unresolved at the base of the tree. Within the danthonioid lineage, there is also some loss of resolution; Merxmuellera setacea becomes part of a polytomy of other Merxmuellera species, and Amphipogon, previously appearing as the basal danthonioid lineage, collapses into the basal polytomy.

The bootstrap topology is once again changed (Figure 2.15). In this topology, the herbaceous bambusoid taxa now appear as the basal clade (with the exception of Brachiaria, explained above). The PACC clade is still retained with good bootstrap support (79.2%; Bremer support value = 9), but the danthonioid lineage is polyphyletic and fragmented into three separate clades (a Pentaschistis - Pentameris - Prionanthium clade, a Chionochloa - basal Merxmuellera clade, and a clade comprising the remainder of the danthonioid taxa.

Informative sites and gaps:

The topology of this consensus tree (Figure 2.16) differs from the tree from the nucleotide data set in only one instance:

Amphipogon is lifted out of the basal polytomy into the basal position of the danthonioid lineage.

Neighbour Joining method:

The topology of the Neighbour Joining tree (Figure 2.17) is much changed from previous trees. As in the previous alignment, Lophatherum appears as the basal grass taxon, followed by Amphipogon and then the Aristideae. Other than these taxa, the relationships of the genera within the major lineages are almost unchanged, but the relationships of the major lineages have once again been shuffled. The pooids, with poor bootstrap support (22%), are once again embedded within the bambusoids, and these taxa are now shown as sister to the danthonioids, which as a clade are well supported (66% bootstrap support). The PACC clade is thus lost in this topology.

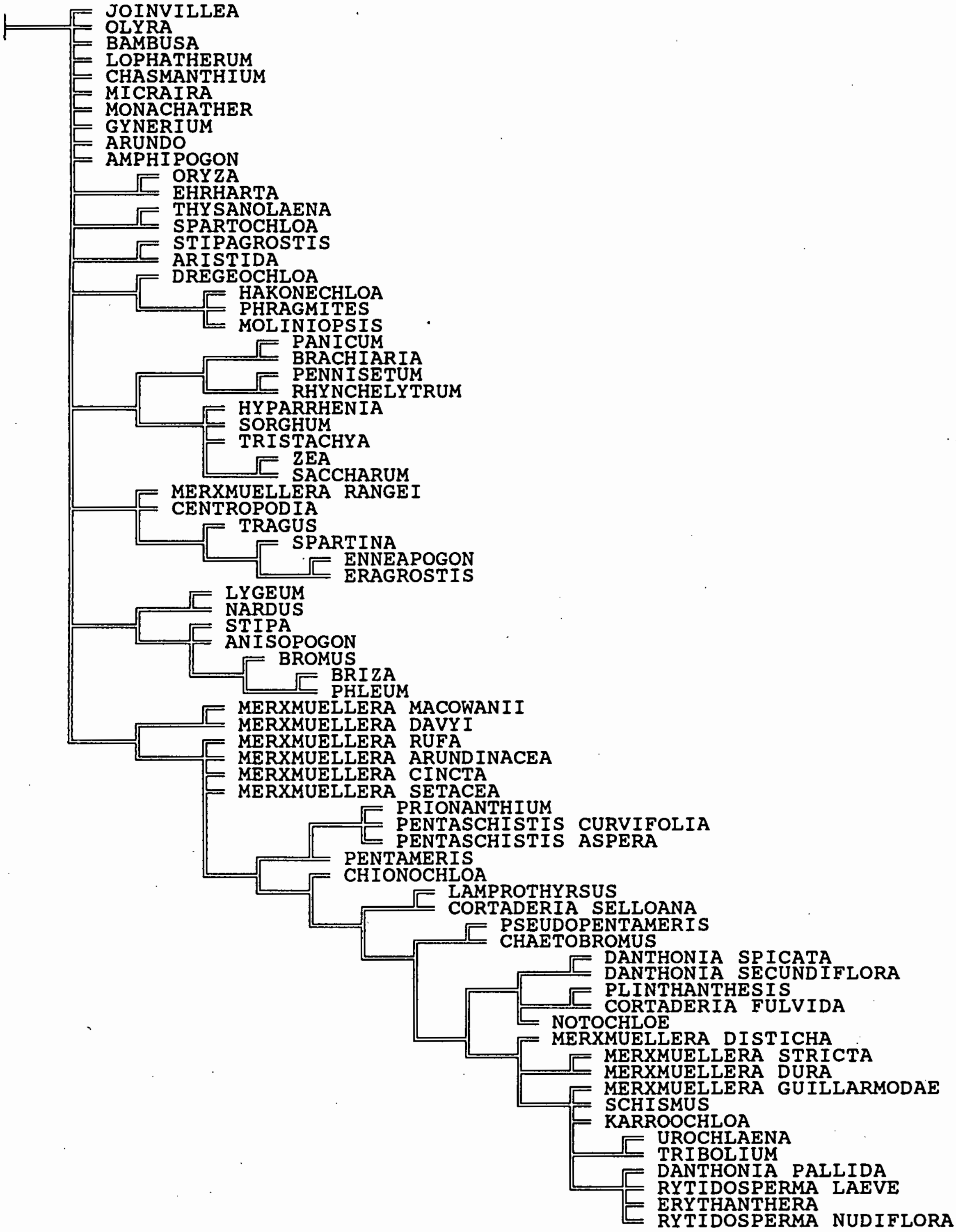


Figure 2.14. The consensus tree from 1285 trees (memory overflow) derived from a m*bb* analysis of the nucleotide sequence data at alignment at stringency 4. Length = 551, ci = 45, ri = 76.

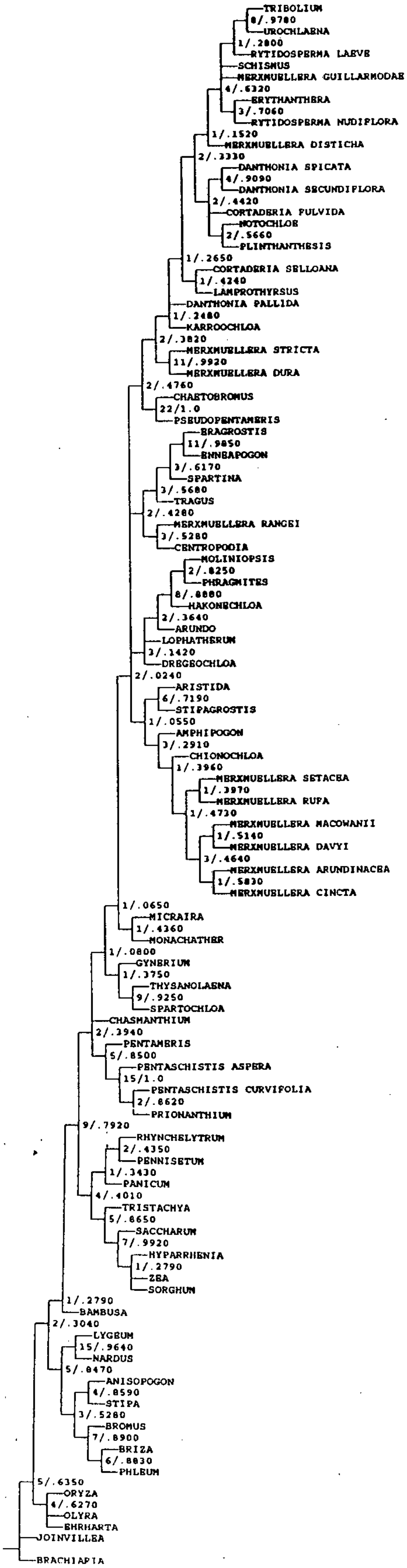


Figure 2.15. Topology showing Bremer support (left of "/") and bootstrap support values produced from 1000 replicates of the nucleotide data aligned at stringency 4.

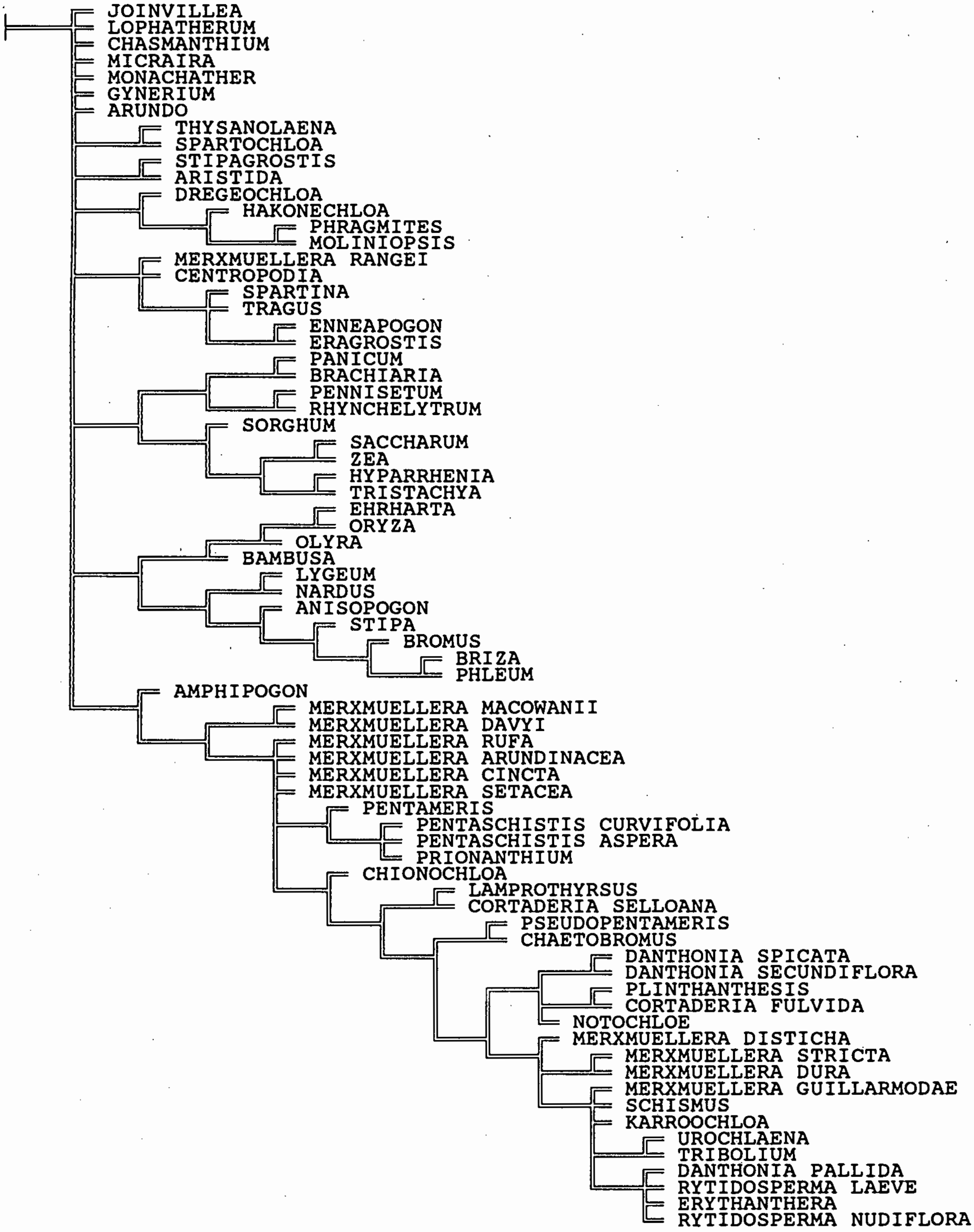


Figure 2.16. Consensus tree of 1286 (memory overflow) equally parsimonious trees obtained from a m*bb* analysis of the nucleotide sequence data at alignment stringency 4, with characters added for gaps. Length = 652, ci = 44, ri = 76.

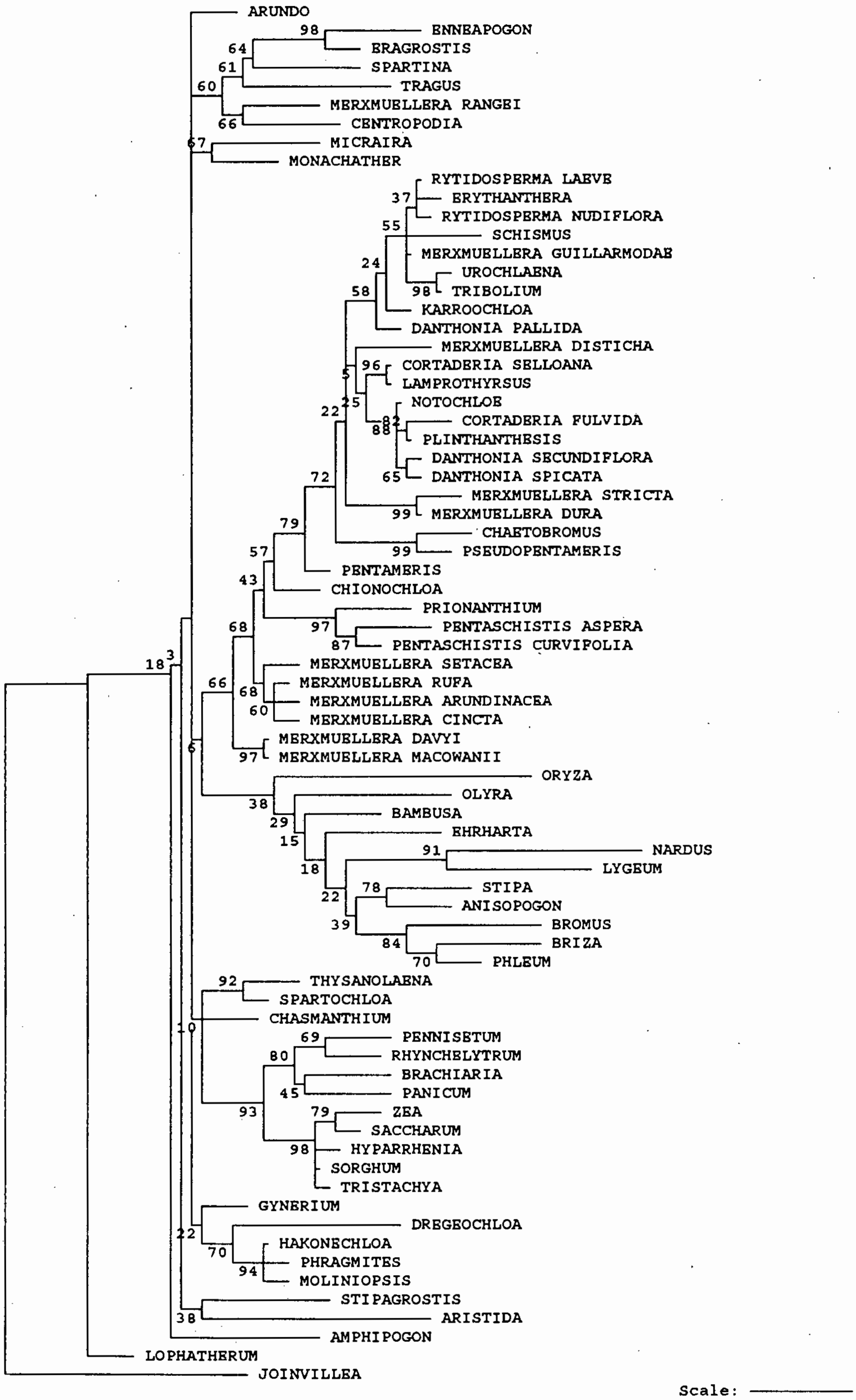


Figure 2.17. The Neighbour Joining tree derived from the nucleotide sequence data at alignment stringency 4 using the Jukes and Cantor correction. Bootstrap support values from 1000 replicates are provided. Scale is approximately equal to a distance of 0.02622.

Phylogenies from alignment at stringency 5

Changes made to this alignment affected three blocks of repeats; one each from the panicoid lineage (including the centothecoid and thysanolaenoid taxa), the pooid clade and a group of four taxa comprising a "phragmitoid" clade (Phragmites, Moliniopsis, Hakonechloa and Dregeochloa).

Two of the repeat blocks that were re-aligned at this stringency came from a region that, in the initial alignments, had sequences from all except one of the taxa sampled. At the initial alignment, seven of the 10 characters derived from this region had a CCI of less than 33, and none of the 10 characters had high CRI values. Bambusoid elements from this region were initially re-aligned in the alignment at stringency 3. Thus only through multiple re-alignment events can the homoplasies from this region be removed from the data. This region is even further expanded in the next alignment level.

Informative sites:

The alignment changes resulted in the loss of only one informative site, but a massive reduction in tree length of 26 steps, suggesting that consistency of both the lost character and those remaining were substantially improved by the re-alignment. These changes resulted in a two percent increase in the ensemble ci, and a one percent increase in the ensemble ri.

In the consensus tree from this analysis (Figure 2.18), the PACC clade is once again resolved. The resolution of the basal polytomy that was present in the tree from the previous alignment has also improved; the individual genera (with the exception of Lophatherum, Bambusa and Olyra) are once more placed within various clades. Amphipogon again appears as the basal taxon in the danthonioid lineage, and the arundinoid (reeds), centothecoid and thysanolaenoid taxa are all shown to be associated with, and basal to, the Panicoideae.

The bootstrap topology (Figure 2.19) at this stringency reverts once again to showing the pooid clade as basal, and the bambusoids as paraphyletic. The PACC clade is retained (72.4% bootstrap support; Bremer support value = 6), as is the fragmentation of the danthonioid clade obtained in the previous stringency.

Informative sites and gaps:

The consensus tree from this analysis is slightly less resolved than the tree from the analysis using nucleotide data only (Figure 2.20). The differences include the collapse of Amphipogon and Dregeochloa into the basal polychotomy from their basal positions in the danthonioid and arundinoid lineages respectively. Furthermore, the two species of Pentaschistis and a representative of Prionanthium form a polychotomy.

Neighbour Joining method:

The tree from this alignment (Figure 2.21) still shows Lophatherum and Amphipogon as the basal two taxa. Also unchanged is the relationship of the pooid taxa (embedded within the bambusoid clade), but this clade is now shown to be sister to the Panicoideae. The PACC clade is thus not retrieved. The danthonioid lineage once again receives reasonable bootstrap support (65%).

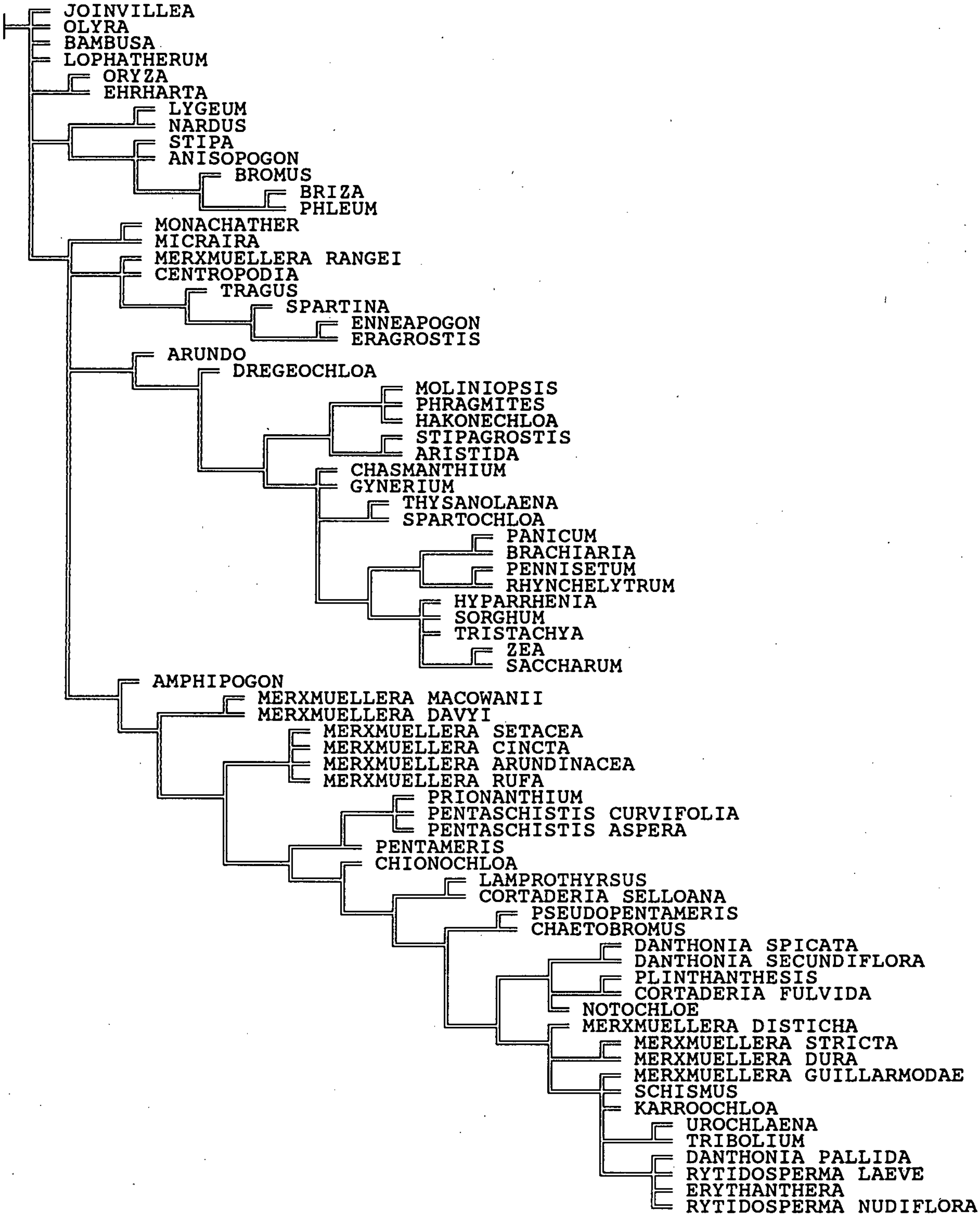


Figure 2.18. Consensus tree of the 1282 trees (memory overflow) derived from a m*bb* analysis of the nucleotide sequence data at alignment stringency 5. Length = 525, ci = 47, ri = 77.

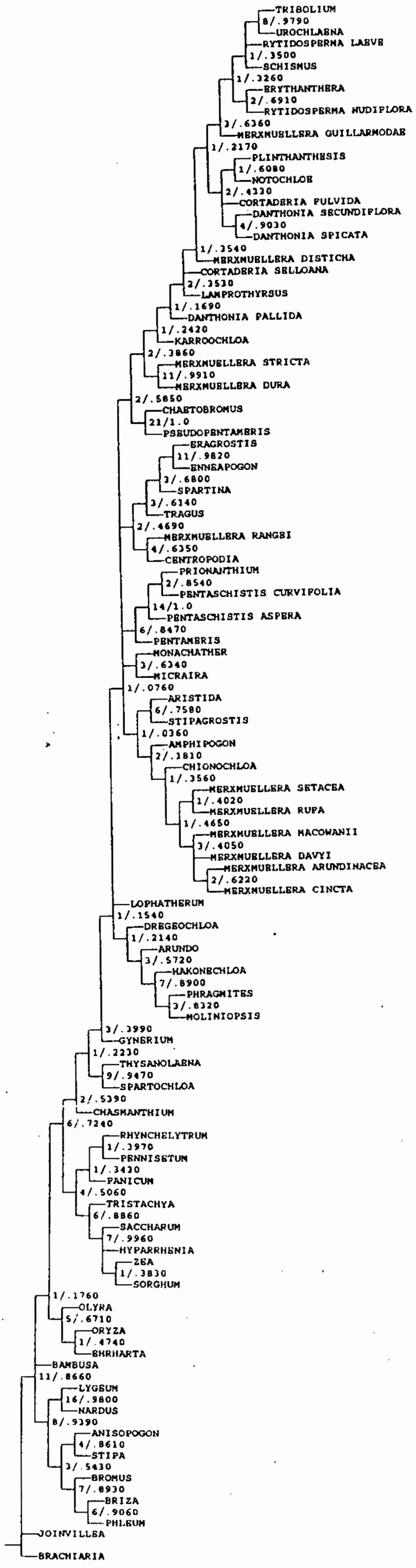


Figure 2.19. Topology showing Bremer support (left of "/") and bootstrap support values produced from 1000 replicates of the nucleotide data set aligned at stringency 5.

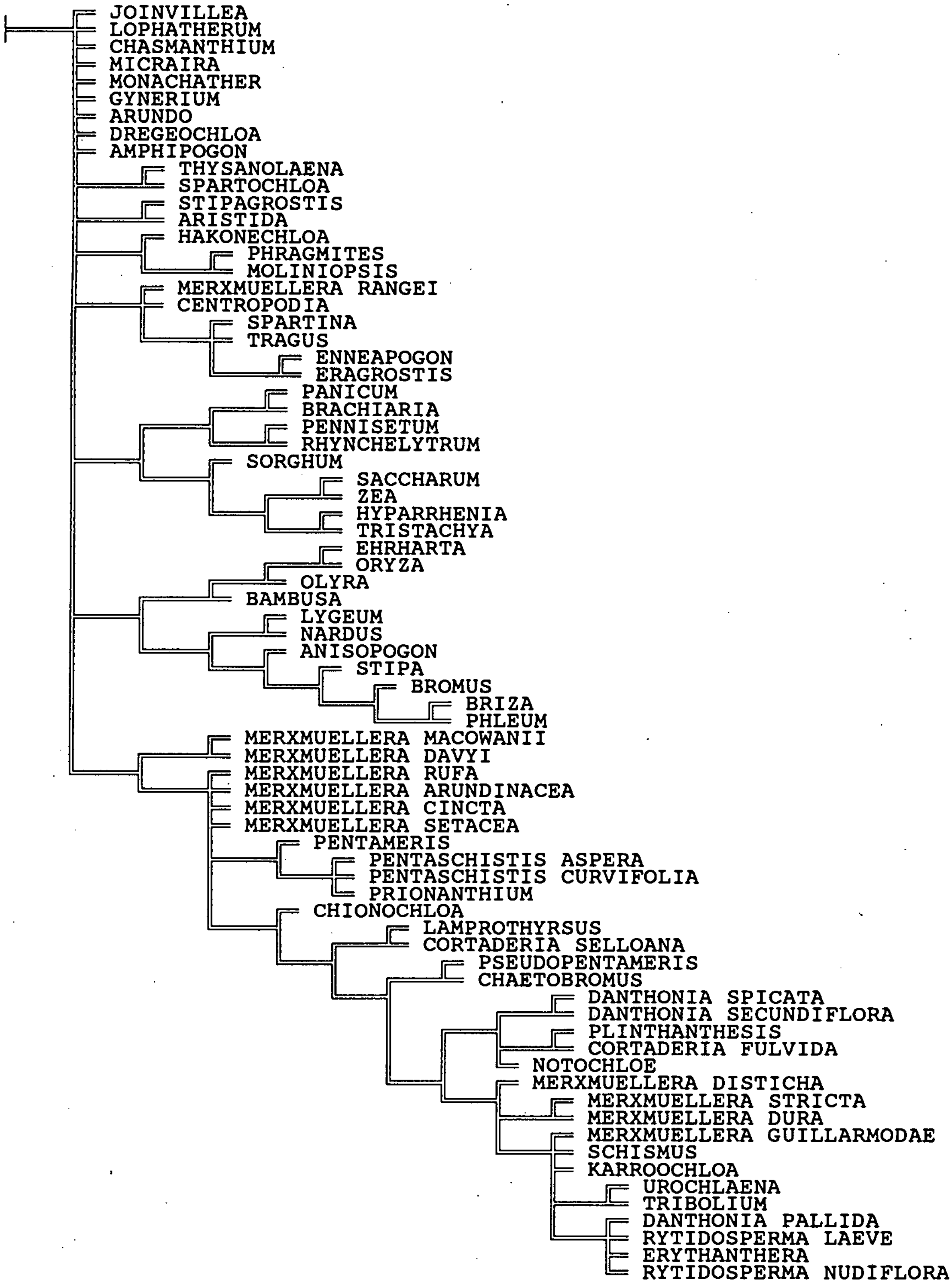


Figure 2.20. Consensus tree of 1289 (memory overflow) equally parsimonious trees obtained from a m*bb* analysis of the nucleotide sequence data at alignment stringency 5, with characters added for gaps. Length = 633, ci = 45, ri = 77.

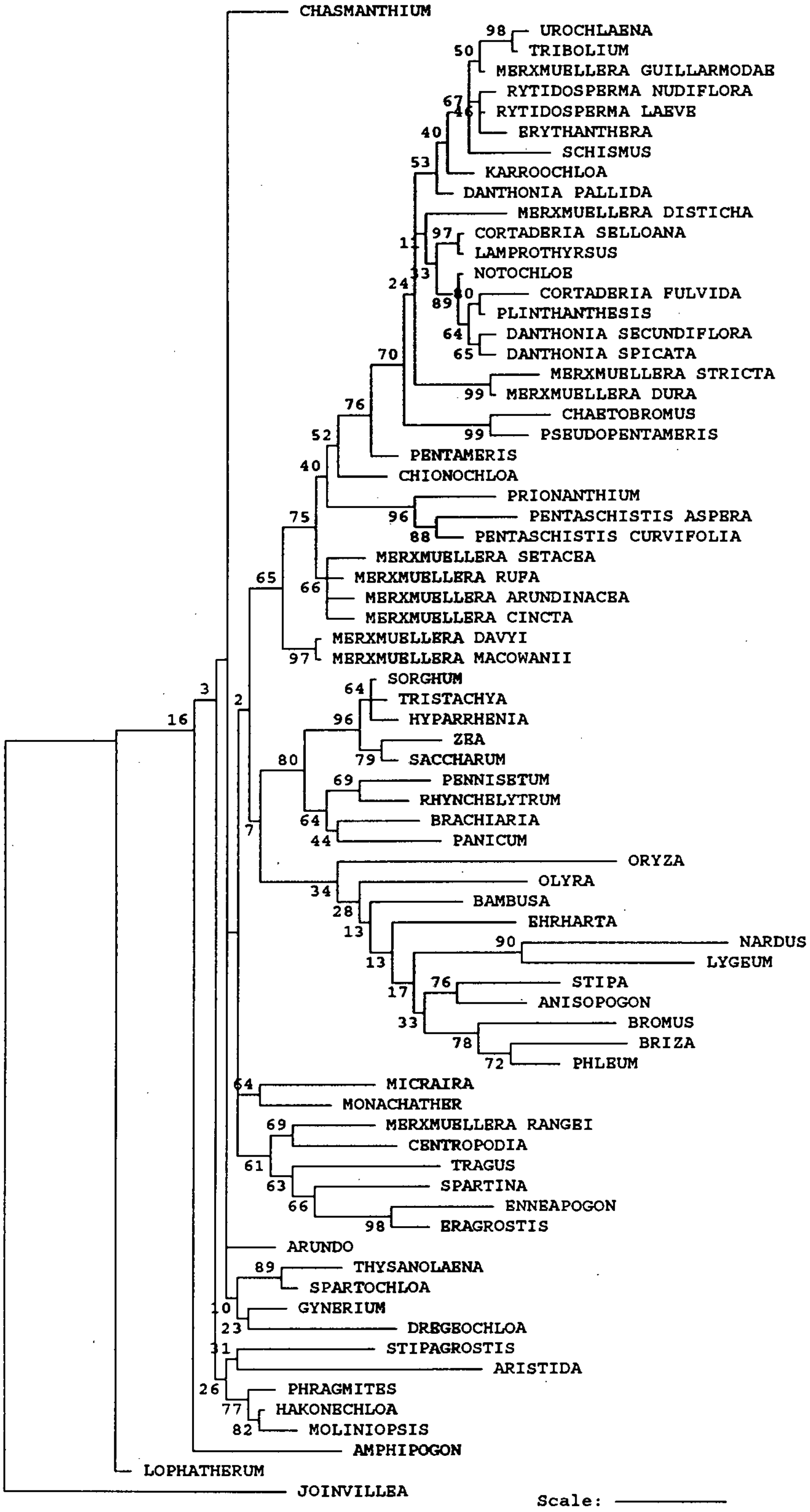


Figure 2.21. The Neighbour Joining tree derived from the nucleotide sequence data at alignment stringency 5 using the Jukes and Cantor correction. Bootstrap support values from 1000 replicates are provided. Scale is approximately equal to a distance of 0.02616.

Phylogenies from alignment at stringency 6

Changes brought about in this alignment affect all or part of the danthonioid lineage. Three groups of repeats are re-aligned.

Informative sites:

These changes reduced the data set by a further four characters. The loss of these characters shortened the tree length by 20 steps, indicating once again that the sites in these regions were homoplasic. The ensemble ri of the trees derived from this alignment has, however, decreased by three percent. This is a result of the loss of clade-specific states that were present in those homoplasic characters which the alignment changes removed from the data. It can be hypothesised that trees from further alignments at yet higher stringencies will have an even more reduced ri.

The relationships among the major clades and the reasonable resolution of these remains unchanged from that shown in the trees from the previous alignments. However, the resolution of the terminals within the danthonioid lineage has decreased, with the presence of a polytomy (Figure 2.22). As the alignment modifications affected these taxa, the alignment changes affected here are in fact detrimental to the information content of the data set. The stringency of the homology assessments of the repeats affected by this alignment is thus too strict, and a loss of characters that would have been consistent within the danthonioid clade has resulted. This loss is also reflected in the lower ensemble retention index.

The bootstrap topology (Figure 2.23) shows the herbaceous bambusoid taxa to be the basal clade. Support for the PACC clade is markedly reduced to 56.2% (Bremer support value = 3), while the danthonioid clade is once again shown to be monophyletic, but with a low bootstrap support of 25.5% (Bremer support value = 2). Other changes in this topology include the removal of Dregeochloa from the arundinoid clade to a position basal to the (Aristideae, Danthonieae) clade and Gynerium, now shown to be associated in a polychotomy with the arundinoid taxa.

Informative sites and gaps:

The topology of the consensus tree from this analysis (Figure 2.24) is less resolved than in previous alignments, the lower portions of the danthonioid lineage collapsing into a polychotomy. This topology is even less resolved than the topology based on nucleotide characters only. This indicates that the addition of single characters coding for the added gaps is ineffective in retaining the information lost in the alignment changes at this extreme alignment stringency.

