Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

GRASS EVOLUTION AND DIVERSIFICATION: A PHYLOGENETIC APPROACH

ph.D. Botany

GRASS EVOLUTION AND DIVERSIFICATION: A PHYLOGENETIC APPROACH

Thesis submitted to the University of Dublin, Trinity College
for the
Degree of Doctor of Philosophy (Ph.D.)

by

Yanis Bouchenak-Khelladi

School of Natural Sciences
Botany Department
University of Dublin, Trinity College

2007

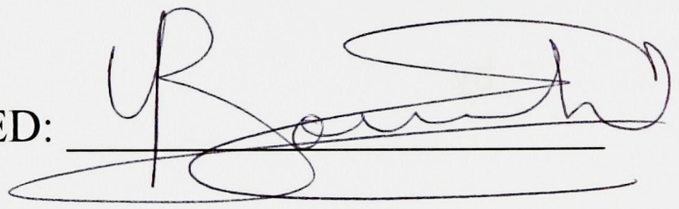
Research conducted under the supervision of

Dr. Trevor R. Hodkinson
School of Natural Sciences, Botany Department, University of Dublin, Trinity College
Dr. Nicolas Salamin
Department of Ecology and Evolution, University of Lausanne
Dr. Vincent Savolainen
Jodrell Laboratory, Molecular Systematics Section, Royal Botanic Gardens Kew, London

DECLARATION

I thereby certify that this thesis has not been previously submitted as an exercise for a degree at this or any other University. This thesis contains research based on my own work, except where otherwise stated.

I grant fully permission to the Library of Trinity College to lend or copy this thesis upon request.

SIGNED: 

ACKNOWLEDGMENTS

I wish to thank Trevor Hodkinson and Vincent Savolainen for all the tremendous support they gave me during the last three years. I wish to thank also Nicolas Salamin for his encouragement and his very useful advices on all aspects of this thesis.

It has been a unique experience to study in Ireland, and I would like to express special acknowledgment to Trevor for all he did for me. He was always available for any kind of request, and he made the difficult life of a PhD student a lot easier thanks to his patience and his understanding. I am also grateful to Vincent for his advice on scientific papers and his open-mind after work. I would like to thank Prof. Mark W. Chase for his hospitality when I was arriving in Kew for the first time and all the members of the Molecular Systematics Section from the Jodrell Laboratory.

I wish to thank Dr. G. Anthony Verboom for the great discussions we had, and will have, at work and outside work, and for his help during my PhD. Finally, many thanks to all the Botany members at the University of Cape Town and at the University of Johannesburg. Also, in South Africa, I want to thank Dr. Michelle and Prof. Herman van der Bank for their kindness and hospitality while I was in Johannesburg, Olivier Maurin, Dr. Renaud Lahaye and Dr. Sylvie Duthoit for their friendship.

Finally, I want to dedicate these three years work to all members of my very large family in Algeria, France and Spain, to all my friends worldwide, and to Caro for her tremendous support these last three years.

This project was funded by Enterprise Ireland Basic Research Grant SC/2003/0437 and Trinity College Travel Awards.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

Phylogenetic inference and DNA sequencing

BI = Bayesian inference

BS = Bootstrap support

CI = Consistency index

CTAB = Cetyl trimethylammonium bromide

GTR = General time reversible

HKY85 = Hasegawa, Kishino and Yano (1985) model of substitution

K81 = Kimura (1981) model of substitution

LBA = Long branch attraction

***matK* = Maturase K**

MCMC = Markov chain monte carlo

MP = Maximum parsimony

NPRS = Non parametric rate smoothing

PP = Posterior probabilities

***rbcL* = Ribulose-1,5-bisphosphate carboxylase large subunit**

RI = Retention index

SH test = Shimodaira and Hasegawa (1999) test

TBR = Tree bisection and reconnection

***trnL-F* = Partial leucine transfer RNA (trnL) exons + trnL intron + trnL-trnF intergenic spacer + partial phenylalanine RNA (trnF) exon**

TVM = Transversional model of substitution (Posada and Crandall, 1998)

Grass systematics and evolution

ACCTRAN = Accelerated transformation

BEP = Clade containing Bambusoideae, Ehrhartoideae and Pooideae

DELTRAN = Delayed transformation

EDL = Early-diverging lineage

ERM =Equal rates Markov process

GPWG = Grass Phylogeny Working Group (2001)

PACCAD = Clade containing Panicoideae, Arundinoideae, Centothecoideae, Chloridoideae, Aristidoideae and Danthonioideae

s.l = sensu lato

s.str. = sensu stricto

ABSTRACT

The growth in size of phylogenetic trees, over the last 20 years, has allowed evolutionary biologists to better test hypotheses about the evolutionary history of organisms, and especially those of species rich taxa such as the grasses. Grasses are one of the most diverse families in the angiosperms, consisting of approximately 10,000 species and 600-700 genera and it is essential to investigate evolution and diversification in this group to advance the understanding of the processes shaping the diversity of its life forms. Therefore, this thesis aimed to provide comprehensive phylogenetic trees of the grass family in order to establish macro-evolutionary hypotheses and investigate patterns and processes of grass diversification.

One aspect of this thesis was to infer the most comprehensive phylogenetic tree of the grasses in order to establish robust phylogenetic relationships among grass lineages. In Chapter 2, a much larger representation of grass diversity (82 % of tribes and 42 % of genera) was included than any previous study. Phylogenetic inferences using DNA sequences of three plastid regions: *rbcL*, *matK* and *trnL-F* were performed using maximum parsimony and Bayesian inferences. The resulting trees resolved most of the subfamily relationships within the BEP (Bambusoideae, Ehrhartoideae and Pooideae) and PACCAD (Panicoideae, Aristidoideae, Centothecoideae, Chloridoideae, Arundinoideae and Danthonioideae) clades, which had previously been unclear, such as, among others: (i) the composition of the BEP clade and the sister-relationship of Ehrhartoideae and Bambusoideae + Pooideae, (ii) the paraphyly of tribe Bambuseae, (iii) the position of *Gynerium* as sister to Panicoideae, and (iv) the monophyly of *Eriachne* + *Micraira*. The thesis also highlights how phylogenetic accuracy has been largely neglected in phylogenetic studies of grasses and other organisms with respect to missing data. It is shown that accuracy can be maintained even with the presence of a relatively large amount of missing data in combined analyses (i.e. 33 % of the taxa lacking one or more genes in the combined analysis). However, bootstrap support values, and to a lesser extent Bayesian inference posterior probabilities, are generally lower in combined gene analyses involving missing data than those not including them. We propose a fully

resolved tree for the grass family at subfamily level and indicate the most likely inter-relationships of all included tribes in our analysis (i.e. 82% of total grass tribes).

A second aspect of this thesis was to use these large phylogenetic trees to test for evolutionary patterns of diversification in the grass family. It is generally hard to determine detailed patterns of grass diversification from previous phylogenetic analyses within the family because of a poor taxon sampling. Thus, Chapter 3 aimed to study the temporal and topological patterns of grass diversification and investigate processes leading to such diversification. A complete generic level phylogenetic tree with 815 genera was generated by compiling molecular and morphological databases and performing topological constraint phylogenetic inferences. This was used to test for statistically significant shifts in diversification rates among lineages with the absence of missing taxa. This was coupled with Bayesian molecular dating methods and geographical and ecological mapping. The approach taken that incorporated different datasets (molecular, morphological, ecological and geographical), which have in common an overlap of taxa, has allowed a more detailed analysis of phylogenetic diversification than previous studies. The results show that (i) the grasses may have undergone at least fifteen differential shifts in diversification among lineages during their evolution, (ii) an African origin of the family is most probable (using Bremer's Area of Ancestral Origin inference method) and this has been estimated to have occurred in the late Cretaceous (around 70 Mya), (iii) the grasses dispersed to all continents by 30 million years after their origin, (iv) major diversification events of the BEP clade members (C_3 grasses) occurred in the Paleocene and Eocene (between 55 and 35 Mya) possibly due to the decline of forested environments, (v) there was a later divergence of the PACCAD clade from the Oligocene (between 35 and 25 Mya), possibly due to an early adaptation to arid habitats with recent dispersals from Africa to Eurasia and to the New World and finally (v) relatively recent diversification events within the PACCAD clade and the expansion of C_4 grasses occurred by the middle Miocene (around 15 Mya).

Among the several potential environmental determinants on the ecological success of open-habitat grasses, climate change and low CO_2 levels during the Cenozoic are the most commonly discussed. Despite these, other disturbances, such as herbivory, may also have limited the abundance of closed-habitats dominated by trees. However, the effect of such selection pressure on grass evolution has not been

previously tested. Therefore, Chapter 4 of this thesis aimed to evaluate if changes in silica body density in grass epidermal cells, which are among the few substances capable of inducing morphological changes to animal mouthparts, are correlated with evolutionary changes in molar morphology of ungulates. It also aimed to reconstruct how silica body density has varied during the evolution of the grasses and to see how this varies among major lineages of grasses (subfamilies). Historical changes in silica body densities were recorded from a dated phylogenetic tree (using maximum likelihood and least square parsimony methods). Such changes were compared using rank correlation analysis with the evolution of lophedness (shearing blades on the 2nd upper molar) of ungulates through the Cenozoic. Based on the results, the overall trend of variation in silica body density through time can be summarized as follows: (i) there are differential responses of grasses in response to increased lophedness of ungulates through the Cenozoic, (ii) increase in silica density is correlated with the adaptation of closed-habitats but a higher sampling is needed to further test this hypothesis, (iii) increase in silica density occured for PACCAD lineages (especially Aristidoideae and Danthonioideae) but not for BEP lineages, and (iv) C_4 grasses may have undergone an increase in silica density in response to increasing grazing rates through the Miocene. The increase in lophedness of late Oligocene-Miocene ungulates is correlated with an increase in silica density of C_4 PACCAD grasses. The Miocene radiations of ungulates evolved dental adaptations to deal with vegetation of low primary productivity. It is then plausible that the Miocene ungulates evolved higher loph numbers on their 2nd upper molar to deal with increasing silica density of C_4 grasses. This study also reveals a phylogenetic approach for evaluating the effects of grazing on grass evolution. The most challenging aspect is the precise selection of traits, which may be correlated with grass evolutionary response to herbivory. Other traits (such as leaf tensility, leaf dry matter content, rhizomatous growth and tannin-like substances) should be analyzed in future studies using a phylogenetic approach to reveal evolutionary trends in grass-ungulate interactions.

The applications of biogeographical, ecological, paleontological and taxonomic data coupled with phylogenetic trees have provided robust perspectives for understanding the evolutionary history of the grasses. It is anticipated that the approach taken in this thesis can be further developed to better understand macro-

evolutionary patterns and processes of this highly important group of organisms, and also be applied to a whole host of other species-rich groups.

nodes even in the absence of a molecular clock. Most of them are likelihood approaches that estimate node ages under explicit or heuristic models of how substitution rates vary among lineages (Thorne, 1998; Yoder and Yang, 2000).

Barraclough and Nee (2001) considered that studies using molecular phylogenetic trees to solve speciation problems rely on fundamental issues, and on several assumptions. First, this approach is based on the reconstruction of evolutionary relationships between species within a clade. The entities considered might not correspond to the taxonomically recognised species (Avise, 2000). Other processes than speciation could explain species diversity and the attributes of present-day species, such as extinction events and/or phenotypic evolution (Barraclough and Nee, 2001). There is generally no record of speciation events involving species that went extinct because phylogeny reconstruction relies on living species (Barraclough and Nee, 2001). However, there are cases where fossils have been included in phylogenetic analyses but despite this, their utility has been low. To obtain an accurate view of speciation in a higher group (such as family level), a large proportion of species from that group should be sampled. Thus, reconstructing species-level phylogenies requires a large sampling effort within the taxonomic group studied. Even though some "targeted" organisms have been heavily sequenced (e.g. such as *Arabidopsis*, *Oryza*, *Homo sapiens sapiens* and *Brachydanio rerio*) (http://www.ncbi.nlm.nih.gov/), a large sequencing effort is needed if one wants to integrate a large number of species within an accurate phylogenetic framework, particularly for species rich groups that represent the highest proportion of the total global species diversity. However, as sequencing technology is becoming faster and cheaper, heavy sampling and sequencing is possible. We are therefore in a position to minimize problems regarding comprehensive inter-species phylogenies. The main practical obstacle will be obtaining samples of the species via field work (Hodkinson et al., 2007a).

1.2 Testing macroevolutionary hypotheses using phylogenetic trees

Collection and analyses of appropriate data to study macro-evolutionary patterns of groups of species diversifying over time should help reveal the evolutionary forces and the genetic changes that have been responsible for these patterns and the production of new species (Coyne and Orr, 2004). Species-level phylogenetic trees allow us to consider the rates of species formation within a clade and the correlation of such rates with morphological or ecological traits (Barraclough and Nee, 2001; Coyne and Orr, 2004). Different evolutionary forces might produce species at different rates. For instance, speciation via sexual selection, various forms of sympatric speciation (e.g. via polyploidy or hybridisation) and founder-effect speciation can occur quickly (Coyne and Orr, 2004). Adaptive radiation in newly colonised areas could initially lead to high rates of speciation that decrease as niches become filled (Schluter, 2000). This implies that a variation in speciation rates occurs over time. This could be, in theory, tested with fossil record data (Coyne and Orr, 2004). However, a number of methods for estimating speciation rates from phylogenetic trees containing all species within a clade have been proposed and used (Nee, 2001). They use information on the time elapsed between branching events and reconstruct, graphically, the number of lineages through time (Barraclough and Nee, 2001). Baldwin and Sanderson (1998) used this approach to estimate speciation rate during the radiation of Hawaiian silverswords (Asteraceae). They found that these organisms have speciated at comparable rates to those observed from fossil evidence during continental radiation. However, it is more complicated to infer speciation rates if the data are not consistent with a constant speciation rate model (Barraclough and Nee, 2001). Sampling and taxonomic artefacts can affect the observed rate of speciation by, for example, underestimating the number of nodes towards the present (Nee et al., 1994). The fact that some groups of organisms have more species than others has interested many evolutionary biologists (Sanderson and Donoghue, 1996). In this perspective, several studies found that several angiosperm lineages have produced more species than others (Sanderson and Donoghue, 1994; Sanderson and Wojciechowski, 1996; Barraclough et al., 1996; Chase et al., 2000; Barraclough and Savolainen, 2001; Savolainen et al., 2002). This may seem obvious because species number varies so highly between genera and higher ranked taxa of angiosperms.

However, identifying specific and statistically supported shifts in diversification requires a phylogenetic approach.

Molecular phylogenetic approaches over the past two or three decades have offered alternative methods that can indirectly study the patterns and processes of diversification (Brooks and McLennan, 1991; Barraclough and Nee, 2001) but which generally require the sampling of all the species from within that group (Barraclough and Nee, 2001). Missing species reduce the sample size used for the reconstruction of speciation events and can introduce bias especially by removing the most recent speciation events (Nee et al., 1994). Therefore, studying diversification patterns and processes using phylogenetic approaches for species-rich groups of organisms remains problematic and a comprehensive inference of the species phylogeny is needed (Hey, 1992; Nee et al., 1994; Sanderson and Donoghue, 1996; Paradis, 2003). Sampling all taxa to reconstruct comprehensive species-level phylogenetic trees is currently not practically possible for many groups of organism mainly because of taxon availability, size of the sequencing effort, and the computationally demanding phylogenetic analyses that are required (Hodkinson et al., 2007b). In such cases, the use of an "exemplar" approach (Yeates, 1995), that is sampling one representative at any taxonomic rank such as genus, tribe, or subfamily, may be considered reliable as long as it includes a high proportion of the overall species diversity (i.e. species number within the group) at any taxonomic rank. Once a comprehensive phylogenetic framework has been achieved, two sources of information are relevant to the study of diversification rates: the topological distribution of species diversity and the temporal distribution of branching events (Chan and Moore, 2005). Topological methods allow the assessment of tree shape and imbalance and hence an assessment of diversification patterns across the lineages (Slowinski and Guyer, 1989a; Slowinski and Guyer, 1989b; Chan and Moore, 2005). Temporal methods could be regarded as offering greater power over topological ones because they incorporate phylogenetic branch lengths and can provide estimates of the timing of diversification (Sanderson and Donoghue, 1996). However, it is often difficult to accurately infer branch-lengths for comprehensive phylogenetic trees (either single/multi-gene inferences or supertree reconstructions). Indeed, supertree methods generally do not provide branch-length estimates but progress is being made in this area (Lapointe and Levasseur, 2004; Moore et al., 2004; Vos and Mooers, 2004). The disadvantages of topological

methods (i.e. lack of branch-length information) in comparison to temporal ones might be counterbalanced to some extent by the advantage that topological analyses can more easily incorporate comprehensive taxonomic sampling; as would be the case with supertrees or with the compilation of molecular and morphological data to incorporate the overall diversity within the group under study. This is because, both temporal and topological methods are sensitive to incomplete and/or non-random taxon sampling (Moore et al., 2004) and trees with better sampling are likely to provide more accurate measures of diversification.

1.3 Grass diversity and classification: a case study

1.3.1 History of grass classification

The Poaceae (grass family) are a lineage with more than 10,000 species and between 600 and 700 genera (Renvoize and Clayton, 1992; Watson and Dallwitz, 1992; GPWG, 2001). They include cereals such as wheat, rice, maize, sugar cane, millet and rye as well as numerous forage grasses such as *Brachiaria*, *Lolium* (rye-grass), *Festuca* (fescue) and *Dactylis* (cocksfoot) (Clayton and Renvoize, 1986). Grasses occur on all continents and are ecologically dominant in many ecosystems such as the African and South American savannas (Shantz, 1954). Because of their ecological and economical importance, grasses have been studied for centuries and efforts to produce a comprehensive taxonomic classification began over 200 years ago (Brown, 1810; Brown, 1814; Calderon and Soderstrom, 1980; Clark et al., 1995). There have been constant changes in classification of Poaceae (GPWG, 2001). Brown (1810, 1814) was the first to attempt to define groups of tribes, or what we call now subfamilies. Several classifications of grasses based on morphological traits were proposed in the 19th century (Calderon and Soderstrom, 1980; Gould and Shaw, 1983; Campbell, 1985). However, a different perspective on grass evolution and relationships began to emerge by the end of the 19th century (van Tieghem, 1897). Additional data on leaf anatomy, embryology and cytology were accumulated and incorporated into evolutionary and classification schemes (van Tieghem, 1897; Prat, 1932). Then, several classification systems were published throughout the 20th century such as (Prat, 1960; Stebbins and Crampton, 1961; Clayton and Renvoize,

1986; Renvoize and Clayton, 1992; Watson and Dallwitz, 1992). The number of sub-families ranged from 2 (Tzvelev, 1989) to 13 (Caro, 1982).

The first molecular phylogenetic trees of the grasses were published by Hamby and Zimmer (1988) and Doebley et al. (1990), and showed a significant support for a Pooideae and a PACC clade (i.e. containing: Panicoideae, Arundinoideae, Centothecoideae and Chloridoideae). Within the last decade, phylogenetic analyses have converged on a set of well-supported relationships within Poaceae (Nadot et al., 1994; Barker et al., 1995; Clark et al., 1995; Duvall and Morton, 1996; Soreng and Davis, 1998; Barker et al., 1999; Hilu et al., 1999; GPWG, 2001). The first molecular phylogenetic trees of the grasses using the *rbcL* plastid region (Hamby and Zimmer, 1988; Doebley et al., 1990) supported the monophyly of a group containing Panicoideae, Arundinoideae, Centothecoideae and Chloridoideae (the PACC clade). However, only nine grass species (with two outgroups) from three recognized subfamilies were sequenced. A larger sample with a total of 47 species representing 26 tribes and six recognized subfamilies, was included by Clark et al. (1995) using the plastid gene *ndhF*. They recovered a tree with two major groups, the PACC and the BEP (containing Bambusoideae, Ehrhartoideae and Pooideae) clades and two early-diverging lineages (lineages sister to the rest of the grasses; hereafter EDL) one containing *Anomochloa* Brong. and *Streptochaeta* Schrad., and the other *Pharus* P. Browne. One of the most significant published combined data analysis consisted of plastid and nuclear DNA sequences, plastid restriction site data and morphological data and included 61 genera, but it represented only 8% of all grass genera (GPWG, 2001). A relatively robust and well-resolved topology was obtained, supporting a PACCAD group (PACC plus Aristidoideae and Danthonioideae), a BEP group and three EDLs (recognized as Anomochlooideae, Pharoideae and Puelioideae, respectively). The most recent grass classification considers 12 subfamilies (Anomochlooideae, Pharoideae, Puelioideae, Bambusoideae, Ehrhartoideae, Pooideae, Aristidoideae, Arundinoideae, Danthonioideae, Centothecoideae, Panicoideae, Chloridoideae) and 5 *incertae sedis* genera (*Eriachne*, *Micraira*, *Streptogyna*, *Cyperochloa*, and *Gynerium*) (GPWG, 2001). Nevertheless, several relationships among lineages within clades of grasses were not adequately resolved and needed more molecular characters to address them (GPWG, 2001). Issues regarding taxon sampling have been raised by the GPWG (2001) such as

comparability (there are varied assemblage of species or genera), and the influence of hypotheses on taxon sampling among tribes or subfamilies. The relationships between many major lineages in the PACCAD clade (Panicoideae, Arundinoideae, Chloridoideae, Centothecoideae, Aristidoideae and Danthonioideae) were not resolved even though the subfamilies were strongly supported as monophyletic (GPWG, 2001). Within this clade, the Centothecoideae is the only one not strongly supported.

The monophyly of the BEP clade is not generally strongly supported (GPWG, 2001). Within this clade, the relationships between the Pooideae subfamily and the PACCAD clade are not clear while Bambusoideae and Ehrhartoideae are strongly supported and related to EDLs such as Anomochlooideae, *Streptochaeta*, Pharoideae, *Guaduella* and *Puelia*. Finally, it is believed that the results of the trees showing the monophyly of the Anomochlooideae may be caused by a long-branch attraction in the phylogenetic reconstruction (Felsenstein, 1978) as both genera (i.e. *Anomochloa* and *Streptochaeta*) occupy long branches in trees (GPWG, 2001). The GPWG (2001) suggest that the inclusion of other species from these genera could help break up the long branches and might have an effect on the monophyletic status of the groups.

1.3.2 Grass diversification in a phylogenetic framework

The sample sizes of all previous phylogenetic analyses of the family, with the exception of supertree reconstructions (Salamin et al., 2002; Hodkinson et al., 2007a), ranged from 11 (Doebley et al., 1990) to more than 100 species (Hsiao et al., 1999). Most analyses with large sample sizes have concentrated on smaller taxonomic units than the whole grass family and have included only a small proportion of grass diversity (Petersen and Seberg, 1997; Duvall et al., 2001; Doust and Kellogg, 2002; Hodkinson et al., 2002; Mathews et al., 2002; Aliscioni et al., 2003). For most groups of organisms, only a few species have been sequenced for many genes and a few genes have been sequenced for many species. Consequently, a supermatrix approach that tries to gather most of the potential data available for phylogenetic purposes very often results in data sets containing large amounts of missing data (Sanderson and Driskell, 2003). Producing large datasets by sampling the same taxon for several genes requires either a centralized effort in a single

laboratory, or a coordinated one among multiple laboratories. Such ventures are rare and they require large amounts of time, money and scientists. However, as suggested by Wiens (2005), it may be possible to reap the benefits of increased taxon sampling without having data for all characters for all taxa (i.e. by incorporating taxa with missing data into analyses). Therefore, increased taxon sampling might be obtained far more readily and cheaply than expected.

It is generally hard to determine detailed patterns of grass diversification from previous phylogenetic analyses because of a poor taxon sampling within the family (Kellogg, 2000; Hodkinson et al., 2007b). Even though there have been great advances in grass phylogenetics, few, or arguably no truly large and comprehensive phylogenetic trees of the family have been produced. Partly because of this, few studies have tried to investigate patterns of diversification in the grasses using a phylogenetic framework (Hodkinson et al., 2007a). Furthermore, studies trying to date and characterize patterns of diversification are scarce and insufficiently detailed within the family (Bremer, 2002).

1.4 Evolution and origin of the grasses

In terms of their evolution, it is believed that grasses originated about 50-70 million years ago (Mya) (Jacobs et al., 1999; Kellogg, 2000). The earliest grass fossils are pollen grains from the Paleocene (from 65 to 55 Mya) of South America and Africa (Jacobs et al., 1999). However, a recent study by Prasad et al. (2005) found the oldest grass fossils in India under the forms of phytoliths (silica bodies on leaf epidermis) in titanosaur coprolites from the late Cretaceous (90 Mya). Due to a poor fossil record, it is not clear how the present-day distribution of grasses was established. It could have occured by a long-distance dispersal across the Atlantic and the Indian oceans or across a continuous Gondwanan equatorial forest (GPWG, 2001; Bremer, 2002). Present-day distribution patterns do not readily indicate a possible origin of grasses (GPWG, 2001). The EDLs are found in tropical regions of South America, Africa and Asia (Judziewicz and Soderstrom, 1989; Soderstrom et al., 1987; Clark et al., 2000), suggesting a Gondwanan origin of the family. This origin has also been dated at about 75 Mya, using the Non Parametric Rate Smoothing (NPRS)

(Sanderson, 1997) molecular dating method by Bremer (2002). Kellogg (2000) and the GPWG (2001) suggested that most of the 10,000 species of grasses had evolved by about 20-25 Mya after the divergence of the subfamilies. Kellogg (2000) indicated that diversification of grasses could have occurred a minimum of 23 Mya after the origin of peculiar morphological characteristics: conventional spikelet, grass embryo, floral morphology and cell alternation in the leaf epidermis. Near global spread of grass-dominated ecosystems is thought to have occurred by the mid-Miocene (Cerling et al., 1997; GPWG, 2001), which corresponds to the establishment of all the major lineages by about 20 to 25 Mya (Jacobs et al., 1999). For instance, in a recent study using phytolith assemblage data, Stromberg (2005) suggests that open-habitat grasses had undergone great taxonomic diversification by the early Oligocene (34 Mya), but became ecologically dominant only by the late Oligocene-early Miocene (between 25 and 20 Mya) in North America. Using carbon isotopic composition of paleosols (Cerling et al., 1997) and fossil tooth enamel evidence (MacFadden and Cerling, 1994), the appearance of C_4 grasses has been estimated to have occurred by 15 Mya; they expanded globally by 7 to 5 Mya (Sage and Monson, 1999).

The C_4 photosynthetic pathway occurs in only 3% of the flowering plant, but in nearly half of the grasses (i.e. PACCAD clade) (Sage and Monson, 1999). It is believed to be the result of the co-occurrence of multiple biochemical and histological characteristics (Kellogg, 2000). The phylogenetic tree of Kellogg (2000) shows that the diversification of the C_4 lineages should have occurred between 15 and 32 Mya. This hypothesis is confirmed by macrofossils and isotopic records (Nambudiri et al., 1978; Kingston et al., 1994; Latorre et al., 1997). Other characteristics acquired later in the evolution of the PACCAD clade may have been more important in its diversification and ecological success but this remains to be tested. Indeed, drought tolerance and the ability to grow in dry open habitats appeared in very recent geological times (Kellogg, 2001).

1.5 Coevolution and grasses

In a recent study, Stromberg (2005) suggested that external factors triggered alterations in vegetation structures during the late Oligocene or early Miocene, allowing the spread of open-habitat grasses. Among the several potential

environmental influences on the ecological success of open-habitat grasses, climate change and low CO_2 levels during the Cenozoic are the most commonly discussed (Stromberg, 2005; Sage and Monson, 1999). Interaction between low CO_2 levels, drought and frequent fires may have promoted the spread of open-habitat grasses at the expense of forest trees (Bond et al., 2003). Other disturbances, such as herbivory, may also have limited the abundance of closed habitats dominated by trees (Stromberg, 2005). The role of herbivory in the evolution of open habitat grasses has not been investigated in detail, even though the spread of grasslands may have been associated with increasing grazing rates throughout the Miocene (Chapman, 1996a).

Coevolution occurs when two or more species influence each other's evolution (Ridley, 2004); it can involve co-adaptations or co-speciation. Co-adaptations between prey and predators, as suggested by (Jerison, 1973), may be the result of reciprocal selective pressure. Predators and prey typically show an evolutionary pattern called escalation (Vermeij, 1999). By escalation, Vermeij (1999) means that the improvement in predatory adaptations may be matched by improvements in prey defences. Coevolutionary processes, sometimes held as isolated and occasional processes (Thompson, 2005), are among the least understood aspects of reciprocal evolutionary change (Thompson, 1982). However, it is recognized that both adaptive radiation and coevolution of species are two of the major processes organizing biodiversity (Lunau, 2004). For instance, the tremendous species diversity exhibited by terrestrial plants and their natural enemies (including viral, bacterial and fungal pathogens, and invertebrate and vertebrate herbivores) has been of a major interest to evolutionary biologists for over a century (Rausher, 1996), who have tried to understand if much of this diversity arose from coevolutionary processes. Coevolution can be detected by adaptive traits related to the coevolutionary partner (Lunau, 2004). Ehrlich and Raven's theory of coevolution (1964) has been the most influential concept in plant/herbivore evolutionary interactions. Even though coevolution refers to a set of processes including taxonomic relationships between interacting species and the persistence of the interactions (Thompson, 1982), two patterns can be considered as the main principles in plant/herbivore coevolution: (i) the selection imposed by herbivores which causes plant populations to diverge in "defensive" characteristics, and (ii) the selection imposed by plant "defensive"

characteristics which cause herbivore populations to diverge in characteristics associated with the exploitation of the host plant (Rausher, 1996).

The evolution of the three-toed horses coincided with the diminution of the tree cover and the development of a savanna type of habitat (Chapman, 1996a). A major and rapid radiation of vertebrate herbivores (Equidae and Bovidae families) is thought to have occurred between 20 and 10 Mya (MacFadden and Cerling, 1994; Hassanin and Douzery, 1999). The emergence of Bovidae is thought to have occured around 20 Mya, and its evolution through the Miocene followed two main episodes: (i) a split between Eurasia and Africa which led to the development of Bovinae (cattle-like bovids) and Antilopinae (gazelles and antelopes) respectively, and, (ii) explosive radiations during the middle Miocene to the early Pliocene (Hassanin and Douzery, 1999). This period was marked by an important global climate change promoting the spread of grasslands and the evolution of bovids adapted to a savanna type habitat (Cerling et al., 1997; Janis et al., 2002). Equidae (horses) underwent a high speciation and diversification during the same period (20-10 Mya) but this was principally centered in North America (MacFadden and Cerling, 1994). The classical explanation, as proposed by MacFadden and Cerling (1994), is an explosive adaptive radiation from low- to high-crowned (hypsodont) horses. The changes in dental morphology of ungulates might have coincided with the diminution of the tree cover and the development of a savanna type habitat (Chapman, 1996a). Jernvall et al. (1996) suggested that Miocene ungulates evolved increasingly disparate crown types together with dietary specialization in more fibrous vegetation. One could suppose that herbivores apply a selective pressure on grasses by grazing, so that grasses adapt by increasing leaf toughness. Reciprocally, herbivores might have evolved tooth structures to cope with an increase in leaf toughness. Graminoid grazing tolerance and the nearly simultaneous increase of grasses and their grazers in the fossil record (Stebbins and Crampton, 1961) suggest that grasses are adapted to herbivory, in perhaps a tightly linked process of coevolution (Coughenour, 1985).

Silica ($SiO_2.nH_2O$) is deposited in large quantities in plants, but is particularly abundant, diverse and distinctive in the Poaceae (Ellis, 1990; Gali-Muhtasib et al., 1992; Theunissen, 1994; Lu and Liu, 2003; Piperno, 2006). Silica bodies are thought to reduce palatability, digestibility and the nutritional value of the forage grasses (Coughenour, 1985; Ellis, 1990; Chapman, 1996b; Massey and Hartley, 2006). As

described by Chapman (1996b), the development of phytoliths (silica bodies in grass epidermal cells) and their persistence could be an adaptation to herbivore dentition changes that had evolved to improve their ability to cope with an increasingly grass-based diet (Massey and Hartley, 2006).

The effects of grazing on grasses have been well documented (Austin et al., 1981; Sala et al., 1986), but they have generally focused only on floristic composition, herbage production or changes in soil environment (Sala et al., 1986; Thurow, 1991). However, the responses of plants in terms of growth, and biomass allocation to long-term grazing remain unclear (Wang, 2004). It is known that herbivores have evolved on the world's grasslands (Chapman, 1996a). Thus, grass-herbivore relationships could be considered as coevolutionary processes. Applying the coevolution theory of Ehrlich and Raven (1964) to grasses and their grazers, one could consider that herbivores apply a selective pressure on grasses by grazing, so that grasses evolve by modifying their epidermal cell structure by adding silica bodies to increase leaf toughness. Concurrently, herbivores might have evolved tooth structures to cope with an increase in leaf toughness. Indeed, diffuse coevolution as in Ehrlich and Raven's model (1964) (i.e. evolution of plant lineages in simultaneous response to suites of herbivore species and vice versa) is thought to be an important process shaping the structure of plant diversity (Farrell and Mitter, 1998).

1.6 Aims of this thesis

The general aim of this thesis was to investigate the evolution of Poaceae and its diversification using a phylogenetic framework. One aspect was to infer the most comprehensive phylogenetic tree of the grasses and to assess infra-familial phylogenetic relationships with and without missing data (Chapter 2). Another aspect was to investigate patterns of diversification (Chapter 3). Chapter 3 also aimed to produce accurate dated trees of the grasses and study biogeographical patterns of diversification. The evolution of grasses in relation to herbivory was also investigated using phylogenetic methods (Chapter 4). A new methodological approach is described to reveal any parallel coevolutionary patterns between grasses and their grazers. The use of phylogenetic trees and their relevance to macro-evolutionary studies is

therefore discussed throughout this thesis. Specific objectives for each chapter are as follows:

Chapter 2:

- To generate and analyze large multi-gene region DNA sequence matrices that include a larger representation of grass family diversity than previous studies
- To investigate the effects of increased taxon sampling
- To study the impact of missing data on the phylogenetic trees

Chapter 3:

- To produce a comprehensive dated tree of the family
- To locate shifts in diversification in space and time
- To correlate shifts in diversification patterns with open versus closed habitat adaptations

Chapter 4:

- To quantify silica body density across grass lineages with a broad sampling of the family and to reconstruct possible pre-historical values using a phylogenetic approach
- To correlate silica density changes with open versus closed habitat adaptations
- To correlate increases in silica density with lophedness (i.e. number of shearing blades on 2nd upper molar) of ungulates

1.7 Structure of the thesis

Two papers taken from two different chapters have already been submitted to peer-reviewed journals. Chapter 2 is the basis of a paper submitted to *Molecular Phylogenetics and Evolution* (Bouchenak-Khelladi et al., submitted) with Nicolas Salamin, Vincent Savolainen, Felix Forest, Michelle van der Bank, Mark W.

Chase and Trevor R. Hodkinson as co-authors. Chapter 3 is the basis of a paper in preparation to *Evolution* (Bouchenak-Khelladi et al., in prep) with Trevor R. Hodkinson, Olivier Francois, Vincent Savolainen and Nicolas Salamin as co-authors. Finally, a paper from chapter 4 is in preparation. At the moment, no paper have been readily accepted or in press.

Also, I have contributed via collaboration with my supervisors on other related aspects of grass research, and more generally on species-rich groups of organisms. This have resulted in two additional publications as book sections: a study/review on grass diversification (Hodkinson et al., 2007a) and a paper applying supertrees to study grass diversification (Hodkinson et al. 2007b).

2) Large multi-gene phylogenetic trees of the grasses (Poaceae): impacts of taxon sampling and missing data

2.1 Introduction

Large and comprehensive phylogenetic trees are highly desirable for classification of organisms, and for studying macro-evolutionary processes (Barraclough and Nee, 2001). Theoretical studies have suggested that large phylogenetic trees can be easier to analyze than previously thought (Hillis, 1996b; Salamin et al., 2005), and empirical studies have also shown that large, combined, multi-gene analyses can correctly infer large phylogenetic trees (Soltis et al., 1999; Savolainen et al., 2000; Chase et al., 2006).

Grasses (Poaceae) are one of the most diverse families in the angiosperms, consisting of approximately 10,000 species and 600-700 genera (Clayton and Renvoize, 1986; Watson and Dallwitz, 1992). Understanding the evolution of such large group of organisms requires comprehensive and robust phylogenetic trees (Kellogg, 2000; Hodkinson et al., 2007a b). Although some advances in this research area have been reached (Grass Phylogeny Working Group (GPWG), 2001; Hodkinson et al., 2007a b), we are still far from a complete grass 'Tree of Life'. Grass classification began almost 200 years ago (Brown, 1810) with most subsequent classifications based largely on morphology and anatomy (Prat, 1932; Stebbins and Crampton, 1961; Clayton and Renvoize, 1986; Tzvelev, 1989; Renvoize and Clayton, 1992; Watson and Dallwitz, 1992). However, these classifications have been revised by studies based on additional molecular evidence. In the last two decades, molecular data have provided numerous and robust phylogenetic hypotheses at the family level (Doebley et al., 1990; Barker et al., 1995; Clark et al., 1995; Duvall and Morton, 1996; Soreng and Davis, 1998; Hilu et al., 1999; GPWG, 2001). The first molecular phylogenetic trees of the grasses using the *rbcL* plastid region (Hamby and Zimmer, 1988; Doebley et al., 1990) supported the monophyly of a group containing Panicoideae, Arundinoideae, Centothecoideae

and Chloridoideae (the PACC clade). However, only nine grass species (with two outgroups) from three recognized subfamilies were sequenced. A larger sample with a total of 47 species representing 26 tribes and six recognized subfamilies, was included by Clark et al. (1995) using the plastid gene *ndhF*. They recovered a tree with two major groups, the PACC and the BEP (containing Bambusoideae, Ehrhartoideae and Pooideae) clades and two early-diverging lineages (hereafter EDL) one containing *Anomochloa* Brong. and *Streptochaeta* Schrad. and the other *Pharus* P. Browne. The most significant combined data analysis consisted of DNA sequences (plastid and nuclear), plastid restriction site data and morphological data and included 61 genera, representing only 8% of all grass genera (GPWG, 2001). A relatively robust and well-resolved topology was obtained, supporting a PACCAD group (PACC plus Aristidoideae and Danthonioideae), a BEP group and three EDLs (recognized as Anomochlooideae, Pharoideae and Puelioideae, respectively).

The sample sizes of all previous phylogenetic analyses of the family, with the exception of supertree reconstructions (Salamin et al., 2002; Hodkinson et al., 2007b), ranged from 11 (Doebley et al., 1990) to more than 100 species (Hsiao et al., 1999). Most analyses with large sample sizes have concentrated on smaller taxonomic units than the whole grass family and have included only a small proportion of grass diversity (Petersen and Seberg, 1997; Duvall et al., 2001; Doust and Kellogg, 2002; Hodkinson et al., 2002; Mathews et al., 2002; Aliscioni et al., 2003).

Even though the major subfamilies of the grasses are well established, major questions remain to be resolved especially regarding the relationships within and between the subfamilies. For example, the phylogenetic relationships among major lineages within the PACCAD clade remain unclear and the placements of certain other taxa are not fully resolved such as *Eriachne* R.Br., *Gynerium* P.Beauv., *Micraira* F.Muell., and *Streptogyna* P.Beauv. (GPWG, 2001). The precise circumscriptions of Arundinoideae, Centothecoideae and Danthonioideae have not been determined and cannot be adequately assessed because of the limited sampling of genera (GPWG, 2001). The monophyly of Anomochlooideae may not have been assessed adequately because of long-branch attraction problems in the phylogenetic reconstructions (Felsenstein, 1984), which may have been responsible for the grouping of *Anomochloa* and *Streptochaeta* (GPWG, 2001).

The tribal inter-relationships of many of the grasses have not been sufficiently well addressed using phylogenetic trees because, as explained by the GPWG (2001), this requires an extensive sampling within each subfamily. Indeed, the monophyly of some previously recognized tribes has not been supported when more taxa were incorporated into phylogenetic analyses (Catalan et al., 1997). It is desirable to include a large number of representatives within tribes or subfamilies to adequately study their monophyly. No previous studies have concentrated specifically on generating large trees of the family with good sampling of tribes and their genera. Furthermore, none have compared such trees with those based on limited taxon sampling in order to check for the consistency of clades when more taxa are added.

For most groups of organism, only a few species have been sequenced for many genes and a few genes have been sequenced for many species. Consequently, a supermatrix approach that tries to gather most of the potential data available for phylogenetic purposes very often results in data sets containing large amounts of missing data (Sanderson and Driskell, 2003). As suggested by Wiens (2005), it may be possible to reap the benefits of increased taxon sampling without having data for all characters for all taxa, and thus increased taxon sampling might be obtained far more readily and cheaply.

Different approaches have been considered to resolve differences in phylogenetic estimates from different data sets (Huelsenbeck et al., 1996). It has been shown that a multi-gene approach often yields more accurate trees than a partitioned one, in particular with recent computational advances, that allow different substitution patterns between the genes considered to be incorporated in the phylogenetic reconstructions (Gadagkar et al., 2005). First proposed by Kluge (1989), the 'total evidence' approach states that all of the independent characters available for the set of species sampled should be combined because different data may interact positively to resolve a phylogenetic tree (Hillis, 1987). However, obvious or problematic heterogeneity across data partitions have often not been taken into account (for instance, between morphological and molecular data, or between different genes) because of computational complexity (Nylander et al., 2004). In response, the recent development of Bayesian inference (hereafter BI)

using Markov chain Monte Carlo (MCMC) has facilitated multi-gene analyses with among-partition heterogeneity (Nylander et al., 2004).

In order to combine multiple sequences for the same set of species with the widest possible sampling, we have sequenced *rbcL*, *matK* and *trnL-F* (*trnL* intron and *trnL-F* intergenic spacer) plastid DNA regions. It has been shown that the combined analyses of *rbcL* and *matK* generate more robust trees for monocotyledon angiosperms than those based on single gene analyses (Tamura et al., 2004). A number of grass *rbcL*, *matK* and *trnL-F* sequences have been published and/or desposited in GenBank/EMBL (http://www.ncbi.nih.gov/ and http://www.ebi.ac.uk, respectively), but the overlap between taxa is not optimized. Therefore, our sequencing effort was done in order to maximize the number of taxa for which the three DNA regions have been sequenced.

The aim of this study was to generate and analyze large multi-gene sequence matrices that include a larger representation of grass diversity than previous studies. This includes a thorough sampling of grasses (82% of tribes and 42% of genera). We discuss the effects of increased taxon sampling on the resolution and support of major clades, compare our results with previous phylogenetic studies at subfamily and tribal levels in the grasses using maximum parsimony (hereafter MP) and BI, and study the impact of missing data on our phylogenetic trees.

2.2 Material and methods

2.2.1 Taxon sampling

Total genomic DNA was extracted from silica-gel stored leaf material collected by T.R. Hodkinson at Trinity College Dublin, Ireland (TCD), from specimens found in the living collection at the Royal Botanic Gardens, Kew, England (Kew) and specimens from the herbarium at TCD (Appendix 2.1). DNA samples were also obtained from the DNA bank at Kew and from the DNA bank at TCD. We analyzed sequences of the *rbcL*, *matK* and *trnL-F* gene regions from 358 Poaceae species in 294 genera, 41 tribes and all 12 subfamilies (Appendix 2.1).

Subfamilial classification follows GPWG (2001) and tribal classification generally follows Watson and Dallwitz (1992), except for the tribal classification of Chloridoideae that follows Clayton and Renvoize (1986). For *rbcL*, we sequenced 61 taxa and downloaded a further 156 sequences from GenBank/EMBL (obtaining 217 taxa in total). For *matK*, we sequenced 94 taxa and downloaded 114 sequences (208 taxa in total), and for *trnL-F*, we sequenced 116 taxa and downloaded 41 sequences (157 taxa in total). In the combined analyses, sequence data from different species within the same genus were combined to create a 'conglomerate' sequence for analysis (see Appendix 2.1 for more details). This was necessary because of insufficient taxon overlap for such analyses. We also included two hybrids: *Cammophila* (*Ammophila* × *Calamagrostis*) and *Triticosecale* (*Triticum* × *Secale*). Two genera, *Elegia* and *Joinvillea,* were selected as outgroups (Appendix 2.1) because they are closely related to the grasses and in the case of *Joinvillea* (Joinvilleaceae) may represent its sister group. *Elegia* belongs to Restionaceae, a relatively large family that is clearly positioned in Poales (Doyle et al., 1992; Chase et al., 1993; Duvall et al., 1993; Bremer, 2002). Ecdeiocoleaceae has also been proposed as the sister group of the grasses (Rudall et al., 2005) but was not included in our analyses because of lack of sequence availability for the three DNA regions studied.

2.2.2 DNA extraction, amplification and sequencing

Between 0.1 and 0.5g of silica-gel or herbarium dried leaf (Chase and Hills, 1991) or up to 1g of fresh leaf (or seed) was used for DNA extraction. Total genomic DNA extracts were prepared following the CTAB method (Doyle and Doyle, 1987; Hodkinson *et al.,* 2007). For herbarium material the CTAB protocol was modified by precipitating the DNA with propan-2-ol instead of ethanol and then storing samples at –20°C for four weeks as long storage periods can increase DNA precipitation (Laslo Cziba, personal communication). All DNA extracts were purified by caesium chloride / ethidium bromide gradient centrifugation ($1.55 gml^{-1}$). The total genomic DNA extracts are held in the DNA Bank at the Jodrell Laboratory Kew (aliquots are available upon request; http://www.kew.org/data/dnaBank/homepage.htm) or the

DNA bank at TCD. For herbarium samples, concentrated DNA extracts were obtained by cleaning 100µl of dialyzed solutions through QIAquick™ Spin Columns (QIAGEN Ltd., UK) with a final elution volume of 50µl.