Neighbour Joining method:

The topology of this tree is once again different from those from previous alignments (Figure 2.25). The basal grass clade is now shown to be (Nardus, Lygeum), followed by Ehrharta. The other bambusoid taxa are split, with some placed basal to the remaining pooid taxa, while Oryza is basal to the PACC clade. Within the poorly supported PACC clade (47% bootstrap support) there is a basal dichotomy between the danthonioid lineage (39% bootstrap support) and the remaining PACC clade taxa (4% bootstrap support). Lophatherum and Amphipogon are placed as the basal lineages of this latter clade.

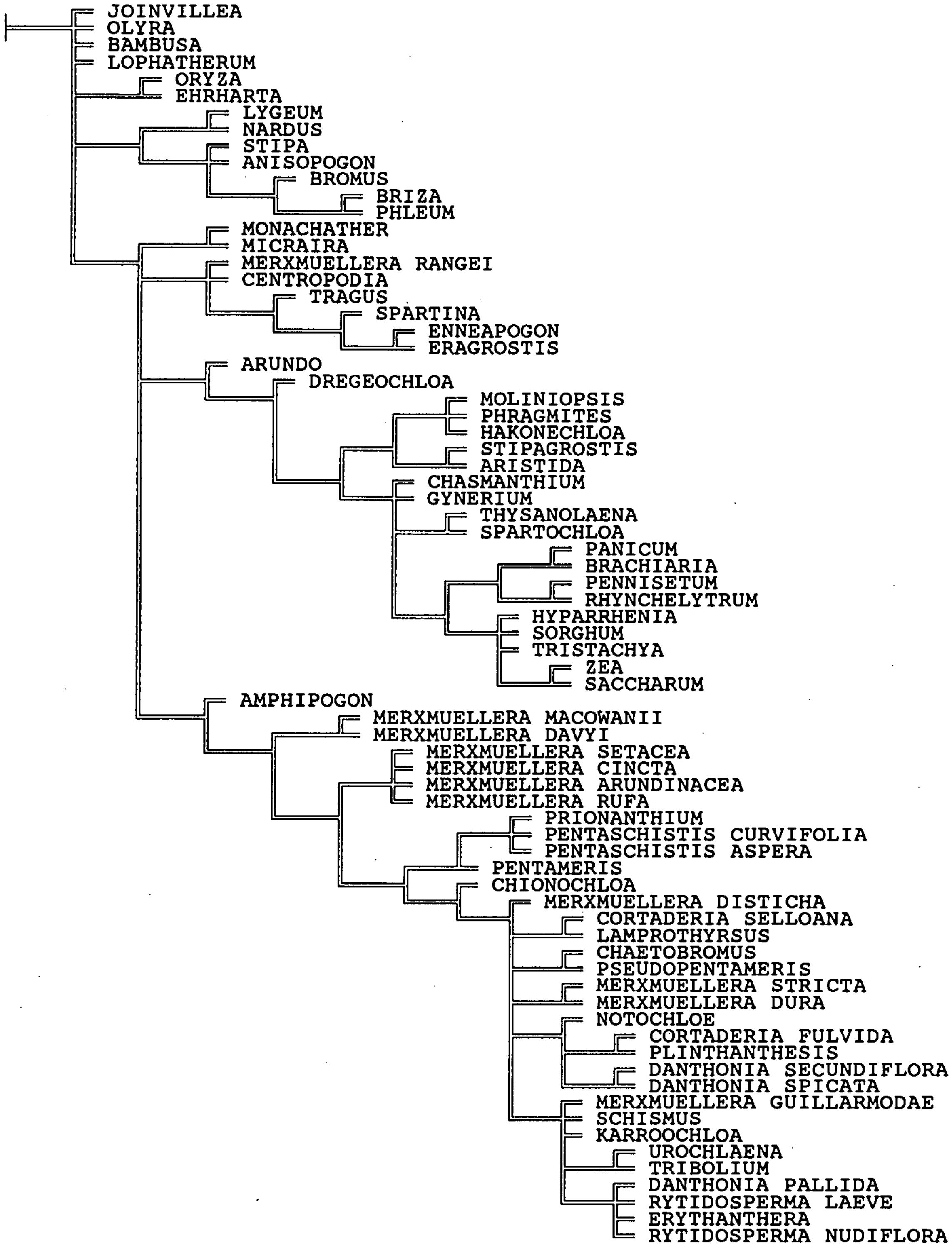


Figure 2.22. Consensus tree of the 1283 trees (memory overflow) derived from a m*bb* analysis of the nucleotide sequence data at alignment stringency 6. Length = 504, ci = 47, ri = 74.

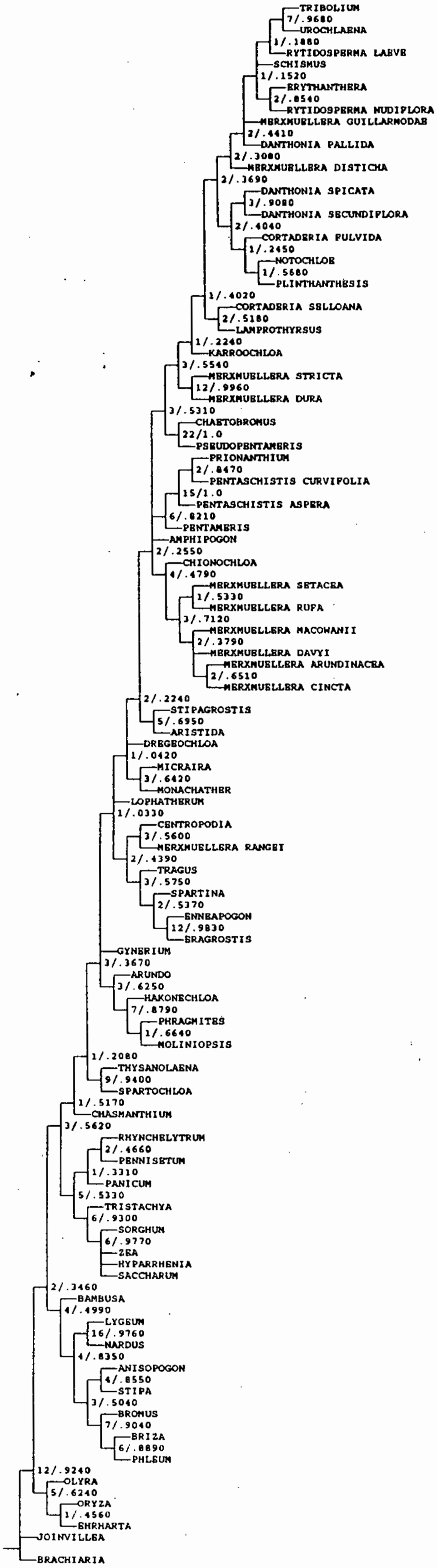


Figure 2.23. Topology showing Bremer support (left of "/") and bootstrap support values produced from 1000 replicates of the data set aligned at stringency 6.

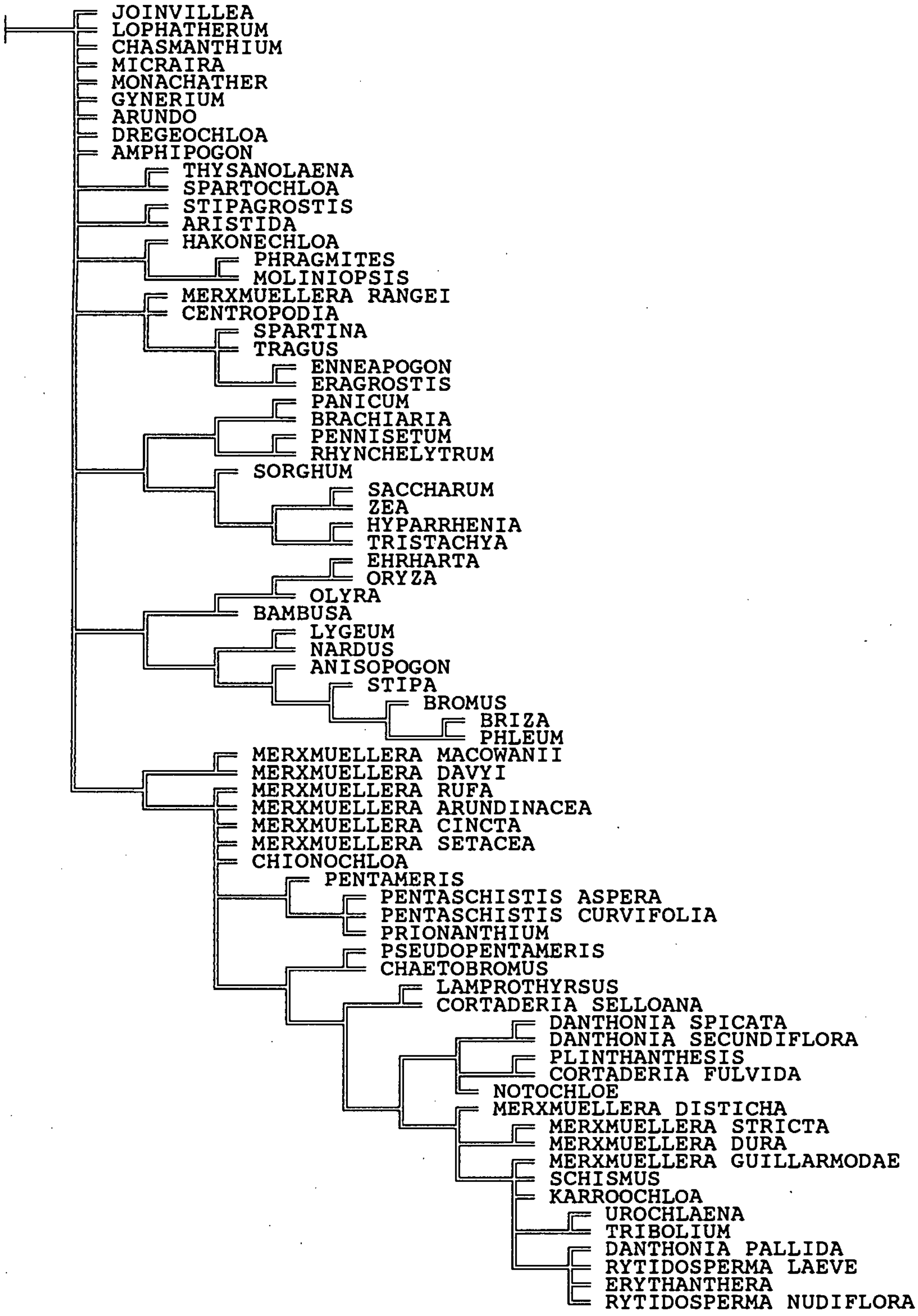


Figure 2.24. Consensus tree of the 1288 (memory overflow) equally parsimonious trees derived from a m*bb* analysis of the nucleotide sequence data at alignment stringency 6, with characters added for gaps. Length = 619, ci = 45, ri = 76.

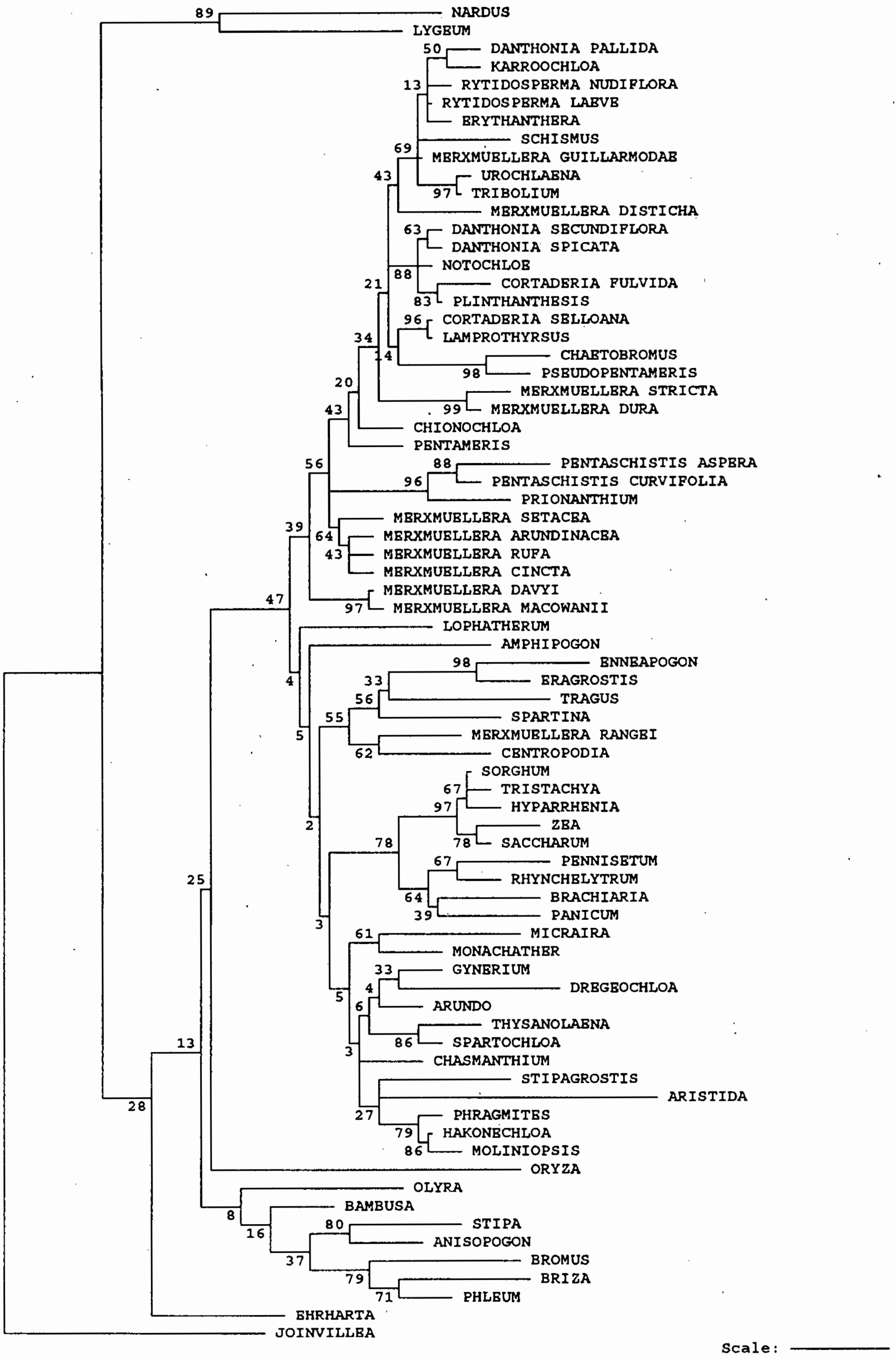


Figure 2.25. The Neighbour Joining tree derived from nucleotide sequence data st alignment stringency 6 using the Jukes and Cantor correction. Bootstrap support values from 1000 replicates are provided. Scale bar is approximatly equal to a distance of 0.0213.

CLUSTAL alignments

These alignments, while being of a comparatively short total length (i.e. the cumulative length of the gaps was short), were highly fragmented. The gaps the program inserted were both large and small, and these were placed such that portions of the sequence were fragmented. Some of these fragments were as small as four base-pairs. The heptameric repeat pattern was not conserved. Furthermore, gaps (implying insertions in some of the sequences) were placed within codons. While it is biologically possible for the slipped strand mispairing process to disrupt codons, it is unlikely that such disruptions would result in functional gene unless the interrupted and modified codon coded for an amino acid with similar characteristics (charge etc.) to the previous (unmodified) codon.

CLUSTAL, and probably many other computerised alignment programs, produce alignments that are not "contextual"; the program does not recognise heptameric repeats, or even codons. The program does, however, produce alignments according to sequence similarity. The results of the phylogenetic analyses of these alignments are discussed below.

CLUSTAL alignment at gap penalty = 25

This alignment resulted in 264 informative sites, 58 more than obtained from the least stringent intuitive alignment. The ensemble ci is slightly less than that obtained from any trees using the successive stringency alignment method (0.43 versus 0.45), but the ensemble ri is substantially higher (0.81 versus 0.77).

The consensus tree from the analysis of these data is shown in Figure 2.26. This tree is less resolved than any of the topologies obtained using intuitive alignment methods. The presence of the major subfamilial clades is similar to that shown in trees obtained using intuitive alignment methods, although there are some notable differences in their composition. The panicoid clade lacks Panicum. The danthonioid clade comprises only some of the taxa included in it in the trees from the alignments at all stringencies, the basal danthonioid taxa collapsing into the basal polytomy. Furthermore, many of the smaller clades, including some of those corresponding to the smaller arundinoid tribes, have collapsed into the basal polytomy. The pooid taxa are polyphyletic, and embedded within the bambusoid clade. This topology is less resolved, but only in a few places incongruent with those previously discussed.

The topology of tree produced by the bootstrapping process (Figure 2.27) also contains some anomalies. As in the bootstrap topologies obtained from the intuitive alignments, Brachiaria is (for the same reasons) retained as the basal taxon. The bambusoid taxa are scattered, and Bambusa is shown to be basal to the (Pseudopentameris, Chaetobromus) clade, which forms part of the danthonioid lineage. The PACC clade is thus not retrieved. Furthermore, elements of the chloridoid and panincoid clades (such as Tragus and Panicum) are placed in other lineages. Of the subfamily clades, only the Pooideae remain intact, with high bootstrap support (98.6%; Bremer support value = 17).

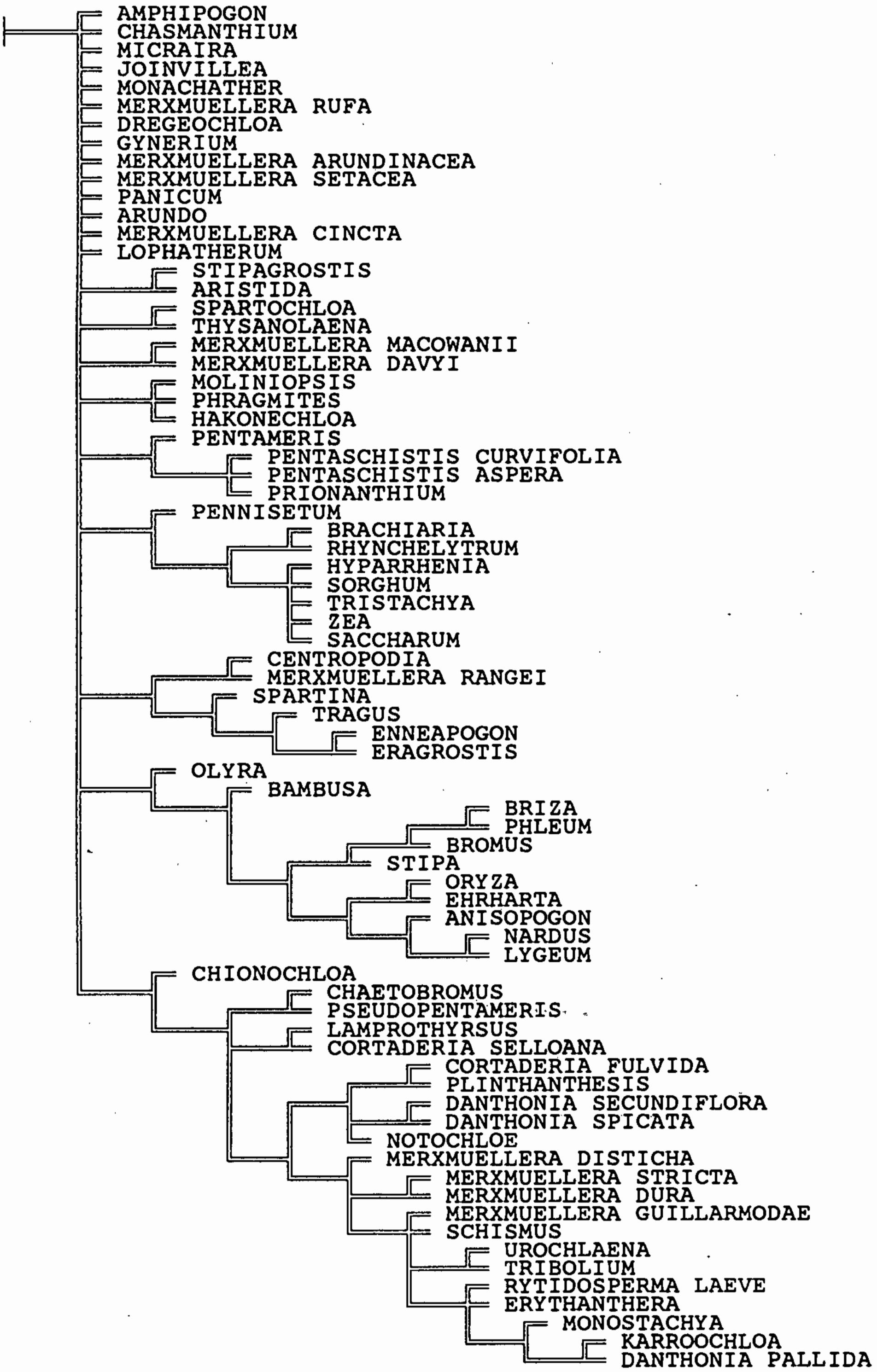


Figure 2.26. The consensus tree of the 1284 (memory overflow) equally parsimonious trees obtained from a m*bb* analysis of the nucleotide sequence data aligned using CLUSTAL, with the gap penalty set at 25. Length = 865, ci = 43, ri = 81.

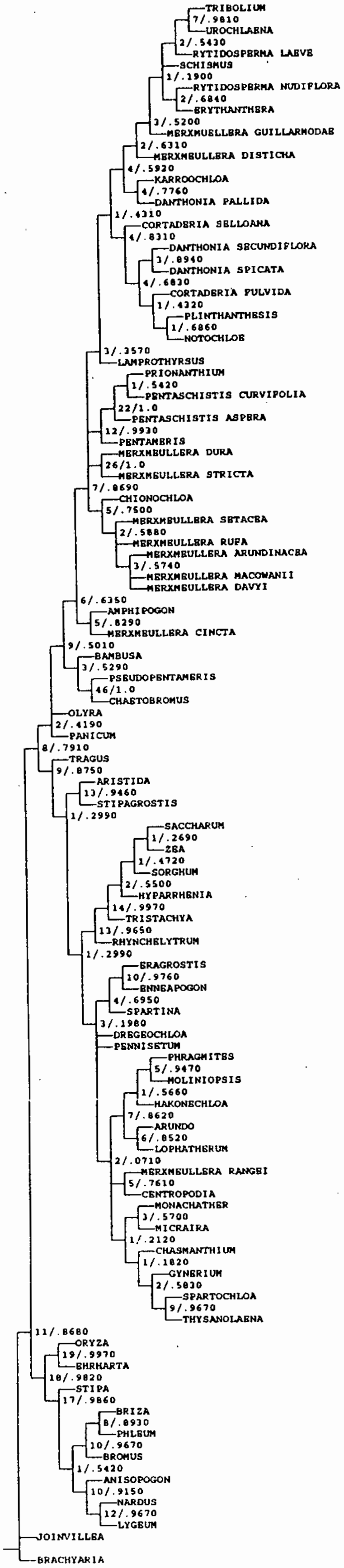


Figure 2.27. Topology showing Bremer support (left of "/") and bootstrap support values produced from 1000 replicates of the nucleotide data set aligned using CLUSTAL, with gap penalty set to 25.

CLUSTAL alignment at gap penalty = 35

This gap penalty was used by Cummings et al. (1994) in their study using rpoC2 sequence data. The alignment yielded 290 informative sites; 84 more than obtained from the least stringent intuitive alignment.

The consensus tree of the equally parsimonious trees from this alignment is poorly resolved (Figure 2.28). The presence of the major subfamilial clades is similar to that shown in trees obtained using intuitive alignment methods, although there are some notable differences in their composition. The panicoid clade lacks Panicum, Pennisetum and Brachiaria, and gains the chloridoid genus Tragus as its basal-most taxon. As in the previous alignment, the danthonioid clade comprises only some of the taxa included in it in trees from the intuitive alignments. In contrast to the tree from the alignment using a gap penalty of 25, the pooid clade is now monophyletic and sister to a clade comprising some of the bambusoid taxa.

In the bootstrap topology from this alignment (Figure 2.29), only two of the major lineages are retained in their entirety; the danthonioid clade (with bootstrap support of 87.4%; Bremer support value = 11) and the pooid clade (99.4% bootstrap support; Bremer support value = 21).

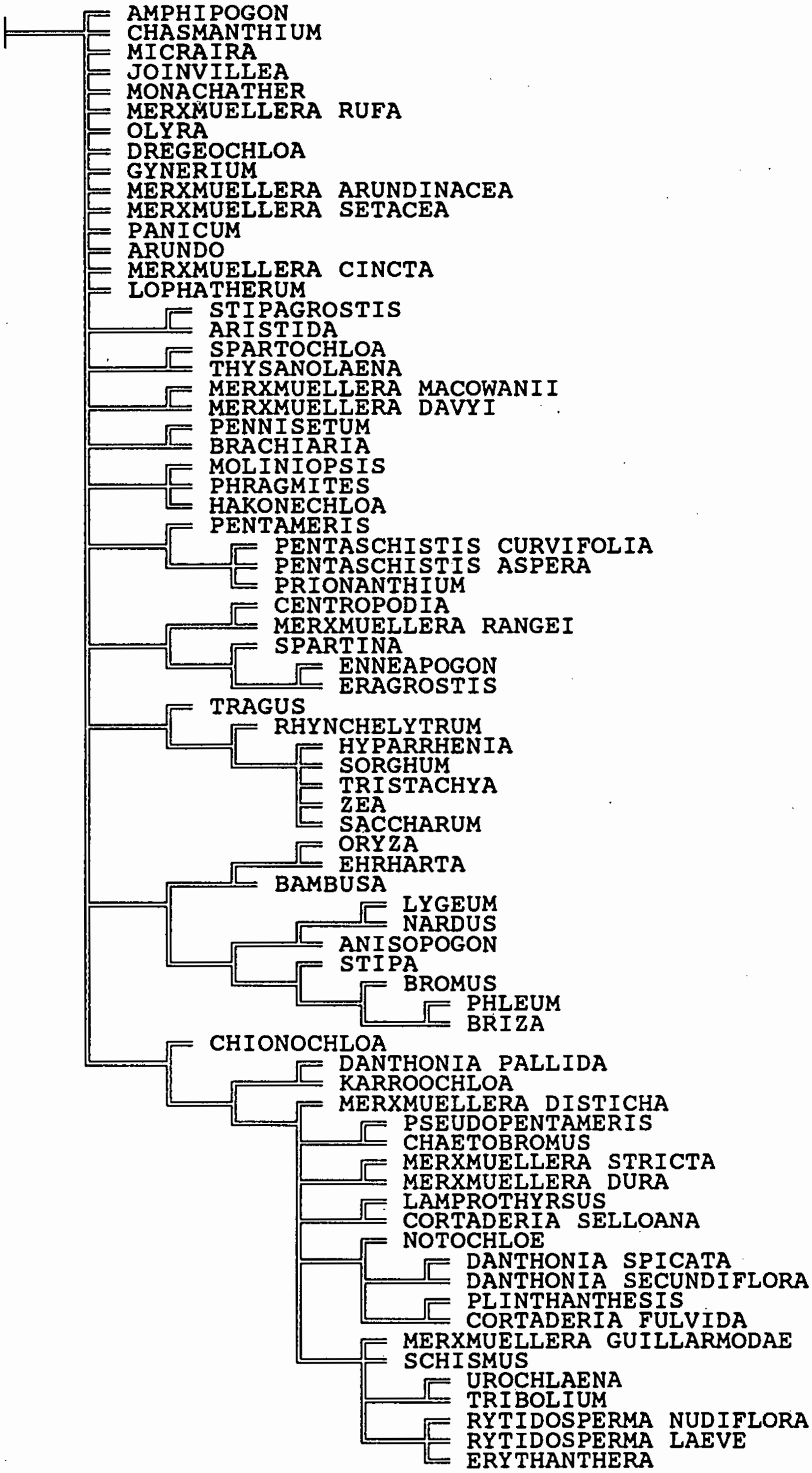


Figure 2.28. The consensus tree of the 1287 (memory overflow) equally parsimonious trees obtained from a m*bb* analysis of the nucleotide sequence data aligned using CLUSTAL, with the gap penalty set at 35. Length = 945, ci = 43, ri = 82.

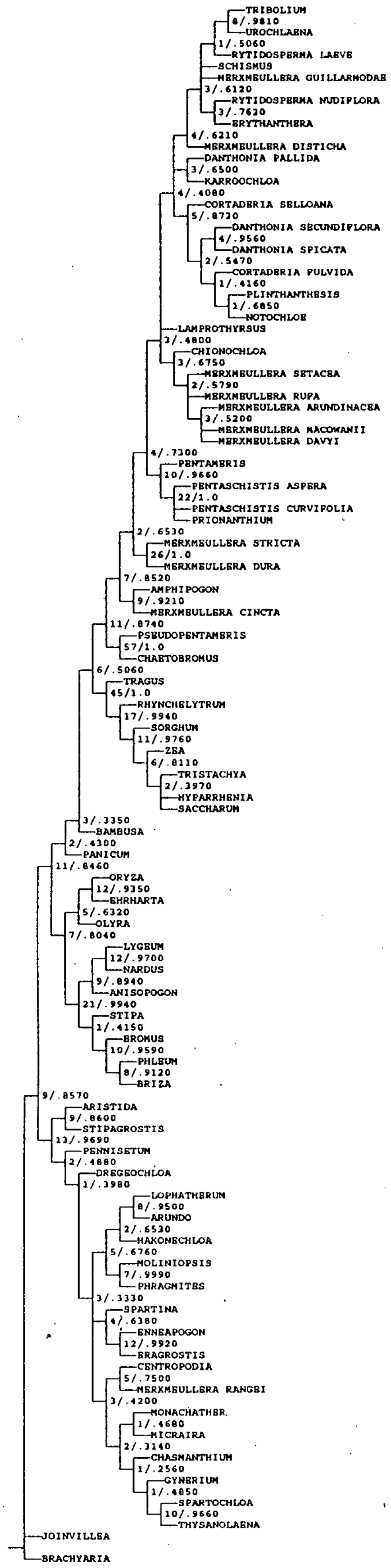


Figure 2.29. Topology showing Bremer support (left of "/") and bootstrap support values produced from 1000 replicates of the nucleotide data set aligned using CLUSTAL, with gap penalty set to 35.

CLUSTAL alignment at gap penalty = 45

Two hundred and ninety two informative sites were extracted from this alignment, two more than when a gap penalty of 35 was used. Parsimony analysis of these data resulted in a tree that was only slightly more resolved than those from the previous alignments. Improved resolution in this tree (Figure 2.30) is mainly found in the danthonioid lineage. The composition of this lineage is now identical to that of the danthonioid lineage as elucidated by the successive stringency alignments. Other anomalies involving the panicoid and chloridoid clades (described above) persist.

The bootstrap topology from this data set is similar to that obtained from the alignment using a gap penalty of 35 (Figure 2.31). The danthonioid and pooid major clades are once again retained intact with good bootstrap support.

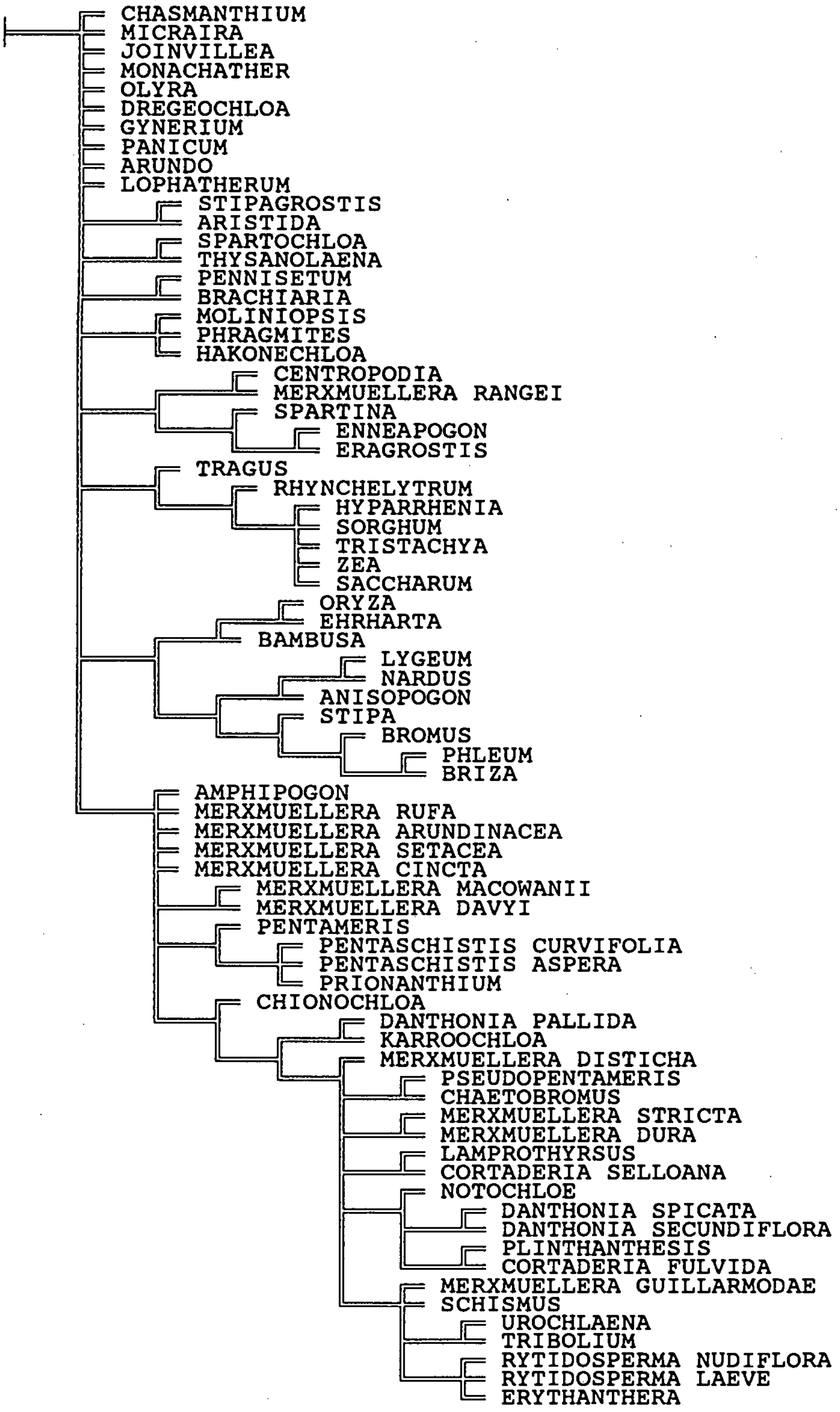


Figure 2.30. The consensus tree of the 1286 (memory overflow) equally parsimonious trees obtained from a m*bb* analysis of the nucleotide sequence data aligned using CLUSTAL, with the gap penalty set at 45. Length = 950, ci = 44, ri = 82.

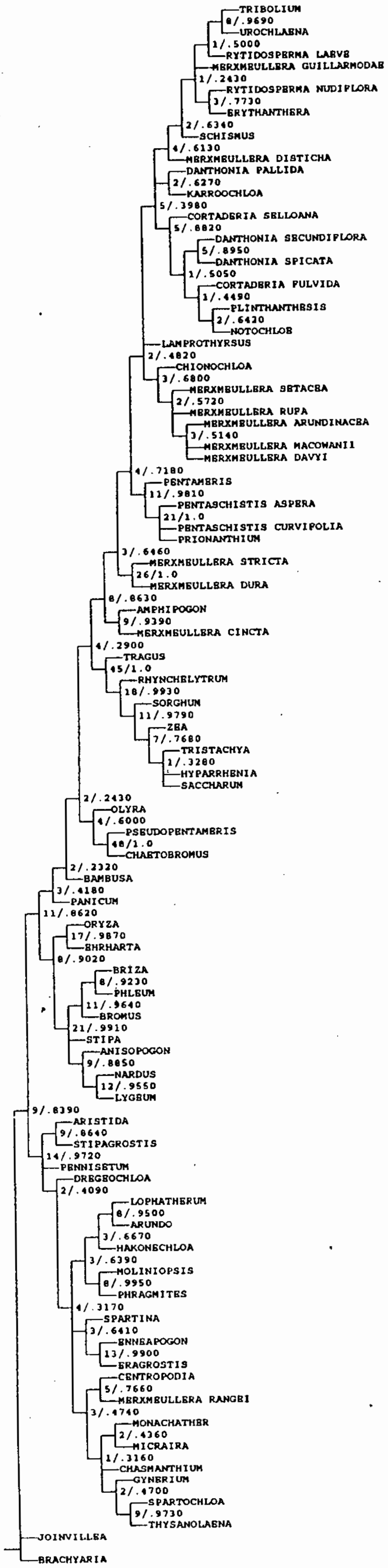


Figure 2.31. Topology showing Bremer support (left of "/") and bootstrap support values produced from 1000 replicates of the nucleotide data set aligned using CLUSTAL, with gap penalty set to 45.

Comparison of the phylogenies from different alignment methods

Comparison of the different trees derived from the different alignments and alignment methods can be carried out in a number of ways. These include comparing the relationships among the major clades, comparing the composition of each of the clades, and comparing the relationships of the taxa within these clades.

Comparison of the composition of the major clades

Inconsistent composition of clades between different phylogenies implies that either (one of) the cladistic methods used is inaccurate, or that these data (quantity or quality) are insufficient to adequately expose cladistic relationships.

When the major clades shown in the various trees are examined, their composition is, with the exception of the trees from the CLUSTAL aligned sequences, quite constant. Any variation consistently involves the same small set of taxa; Amphipogon, Dregeochloa, Arundo and Lophatherum in particular deserve mention in this regard. Problems in the placing of the latter genus are difficult to diagnose. As the sequence was not one obtained in this study, its accuracy cannot be checked. Even when gap characters are added, the resolution of this taxon is not improved.

The smaller clades (those comprising two taxa), are more difficult to examine for consistency, as only one taxon has to be misplaced for the clade to disappear. They are thus either present or absent.

The topologies from the different alignment stringencies are the most consistent in terms of clade composition. The Neighbour Joining topologies from the lower two stringencies are comparable to these, but clade composition undergoes changes with increasing stringency. The CLUSTAL alignments are very different when compared to all the other trees, mainly because of the poor resolution at the base of the trees, and the loss of taxa from the major clades resolved in trees derived from other alignments.

Comparison of the relationships of taxa within the major lineages

Where well resolved, both the nucleotide only and nucleotide and gap data sets are consistent in their depiction of relationships among the taxa within the major clades in all the alignment stringencies. Within the largest clade (the danthonioid lineage) minor differences in resolution are found, but no there are no contradictory topologies. The relationships from these analyses are also generally consistent with those shown in the Neighbour Joining topologies derived from these alignments. However, the relationships among representatives of the other subfamily lineages in the Neighbour Joining trees are not consistent. This may be caused to some extent by the small sample size in these clades.

Owing to the substantial inconsistencies in clade composition in the trees based on the CLUSTAL alignments, the comparison of relationships within the major clades is not possible. However, within the danthonioid lineage, the taxa that are constant to the clade show similar relationships to those shown in the trees from the various alignment stringencies.

Comparison of relationships of the major lineages

The comparison of the relationships between the major clades is limited to some extent by the lack of resolution and the inconsistent clade composition between some of the trees. Figures 2.32(a-u) summarise the relationships of the major clades from the different trees derived using the various alignment and analytical methods. The most resolved topologies are those derived from the least stringent alignments with added characters coding for gaps (Figure 2.32 g,h). The only trend that is detectable from the intuitive alignments is the elucidation of the PACC clade from the basal pooid and bambusoid clades. The relationships within this clade are, however, shown to be variable.

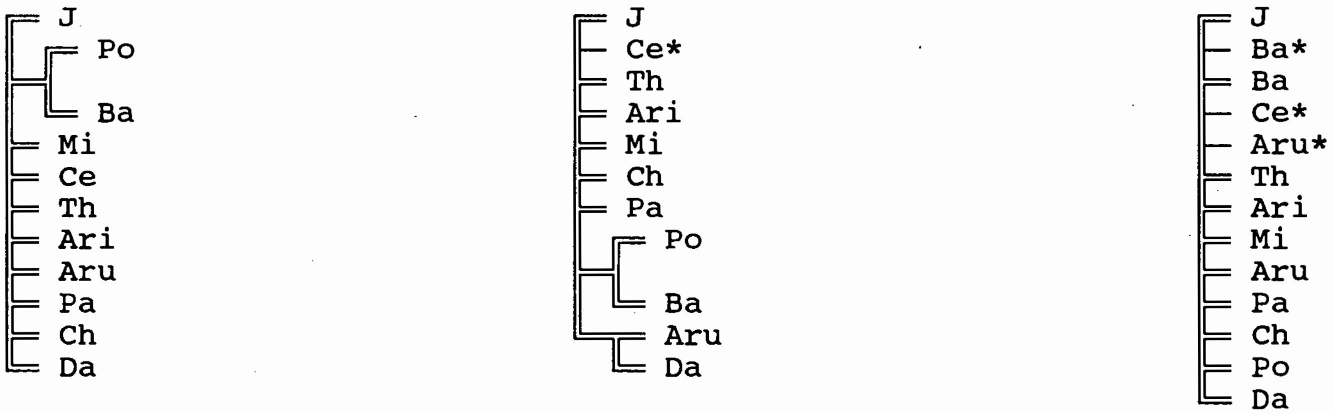
In general, bootstrap support for the major clades as elucidated from the successive stringency alignments is weak. Exceptions to this are the clearly supported pooid clade, and (where present) the PACC clade. Bootstrap support for the smaller lineages (such as the Micraireae and Thysanolaeneae) is generally strong, but the relationships of these clades are equivocal and not well supported. The bootstrap values of the major clades in the distance-based topologies tend to be higher than those from the parsimony-based topologies. The inconsistent and taxonomically unacceptable (because of subfamily fragmentation) topologies from the bootstrap analysis of the computerised alignments suggest that these results should be interpreted with caution, or even rejected.

Is the successive stringency alignment method successful?

Arguments against the use of exceedingly variable sequences for phylogenetic analyses are based on the uncertain positional homologies that result from difficulties in sequence alignment. Although the rpoC2 sequence used here is highly variable, it is felt that alignment problems are effectively addressed because the successive stringency method tests interpretations of positional homology of the repeat units and compares the results on a cladistic basis.

In terms of both resolution and consistency in clade composition, the successive stringency alignments outperform the computerised alignments, especially when maximum parsimony is used to produce the phylogenies. Unfortunately, no statistic or measure is available to support or show this

Figure 2.32 (a-u). The relationships between the major lineages from the various analyses carried out on the rpoC2 sequence data. Figures 32 (a-f) depict the consensus trees from the parsimony analyses of the nucleotide data sets at each alignment stringency; Figures 32 (g-l) depict those trees from the parsimony analysis of the nucleotide data sets at each alignment stringency, with characters added for gaps; Figures 32 (m-r) depict the topologies from the Neighbour Joining analyses at each stringency; Figures 32 (s-u) depict the topologies from the parsimony analysis of the CLUSTAL alignments using different gap penalties. Taxa marked with an asterisk (*) are polyphyletic, and (when split unequally) the smaller of the polyphyletic clades is distinguished with a thin line (-). Abbreviations of the clades: Ari = Aristideae, Aru = Arundineae (sensu Watson 1990), Ba = Bambusoideae, Ce = Centothecoideae, Ch = Chloridoideae, Da = Danthonieae (sensu Watson 1990), J = *Joinvillea* (outgroup), Mi = Micraireae, Pa = Panicoideae, Po = Pooideae, Th = Thysanolaeneae.

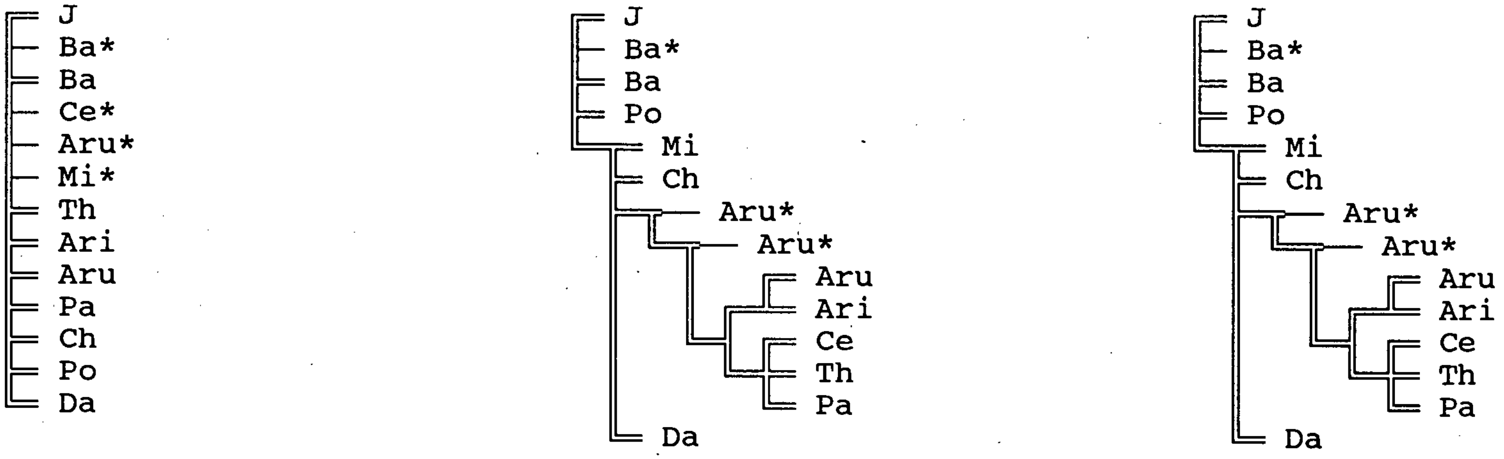


a) Stringency 1. Length = 613, ci = 45, ri = 75

b) Stringency 2. Length = 606, ci = 45, ri = 75

c) Stringency 3. Length = 578, ci = 45, ri=75

Figure 32 cont.

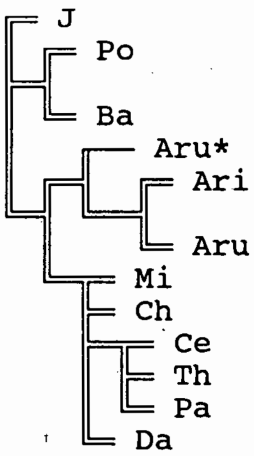


d) Stringency 4. Length = 551, ci = 45, ri = 76

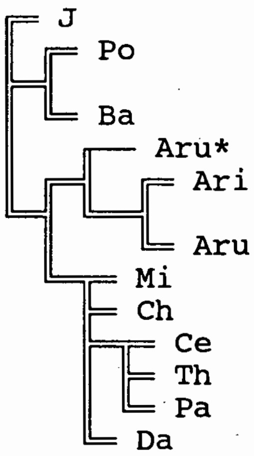
e) Stringency 5. Length = 525, ci = 47, ri = 77

f) Stringency 6. Length = 504, ci = 47, ri = 74

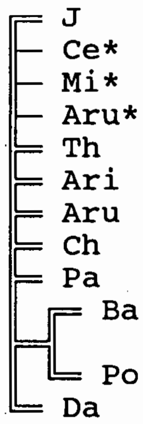
Figure 2.32 cont.



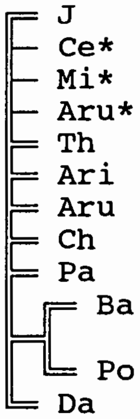
g) Stringency 1 Length = 689, ci = 44, ri = 75



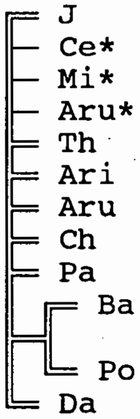
h) Stringency 2. Length = 688, ci = 44, ri = 75



i) Stringency 3. Length = 670, ci = 44, ri = 75



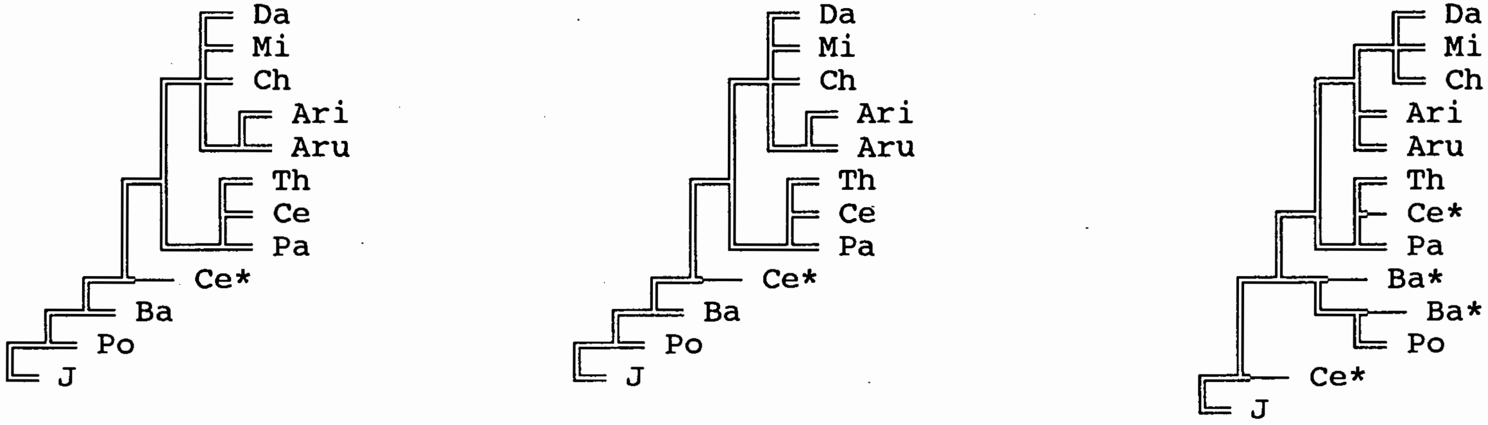
j) Stringency 4. Length = 652, ci = 44, ri = 76



k) Stringency 5. Length = 633, ci = 45, ri = 77

l) Stringency 6. Length = 619, ci = 45, ri = 76

Figure 2.32 cont.

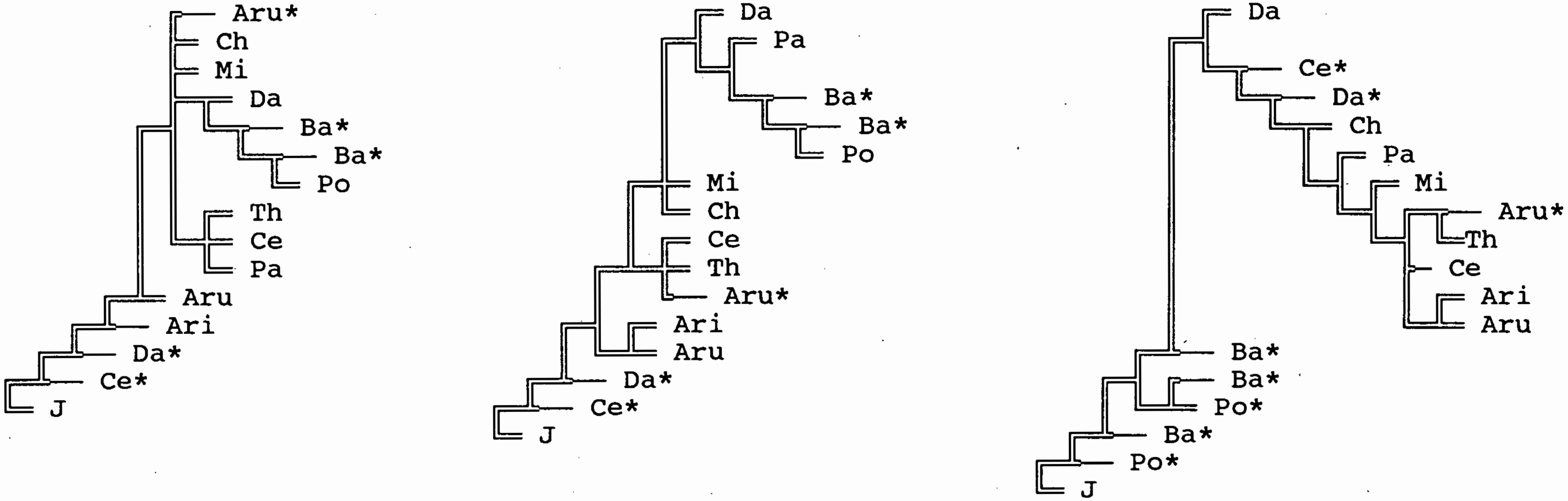


m) Stringency 1 (Neighbour Joining)

n) Stringency 2 (Neighbour Joining).

o) Stringency 3 (Neighbour Joining).

Figure 2.32 cont.

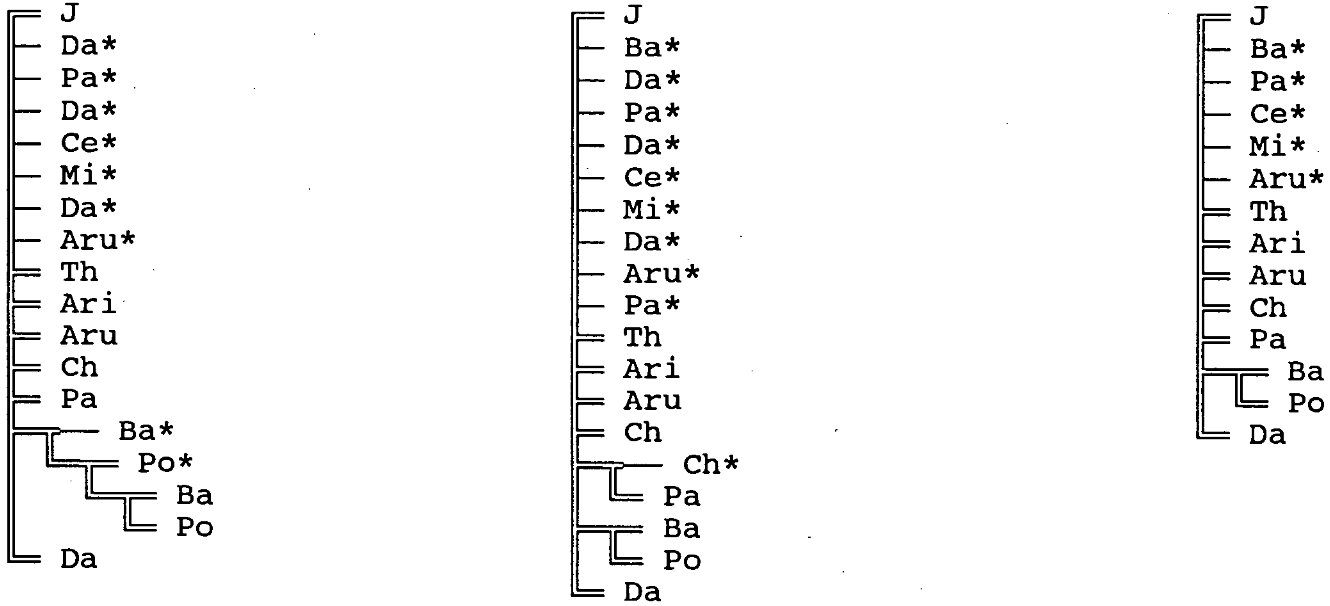


p) Stringency 4 (Neighbour Joining).

q) Stringency 5 (Neighbour Joining).

r) Stringency 6 (Neighbour Joining).

Figure 2.32 cont.



s) CLUSTAL - 25. Length = 865, ci = 43, ri = 81

t) CLUSTAL - 35. Length = 945, ci = 43, ri = 82

u) CLUSTAL - 45. Length = 950, ci = 44, ri = 82 "performance". The method thus has some merit, but it must be remembered that it is based on an initial intuitive alignment deduced using a set of rules constructed in a subjective manner.

The effectiveness of the intuitive alignment method is linked to sample size. Only once a substantial number of taxa had been sequenced could any certainty be placed on the similarity (and thus assumed homology) of the heptameric repeats. This certainty is based on the levels of variation found in sequences from groups of taxa known to be related, such as the chloridoids or panicoids. Thus, as sample size is increased, clade-specific (such as subfamilial) patterns begin to appear, and insights into the extent of variation among and within the repeats in these groups were obtained. However, in poorly represented groups or groups where taxic density is naturally limited, it is difficult to gain any impression of clade-specific patterns which could affect interpretations of positional homology. Thus groups, such as the Bambusoideae (undersampled) and Aristideae (low taxic diversity), possess some unusual repeat motifs. These taxa may be placed in unusual or unexpected positions in the phylogenies because of sub-optimal alignments of these motifs. Where possible, a larger sample across the diversity of these groups would assist in producing a better alignment and thus better phylogenies.

In the successive stringency alignment method, the distinction between alignment modifications made to support biological reality on the one hand (positional homology), and those made to support phenetic reality (similarity) on the other, needs to be made. The majority of the changes made in the alignments at higher stringencies may be viewed as "luxuries" - changes made to satisfy the aesthetic criterion of similarity or to satisfy numerical requirements, rather than the historical criterion of homology. In this regard, the level of coarseness referred to in the section on alignment methods needs to be re-considered. The application of the method at the coarsest level, that of recognising and treating whole repeat units as characters, appears to correspond with biological reality. Changes made on the basis of the finest level (nucleotides), and to a lesser extent the middle level (amino acids) correspond to "luxury" alignment changes.

The addition of binary codes for the gaps in the alignment results in topologies with improved resolution at the lower stringencies. This implies that the gap characters are sufficient to override the noise (homoplasies) in the data, and provide a more resolved topology. At the higher stringencies, these characters did not assist in resolving the tree further unless they were given moderate additional weightings (two or three times the weight of the informative sites; trees not shown).

The limits of rpoC2 data

It is clear from the variety of analyses carried out that sequence data from the insert in the rpoC2 gene is limited in its systematic applicability. It is capable of showing relationships below the level of the subfamily, but cannot resolve relationships among the subfamilies, despite the substantial sample size and variety of alignment and analytical methods employed. This implies that the gene is too variable to be useful at this level, and if subfamilial relationships are to be resolved a more conserved gene sequence is required.

In the next chapter, the conserved plastid rbcL gene is sequenced from taxa selected from each of the various lineages elucidated by the rpoC2 data. Sequence data from this gene are then used in an effort to ascertain the relationships between the subfamilies and other lineages that rpoC2 sequence data could not resolve.

Examination of the variability of the rpoC2 insert

Examination of the five rpoC2 sequences from the two Phragmites species indicates that, despite samples being taken from different continents, the rpoC2 sequences are invariant in terms of insertion and deletion events, and the sequence itself is conserved among the specimens examined. Thus, in comparing these sequences, no alignment problems were encountered.

When compared in a pairwise manner, the sequence divergence among the pairs of taxa is minimal, with the greatest differences being between Moliniopsis and the different samples of the two species of Phragmites (Table 2.4).

Only two phylogenetically informative sites were obtained from the five Phragmites sequences. Phylogenetic analysis of these data (using Moliniopsis as an outgroup) shows the Australian and Japanese samples of P. australis as sister taxa (supported by one of the informative sites) while the Canadian and southern African samples form a second clade, united by the second informative character (Figure 2.33).

A Neighbour Joining analysis using the Jukes and Cantor correction shows that the pandemic P. australis is paraphyletic, as P. karka diverges from within the pandemic species (Figure 2.34). While P. karka may well have evolved from an oriental population of P. australis, only further sampling of both species will show this with any certainty. This would, however, be an interesting study, and would be the first study to use sequence data to assess species concepts in pandemic and endemic taxa.

These observations suggest that the insert is sufficiently stable and conserved for use in molecular systematic studies that use single species as generic representatives. However, sequence data from other larger genera from which more than one species have been sampled indicates that there may be a greater degree of variation between some congeneric species in terms of both indels and point mutations. These genera include Danthonia, Cortaderia, Pentaschistis and Merxmuellera. However, all these genera have been noted as being morphologically diverse, and so the infrageneric sequence variation may reflect poor taxonomic concepts rather than the ineligibility of the sequence for systematic purposes. In these instances, only an expanded sequencing study (such as carried out on Merxmuellera) will indicate taxonomic problems.

Table 2.4. Number of point mutations between the specimens of Phragmites sampled to assess rpoC2 sequence variation in a pandemic species. Taxon abbreviations: Mol = Moliniopsis, P.k = P. karka, P.a = P. australis. Country abbreviations: Ja = Japan, SA = South Africa, Au = Australia, Ca = Canada.

	P.k (Ja)	P.a (Ja)	P.a (Au)	P.a (Ca)	P.a (SA)
Mol	5	6	6	6	6
P.k (Ja)		1	1	1	1
P.a (Ja)			0	2	2
P.a (Au)				2	2
P.a (Ca)					0

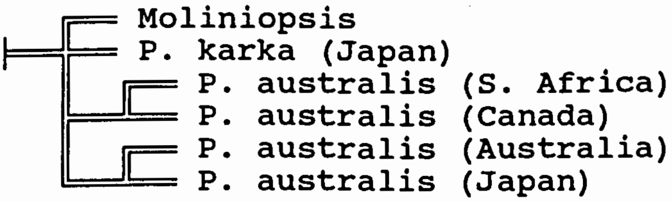


Figure 2.33. The single most parsimonious tree obtained from the analysis of the two phylogenetically informative sites obtained from the set of Phragmites sequences. Length = 2, ci = 100, ri = 100.

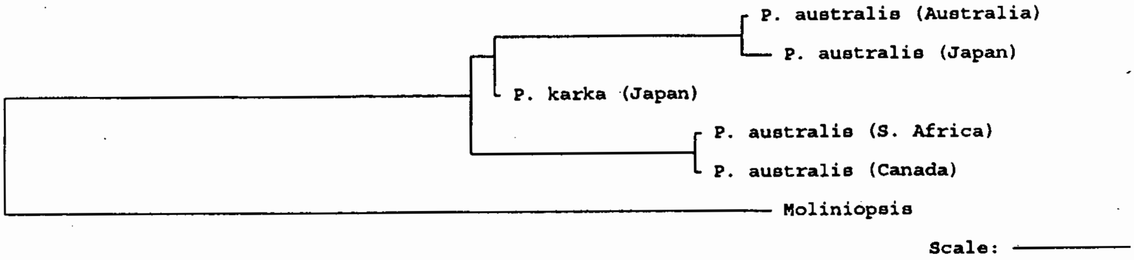


Figure 2.34. The Neighbour Joining analysis using the Jukes and Cantor correction of the samples of two species of Phragmites. Note that in this topology, the oriental P. karka is shown to be derived from within the pandemic P. australis. Scale bar is approximately equal to a distance of 0.00104.

Can the slipped strand mispairing events be tracked along a lineage?

Gaps to be plotted on this cladogram are taken from the alignment at stringency 2. These gaps, corresponding to deletions and insertions, are plotted on the lineage and shown in Figure 2.35. Additional adjustments to the alignment affected the danthonioid lineage only in stringencies 4 and 6, and these changes reflect what was earlier termed "luxuries" in homology assessment, rather than realities.

Where gaps are of consecutive repeats, the individual repeats are coded, a somewhat unsatisfactory method as it is possible that they are the result of a single slipped strand mispairing event. However, this system is retained because not all taxa may share the complete set of consecutive gaps; an event that lost two repeats may not have been the same event that lost only one of the two repeats in another lineage.

In examining the distribution of these, several observations can be made.

1) In the initial alignment there is no danthonioid-specific insert, as there are for the pooid and panicoid lineages, although the presence of such synapomorphies are alignment dependant. However, in the alignment at stringency 2 a danthonioid-specific repeat is recognised by default, as a bambusoid-specific repeat unit is recognised. This repeat unit (numbered "1" in Figure 2.35) has, however, been lost on four separate occasions during the evolution of the lineage.

2) There are a total of 25 repeat-sized or smaller gaps in the lineage. Of these, only one is a reversal within the lineage (the loss of a repeat in _Merxmuellera stricta_ following its gain in the "_Rytidosperma_" clade). Other reversals (a total of 17) are of repeats shared with other lineages. Of these, 14 are restricted to single terminals.

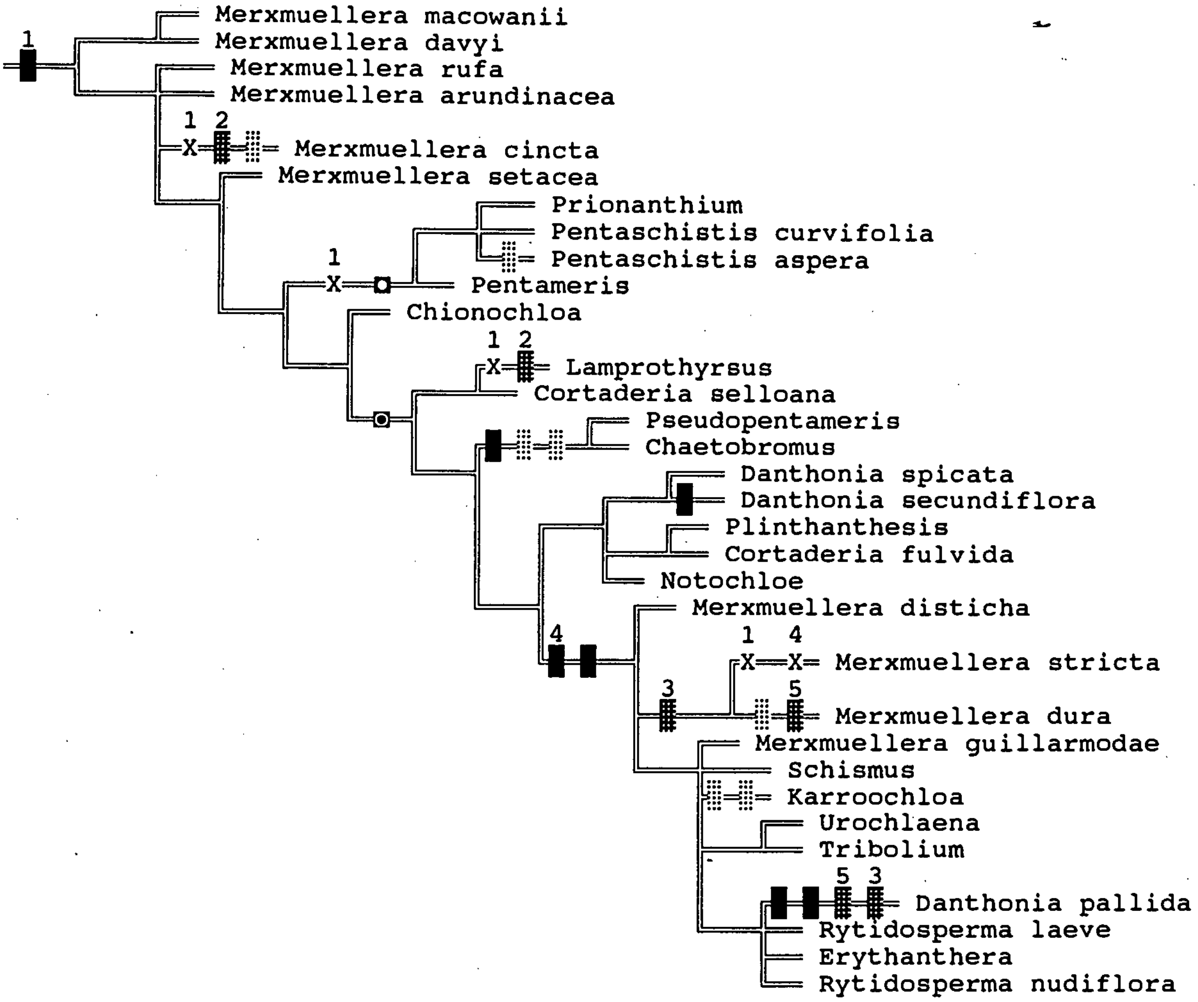


Figure 2.35. The distribution of characters corresponding to gaps in the alignment at stringency 2. Numbers are provided to allow tracking of the homoplasic characters. █ = gains, ░ = single loss event, ▦ = multiple (homoplasic) loss event, X = loss event following gain, □ = clade supported by gaps added at stringency 4 (reversals not shown), ▣ = clade supported by gaps added at stringency 6 (reversals not shown).

3) Other than the danthonioid-specific repeat, there are six novel repeats gained during the evolution of the clade. Three of these are restricted to single terminals and are thus cladistically uninformative.

4) The majority of gaps are for entire heptameric repeats. The only exceptions are a two-codon gain in Danthonia secundiflora, a three-codon loss in the (Chaetobromus Pseudopentameris) clade and a two-codon loss in Merxmuellera dura.