The three DNA regions (*rbcL*, *matK* and *trnL-F*) were amplified using an Applied Biosystems GeneAmp ® PCR System 9700 thermal cycler using the polymerase chain reaction (PCR). Each reaction was carried out using the specific set of primers for each gene (Table 2.1). PCR reaction volumes (50µl) included between 1 and 1.5µl of template DNA (with DNA concentrations mostly ranging from 400 to 1,200 $ng\mu l^{-1}$), 1µl of a 0.4% bovine serum albumin solution, 0.5µl of forward and reverse primers (100 $ng\mu l^{-1}$), 45µl of 1.1x ReddyMix™ PCR Master Mix (1.25 units Thermoprime Plus DNA Polymerase, 75mM Tris-HCl pH8.8, 2.5mM $MgCl_2$, 0.2mM for each dATP, CTP, GTP, TTP, 20mM $(NH_4)_2SO_4$ and between 1.5 and 2µl of sterile ultrapure water (MilliQ).

Cycle sequencing reactions were carried out in a GeneAmp PCR System 9700 thermal cycler using the Applied Biosystems BigDye Terminator v3.1. Cycle Sequencing Kit®. Various strategies were used for sequencing, the most successful being described in Table 2.1. After a series of cleaning with 250µl of 70% ethanol, samples were suspended into 10µl HiDi™ Formamide (Applied Biosystems) and run on an Applied Biosystems 3100 Automated DNA sequencer, and the sequences assembled using AutoAssembler 2.1 (Applied Biosystems).

Table 2.1 Primers used for amplification of rbcL, matK and trnL-F

	PRIMERS	SEQUENCES(5'-3')	SOURCE OR REFERENCE
rbcL	1Forward	ATG TCA CCA CAA ACA GAA ACT AAA GC	DoloresLledo *et al.*, 1998
	724Reverse	TCG CAT GTA CCY GCA GTT	
	627Forward	CAT TTA TGC GCT GGA GAG	
	1504Reverse	GAA TTA CTG ATT TCG CAA C	
matK	19Forward	CGT TCT GAC CAT ATT GCA	Molvray *et al.*, 2000
	9Reverse	GCT AGA ACT TTA GCT CGT A	Hilu *et al.*, 1999
	390Forward	CGA TCT ATT CAT TCA ATA	Cuenoud *et al.*, 2002
	trnK-2Reverse	AAC TAG TCG GAT GGA GTA G	Johnson and Soltis, 1994
TrnL-F	*trnL-F* c	CGA AAT CGG TAG ACG	Taberlet *et al.*, 1991
	trnL-F f	ATT TGA ACT GGT GAC ACG AG	Taberlet *et al.*, 1991

2.2.3 Phylogenetic analyses

Alignment of complete sequences of *rbcL* and *matK* was unambiguous and, thus, done manually. The alignment of *trnL-F* was done using Clustal W (Thompson et al., 1994) with subsequent manual adjustment; sections of ambiguous alignment were excluded from the analysis. MP and BI methods of phylogenetic inference were used as implemented in PAUP*4.0b10 (Swofford, 2002) and MrBayes 3.0b4 (Huelsenbeck and Ronquist, 2001), respectively. All analyses were done on the Trinity Centre for High Performance Computing Cluster (http://www.tchpc.tcd.ie).

2.2.3.1 Single-gene analyses

A separate phylogenetic analysis was performed for each data set (*rbcL*, *matK* and *trnL-F*). Heuristic MP searches with 10,000 replicates of random addition sequence and tree bisection reconnection (hereafter TBR) swapping were performed

for each data set, saving no more than 25 trees for each replicate. Support for clades was assessed using 1,000 bootstrap (Felsenstein, 1985a) replicates with simple addition sequence, TBR swapping, and saving no more than 50 trees for each replicate. Trees resulting from single-gene region analyses were compared to check the support and congruency of the major clades. Also the placement of species forming 'conglomerate' genera in the combined analyses was compared visually to the placement of these taxa in each of the three single-gene region analyses to check for consistency.

2.2.3.2 Combined analyses

The three single-region matrices were amalgamated into a combined matrix. However, before combining data sets, incongruence between the three separate analyses was assessed by comparing the results of the single-gene region analyses, on a node-by-node basis and specifically with respect to levels of bootstrap support following the approach taken by Sheahan and Chase (2000) and Reeves et al. (2001).

Two combined data matrices were subjected to phylogenetic analyses: one with perfect parallel sampling (hereafter DataSet I) and the other with missing data (hereafter DataSet II). DataSet I included 107 taxa sequenced for the three DNA regions (5,070 characters) with no missing sequences. It represents 15% of all grass genera, 51% of all tribes, and 10 subfamilies (Aristidoideae and Puelioideae were not sampled). DataSet II consisted of 294 taxa with missing sequences for either one or two DNA regions (Appendix 2.1). It represents 42% of all grass genera, 82% of tribes and all of the subfamilies. Heuristic MP analyses for the two combined data sets included 10,000 replicates of random sequence addition and TBR swapping, saving no more than 25 trees for each replicate. Robustness was assessed with the bootstrap (Felsenstein, 1985) using 1,000 replicates of simple addition sequence and TBR swapping with a limit of 50 trees for each replicate.

The substitution model used for the three different gene sequences was determined using a hierarchical likelihood ratio test framework as implemented in MODELTEST 3.06 (Posada and Crandall, 1998). The optimal models identified were HKY +Γ+I (Hasegawa et al., 1985) for the *rbcL* data, TVM +Γ+I (Posada and Crandall, 1998) for the *matK* data, and K81 +Γ+I (Kimura, 1981) for the *trnL-F* data.

The two combined matrices were analyzed using BI by partitioning the sequences by DNA region. This allowed independent estimation of parameters for each partition. Site-specific rates of substitution were allowed to vary across partitions (ratepr=variable). The HKY +Γ+I model was used for the *rbcL*, and the more general GTR +Γ+I model (Yang, 1994) was used for the *matK* and the *trnL-F* data. The *matK* and *trnL-F* sequences were analyzed using the GTR substitution model as neither the TVM nor the K81 models can be implemented in MrBayes 3.0b4. Four parallel Markov Chain Monte Carlo chains were run for 2,000,000 generations with trees sampled every 1,000 generations. Two independent analyses were performed to check whether convergence to the same posterior distribution was reached. The first 500 trees were discarded as burn-in. A one-tailed SH test (Shimodaira and Hasegawa, 1999) was performed using RELL bootstraps with 1,000 replicates to test if the Bayesian consensus tree was significantly different from the strict consensus parsimony tree.

2.3 Results

2.3.1 Single-gene analyses

Of the 1,405 sites included in the analysis of *rbcL*, 836 were constant and 283 were potentially parsimony informative. MP analysis of *rbcL* sequences resulted in 775 equally most parsimonious trees of 1,875 steps with a consistency index (CI) of 0.33 and a retention index (RI) of 0.80. The strict consensus tree is shown in Figure 2.1. Bootstrap percentages are indicated above the branches. The *rbcL* data set provides support for the monophyly of the PACCAD clade (73% Bootstrap Support, hereafter % BS), but does not support a BEP clade. Weak bootstrap support values (≥ 50% BS; moderate: ≥80 % BS; strong: ≥ 90% BS) were found for the major subfamilies. Subfamilies Anomochlooideae, Chloridoideae and Pooideae were weakly supported (54, 58 and 52% BS respectively). Aristidoideae and Danthonioideae were well supported with 99 and 84% BS respectively.

Arundinoideae, Bambusoideae, Centothecoideae, Ehrhartoideae and Panicoideae were not supported. However, the inclusion of Centothecoideae within Panicoideae was supported with 66% BS. Pharoideae was supported as an EDL but not strongly supported as sister to the core grasses (57% BS). Puelioideae were grouped with the core grasses (all subfamily lineages except the EDLs) in a polytomy including also Bambusoideae, Pooideae/Ehrhartoideae, and PACCAD.

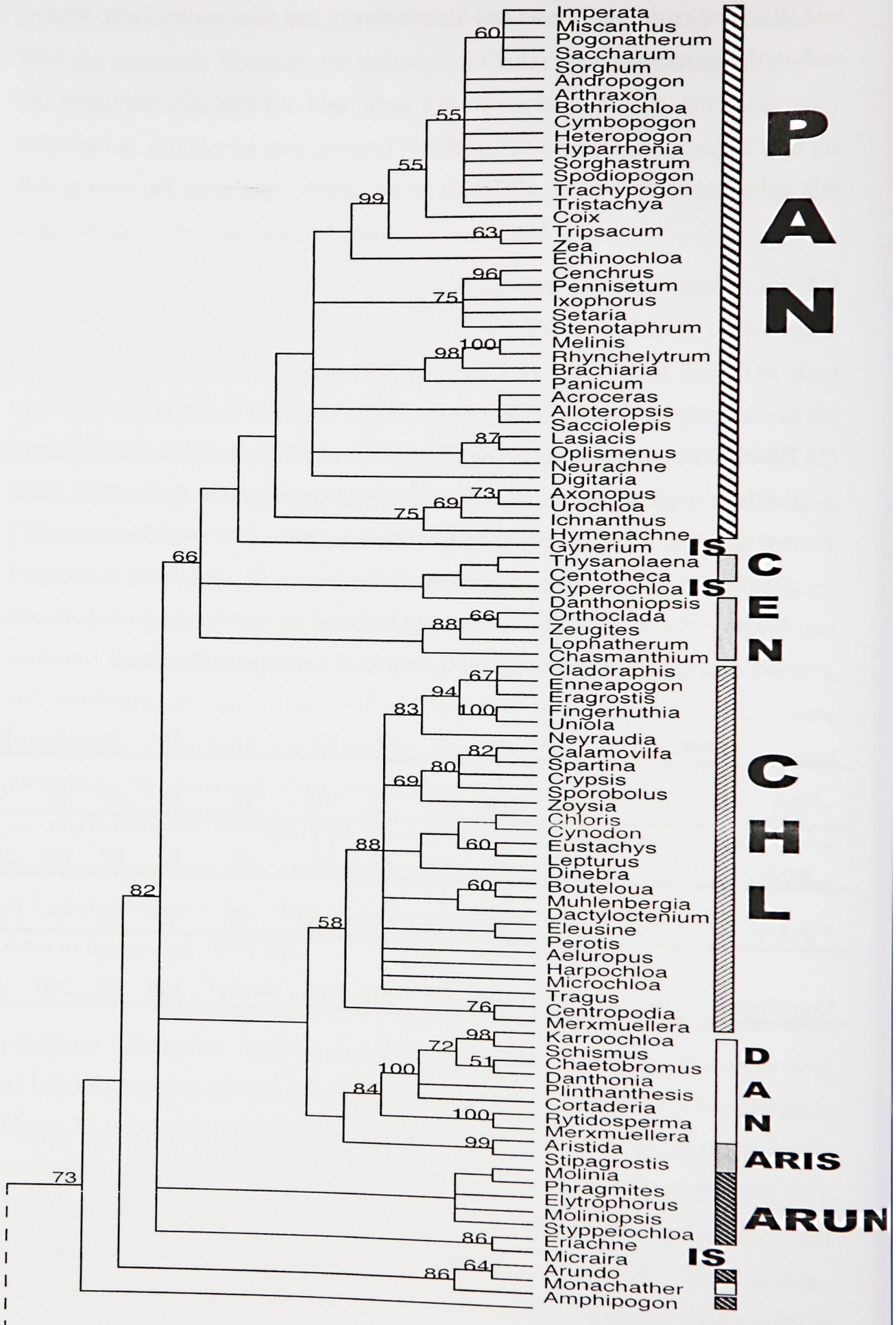
Of the 1,596 sites included in the *matK* analysis, 534 were constant and 742 were potentially parsimony informative. Analysis resulted in 129,975 equally most parsimonious trees of 4,210 steps with a CI of 0.40 and a RI of 0.80. The strict consensus tree is shown in Figure 2.2 with bootstrap support values given above the branches. The results showed strong support for the monophyly of the PACCAD clade (100% BS), but did not support a BEP clade. Bambusoideae (87% BS), Chloridoideae (83% BS), Danthonioideae (90% BS), Ehrhartoideae (100% BS), Panicoideae (78% BS) and Pooideae (94% BS) had well resolved internal structures. However, the internal structure of Chloridoideae was generally poorly resolved and contained a large polytomy. Danthonioideae were sister to Chloridoideae (68% BS), and Bambusoideae were sister to Pooideae (94% BS). Anomochlooideae were paraphyletic. Pharoideae were sister to the core grasses (95% BS) followed successively by *Anomochloa* and *Streptochaeta*. *Puelia* was not sampled for *matK*. Centothecoideae and Arundinoideae, sensu the GPWG (2001), were not supported by bootstrapping nor were they monophyletic on the strict consensus tree. The inclusion of Centothecoideae within Panicoideae was supported with 68% BS. *Gynerium* was sister to Panicoideae (89% BS).

Of the 1,075 included characters of *trnL-F*, 488 were constant and 369 were parsimony informative. Analysis resulted in 96,675 equally most parsimonious trees of 1,693 steps with a CI of 0.54 and a RI of 0.81. The strict consensus tree shown in Figure 2.3 is generally well resolved. The *trnL-F* data provides strong support for the monophyly of PACCAD (99% BS), and weak support for the BEP clade (67% BS). The individual subfamilies Danthonioideae, Ehrhartoideae, Panicoideae and Pooideae were supported (95, 74, 69 and 89% BS respectively). Arundinoideae, Bambusoideae and Centothecoideae were not supported. However, Bambusoideae were present in all equally most parsimonious trees. The inclusion of Centothecoideae within Panicoideae was moderately supported (79% BS). Two

EDLs were retrieved, *Pharus* and *Streptochaeta,* that were successively sister to the rest of the grasses.

*Table 2.2 Bootstrap supports (%) of PACCAD and BEP clades, and subfamilies for the three single-gene analyses with Maximum Parsimony. Pa: Panicoideae, Aris: Aristidoideae, Ch: Chloridoideae, Ce: Centothecoideae, Ar: Arundinoideae, D: Danthonioideae, B: Bambusoideae, E: Ehrhartoideae, Po: Pooideae, Ano: Anomoochloideae, Ph: Pharoideae, Pu: Puelioideae. NS: Not Supported, Ab: Absent (Not sampled), —: Only one representative, *: values from DataSet II.*

	PACCAD	BEP	Pa	Ar	Ch	Ce	Ar	D	B	E	Po	Ano	Ph	P
rbcL	73	NS	NS	99	58	NS	NS	84	NS	NS	52	54	—	N
matK	100	NS	78	—	83	NS	NS	90	87	100	94	NS	—	A
trnL-F	99	67	69	Ab	NS	NS	NS	95	NS	74	89	—	—	A
Combined	100	89	100	98*	100	100	66/NS*	100	98	100	100	—	—	-


Imperata
Miscanthus
Pogonatherum
Saccharum
Sorghum
Andropogon
Arthraxon
Bothriochloa
Cymbopogon
Heteropogon
Hyparrhenia
Sorghastrum
Spodiopogon
Trachypogon
Tristachya
Coix
Tripsacum
Zea
Echinochloa
Cenchrus
Pennisetum
Ixophorus
Setaria
Stenotaphrum
Melinis
Rhynchelytrum
Brachiaria
Panicum
Acroceras
Alloteropsis
Sacciolepis
Lasiacis
Oplismenus
Neurachne
Digitaria
Axonopus
Urochloa
Ichnanthus
Hymenachne
Gynerium
Thysanolaena
Centotheca
Cyperochloa
Danthoniopsis
Orthoclada
Zeugites
Lophatherum
Chasmanthium
Cladoraphis
Enneapogon
Eragrostis
Fingerhuthia
Uniola
Neyraudia
Calamovilfa
Spartina
Crypsis
Sporobolus
Zoysia
Chloris
Cynodon
Eustachys
Lepturus
Dinebra
Bouteloua
Muhlenbergia
Dactyloctenium
Eleusine
Perotis
Aeluropus
Harpochloa
Microchloa
Tragus
Centropodia
Merxmuellera
Karroochloa
Schismus
Chaetobromus
Danthonia
Plinthanthesis
Cortaderia
Rytidosperma
Merxmuellera
Aristida
Stipagrostis
Molinia
Phragmites
Elytrophorus
Moliniopsis
Styppeiochloa
Eriachne
Micraira
Arundo
Monachather
Amphipogon
PAN
CEN
CHL
DAN
ARIS
ARUN

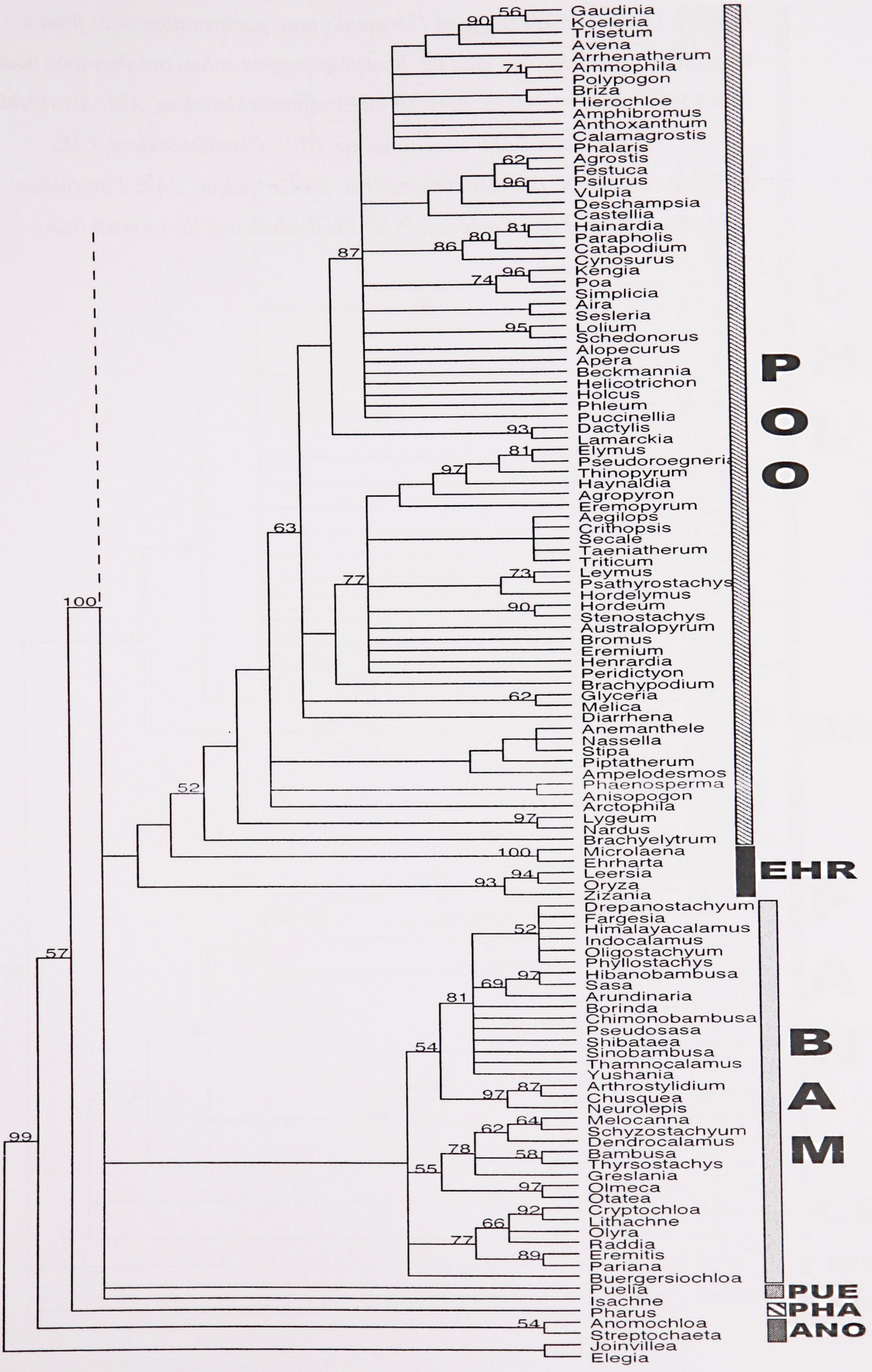
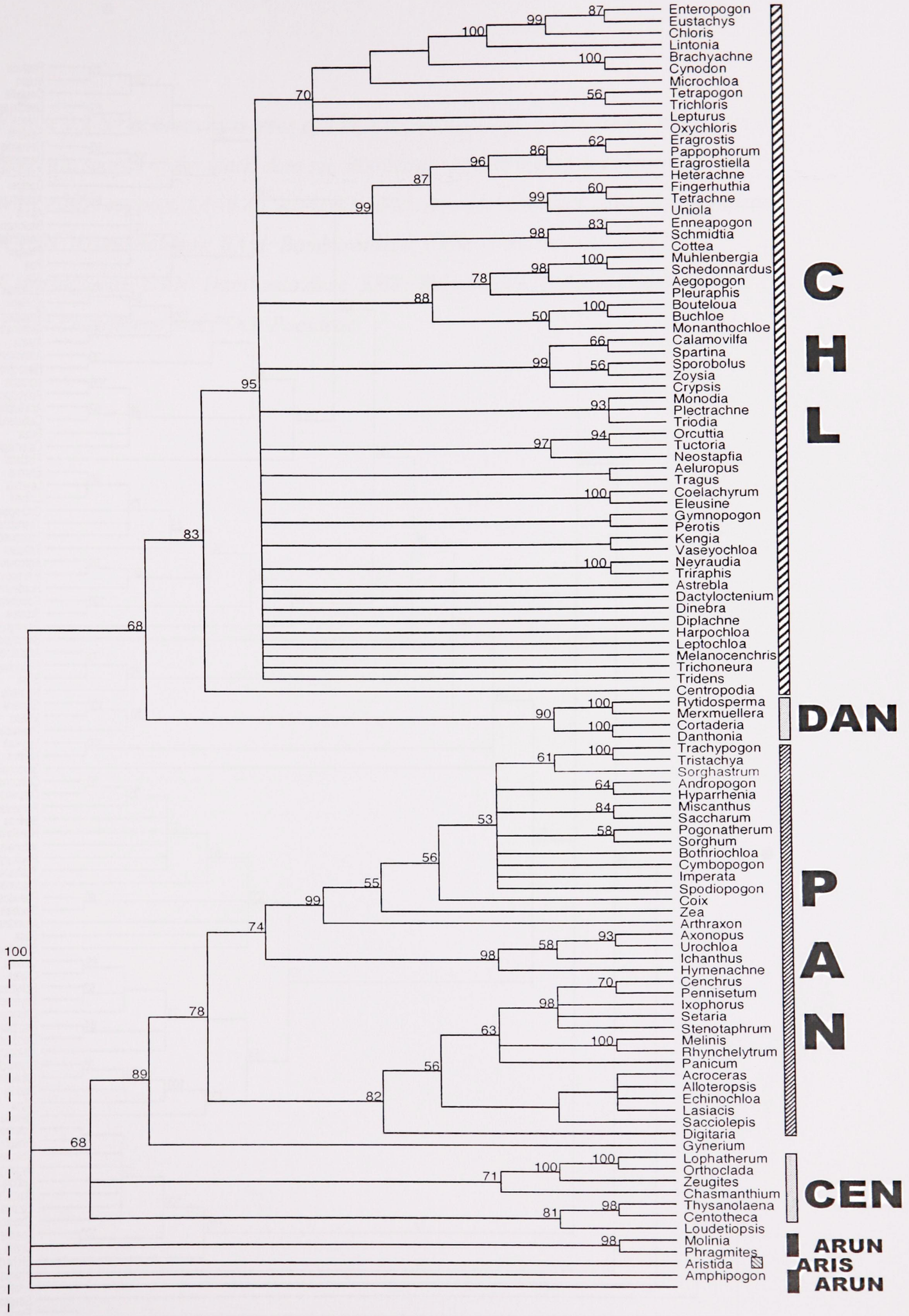
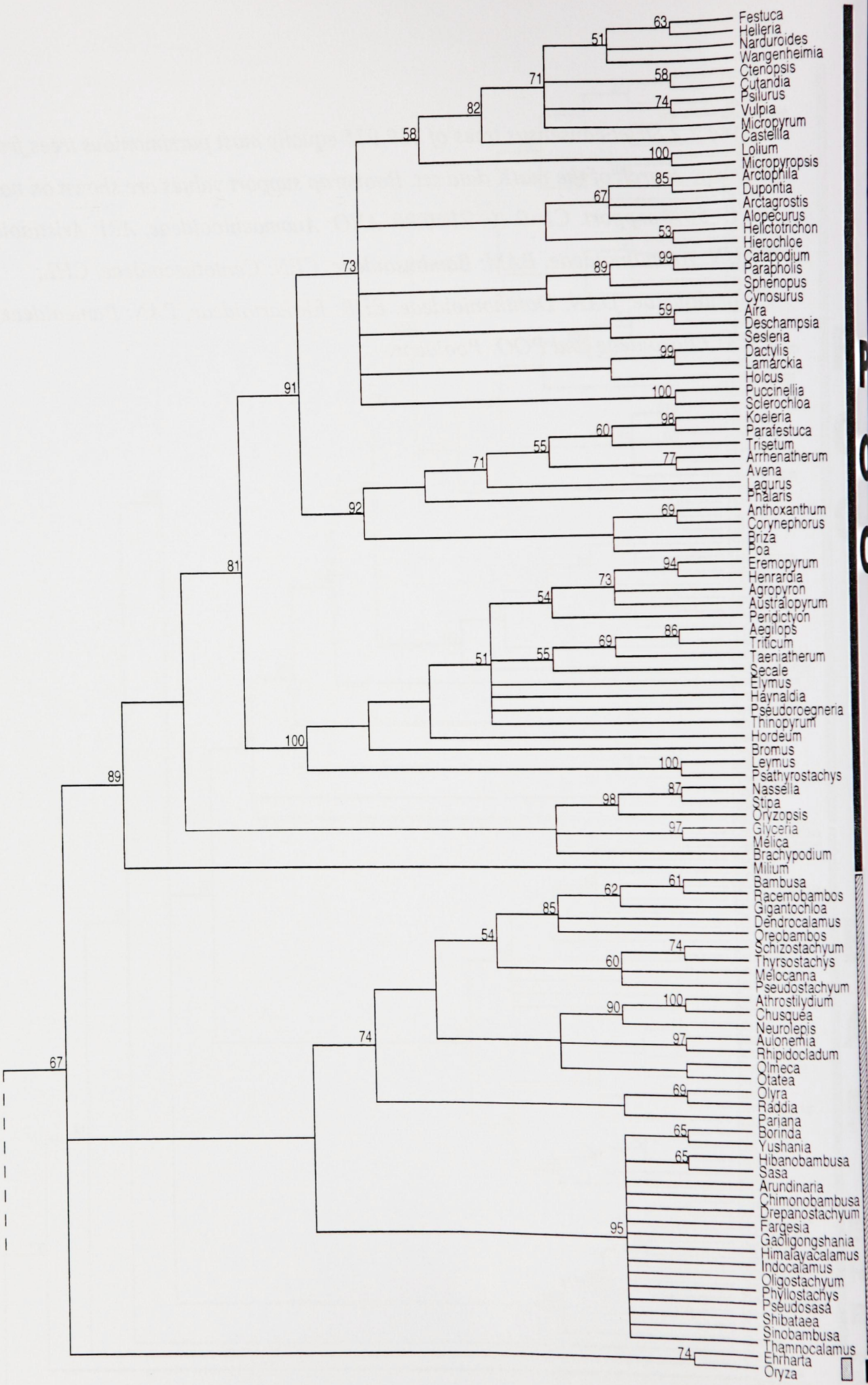

Gaudinia
Koeleria
Trisetum
Avena
Arrhenatherum
Ammophila
Polypogon
Briza
Hierochloe
Amphibromus
Anthoxanthum
Calamagrostis
Phalaris
Agrostis
Festuca
Psilurus
Vulpia
Deschampsia
Castellia
Hainardia
Parapholis
Catapodium
Cynosurus
Kengia
Poa
Simplicia
Aira
Sesleria
Lolium
Schedonorus
Alopecurus
Apera
Beckmannia
Helicotrichon
Holcus
Phleum
Puccinellia
Dactylis
Lamarckia
Elymus
Pseudoroegneria
Thinopyrum
Haynaldia
Agropyron
Eremopyrum
Aegilops
Crithopsis
Secale
Taeniatherum
Triticum
Leymus
Psathyrostachys
Hordelymus
Hordeum
Stenostachys
Australopyrum
Bromus
Eremium
Henrardia
Peridictyon
Brachypodium
Glyceria
Melica
Diarrhena
Anemanthele
Nassella
Stipa
Piptatherum
Ampelodesmos
Phaenosperma
Anisopogon
Arctophila
Lygeum
Nardus
Brachyelytrum
Microlaena
Ehrharta
Leersia
Oryza
Zizania
Drepanostachyum
Fargesia
Himalayacalamus
Indocalamus
Oligostachyum
Phyllostachys
Hibanobambusa
Sasa
Arundinaria
Borinda
Chimonobambusa
Pseudosasa
Shibataea
Sinobambusa
Thamnocalamus
Yushania
Arthrostylidium
Chusquea
Neurolepis
Melocanna
Schyzostachyum
Dendrocalamus
Bambusa
Thyrsostachys
Greslania
Olmeca
Otatea
Cryptochloa
Lithachne
Olyra
Raddia
Eremitis
Pariana
Buergersiochloa
Puelia
Isachne
Pharus
Anomochloa
Streptochaeta
Joinvillea
Elegia
POO
EHR
BAM
PUE
PHA
ANO

Figure 2.1 Strict consensus tree of 775 equally most parsimonious trees from a heuristic search of the rbcL data set. Bootstrap support values are shown on nodes with >50% support. CI=0.33, RI=0.80. ANO: Anomochlooideae, ARI: Aristidoideae, ARUN: Arundinoideae, BAM: Bambusoideae, CEN: Centothecoideae, CHL: Chloridoideae, DAN: Danthonioideae, EHR: Ehrhartoideae, PAN: Panicoideae, PHA: Pharoideae, POO: Pooideae, PUE: Puelioideae and IS: Incertae Sedis.


Enteropogon
Eustachys
Chloris
Lintonia
Brachyachne
Cynodon
Microchloa
Tetrapogon
Trichloris
Lepturus
Oxychloris
Eragrostis
Pappophorum
Eragrostiella
Heterachne
Fingerhuthia
Tetrachne
Uniola
Enneapogon
Schmidtia
Cottea
Muhlenbergia
Schedonnardus
Aegopogon
Pleuraphis
Bouteloua
Buchloe
Monanthochloe
Calamovilfa
Spartina
Sporobolus
Zoysia
Crypsis
Monodia
Plectrachne
Triodia
Orcuttia
Tuctoria
Neostapfia
Aeluropus
Tragus
Coelachyrum
Eleusine
Gymnopogon
Perotis
Kengia
Vaseyochloa
Neyraudia
Triraphis
Astrebla
Dactyloctenium
Dinebra
Diplachne
Harpochloa
Leptochloa
Melanocenchris
Trichoneura
Tridens
Centropodia
Rytidosperma
Merxmuellera
Cortaderia
Danthonia
Trachypogon
Tristachya
Sorghastrum
Andropogon
Hyparrhenia
Miscanthus
Saccharum
Pogonatherum
Sorghum
Bothriochloa
Cymbopogon
Imperata
Spodiopogon
Coix
Zea
Arthraxon
Axonopus
Urochloa
Ichanthus
Hymenachne
Cenchrus
Pennisetum
Ixophorus
Setaria
Stenotaphrum
Melinis
Rhynchelytrum
Panicum
Acroceras
Alloteropsis
Echinochloa
Lasiacis
Sacciolepis
Digitaria
Gynerium
Lophatherum
Orthoclada
Zeugites
Chasmanthium
Thysanolaena
Centotheca
Loudetiopsis
Molinia
Phragmites
Aristida
Amphipogon
CHL
DAN
PAN
CEN
ARUN
ARIS
ARUN


Festuca
Vulpia
Castellia
Deschampsia
Holcus
Sesleria
Hainardia
Parapholis
Catapodium
Dactylis
Alopecurus
Apera
Beckmannia
Helicotrichon
Phleum
Poa
Aira
Gaudinia
Koeleria
Trisetum
Arrhenatherum
Avena
Anthoxanthum
Hierochloe
Agrostis
Polypogon
Ammophila
Calammophila
Briza
Calamagrostis
Phalaris
Elymus
Leymus
Lolium
Crithopsis
Taeniatherum
Triticosecale
Triticum
Agropyron
Eremopyrum
Hordeum
Boissiera
Bromus
Psathyrostachys
Brachypodium
Ampelodesmos
Piptochaetium
Stipa
Nassella
Glyceria
Melica
Phaenosperma
Lygeum
Nardus
Brachyelytrum
Chimonobambusa
Drepanostachyum
Fargesia
Borinda
Himalayacalamus
Indocalamus
Phyllostachys
Yushania
Thamnocalamus
Pseudosasa
Sinobambusa
Arundinaria
Sasa
Shibataea
Olmeca
Otatea
Schyzostachyum
Thyrsostachys
Chusquea
Bambusa
Dendrocalamus
Melocanna
Cryptochloa
Lithachne
Raddia
Olyra
Pariana
Potamophila
Prosphytochloa
Chikusichloa
Hygroryza
Luziola
Rhynchoryza
Zizania
Zizaniopsis
Leersia
Oryza
Ehrharta
Pharus
Anomochloa
Streptochaeta
Elegia
Joivillea
POO
BAM
EHR
PHA
ANO

Figure 2.2 Strict consensus trees of 129,975 equally most parsimonious trees from a heuristic search of the matK data set. Bootstrap support values are shown on nodes with >50% support. CI=0.40, RI=0.80. ANO: Anomochlooideae, ARI: Aristidoideae, ARUN: Arundinoideae, BAM: Bambusoideae, CEN: Centothecoideae, CHL: Chloridoideae, DAN: Danthonioideae, EHR: Ehrhartoideae, PAN: Panicoideae, PHA: Pharoideae and POO: Pooideae.


Festuca
Helleria
Narduroides
Wangenheimia
Ctenopsis
Cutandia
Psilurus
Vulpia
Micropyrum
Castellia
Lolium
Micropyropsis
Arctophila
Dupontia
Arctagrostis
Alopecurus
Hellctotrichon
Hierochloe
Catapodium
Parapholis
Sphenopus
Cynosurus
Aira
Deschampsia
Sesleria
Dactylis
Lamarckia
Holcus
Puccinellia
Sclerochloa
Koeleria
Parafestuca
Trisetum
Arrhenatherum
Avena
Lagurus
Phalaris
Anthoxanthum
Corynephorus
Briza
Poa
Eremopyrum
Henrardia
Agropyron
Australopyrum
Peridictyon
Aegilops
Triticum
Taeniatherum
Secale
Elymus
Haynaldia
Pseudoroegneria
Thinopyrum
Hordeum
Bromus
Leymus
Psathyrostachys
Nassella
Stipa
Oryzopsis
Glyceria
Melica
Brachypodium
Milium
Bambusa
Racemobambos
Gigantochloa
Dendrocalamus
Oreobambos
Schizostachyum
Thyrsostachys
Melocanna
Pseudostachyum
Athrostilydium
Chusquea
Neurolepis
Aulonemia
Rhipidocladum
Olmeca
Otatea
Olyra
Raddia
Pariana
Borinda
Yushania
Hibanobambusa
Sasa
Arundinaria
Chimonobambusa
Drepanostachyum
Fargesia
Gaoligongshania
Himalayacalamus
Indocalamus
Oligostachyum
Phyllostachys
Pseudosasa
Shibataea
Sinobambusa
Thamnocalamus
Ehrharta
Oryza

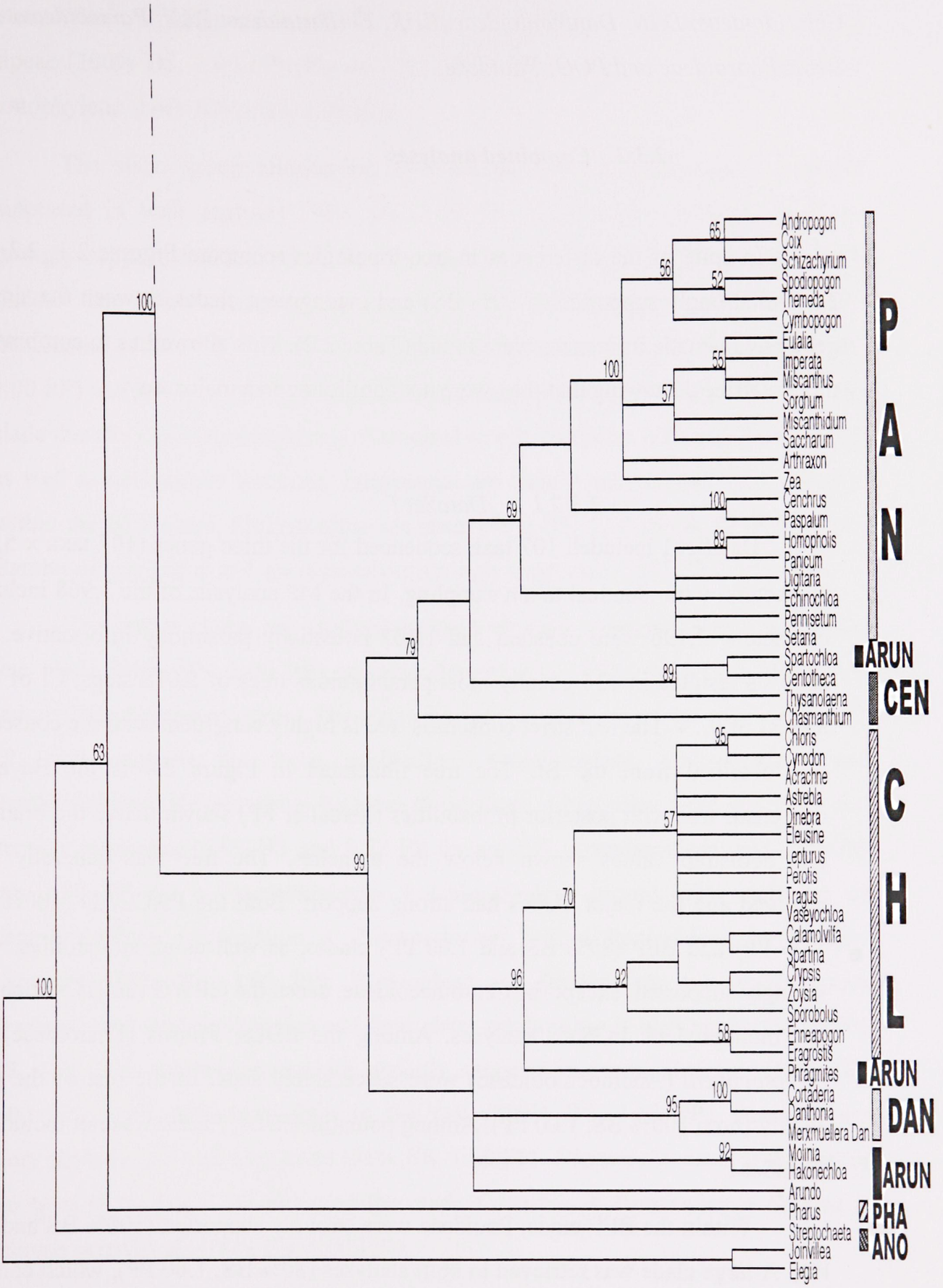


Figure 2.3 Strict consensus trees of 96,675 equally most parsimonious trees from a heuristic search of the trnL-F data set. Bootstrap support values are shown on nodes with >50% support. CI=0.54, RI=0.81. ANO: Anomochlooideae, ARI: Aristidoideae, ARUN: Arundinoideae, BAM: Bambusoideae, CEN: Centothecoideae, CHL:

Chloridoideae, DAN: Danthonioideae, EHR: Ehrhartoideae, PAN: Panicoideae, PHA: Pharoideae and POO: Pooideae.

2.3.2 Combined analyses

In spite of the differences in tree topologies (compare Figures 2.1, 2.2 and 2.3), no strongly supported (> 90% BS) and incongruent clades between the single-gene phylogenetic inferences were found (Table 2.2). This allowed us to combine all three matrices, knowing that they were not conflicting in a major way.

2.3.2.1 DataSet I

DataSet I included 107 taxa sequenced for the three genes (107 taxa × 5,070 characters) with identical taxon sampling. In the MP analysis, of the 3,968 included characters, 2,190 were constant and 1,107 potentially parsimony informative. MP analysis resulted in 134 equally most parsimonious trees of 5,018 steps, CI of 0.47 and RI of 0.74. The MP strict consensus tree is highly congruent with the consensus tree obtained from the BI. The tree illustrated in Figure 2.4 is the Bayesian consensus tree with posterior probabilities (hereafter PP) shown above the branches and bootstrap values shown below the branches. The tree was generally well resolved and the major clades had strong support. Both the PACCAD (100% BS, 1.00 PP) and BEP (89% BS and 1.00 PP) clades, as well as all subfamilies, were strongly supported, except for Centothecoideae, sensu the GPWG (2001), which was not monophyletic in both analyses. Among the EDLs, *Pharus* (Pharoideae) and *Streptochaeta* (Anomochlooideae) were successively sister to the rest of the grass family (both 100% BS, 1.00 PP). Among potential EDLs, *Puelia* was not included in DataSet I.

Within the BEP clade, Pooideae were strongly supported (100% BS and 1.00 PP). A large clade was retrieved in both analyses (80% BS, 1.00 PP), which contains tribes Aveneae, Bromeae, Poeae and Triticeae (Figure 2.4). None of these tribes were found to be monophyletic in both analyses. However, there was a sister relationship (80% BS, 1.00 PP) between a Triticeae + Bromeae + *Lolium* (81% BS, 1.00 PP), and an Aveneae + Poeae clade (96% BS, 1.00 PP). Three lineages were successively

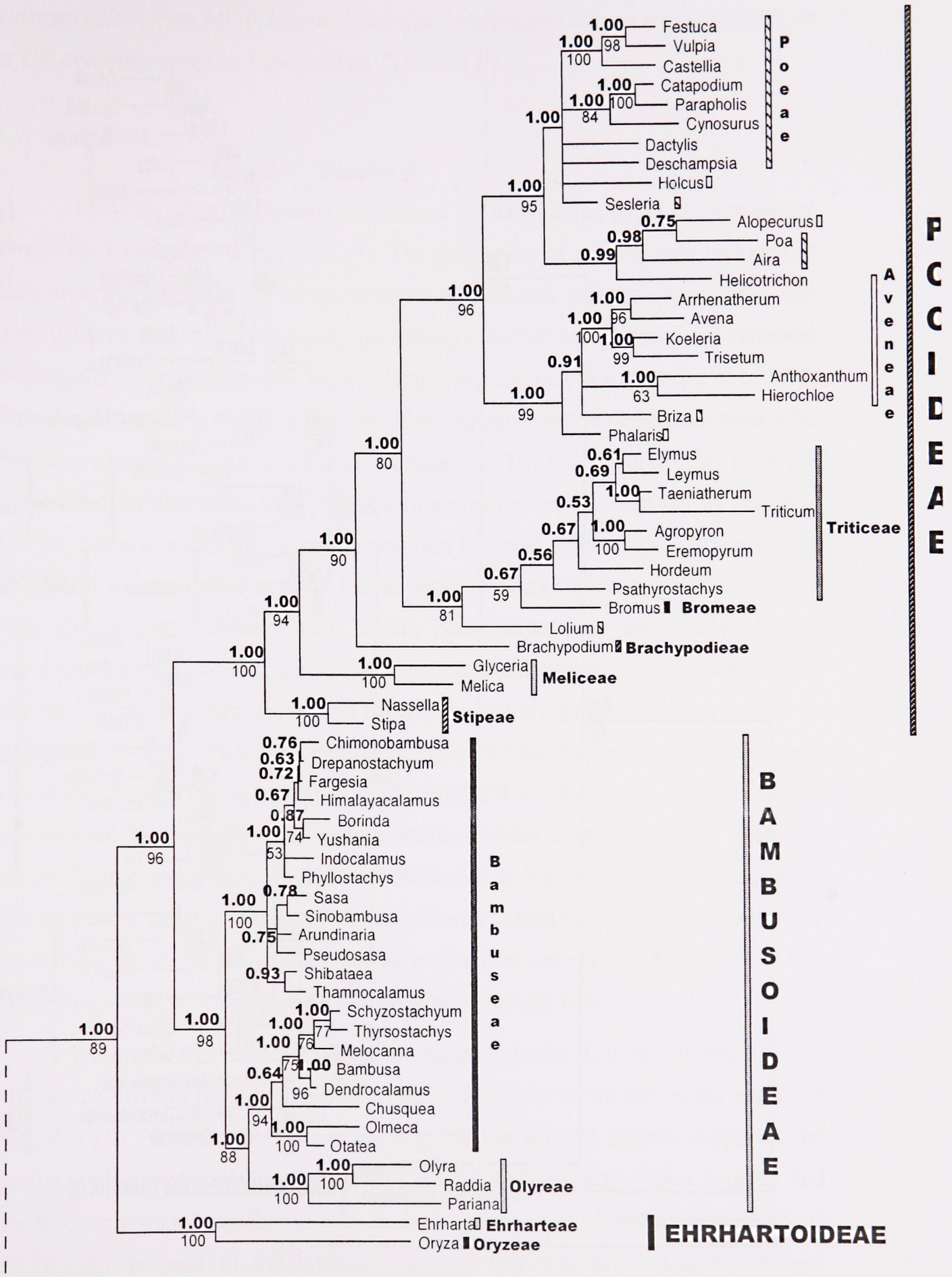
sister to these, Brachypodieae (90% BS, 1.00 PP), Meliceae (94% BS, 1.00 PP) and Stipeae (100% BS, 1.00 PP) (Figure 2.4). Meliceae and Stipeae were supported as monophyletic (both 100% BS, 1.00 PP).

The sister group relationship of Pooideae and Bambusoideae is strongly supported in both analyses (96% BS, 1.00 PP; Figure 2.4). Bambusoideae are strongly supported (98% BS, 1.00 PP), and are divided into two well-supported clades. One clade contains representatives of the temperate bamboos, tribe Bambuseae (100% BS, 1.00 PP). The other contains tropical Bambuseae (94% BS, 1.00 PP) and Olyreae (100% BS, 1.00 PP) as sister groups (88% BS, 1.00 PP). This clade therefore has representatives of tropical woody bamboos (Old and New World) as well as herbaceous bamboos. Bambuseae are clearly not monophyletic. Finally, within the BEP clade, Ehrhartoideae are sister (89% BS, 1.00 PP) to the Pooideae + Bambusoideae group and are themselves strongly supported (100% BS, 1.00 PP).