5) In the more stringent alignments where changes were made that affect the danthonioid lineage (stringencies 4 and 6), the added gaps affect numerous taxa and may be viewed as apomorphies for substantial parts of the lineage. These gaps (shown in Figure 2.35) are not included in the count of gains or losses described above.

Implications for phylogenies based on the rpoC2 insert

The repetitive loss of repeat units results in substantial homoplasy in both the nucleotide and gap characters, while the more conserved, informative data appears to be derived from gains of repeat units. This implies that the presence of clade specific (synapomorphic) units is important, and that these repeats should play a major role in the alignment of the sequence data. This emphasises the importance of examining homology at the level of the repeat units. As noted above, this is the coarsest of the three levels at which homology may be tested, and reflects biological reality, rather than phenetic or aesthetic luxury. It is unfortunate that the only computerised alignment procedure that is based on cladistic methodology, and which accounts for gaps in a cladistic manner (MALIGN), is severely limited by sample size.

The rarity of parallel repeat gains (homoplasies) in comparison to parallel losses suggests that the analysis of the data set of gap characters (as was added to the various alignment stringencies) might best be carried out using Dollo parsimony or relaxed Dollo parsimony methods.

How did the insert evolve?

Igloi et al. (1990) considered that the repeats of the insert evolved by means of slipped strand mispairing, the original template of the repeats being from a region 3′ to the insert. This template region is recognisable in all the grasses sequenced here, as well as the poalean relative Joinvillea.

While the ancestors of the grasses are extinct, certain bambusoid taxa have retained some pleseiomorphic morphological characteristics that the putative grass ancestors possessed. Two such bamboos are Anomochloa and Streptochaeta, each belonging to its own tribe, and both only known from shaded forest areas in South America. DNA of these two genera was obtained from Dr Lynn Clarke (Iowa State University, Ames) and the rpoC2 insert and flanking regions were amplified and sequenced as described in the methods section of this chapter.

The length and composition of the insert of Anomochloa was found to be quite different from that of Streptocheata and the other bambusoid taxa. The insert in the latter genus is substantially longer than in the former (459 versus 210 base pairs). Furthermore, the insert in Anomochloa contains no heptameric repeats. Instead, it has much longer repeats of 11 or 13 amino acids, or even longer. Streptochaeta also possesses 11-amino acid repeat units, but these are dispersed among some 7-amino acid repeats (Figure 2.35). The repeat structure of both these taxa is thus quite different from other bambusoid taxa, in that they both contain numerous longer repeat units. This unusual sequence structure, and the undersampling of the bambusoid lineage, meant that it was not possible to align these sequences with those of the other grasses. For this reason they were excluded from the phylogenetic analysis performed above.

Assuming that Anomochloa and Streptochaeta have retained a primitive form of the insert, it appears that the insert initially comprised repeats of 11 or 13 amino acids, rather than the heptameric construction prevalent in derived taxa. The subsequent accumulation and accretion of 7-amino acid repeats at the expense of 11 and 13-amino acid repeats thus occurred during bambusoid diversification. The accumulation of the 7-amino acid repeat units over the 11 and 13-amino acid repeats may be driven by selection favouring the stability or improved functioning of the protein's alpha helix based on the shorter repeats. Additional sequence data from other putatively primitive bambusoid taxa may help in understanding the evolution of this interesting sequence in the grasses.

Figure 2.35. The amino acid sequences of the insert in the plastid *rpo*C2 gene from *Anomochloa* and *Streptochaeta*, as well as other genera from a range of bambusoid tribes.

Anomochloa

FRYRALNENSE
YE
YRTLEEDSEDE
YEDGTLEDEDGIPEEDSEDE
YEDGIPEEDSEDE

Streptochaeta

FRYRTLEEDSEDE
YETLEDE
YGIPENE
YRTLEDE
YGILEEDSEEE
YGILENE
YRDE
YGILERLEDE
YGILEETRGE
YGILEEDSEEE
YGNPEEDSEDE
YGSPEEDSEDE
YGSPEEDAEDE
YRSPEEDSEDE
YGNPEEDSEEE
YGILEEDSEEE

Bambusa

FRYRTLEDE
YRTLEDE
YRTREEDSEDE
YGSPENK
YRTREDE
YETLEDE
YRTLEDE
YETLEDE
YGILEDE
YETLEDE
YRTREEDSEDE
YGSPENK
YRTREDE
YGTLDEDSEDE
YGSPEEGSEDE
YGTLEEDSEEDSEDE

Figure 2.35 cont.

Oryza	Ehrharta	Olyra
FRYRTLEDE	FRYRTLEDE	FRYRTLEDE
YRTREKDSENE	YRTREDE	YRTREEDSEDE
YGSPENE	YRTREEDSEDE	YGSPENG
YRTREEE	YGSPENE	YRTREKE
CKTLEDE	YRTREEE	YYETLEDE
YRTREEE	YETLEDK	YGVLEDE
YETLEDE	YRTLEDE	YETLEDE
YGIPENE	YETLEDE	YGILEGE
YETLEDE	YGILEDE	YETLEDE
YGILEDE	YETLEDE	YGILEDE
YRTREEESEDE	YGIPENE	YRTREEDSEDE
YGSPENK	YRTREEDSEAE	YRSPENE
YRPREDK	YGSPENE	YRTREDE
YGTLEEDSEDE	YRTREDE	YGTLEEDSEEDSEDE
HGTLEEDSEEDSEEDSEDE	YGTLEEDSEEESEDE	
	YGTLEEDSEEDSEDE	

CONCLUSIONS

The _rpo_C2 sequencing studies, divided into four smaller facies, has shown the following:

1. As found by Cummings _et al_. (1994), the grass specific insert in the _rpo_C2 gene is too variable to resolve the relationships between the grass subfamilies.

Irrespective of alignment and cladistic methodology, the relationships of the subfamilies were consistently unresolved. At best, resolution of the samples into ((Bambusoideae, Pooideae) PACC clade) could be attained.

2. The variability in the sequence data presented difficulties in obtaining satisfactory alignments.

An alignment procedure was developed to test the positional homology of the heptameric repeat units by means of congruence. This procedure produced better alignments than computerised alignment methods (as measured by the resolution of cladograms derived from them). While developed specifically for sequence data comprising repeats originating from slipped strand mispairing, the principles of the method may be applied to other variable sequence data.

3. Phylogenetic analysis of the _rpo_C2 sequence data shows that the subfamily Arundinoideae is polyphyletic.

Lineages corresponding the tribes Arundineae and Danthonieae (both _sensu_ Watson 1990) and the Aristideae, Thysanolaeneae and Micraireae _sensu_ Clayton and Renvoize (1986) are supported. However, with the exception of the Aristideae, the generic composition of these tribes does not conform to any previously described classification. The Micraireae and Thysanolaeneae (previously monotypic tribes) are shown to include other genera (_Monachather_ and _Spartochloa_ respectively), while the Arundineae (the reeds) are shown to lack the genera *Arundo* and *Gynerium*, but include the unusual *Dregeochloa*, previously placed in the Danthonieae *sensu* Watson (1990).

4. The lineage corresponding approximately to the Danthonieae comprises the majority of the genera in the Arundinoideae, but there are some notable absences.

Other than genera shown to be placed in the reedy or arundinoid clade, *Centropodia* and *Merxmuellera rangei* are placed in the Chloridoideae, while *Anisopogon* is shown to be related to *Stipa* (Stipeae) in the Pooideae. There is no suggestion that the Stipeae are a tribe of the Arundinoideae, as proposed by Watson (1990). The relationships of *Amphipogon* are shown to be equivocal, and it appears in the phylogenies obtained here either as the basal-most taxon in the danthonioid lineage, or basal to the Micraireae.

5. The grass specific insert in the *rpo*C2 gene is relatively conserved within a species.

This was clearly shown in the study using a pandemic species to test for variation in samples from diverse geographic localities. Variation that is found at this level takes the form of point mutations rather than insertion or deletion events.

6. The insert is suitable for elucidating intergeneric relationships below the level of the subfamily.

Because of problems with the alignment of the sequences, relationships between the major lineages are not clearly resolved or well supported.

7. Within the well sampled danthonioid lineage, slipped strand mispairing mutational events are phylogenetically informative.

Within the danthonioid lineage it appears that losses of repeats are more frequent and homoplasic than gains. When these events are coded and included in phylogenetic analyses with the nucleotide data, cladogram resolution can be considerably enhanced.

8. The repeat structure of the insert in primitive bambusoid taxa suggests that the heptameric repeat evolved from a larger 11- or 13- amino acid repeat unit.

This result is based on the assumption that morphologically primitive taxa will retain similarly primitive sequence structure. This is not testable using the rpoC2 insert sequence data as alignment of the sequences was found to be impossible.

CHAPTER 3

PHYLOGENETIC ANALYSIS OF

THE rbcL SEQUENCE DATA

INTRODUCTION

The rbcL gene is highly conserved, occasionally varying in length by one or two codons. These mutations tend to occur at the more variable 3′ end of the gene (Clegg 1993). Apart from a few exceptions, the gene is present as a single copy in the plastid genome, and is found in the LSC (Large Single Copy) region.

It has been used extensively in molecular systematic studies, generally at the superfamilial level. Chase et al. (1993) demonstrated the utility of this gene in examining relationships among the higher land plants, and Wolfe et al. (1989) have attempted to date the monocot - dicot divergence by means of a phylogeny based on rbcL sequence data.

As mentioned in the introductory chapter, rbcL was used in one of the first molecular systematic studies, where it was concluded that it was not sufficiently variable to solve relationships between the subfamilies of the grasses (Doebley et al. 1990). These findings may have been influenced by a number of factors:

1) Owing to a possible emphasis on the economically important grasses, only those subfamilies with cereal or crop species were sampled. Three other subfamilies were missing from the analysis.

2) The use of an outgroup (spinach) that is too distant from the study group. This problem was recognised by Doebley et al. (1990) who were unfortunately limited by a choice of dicot outgroup taxa. Subsequent studies have resolved the sister group relationships of the grasses, and the monotypic family Joinvilleaceae is now known to be the closest extant relative of the Poaceae (Campbell and Kellogg 1987; Doyle et al. 1992; Linder and Rudall 1993; Kellogg and Linder, in press). The rbcL sequence for Joinvillea is now available, and is used here as the outgroup.

3) The sequence for Zea used in the initial study was later found to be incorrect. Although the inaccuracies in the sequence were small, they may have affected the results of the relative rate tests that suggested that rbcL evolved at an accelerated rate in the lineage leading to Zea (Doebley et al. 1990).

The latter two points in particular may have resulted in the very large number of characters (296) used in the parsimony analysis conducted by Doebley et al. (1990). However, these authors do not state if this number includes autapomorphies or not. Nonetheless, this number of characters is still substantially higher than the 160 phylogenetically informative characters obtained from the sequences used in the analyses presented here, which is based on more than three times the number of taxa.

These additional taxa were selected from all the lineages elucidated by the phylogenetic analyses of the rpoC2 data. Thus all six subfamilies sensu Clayton and Renvoize (1986) are included. Furthermore, emphasis has been placed on ensuring that sequence data from at least one taxon from each of the possibly polyphyletic clades of the Arundinoideae is obtained. Unfortunately, a lack of sufficient DNA of Micraira precluded its inclusion in this study. However, as Monachather is consistently shown to be its sister taxon in many of the rpoC2 analyses, this genus is conditionally considered to represent this tribe. Details of all the additional taxa sequenced in this study are presented in Table 3.1. Using this enlarged rbcL sequence data set, this study:

1) Re-assesses the phylogeny of the grasses using rbcL sequence data.

2) Re-examines the proposal that rbcL is uninformative about relationships in the grasses.

MATERIALS AND METHODS

The rbcL gene was amplified using the two flanking primers Z-1 and Z-1375R. The rbcL sequencing primers and their sequences were provided by G. Zurawski (DNAX Research Institute). Complete one-directional sequences were obtained for all samples using the primers and strategy outlined in Figure 3.1. However, difficulties in sequencing some taxa necessitated the additional use of selected reverse-strand sequencing primers to read across problematic regions. These primers were complementary to those shown in Figure 3.1. Direct PCR amplification and sequencing was carried out as described in the methods section of the previous chapter.

Sequences obtained here, and existing sequences obtained from GENBANK, were entered into, and manipulated by, DAPSA, which was then used to extract phylogenetically informative sites for phylogenetic analyses.

The methods of phylogenetic analysis used in this study are identical to those described in the previous chapter. The heuristic m* and bb* options of HENNIG86 (Farris 1988) were used to obtain a set of shortest trees. In order to search for the existence of possible islands of most-parsimonious trees, PAUP version 3.1.1 (Phylogenetic Analysis Using Parsimony;

Swofford 1993) was used to conduct a random addition entry search using 500 replicates. Further details of this method are provided in the previous chapter.

Successive weighting (Carpenter 1988) as implemented in HENNIG86 was used to select a single tree from the set of minimal length trees. RNA version 1.1 (Farris 1994) was used to calculate bootstrap values (Felsenstein 1985) and Bremer support indices (Bremer 1988) from 1000 sampling replicates.

MEGA version 1.0 (Kumar _et al_. 1993) was used to produce trees using the Neighbour Joining algorithm (Saitou and Nei 1987) using the Jukes and Cantor correction (Jukes and Cantor 1969). MEGA was also used to obtain bootstrap figures from 1000 replicates for the neighbour joining trees using the Jukes and Cantor correction.

Figure 3.1. The primers and sequencing strategy to obtain a complete one directional sequence. Primers Z-1 and Z-1375R were used to obtain the initial _rbc_L amplification product. Primers Z-1, Z-234, Z-427, Z-674, Z-895, Z-1020, Z-1204 and Z-153R were used routinely, while primers complementary to those shown above were used to sequences problematic templates. Arrows indicate direction of sequencing. Primer sequences are as follows (5'-3'):

Z-1: ATGTCACCACAAACAGAAACTAAAGCAAGT
Z-234: CGTTACAAAGGACGATGCTACCACATCGA
Z-427: GCTTATTCAAAAACTTTCCAAGGCCCGCC
Z-674: TTTATAAATCACAAGCCGAAACTGGTGAAATC
Z-895: GCAGTTATTGATAGACAGAAAAATCATGGT
Z-1020: ACTTTAGGTTTTGTTGATTTATTGCGCGATGATT
Z-1204: TTTGGTGGAGGAACTTTAGGACACCCTTGGGG
Z-153R: AGAAGATTCCGCAGCCACTGCAGCCCCTGCTTC
Z-1375R: AATTTGATCTCCTTCCATATTTCGCA

Table 3.1. Species sampled for *rbc*L sequence variation, arranged according to subfamily and tribe *sensu* Clayton and Renvoize 1986).

Subfamily Tribe Species	Source of material and accession number
Pooideae	
Stipeae	
Stipa dregeana Steud. var. dregeana	Constantia, Cape Town, South Africa McDowell s.n.
Centothecoideae	
Centotheceae	
Chasmanthium latifolium (Michx.) Yates	Cultivated; Missouri Botanical Gardens, St. Louis, USA (Snow, 5944)
Arundinoideae	
Aristideae	
Aristida congesta Roem. and Schult. subsp. barbicollis (Trin. and Rupr.) De Winter	Pretoria Botanic Gardens, South Africa (Barker 1130)
Stipagrostis zeyheri (Nees) De Winter subsp. zeyheri	Cape Point Nature Reserve, South Africa (Barker 1133)

Arundinoideae	
Arundineae	
Amphipogon strictus R. Br.	Kings Tableland, New South Wales, Australia (Linder 5634)
Arundo donax L.	University of Cape Town campus, South Africa (Barker 1131)
Centropodia glauca (Nees) T.A. Cope	Alexander Bay, Northern Cape, South Africa (Linder 5410)
Danthonia spicata Roem. and Schult.	Mt. Desert Island, Maine, USA. (Kellogg, s.n.)
Dregeochloa pumila (Nees) Conert	Alexander Bay, Northern Cape, South Africa (Linder 5408)
Gynerium sagittatum (Aubl.) P. Beauv.	Kew Gardens (originally from Peru; Ref. No. 1991-1276 Kall)
Karroochloa purpurea (L.f.) Conert and Tuerpe	Botterkloof, Cape Province, South Africa (Linder 5360)
Merxmuellera macowanii (Stapf) Conert	Drakensberg mountains, South Africa (Barker 1009)
Moliniopsis japonica (Hack.) Hayata	Utsunomiya University campus, Japan (Kobayashi 1253)
Monachather paradoxus Steud.	Eulo, Queensland, Australia (Moll 1)
Phragmites australis (Cav.) Steud.	Black River, Cape Town, South Africa (Barker 1132)
Plinthanthesis paradoxa (R. Br.) S.T. Blake	Wollongong, NSW, Australia (Linder 5638)
Rytidosperma nudiflorum (Morris) Connor & Edgar	Cradle Mt., Tasmania (Linder 5693)
Arundinoideae	
Thysanolaeneae	
Thysanolaena maxima Kuntze	Royal Botanic Gardens, Kew (originally from Barbados; Ref. No. 1979-3225 Warr)

Chloridoideae Eragrostideae Eragrostis capensis (Thunb.) Trin.	Kenilworth Race Course, Cape Town, South Africa (Barker 1135)
Chloridoideae Pappophoreae Enneapogon scaber Lehm. var. scaber	Leeuw Gamka, Cape Province, South Africa (Barker 1023)
Panicoideae Andropogoneae Hyparrhenia hirta (L.) Stapf	Kirstenbosch Botanic Gardens, Cape Town, South Africa (Barker 1134)
Panicoideae Arundinelleae Tristachya biseriata Stapf	Pretoria Botanic Gardens, National Botanical Institute, South Africa (Barker 1126)

RESULTS

A total of 22 complete rbcL sequences was obtained, and added to 14 sequences extracted from GENBANK. No alignment problems were encountered during the manipulation of these 36 sequences. A total of 1344 base pairs was included in the analysis (regions corresponding to the flanking primers were excluded).

A total of 160 phylogenetically informative sites was extracted from these sequences. Of these, 111 represented the changes at the third codon position. Parsimony analysis using just these sites yielded a substantially resolved tree, whereas analysis using either the first base positions (31 informative sites) or second base positions (18 informative sites) of the codon resulted in numerous highly unresolved trees (trees not shown).

Using HENNIG86 (Farris 1988), an m*bb* analysis found 26 equally parsimonious trees with a length of 489 steps (ci = 42, ri = 67). Successive weighting produced three trees, one of which is shown in Figure 3.2. The search on 500 random input datasets failed to find any trees not found by the HENNIG86 m*bb* analysis, suggesting that the all islands of most-parsimonious trees had been found.

The bootstrap topology, obtained using RNA (Farris 1994) is shown in Figure 3.3.

The tree obtained using Neighbour Joining with the Jukes and Cantor correction differs slightly from the tree based on maximum parsimony, and is shown in Figure 3.4.

DISCUSSION

The 26 equally parsimonious trees differ in the topology within the subfamily Bambusoideae, and the relationships between the chloridoid, panicoid and arundinoid clades. Consequently, the consensus tree retains the Bambusoideae as a basal unresolved group, while the major lineages within the PACC clade form a polytomy. These nodes are marked on the topology shown in Figure 3.2, selected from one of the three trees obtained by successive weighting. The Neighbour Joining tree differs from the parsimony-based phylogeny in the placing of Gynerium and the relationships of the Chloridoideae - Danthonieae - Aristideae clades. These differences are discussed where relevant. A brief discussion of the relationships among and within the subfamilies and other lineages appears below. As in the discussion in the previous chapter, biological and taxonomic implications of the phylogenies are not discussed in detail. These issues are addressed in the final chapter (Chapter 5).

The placing of the Bambusoideae as the basal lineage in the grasses confirms at least to some extent the accuracy of the rooting of the rpoC2 topologies. In the latter topologies the base of the tree was usually unresolved, and at best a ((Bambusoideae, Pooideae), PACC) relationship was resolved.

The Bambusoideae

The rbcL data shows the Bambusoideae to be the basal (and paraphyletic) subfamily of the grasses. The relationship of the Bambusoideae to the Pooideae and PACC clades obtained in the rbcL analysis is consistent with those obtained by Hamby and Zimmer (1988) from rRNA sequence data, but conflicts with the parsimony analysis in the earlier rbcL study (Doebley et al. 1990) which suggested a (Pooideae (Bambusoideae, Panicoideae)) relationship.

The Pooideae

The pooid clade, shown to be monophyletic in Figures 3.2 and 3.4, is shown to be paraphyletic in the RNA-derived bootstrap

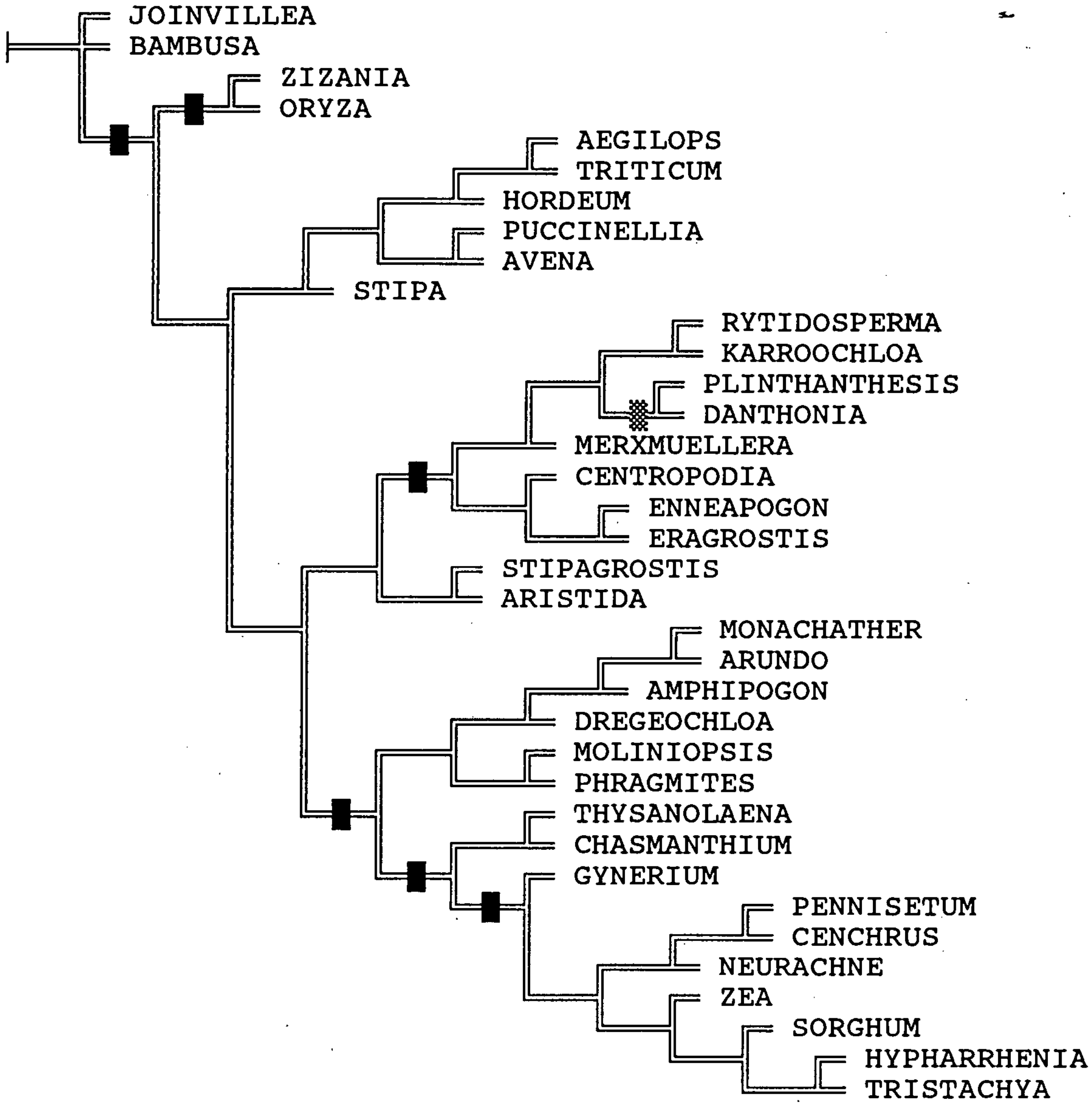


Figure 3.2. One of the 3 equally shortest trees obtained after successive weighting of the rbcL data set. The "▓" marks the node that collapses in the consensus tree of these three trees. The "█" characters mark the nodes that collapse in the consensus tree of the 26 equally shortest trees obtained by the initial m*bb* analysis of the data set.

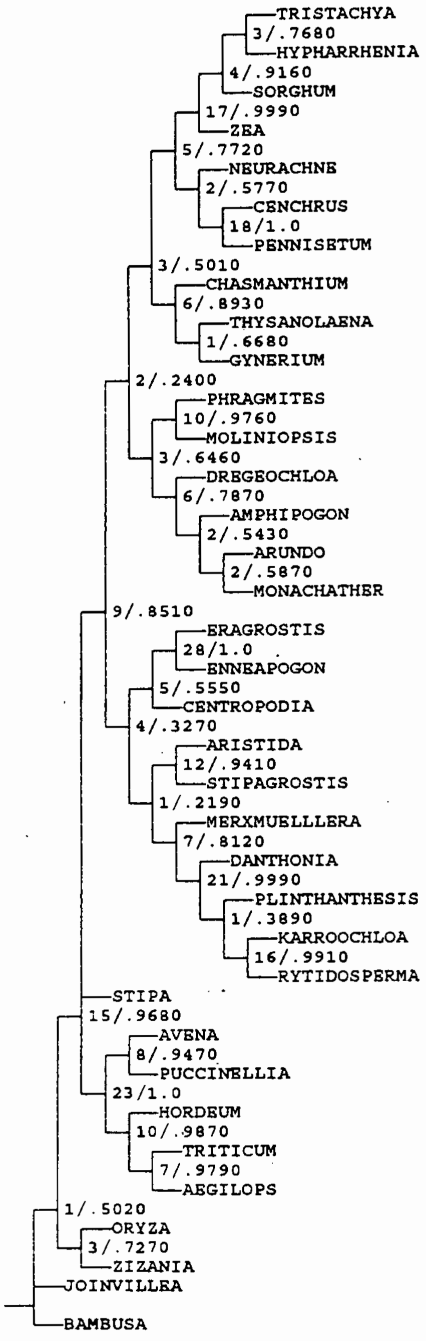


Figure 3.3. The Bremer support values (left of "/") and bootstrap support values produced from 1000 replicates of <u>rbcL</u> sequence data.

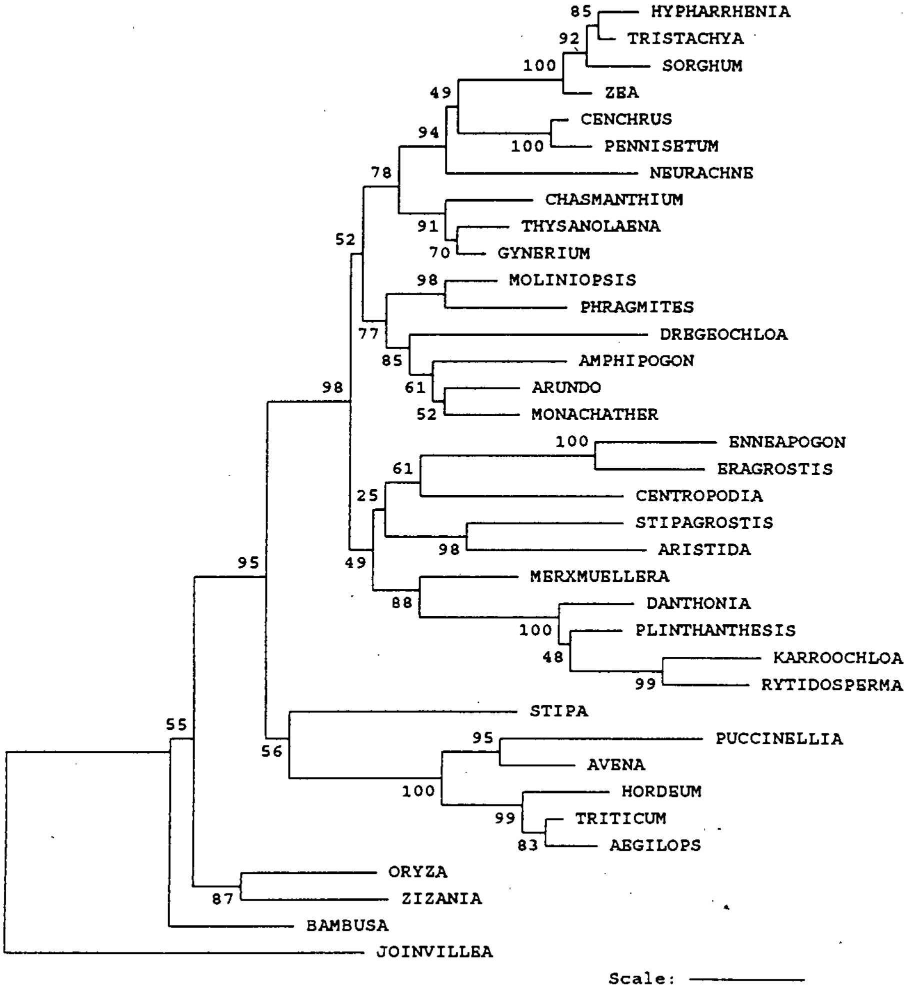


Figure 3.4. The Neighbour Joining tree from the rbcL data using the Jukes and Cantor correction, with bootstrap support values from 1000 replicates provided. Scale bar is approximately equal to the distance of 0.00867.

topology, as Stipa is shown to be paraphyletic to the pooid clade. (Figure 3.3). The core pooid clade is strongly supported, as can be seen from both the high bootstrap values (100%) and Bremer support index (23). In the Neighbour Joining analysis, Stipa is included within the pooids and this clade receives 56% bootstrap support (Figure 3.4).

The association of Stipa (Stipeae) with the pooid clade is contrary to Watson's (1990) classification, where it is placed in the subfamily Arundinoideae. However, as only three of the eight non-stipoid tribes in the Pooideae have been sampled in the rbcL study, the exact relationships of the Stipeae cannot be assessed. The inclusion of other taxa shown to have possible pooid affinities such as Brachyletrum, Nardus and Lygeum (Davis and Soreng 1993) may also affect this relationship.

The PACC clade

The PACC clade is well supported by the rbcL data, with a bootstrap value of 85% (Bremer support index = 9) for this node (Figure 3.3). This value is higher in the Neighbour Joining analysis (98% in Figure 3.4). The relationships between the representatives of the four subfamilies in this clade are, however, not completely resolved in the parsimony analysis. Successive weighting of the data set does, however, resolve the clade into a (Arundineae + Micraireae ((Centothecoideae, Thysanolaeneae), Panicoideae)) clade and a (Aristideae (Danthnoieae, Chloridoideae)) clade.

The relationships of the (Arundineae, PCT) clade are congruent with those obtained from the Neighbour Joining analysis, but the Aristideae are shown to be sister to the chloridoids instead of the danthonioids.

The Panicoideae

The clade comprising genera from the Panicoideae shows strong bootstrap support (77.2% in Figure 3.3, 94% in Figure 3.4). Within this subfamily, the genera sampled are divided into two lineages; the one corresponding to the tribe Paniceae, while the other contains genera from the Andropogoneae and Arundinelleae. The relationships agree with those proposed by Clayton and Renvoize (1986).

The Centothecoideae and Thysanolaeneae

The suggestion by Hilu and Wright (1982) that Thysanolaena might be allied to the Centothecoideae, and suggested by the rpoC2 analyses, is corroborated by rbcL data. In the parsimony analysis, (Figure 3.2), Thysanolaena and Chasmanthium are shown as sister taxa. However, the topology of the bootstrap tree shows Thysanolaena to be sister to Gynerium (66.8%), with Chasmanthium basal to these two taxa (89.3% bootstrap support). The Neighbour Joining tree agrees with this topology, the bootstrap values for the (Gynerium, Thysanolaena) clade being somewhat higher (70%). The support for the inclusion of Chasmanthium basal to these two genera receives 91% bootstrap support (Figure 3.4).

The Arundineae and Micraireae

This clade, which has a bootstrap support of 64.6% in the RNA-derived topology (Figure 3.3) and 77% in the Neighbour Joining tree (Figure 3.4), comprises the Arundineae sensu Watson 1990 as well as the previously danthonioid genera Monachather, Dregeochloa and Amphipogon. As Monachather is tentatively considered to represent the tribe Micraireae sensu Clayton and Renvoize (1986), its presence in this clade renders the Arundineae paraphyletic.

The Aristideae

The two representatives of this arundinoid tribe are well supported as a monophyletic clade (94.1% and 98% bootstrap support in Figures 3.3 and 3.4 respectively).

However, as discussed above, the Neighbour Joining and parsimony analysis are incongruent with respect to the relationships of the Aristideae. Successive weighting resolved the relationship as (Aristideae (Chloridoideae, Danthonieae)). The RNA-derived bootstrap topology (Figure 3.3) shows a (Chloridoideae (Aristideae, Danthonieae)) relationship, but support for this placing of the Aristideae is poor (21.9% bootstrap support), as is the support for the basal position of the Chloridoideae (32.7%).

The Neighbour Joining analysis (Figure 3.4) resolves these relationships as (Danthonieae (Aristideae, Chloridoideae)), but the (Aristideae, Chloridoideae) relationship receives poor bootstrap support (25%), and the support for a basal danthonioid clade is also weak (49%).

These three topologies (Figures 3.2, 3.3 and 3.4) thus retrieve all the possible combinations of relationships between these three lineages. The Neighbour Joining analysis (Figure 3.4) shows the branch length of the (Aristideae, Chloridoideae) clade to be very short. This may be indicative of rapid diversification of these lineages and thus imply that this relationship may never be satisfactorily resolved by *rbc*L sequence data. However, before accepting this explanation, additional samples from both the Aristideae (such as *Sartidia* and other species of the diverse *Aristida* and *Stipagrostis*) and the Chloridoideae (such as the unusual *Merxmuellera rangeii*, suggested by the *rpo*C2 data to be a basal chloridoid lineage) should be included.

The Danthonieae

There is strong bootstrap support (81.2% in Figure 3.3, 88% in Figure 3.4) and Bremer support (7 in Figure 3.3) for the danthonioid clade . In the consensus tree of the 26 equally parsimonious trees, the relationships among Danthonia, Plinthanthesis and the (Rytidosperma, Karroochloa) clade are equivocal. Despite this lack of resolution, the topology of the danthonioid clade is generally congruent with that obtained in the rpoC2 analyses, with Merxmuellera (represented by M. macowanii) being the basal taxon, and Rytidosperma and Karroochloa shown as sister taxa.

The Chloridoideae

The inclusion of Centropodia, previously considered to be danthonioid, at the base of the chloridoid clade is weakly supported by the bootstrap values (55.5% in Figure 3.3, 61% in Figure 3.4), but is supported by anatomical and morphological features discussed in the next chapter, and corroborates the phylogenies obtained from the rpoC2 data. As only two of the five chloridoid tribes recognised by Clayton and Renvoize (1986) have been sampled in the rbcL study, additional sampling may clarify both the tribal relationships and the position of Centropodia within this subfamily.

The utility of rbcL

It is now possible to re-assess the performance of rbcL in elucidating the phylogeny of the grass subfamilies. Using a suitable outgroup and sampling all the major grass lineages, rbcL appears to be sufficiently variable to resolve relationships among most of the subfamilies and tribes of the grass family.

Even below the level of the subfamily, the gene is able to provide phylogenetic information. This is shown by the number of informative sites obtained from the sequences within each of the various lineages. For instance, within the six-taxon pooid clade, 20 phylogenetically informative sites are obtained, the panicoid clade (seven taxa) yields 26 sites, while the danthonioid clade (five taxa) yields 18 informative sites.

The utility of rbcL in solving systematic problems within the grasses thus appears to be dependant on sample size and sample selection. The failure of the data to provide a well resolved phylogeny may not be attributable to the nature of the gene, but rather to insufficient sampling, the absence of certain "missing link" taxa, or to the rapid radiation of several lineages in the evolution of the subfamilies.

CONCLUSIONS

1. Based on the samples sequenced here, the rbcL data supports a (Bambusoideae (Pooideae, PACC)) relationship.

The monophyly of the Bambusoideae is not testable owing to insufficient sampling, and is central to this study. The pooid clade is shown to be monophyletic, strongly supported as indicated by the bootstrap values, and include the Stipeae. The existence of a monophyletic PACC clade as elucidated by Davis and Soreng (1993) is corroborated here.

2. The rbcL sequence data is unable to resolve some of the relationships of the major lineages within the PACC clade.

The presence of a basal trichotomy in the PACC clade in the consensus topology suggests that the gene is insufficiently variable to be taxonomically useful, possibly as a result of an initial rapid radiation during the evolution of the PACC clade. Such a radiation would result in numerous short branches which would not be resolvable with any degree of confidence. The short branch lengths of some of these clades shown in the Neighbour Joining topology suggest that this might be the case.

3. The observation that rbcL does not produce an unambiguous position for some taxa may not be because the gene is too conservative, but rather a result of inadequate sampling.

Where comparable sampling intensity exists (as in the Panicoideae and Danthonieae), the rbcL data retrieves relationships congruent with those obtained using the highly variable rpoC2 insert region. Both rbcL and rpoC2 are, however, plastid genes and congruence between them provides additional confidence that they present an accurate estimate of the plastid phylogeny.

4) In the light of the above results, rbcL is informative about relationships within the family, but this is dependant to some extent on sample selection and taxon density.

CHAPTER 4

ANALYSIS OF COMBINED MOLECULAR DATA SETS

INTRODUCTION

Molecular (i.e. usually DNA-based) data may be combined with other molecular data, or with morphological data. However, the use of plastid-derived molecular data in a combined analysis needs to be handled with care. Very often, knowledge of plastid inheritance systems in the study group is scanty. Furthermore, hybridisation events may serve to obscure relationships between the organisms. As a result of these factors, the plastid phylogeny will not necessarily reflect the organismal phylogeny, resulting in conflicting phylogenetic hypotheses obtained when the plastid data are compared to other data sets.

As outlined by Williams (1994), there are two approaches to resolving such conflict. Either the topologies from the two data sets may be combined (the taxonomic congruence, or consensus approach), or the data sets are combined and re-analysed (the character congruence, or combined data approach). Both these approaches attempt to maximise evidence; one directly from the data, the other from the relationships implied by the data.

The justification of the consensus approach is based on the idea that different data sets are independent (De Queiroz 1993 and references therein). De Queiroz (1993) recommends that consensus be used in instances where different data sets produce conflicting trees that are well supported. He further recommends that where non-independence of characters within the data sets is suspected, and support for the conflicting trees is weak, then both consensus and combined analyses should be carried out. Only in instances where there is independence of characters within data sets should a combined approach be used.

Eernisse and Kluge (1993) list four criticisms of the consensus method. These are

1) The consensus of the fundamental cladograms (those derived individually from each of the two data sets, should there be more than one shortest tree) can be misleading.

2) The different data sets being analysed are weighted equally, but differences in data set size can weight the constituent characters differently.

3) The basis on which consensus of suites of equally parsimonious fundamental cladograms is achieved, is not clear.

4) The partitioning of evidence into classes (such as "molecular" or "morphological") is artificial.

For these reasons, Kluge (1989) and Eernisse and Kluge (1993) support the combined data approach when dealing with more than one data set.

Because they are both plastid genes, sequence data from the rpoC2 and rbcL sequences are not independent. Thus, according to De Queiroz's logic, the approach of analysing the combined data should be adopted (De Queiroz 1993). This approach may be viewed as more useful in that the increased number of characters may allow a closer approximation of the true phylogeny, and it can produce a tree that is more resolved than a consensus tree (De Queiroz 1993).

Other than the above considerations, the practicalities of combining data sets needs to be given some attention. In this and other instances (such as Kellogg and Linder, in press), the data sets have been produced by different researchers on different continents. Potential problems include the limited overlap or commonality of the taxa in each data set, and the fact that data from representatives of the overlapping taxa may possibly (and even probably) not have been taken from the same sample, or even be from the same species. Thus in combining data taken from two different species of a genus (where the genus is the terminal unit in the analysis, or in phenetic terminology the OTU), the assumption that the genus is monophyletic has to be made and that the exemplars chosen are "typical" of that genus.

Molecular data sets available for the Poaceae

The existence of several molecular data sets for the same set of organisms is not common, but the Poaceae and Asteraceae are exceptions. Both families have been the subject of extensive sequencing and restriction site studies. Combined analyses of such data sets are now possible, and have already been carried out for the Poaceae by Kellogg and Linder (in press) and Asteraceae (Olmestead and Sweere 1994).

There are several molecular data sets for the grasses. The overlap between these data sets in terms of the taxa sampled varies, and there has been no collaborative effort among the various researchers to sample a core set of taxa. Thus past studies are either incomplete in terms of taxic diversity, or there has been an over-emphasis on sampling certain groups (as outlined in Table 1.3).

Only three studies have a sufficiently wide sampling range (i.e. include all the subfamilies) to be of any value in a combined analysis. These are the rpoC2 and rbcL sequencing studies presented here, and the cpDNA restriction site mapping study of Davis and Soreng (1993). Other published studies include the rps4 sequencing study (Nadot et al. 1994) and the nuclear rRNA study of Hamby and Zimmer (1988). Work still in progress includes an ndhF sequencing study, concentrating on the Bambusoideae (Clark et al. in prep.) and the continuing site mapping efforts of Davis and Soreng. The rps4 sequences (including some unpublished ones) were obtained from Sophie Nadot. Unfortunately, Nadot's sampling range does not overlap widely with any of the other data sets, and she did not include Joinvillea. Combined analyses with the rps4 data was not carried out because of these factors.

Although the rpoC2, rbcL and site mapping data sets were analysed by Kellogg and Linder (in press), all three data sets have been further expanded, and this has resulted in an increase in "commonality" among the data sets. The terminal unit in these analyses is the genus. As intimated above, the data for these units in each of the data sets may have been taken from different species, and the monophyly of these genera has to be assumed.

MATERIAL AND METHODS

Three sources of plastid cpdna data are available for analysis in a variety of configurations: the rpoC2 and rbcL sequence data, and restriction site mapping data, kindly provided by Jerry Davis (Cornell University, Ithaca). The use of the rpoC2 sequences is problematic, given the differing results of the various alignments analysed above. However, for the purposes of these analyses, the alignment at stringency 2 is used, as it is considered here to be the alignment which best explains the biological realities (homologies) of the heptameric repeats.

Unfortunately, no site mapping or rbcL sequence data is available for Micraira (tribe Micraireae; Arundinoideae). However, as Micraira is consistently placed as sister to

Monachather in the rpoC2 analyses, the relationships of the Micraireae may be tentatively extrapolated on the basis of this relationship.

All analyses were carried out using the m*bb* options in HENNIG86. Because RNA (Farris 1994) will work on sequence data only, bootstrapping (Felsenstein 1985) was performed on all data sets by means of PAUP version 3.1.1 (Swofford 1993) using 500 replicates. The combinable components consensus tree (Bremer 1990) of the strict consensus trees obtained from each of the individual data sets was obtained using the "semi strict" consensus option in PAUP.

RESULTS

Combined rpoC2 and rbcL data

This data set contained 28 taxa and 253 phylogenetically informative characters. Unfortunately, while the remaining subfamilies were well represented, the Pooideae and Centothecoideae are represented by single taxa. The m*bb* analysis resulted in a single shortest tree, shown in Figure 4.1. Bootstrap support values are shown in this figure. Figures 4.2 and 4.3 show the phylogeny from an m*bb* analysis of the two data sets separately, and Figure 4.4 the combinable component consensus tree.

All topologies (Figures 4.1 to 4.4) resolve a monophyletic PACC clade, which receives 100% bootstrap support in the analysis of the combined data (Figure 4.1). One branch of the basal dichotomy in the PACC clade comprises the Centothecoideae, Panicoideae, Thysanolaeneae and Gynerium, with the remaining arundinoid and chloridoid taxa included in the other branch. The inclusion of the Centothecoideae and Thysanolaeneae in the panicoid lineage is thus strongly supported (93%), but the sister relationship between the Centothecoideae and the Thysanolaeneae is not as well supported (58%). The placing of Gynerium basal to the panicoids also receives only moderate support (52%).

The monophyly of the clade comprising the Chloridoideae and polyphyletic Arundinoideae is poorly supported (28%). The Danthonieae are placed sister to the Chloridoideae, a relationship that is weakly supported (56% bootstrap support), and the Aristideae are placed basal to these two lineages; a position that also receives 56% bootstrap support.

The combinable component tree is somewhat less resolved, with the positions of Amphipogon and Dregeochloa being equivocal, and the relationships among the various clades within the PACC clade are not resolved.

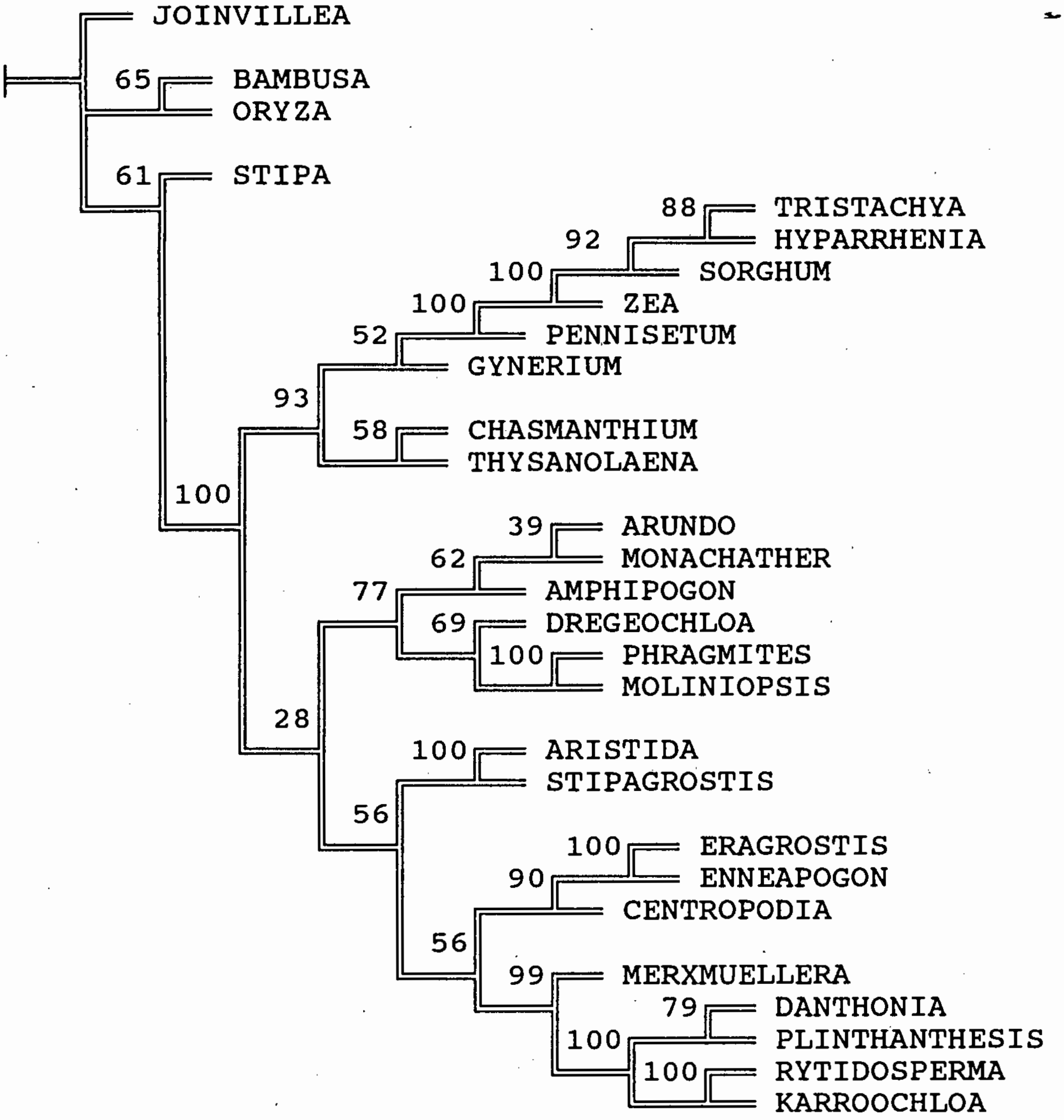


Figure 4.1. The single shortest tree obtained from an m*bb* analysis of the combined rpoC2 and rbcL data sets. Bootstrap support values from 500 replicates are provided. Length = 622, ci = 50, ri = 65.

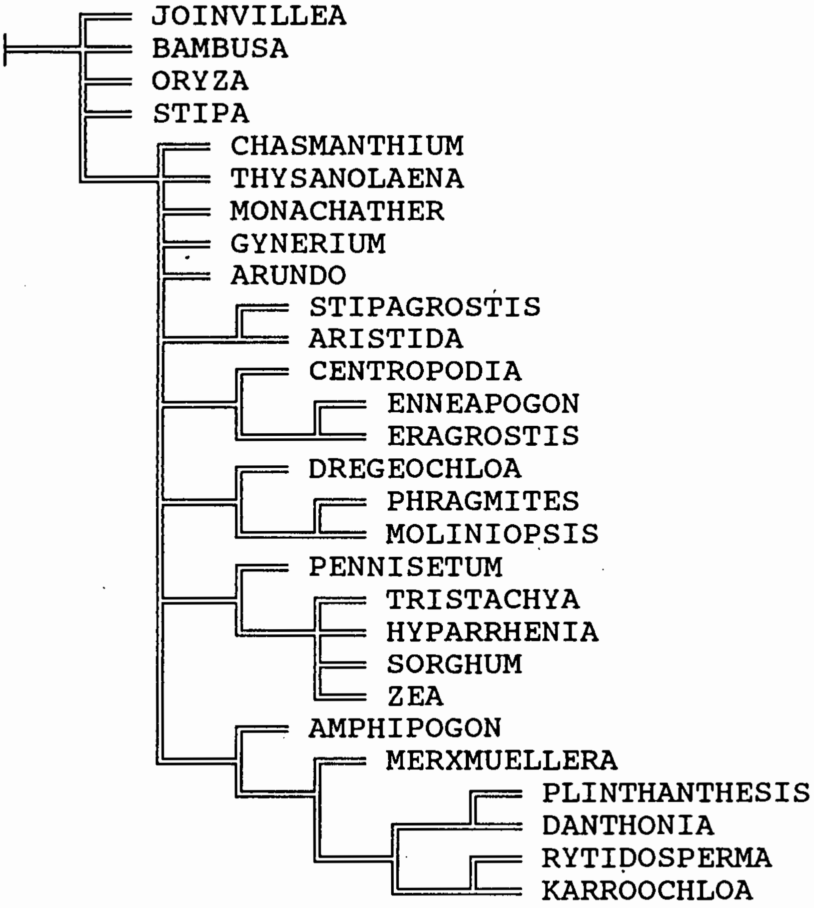


Figure 4.2. The strict consensus tree of the 237 fundamental trees obtained from the analysis of the <u>rpo</u>C2 data set with only those taxa in common with the <u>rbc</u>L data set. Length 237, ci = 56, ri = 67.

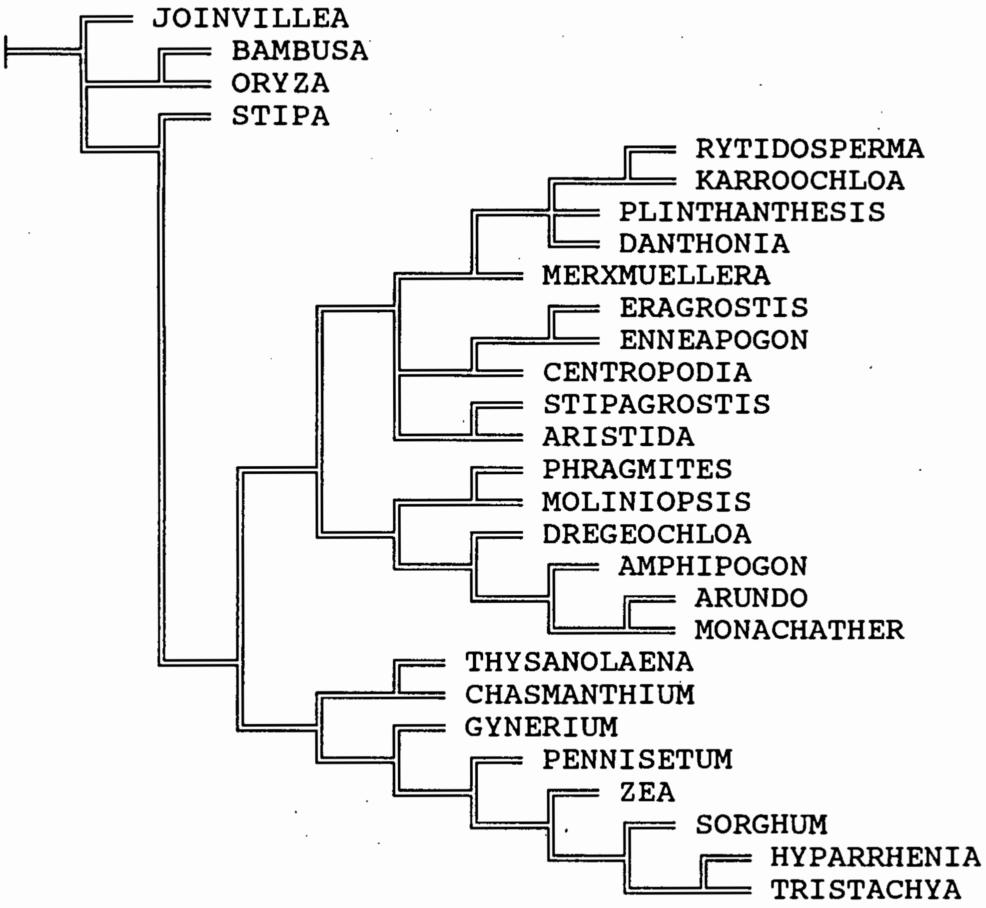


Figure 4.3. The strict consensus tree of the 65 fundamental trees obtained from the analysis of the rbcL data set with only those taxa in common with the rpoC2 data set included. Length = 369, ci = 47, ri = 65.

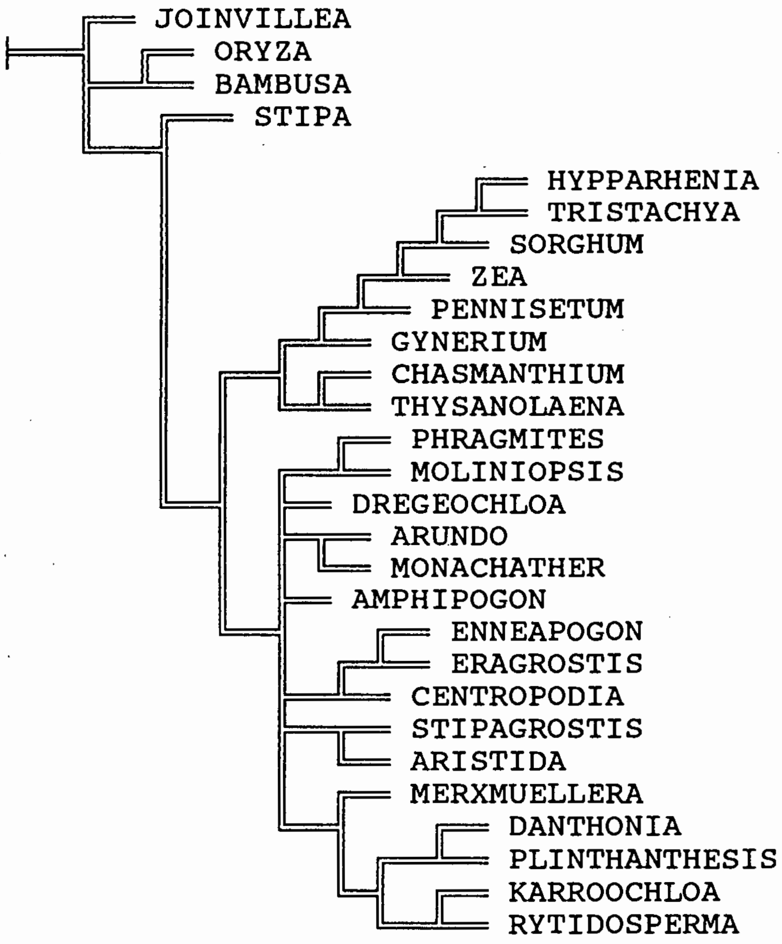


Figure 4.4. The combinable components (semi-strict) consensus tree of two strict consensus trees derived from the fundamental trees of each of the rbcL and rpoC2 data sets.

Combined rpoC2 and site mapping data

This data set comprised information from 27 taxa, and was made up of more site mapping characters than sequence characters (219 versus 130). However, invariant and autapomorphic characters were not removed from the latter data set. Although there were substantial areas of missing site mapping data for Thysanolaena, it was included in the analysis. The m*bb* analysis resulted in a single most parsimonious tree of length 607 steps (ci = 59, ri = 61). This tree is shown in Figure 4.5, which also shows bootstrap values.

These trees show relationships that have previously not been obtained in either the rbcL or rpoC2 studies:

1) The (Danthonia, Rytidosperma) clade appears as the basal group of the PACC clade. While the PACC clade as a whole is well supported (99% bootstrap value), the monophyly of the remaining PACC clade taxa above the (Danthonia, Rytidosperma) clade is not well supported (31% bootstrap support).

2) The reedy (Phragmites, Moliniopsis) clade is well supported (99%), but basal to the rest of the PACC clade (excluding the danthonioid lineage discussed above) and not associated with Arundo.

3) Aristida appears basal to the chloridoid lineage, but this position is poorly supported by the bootstrap analysis (32% bootstrap support).

4) The Arundineae (as delimited in the rbcL study presented in Chapter 3) are paraphyletic, with the Aristideae and Chloridoideae embedded within them. However, this entire clade is very weakly supported (18% bootstrap support).

The topology of the strict consensus tree of the fundamental trees obtained from the analysis of the site mapping data set is poorly resolved (Figure 4.6). In contrast, the strict consensus tree of the fundamental trees from the rpoC2 data shows much more resolution (Figure 4.7), although Arundo is placed in the somewhat unusual position as the basal taxon in the PACC clade. The combinable component consensus tree (Figure 4.8) is well resolved, although the relationships it suggests are unusual.

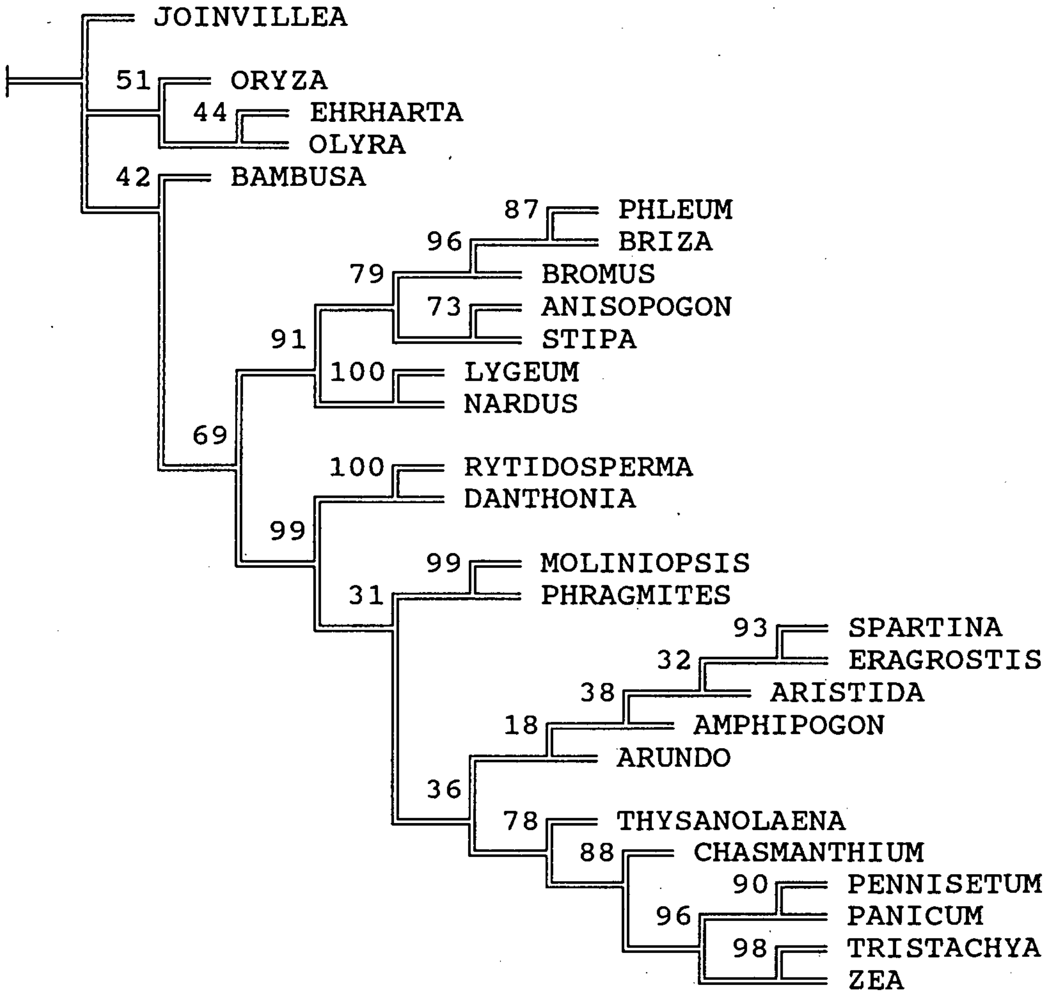


Figure 4.5. The single most parsimonious tree obtained from an m*bb* analysis of the combined rpoC2 and site mapping data. Bootstrap support values from 500 replicates are provided. Length = 607, ci = 49, ri = 61.

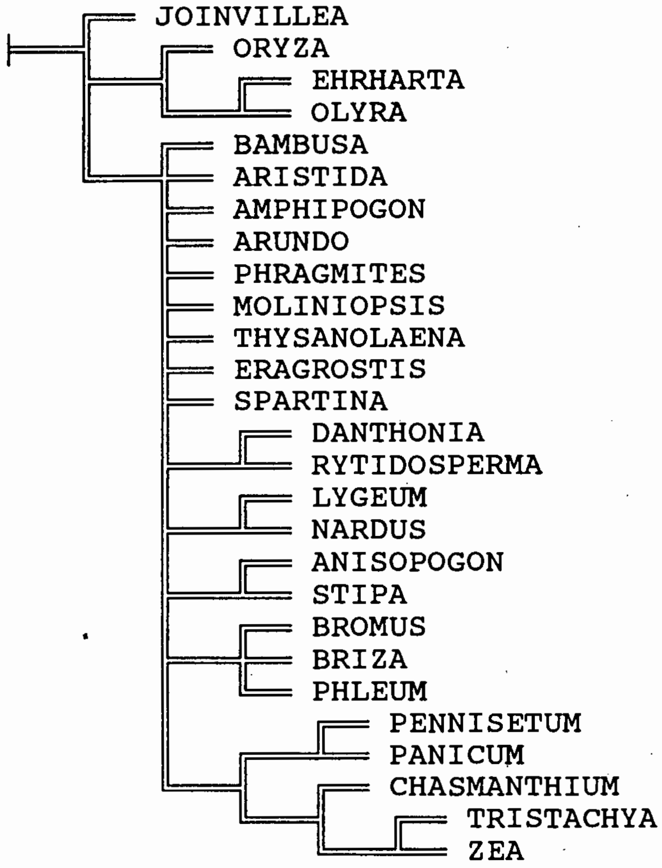


Figure 4.6. The strict consensus tree of the 192 fundamental trees obtained from the analysis of the site mapping data set with only the taxa common to the rpoC2 data set included. Length = 296, ci = 46, ri = 54.

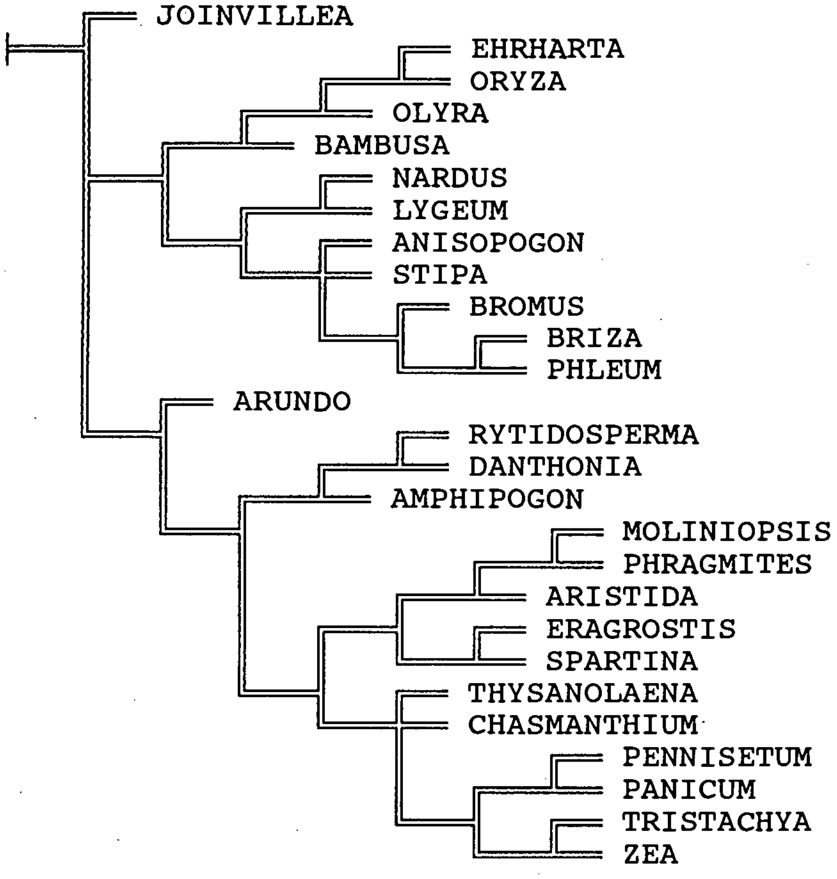


Figure 4.7. The consensus tree of the topology of the 8 equally parsimonious trees obtained from the analysis of the rpoC2 data set with only the taxa common to the site mapping data set included. Length = 300, ci = 53, ri = 68.

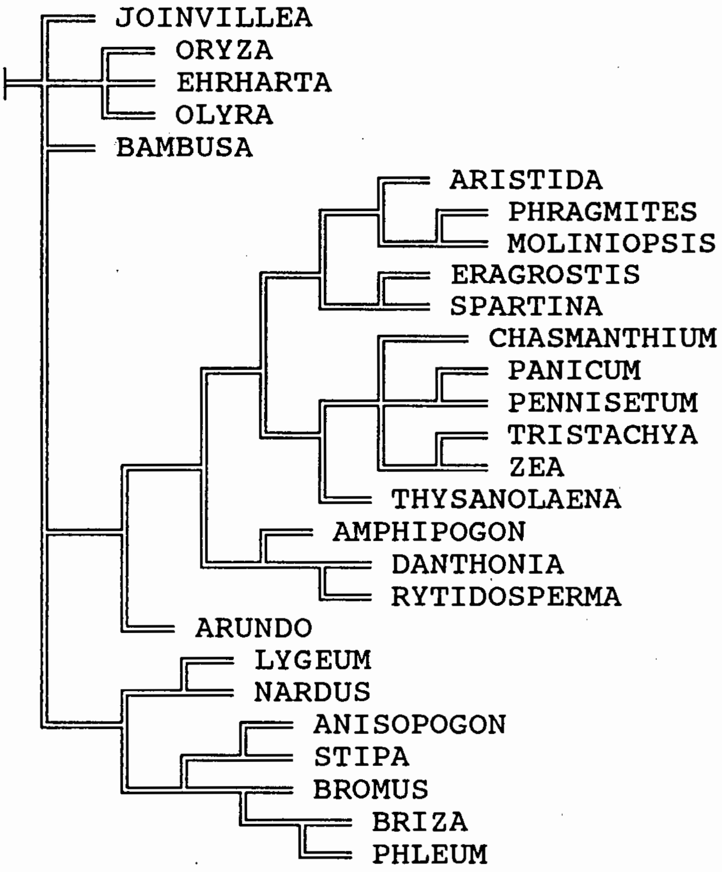


Figure 4.8. The combinable components (semi-strict) consensus tree of strict consensus trees derived from the fundamental trees of each of the rpoC2 and site mapping data sets.