Two major clades are resolved within a strongly supported PACCAD clade. One clade (98% BS, 1.00 PP) contains Panicoideae (100% BS, 1.00 PP) and a paraphyletic Centothecoideae. The other clade is weakly supported (63% BS, 0.72 PP) and contains the three subfamilies Arundinoideae, Chloridoideae and Danthonioideae. Relationships between these three subfamilies were generally not strongly supported (63% BS and 0.72 PP for an EDL Arundinoideae; and 78% BS and 1.00 PP for a Danthonioideae + Chloridoideae group). The inclusion of Centothecoideae (*Centotheca* + *Thysanolaena*) within Panicoideae is strongly supported (98% BS, 1.00 PP). Panicoideae are divided into two major tribes (Paniceae and Andropogoneae) and their sister relationship is well supported (100% BS, 1.00 PP). Andropogoneae are strongly supported as monophyletic (100% BS, 1.00 PP), with *Arthraxon* sister to the rest (100% BS, 1.00 PP). Paniceae are monophyletic and well supported (81% BS, 1.00 PP). *Digitaria* is sister to the rest of the tribe (81% BS, 1.00 PP), and the clade *Cenchrus* + *Pennisetum* + *Setaria* is strongly supported (100% BS, 1.00 PP).

Arundinoideae are monophyletic and strongly supported with BI analysis (1.00 PP) but they are only weakly supported with MP (66% BS). However, the grouping of *Molinia* and *Phragmites* is strongly supported in both analyses (100% BS, 1.00 PP). Chloridoideae and Danthonioideae are sister groups (78% BS, 1.00 PP)

and both are strongly supported (100% BS, 1.00 PP). Tribes within Chloridoideae are polyphyletic. Pappophoreae and *Eragrostis* are sister to the rest of Chloridoideae (100% BS, 1.00 PP). The remaining taxa tend to form a clade with an assemblage between two sister clades (77% BS, 0.99 PP), one containing Chlorideae + Cynodonteae + Eragrostideae + Leptureae (95% BS, 1.00 PP) and the other containing only two tribes Cynodonteae + Eragrostideae (100% BS, 1.00 PP).


Festuca
Vulpia
Castellia
Catapodium
Parapholis
Cynosurus
Dactylis
Deschampsia
Holcus
Sesleria
Alopecurus
Poa
Aira
Helicotrichon
Arrhenatherum
Avena
Koeleria
Trisetum
Anthoxanthum
Hierochloe
Briza
Phalaris
Elymus
Leymus
Taeniatherum
Triticum
Agropyron
Eremopyrum
Hordeum
Psathyrostachys
Bromus
Lolium
Brachypodium
Glyceria
Melica
Nassella
Stipa
Chimonobambusa
Drepanostachyum
Fargesia
Himalayacalamus
Borinda
Yushania
Indocalamus
Phyllostachys
Sasa
Sinobambusa
Arundinaria
Pseudosasa
Shibataea
Thamnocalamus
Schyzostachyum
Thyrsostachys
Melocanna
Bambusa
Dendrocalamus
Chusquea
Olmeca
Otatea
Olyra
Raddia
Pariana
Ehrharta
Oryza
Poeae
Aveneae
Triticeae
Bromeae
Brachypodieae
Meliceae
Stipeae
Bambuseae
Olyreae
Ehrharteae
Oryzeae
POCIDEAE
BAMBUSOIDEAE
EHRHARTOIDEAE

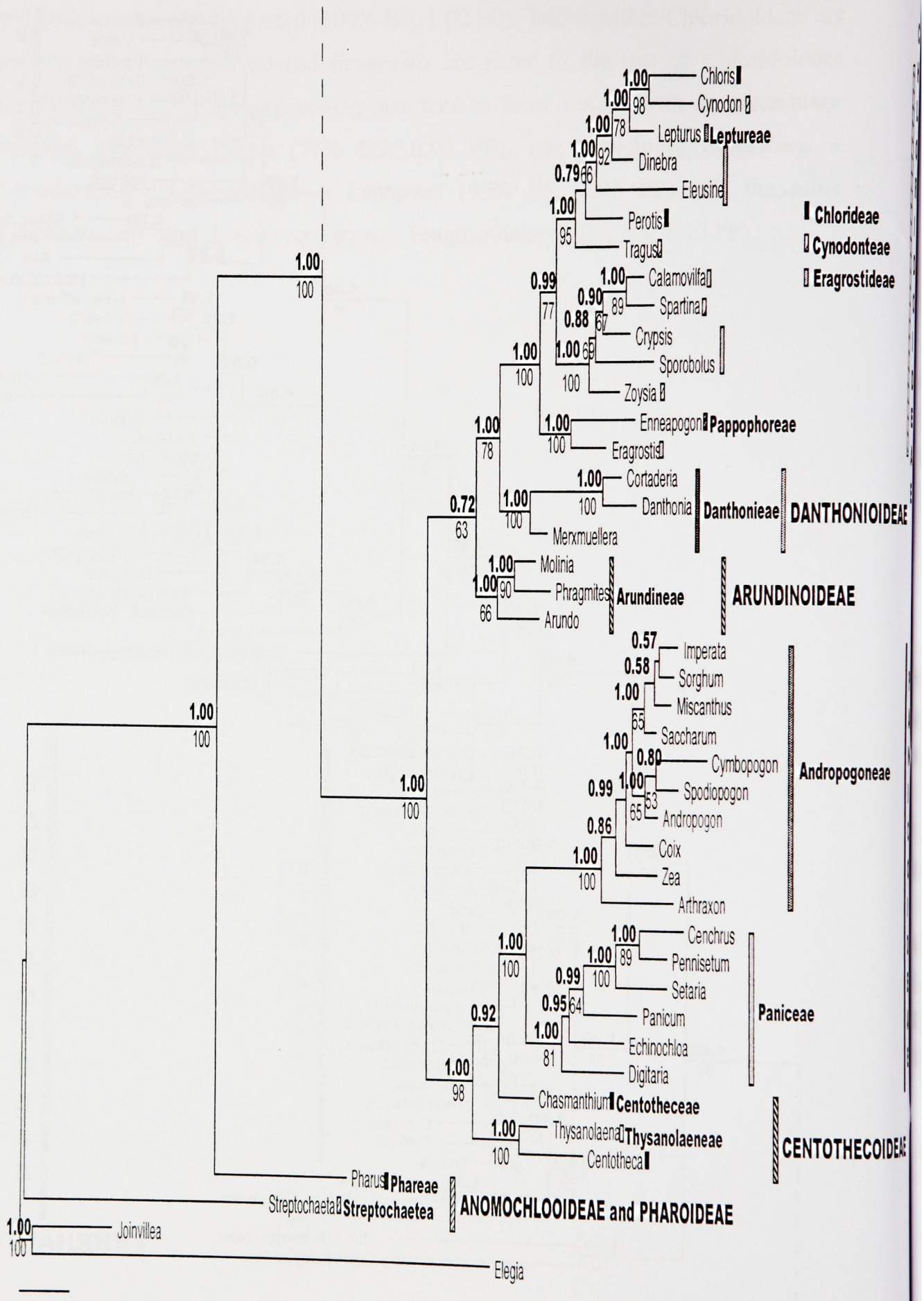


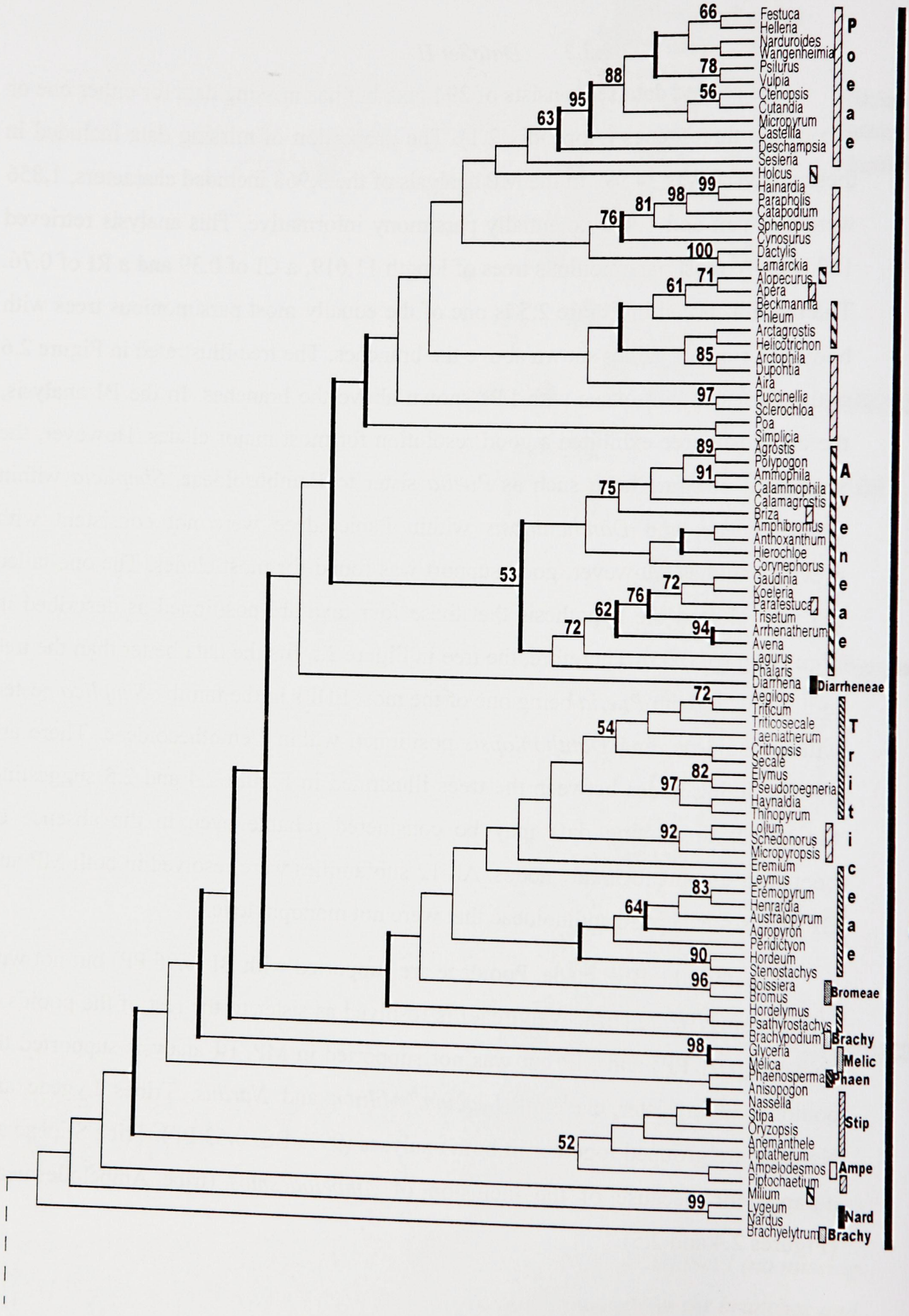
Figure 2.4 Bayesian consensus tree of combined data sets from DataSet I (no missing data) where posterior probabilities for each clade are shown above the branches and

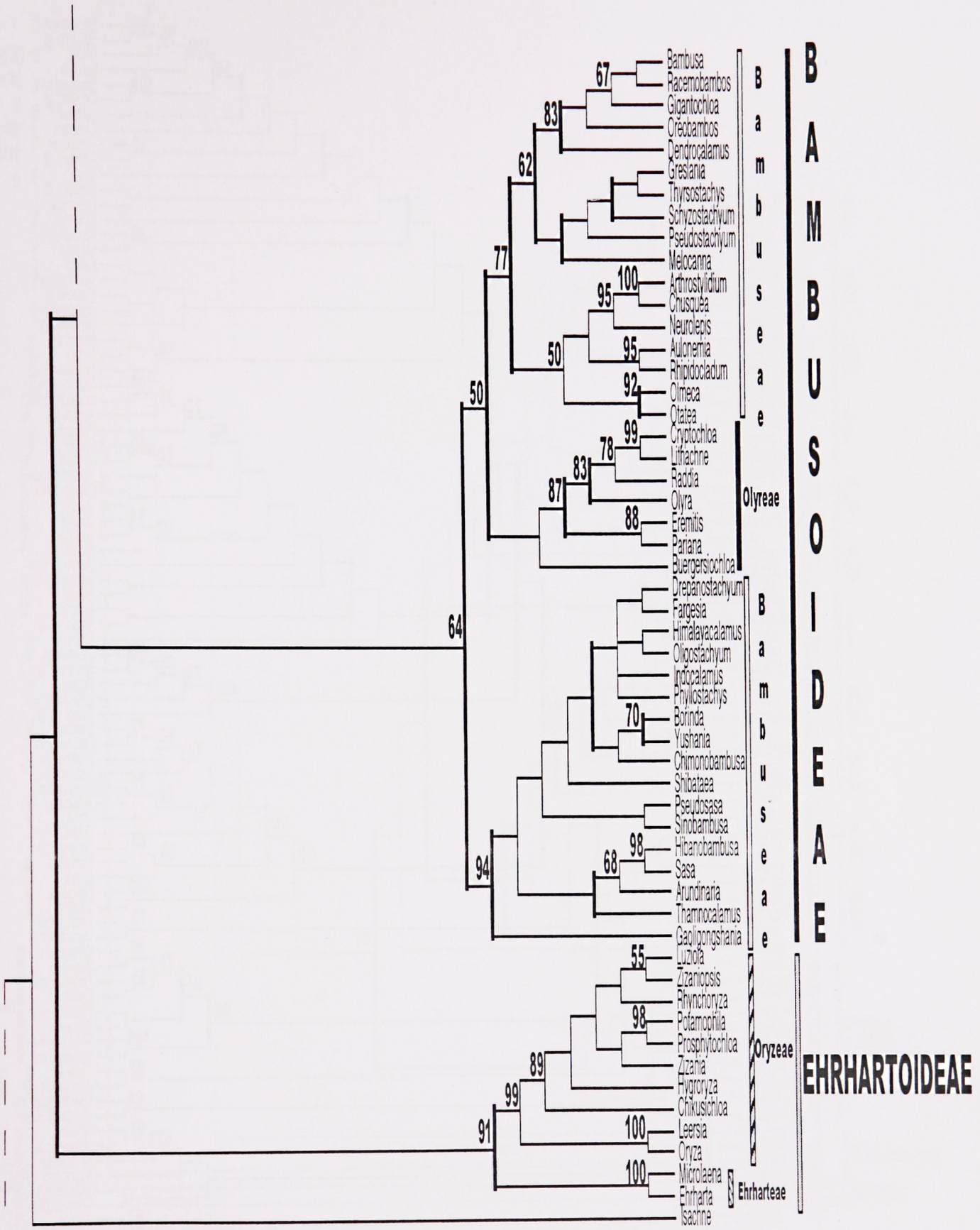
bootstrap values from MP below the branches. (Molecular data matrix is available on the CD accompanying this thesis; folder "DataSet I")

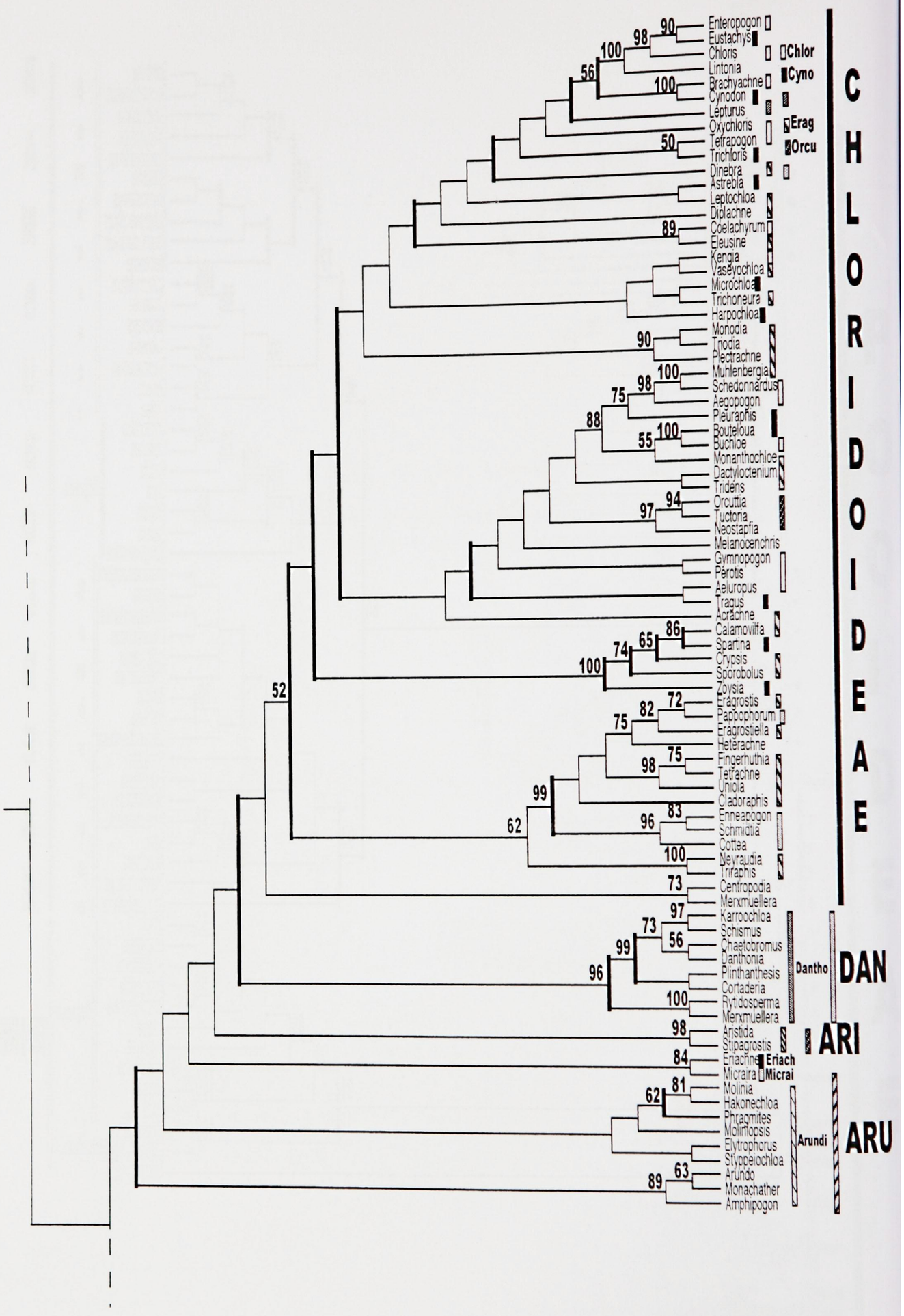
2.3.2.2 DataSet II

The second data set consists of 294 taxa but has missing data for either one or two of the three genes (Appendix 2.1). The proportion of missing data included in these analyses was 34 %. In the MP analysis of the 3,968 included characters, 1,856 were constant and 1,403 potentially parsimony informative. This analysis retrieved 100 equally most parsimonious trees of length 11,019, a CI of 0.39 and a RI of 0.76. The tree illustrated in Figure 2.5 is one of the equally most parsimonious trees with bootstrap support values shown above the branches. The tree illustrated in Figure 2.6 is the Bayesian consensus with PPs shown above the branches. In the BI analysis, the consensus tree exhibited a good resolution for most major clades. However, the placement of some taxa, such as *Puelia* sister to Bambusoideae, *Simplicia* within Chloridoideae, and *Danthoniopsis* within Panicoideae were not consistent with previous studies. However, good support was found for most clades. The one-tailed SH test rejected the hypothesis that these four taxa are positioned as described in Figure 2.6 ($p<0.05$). Therefore, the tree in Figure 2.5 fits the data better than the tree in Figure 2.6, with *Puelia* being one of the most EDLs in the family, *Simplica* nested within Pooideae, and *Danthoniopsis* positioned within Centothecoideae. There are no conflicting nodes between the trees illustrated in Figure 2.4 and 2.5 suggesting the tree with missing data may be considered reliable even in the absence of bootstrap support for many nodes. All 12 subfamilies were resolved in both MP and BI analyses except Arundinoideae that were not monophyletic.

Within the BEP clade, Pooideae are supported with BI (0.96 PP) but not with MP. *Brachyelytrum* (Brachyelytreae) is resolved as sister to the rest of the pooids in both BI (0.91 PP) and MP but was not supported in MP. BI analysis supported the pooids with a sister clade of *Lygeum, Milium* and *Nardus*. Tribes Lygeae and Nardeae are grouped together in both analyses (99% BS, 0.53 PP). Tribe Stipeae are paraphyletic because of the inclusion of *Ampelodesmos* (tribe Ampelodesmeae) (Figures 2.4 and 2.5).


POOIDEAE
Festuca
Hellena
Narduroides
Wangenheimia
Psilurus
Vulpia
Ctenopsis
Cutandia
Micropyrum
Castellia
Deschampsia
Sesleria
Holcus
Hainardia
Parapholis
Catabodium
Sphenopus
Cynosurus
Dactylis
Lamarckia
Alopecurus
Apera
Beckmannia
Phleum
Arctagrostis
Helictotrichon
Arctophila
Dupontia
Aira
Puccinellia
Sclerochloa
Poa
Simplicia
Agrostis
Polypogon
Ammophila
Calammophila
Calamagrostis
Briza
Amphibromus
Anthoxanthum
Hierochloe
Corynephorus
Gaudinia
Koeleria
Parafestuca
Trisetum
Arrhenatherum
Avena
Lagurus
Phalaris
Diarrhena
Aegilops
Triticum
Triticosecale
Taeniatherum
Crithopsis
Secale
Elymus
Pseudoroegneria
Haynaldia
Thinopyrum
Lolium
Schedonorus
Micropyropsis
Eremium
Leymus
Eremopyrum
Henrardia
Australopyrum
Agropyron
Peridictyon
Hordeum
Stenostachys
Boissiera
Bromus
Hordelymus
Psathyrostachys
Brachypodium
Glyceria
Melica
Phaenospermat
Anisopogon
Nassella
Stipa
Oryzopsis
Anemanthele
Piptatherum
Ampelodesmos
Piptochaetium
Milium
Lygeum
Nardus
Brachyelytrum
Poeae
Aveneae
Diarrheneae
Triticeae
Bromeae
Brachy
Melic
Phaen
Stip
Ampe
Nard
Brachy
66
78
88
56
95
63
99
98
81
76
100
71
61
85
97
89
91
75
53
72
76
62
72
94
72
54
82
97
92
83
64
90
96
98
52
99


Bambusa
Racemobambos
Gigantochloa
Orebambos
Dendrocalamus
Greslania
Thyrsostachys
Schyzostachyum
Pseudostachyum
Melocanna
Arthrostylidium
Chusquea
Neurolepis
Aulonemia
Rhipidocladum
Olmeca
Otatea
Cryptochloa
Lithachne
Raddia
Olyra
Eremitis
Pariana
Buergersiochloa
Drepanostachyum
Fargesia
Himalayacalamus
Oligostachyum
Indocalamus
Phyllostachys
Borinda
Yushania
Chimonobambusa
Shibataea
Pseudosasa
Sinobambusa
Hibanobambusa
Sasa
Arundinaria
Thamnocalamus
Gaoligongshania
Luziola
Zizaniopsis
Rhynchoryza
Potamophila
Prosphytochloa
Zizania
Hygroryza
Chikusichloa
Leersia
Oryza
Microlaena
Ehrharta
Isachne
Bambuseae
Olyreae
Bambuseae
Oryzeae
Ehrharteae
BAMBUSOIDEAE
EHRHARTOIDEAE


Enteropogon
Eustachys
Chloris
Lintonia
Brachyachne
Cynodon
Lepturus
Oxychloris
Tetrapogon
Trichloris
Dinebra
Astrebla
Leptochloa
Diplachne
Coelachyrum
Eleusine
Kengia
Vaseyochloa
Microchloa
Trichoneura
Harpochloa
Monodia
Triodia
Plectrachne
Muhlenbergia
Schedonnardus
Aegopogon
Pleuraphis
Bouteloua
Buchloe
Monanthochloe
Dactyloctenium
Tridens
Orcuttia
Tuctoria
Neostapfia
Melanocenchris
Gymnopogon
Perotis
Aeluropus
Tragus
Acrachne
Calamovilfa
Spartina
Crypsis
Sporobolus
Zoysia
Eragrostis
Pappophorum
Eragrostiella
Hetérachne
Fingerhuthia
Tetrachne
Uniola
Cladoraphis
Enneapogon
Schmidtia
Cottea
Neyraudia
Triraphis
Centropodia
Merxmuellera
Karroochloa
Schismus
Chaetobromus
Danthonia
Plinthanthesis
Cortaderia
Rytidosperma
Merxmuellera
Aristida
Stipagrostis
Eriachne
Micraira
Molinia
Hakonechloa
Phragmites
Molinopsis
Elytrophorus
Styppeiochloa
Arundo
Monachather
Amphipogon
Chlor
Cyno
Erag
Orcu
Dantho
Eriach
Micrai
Arundi
CHLORIDOIDEAE
DAN
ARI
ARU

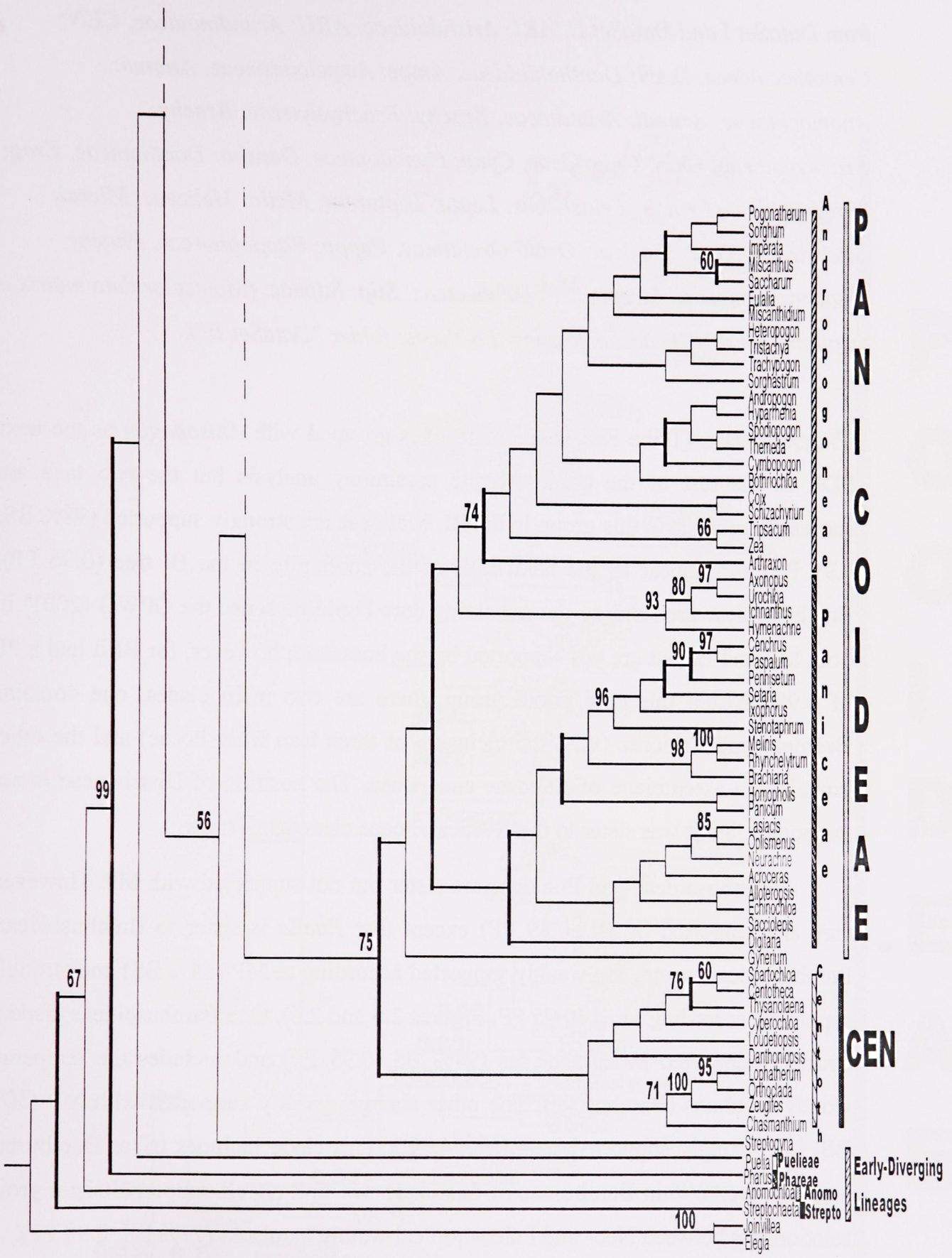


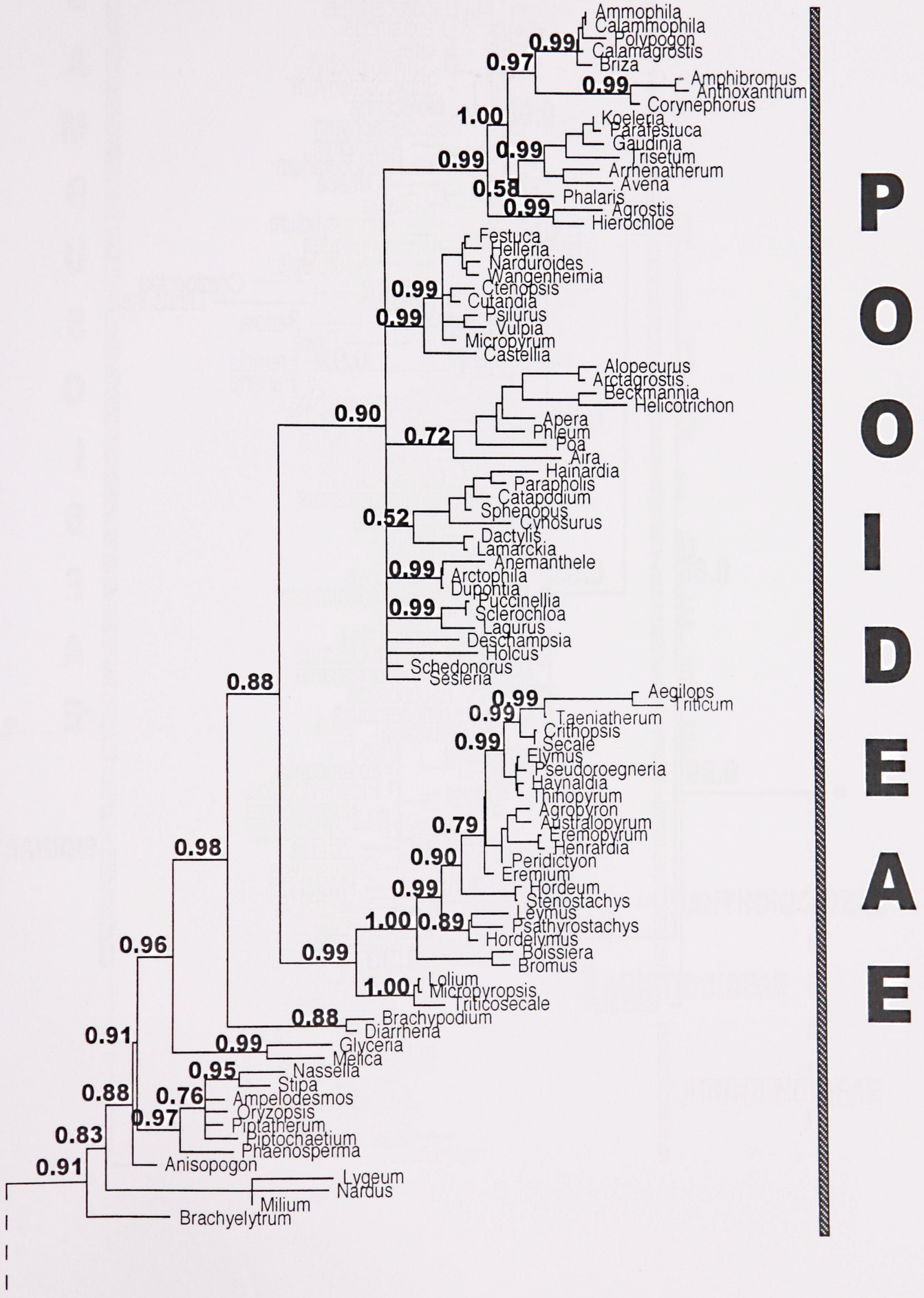
Figure 2.5 One of the 100 equally most parsimonious trees from a heuristic search of DataSet II (33% missing data). Bootstrap values are shown on nodes with >50% support. Vertical bars in ***bold*** *indicate congruent nodes between heuristic searches*

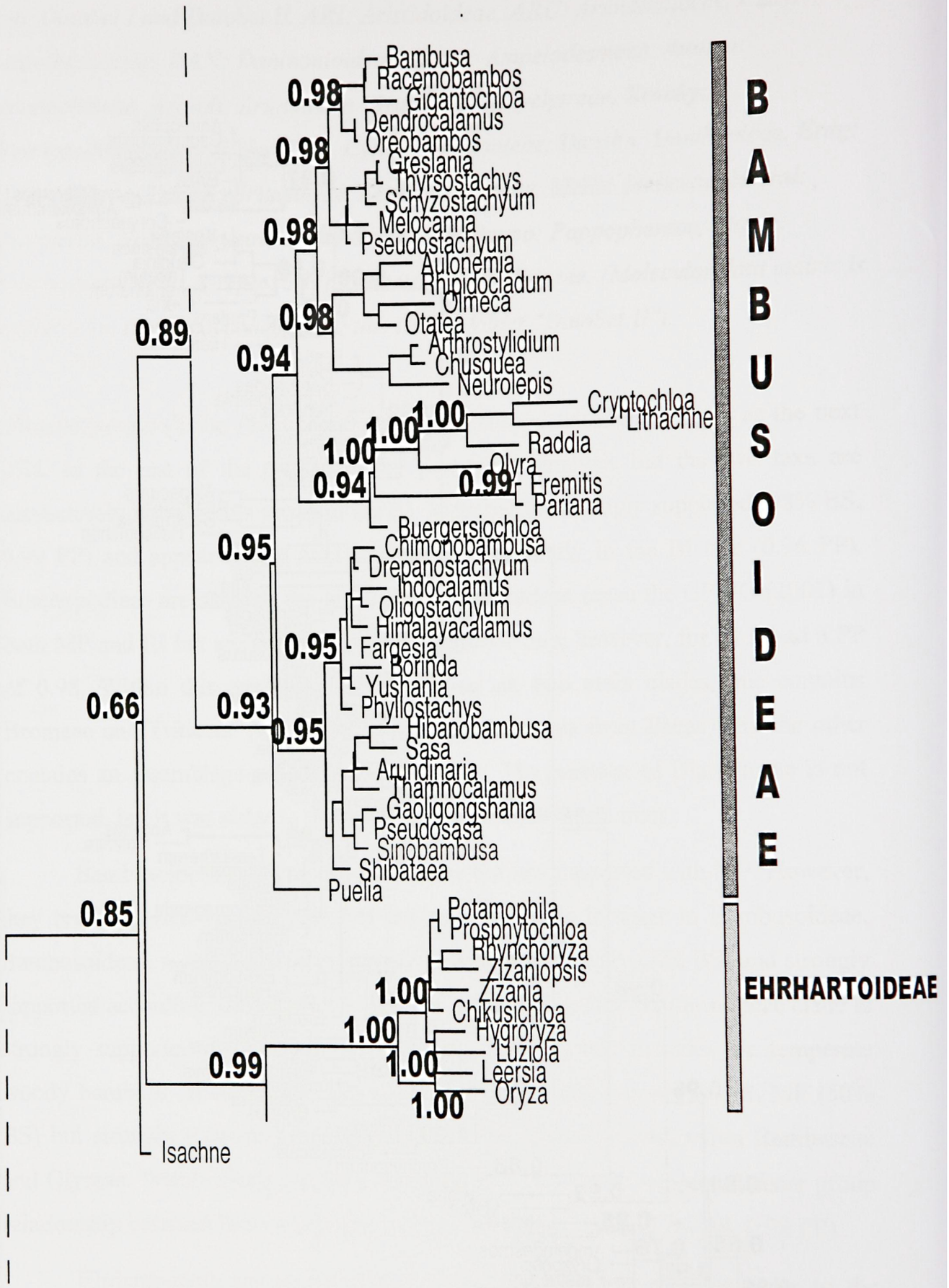
from DataSet I and DataSet II. ***ARI****: Aristidoideae,* ***ARU****: Arundinoideae,* ***CEN****: Centothecoideae,* ***DAN****: Danthonioideae,* ***Ampe****: Ampelodesmeae,* ***Anomo****: Anomochloeae,* ***Arundi****: Arundineae,* ***Brachy****: Brachyelytreae,* ***Brachy****: Brachypodieae,* ***Chlo****: Chlorideae,* ***Cyno****: Cynodonteae,* ***Dantho****: Danthonieae,* ***Erag****: Eragrostideae,* ***Eriach****: Eriachneae,* ***Leptu****: Leptureae,* ***Melic****: Meliceae,* ***Micrai****: Micraireae,* ***Nard****: Nardeae,* ***Orcu****: Orcuttieae,* ***Pappo****: Pappophoreae,* ***Phaen****: Phaenospermateae,* ***Strepto****: Streptochaeteae,* ***Stip****: Stipeae. (Molecular data matrix is available on the CD accompanying this thesis; folder "DataSet II").*

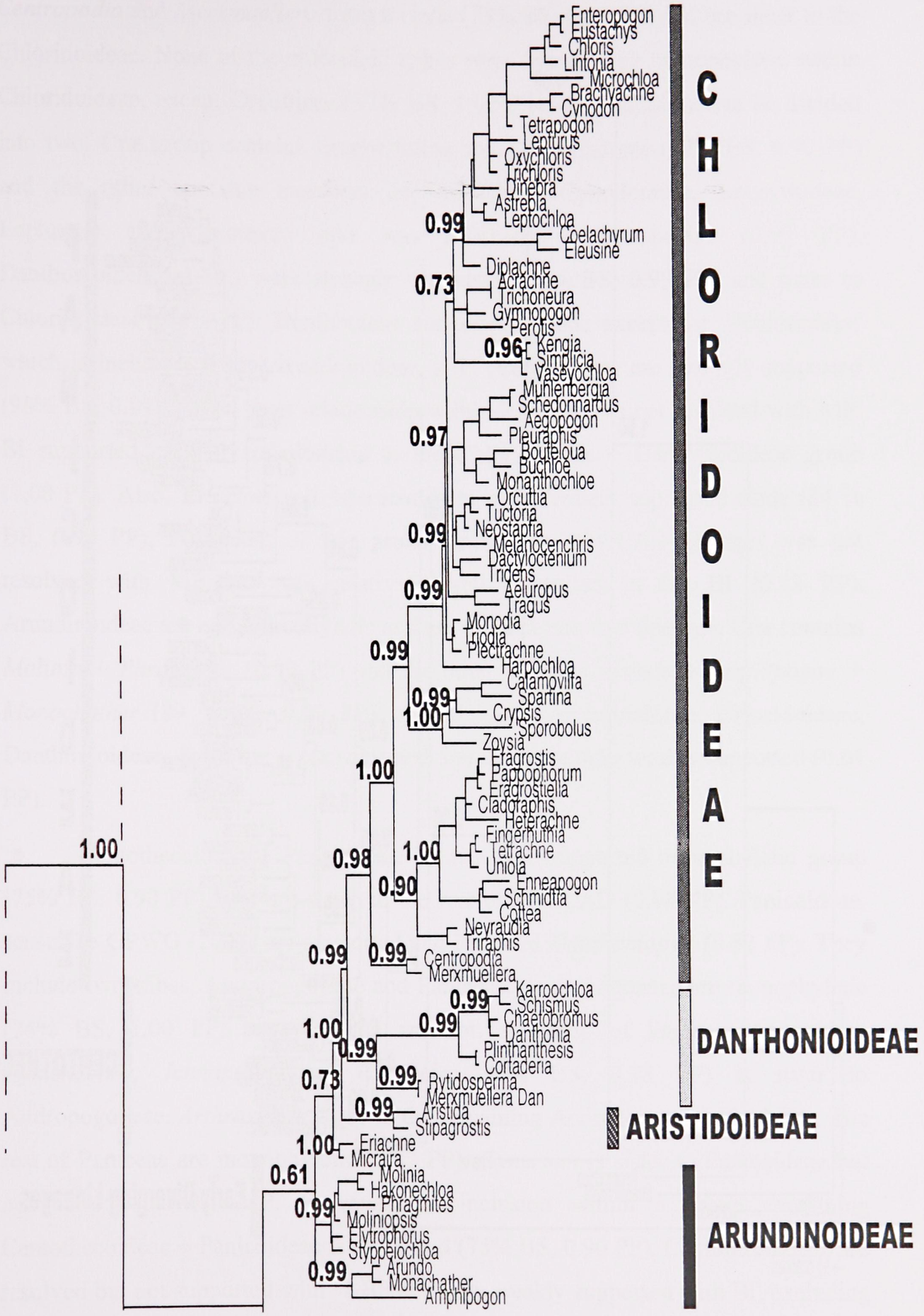
Phaenosperma (Tribe Phaenospermateae) is grouped with *Anisopogon* as the next EDL to the rest of the pooids in the parsimony analysis but the two taxa are successively sister to this group in the BI. Meliceae are strongly supported (98% BS, 0.99 PP) and appear as the next EDL of the subfamily in the BI tree (0.96 PP). Brachypodieae are sister to the remaining core Pooideae sensu the GPWG (2001) in both MP and BI but are not supported by the bootstrap; however, for BI it had a PP of 0.98. Within this core pooid group, there are two main clades, one contains Bromeae and Triticeae (with the inclusion of three taxa from Poeae) and the other contains an assemblage of Aveneae and Poeae. The position of Diarrheneae is not supported, but it was sister to the Aveneae/Poeae clade in all trees.

Bambusoideae and Pooideae are sister but not supported with MP. However, they are supported by BI (0.89 PP) except that *Puelia* is sister to Bambusoideae. Bambusoideae, s. str., are weakly supported according to MP (64% BS) and strongly supported according to BI (0.95 PP; Figures 2.5 and 2.6). One Bambusoideae clade is strongly supported in all analyses (94% BS, 0.95 PP) and includes the temperate woody bamboos (Bambuseae). The other clade is weakly supported with MP (50% BS) but strongly supported by BI (0.94 PP). This clade includes tribes Bambuseae and Olyreae. Within Bambuseae of this clade, we find a well-supported sister group relationship between Neo- and Paleo-tropical woody bamboos (77% BS, 0.98 PP).

Ehrhartoideae are monophyletic (91% BS, 0.99 PP) and are sister to the Pooideae + Bambusoideae group (0.66 PP, including *Puelia*). Within Ehrhartoideae, Ehrharteae and Oryzeae are both monophyletic (91% BS, 0.99 PP). Inclusion of *Isachne* within the BEP clade is not supported with MP and only weakly supported in BI (0.85 PP).


Ammophila
Calammophila
Polypogon
Calamagrostis
Briza
Amphibromus
Anthoxanthum
Corynephorus
Koeleria
Parafestuca
Gaudinia
Trisetum
Arrhenatherum
Avena
Phalaris
Agrostis
Hierochloe
Festuca
Helleria
Narduroides
Wangenheimia
Ctenopsis
Cutandia
Psilurus
Vulpia
Micropyrum
Castellia
Alopecurus
Arctagrostis
Beckmannia
Helicotrichon
Apera
Phleum
Poa
Aira
Hainardia
Parapholis
Catapodium
Sphenopus
Cynosurus
Dactylis
Lamarckia
Anemanthele
Arctophila
Dupontia
Puccinellia
Sclerochloa
Lagurus
Deschampsia
Holcus
Schedonorus
Sesleria
Aegilops
Triticum
Taeniatherum
Crithopsis
Secale
Elymus
Pseudoroegneria
Haynaldia
Thinopyrum
Agropyron
Australopyrum
Eremopyrum
Henrardia
Peridictyon
Eremium
Hordeum
Stenostachys
Leymus
Psathyrostachys
Hordelymus
Boissiera
Bromus
Lolium
Micropyropsis
Triticosecale
Brachypodium
Diarrhena
Glyceria
Melica
Nassella
Stipa
Ampelodesmos
Oryzopsis
Piptatherum
Piptochaetium
Phaenosperma
Anisopogon
Lygeum
Nardus
Milium
Brachyelytrum
0.99
0.97
0.99
1.00
0.99
0.58
0.99
0.99
0.99
0.90
0.72
0.52
0.99
0.99
0.88
0.99
0.99
0.99
0.79
0.90
0.99
1.00
0.89
0.99
1.00
0.88
0.98
0.96
0.91
0.99
0.95
0.88
0.76
0.97
0.83
0.91
POOIDEAE


Bambusa
Racemobambos
Gigantochloa
Dendrocalamus
Oreobambos
Greslania
Thyrsostachys
Schyzostachyum
Melocanna
Pseudostachyum
Aulonemia
Rhipidocladum
Olmeca
Otatea
Arthrostylidium
Chusquea
Neurolepis
Cryptochloa
Lithachne
Raddia
Olyra
Eremitis
Pariana
Buergersiochloa
Chimonobambusa
Drepanostachyum
Indocalamus
Oligostachyum
Himalayacalamus
Fargesia
Borinda
Yushania
Phyllostachys
Hibanobambusa
Sasa
Arundinaria
Thamnocalamus
Gaoligongshania
Pseudosasa
Sinobambusa
Shibataea
Puelia
Potamophila
Prosphytochloa
Rhynchoryza
Zizaniopsis
Zizania
Chikusichloa
Hygroryza
Luziola
Leersia
Oryza
Isachne
0.98
0.98
0.98
0.98
0.89
0.94
1.00
1.00
1.00
0.94
0.99
0.95
0.95
0.66
0.93
0.95
0.85
1.00
1.00
0.99
1.00
BAMBUSOIDEAE
EHRHARTOIDEAE


Enteropogon
Eustachys
Chloris
Lintonia
Microchloa
Brachyachne
Cynodon
Tetrapogon
Leptiurus
Oxychloris
Trichloris
Dinebra
Astrepia
Leptochloa
Coelachyrum
Eleusine
Diplachne
Acrachne
Trichoneura
Gymnopogon
Perotis
Kengia
Simplicia
Vaseyochloa
Muhlenbergia
Schedonnardus
Aegopogon
Pleuraphis
Bouteloua
Buchloe
Monanthochloe
Orcuttia
Tuctoria
Neostapfia
Melanocenchris
Dactyloctenium
Tridens
Aeluropus
Tragus
Monodia
Triodia
Plectrachne
Harpochloa
Calamovilfa
Spartina
Cynosis
Sporobolus
Zoysia
Eragrostis
Pappophorum
Eragrostiella
Cladoraphis
Heteranthe
Fingerhuthia
Tetrachne
Uniola
Enneapogon
Schmidtia
Cottea
Neyraudia
Triraphis
Centropodia
Merxmuellera
Karroochloa
Schismus
Chaetobromus
Danthonia
Plinthanthesis
Cortaderia
Rytidosperma
Merxmuellera Dan
Aristida
Stipagrostis
Eriachne
Micraira
Molinia
Hakonechloa
Phragmites
Moliniopsis
Elytrophorus
Styppeiochloa
Arundo
Monachather
Amphipogon
1.00
0.99
0.73
0.96
0.97
0.99
0.99
1.00
0.99
1.00
0.98
1.00
0.90
0.99
0.99
0.99
1.00
0.99
0.73
0.99
0.99
1.00
0.61
0.99
0.99
CHLORIDOIDEAE
DANTHONIOIDEAE
ARISTIDOIDEAE
ARUNDINOIDEAE

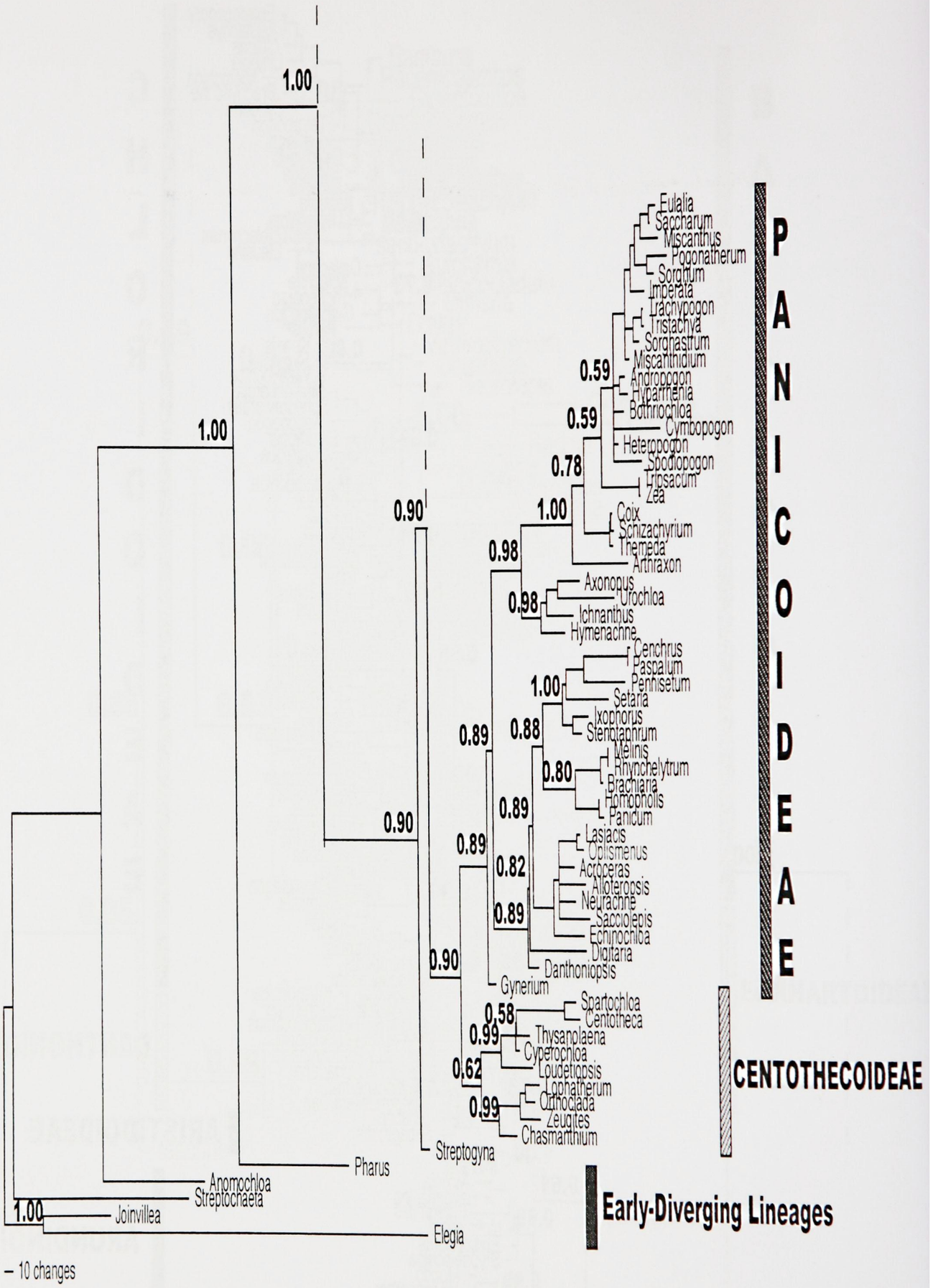


Figure 2.6 Bayesian consensus tree of combined data sets from DataSet II where posterior probabilities are shown above the branches.

Centropodia and *Merxmuellera* form a clade (73% BS, 0.99 PP) and are sister to the Chloridoideae. None of the chloridoid tribes were found to be monophyletic within Chloridoideae, except Orcuttieae (97% BS, 1.00 PP). Chloridoideae can be divided into two. One group contains Eragrostideae and Pappophoreae (62% BS, 0.90 PP) and the other contains members of Chlorideae, Cynodonteae, Eragrostideae, Leptureae and Orcuttieae and was relatively well-supported (0.99 PP). Danthonioideae, s. str., were strongly supported (96% BS, 0.99 PP) and sister to Chloridoideae (0.99 PP). Danthonieae are monophyletic except for *Monachather,* which is included within Arundinoideae, s. l.. Aristidoideae are strongly supported (98% BS, 0.99 PP) but their relationship within PACCAD is not resolved with MP. BI supported its sister relationship to the Chloridoideae + Danthonioideae group (1.00 PP). Also, *Eriachne* and *Micraira* form a moderately supported clade (84 % BS, 0.99 PP). Placement of this group among the PACCAD lineages was not resolved with MP but was relatively well supported in the BI (0.73 PP). Arundinoideae are not monophyletic and are divided into two lineages. One contains *Molinia* + *Phragmites* (0.99 PP) and the other contains *Arundo* + *Amphipogon* + *Monachather* (89 % BS, 0.99 PP). Aristidoideae, Arundinoideae, Chloridoideae, Danthonioideae, *Eriachne* + *Micraira* and *Simplicia* are only weakly supported (0.61 PP).

Centothecoideae + Panicoideae form a well-supported monophyletic group (75% BS, 0.90 PP), and are sister to the rest of PACCAD (0.90 PP). Panicoideae, sensu the GPWG (2001) are supported and included *Danthoniopsis* (0.82 PP). They include two tribes, Andropogoneae and Paniceae. Andropogoneae are monophyletic (74% BS, 1.00 PP) but Paniceae are not. A group of Paniceae (*Axonopus, Hymenachne, Ichnanthus* and *Urochloa;* 93% BS, 0.98 PP) is sister to Andropogoneae. *Arthraxon* is sister to the remaining Andropogoneae (0.89 PP). The rest of Paniceae are monophyletic (0.82 PP). *Gynerium* is sister to Panicoideae but not supported with MP. However, its inclusion within a group containing Centothecoideae + Panicoideae is supported (75% BS, 0.90 PP). Centothecoideae are resolved but not supported with % BS and only weakly supported with BI, excluding *Danthoniopsis* (0.62 PP). *Cyperochloa* (Cyperochloeae) is nested within the subfamily as well as a representative of Arundinelleae, *Loudetiopsis* (0.99 PP). *Streptogyna* is weakly supported as sister to PACCAD with MP (56% BS) and

relatively well supported with BI (0.90 PP). Finally, we retrieved four EDLs, that are successively sister to the rest of the grasses, comprising *Puelia* (no bootstrap support), *Pharus* (99% BS, PP 1.00), *Anomochloa* (67% BS) and *Streptochaeta*. Anomochlooideae are not monophyletic.

2.4 Discussion

2.4.1 Effect of characters and taxon sampling, and missing data

Previous theoretical (Hillis 1996a; Graybeal, 1998) and empirical (Soltis et al., 1999) studies have indicated that large numbers of characters may be necessary to resolve phylogenetic patterns in many groups of organisms. By increasing character number in our study from a range of 283-742 parsimony informative characters in the single gene analyses (results not shown) to a range of 1,107-1,403 in the combined analyses, we found more robust and resolved phylogenetic trees than in individual single-gene analyses (Table 2.2). There is no reason to suggest that our results might have experienced systematic bias after data combination because there is topological convergence of trees with both MP and BI methods. Reducing misleading effects or systematic bias might be achieved by increased taxon sampling (Wiens, 1998; Hillis et al., 2003; Salamin et al., 2005), as it enables a better detection of multiple substitutions at the same nucleotide site. This helps counteract branch-attraction effects and therefore improves phylogenetic inference (Hillis, 1996b). Some empirical studies have also found that data combination (i.e. multi-gene approaches) of multiple sequences from the same taxon (i.e. the multi-gene approach) does improve accuracy of phylogenetic inference (Qiu et al., 1999; Soltis et al., 1999; Bapteste et al., 2002). Our results in the combined analyses (Figures 2.4, 2.5 and 2.6) show high levels of congruence between the phylogenetic inferences with 107 and 294 taxa (the latter increased the proportion of all grass genera sampled from 15 to 42%).

The impact of missing data (i.e. taxa for which there is a proportion of missing character states) has been neglected in phylogenetic analyses (Wiens, 2005). If we compare the phylogenetic inferences of DataSet I (no missing data but fewer taxa) and DataSet II (34% of missing data and more taxa), we observe the same clades in both (Figure 2.5). However, bootstrap percentages differ greatly between the two analyses (compare Figures 2.4 and 2.5).

Trees from DataSet I showed high support for most clades, but clades determined in DataSet II were not as well supported. Wiens (2003) showed that reduced phylogenetic accuracy resulting from the inclusion of missing data (i.e. incomplete taxa) was associated with incorrect placement of only the incomplete taxa; the relationships among the complete taxa (i.e. with no missing data) were estimated almost perfectly. The lack of support for clades associated with analyses of DataSet II may have resulted from the poorly resolved placement of incomplete taxa. The number of characters included for these taxa is crucial to correctly place them on the tree. Wiens (1998), in a simulation study, explained that it was not just the amount of missing data that is important for phylogenetic accuracy, but also how these missing data are distributed among taxa. In our analyses, missing data were biased towards the subfamilies Chloridoideae, Panicoideae and Pooideae, and this is mainly due to their relatively larger size. Consequently, support for these clades has been more affected in DataSet II than the other major clades (Figures 2.5 and 2.6). However, Wiens (2005) argued that adding taxa that are 50% incomplete (i.e. for which only half of the characters are known) might show similar benefits to adding complete taxa under many conditions. Our analyses support this assertion and suggest that adding incomplete taxa might have great benefits as long as their placement are checked for consistency with phylogenetic inferences including only complete taxa (Figures 2.4, 2.5 and 2.6).

There are also differences in the performance of the different phylogenetic methods when incomplete taxa are included: BI seems to be less sensitive than MP to missing data (compare Figures 2.5 and 2.6). Indeed, Bayesian inference produced relatively good support for most internal nodes but there was incorrect placement of four taxa (Figure 2.6). One has to bear in mind that the support for the placement of incomplete taxa in the BI analyses should be interpreted cautiously because BI can

overestimate support (Simmons et al. 2004), and thus the appearance of missing data having less effect on BI could be an illusion.

2.4.2 *Phylogenetic relationships between and within large clades, subfamilies and tribes*

2.4.2.1 EDL

Three EDL were recognized by the GPWG (2001) as successively sister to the rest of the grass family: Puelioideae, Pharoideae and Anomochlooideae (including *Anomochloa* and *Streptochaeta* respectively). In DataSet I, Anomochlooideae represented by *Streptochaeta*, diverges from the deepest node in the tree followed by *Pharus* (Pharoideae); Puelioideae was not sampled. In DataSet II, *Streptochaeta* was also the earliest diverging lineage followed by *Anomochloa*, *Pharus* and *Puelia* (Puelioideae). In BI, *Puelia* was sister to Bambusoideae (Figure 2.6), but this position seems to be influenced by missing data (as discussed above). It was only sequenced for *rbcL* and in this tree (Figure 2.1) formed part of a polytomy with all grasses (except Anomochlooideae and Pharoideae). The *rbcL* tree conflicts with the general pattern of four successively sister groups to the rest of the grasses, as *Streptochaeta* and *Anomochloa* group together (but with only 54% BS). Based on our results, it is more conservative to recognize four EDLs because of the possible paraphyly of Anomochlooideae. However, note that *Anomochloa* is weakly supported as the next EDL after *Streptochaeta* (Figure 2.5). The monophyly of *Anomochloa* and *Streptochaeta* was also not supported in three previous phylogenetic analyses of the family using both molecular and morphological data (Clark et al., 1995; Soreng and Davis, 1998; Hilu et al., 1999). Morphological synapomorphies defining this clade are not easy to find (Clark and Judziewicz, 1996). Only the presence of the adaxial ligule as a fringe of hairs supports its monophyly but this character appears elsewhere in the family (GPWG, 2001). However, Anomochlooideae are mainly characterized by the absence of true grass spikelets, florets, and lodicules (GPWG, 2001). The phylogenetic relationships of Poaceae subgroups have been shown to depend on outgroup selection (Duvall and Morton, 1996), and we expect that the monophyly of Anomochlooideae could be assessed further by using different sets of outgroup taxa. The positions of Pharoideae

and Puelioideae are in agreement with all studies that have included these taxa (Clark et al., 1995; Clark and Judziewicz, 1996; Soreng and Davis, 1998; Clark et al., 2000; GPWG, 2001).

2.4.2.2 PACCAD and BEP clades

Our results show a clear and well-supported BEP-PACCAD bifurcation with both DataSet I and II (Figure 2.4, 2.5 and 2.6). In previous studies, when EDLs are excluded from consideration, the remaining grasses are generally, but not always, split into two lineages (Clark et al., 1995; Soreng and Davis, 1998; GPWG, 2001). Soreng and Davis (1998) recovered a Pooideae + PACCAD clade and the GPWG (2001) found BEP and PACCAD clades. The sister-relationship of BEP + PACCAD is still controversial as few studies have found strong support for this grouping, and no morphological synapomorphies supporting the BEP clade have been identified (GPWG, 2001). According to the GPWG (2001), the lack of sequence data for *Streptogyna* may have affected the assessment of the BEP monophyly in their combined analysis. In our study, *Streptogyna americana* was only sequenced for *rbcL*, and therefore also lacks sequence data to infer an accurate placement in the combined analyses. However, our *rbcL* results (70% BS; Figure 2.1) and the combined analyses (56% BS; Figure 5; 0.90 PP; Figure 2.6) suggest that *Streptogyna* should be placed within the PACCAD clade, in particular with BI that provides good support for its placement.

2.4.2.3 Subfamilies and tribes of the PACCAD clade

Within the PACCAD clade, our results supported the monophyly of the six subfamilies as defined by the GPWG (2001). The monophyly of Chloridoideae was supported and this is in agreement with previous studies (Hilu et al., 1999; GPWG, 2001; Hilu and Alice, 2001). Many clades within the chloridoids were also in agreement with previous analyses. For example, Hilu and Alice (2001) sampled 56 genera and found that *Centropodia* was sister to Chloridoideae. An assemblage of tribes Pappophoreae, Eragrostideae and Uniolineae diverged early in the evolution of the group (Hilu et al., 1999). Also, a clade including *Sporobolus* and *Zoysia* was well supported (Clark et al., 1995; Soreng and Davis, 1998). Our results supported a clade

comprising *Centropodia* and *Merxmuellera*, and it is sister to Chloridoideae in the combined BI (0.98 PP; Figure 2.6) and *rbcL* (58% BS; Figure 2.1) analyses. However, it was not supported by bootstrapping in the MP of the combined data (Figure 2.5) nor in the *matK* parsimony analysis. As suggested by the GPWG (2001), it might be *Merxmuellera rangeii* forming a clade with *Centropodia*. Tribes Eragrostideae and Pappophoreae were not monophyletic but an assemblage of them was found in a group sister to the rest of the chloridoids (Figures 2.4 and 2.5). The next diverging lineage was a group containing *Calamovilfa*, *Crypsis*, *Spartina*, *Sporobolus* and *Zoysia* (Figures 2.4 and 2.5). In my MP analysis of the combined DataSet II (Figure 2.5), there is a lack of resolution at the base of the subfamily that may be accounted for by the amount of missing data. It may also reflect an accelerated rate of radiation/diversification in this group (Hilu and Alice, 2001), which is illustrated by the short branch lengths in the Bayesian analyses (Figures 2.4 and 2.6). Hilu and Alice (2001) found a similar lack of resolution using *matK*. According to our phylogenetic inferences, a sister group relationship is found between Chloridoideae and Danthonioideae (Figures 2.4, 2.5, and 2.6). This finding disagrees with previous studies where Arundinoideae were thought to be the most closely related subfamily to the chloridoids (Clayton and Renvoize, 1986; Hilu and Alice, 2001).

Danthonioideae are well supported in all our analyses with the exclusion of *Monachather* (Figure 2.5). My results are based on 11 representatives of Danthonieae and eight Arundineae. Arundinoideae were paraphyletic or unresolved (Figures 2.5 and 2.6) with a non-supported *Elytrophorus* + *Hackonechloa* + *Molinia* + *Moliniopsis* + *Phragmites* + *Styppeiochloa* clade in the MP analysis (but with a 0.99 PP in the BI), and a well-supported *Amphipogon* + *Arundo* + *Monachather* clade (89% BS, 0.99 PP; Figures 2.5 and 2.6). *Spartochloa* is excluded from these clades (Figure 2.5). It is worth noting that Arundinoideae were weakly monophyletic in DataSet I, but comprise only *Arundo*, *Molinia* and *Phragmites* (Figure 2.4) and is sister to the Danthonioideae and Chloridoideae group (63% BS, 0.72 PP). Previous studies suggested a monophyletic Arundinoideae comprising two clades: one containing tribe Danthonieae and the other tribe Arundineae but their respective bootstrap values were low (Barker et al., 1999). Subsequent phylogenetic analyses have proposed that they would be better treated as two distinct subfamilies,

Arundinoideae and Danthonioideae (GPWG, 2001) but their composition was not precisely determined mainly due to a poor sampling (GPWG, 2001). The monophyly of Arundinoideae was implied using only a few taxa (Hilu et al., 1999), mainly *Arundo* and *Phragmites* (Duvall and Morton, 1996) and the monophyly of Danthonioideae was supported by an unreversed morphological synapomorphy (presence of haustorial synergids) (Verboom et al., 2006) but a greater sampling is necessary before its monophyly can be assessed (GPWG, 2001).

Aristidoideae, represented by *Aristida* and *Stipagrostis*, were well supported in our analyses and look best positioned as sister to a Chloridoideae / Danthonioideae group. However, their position among PACCAD lineages was not always clear. They were sister to the Chloridoideae / Danthonioideae group in DataSet II (1.00 PP; Figure 2.6). They were also present in all equally most parsimonious trees in the MP analysis, but were unsupported by BS. However, they were sister to only Danthonioideae in the strict consensus tree of the *rbcL* analysis but this placement was not supported by bootstrapping (Figure 2.1). With *matK* they were unresolved (Figure 2.2) and it was missing in the *trnL-F* analysis. We can, therefore, only tentatively suggest that Aristidoideae are sister to a Chloridoideae / Danthonioideae group. More taxa and character sampling are therefore required to test this hypothesis.

The positions of *Cyperochloa, Eriachne*, *Gynerium, Micraira* and *Streptogyna* were recognized as *incertae sedis,* representing five distinct tribes by the GPWG (2001). In our study, *Eriachne* and *Micraira* form a well-supported clade (84% BS, 1.00 PP; Figures 2.5 and 2.6) but the position of this group within the PACCAD clade is not always strongly supported (Figures 2.5 and 2.6). In the combined BI analysis, this *Eriachne* + *Micraira* group is sister to a chloridoid + danthonioid + aristidoid group (0.73 PP; Figure 2.6). The same pattern is seen in the MP analysis except that it is not supported by bootstrapping (Figure 2.5). With *rbcL*, *Eriachne* and *Micraira* also group together and form a polytomy with Arundinoideae and the rest of the PACCAD clade (Figure 2.1). It has been hypothesized that *Eriachne* might be placed near the base of the PACCAD group (GPWG, 2001) but our results do not support this hypothesis. However, the presence of an *Eriachne* + *Micraira* clade as sister to a chloridoid + danthonioid + aristidoid group can be interpreted as a novel result. *Streptogyna* was weakly supported as sister to the

PACCAD clade (Figures 2.5 and 2.6; 56% BS, 0.90 PP respectively) and was moderately supported with *rbcL* only (73% BS, Figure 2.1). This contradicts the suggestion that *Streptogyna* might be placed within Ehrhartoideae (GPWG, 2001). *Cyperochloa* was positioned within Centothecoideae (Figure 2.5) but did not receive strong support, except in the combined BI (0.99 PP). Finally, *Gynerium* was sister to Panicoideae in all equally parsimonious trees of the combined matrix of DataSet II (but was not supported by BS or by high PP). With *matK*, it was well supported as sister to Panicoideae (89% BS, Figure 2.2).

Our results suggest a strongly supported sister group relationship between Centothecoideae + Panicoideae and an Aristidoideae + Arundinoideae + Chloridoideae + Danthonioideae + *Eriachne* / *Micraira* group (Figures 2.4 and 2.5). The relationships between the major lineages of the PACCAD clade have not been fully resolved by previous studies of the family (Soreng and Davis, 1998; Hilu et al., 1999; GPWG, 2001). It is possible that a rapid radiation of the PACCAD group has obscured the phylogenetic signal and made the relationships difficult to resolve (GPWG, 2001). There is however consistency among subfamilies in our analyses (Figures 2.4, 2.5 and 2.6) that would support the pattern of a paraphyletic grade of Arundinoideae taxa, an *Eriachne* + *Micraira* group and Aristidoideae as successively sister to a monophyletic Chloridoideae + Danthonioideae group. This pattern could therefore be represented by the following parenthetical notation: Arundinoideae grade (*Eriachne* / *Micraira* (Aristidoideae (Chloridoideae, Danthonioideae))).

Taxa of Centothecoideae grouped with a monophyletic Panicoideae with high support but were not themselves monophyletic (Figures 2.4 and 2.5). The monophyly of Centothecoideae was not found (Figure 2.4) or not well supported (Figures 2.5 and 2.6) regardless of the sample size (DataSets I and II). Centothecoideae can be divided into two clades (Figures 2.4, 2.5 and 2.6). We also retrieve *Loudetiopsis* (tribe Arundinelleae) within Centothecoideae (Figure 2.5). In a previous study, *Loudetiopsis* was found sister to a Centothecoideae + Panicoideae group (Hilu et al., 1999).

Panicoideae were well supported in most of our analyses (Figures 2.2-2.6). Two main clades within Panicoideae can be identified. One contains tribe Andropogoneae (with the inclusion of *Tristachya*) and four representatives of tribe

Paniceae (*Axonopus*, *Ichnanthus*, *Hymenachne* and *Urochloa;* Figure 2.5), which form a strongly supported clade (93% BS, 0.98 PP) that may be sister to the Andropogoneae. This sister-group relationship to Andropogoneae was not supported with BS except in the *matK* dataset (Figure 2.2). However, their sister group status is supported by BI (0.98 PP). The inclusion of *Tristachya* within Andropogoneae is easily explained by convergence in morphological characteristics (spikelets in triads) of tribe Arundinelleae for *Tristachya* (Hilu et al., 1999). The other major clade of Panicoideae contains exclusively representatives of Paniceae except for the inclusion of *Danthoniopsis* (also Arundinelleae) in the BI (Figure 2.6). The monophyly of Arundinelleae has been questioned elsewhere (GPWG, 2001).

2.4.2.4 Subfamilies and tribes of the BEP clade

Within the BEP clade, the three main subfamilies are generally well supported (Figures 2.4, 2.5 and 2.6). In DataSet I (Figure 2.4), Pooideae are strongly supported (100% BS, 1.00 PP). Bambusoideae are monophyletic (98% BS, 1.00 PP) and sister to Pooideae (96% BS, 1.00 PP). Ehrhartoideae were found to be monophyletic (91% BS, 0.99 PP) and sister to the Bambuoideae + Pooideae group (89% BS, 1.00 PP). Tribe Brachyelytreae were sister to the rest of the pooids in both combined analyses of DataSet II. They were found sister to the other pooids in a previous study considering 48 pooid taxa (Catalan *et al.*, 1997) and exhibit some bambusoid characters that are believed to be retained pleisiomorphies. The next earliest diverging lineage in our parsimony analyses of DataSet II (Figure 2.5) and analysis of *rbcL* (Figure 2.1) was a clade that includes tribes Nardeae and Lygeae (Catalan et al., 1997). However, the order of divergence of the EDLs of the Pooideae clade was not conclusive in our analyses (Figures 2.1, 2.2 and 2.5) nor in previous studies (Catalan et al., 1997; GPWG, 2001). However, Brachyelytreae were consistently the EDL in the pooids, and Nardeae + Lygeae (sometimes with the inclusion of *Milium*) and a group of Stipeae genera (including *Ampelodesmos* (Ampelodesmeae)) were generally successively sister to the rest of pooids. It is worth noting that the analysis of DataSet I retrieved tribes Stipeae and Meliceae as the earliest diverging lineage of the pooids because it excludes tribes Brachyelytreae, Nardeae and Lygeae. *Milium* can also be identified as an EDL (Figure 2.5) but it is grouped with Nardeae and Lygeae in the BI (Figure 2.6). *Milium* was included in

tribe Stipeae by Clayton and Renvoize (1986). The general lack of support for the EDLs in the combined MP analysis of DataSet II (Figure 2.5) may be also due to the amount of missing data for those taxa that are incomplete. However, BI posterior probabilities were generally high for most of the early divergences within the pooids (Figure 2.6). Furthermore, once there are complete taxa at the terminal branches (i.e. taxa for which the three DNA regions were sequenced; Figure 2.4), a well-supported core pooid group (80% BS, 1.00 PP) is retrieved with Stipeae being the earliest diverging lineage (100% BS, 1.00 PP). Tribe Meliceae is strongly supported and appears as the next EDL followed by tribe Brachypodieae. Catalan et al. (1997) could not find a supported order of divergence for these tribes. Two main clades were found in the core pooids (Figures 2.4, 2.5 and 2.6): one containing Bromeae + Triticeae and the other containing an assemblage of Poeae + Aveneae taxa. None of these tribes appear monophyletic except tribe Bromeae (*Bromus* and *Boissiera*), which resolves as sister to Triticeae (Figures 2.2, 2.4, 2.5 and 2.6). A more extensive sampling of these tribes will be needed to determine the composition and inter-relationships of many of the major pooid groups.

We find a sister relationship between subfamilies Pooideae and Bambusoideae supported by *matK* sequence data, DataSet I, and BI of DataSet II (Figures 2.2, 2.4 and 2.6). The *rbcL* and *trnL-F* sequence data cannot resolve the relationships of these subfamilies (Figures 2.1 and 2.3). Subfamily Bambusoideae s. str. (Bambuseae and Olyreae), are supported in all our analyses and are divided into two main lineages. One contains exclusively Bambuseae with representatives of temperate woody bamboos, and the other contains tribes Bambuseae, with representatives of neo-tropical and paleo-tropical woody bamboos, and Olyreae containing exclusively herbaceous bamboos. Bambuseae are therefore not monophyletic. Previous phylogenetic studies have found a derived Olyreae lineage from within Bambusoideae s.str. (Clark et al., 1995) and a monophyletic Olyreae / Parianeae as sister to a monophyletic Bambuseae (Kelchner and Clark, 1997). In Kelchner and Clark (1997), 23 bamboo species were sampled whereas we have sampled 43 species. Our results support the findings of Kelchner and Clark (1997) who showed that distinct lineages within the subfamily correspond strongly with geographic divisions. However, we retrieve a sister relationship between herbaceous (Olyreae) and tropical woody bamboos (tribe Bambuseae) rather than an Olyreae /

Bambuseae sister group relationship. It seems therefore that either Olyreae should be included within Bambuseae or some Bambuseae taxa should be included in Olyreae.

The inclusion of Ehrhartoideae within the BEP clade was well-supported in the combined DataSet I (89% BS, 1.00 PP; Figure 2.4) but not in the individual gene analyses (Figures 2.1, 2.2 and 2.3). The subfamily was monophyletic in all our analyses and a sister relationship was found between tribes Ehrharteae and Oryzeae. This was also found by Guo and Ge (2005). Both tribes were supported as monophyletic. Several studies using DNA data have shown that tribe Oryzeae should be considered as a distinct entity (Barker et al., 1995; Clark et al., 1995; Soreng and Davis, 1998; Guo and Ge, 2005), but the inclusion of tribe Ehrharteae was assessed only recently (Hilu et al., 1999; GPWG, 2001). The phylogenetic position of Ehrhartoideae was unclear in other studies (Hilu et al., 1999; GPWG, 2001), but our results support a sister relationship between Ehrhartoideae and Bambusoideae + Pooideae.

Isachne (tribe Isachneae) was found to be sister to the BEP clade but was not supported (Figure 2.5). It received 0.85 PP in the BI. The seemingly incorrect placement of this genus may be due either to the amount of missing data, as only the *rbcL* sequence was available. The *rbcL* analysis could not resolve the position of this taxon (Figure 2.1) as it formed a polytomy with Puelioideae, Bambusoideae, Pooideae/Ehrhartoideae and the PACCAD clade. *Isachne* was considered as a member of Panicoideae (Clayton and Renvoize, 1986; Watson and Dallwitz, 1992), but this cannot be confirmed by our phylogenetic inferences. There is no evidence to embed it in the PACCAD clade but its subfamilial position is unclear.

2.4.3 Conclusion

In this study, we have performed a multi-gene phylogenetic analysis of the grasses with the largest sample size published to date, at tribal and generic levels. It represents a near complete tribal level phylogenetic treatment of the grasses. While there are substantial amount of missing data in our analyses, our phylogenetic inferences showed a considerable topological congruence with our single-gene analyses, and a strongly supported topology with DataSet I. With the exception of

subfamilies Arundinoideae, Centothecoideae and Anomochlooideae whose circumscriptions remain unclear (GPWG, 2001), all subfamilies were resolved as monophyletic as well as most inter-subfamilial relationships. Some of the subfamily relationships within the BEP and PACCAD clades, which remained unclear in the GPWG (2001) paper, have been substantially resolved such as (i) sister-relationship between Panicoideae + Centothecoideae to the rest of the PACCAD, (ii) composition of the BEP clade, (iii) non-monophyly of Bambuseae, (iv) sister-relationship of Ehrhartoideae/Bambusoideae + Pooideae, (v) position of *Gynerium* as sister to Panicoideae, and (vi) monophyly of an *Eriachne* + *Micraira* group. We have provided a summary tree of tribal inter-relationships based on our results that has only been possible by using large trees (this is provided in the final dicussion chapter, Chapter 5 of this thesis).

The lack of BS support for groups determined in our analyses with missing data (DataSet II) reflects the need for a 'better and smarter' data acquisition in grass phylogenetic studies. Two different approaches can be considered to obtain the large trees required to establish inclusive phylogenetic hypotheses and provide a more comprehensive summary of clade's history (Sanderson and Driskell, 2003): (i) the supertree-building methods (Salamin et al., 2002), or (ii) the data acquisition from the widest range of taxa. The latter approach requires us to 'fill the gaps' of DNA data matrices which are now large enough to infer comprehensive phylogenetic trees of the family. This study contributes to continuing progress in grass phylogenetics and raises issues regarding the increase of taxa and the amount of missing data required to reconstruct accurate and robust phylogenetic trees, especially for a very large group of organisms.

3) Diversification of the grasses (Poaceae): a phylogenetic approach to reveal macro-evolutionary patterns

3.1 Introduction

Within the angiosperms, one of the greatest terrestrial radiations of recent geological times (Davies et al., 2004), the grasses are the fifth most diverse family (Renvoize and Clayton, 1992; Watson and Dallwitz, 1992). The earliest records of grass pollen date from the Paleocene of South America and Africa, between 60 and 55 million years ago (Mya hereafter) (Jacobs et al., 1999). These findings, along with the evidence of early-diverging lineages of the grasses (GPWG, 2001) occurring in South America and Africa, and the present-day Gondwanan distribution of many of their lineages, suggest a Gondwanan origin of the family. This origin has also been dated at about 75 Mya, using the Non Parametric Rate Smoothing (NPRS; Sanderson, 1997) molecular dating method (Bremer, 2002). It is also possible that the grasses achieved their Gondwanan distribution by dispersal (Soreng and Davis, 1998), in a similar way to that deduced for Atherospermataceae (Renner et al., 2000).

Near global spread of grass-dominated ecosystems is thought to have occurred by the mid-Miocene (Cerling et al., 1997; GPWG, 2001), which corresponds to the establishment of all the major lineages by about 20 to 25 Mya (Jacobs et al., 1999). Much of grass diversification has therefore occurred in the more recent history of the family (Kellogg, 2000). For instance, in a recent study using phytolith assemblage data, Stromberg (2005) suggests that open-habitat grasses had undergone great taxonomic diversification by the early Oligocene (34 Mya), but became ecologically dominant only by the late Oligocene-early Miocene (between 25 and 20 Mya) in North America. Using carbon isotopic composition of paleosols (Cerling et al., 1997) and fossil tooth enamel evidence (MacFadden and Cerling, 1994), the appearance of C_4 grasses, which constitute nearly half of the total species number of the family, is

thought to have occurred by 15 Mya and expanded globally by 7 to 5 Mya (Sage and Monson, 1999).

The mechanisms leading to present-day patterns of species diversity are of great interest in understanding both the evolutionary histories of living organisms (Paradis, 1997, 1998) and the formation and composition of modern ecosystems. Studies of fossils and their variations through time have provided the main evidence for assessing diversification in geographical space and geological time (Raup et al., 1973) but due to the limited fossil record, this approach is only well suited to a restricted range of phyla (mainly vertebrates, and some groups of invertebrates) (Paradis, 1997). Molecular phylogenetic approaches over the past two decades have offered alternative methods that can indirectly study the patterns and processes of diversification (Brooks and McLennan, 1991; Barraclough and Nee, 2001) but which generally require the sampling of all the species from within that group (Barraclough and Nee, 2001). Missing species reduce the sample size used for the reconstruction of speciation events and can introduce bias especially by removing the most recent speciation events (Nee et al., 1994). In the case of grasses, studies on this scale involve sampling approximately 10,000 species. Studying diversification patterns and processes using phylogenetic approaches for species-rich groups of organisms remains problematic and a comprehensive inference of the species phylogeny is needed (Hey, 1992; Nee et al., 1994; Sanderson and Donoghue, 1996; Paradis, 2003).

It is generally hard to determine detailed patterns of grass diversification from previous phylogenetic analyses because of a poor taxon sampling within the family (Kellogg, 2000; Hodkinson et al., 2007a). Even though there have been great advances in grass phylogenetics, few, or arguably no truly large and comprehensive phylogenetic trees of the family have been produced. However, one of the most recent and most comprehensive phylogenetic analyses of the grasses was done by the GPWG (2001). They produced trees with several well-supported major lineages that allowed classification of the family to be made at the subfamily level. However, it was limited in scope because the sample size only reached 8 % of all grass genera. Supertrees with higher sampling of the grasses have been produced (Salamin et al., 2002; Hodkinson et al., 2007b) but these did not incorporate branch lengths and did not include all the genera in the analyses. Partly because of this, few studies have tried to investigate patterns of diversification in the grasses using a phylogenetic

framework (Hodkinson et al., 2007a).

In addition, studies trying to date and characterize patterns of diversification are scarce and insufficiently detailed within the family (Bremer, 2002). Sampling all taxa to reconstruct comprehensive species-level phylogenetic trees is currently not practically possible mainly because of taxon availability, size of the sequencing effort, and the computationally demanding phylogenetic analyses that are required (Hodkinson et al., 2007b). In such cases, the use of an 'exemplar' approach (Yeates, 1995) (i.e. sampling one representative at any taxonomic rank such as genus, tribe, or subfamily) may be considered reliable as long as it includes the overall species diversity (i.e. species number within the group) at any taxonomic rank. Once a comprehensive phylogenetic framework has been achieved, two sources of information are relevant to the study of diversification rates: the topological distribution of species diversity and the temporal distribution of branching events (Chan and Moore, 2005). Topological methods allow the assessment of tree shape and imbalance and hence an assessment of diversification patterns across the lineages (Slowinski and Guyer, 1989a; Slowinski and Guyer, 1989b; Chan and Moore, 2005). Temporal methods offer greater power over topological ones because they incorporate phylogenetic branch lengths and can provide estimates of the timing of diversification (Sanderson and Donoghue, 1996). The disadvantages of topological methods in comparison to temporal ones, due to the lack of branch length information, might be counterbalanced to some extent by the advantage that topological analyses can more easily incorporate comprehensive taxonomic sampling; as would be the case with supertrees or with the compilation of molecular and morphological data to incorporate the overall diversity within the group under study.

This chapter aims to study the temporal and topological patterns of grass diversification and investigate processes leading to such diversification by (i) testing for shifts in diversification by inferring a complete generic level phylogenetic tree of grass genera including the total number of grass species to avoid sampling bias (a total of 815 exemplars (genera) that include 10,176 species), (ii) producing a comprehensive dated tree of the family, (iii) locating shifts in diversification in space and time, and (iv) trying to correlate these shifts with open versus closed habitat adaptation. To achieve this, a combination of phylogenetic trees, topological tests of

shifts in diversification, molecular dating and geographical and ecological data mapping onto phylogenetic trees were used.

3.2 Material and methods

3.2.1 Taxon and data sampling, phylogenetic analyses and testing shifts in diversification

The sampling list is shown in Appendix 2.1 of this thesis and includes 294 taxa. Two outgroups were chosen, *Elegia* (Restionaceae) and *Joinvillea* (Joinvilleaceae) because they are considered closely related to the grasses and, in the case of Joinvilleaceae, could be their sister group (Duvall and Morton, 1996). However, Ecdeiocoleaceae has also been resolved as sister to the grasses in recent analyses (Bremer, 2002; Rudall et al., 2005). The protocols used for DNA extraction, PCR amplification and sequencing cycles are described in the Material and Methods section of Chapter 2. In order to infer a complete generic level phylogenetic tree, we used data from the Grass Genera of the World DELTA database (Watson and Dallwitz, 1992) within which 436 morphological, anatomical, biochemical and ecological characters are coded for 798 genera. We considered 433 of these as we excluded three ecological attributes (salt-, light- and water-tolerance) for further character mapping.

We performed a MP analysis on a data matrix of all grass genera contained in Watson and Dallwitz (1992). The overall diversity included in our complete generic level phylogenetic tree was 10,176 species (each genus was coded by its number of recognized species). We used the 294 taxa phylogenetic tree (see Figure 5 of Chapter 2; or see details in Bouchenak-Khelladi et al., *submitted*) based on three plastid DNA regions (*rbcL*, *matK* and *trnL-F*), as a topological backbone to constrain the resulting tree(s) into the groups found in that analysis. The remaining 504 missing genera (798 total genera minus the 294 taxa sampled previously) were, in this inference placed according to unweighted and unordered morphological and anatomical characters. The phylogenetic analyses using MP of sequence data to generate the backbone tree of 294 taxa followed procedures given in the Material and Methods section of Chapter 2, as implemented in PAUP*4.0b10 (Swofford, 2002). For the complete generic level

tree (815 taxa), we ran MP analysis for 500 replicates of random addition sequence, using TBR swapping and saving no more than 10 trees per replicate. This analysis was done on the Trinity Centre for High Performance Computing Cluster (http://www.tchpc.tcd.ie). The strict consensus tree was used as a starting tree for TBR swapping while saving no more than 10,000 trees whose length is larger than the starting tree. This was done for successive runs until the resulting tree lengths did not vary. Robustness was assessed with the bootstrap (Felsenstein, 1985a) using 1,000 replicates of random addition sequence and TBR swapping with a limit of 50 trees for each replicate.

The strict consensus tree of 815 genera was then used to test for significant shifts in diversification. To test for shifts in diversification, we used statistical tests of phylogenetic tree shape and a whole-tree likelihood-based test (Moore et al., 2004). Polytomies resulting from the number of terminals (i.e. species) included in each genus were resolved by using an equal-rates Markov (ERM hereafter) random branching process (see below for details). Polytomies had to be removed in this way to allow calculations of tree imbalance, test for significant shifts in diversification and locate these shifts on the phylogenetic tree.

Statistical tests of phylogenetic tree shape. Phylogenetic tree-balance indices were used on the complete generic level phylogenetic tree to test the null hypothesis of constant diversification rates among the tree branches against the hypothesis of among-lineage variation. More specifically, two test statistics were used: the Colless *Ic* index (Colless, 1982; Mooers and Heard, 1997) and the logarithm of Chan and Moore's *M* index (Chan and Moore, 2002; Blum and Francois, 2006). The logarithm was computed as

$$S = S_{i=1}^{n-1} \log (N_i - 1) \qquad (1)$$

where the sum, *S*, ran over the internal nodes, *n* was the total number of taxa, and N_i was the number of descendants from node *i*. The *Ic* and the *S* statistics were usually well-defined for fully resolved topologies. To solve polytomies, we simulated the unknown topologies from the ERM random branching process (Yule, 1924; Kendall,

1948; Harding, 1971), which can be defined as a continuous-time, discrete-state, pure-birth Markov process in which the probability of a branching event is constant through time (Moore et al., 2004). This strategy was conservative as its aim was to preserve the constant rate hypothesis (i.e. this method could be considered more conservative than the same approach applied to trees without polytomies). We preferred using S in comparison to the statistic of Moore et al. (2004) because the tree was very large (10,187 species). For large trees the Gaussian distribution could be considered to be an accurate approximation for the distribution of S under the ERM model ($N(0,1)$ after standardization); see details in Blum and Francois (2006). This good statistical behavior made the values of S easier to interpret than the values of M. In a second stage, a modified S statistic was also considered in order to include the number of species in each genus as additional information.

Shifts in Diversification Rates. Diversification rate shifts were detected on the complete generic level phylogenetic tree (815 genera and 10,176 species) following the same approach as described in Moore et al. (2004) and Ree et al. (2005). This approach was developed in a likelihood framework that evaluated the relative fit of models with one- or two-rate parameters distributed over different parts of a three-taxon tree. The key quantity to compute was a likelihood ratio that represented the relative fit of the one- and two-rate parameter models to the observed diversity partition. This was assessed by the difference in the natural logarithm of the respective likelihood values in homogeneous and heterogeneous diversification rate models. Here we used mathematical arguments, developed by O. Francois (TIMB, INPG-ENSIMAG, France), and now implemented in APTREESHAPE (Bortolussi et al., 2006), taken from the branching process theory to derive a simpler formula for the logarithm of the likelihood ratio. If the ERM branching process is initiated with a single species and allowed to run for a period of time t with a branching rate λ, the probability of realizing n species is according to Harris (1964):

$$P(n \mid \lambda, t) = e^{-\lambda t} (1 - e^{-\lambda t})^{n-1} \qquad (2)$$

Accordingly, the probability of realizing n species partitioned between the left and the right descendants of a single node with ℓ and r species, respectively, under the

heterogeneous two-rate parameter model H_A is

$$P(\ell,r \mid H_A)= P(\ell|\lambda,t)\, P(r \mid l_s, t) \,/\, S_{i=1}^{n-1}\; P(i|\lambda, t)\, P(n - i \mid l_s, t) \qquad (3)$$

where λ is the ancestral rate and l_s the shifted rate. For notation convenience, we assumed that the shift occurred on the right sister branch after the speciation event, but the derivation of the symmetric formula poses no conceptual difficulties. Now, introducing the parameters

$$q= 1- \exp(-\lambda t), \qquad q_s= 1- \exp(-l_s\, t) \qquad (4)$$

Equation (2) can be rewritten as

$$P(\ell,n - \ell| H_A)= (q^{l} q^{n-l}{}_s) \,/\, S_{i=1}^{n-1}\; q^{l} q^{n-l}{}_s \qquad (5)$$

Using $Q= q / q_s$ and assuming $\lambda \neq l_s$, we obtain that

$$P(\ell,n - \ell| H_A)= Q^{l} (1-Q) / (1-Q^{n}) \qquad (6)$$

Under the null-model with uniform branching probability (H_o: $\lambda = l_s$), we recover the classical result of Harding (1971)

$$P(\ell,n - \ell| H_o)= 1 / (n - 1) \qquad (7)$$

According to equations (6) and (7), the log-likelihood ratio $LR_{HA: HO}$ can be written as

$$LR_{HA: HO} = \log P(\ell,n - \ell| H_A) - \log P(\ell,n - \ell \mid H_o)$$

which is equal to

$$LR_{HA: HO}= \ell \log Q - \log (1 - Q^{n}) + \log (n-1) + C \qquad (8)$$

with C a constant independent on ℓ and n. The P-values of the likelihood ratio test can be easily computed by using Harding's result (i.e. sampling ℓ from the uniform distribution, allowing us to implement the computation of P-values efficiently). Finally, diversification rate shifts were tested on the basis of the Δ_1 statistic, which involved computing likelihood ratios at two nested levels (Moore et al., 2004). In order to attenuate the effect of absence of correction for multiple testing, we assessed shifts that led to a minimum of a 100-fold increase in the ancestral rate. The type I errors were also fixed to a low level (1 %). P-values were computed using 10,000 Monte-Carlo replicates of Δ_1 under the ERM model. All tests were performed using the *R* computer package APTREESHAPE (Bortolussi et al., 2006). We compiled the results of the likelihood-based method implemented in APTREESHAPE (with $P \leq 0.01$) and pinpointed them onto the complete generic level phylogenetic tree.