Combined rbcL and site mapping data

This data set comprised 343 characters (124 from the rbcl data 219 from the site mapping data) and 20 taxa. The m*bb* analysis of the combined data set produced two equally parsimonious trees of length 552 (ci = 50, ri = 56), the consensus of which is shown in Figure 4.9. The two shortest trees differ only in the relative positions of Thysanolaena and Chasmanthium.

In this tree, the Bambusoideae are shown to be the basal lineage, although the sister group relationship between Bambusa and Oryza is only moderately supported (52% bootstrap support). The monophyletic Pooideae are well supported (74 % bootstrap value) as is the PACC clade (96%). Within the PACC clade the Aristideae are placed as sister to the Chloridioideae, a position that receives good bootstrap support (76%). The Danthonieae are sister to this clade, at a lower bootstrap support value of 62%. This Danthonieae - Aristideae - Chloridoideae assemblage is sister to the Arundineae (59% bootstrap support). The inclusion of the centothecoid and thysanolaenoid taxa in the panicoid lineage receives bootstrap support of 70%. This topology is thus in general agreement with that obtained from the analysis of the rbcL data alone, both as a complete data set (Figure 3.2 in the previous chapter) and as a subset of taxa common to the site mapping data set (Figure 4.10).

In contrast, the strict consensus tree of the fundamental trees resulting from the analysis of the site mapping data is poorly resolved (Figure 4.11). However, the combinable components consensus tree (Figure 4.12) is well resolved, and its topology is almost identical to the tree based on rbcL data alone (Figure 4.10).

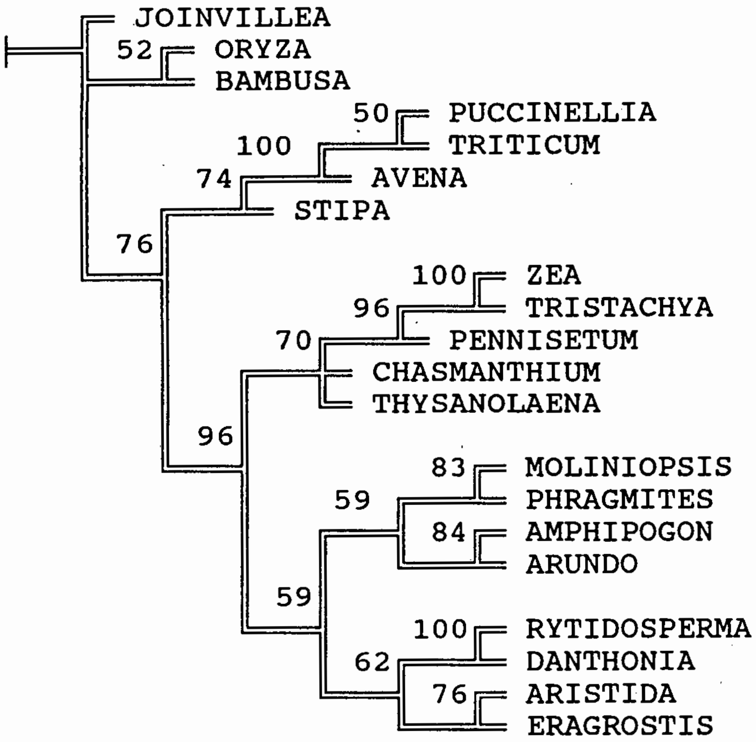


Figure 4.9. The strict consensus tree of the 2 equally parsimonious trees obtained from the combined <u>rbc</u>L and site mapping data set. Length = 552, ci = 50, ri = 56.

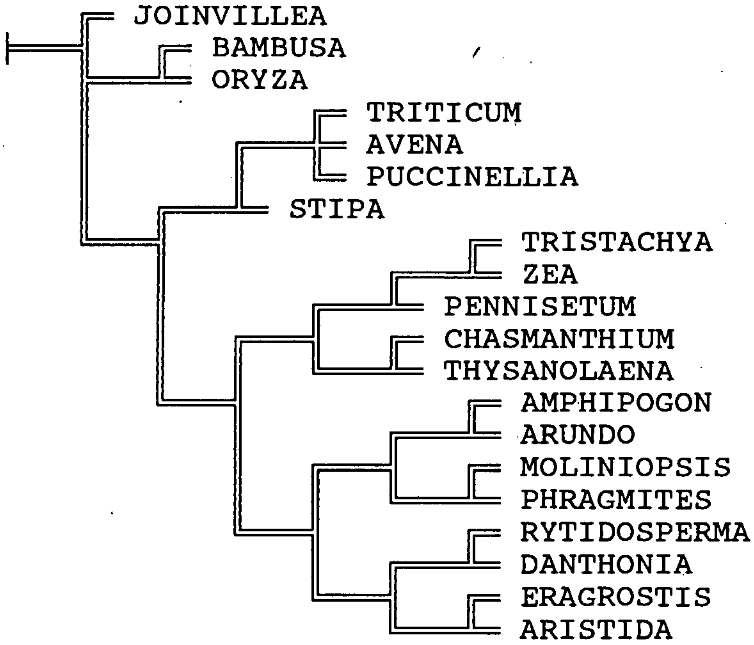


Figure 4.10. The strict consensus tree of the 2 fundamental trees from the analysis of the rbcL data set for the taxa in common with the site mapping data set. Length = 333, ci = 47, ri = 57.

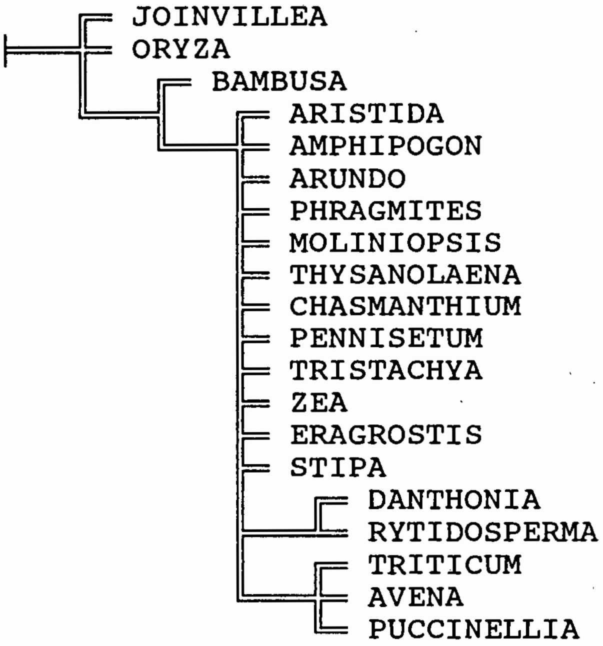


Figure 4.11. The strict consensus tree of the 255 fundamental trees from the analysis of the site mapping data set for the taxa in common with the rbcL data set. Length = 210, ci = 56, ri = 57.

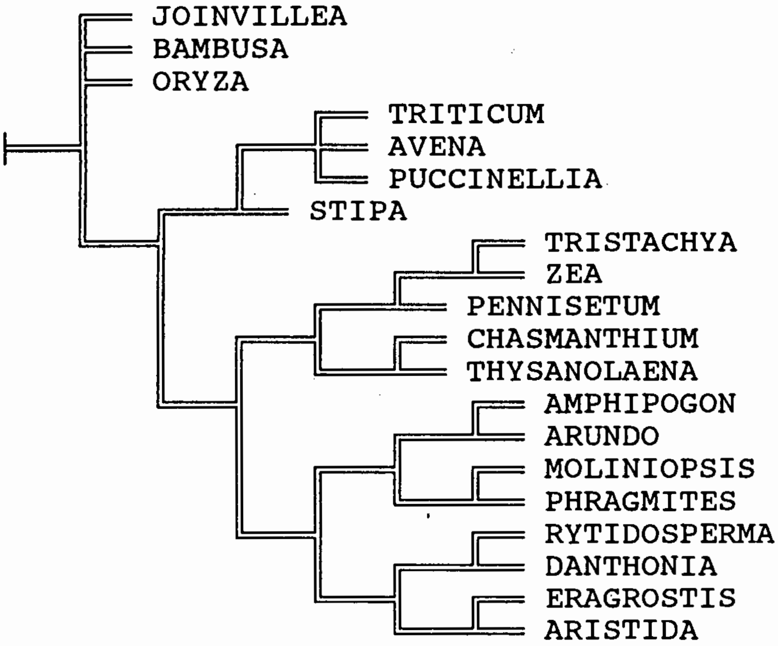


Figure 4.12. The combinable components (semi-strict) consensus tree of the two strict consensus trees derived from the fundamental trees of each of the rbcL and site mapping data sets.

Combined rbcL, rpoC2 and site mapping data

This data set comprised 17 taxa (including the outgroup, Joinvillea) and a total of 404 characters. Of the six subfamilies, the Pooideae, Chloridoideae and Centothecoideae are each represented by only one taxon.

The m*bb* analysis of this data set retrieved a single shortest tree, shown in Figure 4.13. The topology of this tree reflects what has become a fairly common set of relationships between the three major clades: (Bambusoideae (Pooideae, PACC clade)). The PACC clade receives 100% bootstrap support. Within this clade, the Aristideae are shown to be sister to the Chloridoideae (67% bootstrap support), while the danthonioid lineage is sister to these two taxa, although this relationship is weakly supported (33% bootstrap support). The Arundineae are sister to the (Danthonieae (Aristideae, Chloridoideae)) clade, but this relationship is also poorly supported by the bootstrap analysis (36%). The remaining tribe of the Arundinoideae, the Thysanolaeneae, is shown once again to be associated with the basal lineage of the (Panicoideae, Centothecoideae) clade, a position that is well supported (85% bootstrap support).

Of the strict consensus trees of the fundamental trees from the individual data sets (Figures 4.14, 4.15 and 4.16), only the rpoC2 (Figure 4.14) and rbcL topologies (Figure 4.15) are well resolved. Once again, the site mapping data is unable to resolve the relationships of the taxa common to these three data sets. The combinable component consensus tree from these three strict consensus trees is not well resolved (Figure 4.17). The PACC clade is retained, and only the large clade within it comprises the ((Thysanolaeneae, Centothecoideae) Panicoideae). The remaining taxa are placed in two-taxon relationships or part of the PACC polytomy.

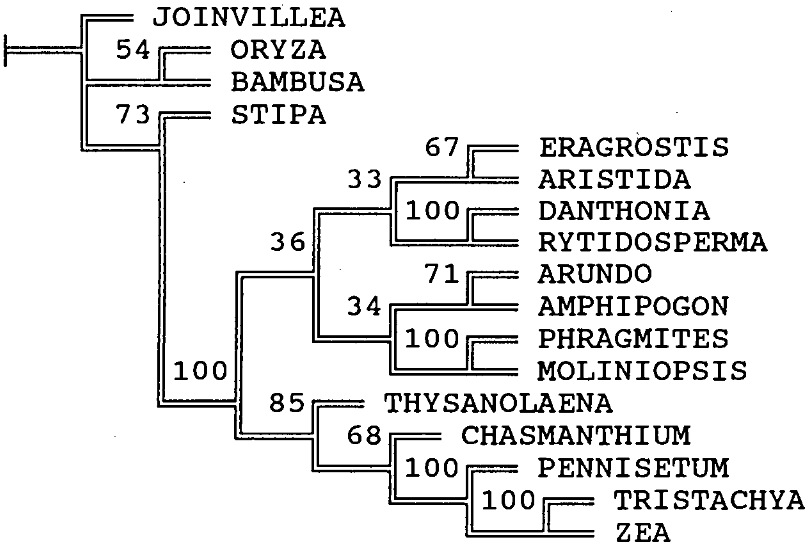


Figure 4.13. The single shortest tree obtained from the analysis of the combined rpoC2, rbcL and site mapping data sets; a total of 404 characters for 17 taxa. Length = 581, ci = 55, ri = 55.

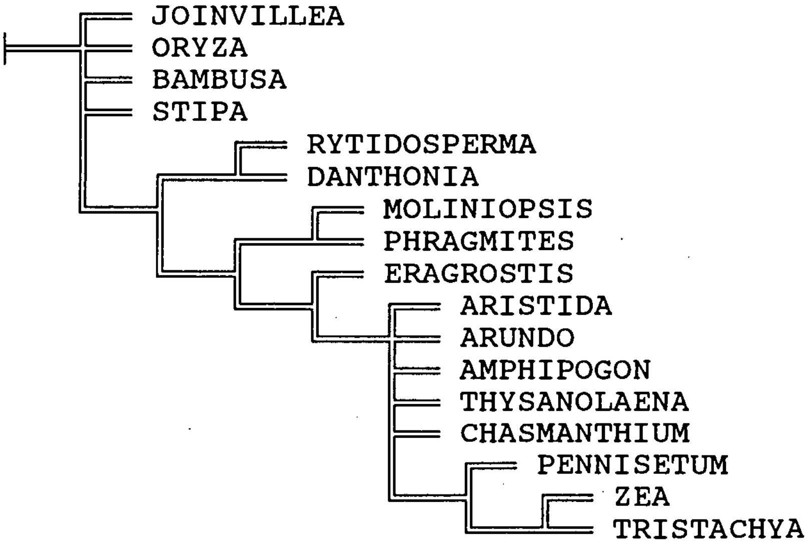


Figure 4.14. The strict consensus tree of the 70 fundamental trees from the analysis of the rpoC2 data set for the taxa in common with the rbcL and site mapping data sets. Length = 143, ci = 62, ri = 65.

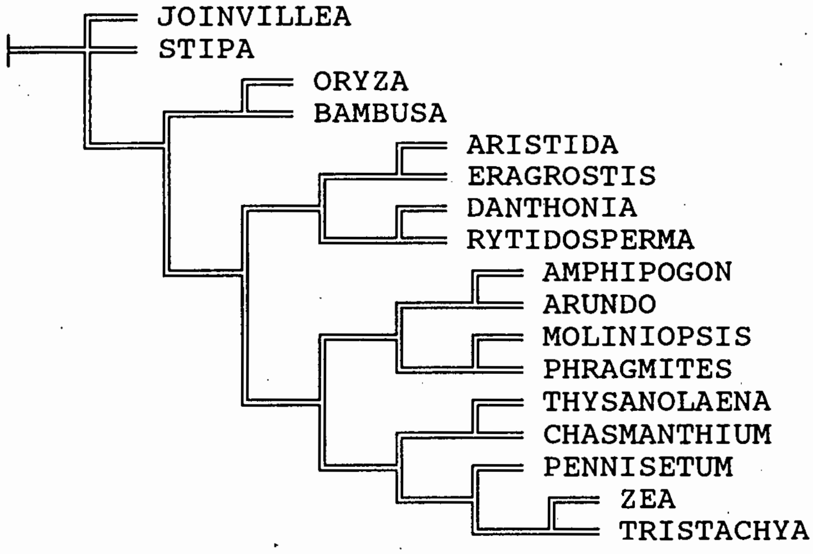


Figure 4.15. The single most parsimonious tree obtained from the analysis of the *rbcL* data set for the taxa in common with the *rpoC2* and site mapping data sets. Length = 261, ci = 50, ri = 55.

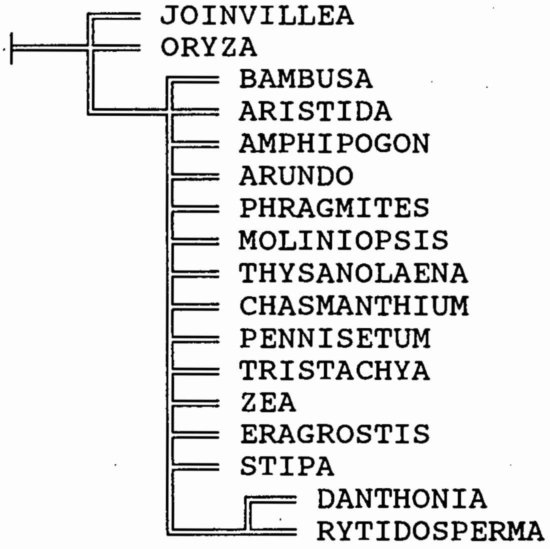


Figure 4.16. The strict consensus tree of the 308 fundamental trees from the analysis of the site mapping data set for the taxa in common with the *rbcL* and *rpoC2* data sets.
Length = 156, ci = 62, ri = 55.

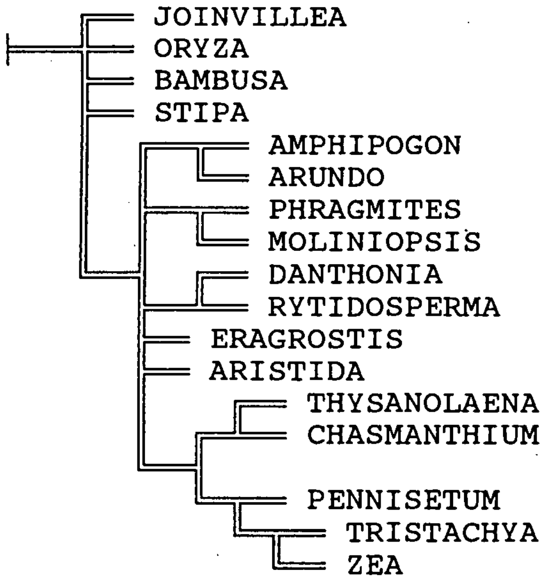


Figure 4.17. The combinable components (semi-strict) consensus tree of the three strict consensus trees derived from the fundamental trees of each of the rbcL, rpoC2 and site mapping data sets.

DISCUSSION

The analysis of the combined data sets results in well resolved topologies, although the bootstrap support for some nodes may be weak. In contrast, the combinable component consensus approach results in somewhat less resolved topologies.

The resolution of a tree may be "measured" by comparing the number of nodes on the tree (N_{obs}) to the theoretical maximum number of nodes (N_{max}), where N_{max} = (number of taxa - 1).

$$\text{Percent Resolution} = \frac{N_{obs}}{N_{max}} \times 100 \qquad \text{(equation 1)}$$

When this measure is calculated for the three possible two-data set combinations, it is found that the combinable component consensus trees (Figures 4.4, 4.8 and 4.12) are all about 80% resolved (Table 4.1). This is in contrast to the resolution of the trees from the analysis of the combined data (94 - 100% resolved; Table 4.2).

The almost identical degree of resolution of the combinable components topologies (approximately 80%) indicates that the contributing data sets are equally congruent (or incongruent), and no single data set is disabling the analysis. The 20% lost resolution may thus be used as a measure of the incongruence of the data sets. If each data set is contributing to the incongruence equally, then they are each only 10% incongruent from the true topology. However, if one of the data sets reflects the true topology, then the other data set is 20% incongruent.

Table 4.1. The percentage resolution of the combinable-component topologies of each two-data set combinations as measured by equation 1, given in the text above.

	rpoC2	site mapping
rbcL	81.98%	84.21%
rpoC2		80.76%

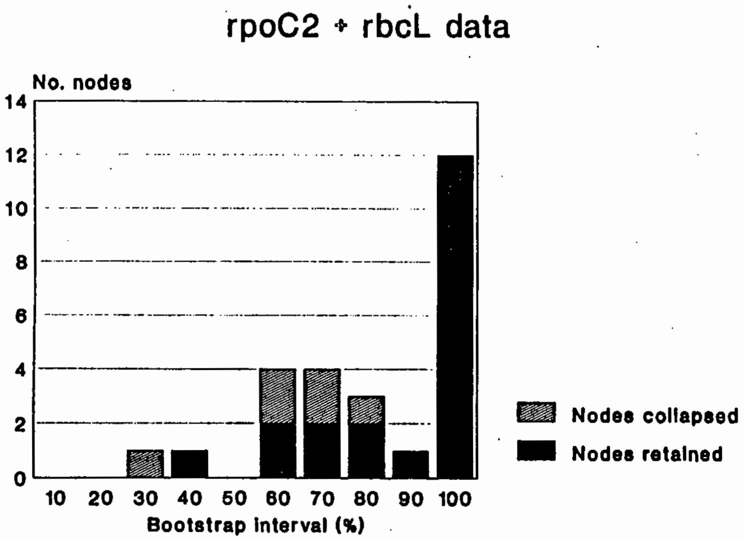
Table 4.2. The percentage resolution of the topologies from the analysis of the combined data of each of the two-data set combinations, as measured by equation 1, given in the text above.

	rpoC2	site mapping
rbcL	100.00%	94.73%
rpoC2		100.00%

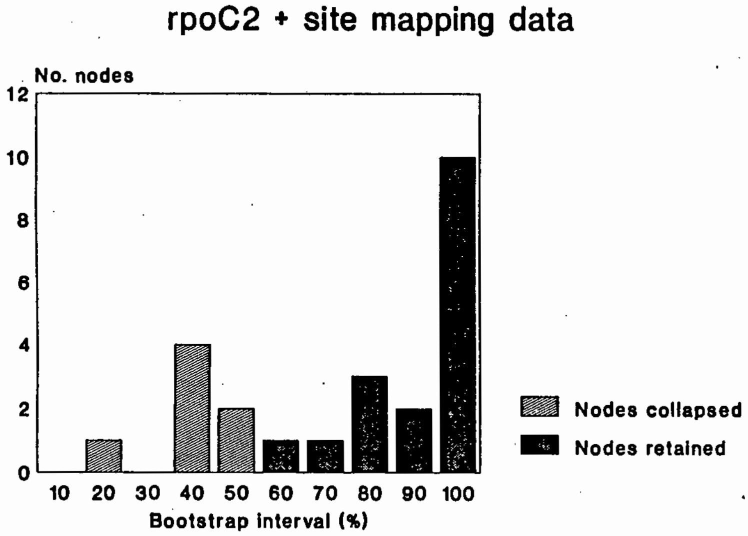
Thus, if resolution is to be used as a measure of usefulness, analysis of combined data performs better than using a consensus approach. However, as the consensus approach is expected to produce a more conserved estimate of phylogeny (Hillis 1987), this result is not unexpected. It is, however, interesting to note which nodes collapse in the combinable component topologies and relate this to the bootstrap support these nodes receive in the combined data analysis. This may be simply illustrated by means of a histogram that plots the number of nodes against a bootstrap support interval, and noting the classes which contain collapsed nodes in the combinable component topologies.

These histograms (Figures 4.18(a-c) all indicate that it is the nodes with the lower bootstrap support that are likely to collapse in the combinable component topologies. No absolute value can be put on the boundary of bootstrap support that suggests a node would collapse in a combinable components

4.18(a)



4.18(b)



4.18(c)

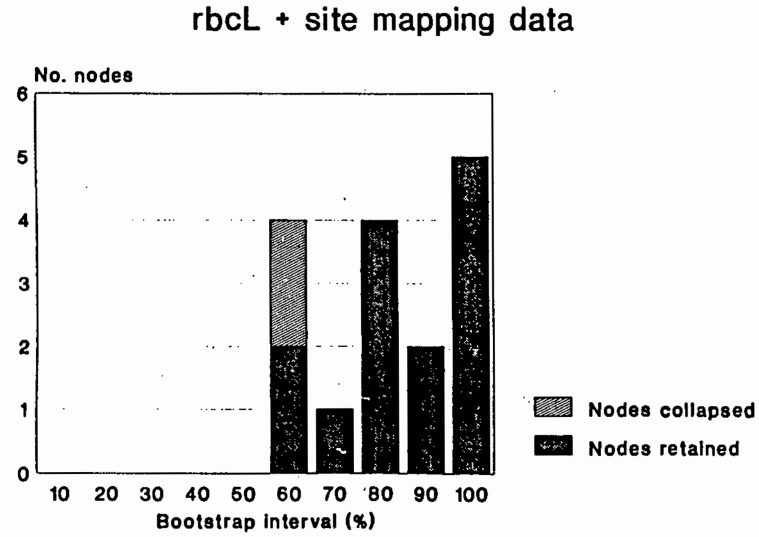


Figure 4.18(a-c). Histograms plotting the number of nodes which collapse in the three combinable component consensus topologies against the bootstrap support value those nodes receive in each of the analyses of the combined data sets. Figure 4.18(a): combination of rpoC2 and rbcL data; Figure 4.18(b): combination of rpoC2 and site mapping data; Figure 4.18(c): combination of rbcL and site mapping data.

topology, as this is data-set specific. However, the upper bound for bootstrap values of nodes likely to collapse is 77% (Figure 4.1). Generally, these values are less than 70%, which lends support to Hillis and Bull's (1993) contention that, under conditions thought to be typical of most phylogenetic analyses, bootstrap values provide conservative estimates of the probability of correctly inferring a clade. Hillis and Bull (1993) found that bootstrap values in excess of 75% translated into a probability greater than 95% that the corresponding clade is real, or <u>accurate</u> to use their terminology. In contrast, the clades retained in the combinable component topology indicate <u>repeatability</u> (the probability that a specified group will be found in an analysis of an <u>independent</u> sample of characters; Hillis and Bull, 1993). There is thus an intricate relationship between accuracy (resolution of combinable component topology) and repeatability (bootstrap values of, in this instance, topologies from combined data sets).

As no single data set seems to be overly incongruent with the others, the unusual topology obtained from the analysis of the combined <u>rpo</u>C2 and site mapping data, and the low bootstrap support values of some of the nodes (Figure 4.5) needs an alternative explanation. An examination of the tree obtained from an analysis of the site mapping data (for the subset of common taxa) shows that the data are unable to resolve many of the relationships (Figure 4.6). This suggests that there is substantial homoplasy in the site mapping data set, some soft polytomies, or both. The topology derived from the <u>rpo</u>C2 data subset for this combination is unusual but well resolved (Figure 4.7). The combinable components topology (Figure 4.8) is well resolved, but shows a very different set of relationships when compared to the other topologies from this combination. Furthermore, this topology contains relationships not shown by the analysis of the combined data. This suggests that one (or both) data sets are suggesting an hypothesis of relationships based on false signals and is what Felsenstein

(1978) called "positively misleading". Despite the fact that those supporting combined analysis argue that the increase in characters will resolve conflict among the characters and result in a better estimate of phylogeny (Barrett et al. 1991), the result of the rpoC2 / site mapping combination (Figure 4.5) is in conflict with all other topologies obtained in this and previous chapters.

In contrast, the combination of the rbcL data with the site mapping data results in a well resolved and generally well supported topology (Figure 4.9). However, the analysis of the site mapping data (from taxa common to the rbcL data set) once again produces a tree that is very poorly resolved (Figure 4.11). Nonetheless, the combinable components topology is well resolved (Figure 4.12) and not in conflict with other topologies. In this instance, the well resolved rbcL topology appears to reinforce one of the possible topologies that the site mapping data can support. In combination, the rbcL data is sufficiently informative to override any character conflict that might otherwise reduce the resolution of the resultant tree (Figure 4.5). The rpoC2 data may not be as powerful in this regard, resulting in the unusual topology from the analysis of the combined rpoC2 and site mapping data.

The success of the rbcL data over both the rpoC2 and site mapping data, in terms of the provision of a phylogenetic signal that improves resolution, highlights once again the fact that the rpoC2 data, and to a lesser extent the site mapping data, is not informative at the deeper hierarchical levels in the family. This reinforces the conclusion that rbcL is useful for molecular systematic studies at the subfamily level in the grasses (Chapter 3).

Each data set thus has its own strengths, which some combinations utilise or emphasise (almost in a "synergistic" manner) to greater effect than other combinations. The "strength" of each data set correlates to the ability of the data to resolve relationships at various hierarchical levels. In this regard, the rbcL data (the most conservative) is very strong, being able to resolve well supported nodes at all levels. In contrast, the variable rpoC2 data is unable to resolve the deeper nodes, or (if it does) these nodes are not well supported. The combination of these two sets of sequence data results in a well resolved topology with good bootstrap support at most of the nodes (no bootstrap values less than 50% in Figure 4.1). In this topology, the rbcL data provides strong support for the basal nodes, while the rpoC2 data gives added support to the relationships of the terminals.

Although they make no mention of the performance of their data in this respect, Davis and Soreng's (1993) site mapping data may be viewed as intermediate between the rbcL and rpoC2 data. Although this statement cannot be readily supported in any statistical manner, a comparison of the degree of resolution in the consensus trees from the rbcL study and site mapping study (Davis and Soreng 1993) bears this out. Both studies had comparable sample sizes (36 versus 31 genera respectively) and number of characters (160 versus 155 respectively). The rbcL topology seemingly resolves basal relationships in the grasses more clearly than the site mapping data set, but parts of the PACC clade are not resolved by the rbcL data, which the site mapping data may be able to accomplish. Unfortunately, the different sampling emphases of these two studies make this comparison somewhat spurious. Of greater informativeness is the fact that the ensemble consistency indices for the rbcL derived trees are somewhat higher, (42 versus 39), as are the ensemble retention indices (67 versus 61), indicating that the rbcL data contains a stronger phylogenetic signal.

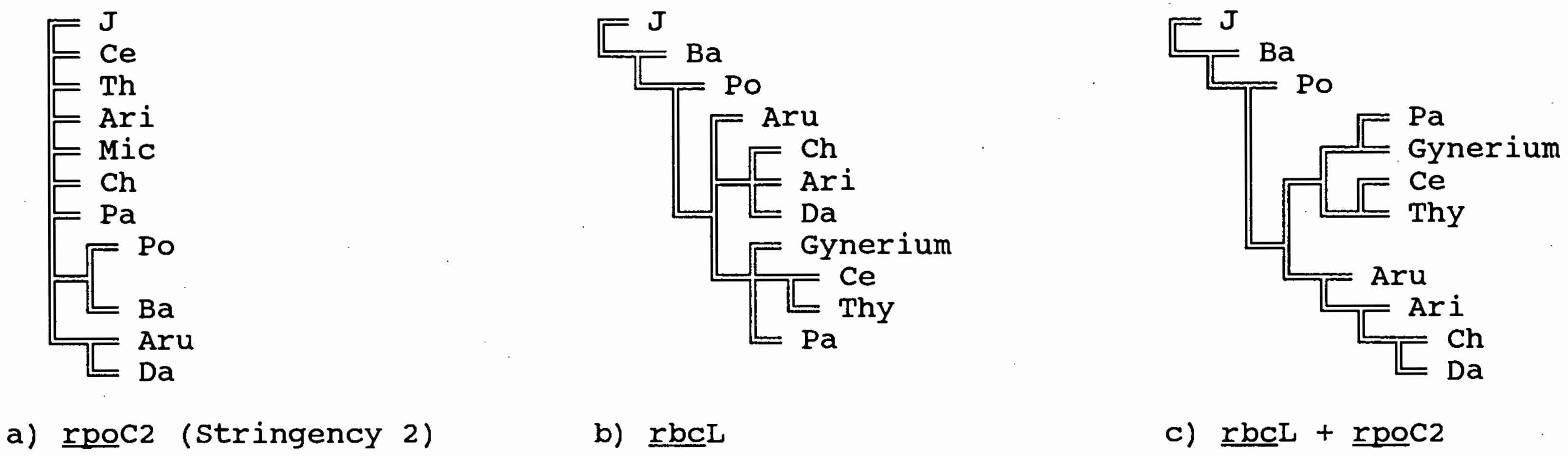
CONCLUSIONS

Owing to the interdependence of the plastid data sets, the analysis of combined data sets is recommended (De Queiroz 1993). This recommendation is supported here, as the analysis of the various combinations of molecular data sets produced more resolved topologies than the consensus approach, although the topologies from the two approaches were (in most cases) congruent. The degree of resolution of the combinable component topologies indicated that all data sets were equally incongruent, and that the retention of particular nodes was related to the bootstrap support value that those nodes enjoyed in the combined analysis topology.

The most resolved topologies (resulting from both combinable component consensus and the analysis of combined data) were obtained when at least one of the data sets comprised conservative characters. In this study, the <u>rbc</u>L data set was the best source of these characters. However, some data sets still have an imbalance in the sampling density of some lineages, and there is a lack of commonality in others. In spite of this, the analyses of the combined data clarifies the relationships of many of the lineages of the grasses.

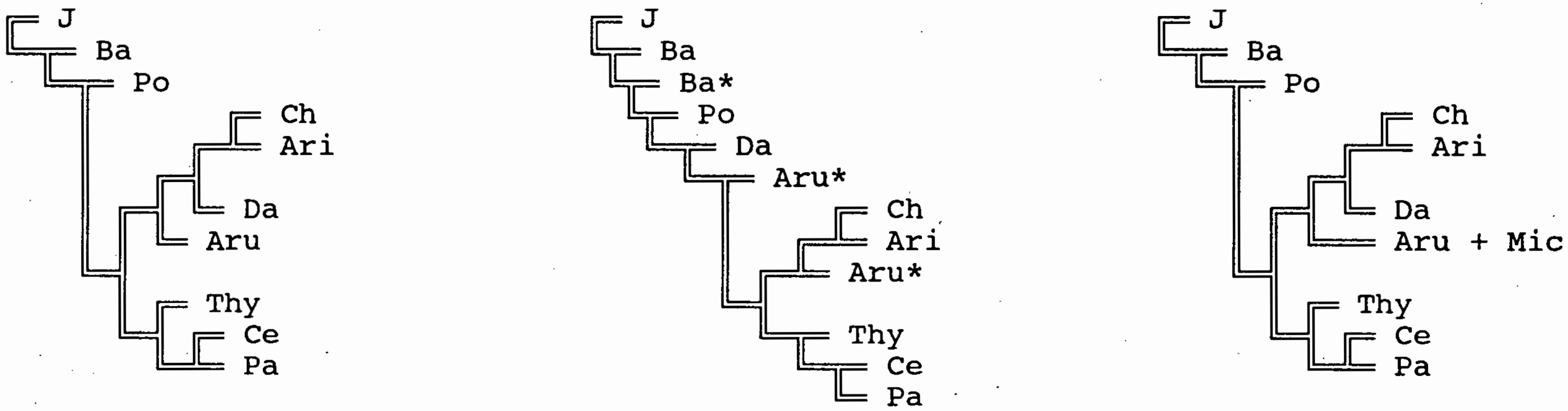
The analyses indicate that the Bambusoideae are the basal lineage of the grasses, and that the PACC clade is terminal and monophyletic. Figure 4.18(a-e) summarises and compares the various topologies from analyses of the combined data sets. Figure 4.18(f) shows the extrapolated relationships of the lineages of the Arundinoideae, including the Micraireae.