3.2.2 Molecular dating

Molecular dating was done on a 110 taxa phylogenetic tree, inferred with three DNA plastid regions: *rbcL*, *matK* and *trnL-F*, because molecular dating is best done on matrices with a limited amount of missing data especially for likelihood-based methods. This tree was found congruent with the 294 taxa phylogenetic tree (see Results section of Chapter 2 of this thesis, or see details in Bouchenak-Khelladi et al., submitted) used as a topological backbone for inferring the complete generic level phylogenetic tree shown in Figures 3.1, 3.2 and 3.3. The methods used to construct the 110 taxa tree are described in the Material and Methods section of Chapter 2 and in Bouchenak-Khelladi et al. (*submitted*). However, three taxa were added to DataSet I (see Material and Methods section of Chapter 2) in order to optimize the taxonomic representation of our sampling (i.e. representatives of Anomochlooideae and Aristidoideae were added). Bayesian methods (Thorne, 1998; Kishino et al., 2001; Thorne and Kishino, 2002) were used to estimate divergence times using MULTIDIVTIME (available from J. Thorne, North Carolina State University, http://stagen.ncsu.edu/thorne/multidivtime.html). This approach relaxes the assumption of a strict molecular clock with continuous autocorrelation of substitution rates across the phylogenetic tree. We performed a 'partitioned approach' in which

branch lengths were estimated from each dataset separately to account for differences in substitution processes across plastid regions. First, ESTBRANCHES was run to estimate branch lengths from the data as well as from the fixed tree topology using the F84 (Felsenstein, 1984) model of sequence evolution with rates allowed to vary among sites following a discrete gamma distribution with four categories along with their variance-covariance matrix. Parameter values for the F84 + Γ were estimated using the BASEML program in PAML version 3.14 (Yang, 2000).

Next, the outgroup taxa (*Elegia* and *Joinvillea*) were pruned from the tree, and MULTIDIVTIME was used to estimate the prior and posterior ages of branching events, their standard deviations, and the 95 % credibility intervals via MCMC. The Markov Chain was run for 1,000,000 generations and sampled every 1,000 generations after an initial burn-in of 100,000 cycles. To check for convergence of the MCMC, analyses were run from three different starting points. The following priors were used in these analyses: 80 Mya (SD=20.0 Mya) for the expected time between the tips and root; the rate at the root node (rtrate parameter) in substitutions per site per million years was calculated as the median of all rates from the ingroup root to the tips; the parameter that determines the magnitude of autocorrelation per million years was calculated as 1 over the root to tips parameter (rttm) (as suggested by J. Thorne); and 250 Mya for the largest value of the time units between the root and the tips. We set all standard deviations in priors equal to the prior mean, except the time of the root to tips parameter (rttm).

At least some reference fossil records are necessary for dating, and it is desirable to find several reference fossils that may be attached to the lower nodes of the tree. Despite the fact that the fossil record of the grasses is extremely poor, Crepet & Feldman (1991) identified the earliest remains of true grasses as typical grass spikelets from the Paleocene-Eocene boundary (55 Mya). The spikelets have two-glumes and two flowers and are characteristic of the BEP and PACCAD clades in Poaceae (Crepet and Feldman, 1991; GPWG, 2001). This date was then attached above the root node, at the ancestral node of the BEP + PACCAD clade, with a minimum age of 50 Mya and a maximum age of 55 Mya. Also, two fossil grasses were identified and dated by Dugas & Retallack (1993) in southwestern Kenya as members of subfamilies Panicoideae and Chloridoideae. Those have been dated at about 14 Mya (middle Miocene). The first fossil was identified as a member of

Panicoideae to the genus *Cleistochloa* based on the occurrence of dumbbell-shaped silica bodies and stomata with parallel to low-domed subsidiary cells (Watson and Dallwitz, 1989). The latter feature is highly significant for taxonomic classification (Metcalfe, 1960). The second fossil was identified as a member of Chloridoideae based on the combination of round silica bodies and triangular stomatal subsidiary cells (Dugas and Retallack, 1993). The authors narrowed the identification to *Distichlis* (Dugas and Retallack, 1993). These dates were attached according to the position of these genera in the complete generic level phylogenetic tree, and then were pinpointed on the 110 taxa phylogenetic tree with a minimum age of 12 Mya and a maximum age of 16 Mya for both. We chose to calibrate these nodes with maximum ages using the fossils described above. It is not clear how these assumptions are consistent with the fossil record. The nodes to calibrate were chosen according to the shortest branch length found between the 110 taxa tree and in the complete generic level phylogenetic tree where the point of divergence of *Cleistochloa* and *Distichlis* occurred (see a similar approach taken by Bininda-Emonds et al., 1999). The root node was calibrated between 60 and 100 Mya based on a previous dating of the origin of the family at about 75 Mya (Bremer, 2002).

3.2.3 Geographical and ecological data mapping

In order to locate shifts in diversification in geographical space, we wanted to consider the center of origin of the grass family and then infer a possible paleo-biogeographic scenario for dispersal routes from this area. Many grass genera sampled exhibit a cosmopolitan distribution, which makes inferences of vicariance and/or dispersion events difficult and several combinations of events, if not all, could be optimized. We tested for the most likely center of origin of the grasses using Bremer's Ancestral Area Analysis (Bremer, 1992). We used this method with the strict consensus tree analysis and tested for the possible center of origin with the strict consensus tree (complete generic level phylogenetic tree) as well as with ten trees sampled randomly among the 100 equally most parsimonious. We then inferred a possible paleo-biogeographic scenario of dispersal routes by mapping geographical distribution data on the tree. Bremer's method (1992) is based on comparing the number of gains and losses of an area on the cladogram. Every event of colonization is a gain. It is assumed that the more gains (or colonization events) that are required for

an area, the less likely it is that it has been colonized and thus the more likely it is that it was part of the ancestral area. Alternatively, all the differences in present day geographical distributions could have arisen through losses by extinction or fragmentation followed by vicariance, thus we may expect that with the more losses for an area, the less likely the area is to be ancestral (Bremer, 1992). The higher the ratio between gains and losses is, the more likely it is that an area is part of the ancestral area. Therefore, the most repeated area closest to the base of the tree is selected as being part of the ancestral area. In this study, we considered a set of geographical units and tested for the most likely ancestral areas. The set comprised seven different areas: (1) Africa, (2) Asia, (3) Australia, New Caledonia and the Pacific archipelagos, (4) Tropical south-east Asia, (5) Europe, (6) North America, Caribbean and (7) South America (see Figure 3.1 for details). All taxa were coded for presence or absence in each unit area of the genus distribution. To reconstruct paleo-biogeographic scenarios, we used character optimizations using MACCLADE v.4.08 (Maddison and Maddison, 2005) with unweighed and unordered characters. In case of equivocal tracings, both DELTRAN (delay changes) and ACCTRAN (accelerated changes) were used to resolve tracings. We mapped each geographical area (coded as presence or absence of taxa) independently on the tree. It was not possible to use other methods, such as DIVA (Ronquist, 1996) because the sample size of our complete generic level phylogenetic tree is not manageable by DIVA (Ronquist, 1996), which can handle no more than 180 taxa (see DIVA 1.1 User's manual; Ronquist, 1996).

We used one ecological attribute 'light tolerance', coded for each genus as a measure for adaptation to open or closed habitats, in our studies of diversification by mapping it onto our complete generic level phylogenetic tree to check if contrasting traits occurred at nodes for which we detected shifts in diversification. In the case of equivocal tracings, both DELTRAN and ACCTRAN were used to resolve tracings using MACCLADE v.4.08 (Maddison and Maddison, 2005).

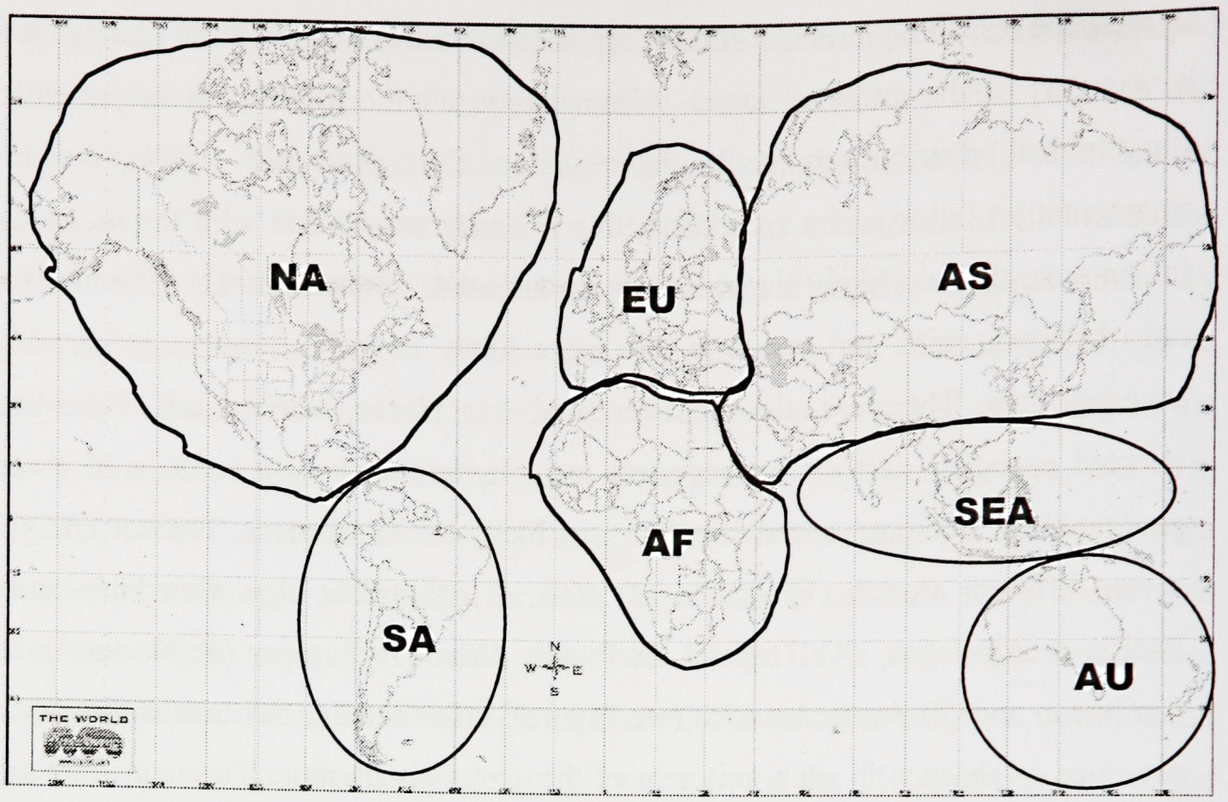
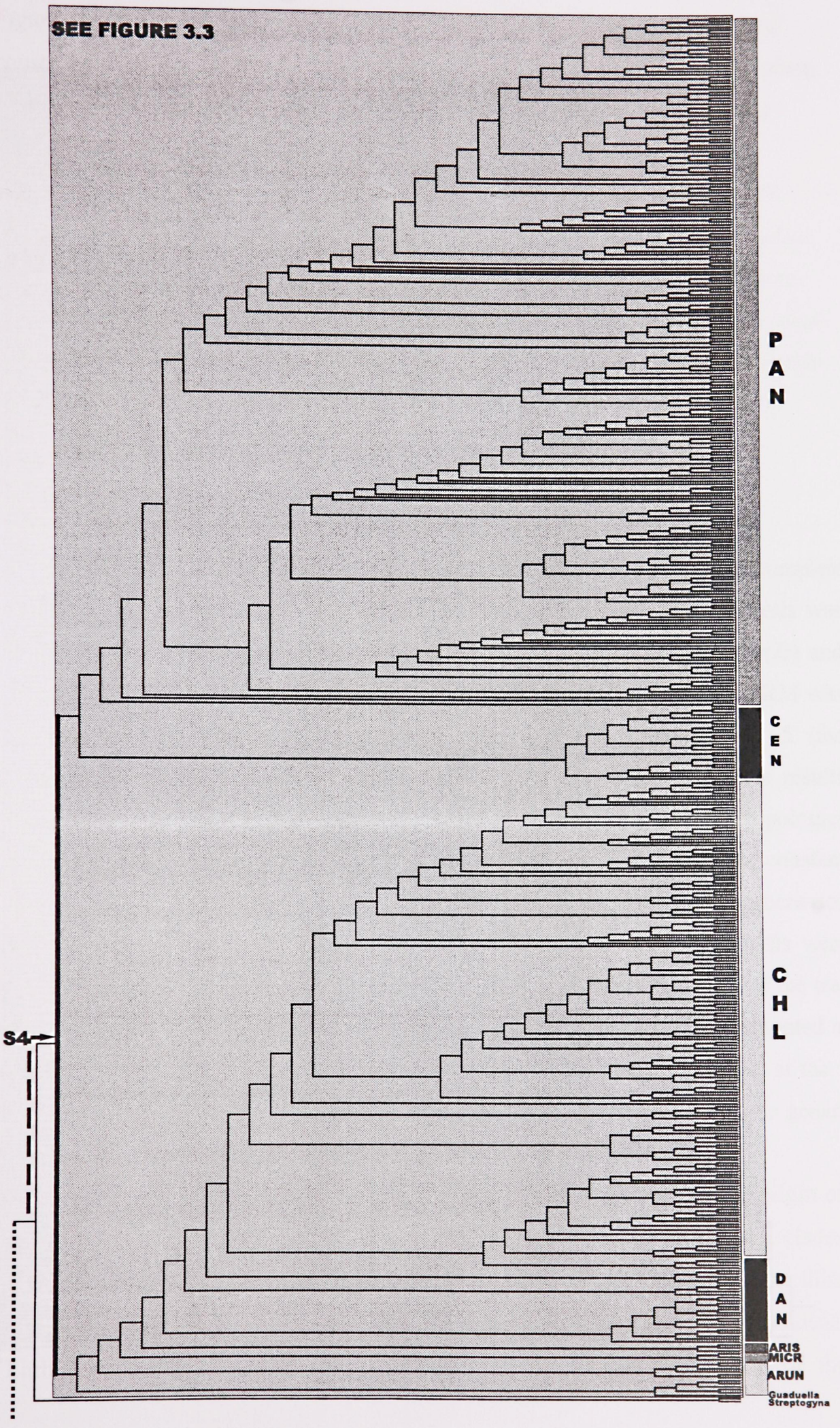


Figure 3.1 Distribution of the geographical areas coded for all grass genera. AF: Africa, AS: Asia, AU: Australia, EU: Europe, NA: North America, SA: South America and SEA: South-East Asia (map taken from: http://www.amaps.com/mapstoprint/)

3.3 Results

3.3.1 Complete generic level phylogenetic tree

Of the 433 morphological and anatomical characters, 374 were parsimony informative, 54 were uninformative and five constant. After a series of TBR swapping, the 100 shortest trees found had 10,827 steps with a Consistency Index (CI) of 0.063 and a Retention Index (RI) of 0.605. The CI is very correlated with the dataset size and is found very low. The phylogenetic tree is summarized in Figure 3.1 and is also available online in the TREEBASE website (http://www.treebase.org/treebase/) with the reference SN3293. The two major clades found are shown in more detail in Figures 3.2 and 3.3. No bootstrap support was found for all the clades in the tree.


SEE FIGURE 3.3
PAN
CEN
CHL
DAN
ARIS
MICR
ARUN
Guaduella
Streptogyna
S4

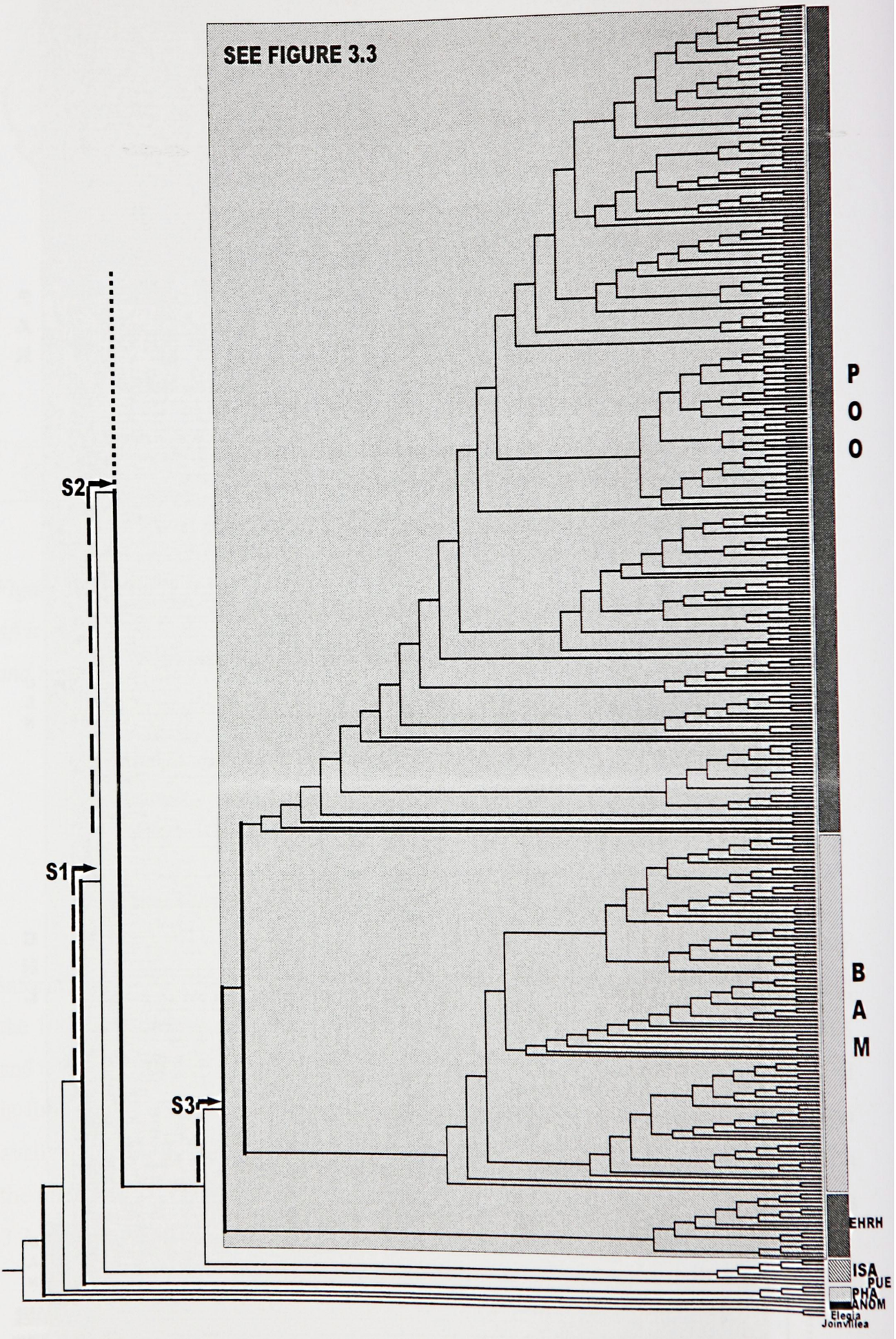

SEE FIGURE 3.3
S2
S1
S3
POO
BAM
EHRH
ISA
PUE
PHA
ANOM
Elegia
Joinvillea

Figure 3.2 One of 100 equally most parsimonious trees found from a topological constraint analysis. Dashed arrows indicate an increase in diversification rate along the branch. Bold vertical bars indicate congruent nodes with the dated 110 taxa phylogenetic tree shown in Figure 3.4. S1, S2, S3 and S4 indicate increase in diversification rates (see Table 3.1 for details). ANOM: Anomochlooideae, ARIS: Aristidoideae, ARUN: Arundinoideae, BAM: Bambusoideae, CEN: Centothecoideae, CHL: Chloridoideae, DAN: Danthonioideae, EHRH: Ehrhartoideae, ISA: Isachne clade, MICR: Micraira clade, PAN: Panicoideae, PHA: Pharoideae, POO: Pooideae, PUE: Puelioideae. Elegia and Joinvillea are outgroups. (Morphological data matrix is available on the CD accompanying this thesis; folder "Morphological_data")

3.3.2 Shifts in diversification rates

Statistical tests based on tree-balance measures were applied to the complete generic level phylogenetic tree, and the constant diversification rate hypothesis was strongly rejected. The tests based on the *Ic* and *S* statistics yielded $P<10^{-4}$ (*Ic*) and $P<10^{-10}$ (*S*). The *Ic* was estimated to 14,421 (the expected value under the ERM was computed as *Ic*= 4,553), and the standardized value of *S* was computed to 19.65 (the values expected under the ERM model have the *N(0,1)* distribution). These results showed that the tree contained levels of imbalance that differed from ERM topologies significantly. In addition, the *S* index was used to test the null hypothesis of constant diversification rates given the number of species in each genus. The *P*-value increased slightly but the test was still significant at $P<10^{-4}$. Diversification rate shifts were located using the D_1 test statistic (Moore et al., 2004) at each internal node of the tree that detected 100-fold shifts in diversification. Among the 813 *P*-values computed at each internal node except the root, fifteen were detected as being significant at the 1 percent level (Figures 3.2, 3.3 and 34). The sister clade's description and the genera involved in shifts in diversification are described in Table 3.1.

The first two shifts in diversification were found successively at the origin of the core grasses, defined as the ancestral node of the BEP and PACCAD clades excluding the early-diverging lineages (S1 and S2: Figure 3.2; Table 3.1). The third one was found at the origin of the BEP clade that includes Bambusoideae, Ehrhartoideae and Pooideae (S3: Figure 3.3; Table 3.1). At the origin of the

Bambusoideae, a significant shift in diversification was found (S5: Figure 3.3; Table 3.1).

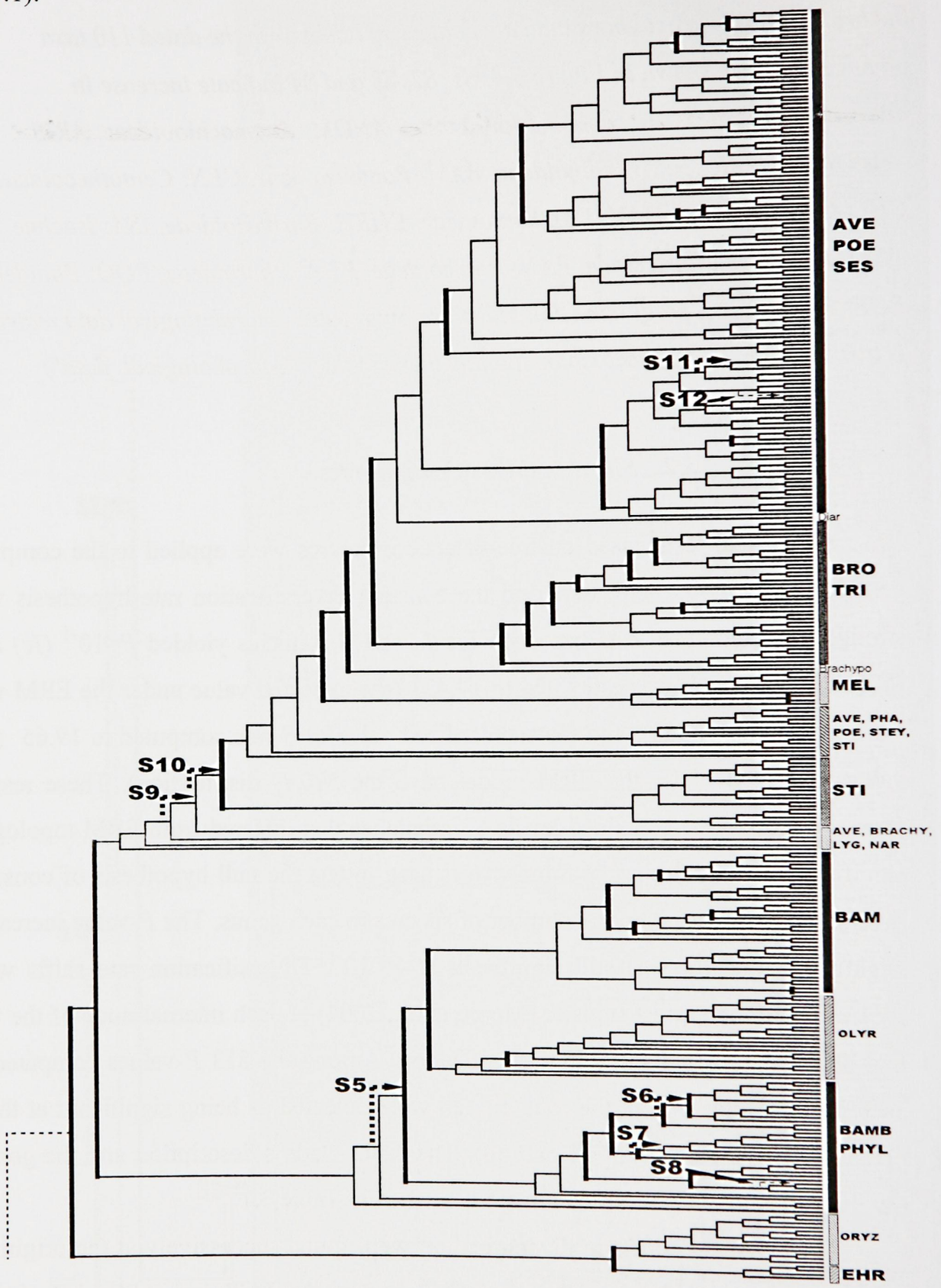


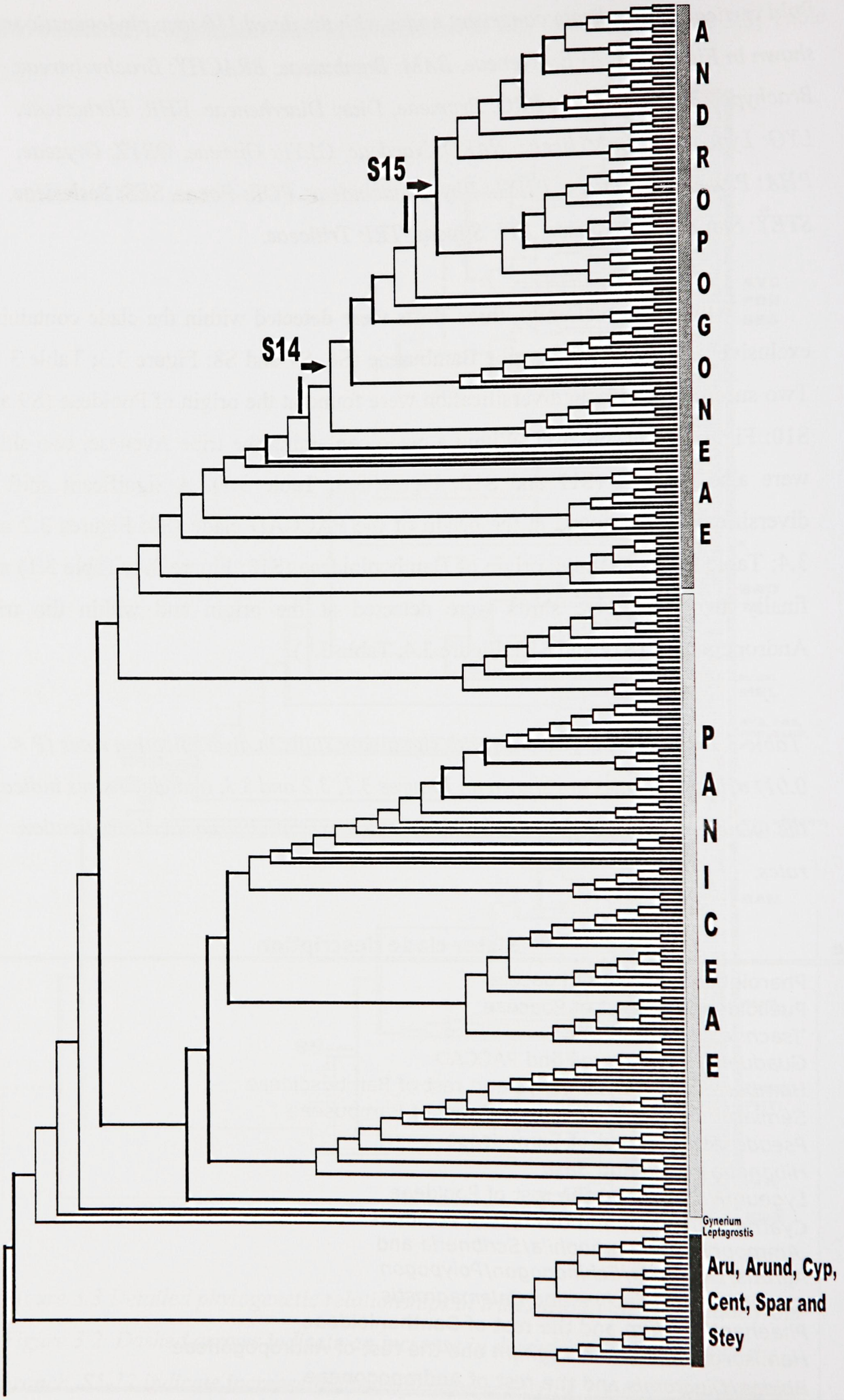
Figure 3.3 Detailed phylogenetic relationships of tribes within the BEP clade from Figure 3.2. Dashed arrows indicate an increase in diversification rate along the branch. S5-12 indicate increase in diversification rates (see Table 3.1 for details).

Bold vertical bars indicate congruent nodes with the dated 110 taxa phylogenetic tree shown in Figure 3.5. AVE: Aveneae, BAM: Bambuseae, BRACHY: Brachyelytreae, Brachypo: Brachypodieae, BRO: Bromeae, Diar: Diarrheneae, EHR: Ehrharteae, LYG: Lygeae, MEL: Meliceae, NARD: Nardeae, OLYR: Olyreae, ORYZ: Oryzeae, PHA: Phaenospermateae, PHYL: Phyllorhachideae, POE: Poeae, SES: Seslerieae, STEY: Steyermarkochloeae, STI: Stipeae, TRI: Triticeae.

Within this subfamily, three shifts were detected within the clade containing exclusively members of the tribe Bambuseae (S6, S7 and S8: Figure 3.3; Table 3.1). Two successive shifts in diversification were found at the origin of Pooideae (S9 and S10: Figure 3.3; Table 3.1). Within a clade containing the tribe Aveneae, two shifts were also detected (S11 and S12: Figure 3.3; Table 3.1). A significant shift in diversification was found at the origin of the PACCAD clade (S4: Figures 3.2 and 3.4; Table 3.1), one at the origin of Danthonioideae (S13: Figure 3.4; Table 3.1) and finally two successive shifts were detected at the origin and within the tribe Andropogoneae (S14 and S15: Figure 3.4; Table 3.1).

Table 3.1 Poaceae sister clades with significant shifts in diversification rates ($P < 0.01$) using Δ_1. Nodes are shown on Figures 3.1, 3.2 and 3.3, and clade sizes indicate the number of species of the sister clades which exhibit differential diversification rates.

Node	Sister clade description	Clade sizes
S1	Pharoideae and rest of Poaceae	20/10,18
S2	Puelioideae and rest of Poaceae	6/10,17
S3	'*Isachne* clade' and BEP	114/4,61
S4	*Guaduella/Streptogyna* and PACCAD	10/5,44
S5	*Humbertochloa/Phyllorachys* and rest of Bambusoideae	3/976
S6	*Semiarundinaria/Shibataea* and rest of Bambuseae	9/95
S7	*Pseudosasa* and rest of Bambuseae	8/91
S8	*Hibanobambusa* and *Sasa*	1/50
S9	*Lygeum/Nardus* and the rest of Pooideae	2/3,506
S10	*Cyathopus/Milium* and the rest of Pooideae	5/3,501
S11	*Ammophila/Calammophila/Scribneria* and *Agrotis/Deyeuxia/Echinopogon/Polypogon*	4/265
S12	*Sinochasea/Triplachne* and *Calamagrostis*	2/230
S13	*Phaenanthoecium* and the rest of Danthonioideae	1/250
S14	*Hemisorghum/Pseudosorghum* and the rest of Andropogoneae	4/896
S15	*Bhidea/Diectomis* and the rest of Andropogoneae	4/728


S15
S14
ANDROPOGONEAE
PANICEAE
Gynerium
Leptagrostis
Aru, Arund, Cyp,
Cent, Spar and
Stey

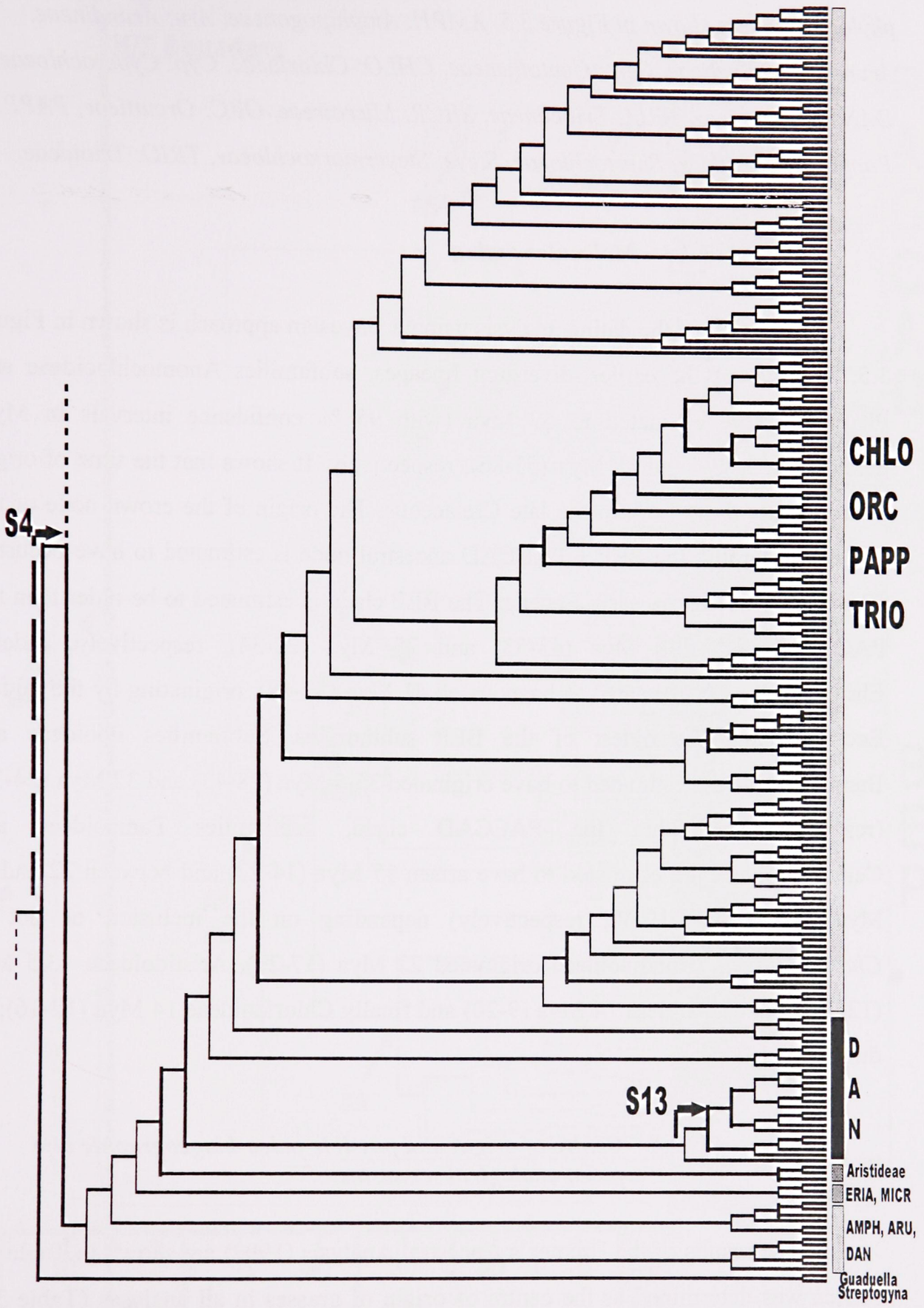


Figure 3.4 Detailed phylogenetic relationships of tribes within the PACCAD clade from Figure 3.2. Dashed arrows indicate an increase in diversification rate along the branch. S4, S13, S14 and S15 indicate increase in diversification rates (see Table 3.1 for details). Bold vertical bars indicate congruent nodes with the dated 110 taxa

phylogenetic tree shown in Figure 3.5. AMPH: Amphipogoneae, Aru: Arundineae, Arund: Arundinelleae, Cent: Centotheceae, CHLO: Chlorideae, Cyp: Cyperochloeae, DAN: Danthonieae, ERIA: Eriachneae, MICR: Micraireae, ORC: Orcuttieae, PAPP: Pappophoreae, Spar: Spartochloeae, Steye: Steyermarkochloeae, TRIO: Trioideae.

3.3.3 Molecular dating

The result of the dating analysis using a Bayesian approach is shown in Figure 3.5. The age of the earliest-diverging lineages, subfamilies Anomochlooideae and Pharoideae, are estimated to 67 Mya (with 95 % confidence intervals in Mya, hereafter: (55-83)) and 68 Mya (53-86), respectively. It shows that the time of origin of the family dates back to the late Cretaceous. The origin of the crown node of the rest of the family, the BEP + PACCAD ancestral node is estimated to have occurred 52 Mya (50-55) in the early Eocene. The BEP clade is estimated to be older than the PACCAD clade: 48 Mya (43-53) and 28 Mya (22-34), respectively. Indeed, Ehrhartoideae are estimated to have arisen 41 Mya (32-48), originating by the middle Eocene, being the oldest of the BEP subfamilies. Subfamilies Pooideae and Bambusoideae are estimated to have originated 35.5 Mya (28-43) and 32 Mya (24-39) (respectively). Within the PACCAD clade, subfamilies Panicoideae and Centothecoideae are estimated to have arisen 17 Mya (14-22) and between 22 and 24 Mya (17-28 and 19-30, respectively) depending on the inclusion or not of *Chasmanthium*. Arundinoideae originated 22 Mya (17-29), Aristidoideae 13.5 Mya (12-24), Danthonioideae 14 Mya (9-20) and finally Chloridoideae 14 Mya (12-16); all during the Miocene.

3.3.4 Centre of origin and possible paleo-biogeographic and paleo-ecological scenarios

The results of the Bremer's Ancestral Analysis (1992) are shown in Table 3.2. Africa was determined as the centre of origin of grasses in all analyses (Table 3.2). Based on Africa as the ancestral area, the most plausible biogeographic pathway, using both ACCTRAN and DELTRAN resolving options, is illustrated in Figures 3.5 and 3.6. The earliest-diverging lineages diverged to South America possibly by vicariance, between 65 and 70 Mya, before or during the break-up of Africa and South America

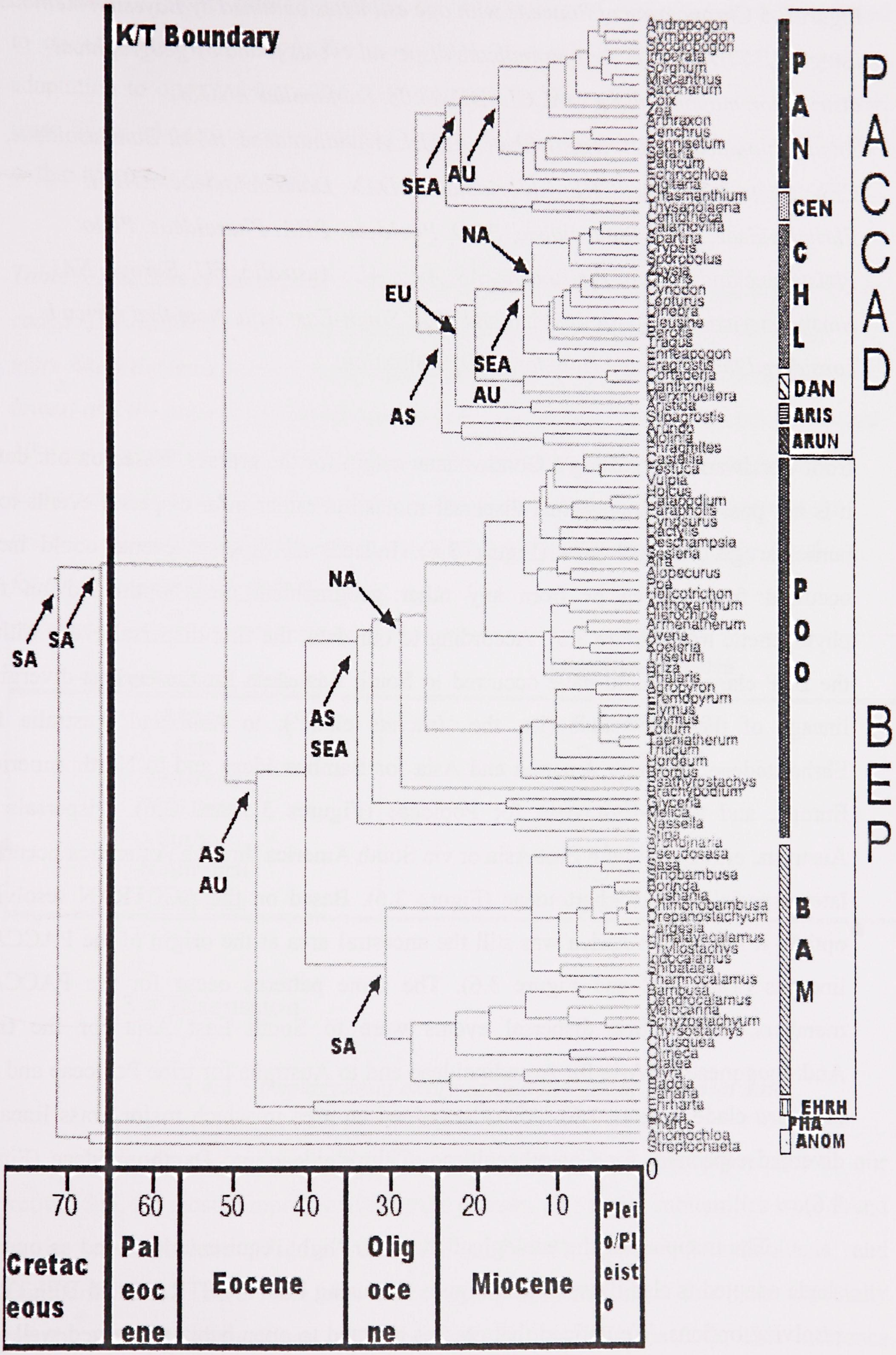

K/T Boundary
Andropogon
Cymbopogon
Spodiopogon
Imperata
Sorghum
Miscanthus
Saccharum
Coix
Zea
Arthraxon
Cenchrus
Pennisetum
Setaria
Panicum
Echinochloa
Digitaria
Chasmanthium
Thysanolaena
Centotheca
Calamovilfa
Spartina
Crypsis
Sporobolus
Zoysia
Chloris
Cynodon
Lepturus
Dinebra
Eleusine
Perotis
Tragus
Enneapogon
Eragrostis
Cortaderia
Danthonia
Merxmuellera
Aristida
Stipagrostis
Arundo
Molinia
Phragmites
Castellia
Festuca
Vulpia
Holcus
Catapodium
Parapholis
Cynosurus
Dactylis
Deschampsia
Sesleria
Aira
Alopecurus
Poa
Helictotrichon
Anthoxanthum
Hierochloe
Arrhenatherum
Avena
Koeleria
Trisetum
Briza
Phalaris
Agropyron
Elymus
Lolium
Taeniatherum
Triticum
Hordeum
Bromus
Psathyrostachys
Brachypodium
Glyceria
Melica
Nassella
Stipa
Arundinaria
Pseudosasa
Sasa
Sinobambusa
Borinda
Yushania
Chimonobambusa
Drepanostachyum
Fargesia
Himalayacalamus
Phyllostachys
Indocalamus
Shibataea
Thamnocalamus
Bambusa
Dendrocalamus
Melocanna
Schyzostachyum
Thyrsostachys
Chusquea
Olmeca
Otatea
Olyra
Buddia
Pariana
Ehrharta
Oryza
Pharus
Anomochloa
Streptochaeta
SEA
AU
NA
EU
SEA
AU
AS
NA
AS
SEA
AS
AU
SA
SA
SA
P
A
N
CEN
C
H
L
DAN
ARIS
ARUN
P
O
O
B
A
M
EHRH
PHA
ANOM
PACCAD
BEP
0
70
60
50
40
30
20
10
Cretaceous
Paleocene
Eocene
Oligocene
Miocene
Pleio/Pleisto

Figure 3.5 Chronogram of Poaceae with age estimates obtained by Bayesian methods using MULTIDIVTIME. Arrows indicate dispersal events found by geographical distribution mapping using MACCLADE v.4.08. Subfamilies ANOM: Anomochlooideae, ARI: Aristidoideae, ARUN: Arundinoideae, BAM: Bambusoideae, CEN: Centothecoideae, CHL: Chloridoideae, DAN: Danthonioideae, EHRH: Ehrhartoideae, PAN: Panicoideae, POO: Pooideae, PHA: Pharoideae. Pleio: Pleistocene and Pleisto: Pleistocene. AS: Asia, AU: Australia, EU: Europe, NA: North America, SA: South America and SEA: South-East Asia. Note that Africa is considered the centre of origin for this optimization.

from Gondwana suggesting a Gondwnanan origin for the grasses. Based on our data, it is not possible to infer actual dispersal routes but rather infer dispersal events to a particular geographical area (Figure 3.6). Indeed, all dipersal events could have occurred from Africa or from any other geographical areas optimized on the phylogenetic tree (Figure 3.5). According to our data, the first dispersal event within the BEP clade seems to have occurred to South East Asia for the earliest-diverging lineage of the BEP clade (i.e. the '*Isachne* clade'), to Asia and Australia for Ehrhartoideae, to South America and Asia for Bambusoideae and to North America, Europe, and South-East Asia for Pooideae (Figures 3.5 and 3.6). Dispersals to Australia, either via South East Asia or via South America through Antarctica occurred later especially for Ehrhartoideae (Figure 3.6). Based on the ACCTRAN resolving option, it seems that Africa was still the ancestral area at the origin of the PACCAD lineages (Table 3.2 and Figure 3.6). The same patterns occur for the PACCAD members, as the first dispersal events were to South East Asia for the tribe Andropogoneae, to Asia for Arundinoideae and to Australia for tribe Paniceae and the *Micraira* clade (Figure 3.6). Africa remained an area in which major grass lineages diverged especially for Centothecoideae, Chloridoideae and Danthonioideae (Figure 3.6).

The mapping of the ecological character 'light requirement' coded as open or shade adapted is also illustrated in Figure 3.6, using both ACCTRAN and DELTRAN resolving options. We noticed that grasses adapted to open habitat diverged well after the origin of the family and during or after the origin of the two major clades (Figure 3.6). Indeed, the ancestors of Pooideae seem to have been adapted to open habitats.

With the ACCTRAN resolving option, the '*Isachne* clade' seems to have been adapted to open habitat well before the divergence of Pooideae (Figure 3.6), making the adaptation to open habitat older. The ancestors at the origin of the PACCAD clade were likely adapted to open habitats before the divergence of the major subfamilies within this clade (Figure 3.6).

Table 3.2 Results of the Bremer's Ancestral Analysis for the seven geographical areas coded. The likelihood ratio is calculated as gains/losses, the higher the ratio, the more likely the area belongs to the ancestral area of origin. Ranges indicate the lowest and the highest likelihood ratio found in all analyses (strict consensus tree and 10 trees sampled randomly within the equally most parsimonious). Africa has the highest likelihood ratio and may therefore be the most likely centre of origin of the grasses. (Geographical data matrix is available on the CD accompanying this thesis; folder "Geographical_data")

Geographical Areas	Likelihood Ratio
Africa	**0.864-0.883**
Asia	0.622-0.645
Australia	0.481-0.507
S.E.Asia	0.505-0.522
Europe	0.493-0.517
N.America	0.534-0.562
S.America	0.596-0.614

3.4 Discussion

3.4.1 Phylogenetic analyses and the use of temporal data

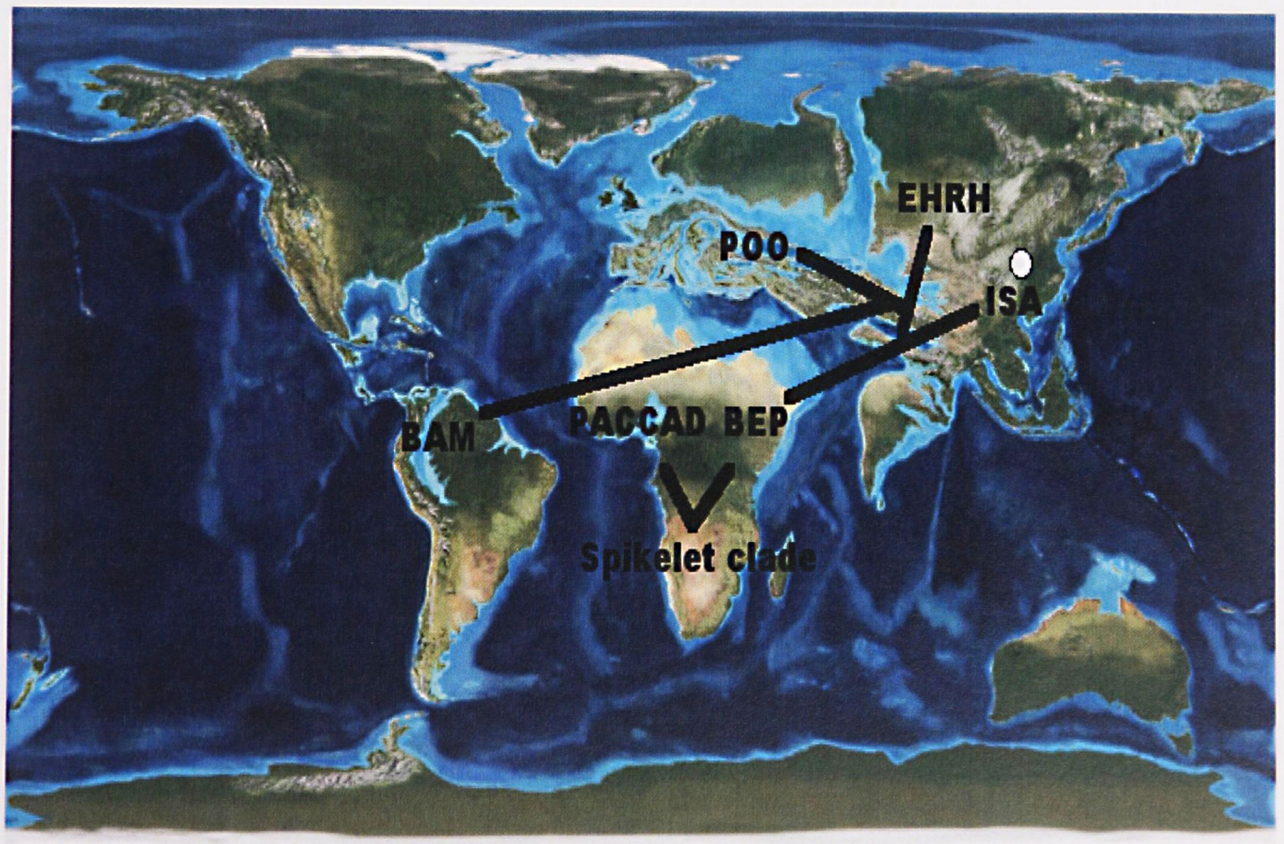
In terms of taxon numbers, the trees presented in this paper are, to our knowledge, the most comprehensive for the grasses. The major subfamilies were found to be monophyletic (Figures 3.2, 3.3 and 3.4), except Arundinoideae and Danthonioideae. There were also some odd placements of a few taxa especially members of Centothecoideae and tribe Arundinelleae (Figure 3.4). This inference allowed us to test for diversification shifts by removing bias with regards to incomplete sampling and missing taxa.

The lack of temporal data using such phylogenetic tree was counterbalanced by the use of molecular dating based on a 110 taxa subtree (Figures 3.2-3.4). It enabled us to provide a timescale thanks to the congruent nodes between the complete generic level and the dated phylogenetic trees.

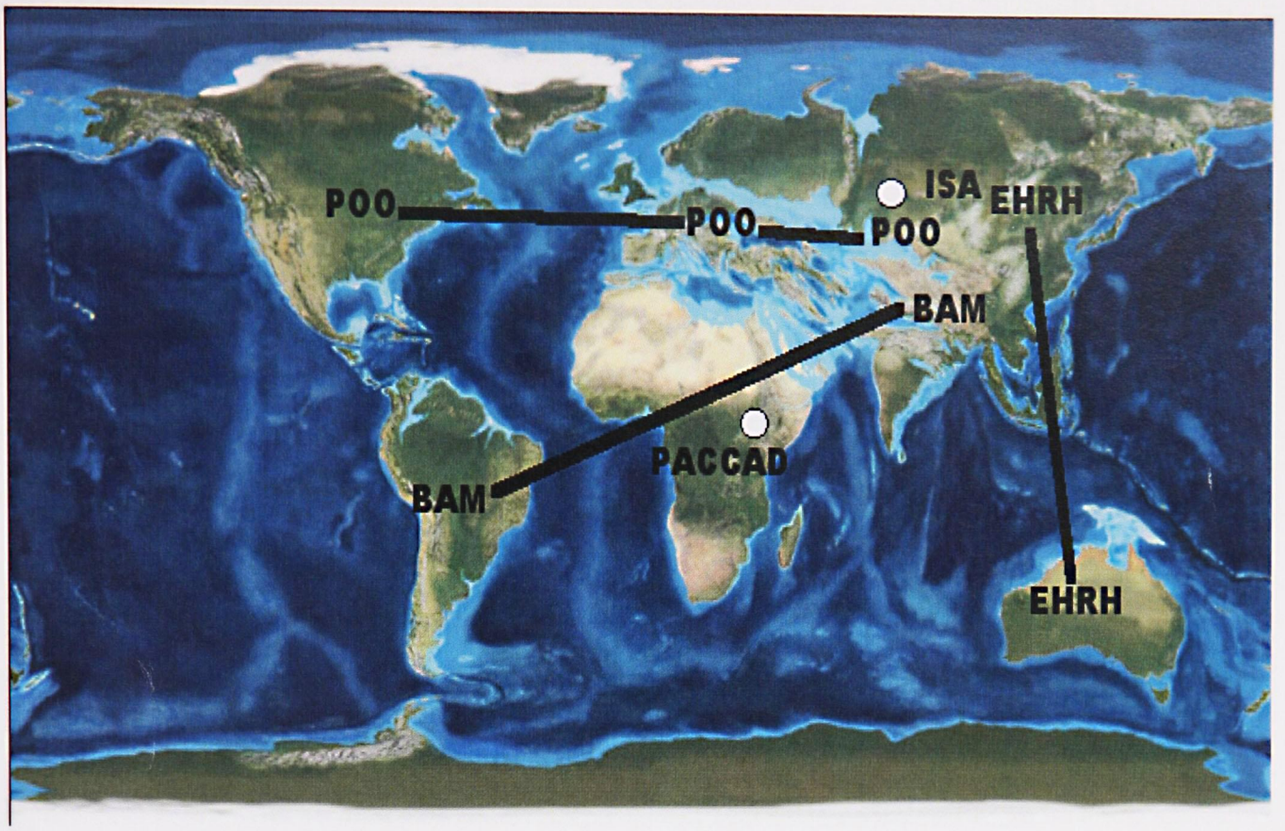
A


ORIGIN of
POACEAE
EDLs

B


EHRH
POO
ISA
BAM
PACCAD BEP
Spikelet clade

C



D

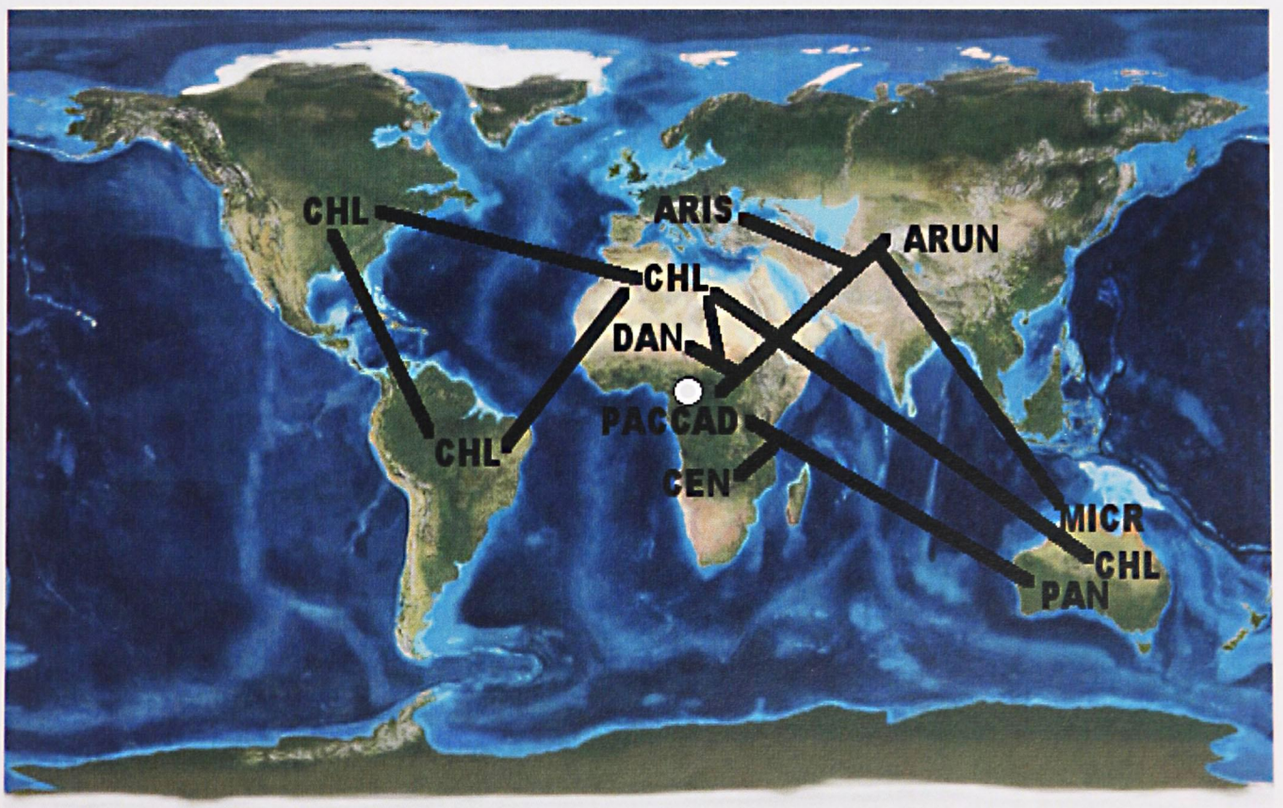


Figure 3.6 Biogeographical history of grasses (Poaceae) showing dispersal events of the main lineages, with geographical areas optimized using MACCLADE v.4.0. (A): Around ***65 Mya****, African origin of Poaceae; EDLs: Early-Diverging Lineages. (B):*

*Around **50 Mya** (Eocene), Spikelet Clade, BEP and PACCAD according to the GPWG (2001), BAM: Bambusoideae, EHRH: Ehrhartoideae, ISA: 'Isachne clade', POO: Pooideae. (C): Around **35 Mya** (Oligocene). (D): Around **20 Mya** (Miocene), ARIS: Aristidoideae, ARUN: Arundinoideae, CHL: Chloridoideae, CEN: Centothecoideae, DAN: Danthonioideae, MICR: 'Micraira clade' and PAN: Panicoideae. White circles indicate the appearance of adaptation to open-habitat along the branches (maps taken from http://jan.ucc.nau.edu/~rcb7/).*

3.4.2 Grass evolution and diversification scenario

According to our molecular dating (Figure 3.5), the time of origin of grasses is estimated to the late Cretaceous (around 72 Mya), before the Cretaceous/Tertiary extinction event (K/T boundary). It is believed that terrestrial plants passed through the K/T boundary, with only minor taxonomic richness in comparison to today (Macleod et al., 1997). Understory vegetation may have survived the event (Sweet, 2001). Our results indicate that the grasses may have originated in Africa (Table 3.2), suggesting a Gondwanan origin of the family. This general Gondwanan origin hypothesis agrees with Bremer (2002) but the inference of an African origin is a novel result. Subfamilies Anomochlooideae and Pharoideae, estimated to have originated around 67-68 Mya, have a South American and pantropical distribution respectively, and our results may indicate that grasses were distributed in South America as early as the Paleocene (Figures 3.3, 3.5 and 3.6). Whether Anomochlooideae and Pharoideae achieved their South American by dispersal or vicariance is not clear based on these results. Indeed, it may not be possible to disperse by vicariance as Africa and South America already broke up in the Paleocene. Diversification in this case could have been stimulated by vicariance.

A shift in diversification rate (Figure 3.3) is found in the grasses at the crown node above the earliest-diverging lineages (Anomochlooideae, Pharoideae, Puelioideae and a clade including, among others *Suddia* and *Limnopoa*) between the early Paleocene and early Eocene (Figure 3.5). This corresponds to the so-called 'spikelet clade', a group with typical grass spikelets, including the BEP and PACCAD clades (GPWG, 2001). The correlation of this trait with increased diversification rate is not tested, and one cannot state that this trait caused a shift in diversification. However, it

is clear that the great diversity of morphological forms that exist in the two major clades are based upon the 'basic design' of the standard spikelet. It enabled them to diversify their inflorescences in a large number of ways and can be seen as analogous to other floral/inflorescence types such as thos found in orchids (Orchidaceae) flowers and the inflorescences of the daisy family (Asteraceae).

Anomochlooideae, Pharoideae and Puelioideae inhabit shaded tropical or warm temperate forest understories (GPWG, 2001) suggesting that proto-grasses might have inhabited similar forested environments by the early Tertiary. They then adapted to open habitats during the Tertiary following the opening of Paleocene and Eocene forested environments from the early to middle Tertiary (Jacobs et al., 1999). According to our geographical mapping using both ACCTRAN and DELTRAN resolving options, we found that South America might not have been an area in which members of the 'spikelet clade' began to evolve and diverge. Furthermore, one of the earliest-diverging lineages of the grasses, Puelioideae has an African distribution; so to do the early-diverging lineages of the BEP clade. This indicates that South America might have been colonized by Bambusoideae well after the breakup of Gondwana (Figure 3.6).

According to our molecular dating method, the BEP clade is considerably older than the PACCAD clade (45 versus 25 Mya), and underwent several shifts in diversification rates before the origin of the PACCAD grasses (Figures 3.4 and 3.5). A significant shift in diversification rate was detected at the origin of the BEP lineages above the *Isachne* clade between 50 and 40 Mya during the Eocene (Figures 3.4 and 3.5). This was coupled with a possible dispersal event to South East Asia and possible adaptation to open habitats for the *Isachne* clade members (Figure 3.6). A recent study found grass phytoliths preserved in coprolites (suspected titanosaur sauropod dung) from late Cretaceous in India (Prasad et al., 2005). Prasad et al. (2005) extracted several grass phytoliths, which seem to correspond to at least four morphotypes of extant grass subclades: Bambusoideae, Ehrhartoideae, PACCAD and/or Pooideae. These findings would imply that the BEP clade (GPWG, 2001) had diversified by the late Cretaceous and that typical pooids and/or PACCAD grasses also occurred at this time (Prasad et al., 2005). However, based on our molecular dating, a late Cretaceous diversfication of the BEP clade would date the origin of grasses at about 120 Mya. If PACCAD grasses would have occurred at this period that would date the origin of the

family back to around 140 Mya, which would contradict a recent molecular dating of the angiosperms suggesting a middle Jurassic to an early Cretaceous origin (between 180 and 140 Mya) (Bell et al., 2005). But, Prasad et al. (2005) states that we need to rely on the assumption that the ingroup taxa can be used to inform plesiomorphic character states (i.e. silica bodies morphotypes) within the family. Nonetheless, the findings of Prasad et al. (2005) do confirm that grasses were extant before the K/T extinction, at least the basal-most grass taxa. Also, this study suggests that grasses had already spread to India by the late Cretaceous (Prasad et al., 2005). Indeed, Asia and Africa were directly connected to each other as early as the Paleocene (Raven and Axelrod, 1974). Colonization of Asia by grasses might, therefore, have occurred via India, which was connected to Asia by the present-day south-eastern Asia land mass, as early as the late Cretaceous-early Tertiary (around 65 Mya) (Klootwijk et al., 1992).

An increase in diversification rate was also found at the origin of Bambusoideae around 30 Mya in the middle Oligocene (S5; Figures 3.4 and 3.5). The spread to South America, as suggested by our geographical mapping, could have occurred along a dispersal route through Eurasia which was connected to North America by the North Atlantic Land Bridge (NALB hereafter) until the end of the Eocene (Tiffney and Manchester, 2001) and finally to South America which was connected by the Greater Antilles, the Bahamas platform, the Aves Ridge which is submerged today, and perhaps the site of the present Lesser Antilles (Rage, 1996). This scenario of a spread from the Old to the New World via the NALB has already been stated for other tropical angiosperm clades (Chanderbali et al., 2001). Furthermore, Stromberg (2005) found the earliest evidence for Bambusoideae in the fossil record around the early Oligocene in North America, which is concordant with dispersal to South America from Asia across North America. We may speculate that this migration of closed-habitat types, such as bamboos, across a latitudinal gradient might be explained by the global warming trend during the first half of the Tertiary that lead to the expansion of tropical and paratropical forests into higher latitudes (Wing, 1987). Three shifts in diversification (S6, S7 and S8) were detected in Bambusoideae within the tribe Bambuseae between 10 and 5 Mya in the late Miocene (Figures 3.4 and 3.5). However, these shifts are not large in comparison to other detected shifts in terms of species number and are all located within the temperate woody bamboos. It seems likely that these increases in diversification occurred in Asia

and South East Asia (Figure 3.6).

Two successive shifts in diversification rates were found at the origin of Pooideae (S9 and S10; Figure 3.4) between 40 and 30 Mya in the late Eocene-early Oligocene (Figure 3.5). There was therefore an increased diversification following the divergence of the majority of the pooids from their sister lineages represented by the tribes Brachyelytreae, Lygeae and Nardeae, and the genera *Cyathopus* and *Milium*. These shifts seem to have occurred during the spread to North America and Europe (Figure 3.6), when the ancestors of Pooideae became adapated to open habitats. This common finding in phylogenetic studies (i.e. South East Asian and East Asian taxa being sister to derived New World taxa: as it is the case within the BEP clade in our study) often leads to the conclusion that the Bering Land Bridge (BLB hereafter) was a major dispersal route to the New World (Li et al., 2000). A dual spread from Europe to North America and from East Asia to North America, via the NALB and the BLB respectively, could have also occurred (Tiffney and Manchester, 2001). The diversification to open habitats by the ancestors of Pooideae is in agreement with an increase in the abundance of C_3 grasses in the middle Tertiary (Jacobs et al., 1999), in response to a stepwise climatic deterioration occurring in Europe and North America leading to the disintegration of the forest cover (Prothero and Berggren, 1992; Knobloch et al., 1993). As suggested by Stromberg (2005), some taxonomic diversification (as shown in Figures 3.2 and 3.4) occurred before they were first recorded in North America following a dispersal route from Africa, Asia, North America and then South America. Two shifts in diversification (S11 and S12; Figure 3.4) were also detected in very recent geological times, probably during the Pliocene or the Pleistocene, within the tribe Aveneae (Figure 3.4).

The origin of the PACCAD clade is estimated to have occurred between 34 and 22 Mya in the middle Oligocene-early Miocene (Figure 3.5). Diversification of PACCAD lineages occurred in more recent geological times than BEP lineages. The PACCAD ancestors were inferred to have originated in Africa and were adapted to open habitats (Figure 3.6). A shift in diversification rate was found at the origin of the PACCAD clade (S4; Figures 3.2 and 3.4). It is hard to determine the factors that may have driven this major diversification in the grasses from the early Oligocene to the early Miocene, but it was possibly linked to the adaptations to open habitats (Figures 3.3 and 3.5). The Oligocene-Miocene periods were considerably drier than the rest of

the Tertiary and these factors might have had an effect in the decrease of the forest cover and the expansion of open habitats such as savannas mainly in the Miocene (Janis, 1993). Several dispersal events seem to have occurred at the origin of the PACCAD subfamilies (Figures 3.5 and 3.6). According to our results, the first dispersals occurred to Asia, South East Asia and Australia (Figure 3.6). The spread to Europe, North America and South America occurred between 25 and 15 Mya possibly following the same dispersal routes as the BEP lineages. Stromberg (2005) found that open-habitat types, typically PACCAD grasses, spread at the expense of closed-habitat types during the late Oligocene and the early Miocene. According to our analysis, the adaptation to open-habitats was triggered either in Africa or at some time during the spread of grasses to Asia and South East Asia well before they became ecologically dominant in North America (Stromberg, 2005).

At least two major dispersals of PACCAD clade lineages occurred to Australia, one for the '*Micraira* clade' and the other for the tribe Paniceae both dated around 20 Mya in the Early Miocene (Figures 3.4, 3.5 and 3.6). The earliest record of grasses from Australia is in a middle Eocene pollen flora (Frakes and Vickers-Rich, 1991). One explanation could favor a vicariance hypothesis which states that grasses might have inhabited Australia in the Cretaceous via a dispersal route from South America via Antarctica as it has been shown for other plants (Sanmartin and Ronquist, 2004), birds (Ericson et al., 2002) and mammals (Benton, 1985). This scenario cannot be validated with our data. Australia appears to have had a dominant tropical climate during the Paleogene followed by cooling reflecting general global trends starting in the Oligocene (Singh, 1988). It may explain the early record of grass adapted to tropical forest by the Eocene, followed by extinctions due to the global cooling starting in the Oligocene.

Two shifts in diversification rates are found at the origin of the tribe Andropogoneae (subfamily Panicoideae) (S14 and S15; Figure 3.2). The Andropogoneae is a highly successful tribe in terms of species numbers (986 species) and its members are exclusively using the C_4 photosynthetic pathway (Watson and Dallwitz, 1992). These increases in diversification occurred in the middle Miocene and may be correlated with an increase in aridity, and with decreasing CO_2 concentrations in the atmosphere (Cerling et al., 1998). Even though C_4 taxa only appeared at about 15-10 Mya, ancestors of the PACCAD clade may have diverged and adapted in

restricted arid environments as early as the Eocene period. The diversification of drought-adapted PACCAD grasses only occurred about 20 million years later and was possibly due to a global increase in aridity. However, another explanation for the late diversification of the PACCAD clade could be major extinction events that may have occurred at the Eocene/Oligocene boundary, especially in Europe and North America (Prothero and Berggren, 1992). These extinction events might have especially affected warm-adapted species (Wolfe, 1992).

3.4.3 Conclusion

Our approach to incorporate different datasets (molecular, morphological, ecological and geographical), which have in common an overlap of taxa, have helped allow a more detailed analysis of phylogenetic diversification than previous studies. The results are enriched by paleontological and taxonomical data. It has helped reveal macro-evolutionary patterns and has proven to be a very powerful approach for dealing with species-rich group of organisms such as grasses. The use of the comprehensive phylogenetic tree has allowed us to test hypotheses in regards to extrinsic processes leading to diversification and eventually to help explain the present-day ecological success of these grasses. However, one has to bear in mind that the dating and the geographical analyses included in this study have relied on several assumptions: (i) the method used for the calibration consider maximum ages on certain nodes and (ii) the parsimony reconstruction of geographical areas does favor dispersal and do not consider possible extinctions when retrieving the ancestral area. If these were to be relaxed, it is not clear whether the present scenario would be retained.

This study of grasses has detected fifteen differential shifts in diversification among lineages during their evolution. Grasses also seem to have dispersed to all continents by 30 million years after their Gondwanan origin in the late Cretaceous. This is consistent with paleobotanical, paleofaunal, and stable carbon isotope records (Jacobs et al., 1999). Major events in the evolution of the grasses include: (1) major diversification of the BEP clade members (C_3 grasses) in the Paleocene and Eocene (between 55 and 35 Mya) possibly due to the decline of forested environments, with dispersal to Asia and subsequently to the New World, (2) later divergence of the PACCAD clade from the Oligocene (between 35 and 25 Mya), possibly due to an early adaptation to arid habitats with recent dispersals in the early Miocene to Eurasia

and to the New World, (3) diversification of grasses to become ecologically dominant in open environments between 30 and 20 Mya (Oligocene-Miocene transition), possibly due to initial adaptations to open habitats in Africa and Asia followed by numerous dispersals, and finally (4) relatively recent diversification within the PACCAD clade and the expansion of C_4 grasses occurring by the middle Miocene (between 15 and 10 Mya). Indeed, drought tolerance and the ability to grow in dry open habitats appeared long after the origin of grasses (Kellogg, 2001). As is shown in this study, the shift into open habitat occurred in recent geological times and may have led to major diversification events (Kellogg, 2001). However, limitations have been encountered for testing intrinsic factors leading to diversification among lineages. It has been proposed that key innovations (morphological and/or anatomical traits) might influence the rate of production of new species (Maynard Smith and Szathmary, 1995). However, it was not possible to test such hypothesis because of the high number of genera within the grasses which exhibit polymorphic morphological and/or anatomical characters.

4) Coevolution of grasses and ungulates

4.1 Introduction

Grasses (Poaceae) are of immense importance, both ecologically as they cover $3/4^{th}$ of the earth's surface (Shantz, 1954) and economically in the form of cereals and forage resources (Clayton and Renvoize, 1986). The evolution of the grasses and grasslands played a fundamental role in the formation of many modern ecosystems (Jacobs et al., 1999) and despite over a century of research, the patterns and processes that drove grass evolution are largely unknown (Stromberg, 2005). Even though the fossil record of grasses is extremely poor for most of the Cenozoic (Jacobs et al., 1999), there is enough evidence to show that the major radiation of the grasses and the establishment of all their major lineages had occurred by the mid-Miocene, between 15 and 25 million years ago (Mya) (Cerling et al., 1997; Jacobs et al., 1999; GPWG, 2001). The spread of grass-dominated ecosystems is also believed to have occurred by this time (Kellogg, 2000; GPWG, 2001). A simultaneous taxonomic proliferation of the family and rise to ecological dominance long after its origin is thought to have been stimulated by changes in global and regional climates towards increased seasonal aridity during the Miocene and the Pliocene (between 25 and 5 Mya) (Wing, 1998). In a recent study, Stromberg (2005) suggested that external factors triggered alterations in vegetation structures during the late Oligocene or early Miocene, allowing the spread of open-habitat grasses. Among the several potential environmental influences on the ecological success of open-habitat grasses, climate change and low atmospheric CO_2 levels during the Cenozoic are the most commonly discussed (Sage and Monson, 1999; Stromberg, 2005). Interaction between low atmospheric CO_2 levels and frequent fires may have promoted the spread of open-habitat grasses at the expense of forest trees (Bond et al., 2003).

Other factors, such as herbivory, may also have limited the abundance of closed-habitats that were dominated by trees (Stromberg, 2005). The role of herbivory in the evolution of open-habitat grasses has not been investigated in detail, even though the spread of grasslands may have been associated with increasing grazing rates throughout the Miocene (Chapman, 1996a).

A major and rapid radiation of vertebrate herbivores (Equidae and Bovidae) has occurred between 20 and 10 Mya (MacFadden and Cerling, 1994; Hassanin and Douzery, 1999). The emergence of the Bovidae family is thought to have occured around 20 Mya, and its evolution through the Miocene followed two main episodes: (i) a split between Eurasia and Africa which led to the development of Bovinae (cattle-like bovids) and Antilopinae (gazelles and antelopes), respectively, and, (ii) explosive radiations of Bovidae lineages during the middle Miocene to the early Pliocene (Hassanin and Douzery, 1999). This period was marked by an important global climate change promoting the spread of grasslands and the evolution of bovids adapted to a savanna-type habitat (Cerling et al., 1997; Janis et al., 2002). Equideae (horses) underwent high speciation and diversification during the same period (20-10 Mya), but this was principally centered in North America (MacFadden and Cerling, 1994). The classical explanation, as proposed by MacFadden and Cerling (1994), is an explosive adaptive radiation from low- to high-crowned (hypsodont) horses. The changes in dental morphology of ungulates might have coincided with the diminution of the tree cover and the development of a savanna type of habitat (Chapman, 1996a). Jernvall et al. (1996) suggested that Miocene ungulates evolved increasingly disparate crown types together with dietary specialization in more fibrous vegetation. One could suppose that herbivores apply a selective pressure on grasses by grazing, so that grasses evolve increasing leaf toughness in response. Reciprocally, herbivores might have evolved particular tooth structures to cope with an increase in leaf toughness. Indeed, graminoid grazing tolerance and the nearly simultaneous increase of grasses and mammalian grazers in the fossil record (Stebbins and Crampton, 1961) suggest that grass herbivory tolerance may have resulted from a coevolutionary process with vertebrate grazers (Coughenour, 1985). The effects of grazing on grasses have been documented (Austin et al., 1981; Sala et al., 1986), but they have generally focused only on floristic composition, herbage production or changes in soil environment (Sala et al., 1986), and not on coevolutionary aspects of their development.

As described by Chapman (1996a), the development of phytoliths (silica bodies in grass epidermal cells) and their persistence could be a consequence of herbivore dentition changes to improve ability to cope with an increasingly grass-based diet. Silica bodies are thought to reduce palatability, digestibility and the nutritional value of the forage grasses (Coughenour, 1985; Ellis, 1990; Chapman,

1996a; Massey and Hartley, 2006). Silica bodies are among the few substances capable of inducing morphological changes to animal mouthparts (Piperno, 2006). It has been shown that prairie voles consumed less grass of high silica content, suggesting that silica bodies act as effective deterrent to mammalian herbivores (Gali-Muhtasib et al., 1992). Silica ($SiO_2.nH_2O$) is deposited in large quantities in plants, being particularly abundant, diverse and distinctive in the grass family (Theunissen, 1994; Piperno, 2006). These microscopic silica bodies precipitate in or between cells of living tissues (Lu and Liu, 2003). The morphology and taxonomy of silica bodies has been the subject of many studies (Motomura et al., 2002; Lu and Liu, 2003; Wang, 2004; Stromberg, 2005; Piperno, 2006), but none has tried to investigate the evolution of this trait in relation to grass evolutionary history. Coughenour (1985) suggested that grasses and large grazing herbivores evolved together, but recognizes that it is difficult to show which traits arose predominantly because of grazing mainly because we are unable to determine from the fossil record the precise origin of graminoid traits in relation to herbivore evolution.

An approach to investigate this problem is to use phylogenetic methods to reconstruct historical changes (Losos, 1999) in silica concentration and to evaluate whether these changes are correlated with the timing and pattern of dental morphology throughout the evolution of ungulates. Teeth offer good opportunities to link morphology to ecology through diet and thus also the opportunity to study the rise of herbivory (Jernvall et al., 1996). No studies have tried to investigate if the spread of open-habitat grasses is correlated with increasing grazing rates throughout the Miocene, and if grasses underwent changes in their silica content to cope with large herbivore pressures.

Therefore, this study aims to (i) quantify silica density of leaf epidermal tissue across grass lineages and reconstruct historical changes throughout the Cenozoic, (ii) test if silica density changes vary among grass lineages as a process of coping with increased grazing rates by correlating these historical changes in silica densities with evolutionary changes in the molar morphology of ungulates, and (iii) correlate silica density changes across lineages with the appearance of open and closed-habitat type grasses to investigate if the spread of open-type grasses is linked with the evolution of ungulates.

4.2 Material and methods

4.2.1 Taxa sampling and silica density measurements

Leaves of grass specimens were sampled from herbarium vouchers at the Royal Botanic Gardens Kew, UK (Kew) and at the Botany Department of Trinity College Dublin (Appendix 4.1). We sampled the same species of each genus that were used for phylogenetic inferences and molecular dating methods in chapters two and three of this thesis (Appendices 2.1 and 4.1). However, we were restricted in our sampling by the availability of the specimens in herbaria. We selected the total sample to have, as far as possible, a broad representation of grass lineages (i.e. with representatives of most of the subfamilies defined by the GPWG, 2001).

Segments of mature leaf blades were boiled to hydrate the desiccated material and fixed in FAA (Formaldehyde, Ethanol 70% and Acetic Acid solution) (Johansen, 1940). We used the manual scraping method to prepare abaxial epidermal scrapes (Metcalfe, 1960) by leaving samples for 30 seconds in 3.5% Sodium Hypochlorite and manually scrapping off mesophyll with a scalpel blade. Samples were then stained for five minutes in a solution containing red safranin and Alcyan blue (Tolivia and Tolivia, 1987). Epidermal scrapes were washed in water and placed in ethanol solutions at five concentrations (50, 75, 90, 96 and 100%, respectively) for one minute. Finally, they were transfered in xylene and mounted on microscopic slides for further image analysis. Only one epidermal scrape was mounted for each species. The epidermal anatomical structure was recorded photographically using an OLYMPUS® DP25 Digital Camera. Silica density measurements were done using an image analysis software (Olympus C.A.S.T. Stereology System®). It allowed us to quantify the area in μm^2 of 10 randomly chosen silica bodies (see Figure 4.1). Using quadrats of known area (Figure 4.1), we randomly selected 10 quadrats and counted the number of silica bodies present in each. Whenever a silica body (or part of it) was within the limit of a quadrat, it was taken into account. Silica density was then calculated as:

*Silica Density Index (SDI hereafter) = (n * a) / A*

Where n is the number of silica bodies in 10 quadrats, a the average area of one silica body (μm^2) and A the area of 10 quadrats (μm^2). This was done for 148 leaf epidermis samples. The potential intra-specific and infra-generic variances of silica density were not taken into account. Estimating such variances would ideally require sampling all species for each genus, and consider species replicates for evaluating intra-specific variation. It was not practically possible to perform such a sampling strategy.

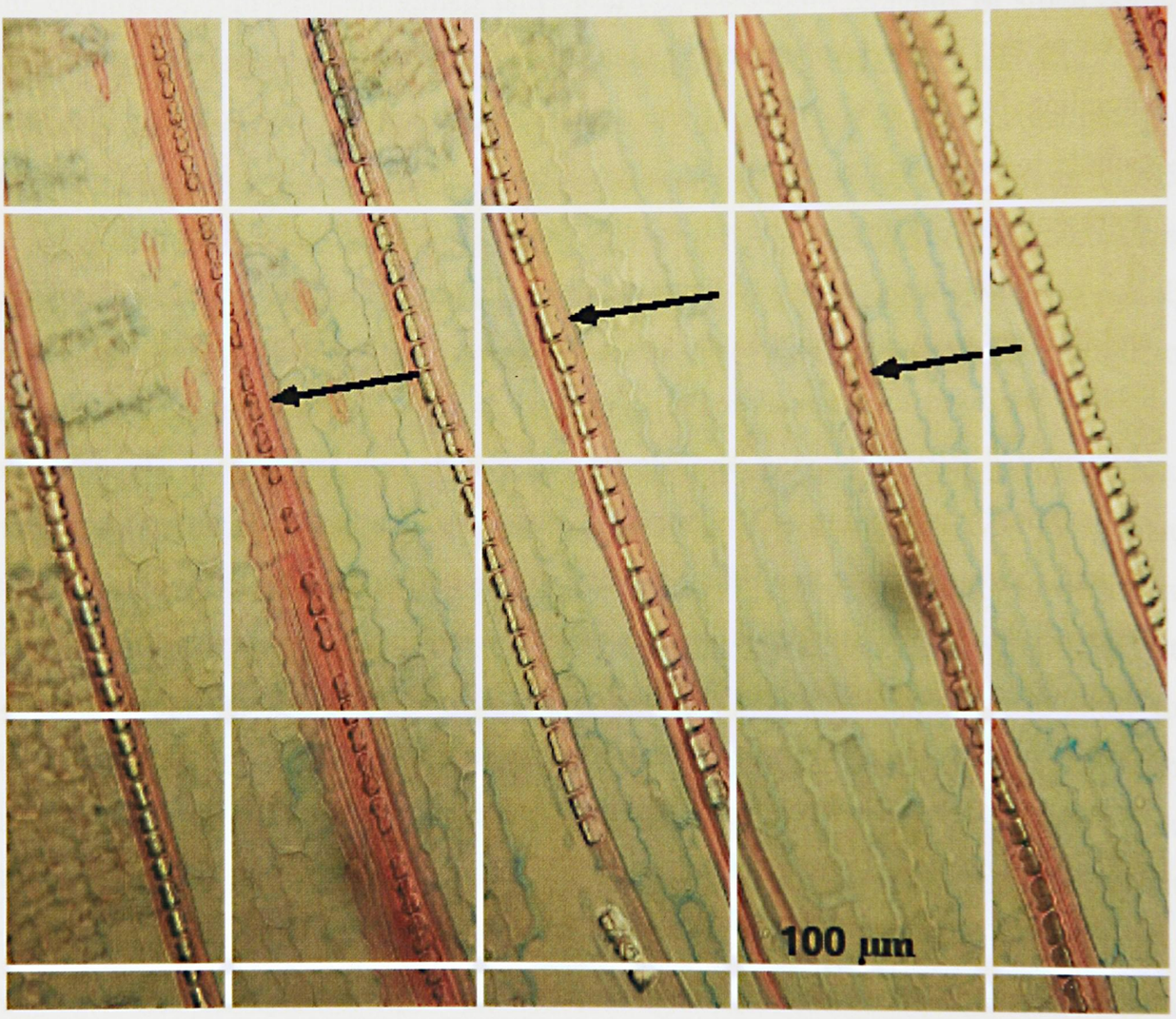


Figure 4.1 Grass abaxial epidermal scrape of Pharus latifolia. Arrows indicate individual silica bodies in the intercoastal short cells. Quantification of silica density was done by applying quadrats of known area (100µm x 100µm) and calculating the area of 10 randomly chosen silica bodies with an image software analysis. (Grass epidermal pictures are available on the CD accompanying this thesis; folder "Grass_leaves_pictures")

4.2.2 *Phylogenetic inferences, molecular dating and ancestral state reconstructions*

The sampling list for the phylogenetic inferences is shown in Chapter 2 of this thesis (section Material and Methods, DataSet I) but only includes 90 taxa because we excluded taxa for which we were not able to collect silica density data. We used the chronogram shown in Figure 3.4 and pruned 20 taxa from the phylogenetic tree. We used a second data set, DataSet II (see Material and methods section of Chapter 2 of this thesis), to check if increased taxon sampling has an effect on ancestral characters reconstructions, as suggested by Ackerly (2000). We pruned 146 taxa from this for which we were not able to collect silica data. The protocols used for DNA extraction, PCR amplification and sequencing cycles are described in the Material and Methods section of Chapter 2. The molecular dating method is described in the Material and Methods section of Chapter 3.

We reconstructed ancestral states of silica density using both maximum likelihood (Schluter et al., 1997) and local squared change parsimony (e.g. PIC method) (Felsenstein, 1985b) methods implemented in APE (Paradis et al., 2004) for both data sets (i.e. 90 taxa chronogram and 148 taxa phylogenetic tree with branch length set to 1.0). Then, we plotted silica density through time, from 72 Mya to 10 Mya, at the nodes of Aristidoideae, Arundinoideae, Bambusoideae, Centothecoideae, Chloridoideae, Danthonioideae, Panicoideae and Pooideae subfamilies, sensus the GPWG (2001) to check if changes in silica density were more pronounced for particular lineages. Also, this method allowed us to keep the information provided by the phylogenetic tree structure of grass lineages through time, which takes into account the cladogenesis of the family.

4.2.3 *Correlation of silica density changes of grasses with molar tooth evolution of ungulates*

We referred to a study done by Jernvall et al. (1996) who performed an analysis of molar crown types of the Artiodactyla, Perissodactyla and archaic ungulates during the Cenozoic period. Using the morphological type of the upper second molar as a discrete crown type, they were able to quantify lophedness (i.e. lophs are defined as shearing blades) by tabulating the number of lophs among the crown types and dividing this by the number of crown types for each land mammal

age (Jernvall et al., 1996). Lophs are best developed in herbivores consuming fibrous plant foods such as grasses (Jernvall et al., 1996), suggesting that loph numbers could be correlated with specialized herbivory in fossil taxa. However, the fossil collections used were only available for North America, Europe and Asia, which did not allow us to compare biogeographical data between grasses and ungulates. We obtained the average of lophedness of the three geographical areas between 72 and 10 Mya using the graph shown in Figure 4 from the study of Jernvall et al. (1996).

We were then able to compare the historical silica density values of grasses with average lophedness of ungulates at the same geological times (i.e. at the time when grass lineage divergences occurred). To correlate silica density and average lophedness through time, we used the Spearman's rank correlation coefficient (Spearman, 1904) that is a non-parametric measure of correlation. The test results include the estimated Spearman's rank-order correlation coefficient (rho) and the p-value (both one- and two-tailed). We performed this test for historical silica density values of each subfamily.

4.2.4 Correlation of silica density changes with grass adaptation to open / closed habitats

In order to correlate silica density with light tolerance characters and to test if silica density among grass lineages is correlated with adaptations to open or closed habitats, we used data from the Grass Genera of the World DELTA database (Watson and Dallwitz, 1992), which categorises all grass genera (i.e. 798 genera) as occupying 'open' or 'closed' habitat. However, out of the 90 genera sampled in the chronogram, 27 were coded as missing because they contain high numbers of species, which could occur in both open and/or closed environments. Assuming the monophyly of genera, we ran three comparative analyses within which polymorphic genera (i.e. coded with missing data in the Grass Genera of the World DELTA database (Watson and Dallwitz, 1992)) were coded as (i) 'open', (ii) 'closed' and (iii) 'open+closed'. We transformed the silica density data into a log-normal distribution as the silica measurements were not normally distributed and performed a Generalized Estimating Equations (GEE hereafter) comparative analysis, implemented in APE (Paradis et al., 2004), that allows us to test the correlation between continuous and discrete variables (Paradis and Claude, 2002)

4.3 Results

4.3.1 Ancestral silica density among grass lineages

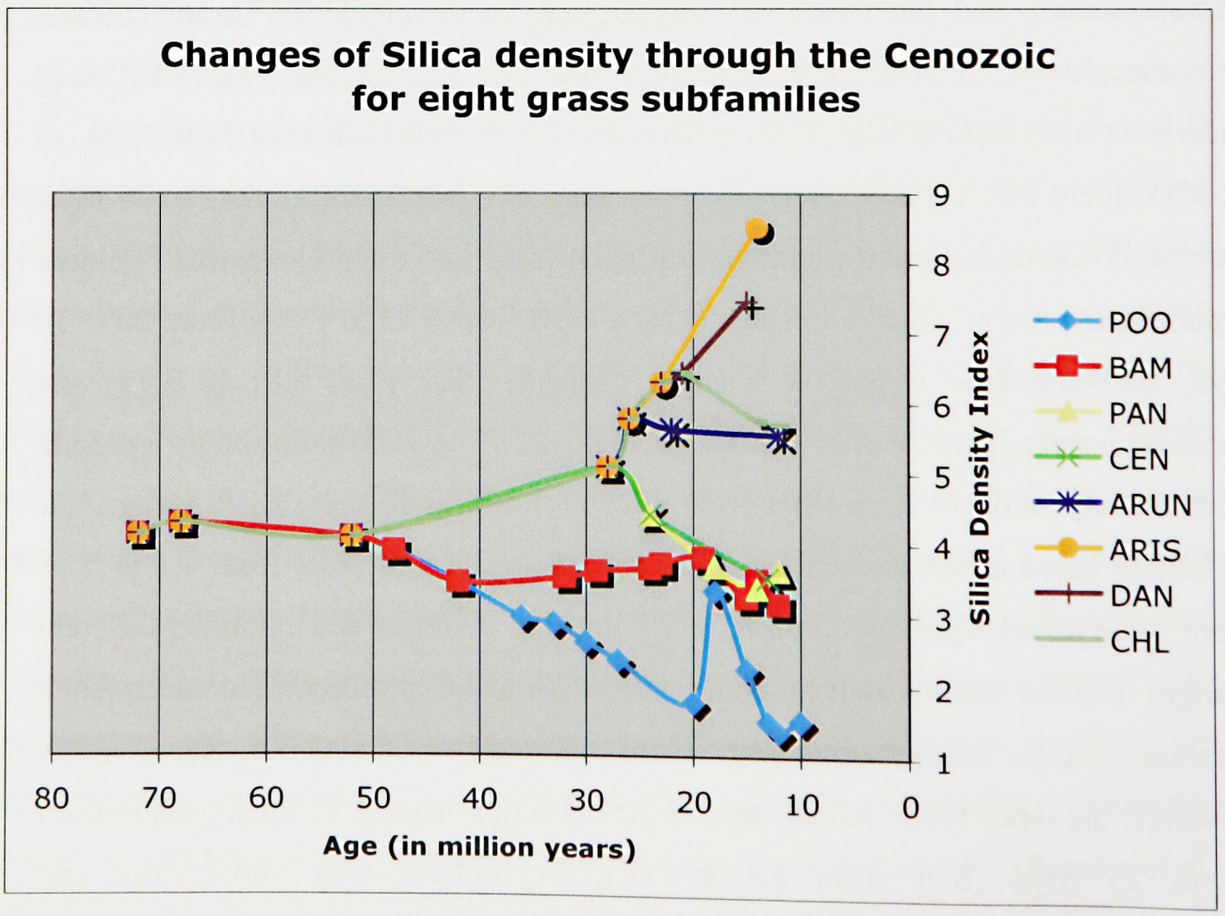
Changes in silica density among grass lineages through time are shown in Figure 4.2. We only presented results from the 90 taxa chronogram because both data sets (90 and 148 taxa) resulted in the same changes in silica density among lineages. The consistency in locating increased shifts in silica density in all analyses and the lack of branch length estimates for the 148 taxa phylogenetic tree led us to rely only on the chronogram (Figure 4.3). Both ML (Schluter et al., 1997) and PIC (Felsenstein, 1985b) reconstruction methods yielded the same changes in silica density.