The relationships between the Centothecoideae, Panicoideae and Thysanolaeneae are comparatively stable and well supported. However, relationships of the other lineages within the PACC clade (the Danthonieae, Chloridoideae, Arundineae and Aristideae) are uncertain. Bootstrap support values for various relationships among these clades are always poor to moderate at best. Only through a combination of additional sampling and an increase in the number of characters from conserved genes can this relationship possibly be resolved. Further discussion of these relationships may be found in the next chapter.



a) rpoC2 (Stringency 2) b) rbcL c) rbcL + rpoC2

Figure 4.18. A summary of the relationships among the various lineages and genera of the grasses, as elucidated by the various molecular data sets and analyses of combinations thereof. **Key**: Ari = Aristideae, Aru = Arundineae, Ba = Bambusoideae, Ch = Chloridoideae, Ce = Centothecoideae, Da = Danthonieae, J = *Joinvillea* (outgroup), Mic = Micraireae, Po = Pooideae, Pa = Panicoideae, Thy = Thysanolaeneae. Abbreviations marked with an asterisk ("*") indicate para- or polyphyletic clades.



d) rbcL + site mapping;
rbcL + rpoC2 + site mapping

e) rpoC2 + site mapping

f) Extrapolated relationships

Figure 4.18 cont.

CHAPTER 5

TAXONOMIC IMPLICATIONS OF THE MOLECULAR PHYLOGENIES

The analysis of the data sets presented in this thesis, both separately and in combination with other molecular data, corroborates the study of Davis and Soreng (1993) which elucidated three main lineages in the grasses. These are the subfamilies Bambusoideae (possibly para- or polyphyletic), Pooideae, and a group they termed the PACC clade, comprising the remaining subfamilies (Panicoideae, Arundinoideae, Centothecoideae and Chloridoideae).

While the analyses of the various data sets are inconclusive about some of the relationships of the major lineages of the PACC clade, insight into the composition and affinities of the tribes of the Arundinoideae has nonetheless been gained. This chapter discusses the lineages of the Arundinoideae in detail, their relationships and their composition as elucidated by the molecular studies. The other subfamilies are, however, briefly discussed.

The discussion on the relationships among the subfamilies and tribes is based on the results of the rbcL and combined data analyses, as these provided a more resolved topology at this level than the rpoC2 phylogenies. However, the wider sampling range in the rpoC2 study allows for a more detailed discussion of the relationships within these lineages. Unless otherwise stated, the discussion of the results from the analyses of the rpoC2 data are based on most resolved topology (alignment at stringency 2, with characters coding for gaps; Chapter 2, Figure 2.5). This topology is a strict consensus tree of trees found using HENNIG86's bb* search option on trees obtained from 500 random order input searches carried out as described in Chapter 2.

The subfamily Bambusoideae

Unfortunately, there is insufficient overlap between the taxa sampled in the rbcL and rpoC2 studies to allow a comparison of the phylogenies based on these two genes. The lack of resolution among the three samples of the Bambusoideae in the rbcL phylogeny reflects either a paraphyletic Bambusoideae, or inadequate sampling. The rbcL data shows the Bambusoideae to be the basal lineage of the grasses, while the rpoC2 phylogeny shows the Bambusoideae to be sister to the Pooideae. This latter relationship may be alignment dependant, as it becomes equivocal at high alignment stringencies. Furthermore, the sample size of this lineage is insufficient to adequately assess the homologies of the repeat units. Thus an accurate alignment for this group was not feasible. However, the variability of the rpoC2 insert makes it an ideal gene for examining the phylogeny within the Bambusoideae.

The relationship of the Bambusoideae to the Pooideae and PACC clades obtained in the rbcL and combined analysis is consistent with those obtained by Hamby and Zimmer (1988) from rRNA sequence data; i.e. ((Bambusoideae (Pooideae, PACC)). However, Doebley et al. (1990) using rbcL, and Davis and Soreng (1993), using plastid site mapping data, provide evidence for a (Pooideae (Bambusoid, PACC)) relationship. However, no data set provides strong evidence favouring any particular grouping. The restriction site data (Davis and Soreng 1993) provides only weak evidence for the conflicting (Pooideae (Bambusoid, PACC)) grouping, while the parsimony analysis of the rbcL data shows weak bootstrap support (40%) for the (Pooideae, PACC) relationship (56% in the Neighbour Joining analysis). The Bremer support indices indicate that this relationship is lost in trees only one step longer. These weak and conflicting hypotheses of bambusoid relationships are reviewed in detail by Kellogg and Watson (1993) and Kellogg and Linder (in press).

The subfamily Pooideae

This includes the "core" pooids (the Pooideae) as well as Stipa (Stipeae), Nardus (Nardeae) and Lygeum (Lygeae). As the rbcL study did not include the latter two taxa, Stipa appears as the basal taxon in the pooid clade in topologies based on this gene. In these phylogenies, the "core" pooid clade is strongly supported. However, only three of the eight non-stipoid tribes in the subfamily have been sampled. In the rpoC2 study, Lygeum and Nardus are consistently placed at the base of the pooid assemblage. The rpoC2 insert in these two taxa has a typical pooid heptameric repeat structure, and thus supports a pooid placement for these tribes.

The Stipeae, Nardeae and Lygeae have previously been included in Arundinoideae by Watson (1990). In the rpoC2 sequencing study, the Nardeae and Lygeae are shown to be the sister taxa, and comprise the basal lineage of the Pooideae, which as a subfamilial clade receives good bootstrap support (88.6% in Figure 2.8). In this regard, the classification of Clayton and Renvoize (1986) is corroborated.

Although bootstrap support for a (Stipeae, Pooideae) grouping is weak or equivocal in the rbcL study (56% in Figure 3.4), this clade is retrieved by the combined analyses, as well as the restriction site data (Davis and Soreng 1993), rpoC2 data and morphological data (Kellogg and Watson 1993). There is thus ample evidence supporting the inclusion of the Stipeae in the Pooideae rather than the Arundinoideae. The rpoC2 data further indicates that Anisopogon, a genus previously considered to be arundinoid, is allied to Stipa, irrespective of alignments used. However, owing to the incomplete sampling of the pooid tribes, this may not be a true reflection of their relationship. Other evidence supporting the affinity of Anisopogon with the Stipeae includes data on the amino acid composition of the endosperm (Yeoh and Watson 1981) and the presence of three large lodicules (Clayton and Renvoize 1986). The delimitation and composition of the Stipeae and its position in the pooid clade needs to be further investigated.

The PACC Clade

The majority of the taxa in both the rbcL and rpoC2 studies fall into what Davis and Soreng (1993) call the PACC clade. With the exception of Anisopogon, all the members of the Arundinoideae sensu Clayton and Renvoize (1986) are placed within this assemblage.

Both the rbcL and, to a lesser extent, the rpoC2 data support the monophyly of the PACC clade. The relationships between the representatives of the four subfamilies in this clade are, however, incompletely resolved. Analysis of combined data sets suggests a ((Thysanolaeneae (Panicoideae, Centothecoideae) (Arundineae (Danthonieae (Aristideae, Chloridoideae))) relationship (Figure 4.18d, f). However, bootstrap support for some of these relationships is moderate at best.

In examining the PACC clade, the fragmentation of the subfamily Arundinoideae becomes obvious. Firstly, and most obviously, two large clades comprising arundinoid genera may be recognised. These correspond approximately to the tribes Danthonieae and Arundineae as delimited by Watson (1990). Secondly, several "arundinoid" tribes and genera are shown to be affiliated with other subfamilies. Relationships within and between these groups and genera are discussed below.

The subfamily Panicoideae

The structure of the rpoC2 sequence data indicates that all genera sampled in this clade possess clade-specific heptameric repeats, emphasising once again the utility of the gene at the level of the subfamily and below (see figure 2.2). Within the subfamily, the genera sampled are divided into two lineages; one corresponding to the tribe Paniceae, the other to the (Andropogoneae, Arundinelleae). The inclusion of the Arundinelleae as a terminal clade within the Andropogoneae (as suggested by the rpoC2 data) makes the latter a paraphyletic tribe, and is contrary to the classification presented by Clayton and Renvoize (1986, their figure 5) and Kellogg and Watson (1993), who consider the Arundinelleae to be the basal lineage in the Andropogoneae.

The subfamily Chloridoideae

Kellogg and Campbell (1987) feel that the character generally used to define this subfamily (microhairs with inflated distal cells) is weak, and their analysis indicates that the subfamily may not be monophyletic. However, much of the data for this subfamily in their analysis was unknown, and had to be coded as missing. In all the analyses carried out in this study, the few chloridoid representatives were always resolved as a monophyletic clade.

As discussed in chapter 3, the relationships between this subfamily and the Aristideae and Danthonieae are not resolved with any certainty by the rbcL sequence data. The analysis of the combined data (rbcL, rpoC2 and site mapping data) suggest the (Aristideae, Chloridoideae) relationship with reasonable bootstrap support (67%; Figure 4.13). However, the alternative (Chloridoideae, Danthonieae) is in agreement with several recent classifications based on both morphology (Clayton and Renvoize 1986, Watson et al. 1985, Watson 1990) and molecular data (Davis and Soreng 1993, Cummings et al. 1994).

Of particular interest is the consistent placement (in both *rbc*L and *rpo*C2 derived topologies) of the genus *Centropodia*, which appears basal to the rest of the Chloridoid taxa, and *Merxmuellera rangei*, which is associated with *Centropodia* in the *rpo*C2 analyses. As only two of the five chloridoid tribes recognised by Clayton and Renvoize (1986) have been sampled in the *rbc*L study, and three tribes in the *rpo*C2 study, additional sampling may clarify both tribal level relationships and the position of *Centropodia* and *Merxmuellera rangei* within this clade.

Centropodia has previously been placed in the danthonioid lineage. The inclusion of *Centropodia* in the Chloridoideae receives good bootstrap support (90%; Figure 4.1) in the analysis of the combined *rpo*C2 and *rbc*L data sets. This placing is supported by the fact that *Centropodia* possesses C_4 anatomy and metabolic pathway (Ellis 1984, Tomlinson 1985), a feature typical of the Chloridoideae. However, it lacks the characteristic egg-shaped microhairs of this subfamily (Renvoize 1981). However, as mentioned above, this character is considered by Kellogg and Campbell (1987) to be weak. Nonetheless, the exclusion of *Centropodia* from the danthonioids is supported by the observation that it lacks haustorial synergids that characterise the Danthonieae (Verboom *et al*. 1994).

According to the *rpo*C2 analyses, *Merxmuellera rangei* (another "danthonioid" taxon) is consistently placed at the base of the chloridoid lineage and may be associated with *Centropodia*. The Neighbour Joining topologies show these two as sister taxa. However, bootstrap support for this relationship from both parsimony and distance based methods is not very strong (53.8% in Figure 2.8, 65% in Figure 2.9). Unlike *Centropodia*, *M. rangei* is a C_3 species. Ellis (1982) commented on the anatomical distinctness of *M. rangei*, the leaf blade of which is almost cylindrical or terete in transverse section with a small adaxial groove. Furthermore, the presence of three orders of vascular bundles was noted as being atypical of the other Merxmuellera species, and Ellis suggested that M. rangei warranted elevation to generic status. The lack of haustorial synergids (G. Verboom, pers. comm.) is further evidence favouring the exclusion of this species from the danthonioid lineage (in which all con-generics examined to date are placed). In accordance with Ellis's (1982) suggestion, this unusual species is to be placed in a genus of its own (Barker et al., in prep.)

The subfamily Centothecoideae

The Centothecoideae, elevated to subfamily status by Soderstrom (1981) has been variously placed as a tribe in the Bambusoideae (Watson 1990) or Arundinoideae (Renvoize 1981). Although considered by Clayton and Renvoize (1986) to be related to the Arundinoideae, Kellogg and Campbell (1987) suggested that this small subfamily was related to the Panicoideae; a suggestion supported by all molecular data sets.

In the rbcL study, the Centothecoideae are represented by only one taxon (Chasmanthium). In the rpoC2 study, a second representative is included; Lophatherum. However, this taxon never appears as sister to Chasmanthium in any of the consensus trees obtained, but occupies a variety of positions. Generally it is placed in the basal-most position in the cladograms. Forcing the monophyly of the Centothecoideae by means of the Dos Equis option in HENNIG86 in these topologies does not change the tree length (except the alignment at stringency 5 where the tree length increases by one step). The rpoC2 sequence for Lophatherum is 42 base pairs (2 heptameric repeats) shorter than that of Chasmanthium, and thus possibly doesn't contain informative sites that unambiguously place it in a more resolved position.

The subfamily Arundinoideae, tribe Thysanolaeneae

The suggestion by Hilu and Wright (1982) that Thysanolaena might be allied to the Centothecoideae, based on morphological data, is corroborated by both the rbcL and rpoC2 analyses. The Neighbour Joining tree based on rbcL data differed from this topology, including Gynerium as sister to Thysanolaena, with the centothecoid representative basal to these two genera (Figure 3.5).

The rpoC2 analysis indicated that the little known Australian genus Spartochloa is sister to Thysanolaena. This relationship is retrieved in all topologies, irrespective of alignment or cladistic method employed. Furthermore, the clade receives strong bootstrap support (93.9% in Figure 2.8, 93% in Figure 2.9).

The status of Thysanolaena in past tribal classifications has been equivocal (Table 1.1). Watson (1990) places it in the Arundineae along with Phragmites and Arundo, but Clayton and Renvoize (1986) place it in its own tribe, with the caveat that tribal status may overstate the differences between this genus and Phragmites. Renvoize (1981), on the basis of a phenetic analysis of anatomical characters, considered the genus to be peripheral to the rest of the subfamily, and suggested a relationship with Gynerium and Phragmites.

Spartochloa was considered by Watson (1990) to be sufficiently distinct to warrant its own tribe within the Arundinoideae, the only classification that recognises this genus as being excluded from the Danthonieae and Arundineae. Thysanolaena and Spartochloa have previously never been placed outside the Arundinoideae, although Hilu and Wright (1982) have suggested that the former genus may be related to the Centothecoideae.

The relationship between the (Thysanolaena, Spartochloa) clade and the Centothecoideae as indicated by both the rbcL and rpoC2 analyses needs further investigation. Details of the embryo anatomy, which provide one of the unique features of the Centothecoideae and which is used by Clayton and Renvoize (1986) to justify its subfamilial status, has not been documented for Thysanolaena or Spartochloa. Leaf anatomical features may also be useful, and these have been documented for both the Centothecoideae (Renvoize 1986b) and Arundinoideae, including Thysanolaena (Renvoize 1986a). However, these descriptions are somewhat brief and no mention is made of any bicellular, obliquely joined microhairs in Thysanolaena (a general centothecoid character). Tateoka et al. (1959) describe the microhairs of Thysanolaena as being "distinctive", but do not expand on this statement.

Gynerium

The placing of the monotypic genus Gynerium in both the rbcL and rpoC2 analyses is equivocal. In the rbcL analysis, it appears that the taxon is associated with the base of either the panicoid or centothecoid lineages. In contrast, the Neighbour Joining analysis indicates a sister group relationship between Gynerium and Thysanolaena, a topology which receives strong bootstrap support (70%).

In the rpoC2 analyses, the uncertainty of the position of Gynerium is indicated by its variable positions in the trees from different alignment stringencies and cladistic methods. When most resolved, the topologies show it to be basal to the reedy genera or in a polychotomy with the Panicoideae, Centothecoideae, Thysanolaena and Spartochloa. The latter relationship is supported by the rbcL-based topologies, and analysis of the combined rbcL and rpoC2 sequence data (Figure 4.1) indicates that it is basal to the Panicoideae, although this relationships is weakly supported (52% bootstrap support).

In contrast, Renvoize (1981, 1986a), Clayton and Renvoize (1986) and Kalliola and Renvoize (1994) consider Gynerium to be closely related to the other reedy arundinoid genera, in particular Arundo. In those classifications that make the distinction between the Arundineae and Danthonieae (eg. Watson 1990), Gynerium has always been placed in the Arundineae. Despite the fact that the molecular data cannot resolve the relationships of Gynerium beyond doubt, the data are clearly not consistent with a close relationship with the other reedy genera.

Unfortunately, nothing is known about the embryology of Gynerium, and leaf anatomical studies (Renvoize 1981, 1986a; Kalliola and Renvoize 1994) are limited by a lack of comparable data. Thus the novel placement of Gynerium elucidated by the plastid molecular phylogenies cannot be easily assessed using these characters.

Spartochloa, Gynerium and Thysanolaena, previously genera in the Arundinoideae, are shown here to have affinities with the Centothecoideae and Panicoideae. The Thysanolaeneae is thus not a tribe of the Arundinoideae. It also is no longer monotypic, as it is shown to comprise two genera. Whether it should be retained in this context as a tribe of the Centothecoideae can only be assessed in conjunction with a thorough examination of the latter subfamily.

The subfamily Arundinoideae, tribes Arundineae sensu stricto and Micraireae.

The past recognition of the Arundineae and Danthonieae as separate tribes by some taxonomists (eg. Watson 1990) is supported by both the rpoC2 and rbcL phylogenies. Substantial dissimilarity between these two lineages has also been shown using immunological cross-reactions of prolamins (Hilu and Esen 1990), which indicated that Phragmites was distantly related to other danthonioid and aristidoid representatives. Despite this evidence showing the distinctness of Phragmites, Hilu and Esen (1990) support the retention of a single tribe for these taxa (Arundineae sensu Clayton and Renvoize 1986).

This clade, which is weakly supported by the rbcL data (Figure 3.3) and only moderately supported in the Neighbour Joining tree (Figure 3.4), includes the reedy genera Phragmites and Moliniopsis (which appear as sister taxa) and Arundo, Dregeochloa, Amphipogon and Monachather. None of the latter three genera are at all reed-like, and their unusual relationship to the reedy genera suggested here has not been previously proposed.

In contrast to the rbcL phylogenies, the rpoC2 topologies do not consistently place Arundo into any one clade. Instead, it appears basal to one or more of the panicoid, centothecoid and aristidoid groups. However, the rpoC2 trees clearly support a relationship between Moliniopsis, Phragmites and Hakonechloa. Dregeochloa, traditionally considered to be a danthonioid, is shown in some topologies to also be included in this clade. However, the position of this genus is alignment dependant and sometimes it occupies a less resolved basal position within the general phragmitoid - centothecoid - panicoid lineage.

The rpoC2 topologies consistently show Monachather as sister to Micraira (tribe Micraireae sensu Clayton and Renvoize 1986). However, the bootstrap support for this relationship varies, and may be quite weak (40.2% in Figure 2.8) but higher in the Neighbour Joining analysis (72% in Figure 2.9). Both the parsimony and Neighbour Joining analyses do not resolve the relationships of the (Monachather, Micraira) clade with any consistency. Amphipogon is shown basal to these two taxa in trees from some of the alignments, and basal to the danthonioid lineage in others. Assuming the relationship between Micraira (tribe Micraireae) and Monachather, as suggested by the rpoC2 data is correct, then the inclusion of Monachather (and by default Micraira) within the Arundineae as suggested by the rbcL data renders the latter tribe paraphyletic. It thus becomes imperative that a rbcL sequence for Micraira be obtained so that the suggested relationship between these two tribes can be confirmed.

In the rbcL-derived phylogenies, Dregeochloa is placed within the Arundineae where it appears to be more related to Arundo, Amphipogon and Monachather than to either Phragmites or Moliniopsis, as shown in the rpoC2 topologies. It is nonetheless never associated with the Danthonieae in any phylogeny obtained, and lacks haustorial synergids characteristic of that lineage (Verboom et al. 1994).

Dregeochloa and Micraira exist in very arid areas of southern Africa and Australia respectively. Dregeochloa has an unusual leaf anatomy (Ellis 1977, Tomlinson 1985), and was given generic status partly on the basis of unusual fruit morphology (Conert 1966). It also has an unusual, cushion-like growth form.

Micraira, characterised in part by the presence of spiral phyllotaxy, was placed it in its own subfamily (Lazarides 1979). Species of this genus behave like resurrection plants, indicating a physiology and metabolism adapted to extreme seasonal aridity. Renvoize (1986a) considers the leaf anatomy of Micraira to be typical of the Arundinoideae, postulating it to be a derivative of the Arundineae (in a broad sense), while relationships with Aristida have also been proposed (Clifford 1964, Pilger 1954).

The phenetic analysis carried out by Hilu and Wright (1982) placed Amphipogon in the Arundinoideae basal to Arundo, Phragmites and Cortaderia. However, their small sample size in this clade (six taxa) precludes any detailed discussion of the phylogeny of the subfamily. In contrast, support for a pooid placing for Amphipogon was found by Kellogg and Campbell (1987) in their cladistic analysis of mainly morphological characters. Renvoize (1981, 1986a) considered Amphipogon to have an anomalous leaf blade anatomy, in that it possessed papillate long cells and lacked microhairs, the latter character suggesting a position in the Pooideae. Watson and Dallwitz (1992) comment that the spikelet form (especially the lemma) and microhairs (chloridoid or Enneapogon type) are reminiscent of Enneapogon, thus suggesting chloridoid affinities for this genus. The fruit has a free pericarp, a feature it shares with Dregeochloa (Watson and Dallwitz 1992).

Tomlinson (1985) suggested that the leaf anatomy of Monachather was unusual among the Danthonieae she studied. Known from arid areas of Australia, Vickery (1956) considered Danthonia bipartita (=Monachather paradoxus) to be distantly related to the other Australian Danthonia species. Connor and Dawson (1993) note that Monachather is extremely specialised, and usually cleistogamous. Both Monachather and Dregeochloa posses bulliform cells that penetrate deeply into the mesophyll (Tomlinson 1985), a character that provides some support of the relationships suggested by the molecular phylogeny. Monachather has also been recorded as having a germination flap in the lemma (Johnson and Watson 1981) and small chromosomes (Abele 1959). Unfortunately, comparable data on these features from the other genera in this clade is lacking. Watson and Dallwitz (1992) consider certain features in this grass to be eupanicoid (such as microhairs and costal silica bodies).

Despite the rpoC2 data suggesting a somewhat fragmented set of relationships, Arundo, Phragmites, Moliniopsis, Hakonechloa, Dregeochloa, Amphipogon, Monachather and (by extrapolation) Micraira might well constitute a monophyletic lineage. Unlike the danthonioid clade, where a cryptic but nonetheless sound morphological character corroborates the lineage, no such character has yet been found for this putative clade.

The subfamily Arundinoideae, tribe Aristideae

As discussed in chapters 2 & 3, the relationship of this tribe to the Danthonieae and Chloridoideae remains unresolved. Recent classifications (listed in Table 1.1) include the Aristideae in the Arundinoideae, but previously De Winter

(1962), Clayton (1978) and Hilu and Wright (1982) have placed them in the Chloridoideae, although the latter authors do not consider the tribe to be typical of the subfamily. Hilu and Esen (1990) used immunological methods to place Aristida at the base of the arundinoid clade; a clade that included Phragmites and several other danthonioid genera. However, this study was limited by a lack of samples from the other PACC clade subfamilies, so relationships beyond the danthonioid, aristidoid and arundinoid taxa were not tested.

The variety of leaf anatomy and corresponding metabolic pathways within the genera in the Aristideae is unusual, ranging from C_3 to unique forms of the C_4 pathway (De Winter 1962; Clayton and Renvoize 1986). The embryo morphology of the Aristideae (except Sartidia) is identical to that of the arundinoid and danthonioid genera (P-PF; Reeder 1957). This pattern differs from that of the chloridoids, which posses an epiblast (P+PF), while Sartidia has a P-FF embryo type. However, embryo morphology is unhelpful in solving the discrepancies in the placing of the Aristideae, as the P-PF embryo type appears to be the basal condition for all the basal lineages in the PACC clade (Klak and Linder, in prep.). The acquisition of additional morphological and molecular (i.e. rbcL sequence) data from Sartidia may assist in the resolving of the relationships of the Aristideae.

The subfamily Arundinoideae, tribe Danthonieae

The results of the rbcL analysis indicate that there is strong bootstrap support for the danthonioid clade. In the rpoC2 study, 32 taxa were consistently placed in this clade, which receives 50% bootstrap support in the Neighbour Joining analysis (Figure 2.9) and 49% in the RNA bootstrap topology (Figure 2.8), although this clade includes Amphipogon. The relationships among these taxa as elucidated by the parsimony analyses were little changed by the various alignment stringencies, but resolution decreased at the higher stringencies. There were, however, differences in topology between the Neighbour Joining and parsimony-based trees. This clade, discussed below, is shown in Figure 5.1. The bootstrap values and differences in topology between the trees derived from various cladistic methods are discussed where relevant, and are based on Figures 2.8 and 2.9.

In addition to the molecular support for a monophyletic danthonioid group (excluding *Centropodia*, *Amphipogon*, *Micraira*, *Monachather*, *Dregeochloa*, *Merxmuellera rangei*, *Spartochloa* and *Anisopogon*), all taxa in the clade examined to date possess haustorial synergids in the megagametophyte (Philipson 1977, Philipson and Connor 1984, Verboom *et al*. 1994). The danthonioid clade is thus defined in a narrower context than previously used, but approximates the tribe Danthonieae *sensu* Watson (1990).

Within this clade, the basal taxa are African, comprising species of *Merxmuellera*. However, the *rpo*C2 molecular phylogeny suggests that *Merxmuellera* is a polyphyletic assemblage, as there are species scattered throughout the danthonioid lineage from the basal to terminal clades. The basal-most clade comprises *M. macowanii* and *M. davyi* (the type species of *Merxmuellera*). This two-taxon clade receives good bootstrap support (98%) in the Neighbour Joining analysis. Other species (*M. cincta*, *M. arundinacea*, *M. setacea* and *M. rufa*) form a second, poorly resolved basal group, while yet other species are associated with the terminal (*Rytidosperma sens. lat.* (*Karroochloa*, *Schismus*),(*Tribolium*, *Urochlaena*)) clade. Anatomical and morphological diversity within *Merxmuellera* lends support to hypotheses of suspected polyphyly (Conert 1970, 1971, Ellis 1982, 1983, Barker in Gibbs Russell *et al*. 1990).

Above the basal assemblage of *Merxmuellera* species is a dichotomy leading to a clade comprising *Pentameris*, *Pentaschistis* and *Prionanthium*. This clade is distinct on both molecular (unusual repeat motifs) and morphological grounds (2-flowered spikelets), and it receives strong bootstrap support (>90%) in both bootstrap analyses. However, the Neighbour Joining topology places this clade between the basal two-taxon Merxmuellera clade and the other basal Merxmuellera species. This latter clade of four species is well supported in the Neighbour Joining analysis (74% bootstrap support in Figure 2.9, but only 39.1% in Figure 2.8).

Chionochloa is the most basal non-African genus in the danthonioid clade, and is placed above these basal African clades, a position which is well supported (75% in Figure 2.9). All except one species in this genus are known from New Zealand (Connor 1991).

There are two large terminal clades. One corresponding approximately to the genus Rytidosperma sensu Clayton and Renvoize (1986), the other clade comprising Danthonia and several other Australian genera

The Rytidosperma clade contains the Australasian and South American species of Danthonia that have been placed in various genera such as Erythanthera (Zotov 1963), as well as Danthonia pallida, a species placed in Chionochloa by Connor and Edgar (1979). These taxa are related to an african group comprising the clades (Karroochloa, Schismus) and (Tribolium, Urochlaena), as well as three species of Merxmuellera. Of these latter genera, Clayton and Renvoize (1986) considered only Karroochloa and Merxmuellera to be part of their expanded concept of Rytidosperma. Note, however, the Neighbour Joining topology (Figure 2.9) differs in the placing of some of these species of Merxmuellera. Nonetheless, the bootstrap support for this clade in this tree is 50%. The topology of the tree from the RNA bootstrap analysis (Figure 2.8) does not retrieve the same clade, and taxa from the Rytidosperma and Danthonia clades are mixed.

The phylogeny obtained here (Figures 2.7 and 2.9) implies that Clayton and Renvoize's (1986) concept of Rytidosperma should be restricted to the Australasian and South American danthonioid genera, a view also held by Connor and Edgar, (1979). The African Karroochloa and Merxmuellera should thus be retained, although the latter genus clearly needs further taxonomic attention. The alternative approach would be to consider the entire clade (i.e. including Karroochloa, Schismus, Tribolium and Urochlaena) and the relevant species of Merxmuellera) to be one genus. However, with the exception of Merxmuellera and Urochlaena, which should be included in Tribolium (Linder in prep.), the genera as presently defined are adequately circumscribed, and any taxonomic changes should only be made once a species level phylogenetic analysis has been carried out.

The northern hemisphere Danthonia spicata (and its South American congener D. secundiflora) are sister to the Australasian Notochloe, Plinthanthesis and the New Zealand Cortaderia fulvida. This clade has a high bootstrap support (90%) in the Neighbour Joining tree (Figure 2.9), but a much lower support in the RNA bootstrap topology (49.3%)

The placing of the Australian Danthonia pallida in the Rytidosperma clade rather than with the mainly northern hemisphere Danthonia species supports the contention that the southern hemisphere Danthonia species are not related to the northern hemisphere Danthonia species (Blake 1972, Connor and Edgar 1979, Zotov 1963). Notochloe and Plinthanthesis (both previously Australasian Danthonia's) share an unusual leaf anatomy. There is thus a genuine link between northern and southern hemisphere "Danthonia" segregates through Notochloe and Plinthanthesis, but the relationship of these two genera to other Australasian "Danthonia" species is distant. The inclusion of one of the Cortaderia species in this clade is perplexing, and further discussion on this genus appears below.

Two southern African taxa, Pseudopentameris and Chaetobromus, are found basal to the (Rytidosperma, Danthonia) assemblage. These two genera share several caryopsis features that support this relationship (Barker 1994), which receives strong bootstrap support (100% in Figure 2.8, 99% in Figure 2.9).

Basal to this assemblage is a clade containing two South American taxa; a second species of Cortaderia (C. selloana) and Lamprothyrsus. Note, however, that the Neighbour joining analysis (Figure 2.9) places these two taxa within and basal to the Danthonia discussed above, although this receives weak bootstrap support (30%). The polyphyly of Cortaderia along continental lines is unexpected, and this genus warrants further investigation. Forcing the monophyly of Cortaderia by means of the Dos Equis option in HENNIG86 increases the tree length by six steps, while forcing the sister group relationship of Notochloe to Plinthanthesis requires only one extra step. A possible explanation for this result is that Cortaderia fulvida is of hybrid origin, the maternal (i.e. plastid) ancestor coming from the Danthonia - Notochloe - Plinthanthesis clade. Alternatively, the assumption that Cortaderia is monophyletic is incorrect (as is found for here for Merxmuellera). Sequencing of additional species of Cortaderia will demonstrate if this is the case, as would a phylogeny based on nuclear characters.

Support for the plastid phylogeny of the Danthonieae

The rpoC2 phylogeny is nothing more than a gene tree. The relationships of the Danthonieae discussed above may not agree with those suggested by morphological or anatomical data, derived ultimately from the nuclear genome. An independently derived phylogeny based on nuclear encoded characters is thus desirable for comparative purposes. Unfortunately, no such phylogeny is available, as the characters on which it is to be based are still being assessed and data collected.

However, an alternative approach may be used to test the accuracy of the plastid phylogeny. This method assesses the consistency of (nuclear coded) characters when they are plotted on the plastid phylogeny. Should these characters be consistent with the topology, then the plastid phylogeny (hypothesis) is not rejected. Furthermore, it becomes a powerful predictor of character distributions of the characters in taxa from which they are unknown.

This process is presently limited by the availability of good characters (the homology of which can be or has been tested), and only a few are available. These include characters taken from the megagametophyte (Verboom et al. 1994) and aspects of fruit morphology (Barker 1994). These characters are plotted on the topology shown in Figure 5.1. Unfortunately, these data are not known for all the taxa shown in the figure, so the characters are distributed according to those taxa for which information is available. As can be seen in Figure 5.1, the data obtained from the haustorial synergids provides support for relatively basal nodes in the phylogeny, whereas the caryopsis data tends to support a number of two-taxon clades. With the exception of two reversals, these characters are consistent with the plastid phylogeny, indicating that (on the basis of these few characters) the plastid phylogeny is likely to be an accurate assessment of organismal phylogeny. Because of the confidence this congruence provides, the molecular phylogeny is tentatively taken to indicate organismal relationships, on which the classification proposed below is based.

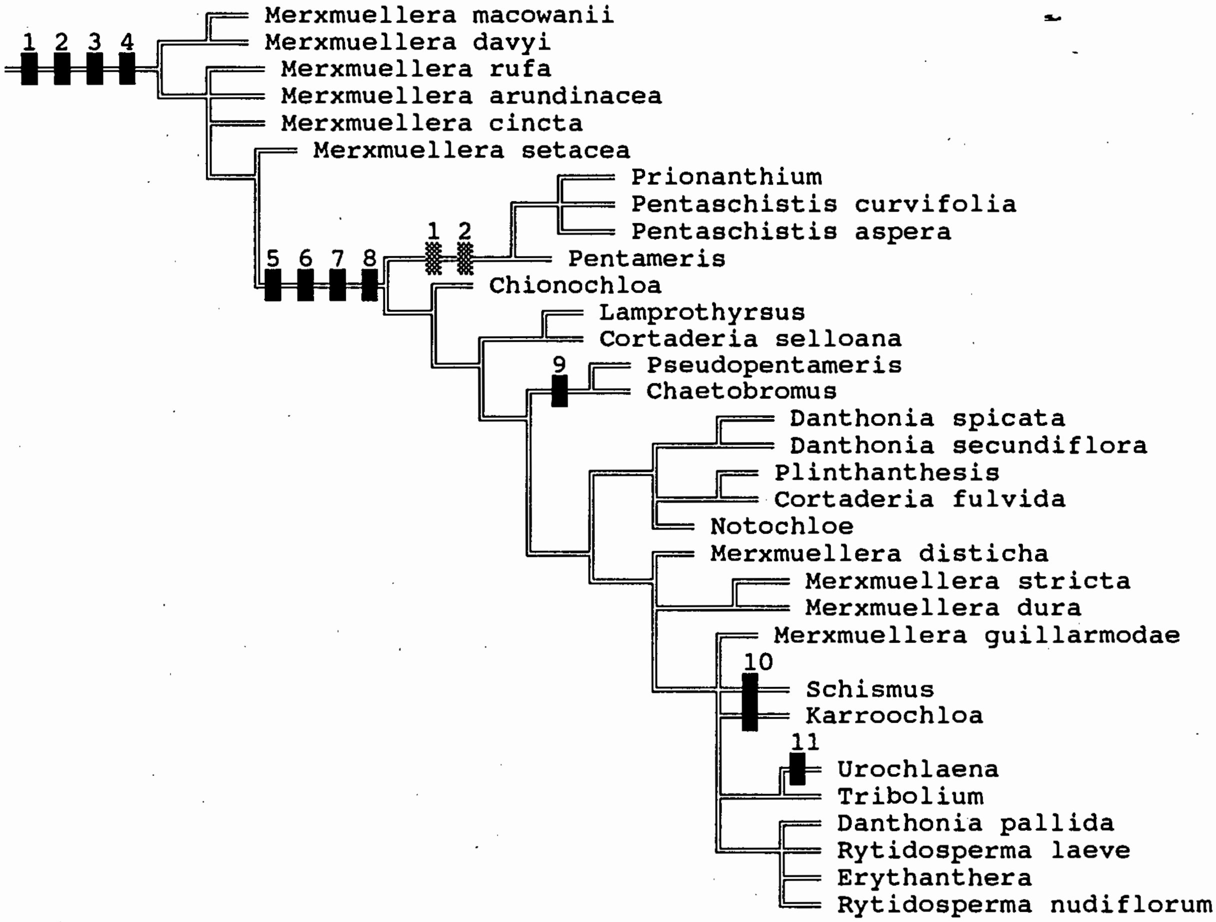


Figure 5.1. The topology of the danthonioid clade based on the analysis of the rpoC2 nucleotide sequence data aligned at stringency 2. Morphological characters that are consistent with this topology are plotted on the tree.