We found that grass ancestors did not undergo a change in silica density index (SDI) between 72 and 52 Mya (SDI from 4.14 to 4.10) (Figure 4.2). At 50 Mya, grasses diverged into two major lineages: the BEP (represented by Pooideae and Bambusoideae), and the PACCAD (represented by Aristidoideae, Arundinoideae, Centothecoideae, Chloridoideae, Danthonioideae and Panicoideae). The BEP clade underwent a decrease in silica density from 50 to 42 Mya (SDI from 4.10 to 3.45). At 42 Mya, there was a split between Bambusoideae and Pooideae. From 42 to 20 Mya, Bambusoideae underwent a slight increase (SDI from 3.45 to 3.81) whereas Pooideae experienced a steep decrease in silica density (SDI from 3.45 to 1.75). From 20 to 10 Mya, Bambusoideae undertook a slight decrease (SDI from 3.81 to 3.15) and Pooideae a steep increase until 18 Mya (SDI from 1.75 to 3.33) followed by a drop in silica density until 10 Mya (SDI from 3.33 to 1.50). Unlike the BEP lineage, the PACCAD faced a rather steep increase from 50 to 28 Mya (SDI from 4.1 to 5.1). From 28 Mya, all representative PACCAD lineages diverged and exhibited different changes in silica density. Indeed, both Panicoideae and Centothecoideae underwent a decrease in silica density (SDI from 5.1 to 3.6, and from 5.1 to 3.5, respectively). However, the other PACCAD lineages experienced an increase in silica density from 28 to 25 Mya (SDI from 5.1 to 5.8), in contrast with Panicoideae and Centothecoideae. Arundinoideae faced a very slight decrease (SDI from 5.8 to 5.5) from 25 to 12 Mya. Chloridoideae underwent an increase in silica density from 28 to

22 Mya (from 5.1 to 6.4) followed by a decrease until 12 Mya (from 6.4 to 5.7) (Figure 4.2). In contrast, both Aristidoideae and Danthonioideae exhibited a very steep increase in silica density from 28 to 14 Mya (from 5.1 to 8.5 and from 5.1 to 7.6, respectively) (Figure 4.2).

According to our results, the BEP and PACCAD lineages seemed to have experienced differential changes in silica body densities since about 50 Mya (Figure 4.2). Within the BEP clade, Bambusoideae evolved higer silica density than Pooideae from the Oligocene through the Miocene (Figure 4.2). Within the PACCAD clade, even though all lineages experienced increase in silica density from the Oligocene to the Miocene, Panicoideae and Centothecoideae underwent a steep decrease during the early Miocene whereas Arundinoideae and Chloridoideae faced a decrease in the late Miocene. Only Aristidoideae and Danthonioideae undertook an increase in silica density through the whole Cenozoic era with a steeper increase during the Miocene (Figure 4.2).

A



B

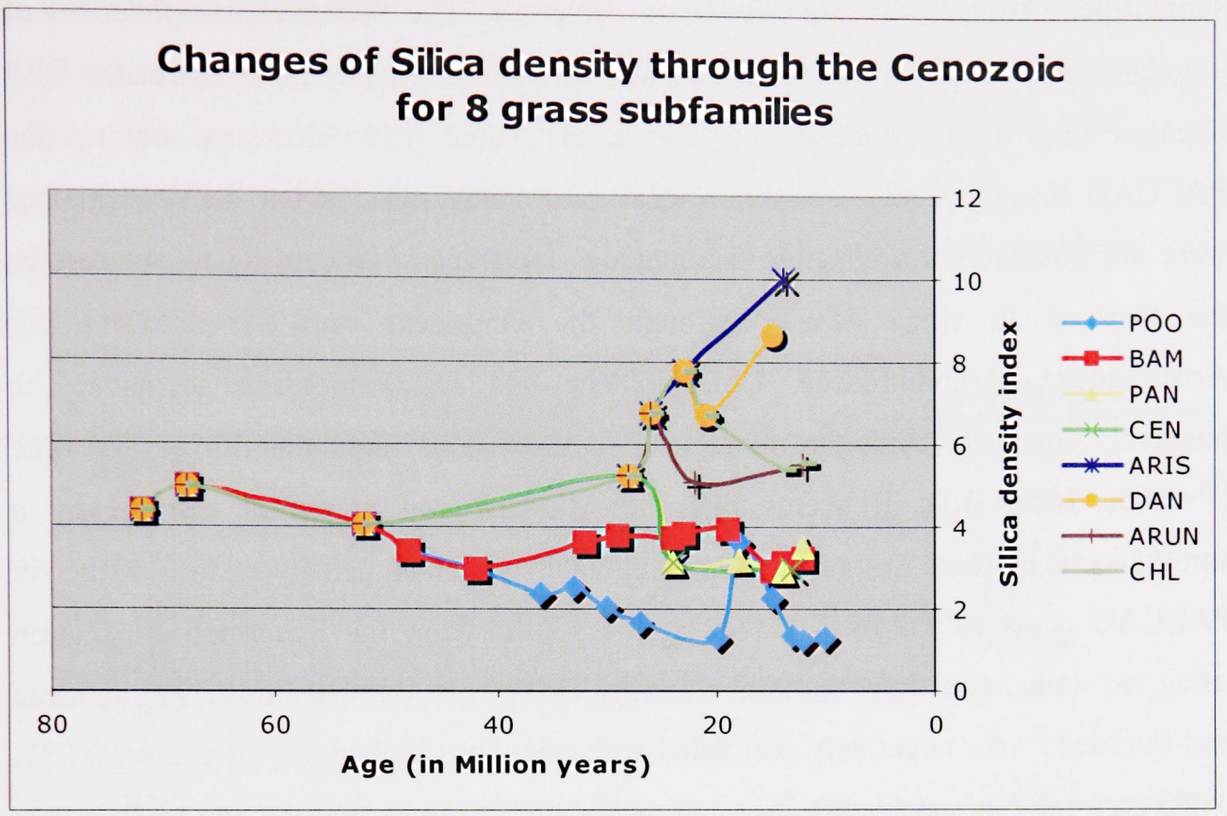


Figure 4.2 Change is silica density through the Cenozoic for the major grass subfamilies with ML (A) and PIC (B) reconstruction methods. ARIS: Aristidoideae, ARUN: Arundinoideae, BAM: Bambusoideae, CEN: Centothecoideae, CHL: Chloridoideae, DAN: Danthonioideae, PAN: Panicoideae and POO: Pooideae.

4.3.2 Correlation of silica density changes with the molar evolution of ungulates

According to the compilation of the average lophedness for the three geographical areas from Jernvall et al. (1996), the average lophedness of ungulates increased over time in North America, Europe and Asia, which accounted for the radiation of forms with many lophs in the latter part of the Cenozoic (Figure 4.4). Following the rise in average lophedness in the Paleocene and the Eocene, the late Eocene-Oligocene was characterized by a slight drop in lophedness (Figure 4.4). From the late Oligocene throughout the Miocene, the data of Jernvall et al. (1996) implies that average lophedness gradually increased to the modern ungulate value (Figure 4.4).

The results of the Spearman rank correlation, based on the data compilation of Jernvall et al. (1996) and our ancestral silica density data for the eight grass subfamilies (Figure 4.2), are shown in Table 4.1. The increase in lophedness of ungulates was not positively correlated with an increase in silica density for the BEP lineages (e.g. Bambusoideae and Pooideae) (Table 4.1). However, among the PACCAD lineages, only Centothecoideae and Panicoideae silica density changes were not positively correlated with ungulates lophedness. According to our results, the increase in silica density through the Oligocene and the Miocene for Aristidoideae, Arundinoideae, Chloridoideae and Danthonioideae seems to be positively correlated with the increase in lophedness of ungulates during this time ($P<0.05$; Table 4.1). It seems that the Miocene was a period marked by a simultaneous increase in lophedness of ungulates and silica density of four of the six PACCAD grass subfamilies. However, the results from the Spearman correlation using two-tailed test show that silica density changes in Bambusoideae, Panicoideae and Pooideae were negatively correlated with ungulates lophedness (p-values: 0.021, 0.003 and 0.045 respectively). To what extent such patterns may be due to chance is not clear in this study. Indeed, no randomization procedure was performed to test if these correlations result from a random process.

4.3.3 Correlation of silica density with open versus closed habitats

The results of the GEE correlation analysis of silica density with open versus closed habitats are shown in Table 4.2. When missing data were coded as 'open', we found that increase in silica density was correlated with 'closed'-habitats adapted grasses ($P<0.05$) (Table 4.2). When missing data were coded as 'closed', increase in silica density was also significantly correlated with 'closed'-habitat types ($P<0.01$) (Table 4.2). Finally, when missing data were coded as 'open+closed', we again found that increase in silica density was significantly correlated with 'closed'-habitat type ($P<0.01$) (Table 4.2). These results suggested that any increase in silica density was more likely to occur for taxa adapted to 'closed' or forested environments.

Table 4.1 Summary of the results of the Spearman correlation analysis (only one-tailed p values are shown) comparing increase in silica density and increase in ungulates lophedness at each class age. Significant p-values are indicated in bold. ARIS: Aristidoideae, ARUN: Arundinoideae, BAM: Bambusoideae, CEN: Centothecoideae, CHL: Chloridoideae, DAN: Danthonioideae, PAN: Panicoideae and POO: Pooideeae.

	Estimate (rho)	p-value
ARIS	0.873	**0.005**
ARUN	0.691	**0.043**
BAM	-0.772	0.999
CEN	-0.116	0.587
CHL	0.655	**0.028**
DAN	0.916	**0.001**
PAN	-0.611	0.946
POO	-0.787	0.999

Table 4.2 Summary of the results of the GEE analysis comparing silica density and light tolerance character, when we coded missing data as (i) 'open', (ii) 'closed' and (iii) 'open and closed'. Significant p-values are indicated in bold.

Missing data		Estimate	Standard Error	P-value
?=Open	(Intercept)	3.527	1.364	0.0192
	Open	0.848	1.025	0.4192
	Closed	0.694	0.286	**0.0267**
?=Closed	(Intercept)	3.274	1.252	0.0181
	Open	-0.117	0.319	0.7173
	Closed	1.321	0.301	**0.0004**
?=Open+Closed	(Intercept)	2.822	1.358	5.30E-02
	Open	0.846	1.003	4.11E-01
	Closed	1.407	0.261	**4.95E-05**

4.4 Discussion

4.4.1 Ancestral reconstructions

Both reconstruction methods, ML (Schluter et al., 1997) and PIC (i.e. Phylogentic Independent Contrasts; Felsenstein, 1985b), inferred the same changes in silica density (Figure 4.2). When reconstructing ancestral characters, one option is to use different methods (Losos, 1999) and check if they produce different results. Ryan and Rand (1999) in their study considering the extent to which different methods of estimating ancestral states yielded different reconstructions (the two methods studied were squared-change and local squared-change parsimony), found that the two methods used did not alter qualitatively their conclusions. To a larger extent, it is important to determine whether different assumptions (which relies on the different reconstruction models) result in statistically different estimates of the same character (Ryan and Rand, 1999). We did not test if ML and PIC resulted in statistically different estimates of silica density, but the results from the two methods did not differ qualitatively (Figure 4.2). In terms of taxon sampling, we found no differences in changes in silica density among grass lineages between the two trees under study (i.e. same shifts found with the 90 taxa chronogram and the 148 taxa phylogenetic tree). As suggested by Ackerly (2000), there would be no a priori reason to include all known taxa in a clade but rather use subsamples drawn from larger clades, as it is the case with the 90 taxa chronogram (Figure 4.3).

According to our results, the eight grass subfamilies sampled did not exhibit parallel and consistent changes in silica density (Figure 4.2). Indeed members of the BEP clade did undergo decrease in silica density from 50 Mya to the late Miocene (Figure 4.2). In contrast, the PACCAD lineages (four of six lineages) seem to have undergone a continuous increase in silica density during the Cenozoic (Figure 4.2). They diversified in the late Oligocene-early Miocene (Figure 4.3) and exihibited a steep increase in silica density for most of the Miocene, with the exception of Centothecoideae and Panicoideae (Figure 4.2 and Table 4.1).

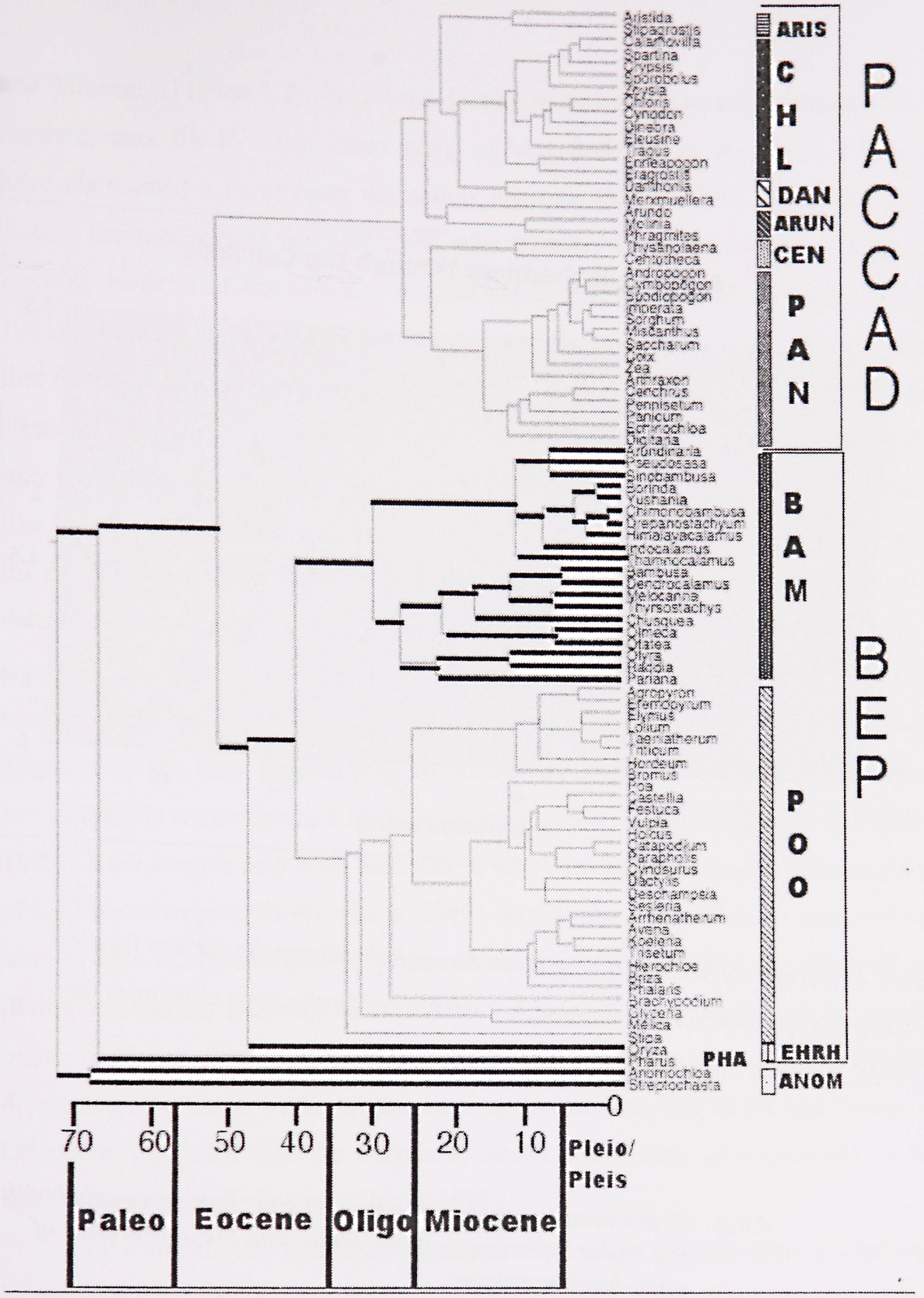


Figure 4.3 90 taxa chronogram used for ancestral characters reconstructions. Black horizontal bars indicate the optimization of closed (or forested) habitat (see Results section of Chapter 3 for details). Paleo: Paleocene, Oligo: Oligocene, Plio: Pliocene and Pleis: Pleistocene. ANOM: Anomoochloideae, ARIS: Aristidoideae, ARUN: Arundinoideae, BAM: Bambusoideae, CEN: Centothecoideae, CHL: Chloridoideae, DAN: Danthonioideae, EHRH: Ehrhartoideae, PAN: Panicoideae, PHA: Pharoideae and POO: Pooideae.

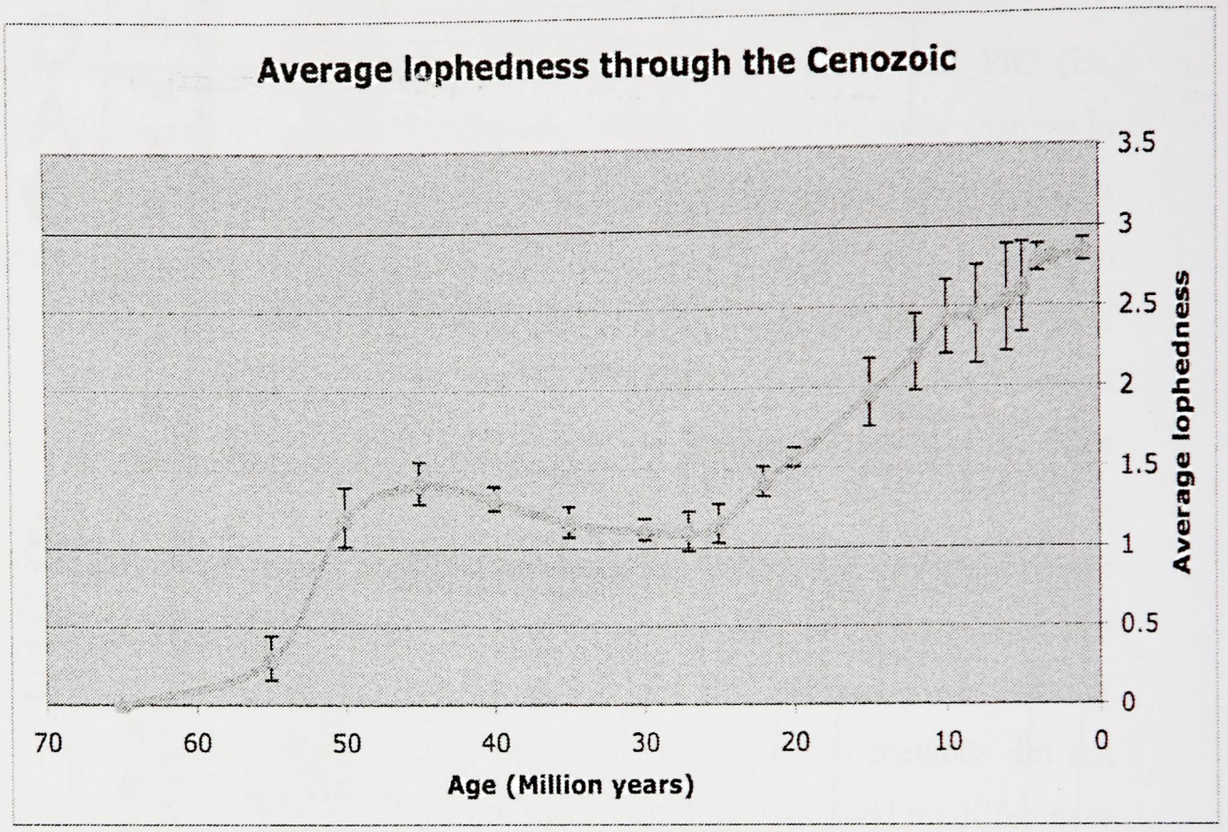


Figure 4.4 Average lophedness of ungulates through the Cenozoic; data of fossils from Asia, Europe and North America are summarized by the same class ages than Figure 4.2. Vertical bars indicate standard errors of the mean.

4.4.2 Evolutionary patterns in silica density changes coupled with the molar lophedness of ungulates and the appearance of open-habitat grasses

Assuming that silica density changes reflect grass response to herbivore pressure, Bambusoideae and Pooideae seem to have responded to increased herbivore pressure by lowering their silica density during the Cenozoic (Table 4.1). Even though the average lophedness of Asian, European and North American ungulates rose in the Eocene and troughout the Miocene (Jernvall et al., 1996; see also Figure 4.4), BEP grasses (especially pooids) did not undergo an increase in silica density (Figure 4.2). We rather found a decrease in silica density from the Paleocene until the

late Miocene (Figure 4.2). Paleocene forests were denser possibly because of the absence, since the K/T extinction event, of large herbivorous dinosaurs, that would have maintained a more open vegetation structure (Janis, 1993). According to our results, the ancestors of the BEP clade were adapated to closed habitats (Figure 4.3). Possibly because of the absence of grass-eating mammals until the Eocene (Janis, 1993), they did not need to increase their silica density. However, there is evidence that titanosaurid sauropods ate grasses, as grass phytoliths were found in coprolites (Prasad et al., 2005). By at least 65 Mya, grasses and to a more extent angiosperms may have experienced herbivore pressure such that they evolved mechanical defences like silica bodies (Piperno and Sues, 2005). This is in contradiction with our results as the first Paleocene grass lineages (e.g. BEP lineages) did seem to decrease their silica density. The overall decrease in silica density of the BEP lineages throughout the Cenozoic may reflect the absence of diversified grass-eating mammals adapted to open-habitats until the Miocene (Janis, 1993). Perissodactyls and artiodactyls appeared by the early Eocene possibly in Asia (Stucky, 1990; Janis, 1989), but only perissodactyls were abundant in the early and middle Eocene (Prothero and Schoch, 1989). Their success may have been linked with their ability to process fibrous foliage of the Eocene tropical forests (Janis, 1989). As a consequence, Eocene grasses did not have to cope with heavy grazing pressures because plant diversity was relatively high during this period (Jernvall et al., 1996), allowing Eocene ungulates to process a higher diversity of foliage resources than Miocene ungulates adapted to open-habitats dominated by grasses. Furthermore, the steep climatic cooling in the late Eocene and Oligocene suggested that the Oligocene epoch was a time of a decrease in large herbivores taxonomic diversity (Janis, 1989).

In contrast, the PACCAD lineages seem to have undergone a continuous increase in silica density during the Cenozoic (Figure 4.2). They diversified in the late Oligocene-early Miocene (Figure 4.3) and exhibited a steep increase in silica density for most of the Miocene, with the exception of Centothecoideae and Panicoideae (Figure 4.2). The relative dryness of the Oligocene (Singh, 1988), as well as the radiation of artiodactyls may have restricted the range of closed-habitats (Janis, 1993). Indeed, the ancestors of the PACCAD clade were likely adapted to open-habitats in the late Oligocene (Figure 4.3), and showed a slight increase in silica density (Figure 4.2). Terrestrial forms in the Oligocene/Miocene transition were more

diverse than the Paleocene faunas (Collinson and Hooker, 1991), and this difference indicates that the opening of the understory permitted diversification of mammals with a greater availability of leaves as a food resource (Janis, 1993). According to Stebbins (1981), grasses that grew in drier and more open habitats (as are C_4 PACCAD grasses) evolved firmer and more silicaceous leaves. Indeed, from the late Oligocene to the late Miocene, PACCAD grasses evolved higher silica density (Figure 4.2). Also, the average lophedness of ungulates gradually increased from the late Oligocene through the Miocene (Figure 4.4). It is tempting, based on these results, to suggest that this trend for PACCAD grasses may be linked to the spread of savannas and savanna-adapted mammals (Webb, 1989). According to Jernvall et al. (1996), the Miocene radiations of ungulates evolved dental adaptations to deal with vegetation of low primary productivity. It is then plausible that the Miocene ungulates evolved higher loph numbers on their 2nd upper molar to deal with increasing silica density of C_4 grasses. Interestingly, not only exclusive C_4 subfamilies faced an increase in silica density, but also subfamilies including both C_3 and C_4 grasses such as danthonioids and aristidoids (Figure 4.2). Based on our results, Aristidoideae, Arundinoideae, Chloridoideae and Danthonioideae evolved higher silica density throught the Cenozoic and these increases were correlated with increase in ungulates lophedness (Table 4.1). It is plausible that these grasses did undergo heavier herbivory pressures than the BEP lineages because of a continuous drying trend during the Miocene, which suggests an expansion of open-habitats (Janis, 1993), and the predominance of C_4 grasses (see Results section of Chapter 3 of this thesis). Unlike the other PACCAD lineages, changes in silica density of Panicoideae, which includes C_4 grasses, were negatively correlated with the changes in lophedness of ungulates (Table 4.1). One can speculate that either the spread of Panicoideae into other geographical areas (such as Australia, see Results section of Chapter 3 of this thesis) may have resulted in lower grazing pressures during the middle to late Miocene, or that Panicoideae may have responded differently (i.e. by increased fibre content, rhizomatous growth…) to heavy grazing pressures.

It seems that the appearance of open-habitat ecosystems, such as savannas, in the Miocene (Ceriling at al., 1997; Jacobs et al., 1999; Stromberg, 2005) was accompanied by an overall increase of silica density for some PACCAD grasses. According to our results, it is rather an adaptation to forested environments that is

linked with increase in silica density (Table 4.2). According to the results from Chapter 3 from this thesis (Results section), we found that the appearance of lineages adapted to open-habitats occurred in the middle Oligocene for Pooideae and in the late Oligocene-early Miocene for PACCAD (Figure 4.3). The fact that the pooids ancestors were adapted to open-habitats and underwent a constant and gradual decrease in silica density might explain this result. Also, the number of pooids (open-habitat type grasses) sampled is high in our phylogenetic tree (i.e 30 pooid taxa versus 36 PACCAD taxa), the decrease in silica density they exhibited tend to bias the trend found for the PACCAD lineage. A larger sample size in all subfamilies sampled might help counteract this effect. Also, light tolerance, as it is labelled in the World Grass Genera database Watson and Dallwitz (1989), may not be linked to silica density in grasses. Other ecological and/or morphological traits may be correlated with silica density in grasses, and this needs further investigation. It is believed that silica enhances plant growth by protecting against detrimental effects of abiotic and biotic stressses (Epstein, 1999). In *Oryza sativa*, instances of growth and development were positively affected by silica concentrations (Mitsui and Takatoh, 1963). Silica accumulation in cell walls may add mechanical strength (Epstein, 1999), and might be triggered by physiological plant activity (Motomura et al., 2002). Adaptation to forested environments may require additional mechanical strength for light competition and to deter herbivory in understories. Nonetheless, any attempt to correlate silica density with single ecological traits may lead to erroneous results, as it is very plausible that silica density may be linked to a set of physiological and morphological traits.

Even though, deterrence of ruminant herbivory by silica bodies has not been unequivocally demonstrated (expect the study by Massey and Hartley (2006) which showed that silica density may reduce leaf palatability), this study reveals a parallel pattern between grass silica content and molar evolution of ungulates at least during the Miocene. Alternatively, one can speculate that grasses may tolerate grazing pressures by lowering their silica contents (by physiological control), which would allow them to allocate more energy to other traits that deter grazing pressures (such as rhizomatous growth, increased contents of fibres and tannin-like substances) (Coughenour, 1985; Ellis, 1990). This study also reveals a possible phylogenetic approach for evaluating the effects of grazing on grass evolution. The most

challenging aspect is the precise selection of traits, which may be correlated with grass evolutionary response to herbivory. Also, even though the role of silica bodies on leaf palatability has been discussed in the literature (Coughenour, 1985; Chapman, 1996a), its relation to leaf toughness and its implications in modifying herbivores mouth counterparts have not been tested and remain unclear.

This study does not consitute conclusive evidence for the presence of selective pressure leading to increased silica density implied by increased grazing rates in the Miocene. Other traits (such as leaf tensility, leaf dry matter content, rhizomatous growth and tannin-like substances) should be analyzed using a phylogenetic approach to reveal any coevolutionary trends in grass-ungulate interactions.

4.4.3 Limitations and further perspectives

We have found evidence of a general parallel increase in ungulates lophedness and silica density of PACCAD grasses through the Cenozoic. The results found in this study also show that: (i) there are differential evolutionary responses of major clades of grasses in relation to increased lophedness of ungulates through the Cenozoic, (ii) an increase in silica density is correlated with adaptation to closed-habitats, (iii) higher silica density changes occured for PACCAD lineages than BEP (especially for Aristidoideae and Danthonioideae grasses), and (iv) C_4 grasses may have undergone an increase in silica density in response to increasing grazing rates through the Miocene. Nevertheless, testing the hypothesis of ungulates (and in a broader sense hypsodont mammals) and grass coevolution remains a very difficult task and major methodological limitations arose. Firstly, only one species was represented for each genus. One has to consider that there may be intra-generic, or even intra-specific variation in silica density and no studies have investigated this issue. Secondly, no biogeographical framework was used to compare ungulates and evolution of grasses because no fossil data were sampled from Africa, Australia and South America by Jernvall et al. (1996), whereas Africa, according to the results of this thesis (see Results section of Chapter 3 of this thesis), is the centre of origin of grasses and remains a geographical area from where grasses dispersed: the first colonization event appears to be to South America (see Results section of Chapter 3 of this thesis). Thirdly, one has to bear in mind that the assumption of the Brownian motion model used to model continuous characters is that the mean of the character reconstructed

from the tips to the root of the phylogenetic tree is staying the same (Butler and King, 2004). It only allows for larger variance as time increases (Butler and King, 2004). That supposes that there is no trend through time when reconstructing ancestral silica density reconstructions. One alternative would be to specificaly model the evolution of silica density through time, which would suppose to imply empirically the evolution of silica density through time among grass lineages. Also, no randomization procedure was performed to test whether the patterns found were due to random processes or not. Finally, selecting adaptive traits that may be correlated with herbivory defence is a rather difficult issue. Other anatomical characters, such as fibre density or tannin-like substances are thought to act as deterrent against herbivores (Ellis, 1990). The applications of biogeographical, ecological, paleontological and taxonomic data coupled with phylogenetic trees would provide more robust perspectives in understanding such parallel evolutionary patterns between herbivores and grasses.

5) Final discussion

5.1 Grass phylogenetic relationships

The results of the large multi-gene (*rbcL*, *matK* and *trnL-F* intron and intergenic spacer) phylogenetic analyses presented in this thesis have offered support to many previous hypotheses of relationships within the family and helped resolve relationships that were previsouly unclear. Three EDLs were recognized by the GPWG (2001) as successively sister to the rest of the grass family: Puelioideae, Pharoideae and Anomochlooideae (including *Anomochloa* and *Streptochaeta* respectively). Based on our results, using MP and BI, it is more conservative to recognize four EDLs (*Anomochloa, Streptochaeta, Pharus* and *Puelia*) because of the possible paraphyly of Anomochlooideae. *Anomochloa* is weakly supported as the next EDL after *Streptochaeta* in both MP and BI (Figures 2.5, 2.6 and 5.1). The monophyly of *Anomochloa* and *Streptochaeta* was also not supported in three previous phylogenetic analyses of the family using both molecular and morphological data (Clark et al., 1995; Soreng and Davis, 1998; Hilu et al., 1999). The positions of Pharoideae and Puelioideae are in agreement with all studies that have included these taxa (Clark et al., 1995; Clark and Judziewicz, 1996; Soreng and Davis, 1998; Clark et al., 2000; GPWG, 2001). They are the next two sister groups to the rest of the grasses after Anomochlooideae.

Our results show that the largest division in the family is a well-supported BEP-PACCAD bifurcation (Figure 5.1). According to Watson and Dallwitz (1992), these two clades contain 4,745 and 5,406 species respectively. In previous studies, when EDLs are excluded from consideration, the remaining grasses are also generally, but not always, split into two lineages (Clark et al., 1995; Soreng and Davis, 1998; GPWG, 2001). Soreng and Davis (1998) recovered a Pooideae + PACCAD clade and the GPWG (2001) found BEP and PACCAD clades. The sister-relationship of BEP + PACCAD can therefore be considered as controversial as few studies have found strong support for this pattern, and no morphological synapomorphies supporting the BEP clade have been identified (GPWG, 2001). Our

results offer strong support for this hypothesis especially using the multi-gene analyses shown on Figure 2.4 with 100% BS and 1.00 PP.

Within the PACCAD clade, the six subfamilies as defined by the GPWG (2001) are supported by our results. The monophyly of Chloridoideae was supported and this is in agreement with previous studies (Hilu et al., 1999; GPWG, 2001; Hilu and Alice, 2001). According to our phylogenetic inferences, a sister group relationship is found between Chloridoideae and Danthonioideae (Figure 5.1). This finding disagrees with some previous studies where Arundinoideae were thought to be the most closely related subfamily to the chloridoids (Clayton and Renvoize, 1986; Hilu and Alice, 2001). However, the monophyly of the arundinoids is not supported by many studies and this result is not surprising in the sense that Danthonioideae are a subset of Arundinoideae s.l. (GPWG, 2001; see discussion below). Danthonioideae are well supported in all our analyses with the exclusion of *Monachather* (Figure 5.1).

It is worth noting that Arundinoideae s.str. were not monophyletic and are distributed in lineages sister to the Aristidoideae, Danthonioideae, *Eriachne* + *Micraira*, and Chloridoideae group (Figure 5.1). Previous studies suggested a monophyletic Arundinoideae comprising two clades: one containing tribe Danthonieae and the other Arundineae but their respective bootstrap values were low (Barker et al., 1999). Subsequent phylogenetic analyses have proposed that they would be better treated as two distinct subfamilies, Arundinoideae and Danthonioideae (GPWG, 2001) but their composition was not precisely determined mainly due to a poor sampling (GPWG, 2001). The subfamily Aristidoideae, represented by *Aristida* and *Stipagrostis* in our analyses, were well supported and positioned as sister to a Chloridoideae/Danthonioideae group (Figure 5.1) although there was only low support for this. We can only tentatively suggest that Aristidoideae are sister to a Chloridoideae/Danthonioideae group. More taxa and character sampling are therefore required to confirm or refute this hypothesis. In our study, *Eriachne* and *Micraira* form a well-supported clade (Figure 5.1) but the position of this group within the PACCAD clade is not always strongly supported. The presence of an *Eriachne* + *Micraira* clade as sister to a chloridoid + danthonioid + aristidoid group is a novel result.

The relationships between major lineages of the PACCAD clade have not been fully resolved by previous studies of the family (Soreng and Davis, 1998; Hilu et al., 1999; GPWG, 2001). Our results suggest a strongly supported sister group relationship between Centothecoideae + Panicoideae and an Aristidoideae + Arundinoideae + Chloridoideae + Danthonioideae + *Eriachne*/*Micraira* group (Figure 5.1). The monophyly of Centothecoideae was not well supported regardless of the sample size (Figure 5.1). We also retrieve *Loudetiopsis* (tribe Arundinelleae) within Centothecoideae (Figure 5.1). In a previous study, *Loudetiopsis* was found sister to a Centothecoideae + Panicoideae group (Hilu et al., 1999). Panicoideae were well supported in most of our analyses (Figure 5.1). Two main clades within Panicoideae can be identified. One contains tribe Andropogoneae (with the inclusion of *Tristachya*) and four representatives of tribe Paniceae, which form a strongly supported clade that may be sister to the Andropogoneae (Figures 2.5 and 2.6).

Within the BEP clade, the three main subfamilies were generally well supported (Figure 5.1). Pooideae were strongly supported (Figures 2.1-2.6 and 5.1), Bambusoideae were monophyletic and sister to Pooideae, and Ehrhartoideae were found to be monophyletic and sister to the Bambusoideae + Pooideae group (Figure 5.1). Brachyelytreae were consistently the EDL in the pooids, and Nardeae + Lygeae and a group of Stipeae genera were generally successively sister to the rest of the pooids (Figure 5.1). Meliceae were strongly supported and were resolved as the next successive sister group to the rest of the pooids, followed by Brachypodieae (Figure 5.1). Catalan et al. (1997) could not find a supported order of divergence for these tribes. Our results resolved two main clades in the core pooids: one containing Bromeae + Triticeae and the other containing an assemblage of Poeae + Aveneae taxa. None of these tribes were monophyletic except Bromeae, which resolves as sister to Triticeae (Figure 5.1). A more extensive sampling of these tribes will be needed to determine the composition and inter-relationships of many of the major pooid groups.

Subfamily Bambusoideae s. str. (Bambuseae and Olyreae), were supported in all our analyses and were divided into two main lineages (Figure 5.1). One contains exclusively tribe Bambuseae with representatives of temperate woody bamboos, and the other contains Bambuseae, with representatives of neo-tropical and paleo-tropical woody bamboos, and Olyreae containing exclusively herbaceous bamboos.

Bambuseae are therefore not monophyletic. Previous phylogenetic studies have found a derived Olyreae lineage from within Bambusoideae s.str. (Clark et al., 1995) and a monophyletic Olyreae/Parianeae as sister to a monophyletic Bambuseae (Kelchner and Clark, 1997). Our results show that either Olyreae should be included within Bambuseae or some Bambuseae taxa should be included in Olyreae.

Ehrhartoideae were monophyletic in all our analyses and a sister relationship was found between tribes Ehrharteae and Oryzeae (Figure 5.1). Several studies using DNA data have shown that Oryzeae should be considered as a distinct entity (Barker et al., 1995; Clark et al., 1995; Soreng and Davis, 1998; Guo and Ge, 2005), but the inclusion of Ehrharteae was assessed only recently (Hilu et al., 1999; GPWG, 2001). The phylogenetic position of Ehrhartoideae was unclear in other studies (Hilu et al., 1999; GPWG, 2001), but our results support a sister relationship between Ehrhartoideae and Bambusoideae + Pooideae.

Finally, *Isachne* (tribe Isachneae) was found to be sister to the BEP clade but its position was not supported (Figure 5.1). The seemingly incorrect placement of this genus may be due either to the amount of missing data or to a misidentification (P-A Christin, personal comm.). *Isachne* has been considered as a member of Panicoideae (Clayton and Renvoize, 1986; Watson and Dallwitz, 1992), but this cannot be confirmed by our phylogenetic inferences. There is no evidence to embed it in the PACCAD clade and its subfamilial position is unclear.

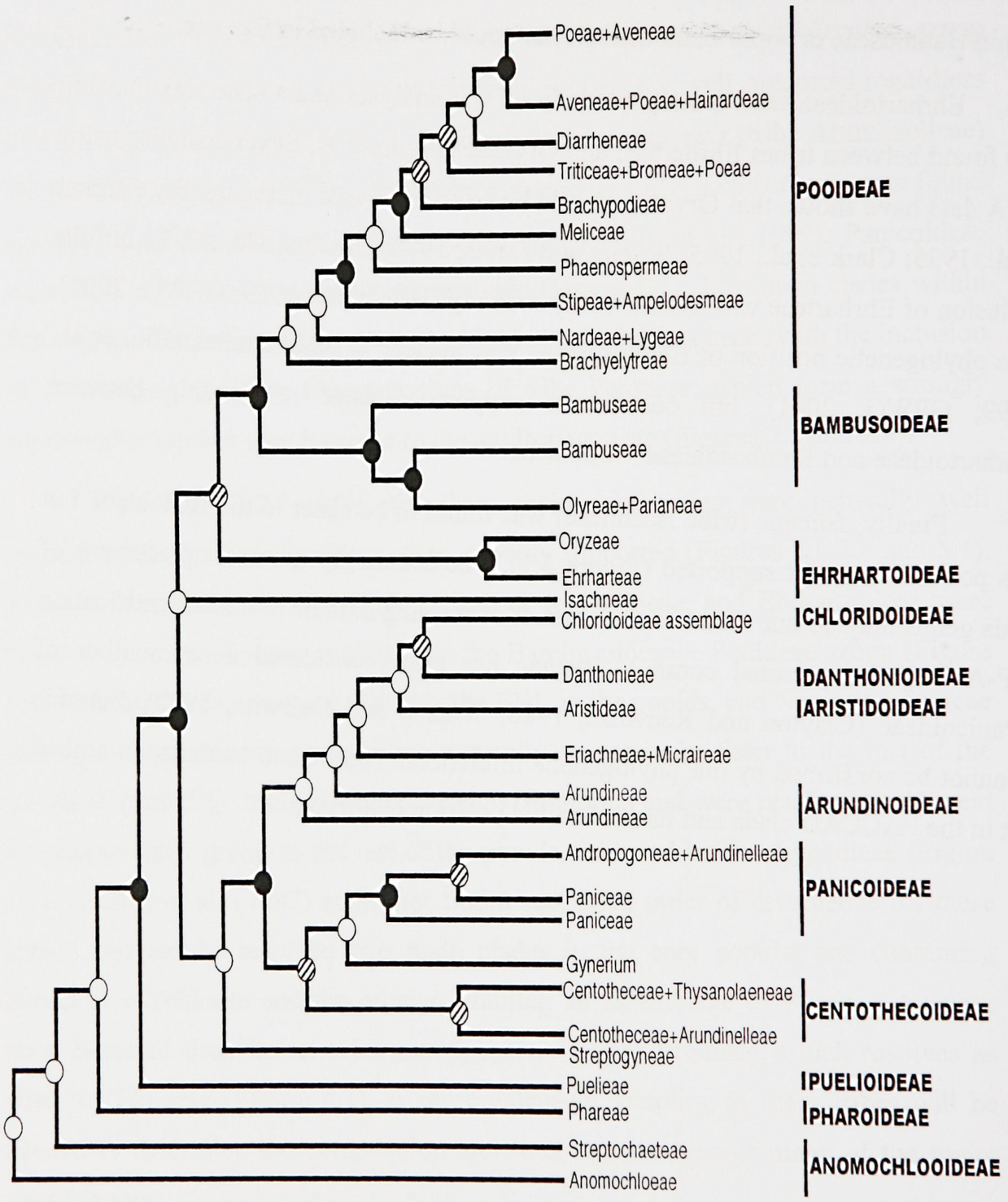


Figure 5.1 Summary tree showing sub-familial and inter-tribal relationships based on DataSets I and II. ○: weak support (<50% BS-75% BS; <0.80 PP), ⊘: moderate support (75% BS-90% BS; 0.80 PP-0.90 PP), and ●: strong support (>90% BS; >0.90PP).

5.2 Impacts of missing data on large phylogenetic trees

Previous theoretical (Hillis 1996a; Graybeal, 1998) and empirical (Soltis et al., 1999) studies have indicated that large numbers of characters may be necessary to resolve phylogenetic patterns in many groups of organisms. By increasing character number in our study from a range of 283-742 parsimony informative characters in the single gene analyses to a range of 1,107-1,403 in the combined analyses, we found more robust and resolved phylogenetic trees than with individual single-gene analyses (see section 2.3.2 of this thesis). Reducing misleading effects or systematic bias might be achieved by increased taxon sampling (Wiens, 1998; Hillis et al., 2003; Salamin et al., 2005). Some empirical studies have also found that data combination of multiple sequences from the same taxon (i.e. the multi-gene approach) does improve accuracy of phylogenetic inference (Qiu et al., 1999; Soltis et al., 1999; Bapteste et al., 2002). Our results in the combined analyses show high levels of congruence between the phylogenetic inferences with 107 and 294 taxa (the latter increased the proportion of all grass genera sampled from 15 to 42%; see section 2.3.2 of this thesis).

The impact of missing data (i.e. taxa for which there is a proportion of missing character states) has been neglected in phylogenetic analyses (Wiens, 2005). Trees from DataSet I showed high support values for most clades but the phylogenetic groupings determined from DataSet II were not as well supported. Wiens (2003) showed that reduced phylogenetic accuracy, resulting from the inclusion of missing data (i.e. incomplete taxa) was associated with incorrect placement of only the incomplete taxa. The lack of support of clades associated with analyses of DataSet II may have resulted from the poorly resolved placement of incomplete taxa. However, Wiens (2005) argued that adding taxa that are 50% incomplete might show similar benefits to adding complete taxa under many conditions. Our analyses support this assertion, with 33% of missing data, and suggest that adding incomplete taxa might show great benefits as long as their placement can be compared and checked for consistency with phylogenetic inferences including only complete taxa.