Character 1: Presence of haustorial synergids
Character 2: Starch in synergids
Character 3: Expanded micropyle
Character 4: Nucellar epidermis incomplete at micropyle
Character 5: Inner integument reduced to 1-2 cells thick
Character 6: Outer integument reduced to a collar
Character 7: Micropylar region of outer integument thickened
Character 8: Micropyle not conspicuously oblique
Character 9: Caryopsis with long, canaliculate hilum
Character 10: Caryopsis obovate, with large embryo.
Character 11: Caryopsis ovate, embryo large, pericarp flaking

Hierarchical status of the lineages of the Arundinoideae

On the basis of the various analyses of the molecular data, both separately and in combination, the subfamily Arundinoideae is shown to be polyphyletic. None of the classifications compared in the introduction (Clayton and Renvoize 1986, Conert 1987 and Watson 1990) is shown to be totally accurate. In terms of the identification of monophyletic lineages, Watson's (1990) classification probably performs the best, but is erroneous in the inclusion of pooid elements (Stipeae, Lygeae and Nardeae). The agglomeration of genera placed in the four tribes of the Arundinoideae by Clayton and Renvoize (1986) excludes these pooid elements. On the basis of the molecular phylogenies, these genera may still be divided into four lineages, but the composition of these lineages is somewhat different. These four lineages correspond to the Thysanolaeneae, Aristideae, Danthonieae _sensu_ Watson (1990) and the clade comprising the Micraireae, Arundineae _sensu_ Watson (1990) and other "danthonioid" genera.

As indicated by the _rpo_C2 analyses, the composition of the Thysanolaeneae is modified with the inclusion of _Spartochloa_. Hilu and Wright (1982) strongly support the retention of the Thysanolaeneae as a tribe of the Centothecoideae. The retention of tribal status for this clade is dependant on a detailed analysis inclusive of the Centothecoideae and basal panicoid lineages and _Gynerium_, but provisionally adopted in the classification proposed below.

The retention of the present tribal status of the Aristideae and Danthonieae is dependant on the resolution of the relationships of these two lineages with the Chloridoideae. Owing to the substantial morphological and anatomical differences between these three lineages, the recognition of these as separate subfamilies is suggested.

The Danthonioideae and Aristidoideae are thus proposed as new subfamilies. Although this solution is preferred, the alternative (depending to some extent on the resolution of the relationships of the Aristideae) is that either one or both of the Aristideae and Danthonieae are retained as tribes within an enlarged and heterogenous subfamily Chloridoideae.

The Chloridoideae would be retained in an emended form, as the inclusion of *Centropodia* and *Merxmuellera rangei* at the base of the Chloridoideae suggests that an additional tribe needs to be erected in this subfamily; the Centropodieae.

The Arundinoideae as a subfamily, and the Arundineae *sensu* Clayton and Renvoize (1986), is severely depleted. Watson's (1990) concept of the tribe is more accurate, but the loss of *Gynerium*, and the inclusion of other "danthonioid" taxa means that this subfamily Arundinoideae needs to be re-delimited in a much narrower context. The inclusion of the Micraireae in this clade renders the tribe Arundineae paraphyletic. Until a complete understanding of the composition and relationships within this clade is obtained, the continued recognition of the tribe Micraireae should be abandoned.

The observation that *Phragmites* and *Moliniopsis* (and, on the basis of the *rpo*C2 analyses, *Hakonechloa*) form a separate clade from *Arundo* and the other taxa in the lineage suggests that this is a natural unit, and should recognised as a tribe within the re-delimited Arundinoideae; the Phragmiteae. The lineage comprising *Arundo*, *Monachather*, *Dregeochloa*, *Micraira* and *Amphipogon* is retained as the tribe Arundineae. The recognition of two tribes within what is now a small subfamily might be viewed as unnecessary, but would be supported by the contention that *Phragmites* is only distantly related to *Arundo* (Clayton and Renvoize 1986).

Affinities of arundinoid taxa not included in this study

Of the genera placed by Clayton and Renvoize (1986) into the subfamily Arundinoideae, 12 are not included in this study. This number is higher if the subfamilial delimitation adopted by Watson (1990) is followed (20 genera). These genera are mostly small, and from localised regions of Africa or the Indian subcontinent. Many of these genera are poorly collected and important features such as their embryology and morphology of the megagametophyte are unknown.

Owing to the retention of plesiomorphic morphological characters in more than one of the above lineages, such as the presence of a geniculate central awn of the lemma, present in *Dregeochloa* (Arundineae; Arundinoideae), *Merxmuellera rangei* (Centropodieae; Chloridoideae) and many of the genera of the Danthonioideae, assessing the placement of the excluded genera on the basis of such characters is not possible. Only the confirmation of the existence of haustorial synergids in the megagametophyte can, at present, confirm a position in the Danthonioideae. Consistent morphological characteristics of the other previously arundinoid lineages are, at present lacking, and the only reliable data that can place them with any degree of confidence is molecular data. In the absence of suitably collected fresh material, efforts to obtain usable DNA from herbarium specimens of these genera should thus be made. Only by these means can this survey of the old, polyphyletic Arundinoideae be completed.

A new classification

Should the molecular phylogenies (i.e. the gene trees) obtained in this thesis reflect the organismal phylogeny, then the classification of the Poaceae requires amendment. The Poaceae would comprise eight subfamilies. Of these, the Bambusoideae and Panicoideae would be unchanged from the circumscription provided by Clayton and Renvoize (1986). The Pooideae would include the Stipeae, to which the previously danthonioid genus *Anisopogon* is tentatively added.

New subfamilies and tribes, and changes to the other existing subfamilies that would be required are as follows (changes given in **bold**):

Subfamily Centothecoideae (emend)
- Tribe Centotheceae
- **Tribe Thysanolaeneae (trib. nov.)**

Subfamily Chloridoideae (emend)
- Tribe Pappophoreae
- Tribe Orcuttieae
- Tribe Eragrostideae
- Tribe Leptureae
- Tribe Cynodonteae
- **Tribe Centropodieae (trib. nov.)**

Subfamily Aristidoideae (subfam. nov.)
- Tribe Aristideae

Subfamily Danthonioideae (subfam. nov.)
- **Tribe Danthonieae (emend)**

Subfamily Arundinoideae (emend.)
- **Tribe Arundineae (emend.)**
- **Tribe Phragmiteae (trib. nov)**

REFERENCES

ABELE, K. 1959. Cytological studies in the genus Danthonia. Transactions and Proceedings of the Royal Society of South Australia 82: 163-173.

AVDULOV, N.P. 1931. Karyosystematic studies in the grass family. The Bulletin of Applied Botany, Genetics and Plant Breeding suppl. 44: 1-428.

BACHMANN, B., W. LUKE and G. HUNSMANN. 1990. Improvement of PCR amplified DNA sequencing with the aid of detergents. Nucleic Acids Research 18: 1309.

BALDWIN, B.G. 1992. Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: An example from the Compositae. Molecular Phylogenetics and Evolution 1: 3-16

BALDWIN, B.G. 1993. Molecular phylogenetics of Calycadenia (Compositae) based on ITS sequences of nuclear ribosomal DNA: Chromosomal and morphological evolution re-examined. American Journal of Botany 80: 222-238.

BARRETT, M., M.J. DONOGHUE and E. SOBER. 1991. Against consensus. Systematic Zoology 40: 486-493.

BARKER, N.P. 1994. External fruit morphology of southern African Arundineae (Arundinoideae: Poaceae). Bothalia 14: 55-66.

BLAKE, S.T. 1972. Plinthanthesis and Danthonia and a review of the Australian species of Leptochloa (Gramineae). Contributions from the Queensland Herbarium 14: 1-19.

BLASKO, K., S.A. KAPLAN, K.G. HIGGINS, R. WOLFSON and B.B. SEARS. 1988. Variation in copy number of a 24-base pair tandem repeat in the chloroplast DNA of Oenothera hookeri strain Johansen. Current Genetics 14: 287-292.

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

BREMER, K. 1990. Combinable component consensus. Cladistics 6: 369-372.

BREMER, B., R.G. OLMSTEAD, L. STRUWE and J.A. SWEERE. 1994. rbcL sequences support exclusion of Retzia, Desfontainia, and Nicodemia from the Gentianales. Plant systematics and evolution 190: 213-230.

BRUNSFIELD, S.J., P.S. SOLTIS, D.E. SOLTIS, P.A. GADEK, C.J. QUINN, D.D. STRENGE and T.A. RANKER. 1994. Phylogenetic relationships among the genera of Taxodiaceae and Cupressaceae: Evidence from rbcL sequences. Systematic Botany 19: 253-262.

BULT, C.J. and E.A. ZIMMER. 1993. Nuclear ribosomal RNA sequences for inferring tribal relationships within Onagraceae. Systematic Botany 18: 48-63.

CAMPBELL, C.S. 1985. The subfamilies and tribes of Gramineae (Poaceae) in the southeastern United States. Journal of the Arnold Arboretum 66: 123-199.

CAMPBELL, C.S. and E.A. KELLOGG. 1987. Sister group relationships of the Poaceae. In: Grass systematics and evolution, eds. T.R. Soderstrom, K.W. Hilu, C.S. Campbell, and M.E. Barkworth. Washington, DC: Smithsonian Institution.

CARPENTER, J.M. 1988. Choosing among multiple equally parsimonious cladograms. Cladistics 4: 291-296.

CASANOVA, J.-L., C. PANNETIER, C. JAULIN and P. KOURILSKY. 1990. Optimal conditions for directly sequencing double-stranded PCR products with Sequenase. Nucleic Acids Research 18: 4028

CHASE, M.W. and H.H. HILLS. 1991. Silica gel: an ideal material for field preservation of leaf samples for DNA studies. Taxon 40: 215-220.

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. HEDREN, 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. GOLENDERG, G.H. LEARN, S.W. GRAHAM, S.C.H. BARRETT, S. DAYANANDAN, AND V.A. ALBERT. 1993. Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid gene rbcL. Annals of the Missouri Botanic Gardens 80: 528-580

CHOO, M.K., R.J. SORENG and J.I. DAVIS. 1994. Phylogenetic relationships among Puccinellia and allied genera of Poaceae as inferred from chloroplast DNA restriction site variation. American Journal of Botany 81: 119-126.

CLARK, L.G., W. ZHANG and J.F. WENDEL. submitted. A phylogeny of the grass family (Poaceae) based on ndhF sequence data.

CLARK, W.D., B.S. GAUT, M.R. DUVALL and M.T. CLEGG. 1993. Phylogenetic relationships of the Bromeliiflorae - Commeliniflorae - Zingiberiflorae complex of monocots based on rbcL sequence comparisons. Annals of the Missouri Botanic Gardens 80: 987-998.

CLAYTON, W.D. 1978. Gramineae. In: Flowering plants of the world. Ed. V.H. Heywood. Mayflower books. N.Y.

CLAYTON, W.D. and S.A. RENVOIZE. 1986. Genera graminum. Grasses of the world. Kew Bull. Additional series XIII. Her Majesty's Stationary Office, London.

CLEGG, M.T. 1993. Chloroplast gene sequences and the study of plant evolution. Proceedings of the National Academy of Sciences, USA 90: 363-367.

CLIFFORD, H.T. 1964. The systematic position of the grass genus Micraira F. Muell. University of Queensland Papers, Department of Botany 4: 87-94

CONERT, H.J. 1961. Die systematik und anatomie der Arundineae. Weinheim, J. Cramer.

CONERT, H.J. 1966. Dregeochloa, eine neue Gattung der Gramineen (Gramineae, Arundinoideae, Danthonieae). Senckenbergiana Biologica 47: 335-343.

CONERT, H.J. 1970. Merxmuellera, eine neue gatting der Gramineen (Poaceae: Arundinoideae). Senckenbergiana Biologica 51: 129-133.

CONERT, H.J. 1971. The genus Danthonia in Africa. Mitteilungen Botanische Staatssammlung, Munchen 10: 299-308.

CONERT, H.J. 1987. Current concepts in the systematics of the Arundinoideae. In: Grass systematics and evolution, eds. T.R. Soderstrom, K.W. Hilu, C.S. Campbell, and M.E. Barkworth. Washington, DC: Smithsonian Institution.

CONNOR, H.E. 1991. Chionochloa Zotov (Gramineae) in New Zealand. New Zealand Journal of Botany. 29: 219-283.

CONNOR, H.E. and M.I. DAWSON. 1993. Evolution of reproduction in Lamprothyrsus (Arundineae: Gramineae). Annals of the Missouri Botanic Garden 80: 512-517.

CONNOR, H.E. and E. EDGAR. 1979. Rytidosperma Steudel (Notodanthonia Zotov) in New Zealand. New Zealand Journal of Botany 17: 311-337.

CONTI, E., A. FISCHBACH and K. SYTSMA. 1993. Tribal relationships in Onagraceae: Implications from rbcL sequence data. Annals of the Missouri Botanic Gardens 80: 672-685.

CUMMINGS, M.P., L.M. KING and E.A. KELLOGG. 1994. Slipped strand mispairing in a plastid gene: rpoC2 in grasses (Poaceae). Molecular Biology and Evolution 11: 1-8.

CURRY, J. and M.K. WALKER-SIMMONS. 1993. Unusual sequences of group 3 LEA (II) mRNA inducible by dehydration stress in wheat. Plant Molecular Biology 21: 907-912.

DAVIS, J.I. and R.J. SORENG. 1993. Phylogenetic structure in the grass family (Poaceae) as inferred from chloroplast DNA restriction site variation. American Journal of Botany 80: 1444-1454

DE PINNA, M.C.C. 1991. Concepts and tests of homology in the cladistic paradigm. Cladistics 7: 367-394.

DE QUEIROZ, A. 1993. For consensus (sometimes). Systematic Biology 42: 368-372.

DE WINTER, B. 1962. The South African Stipeae and Aristideae (Gramineae). (An anatomical, cytological and taxonomic study). Bothalia 8: 201-404

DJIAN, P. AND H. GREEN 1989. Vectorial expansion of the involucrin gene and the relatedness of the hominoids. Proceedings of the American Academy of Sciences, USA 86: 8447-8451.

DOEBLEY, J., M. DURBIN, E.M. GOLENBERG, M.T. CLEGG and D. POW MA. 1990. Evolutionary analysis of the large subunit of carboxylase (rbcL) nucleotide sequence among the grasses (Gramineae). Evolution 44: 1097-1108

DOYLE, J.J. and J.L. DOYLE. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19: 11-15

DOYLE, J.J., J.I. DAVIS, R.J. SORENG, D. GARVIN and M.J. ANDERSON. 1992. Chloroplast DNA inversions and the origin of the grass family (Poaceae). Proceedings of the National Acadamy of Sciences, USA 89: 7722-7726

DURE, L., M. CROUCH, J. HARANDA, T-H. D. HO, J. MUNDY, R QUATRANO, T. THOMAS and Z. R. SUNG. 1989. Common amino acid sequence domains among the LEA proteins of higher plants. Plant Molecular Biology 12: 475-486.

EERNISSE, D.J. and A.G. KLUGE. 1993. Taxonomic congruence versus total evidence, and Amniote phylogeny inferred from fossils, molecules, and morphology. Molecular Biology and Evolution 10: 1170-1195.

ELLIS, R.P. 1977. Leaf anatomy of the South African Danthonieae (Poaceae). I. The genus Dregeochloa. Bothalia 12: 209-213

ELLIS, R.P. 1982. Leaf anatomy of South African Danthonieae (Poaceae). VII. Merxmuellera dura and M. rangei. Bothalia 14: 95-99.

ELLIS, R.P. 1983. Leaf anatomy of the South African Danthonieae (Poaceae). VII. Merxmuellera decora, M. lupulina and M. rufa. Bothalia 14: 197-203.

ELLIS, R.P. 1984. Leaf anatomy of the South African Danthonieae (Poaceae). IX Asthenatherum glaucum. Bothalia 15: 153-159

ELLIS, R.P. 1987. A review of comparative leaf blade anatomy in the systematics of the Poaceae: the past twenty-five years. In: Grass systematics and evolution, eds. T.R. Soderstrom, K.W. Hilu, C.S. Campbell, and M.E. Barkworth. Washington, DC: Smithsonian Institution.

FARRIS, J.S. 1988. HENNIG86 version 1.5.

FARRIS, J.S. 1989. The retention index and the rescaled consistency index. Cladistics 5: 417-419.

FARRIS, J.S. 1994. RNA version 1.01. Naturhistoriska Riksmuseet, Stockholm.

FELSENSTEIN, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Systematics Zoology 27: 401-410.

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

FITCH, W.M. 1970. Distinguishing homologous from analogous proteins. Systematic Zoology 19: 99-113.

GATESY, J., R. DESALLE and W. WHEELER. 1993. Alignment-ambiguous nucleotide sites and the exclusion of systematic data. Molecular Phylogenetics and Evolution 2: 152-157.

GIBBS RUSSELL, G.E., L. WATSON, M. KOEKEMOER, L. SMOOK, N.P. BARKER, H.M. ANDERSON and M.J. DALLWITZ. 1990. Grasses of southern Africa. Memoirs of the Botanical Survey of South Africa No. 58. National Botanical Institute, Pretoria, South Africa.

GIANNASI, D.E., G.R. ZURAWSKI, G. LEARN, and M.T. CLEGG. 1992. Evolutionary relationships of the Caryophyllidae based on comparative *rbc*L sequences. Systematic Botany 17: 1-15

GOLENBERG, E.M., M.T. CLEGG, M.L. DURBIN, J. DOEBLEY and D.P. MA. 1993. Evolution of a noncoding region of the chloroplast Genome, Molecular Phylogenetics and Evolution 2: 52-64..

HAMBY, R.K. and E.A. ZIMMER. 1988. Ribosomal RNA sequences for inferring phylogeny within the grass family (Poaceae). Plant Systematics and Evolution 160: 29-37

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

HEIN, J. 1989a. A new method that simultaneously aligns and reconstructs ancestral sequences for any number of homologous sequences when a phylogeny is given. Molecular Biology and Evolution 6: 649-668.

HEIN, J. 1989b. A tree reconstruction method that is economical in the number of pairwise comparisons used. Molecular Biology and Evolution 6: 669-684.

HENNIG, W. 1966. Phylogenetic systematics. University of Illinois Press, Urbana, Illinois.

HIGGINS, D.G. and P.M. SHARP. 1988. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene 73: 237-244.

HILLIS, D.M. 1987. Molecular versus morphological approaches to systematics. Annual Review of Ecology and Systematics 18: 23-42.

HILLIS, D.M. 1994. Homology in molecular biology. In: Homology: the Hierarchical basis of comparative biology. Ed. B.K. Hall. Academic Press, San Diego.

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

HILU, K.W. and A. ESEN. 1990. Prolamins in systematics of Poaceae subfam. Arundinoideae. Plant Systematics and Evolution 173: 57-70.

HILU, K.W. and K. WRIGHT. 1982. Systematics of Gramineae: a cluster analysis study. Taxon 31: 9-36

IGLOI, G.L., A. MEINKE, I. DöRY and H. KöSSEL. 1990. Nucleotide sequence of maize chloroplast rpo B/C1/C2 operon: Comparison between the derived protein primary structures from various organisms with respect to functional domains. Molecular and General Genetics 221: 379-394.

JOHNSON, C.R. and L. WATSON. 1981. Germination flaps in grass lemmas. Phytomorphology 31: 78-85.

JOHNSON, L.A. and D.E. SOLTIS. 1994. matK sequences and phylogenetic reconstruction in Saxifragaceae s. str. Systematic Botany 19: 143-156

JUKES, T.H. and C.R. CANTOR. 1969. Evolution of protein molecules. In: Mammalian Protein Metabolism. Ed. H.N. Munro. Academic Press, New York

KALLIOLA, R. and S.A. RENVOIZE. 1994. One or more species of Gynerium? (Poaceae). Kew Bulletin 49: 305-320.

KELLOGG, E.A. and C.S. CAMPBELL. 1987. Phylogenetic analysis of the Gramineae. In: Grass systematics and evolution, eds. T.R. Soderstrom, K.W. Hilu, C.S. Campbell, and M.E. Barkworth. Washington, DC: Smithsonian Institution.

KELLOGG, E.A. and L.E. WATSON. 1993. Phylogenetic studies of a large data set. I. Bambusoideae, Andropogonodae, and Pooideae (Gramineae). Botanical Review 59: 273-343

KELLOGG, E.A. and H.P. LINDER. In press. Phylogeny of Poales. Proceedings of the Monocot symposium, Kew.

KLUGE, A.G. 1989. A concern for evidence and a phylogenetic hypothesis of relationships among Epicrates (Boidae, Serpentes). Systematic Zoology 38: 7-25.

KUMAR, S., K. TAMURA and M. NEI. 1993. MEGA: Molecular Evolutionary Genetics Analysis, version 1.0. The Pennsylvania State University, University Park, PA 16802.

LAZARIDES, M. 1979. Micraira F. Muell. (Poaceae; Micrairoideae). Brunonia 2: 67-84

LES, D.H., D.K. GARVIN and C.F. WIMPEE. 1991. Molecular evolutionary history of ancient aquatic angiosperms. Proceedings of the National Academy of Sciences, USA 88: 10119-10123.

LEVINSON, G. and G.A. GUTMAN. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Molecular Biology and Evolution 4: 203-221.

LI, W.-H. and D. GRAUR. 1991. Fundamentals of molecular evolution. Sinauer Associates, Inc. Sunderland, Massachusetts.

LINDER, H.P. and P.J. RUDALL. 1993. The megagametophyte in Anarthria (Anarthriaceae, Poales) and its implications for the phylogeny of the Poales. American Journal of Botany 80: 1455-1464

MADDISON, D.R. 1991. The discovery and importance of multiple islands of most-parsimonious trees. Systematic Zoology 40: 315-328.

MINDELL, D.P. 1991. Aligning DNA sequences: homology and phylogenetic weighting. In: Phylogenetic analysis of DNA sequences. Eds. M.M. Miyamoto and J. Cracraft. Oxford University Press, New York.

MIYAMOTO, M.M. and J. CRACRAFT. 1991. Phylogenetic inference, DNA sequence analysis, and the future of molecular systematics. In: Phylogenetic analysis of DNA sequences. Eds. M.M. Miyamoto and J. Cracraft. Oxford University Press, New York.

NADOT, S., R. BAJON and B. LEJEUNE. 1994. The chloroplast gene rps4 as a tool for the study of Poaceae phylogeny. Plant Systematics and Evolution 191: 27-38.

NEEDLEMAN, S.B. and C.D. WUNSCH. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. Journal of Molecular Biology 48: 443-453.

NEFF, N.A. 1986. A rational basis for a priori character weighting. Systematic Zoology 35: 110-123.

NEI, M. 1991. Relative efficiencies of different tree making methods for molecular data. In: Phylogenetic analysis of DNA sequences. Eds. M.M. Miyamoto and J. Cracraft. Oxford University Press, New York.

NELSON, G. 1994. Homology and systematics. In: Homology: the Hierarchical basis of comparative biology. Ed. B.K. Hall. Academic Press, San Diego.

NEVSKI, S.A. 1937. Act. Inst. Bot. Acad. Sci. U.R.S.S. ser. 1., fasc. 4: 223.

NIXON, K.C. 1992. CLADOS version 1.2.

OLMSTEAD, R.G. and SWEERE, J.A. 1994. Combining data in phylogenetic systematics: an empirical approach using three molecular data sets in the Solanaceae. Systematic Biology 43: 467-481.

OLMSTEAD, R.G., H.J. MICHAELS, K.M. SCOTT and J.D. PALMER. 1992. Monophyly of the Asteridae and identification of their major lineages inferred from DNA sequences of rbcL. Annals of the Missouri Botanic Gardens 79: 249-265.

PAGE, R.D.M. 1989. Comments on component-compatibility in historical biogeography. Cladistics 5: 167-182.

PATTERSON, C. 1982. Morphological characters and homology. In: Problems of phylogenetic reconstruction. Eds. K.A. Joysey and A.E. Friday. Academic Press, London.

PATTERSON, C. 1988. Homology in classical and molecular biology. Molecular Biology and Evolution 5: 603-625.

PENNY, D., M.D. HENDY, E.A. ZIMMER and R.K. HAMBY. 1990. Trees from sequences: Panacea or Pandora's box? Australian Systematic Botany 3: 21-38.

PHILIPSON, M.M. 1977. Haustorial synergids in Cortaderia (Gramineae). New Zealand Journal of Botany 15: 777-778.

PHILIPSON, M.N. and H.E. CONNOR. 1984. Haustorial synergids in danthonioid grasses. Botanical Gazette 145: 78-82.

PILGER, R. 1954. Das System der Gramineae. Botanischer Jahrbucher 76: 281-384.

PLATNICK, N.I. 1989. An empirical comparison of microcomputer parsimony programs, II. Cladistics 5: 145-161.

PRAT, H. 1936. La systematique des Graminees. Annales des Sciences Naturelles, Botanique ser. 10, 18: 165-258.

RENVOIZE, S.A. 1981. The subfamily Arundinoideae and its position in relation to a general classification of the Gramineae. Kew Bulletin 36: 85-102

RENVOIZE, S.A. 1986a. A survey of leaf-blade anatomy in grasses VIII. Arundinoideae. Kew Bulletin 41: 323-338

RENVOIZE, S.A. 1986b. A survey of leaf blade anatomy in grasses IX. Centothecoideae. Kew Bulletin 41: 339-342

RETTIG, J.H., H.D. WILSON and J.R. MANHART. 1992. Phylogeny of the Caryophyllales - gene sequence data. Taxon 41: 210-209

REEDER, J.R. 1957. The embryo in grass systematics. American Journal of Botany 44: 756-768.

SAITOU, N. and M. NEI. 1987. The neighbour joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4: 406-425

SAVARD, L., M. MICHAUD and J. BOUSQUET. 1993. Genetic diversity and phylogenetic relationships between birches and alders using ITS, 18S rRNA, and rbcL gene sequences. Molecular Phylogenetics and Evolution 2: 112-118.

SIEBERT, D.J. 1992. Tree statistics; trees and 'confidence'; consensus trees; alternatives to parsimony; character weighting; character conflict and its resolution. In: Cladistics: a practical course in systematics. P.L. Forey, C.J. Humphries, I.J. Kitching, R.W. Scotland, D.J. Siebert and D.M. Williams. Clarendon Press, Oxford.

SNEATH, P.H.A. and R.R. SOKAL. 1973. Numerical taxonomy: the principles and practice of classification. W,H, Freeman, San Francisco.

SODERSTROM, T.R. 1981. The grass subfamily Centostecoideae. Taxon 30: 614-616.

SOLTIS, D.E., P.S. SOLTIS, M.T. CLEGG and M.DURBIN. 1990. rbcL sequence divergence and phylogenetic relationships in Saxifragaceae sensu lato. Proceedings of the National Acadamy of Sciences, USA. 87: 4640-4644.

SOLTIS, D.E., D.R. MORGAN, A. GRABLE, P.S. SOLTIS and R. KUZOFF. 1993. Molecular systematics of Saxifragaceae *sensu stricto*. *American Journal of Botany* 80: 1056-1081.

SORENG, R.J., J.I. DAVIS and J.J. DOYLE. 1990. A phylogenetic analysis of chloroplast DNA restriction site variation in Poaceae subfam. Pooideae. *Plant Systematics and Evolution* 172: 83-97.

STEELE, K.P. and R. VILGALYS. 1994. Phylogenetic analysis of Polemoniaceae using nucleotide sequences of the plastid gene *mat*K. *Systematic Botany* 19: 126-142.

STEVENS, P.F. 1984. Homology and phylogeny: morphology and systematics. *Systematic Botany* 9: 395-409.

SUH, Y., L.B. THIEN, H.E. REEVE and E.A. ZIMMER. 1993. Molecular evolution and phylogenetic implications of internal transcribed spacer sequences of ribosomal DNA in Winteraceae. *American Journal of Botany* 80: 1042-1055.

SWENSEN, S.M. and B.C. MULLIN. 1994. Phylogenetic affinities of Datiscaceae based on an analysis of nucleotide sequences from the plastid *rbc*L gene. *Systematic Botany* 19: 157-168.

SWOFFORD, D.L. 1993. PAUP: Phylogenetic Analysis Using Parsimony. Version 3.1.1. Smithsonian Institution, Washington DC.

SWOFFORD, D.L. and G.J. OLSEN. 1990. Phylogeny reconstruction. Pp. 411-501 In: *Molecular Systematics*. Eds. D.M. Hillis and C. Moritz. Sinauer Associates, Inc. Sunderland, Massachusetts, USA.

SWOFFORD, D.L. and D.P. BEGLE. 1993. PAUP. Phylogenetic analysis using parsimony (manual to version 3.1). Smithsonian Institution, Washington DC.

TATEOKA, T. 1957. Miscellaneous papers on the phylogeny of Poaceae (10). Proposition of a new phylogenetic system of Poaceae. Journal of Japanese Botany 29: 341-347.

TATEOKA, T., S. INOUE and S. KAWANO. 1959. Notes on some grasses. IX. systematic significance of bicellular microhairs of leaf epidermis. Botanical Gazette 121: 80-91

TOMLINSON, K.L. 1985. Comparative anatomical studies in Danthonia sensu lato (Danthonieae: Poaceae). Aliso 11: 97-114.

VERBOOM, G.A., H.P. LINDER and N.P. BARKER. 1994. Haustorial synergids: An important character in the systematics of danthonioid grasses (Arundinoideae: Poaceae). American Journal of Botany 81: 1601-1610.

VICKERY, J.W. 1956. A revision of the Australian species of Danthonia DC. Contributions from the New South Whales National Herbarium 2: 240-325

WAGNER, G.P. 1986. The systems approach: an interface between development and population genetic aspects of evolution. In: Patterns and processes in the history of life. Eds. D.M. Raup and D. Jablonski. Springer, Berlin.

WATERS, D.A., M.A. BUCHEIM, R.A. DEWEY and R.L. CHAPMAN 1992. Preliminary inferences of the phylogeny of bryophytes from nuclear-encoded ribosomal RNA sequences. American Journal of Botany 79: 459-466.

WATSON, L. 1987. Automated descriptions of grass genera. Pp 343-354. In: Grass systematics and evolution. Eds. T.R. Soderstrom, K.W. Hilu, C.S. Campbell, and M.E. Barkworth. Washington, DC: Smithsonian Institution.

WATSON, L. 1990. The grass family, Poaceae. Pp. 1-31 in Reproductive versatility in the grasses. Ed. G.P. Chapman Cambridge: Cambridge University Press.

WATSON, L. AND H.T. CLIFFORD. 1976. The major groups of Australian grasses: A guide to sampling. Australian Journal of Botany 24: 489-507.

WATSON, L. and M.J. DALLWITZ. 1982. Grass genera of the world: illustrations of characters, descriptions, classification, interactive identification, information retrieval. Research School of Biological Sciences, Australian National University, Canberra.

WATSON, L. and M.J. DALLWITZ. 1992. The grass genera of the world. CAB International, Wallingford, UK.

WATSON, L., M.J. DALLWITZ, and C.R. JOHNSTONE. 1986. Grass genera of the world: 728 detailed descriptions from an automated database. Australian Journal of Botany 34: 223-230.

WHEELER, W. and D. GLADSTEIN. 1993. MALIGN Version 1.89. Dept. of Invertebrates, American Museum of Natural History.

WILCOX, L.W., P.A. FUERST and G.L. FLOYD. 1993. Phylogenetic relationships of four Charophyceaen green algae inferred from complete nuclear-encoded small subunit RRNA gene sequences. American Journal of Botany 80: 1029-1033.

WILLIAMS, D.M. 1992. DNA analysis: theory. Pp. 89-101. In: Cladistics: a practical course in systematics. P.L. Forey, C.J. Humphries, I.J. Kitching, R.W. Scotland, D.J. Siebert and D.M. Williams. Clarendon Press, Oxford.

WILLIAMS, D.M. 1994. Combining trees and combining data. Taxon 43: 449-453.

WILLIAMS, S.E., V.A. ALBERT and M.W. CHASE. 1994. Relationships of Droseraceae: A cladistic analysis of *rbcL* sequence and morphological data. *American Journal of Botany* 81: 1027-1037.

WOLFSON, R., K.G. HIGGINS and B. SEARS. 1991. Evidence for replication slippage in the evolution of *Oenothera* chloroplast DNA. *Molecular biology and evolution* 8: 709-720

WOJCIECHOWSKI, M.F., M.J. SANDERSON, B.G. BALDWIN and M.J. DONOGHUE. 1993. Monophyly of aneuploid *Astragalus* (Fabaceae): Evidence from nuclear ribosomal DNA internal transcribed spacer sequences. *American Journal of Botany* 80: 711-722.

WOLFE, K.H., M. GOUY, Y.-W. YANG, P.M. SHARP and W.-H. LI. 1989. Date of the monocot - dicot divergence estimated from chloroplast DNA sequence data. *Proceedings of the National Academy of Science, USA* 86: 6201-6205.

YEOH, H.-H. and L. WATSON. 1981. Systematic variation in amino acid compositions of grass caryopses. *Phytochemistry* 20: 1041-1051.

ZOTOV, V.D. 1963. Synopsis of the grass subfamily Arundinoideae in New Zealand. *New Zealand Journal of Botany* 1: 78-136.