5.3 Grass diversification patterns

In terms of taxon numbers, the trees presented in this thesis are, to my knowledge, the most comprehensive for the grasses. These inferences allowed me to test for diversification shifts by removing bias with regards to incomplete sampling and missing taxa. The lack of temporal data using such phylogenetic tree was counterbalanced by the use of molecular dating method based on a subtree (see section 3.3.3 of Chapter 3 of this thesis). It enabled me to provide a timescale thanks to the congruent nodes between the complete generic level and the dated phylogenetic trees.

According to my molecular dating, the time of origin of grasses is estimated to the late Cretaceous (around 72 Mya), before the Cretaceous/Tertiary extinction event (K/T boundary). It is believed that terrestrial plants passed through the K/T boundary, with only minor taxonomic richness in comparison to today (Macleod et al., 1997). However, a recent study by Prasad et al. (2005) supports my finding and suggests that grasses may have diverged before the K/T boundary, and that at least some of the major subclades had already diversified. This contradicts recent molecular dating of the Angiosperms (Bell et al., 2005), even though it reveals that grasses have already dispersed from Africa by the late Cretaceous. My results indicate that the grasses may have originated in Africa, suggesting a Gondwanan origin of the family (see section 3.2.4 of Chapter 3 for details). This general Gondwanan origin hypothesis agrees with Bremer (2002), but the inference of an African origin is a novel result. However, the new finding of Prasad et al. (2005) would need further investigation in light of the results of this thesis. Indeed, it would be rather interesting to perform a "new" molecular dating using calibration from these newly found grass fossils and check for consistency with the age of the angiosperms recently inferred (Bell et al., 2005).

This study of grasses has detected fifteen statistically significant differential shifts in diversification (see section 3.2.2 section of Chapter 3 for details) among lineages during their evolution. Grasses also seem to have dispersed to all continents by 30 million years after their Gondwanan origin in the late Cretaceous. This is consistent with paleobotanical, paleofaunal, and stable carbon isotope records (Jacobs et al., 1999). My results indicate several major events in the evolution of the grasses including: (1) major diversification of the BEP clade members (C_3 grasses) in the Paleocene and Eocene (between 55 and 35 Mya) possibly due to the decline of forested environments, with dispersal routes from Africa to Asia and subsequently to

the New World, (2) later divergence of the PACCAD clade from the Oligocene (between 35 and 25 Mya), possibly due to an early adaptation to arid habitats with recent dispersals from Africa to Eurasia and to the New World, (3) diversification of grasses to become ecologically dominant in open environments, possibly due to adaptations to open habitats followed by numerous dispersals, and finally (4) relatively recent diversification within the PACCAD clade and the expansion of C_4 grasses occurring by the middle Miocene (between 15 and 10 Mya). Trying to correlate shifts in diversification with morphological and/or ecological characters, as attempted by Salamin and Davies (2004), remains a difficult task. Indeed, it is highly probable that sets of characters may be linked to higher diversification rates (Salamin and Davies, 2004). However, performing statistical tests is problematic because the number of nodes that exhibit significant shifts in diversification coupled with contrasting traits is small.

5.4 The case of grazers and grasses: a phylogenetic approach to the study of coevolution

A number of factors could be responsible for patterns of diversification in grass evolution. No studies have tried to assess if herbivores have had an impact on grass evolutionary history or to assess their coevolutionary processes in general. As described by Chapman (1996a), the development of phytoliths in grasses (silica bodies in epidermal cells) and their persistence could be a consequence of herbivore dentition changes to improve their ability to cope with an increasingly grass-based diet.

Silica bodies are among the few substances capable of inducing morphological changes to animal mouthparts (Piperno, 2006). Based on our results, the overall trend of variation in silica density through time can be summarized as follows: (i) there are differential responses of grasses in response to increased lophedness of ungulates through the Cenozoic, (ii) increase in silica density is correlated with the adaptation of grasses to closed-habitats, (iii) higher silica density changes occurred over time for PACCAD lineages than BEP, and (iv) C_4 grasses may have evolved increased silica density in response to increased grazing rates through the Miocene.

Indeed, the average lophedness of ungulates gradually increased from the late Oligocene through the Miocene. According to Stebbins (1981), grasses that grew in drier and more open-habitats (as are C_4 PACCAD grasses) evolved firmer and more silicaceous leaves. Indeed, from the late Oligocene to the late Miocene, PACCAD grasses evolved higher silica densities (Figure 4.2). Also, the average lophedness of ungulates gradually increased from the late Oligocene through the Miocene (Figure 4.4). It is tempting, based on these results, to suggest that this trend for PACCAD grasses may be linked to the spread of savannas and savanna-adapted mammals (Webb, 1989). According to Jernvall et al. (1996), the Miocene radiations of ungulates evolved dental adaptations to deal with vegetation of low primary productivity. It is then plausible that the Miocene ungulates evolved higher loph numbers on their 2^{nd} upper molar to deal with increasing silica density of C_4 grasses.

In contrast, the overall decrease in silica density of the BEP lineages throughout the Cenozoic may reflect the absence of diversified grass-eating mammals adapted to open-habitats until the Miocene (Janis, 1993). Perissodactyls and artiodactyls appeared by the early Eocene possibly in Asia (Stucky, 1990; Janis, 1989), but only perissodactyls were abundant in the early and middle Eocene (Prothero and Schoch, 1989). Their success may have been linked with their ability to process fibrous foliage of the Eocene tropical forests (Janis, 1989). As a consequence, Eocene grasses did not have to cope with heavy grazing pressures because plant diversity was relatively high during this period (Jernvall et al., 1996), allowing Eocene ungulates to process a higher diversity of foliage resources than Miocene ungulates adapted to open-habitats.

Even though, the deterrence of ruminant herbivory by silica bodies has not been demonstrated (Coughenour, 1985), this study reveals a parallel pattern between grass silica content and molar evolution of ungulates at least during the Miocene. Alternatively, one can speculate that grasses may tolerate grazing pressures by lowering their silica contents (by physiological control), which would allow to allocate more energy to other traits that deter grazing pressures (such as rhizomatous growth, increased fibre content and tannin-like substances) (Coughenour, 1985; Ellis, 1990). It is also possible that increased lophedness may be related to increasing fibrous content of the vegetation rather than silica density. Coughenour (1985)

suggests that the primary function of silica could be non-defensive even though it invariably wears down herbivore teeth. Our data support this hypothesis. Grasses lack chemical defenses against herbivory, and high concentrations of silica could be primarily involved in providing structural support and drought resistance rather than a means to cope with herbivory (Coughenour, 1985). Nevertheless, an array of non-chemical traits may help deter or tolerate grazing (Coughenour, 1985). The most challenging aspect for estimating the consequences of herbivory on grass evolution remains the precise selection of traits, which may be correlated with grass evolutionary response to herbivory.

5.5 Conclusions and perspectives

In this thesis, a multi-gene phylogenetic analysis of the grasses has been conducted with the largest sample size produced to date at tribal and generic levels. It represents a near complete tribal level phylogenetic treatment of the grasses. While there is a substantial amount of missing data in some of the combined analyses, the phylogenetic inferences showed a considerable topological congruence with our single-gene analyses, and the strongly supported topology with DataSet I (i.e. no missing data). The lack of BS support for groups determined in our analyses with missing data (DataSet II) reflects the need for a 'better and smarter' data acquisition in grass phylogenetic studies. The latter approach requires us to 'fill the gaps' of DNA data matrices which are now large enough to infer comprehensive phylogenetic trees of the family.

Our approach to incorporate different datasets (molecular, morphological, ecological and geographical), which have in common an overlap of taxa, have allowed a more detailed analysis of phylogenetic diversification than previous studies. It has helped reveal macro-evolutionary patterns and has proven to be a very powerful approach for dealing with species-rich groups of organisms such as grasses. The use of the comprehensive phylogenetic tree has allowed us to test hypotheses in regards to extrinsic and intrinsic processes leading to diversification and eventually to the present-day ecological success of these grasses. However, limitations have been encountered for testing intrinsic factors leading to diversification among lineages. We are currently developing methods to allow us to better analyse the contribution of such factors.

Finally, testing hypotheses of ungulates (and in a broader sense hypsodont mammal) and grass coevolution remains a very difficult task. Methodological limitations arose in the study presented in this thesis. However, our results show that C_4 PACCAD grasses evolved higher silica density of their leaf epidermal cells from the Oligocene throughout the Miocene. This increase was found to be correlated with increased lophedness of ungulates. It is plausible that C_4 grasses deter grazing by increasing their silica content whereas C_3 grasses (e.g. BEP lineages) did not deter grazing by increasing their silica content. One could also suppose that the appearance of open-habitats, such as savannas in the early Miocene, was correlated with increasing grazing pressures and therefore with increased silica density for open-habitat grasses. However, based on our sampling, our results do not support this hypothesis but rather show that increased silica density was correlated with adaptation to the closed-habitats. Larger sample size is necessary to further test this hypothesis. Selecting adaptive traits that may be correlated with herbivory defence is a rather difficult issue. Other anatomical characters, such as density of fibres or tannin-like substances, are thought to act as deterrent against herbivores (Ellis, 1990) and could, on the basis of our results, be more important in terms of grass-ungulate coevolution.

The applications of biogeographical, ecological, paleontological and taxonomic data coupled with phylogenetic trees provide robust perspectives to understanding evolutionary history of grasses, and to a larger extent species-rich groups of organisms. It is hoped that the approach taken in this thesis to use large phylogenetic trees to study macroevolutionary patterns and processes, will further our knowledge of this invaluable group of plants, and offer ways of better understanding other similarly species rich groups of organisms.

References

ACKERLY, D. D. (2000) Taxon sampling, correlated evolution, and independent contrasts. *Evolution,* 54, 1480-1492.

ALISCIONI, S. S., GUISSANI, L. M., ZULOAGA, F. O. & KELLOGG, E. A. (2003) A molecular phylogeny of *Panicum* (Poaceae: Paniceae): tests of monophyly and phylogenetic placement within the Panicoideae. *American Journal of Botany*, 90, 796-821.

AUSTIN, M. P., WILLIAMS, O. B. & BELBIN, L. (1981) Grassland dynamics under sheep grazing in an Australian mediterranean type climate. *Vegetatio,* 47, 201-211.

AVISE, J. C. (2000) *Phylogeography: the history and formation of species*, Harvard University Press, Boston, USA.

BALDWIN, B. G. & SANDERSON, M. J. (1998) Age and rate of diversification of the Hawaiian silversword alliance (Compositae). *Proceedings of the National Academy of Science USA,* 95, 9402-9406.

BAPTESTE, E., BRINKMANN, H., LEE, J. A., MOORE, D. V., SENSEN, C. W., GORDON, P., DURUFLE, L., GAASTERLAND, T., LOPEZ, P., MULLER, M. & PHILIPPE, H. (2002) The analysis of 100 genes supports the grouping of three highly divergent amoebae: *Dictyostelium*, *Entamoeba*, and *Mastigamoeba. Proceedings of the National Academy of Science USA,* 99, 1414-1419.

BARKER, N. P., LINDER, H. P. & HARLEY, E. H. (1995) Polyphyly of Arundinoideae (Poaceae): evidence from *rbcL* sequence data. *Systematic Botany*, 20, 423-435.

BARKER, N. P., LINDER, H. P. & HARLEY, E. H. (1999) Sequences of the grass-specific insert in the chloroplast *rpo*C2 gene elucidate generic relationships of the Arundinoideae (Poaceae). *Systematic Botany*, 23, 327-350.

BARRACLOUGH, T. G., HARVEY, P. H. & NEE, S. (1996) Rate of *rbcL* gene sequence evolution and species diversification in flowering plants (Angiosperms). *Proceedings of the Royal Society London Biological Sciences,* 263, 589-591.

BARRACLOUGH, T. G. & NEE, S. (2001) Phylogenetics and speciation. *Trends in Ecology and Evolution,* 16, 391-399.

BARRACLOUGH, T. G. & SAVOLAINEN, V. (2001) Evolution rates and species diversity in flowering plants. *Evolution,* 55, 677-683.

BELL, C. D., SOLTIS, D. E. & SOLTIS, P. S. (2005) The age of the angiosperms: a molecular timescale without a clock. *Evolution*, 59, 1245-1258.

BENTON, M. J. (1985) First marsupial fossil from Asia. *Nature,* 318, 313.

BERLOCHER, S. H. (1998) Can sympatric speciation be proven from phylogenetic and biogeographic evidence? *In* Howard D.J. (Ed.) *Endless forms: species and speciation.* Oxford University Press, Oxford, UK, 3-15.

BININDA-EMONDS, O. R. P., GITTLEMAN, J. L. & PURVIS, A. (1999) Building large trees by combining phylogenetic information: a complete phylogeny of the extant Carnivora (Mammalia). *Biological Review*, 74, 143-175.

BLUM, M. G. B. & FRANCOIS, O. (2006) Which random processes describe the Tree of Life? A large-scale study of phylogenetic tree imbalance. *Systematic Biology,* 55, 685-691.

BOND, W. J., MIDGLEY, G. F. & WOODWARD, F. I. (2003) The importance of low atmospheric CO_2 and fire in promoting the spread of grasslands and savannas. *Global Change Biology,* 9, 973-982.

BORTOLUSSI, N., DURAND, E., BLUM, M. G. B. & FRANCOIS, O. (2006) ApTreeshape: statistical analysis of phylogenetic treeshape. *Bioinformatics*, 22, 363-364.

BREMER, K. (1992) Ancestral areas: a cladistic reinterpretation of the center of origin concept. *Systematic Biology,* 41, 436-445.

BREMER, K. (2002) Gondwanan evolution of the grass alliance of families (Poales). *Evolution,* 56, 1374-1387.

BROOKS, D. R. & MCLENNAN, D. A. (1991) Phylogeny, Ecology, and Behaviour: a Research Program in Comparative Biology. *The Quarterly Review of Biology*, 66, 336-337.

BROWN, R. (1810) *Prodomus florae Novae Hollandiae et insulae Van-Diemen,* London, UK.

BROWN, R. (1814) General remarks, geographical and systematical, on the botany of Terra Australis. *In* Flinders M. (Ed.) *A Voyage to Terra Australis.* London, UK, 533-612.

BUTLER, M. A. & KING, A. A. (2004) Phylogenetic comparative analysis: a modeling approach for adapative evolution. *American Naturalist*, 164, 683-695.

CALDERON, C. E. & SODERSTROM, T. R. (1980) The genera of Bambusoideae (Poaceae) of the American continent: keys and comments. *Smithsonian,* 44, 1-27.

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

CARO, J. A. (1982) Sinopsis taxonomica de las gramineas argentinas. *Dominguezia,* 4, 1-51.

CATALAN, P., KELLOGG, E. A. & OLMSTEAD, R. G. (1997) Phylogeny of Pooideae subfamily based on chloroplast *ndhF* gene sequences. *Molecular Phylogenetics and Evolution,* 8, 150-166.

CERLING, T. E., HARRIS, J. M., MACFADDEN, B. J., LEAKEY, M. G., QUADE, J., EISENMANN, V. & EHLERINGER, J. R. (1997) Global vegetation change through the Miocene/Pliocene boundary. *Nature,* 389, 153-158.

CERLING, T. E., EHLERINGER, J. R. & HARRIS, J. M. (1998) Carbon dioxide starvation, the development of C_4 ecosystems, and mammalian evolution. *Phylosophical Transactions of the Royal Society, B,* 353, 159-171.

CHAN, K. M. A. & MOORE, B. R. (2005) SYMMETREE: whole-tree analysis of differential diversification rates. *Bioinformatics,* 21, 1709-1710.

CHANDERBALI, A. S., VAN DER WERFF, H. & RENNER, S. S. (2001) Phylogeny and historical biogeography of Lauraceae: evidence from the chloroplast and nuclear genomes. *Annals of the Missouri Botanical Garden,* 88, 104-134.

CHAPMAN, G. P. (1996a) From extinct to present-day grasses. *In* Chapman G.P. (Ed.) *The Biology of Grasses.* CAB international, Oxon, UK, 112-123.

CHAPMAN, G. P. (1996b) *The biology of grasses,* Chapman G.P. (Ed.) CAB international, Oxon, UK.

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

CHASE, M. W., FAY, M. F. & SAVOLAINEN, V. (2000) Higher-level classification in the angiosperms: new insights from the perspective of DNA sequences data. *Taxon,* 49, 685-704.

CHASE, M. W., FAY, M. F., SOLTIS, D. E. & SOLTIS, P. S. (2006) Simple phylogenetic tree searches easily succeed with large matrices of single genes. *Taxon,* 55, 573-578.

CLARK, L. G., ZHANG, W. & WENDEL, J. F. (1995) A phylogeny of the grass family (Poaceae) based on *ndhF* sequence data. *Systematic Botany,* 20, 436-460.

CLARK, L. G. & JUDZIEWICZ, E. J. (1996) The grass subfamilies: Anomoochlooideae and Pharoideae (Poaceae). *Taxon,* 45, 641-645.

CLARK, L. G., KOBAYASHI, M., MATHEWS, S., SPANGLER, R. E. & KELLOGG, E. A. (2000) The Puelioideae, a new subfamily of Poaceae. *Systematic Botany,* 25, 181-187.

CLAYTON, W. D. & RENVOIZE, S. A. (1986) *Genera Graminum, Grasses of the World,* Her Majesty's Stationery Office (Ed.), London, UK.

COLLESS, D. H. (1982) Review of phylogenetics: the theory and practice of phylogenetic systematics. *Systematic Zoology*, 31, 100-104.

COLLINSON, M. E. & HOOKER, J. J. (1991) Fossil evidence of interactions between plants and plant-eating mammals. *Philosophical Transactions of the Royal Society, B,* 333, 197-208.

COUGHENOUR, M. B. (1985) Graminoid responses to grazing by large herbivores: Adaptations, Exaptations, and Interacting processes. *Annals of the Missouri Botanical Garden,* 72, 852-863.

COYNE, J. A. & ORR, H. A. (2004) *Speciation.* Sinauer (Ed.), Publishers Sunderland, Massachusetts, USA.

CRAWFORD, D. J. (1978) Flavonoid chemistry and angiosperm evolution. *Botanical Review,* 44, 431-456.

CREPET, W. L. & FELDMAN, G. D. (1991) The earliest remains of grasses in the fossil record. *American Journal of Botany,* 78, 1010-1014.

CUENOUD, P., SAVOLAINEN, V., CHARTROU, L. W., POWELL, M., GRAYER, R. J. & CHASE, M. W. (2002) Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid *rbcL*, *atpB*, and *matK* DNA sequences. *American Journal of Botany*, 89, 132-144.

DAVIES, T. J., BARRACLOUGH, T. G., CHASE, M. W., SOLTIS, P. S., SOLTIS, D. E. & SAVOLAINEN, V. (2004) Darwin's abominable mystery: insights from a supertree of the angiosperms. *Proceedings of the National Academy of Science USA,* 101, 1904-1909.

DIAL, K. P. & MARZLUFF, J. M. (1989) Nonrandom diversification within taxonomic assemblages. *Systematic Zoology,* 38, 26-37.

DOEBLEY, J., DURBIN, M., GOLENBERG, E. M., CLEGG, M. T. & MA, D. P. (1990) Evolutionary analysis of the large subunit of carboxylase (*rbcL*) nucleotide sequence data among the grasses (Poaceae). *Evolution,* 44, 1097-1108.

DOLORESLLEDO, M., CRESPO, M. B., CAMERON, K. M., FAY, M. F. & CHASE, M. W. (1998). Systematics of Plumbaginaceae based upon cladistic analysis of *rbcL* sequence data. *Systematic Botany*, 23, 21-29.

DOUST, A. N. & KELLOGG, E. A. (2002) Inflorescence diversification in the panicoid "bristle grass" clade (Paniceae, Poaceae): evidence from molecular phylogenies and developmental morphology. *American Journal of Botany*, 89, 1203-1222

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

DOYLE, J.J. and DOYLE, J. L. (1987). A rapid isolation procedure for small quantities of fresh leaf tissue. *Phytochemistry Bulletin*, 19, 65-66.

DUGAS, D. P. & RETALLACK, G. J. (1993) Middle Miocene fossil grasses from Fort Ternan, Kenya. *Journal of Paleontology,* 67, 113-128.

DUVALL, M. R. & MORTON, B. R. (1996) Molecular phylogenetics of Poaceae: an expanded analysis of *rbcL* sequence data. *Molecular Phylogenetics and Evolution,* 5, 352-358.

DUVALL, M. R., NOLL, J. D. & MINN, A.H. (2001) Phylogenetics of Paniceae. *American Journal of Botany*, 88, 1988-1992

EHRLICH, P. R. & RAVEN, P. H. (1964) Butterflies and plants: a study in coevolution. *Evolution,* 18, 586-608.

ELLIS, R. P. (1990) Tannin-like substances in grass leaves. *Memoirs of the Botanical Survey of South Africa.* Pretoria, South Africa.

EPSTEIN, E. (1999) Silicon. *Annual Review of Plant Physiology and Plant Molecular Biology,* 50, 641-664.

ERICSON, P. G. P., CHRISTIDIS, L., COOPER, A., IRESTEDT, M., JACKSON, J., JOHANSSON, U. S. & NORMAN, J. A. (2002) A Gondwanan origin of passerine birds supported by DNA sequences of the endemic New Zealand wrens. *Proceedings of the Royal Society London, B,* 269, 235-241.

EYDE, R. H. (1975) The bases of angiosperm phylogeny: Floral anatomy. *Annals of the Missouri Botanical Garden,* 62, 521-537.

FARRELL, B. D. & MITTER, C. (1998) The timing of insect/plant diversification : might *Tetraopes* (Coleoptera : Cerambycidae) and *Asclepias* (Asclepiadaceae) have co-evolved? *Biological Journal of the Linnean Society,* 63, 553-577.

FELSENSTEIN, J. (1978) Cases in which parsimony and compatibility methods will be positively misleading. *Systematic Zoology,* 27, 401-410.

FELSENSTEIN, J. (1984) Distance methods for inferring evolutionary trees: a justification. *Evolution,* 38, 16-24.

FELSENSTEIN, J. (1985a) Confidence limits on phylogenies: An approach using the bootstrap. *Evolution,* 39, 783-791.

FELSENSTEIN, J. (1985b) Phylogenies and the comparative method. *American Naturalist,* 125, 1-15.

FRAKES, L. A. & VICKERS-RICH, P. (1991) Paleoclimatic setting and paleogeographic links of Australia in the Phanerozoic. *In* Vickers-Rich, P., Baird, R. F. & Rich, T. H. (Ed.) *Vertebrate Paleontology of Australasia.* Pioneer Design Studios, Melbourne, Australia, 111-146.

GADAGKAR, S. R., ROSENBERG, M. S. & KUMAR, S. (2005) Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. *Journal of Experimental Zoology part B-Molecular and Development al Evolution*, 304, 64-74.

GALI-MUHTASIB, H. U., SMITH, C. C. & HIGGINS, J. J. (1992) The effects of silica in grasses on the feeding behavior oh the prairie vole, *Microtus ochrogatser. Ecology,* 73, 1724-1729.

GOULD, F. W. & SHAW, R. B. (1983) *Grass Sytematics.* Texas A&M University Press, 2nd edition, College Station, TX, USA.

GPWG (2001) Phylogeny and subfamilial classification of the grasses (Poaceae). *Annals of the Missouri Botanical Garden,* 88, 373-457.

GRAYBEAL, A. (1998) Is it better to add taxa or characters to a difficult phylogenetic problem? *Systematic Biology*, 47, 9-17.

GUO, Y. L. & GE, S. (2005) Molecular phylogeny of Oryzeae (Poaceae) based on DNA sequences from chloroplast, mitochondrial, and nuclear genomes. *American Journal of Botany*, 92, 1548-1558.

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

HARDING, E. F. (1971) The probabilities of rooted tree-shapes generated by random bifurcation. *Advances in Applied Probability,* 3, 44-77.

HARRIS, T. E. (1964) *The theory of branching Process.* Prentice-Hall, Englewood Cliffs, New Jersey, USA.

HASEGAWA, M., KISHINO, H. & YANO, T. (1985) Dating the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution,* 22, 160-174.

HASSANIN, A. & DOUZERY, E. J. P. (1999) The tribal radiation of the family Bovidae (Artiodactyla) and the evolution of the mitochondrial *cytochrome b* gene. *Molecular Phylogenetics and Evolution,* 13, 227-243.

HENNIG, W. (1966) *Phylogenetic Systematics*, University of Illinois Press. Illinois, USA.

HEY, J. (1992) Using phylogenetic trees to study speciation and extinction. *Evolution,* 46, 627-640.

HICKEY, L. J. & WOLFE, J. A. (1975) The bases of angiosperm phylogeny: vegetative morphology. *Annals of the Missouri Botanical Garden,* 62, 538-589.

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

HILLIS, D. M. (1996a) Inferring complex phylogenies. *Nature,* 383, 130-131.

HILLIS, D. M. (1996b) *Molecular Systematics,* Sinauer Associates Inc., Massachusetts, USA.

HILLIS, D. M., POLLOCK, D. D., MCGUIRE, J. A. & ZWICKL, D. J. (2003) Is sparse taxon sampling a problem for phylogenetic inference? *Systematic Biology*, 52, 124-126.

HILU, K. W. & ALICE, L. A. (2001) A phylogeny of Chloridoideae (Poaceae) based on *matK* sequences. *Systematic Botany*, 26, 386-405.

HILU, K. W., ALICE, L. A. & LIANG, H. (1999) Phylogeny of Poaceae inferred from *matK* sequences. *Annals of the Missouri Botanical Garden,* 86, 835-851.

HODKINSON, T. R., CHASE, M.W., DOLORESLLEDO, M. D., SALAMIN, N. & RENVOIZE, S. A. (2002) Phylogenetics of *Miscanthus*, *Saccharum* and related genera (Saccharinae, Andropogoneae, Poaceae) based on DNA sequences from *ITS* nuclear ribosomal DNA and plastid *trnL* intron and *trnL-F* intergenic spacer. *Journal of Plant Research*, 115, 381-392.

HODKINSON, T. R., SALAMIN, N., CHASE, M. W., BOUCHENAK-KHELLADI, Y., RENVOIZE, S. A. & SAVOLAINEN, V. (2007a) Large Trees, Supertrees and diversification of the grass family. *Aliso*, 23, *in press.*

HODKINSON, T. R., SAVOLAINEN, V., JACOBS, S. W., BOUCHENAK-KHELLADI, Y., KINNEY, M. S. & SALAMIN, N. (2007b) Supersizing: Progress in documenting and understanding grass richness. *In* Hodkinson T.R. and Parnell J.A.N. (Eds) *Reconstructing the Tree of Life: taxonomy and systematics of species Rich Groups*, Systematics Association Special Series Volume 72, CRC Press, Boca Raton, FL, USA, 279-298.

HSIAO, C., JACOBS, S. W. L., CHATTERTON, N. J. & ASAY, K. H. (1999) A molecular phylogeny of the grass family (Poaceae) based on the sequences of nuclear ribosomal DNA (*ITS*). *Australian Systematic Botany,* 11, 667-688.

HUELSENBECK, J. P., BULL, J. J. & CUNNINGHAM, C. W. (1996) Combining data in phylogenetics analysis. *Trends in Ecology and Evolution*, 11, 152-158.

HUELSENBECK, J. P. & RONQUIST, F. (2001). MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17, 754-755.

JACOBS, B. F., KINGSTON, J. D. & JACOBS, L. L. (1999) The origin of grass-dominated ecosystems. *Annals of the Missouri Botanical Garden,* 86, 590-643.

JANIS, C. M. (1989) A climatic explanation for patterns of evolutionary diversity in ungulate mammals. *Palaeontology,* 32, 463-481.

JANIS, C. M. (1993) Tertiary mammal evolution in the context of changing climates, vegetation, and tectonic events. *Annual Review of Ecology and Systematics,* 24, 467-500.

JANIS, C. M., DAMUTH, J. & THEODOR, J. M. (2002) The origins and evolution of the North American grassland biome: the story from the hoofed mammals. *Palaeogeography, Palaeoclimatology, Palaeoecology,* 177, 183-198.

JEFFROY, O., BRINKMANN, H., DELSUC, F. & PHILIPPE, H. (2006) Phylogenomics: the beginning of incongruence? *Trends in Genetics,* 22, 225-231.

JERISON, H. J. (1973) *Evolution of the brain and intelligence,* Academic Press, New York, USA.

JERNVALL, J., HUNTER, J. P. & FORTELIUS, M. (1996) Molar tooth diversity, disparity, and ecology in Cenozoic Ungulate radiations. *Science,* 274, 1489-1492.

JOHANSEN, D. A. (1940) *Plant Microtechnique,* McGraw-Hill, NY, USA.

JOHNSON, L. A. & SOLTIS, D. E. (1994) *MatK* DNA sequences and phylogenetic reconstruction in Saxifragaceae s.str. *Systematic Botany*, 19, 143-156.

JUDZIEWICZ, E. J. & SODERSTROM, T. R. (1989) Morphological, anatomical, and taxonomic studies in *Anomochloa* and *Streptochaeta* (Poaceae: Bambusoideae). *Smithsonian,* 68, 1-52.

KELCHNER, S. A. & CLARK, L. G. (1997) Molecular evolution and phylogenetic utility of the chloroplast *rpl16* intron in *Chusquea* and the Bambusoideae (Poaceae). *Molecular Phylogenetics and Evolution*, 8, 385-397.

KELLOGG, E.A. (2001) Evolutionary history of the grasses. *Plant Physiology*, 125, 1198-1205.

KELLOGG, E. A. (2000) The grasses: a case study in macroevolution. *Annual Review of Ecology and Systematics,* 31, 217-238.

KENDALL, D. G. (1948) On the generalized birth-and-death process. *Annals of Mathematical Statistics,* 19, 1-15.

KIMURA, M. (1981) Estimation of evolutionary distances between homologous nucleotide sequences. *Proceeding of the National Acadmy of Science USA,* 78, 454-458.

KINGSTON, J. D., MARINO, B. D. & HILL, A. (1994) Isotopic evidence for Neogene hominid paleoenvironments in the Kenya Rift Valley. *Science,* 264, 955-959.

KISHINO, H., THORNE, J. L. & BRUNO, W. J. (2001) Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Molecular Biology and Evolution,* 18, 352-361.

KLOOTWIJK, C. T., GEE, J. S., PEIRCE, J. W., SMITH, G. M. & MCFADDEN, P. L. (1992) An early India-Asia contact: paleomagnetic constraints from Ninetyeast Ridge, ODP Leg 121. *Geology,* 20, 395-398.

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

KNOBLOCH, E., KVACEK, Z., BUZEK, C., MAI, D. H. & BATTEN, D. J. (1993) Evolutionary significance of floristic changes in the Northern Hemisphere during the late Cretaceous and Paleogene, with particular reference to Central Europe. *Review of Paleobotany and Palynology,* 78, 41-54.

LAMARCK, J-B. P. (1809) *Philosophie zoologique*. Dentu, Paris, France.

LAPOINTE, F. J. & LEVASSEUR, C. (2004) Everything you always wanted to know about the average consensus, and more. *In* Bininda-Emonds, O. R. P. (Ed.) *Phylogenetic supertrees: combining information to reveal the Tree of Life.* Kluwer Academic, Dordrecht, Netherlands, 87-105.

LATORRE, C., QUADE, J. & MCINTOSH, W. C. (1997) The expansion of C4 grasses and global change in the late Miocene: Stable isotope evidence from the Americas. *Earth and Planetary Science Letters,* 146, 83-96.

LEHMANN, E.L. and D'ABRERA H.J.M. (1998). Non parametrics: statistical methods based on ranks. Prentice-Hall, Englewood Cliffs, NJ, USA.

LI, J., BOGLE, A. L., KLEIN, A. S. & DONOGHUE, M. J. (2000) Phylogeny and biogeography of *Hamamelis* (Hamamelidaceae). *Harvard Papers in Botany,* 5, 171-178.

LOSOS, J. B. (1999) Uncertainty in the reconstruction of ancestral character states and limitations on the use of phylogenetic comparative methods. *Animal Behaviour,* 58, 1319-1324.

LU, H.. & LIU, K. (2003) Phytoliths of common grasses in the coastal environments of southeastern USA. *Estuarine, Coastal and Shelf Science*, 58, 587-600.

LUNAU, K. (2004) Adaptive radiation and coevolution-pollination biology case studies. *Organisms, Diversity and Evolution,* 4, 207-224.

LYDEARD, C. (1993) Phylogenetic analysis of species richness: has viviparity increased the diversification of actinopterygian fishes? *Copeia*, 514-518.

MACFADDEN, B. J. & CERLING, T. E. (1994) Fossil horses, carbon isotopes and global change. *Trends in Ecology and Evolution,* 9, 481-486.

MACLEOD, N., RAWSON, P. F., FOREY, P. L., BANNER, F. T., BOUDAGHERFADEL, M. K., BOWN, P. R., BURNETT, J. A., CHAMBERS, P., CULVER, S., EVANS, S. E., JEFFERY, C., KAMINSKI, M. A., LORD, A. R., MILNER, A. C., MORRIS, N., OWEN, E., ROSEN, B. R., SMITH, A. B., TAYLOR, P. D., URGUHART, E. & YOUNG, J. R. (1997) The Cretaceous-Tertiary biotic transition. *Journal of the Geological Society,* 154, 265-292.

MADDISON, D. R. & MADDISON, W. P. (2005) MacClade 4. Version 4.08 for OSX. *In* Sinauer Associates (Ed.), Sunderland, Massachusetts, USA.

MASSEY, F.P. & HARTLEY, S.E. (2006) Experimental demonstration of the antiherbivore effects of silica in grasses: impacts ofn foliage digestibility and vole growth rates. *Proceedings of the Royal Society B-Biological Sciences*, 273, 2299-2304.

MATTHEWS, S., SPANGLER, R. E., MASON-GRAMER, R. J. & KELLOGG, E. A. (2002) Phylogeny of Andropogoneae inferred from *phytochrome b*, *gbbsi* and *ndhF*. *International Journal of Plant Science*, 163, 441-450.

MAYNARD SMITH, J. & SZATHMARY, E. (1995) *The major transitions in evolution.* Oxford University Press, Oxford, UK.

METCALFE, C. R. (1960) *Anatomy of the Monocotyledons. Volume I. Gramineae,* Oxford, UK.

MITSUI, S. & TAKATOH, H. (1963) Nutritional study of silicon in graminaceous crops. *Soil Science and Plant Nutrition,* 9, 49-53.

MOLVRAY, M. P., KORES, P. J. & CHASE, M. W. (2000) Polyphyly of mycoheterotrophic orchids and functional inflyences on floral and molecular characters. *In* Wilson, K. L., Morrison, D.A. (Eds), *Monocots: Systematics and Evolution*, CSIRO, Melbourne, Australia, pp. 441-448.

MOOERS, A. O., & HEARD, S. B. (1997) Inferring evolutionary process from phylogenetic tree shape. *The Quarterly Review of Biology*, 72, 31-54.

MOORE, B. R., CHAN, K. M. A. & DONOGHUE, M. J. (2004) Detecting diversification rate variation in supertrees. *In* Bininda-Emonds, O. R. P. (Ed.) *Phylogenetic supertrees: combining information to reveal the trees of life.* Kluwer Academic, Dordrecht, Netherlands, 487-532.

MOTOMURA, H., MITA, N. & SUZUKI, M. (2002) Silica accumulation in long-lived leaves of *Sasa veitchii* (Carriere) Rehder (Poaceae-Bambusoideae). *Annals of Botany,* 90, 149-152.

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

NAMBUDIRI, E. M. V., TIDWELL, W. D., SMITH, B. N. & HEBBERT, N. P. (1978) A C4 plant from the Pliocene. *Nature,* 276, 816-817.

NEE, S., HOLMES, E. C., MAY, R. M. & HARVEY, P. H. (1994) Extinction rates can be estimated from molecular phylogenies. *Philosophical Transactions of the Royal Society, B,* 344, 77-82.

NEE, S. (2001) Inferring speciation rates from phylogenies. *Evolution,* 55, 661-668.

NYLANDER, J. A. A., RONQUIST, F., HUELSENBECK, J. P. & NIEVES-ALDREY, J. L. (2004) Bayesian phylogenetic analysis of combined data. *Systematic Biology*, 53, 47-67.

PARADIS, E. (1997) Assessing temporal variations in diversification rates from phylogenies: estimation and hypothesis testing. *Proceeding of the Royal Society London, B,* 264, 1141-1147.

PARADIS, E. (1998) Detecting shifts in diversification rates without fossils. *American Naturalist,* 152, 176-187.

PARADIS, E. (2003) Analysis of diversification: combining phylogenetic and taxonomic data. *Proceeding of the Royal Society London, B,* 270, 2499-2505.

PARADIS, E. & CLAUDE, J. (2002) Comparative data and GEE. *Journal of Theoretical Biology,* 218, 175-185.

PARADIS, E., CLAUDE, J. & STRIMMER, K. (2004) APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics,* 20, 289-290.

PETERSEN, G. & SEBERG, O. (1997) Phylogenetics analysis of Triticeae (Poaceae) based on *rpoA* sequence data. *Molecular Phylogenetics and Evolution*, 7, 217-230.

PIPERNO, D. R. (2006) *Phytoliths: a comprehensive guide for Archeologists and Paleoecologists,* Lanham, MD, USA.

PIPERNO, D. R. & SUES, H. S. (2005) Dinosaurs dined on grass. *Science,* 310, 1126-1128.

POSADA, D. & CRANDALL, K. (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics,* 14, 817-818.

PRASAD, V., STROMBERG, C. A. E., ALIMOHAMMADIAN, H. & SAHNI, A. (2005) Dinosaur coprolites and the early evolution of grasses and grazers. *Science,* 310, 1177-1180.

PRAT, H. (1932) L'epiderme des graminees: etude anatomique et systematique. *Annales des Sciences Naturelles-Botanique,* 14, 117-324.

PRAT, H. (1960) Vers une classification naturelle des Graminees. *Bulletin de la Societe Botanique France,* 107, 32-79.

PRICE, R.G. (2006) Understanding evolution: History, theory, evidence and implications. Price R. G. (Eds), *www.rationalevolution.net*, USA.

PROTHERO, D. R. & BERGGREN, W. A. (1992) *Eocene-Oligocene climatic and biotic evolution,* Princeton, NJ, USA.

PROTHERO, D. R. & SCHOCH, R. M. (1989) *The evolution of Perissodactyls,* Oxford University Press, Oxford, UK.

PURVIS, A. (1996) Using interspecies phylogenies to test macroevolutionary hypotheses. *In* Harvey P.H., Leigh Brown A.J., Maynard Smith J. and Nee S. (Eds.), *New Uses for New Phylogenies.* Oxford University Press, Oxford, UK, 153-168.

QIU, Y. L., LEE, J., BERNASCONI-QUADRONI, F., SOLTIS, D. E., SOLTIS, P. S., ZANIS, M., ZIMMER, E. A., CHEN, Z., SAVOLAINEN, V. & CHASE, M. W. (1999) The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature,* 402, 404-407.

RAGE, J.-C. (1996) Terrestrial trans-Tethyan dispersals: An overview. *In First European Workshop on Vertebrate Paleontology, DGF Online Series Volume 1,* Copenhagen, Denmark.

RAUP, D. M., GOULD, S. J., SCHOPF, T. J. M. & SIMBERLOFF, D. S. (1973) Stochastic models of phylogeny and the evolution of diversity. *Journal of Geology,* 81, 525-542.

RAUSHER, M. D. (1996) Genetic analysis of coevolution between plants and their natural enemies. *Trends in Genetics,* 12, 212-217.

RAVEN, P. H. & AXELROD, D. I. (1974) Angiosperm biogeography and past continental movements. *Annals of the Missouri Botanical Garden,* 61, 539-637.

REE, R. H., MOORE, B. R., WEBB, C.O. & DONOGHUE, M. J. (2005) A likelihood framework for inferring the evolution of geographical range on phylogenetic trees. *Evolution*, 59, 2299-2311

REEVES, G., CHASE, M. W., GOLDBLATT, P., RUDALL, P., FAY, M. F., COX, A. V., LEJEUNE, B. & SOUZA-CHIES, T. (2001) Molecular systematics of Iridaceae: evidence from four plastid DNA regions. *American Journal of Botany*, 88, 2074-2087.

RENNER, S. S., FOREMAN, D. B. & MURRAY, D. (2000) Timing transantartic disjunctions in the Atherospermataceae (Laurales): Evidence from coding and noncoding chloroplast sequences. *Systematic Biology,* 49, 579-591.

RENVOIZE, S. A. & CLAYTON, W. D. (1992) Classification and evolution of the grasses. *In* Chapman G.P. (Ed.) *Grass evolution and domestication.* Cambridge University Press, Cambridge, UK, 3-37.

RIDLEY, M. (2004) *Evolution.* 3rd edition, Blackwell Publishing Company, Oxford, UK.

RONQUIST, F. (1996) DIVA, ver. 1.1. Computer program and manual available by anonymous FTP from Uppsala University (ftp.uu.se or ftp.systbot.uu.se).

RUDALL, P., STUPPY, W., CUNNIFF, J., KELLOGG, E. A. & BRIGGS, B. G. (2005) Evolution of reproductive structures in grasses (Poaceae) inferred by sister-group comparison with their putative closest living relatives, Ecdeiocoleaceae. *American Journal of Botany,* 92, 1432-1443.

RYAN, M.J.& RAND A.S. (1999) Phylogenetic influences on mating call preferences in female tœngara frogs (*Physalaemus pustulosus*). *Animal Behaviour*, 57, 945-956.

SAGE, R. F. & MONSON, R. K. (1999) *C4 plant biology.* Academic Press San Diego, CA, USA.

SALA, O. E., OESTERHEDT, M., LEON, R. J. & SORIANO, A. (1986) Grazing effects upon plant community structure in subhumid grassland of Argentina. *Vegetatio,* 67, 27-32.

SALAMIN, N., HODKINSON, T. R. & SAVOLAINEN, V. (2002) Building supertrees: an empirical assessment using the grass family (Poaceae). *Systematic Biology,* 51, 136-150.

SALAMIN, N. & DAVIES, T. J. (2004) Using supertrees to investigate species richness in grasses and flowering plants. *In* Binida-Emonds, O.R.O. (Ed) *Phylogenetic supertrees: combining information to reveal the Tree of Life,*

Computational Biology Series vol.3, Kluwer Academic Publishers, Netherlands, 461-487.

SALAMIN, N., HODKINSON, T. R. & SAVOLAINEN, V. (2005) Towards building the tree of life: a simulation study for all angiosperm genera. *Systematic Biology,* 54, 183-196.

SANDERSON, M. J. & BHARATHAN, G. (1993) Does cladistic information affect inferences about branching rates? *Systematic Biology,* 42, 1-17.

SANDERSON, M. J. & DONOGHUE, M. J. (1994) Shifts in diversification rate with the origin of angiosperms. *Science,* 264, 1590-1593.

SANDERSON, M. J. & DONOGHUE, M. J. (1996) Reconstructing shifts in diversification rates on phylogenetic trees. *Trends in Ecology and Evolution,* 11, 15-19.

SANDERSON, M. J. & WOJCIECHOWSKI, M. F. (1996) Diversification rates in a temperate legume clade-why are there so many species of *Astragalus* (Fabaceae)? *American Journal of Botany,* 83, 1488-1502.

SANDERSON, M. J. (1997) A nonparametric approach to estimating divergence times in the absence of rate constancy. *Molecular Biology and Evolution,* 14, 1218-1231.

SANDERSON, M. J. & DRISKELL, A. C. (2003) The challenges of constructing large phylogenetic trees. *Trends in Plant Science,* 8, 374-379.

SANMARTIN, I. & RONQUIST, F. (2004) Southern hemisphere biogeography inferred by event-based models: Plant versus Animal patterns. *Systematic Biology,* 53, 216-243.

SAVOLAINEN, V., HEARD. S.B., POWELL, M.P., DAVIES, T.J. & MOOERS, A.O. (2002) Is cladogenesis heritable? *Systematic Biology,* 51, 835-843.

SAVOLAINEN, V. & CHASE, M. W. (2003) A decade of progress in plant molecular phylogenetics. *Trends in Genetics,* 19, 717-724.

SAVOLAINEN, V., CHASE, M. W., HOOT, S. B., MORTON, C. M., SOLTIS, D. E., BAYER, C., FAY, M. F., DE BRUIJN, A. Y., SULLIVAN, S. & QIU, Y. L. (2000) Phylogenetics of flowering plants based on combined analysis of plastid *atpB* and *rbcL* gene sequences. *Systematic Biology,* 49, 306-362.

SCHLUTER, D. M. (2000) *The ecology of adaptive radiation.* Oxford University Press, Oxford, UK.

SCHLUTER, D. M., PRICE, T. & MOOERS, A. O. (1997) Likelihood of ancestor states in adaptive radiation. *Evolution,* 51, 1699-1711.

SHANTZ, H. L. (1954) The place of grasslands in the earth's cover of vegetation. *Ecology,* 35, 143-145.

SHEAHAN, M. C. & CHASE, M. W. (2000) Phylogenetic relationships within Zygophyllaceae based on DNA sequences of three plastid regions, with special emphasis on Zygophylloideae. *Systematic Botany*, 25, 371-384.

SHIMODAIRA, H. & HASEGAWA, M. (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution*, 16, 1114-1116.

SIMMONS, M. P., PICKETT, K. M. & MIYA, M. (2004) How meaningful are Bayesian support values? *Molecular Biology and Evolution,* 21, 188-199.

SINGH, A. (1988) History of aridland vegetation and climate: a global perspective. *Biological Review,* 63, 156-198.

SLOWINSKI, J. B. & GUYER, C. (1989a) Testing the stochasticity of patterns of organismal diversity: an improved null model. *American Naturalist*, 134, 907-921.

SLOWINSKI, J. B. & GUYER, C. (1989b) Testing null models in questions of evolutionary success. *Systematic Zoology*, 38, 189-191.

SODERSTROM, T. R., ELLIS R.P. & JUDZIEWICZ, E. J. (1987) The Phareae and Streptogyneae of Sri Lanka: a morphological-anatomical study. *Smithsonian,* 65, 1-26.

SOLTIS, P. S., SOLTIS, D. E. & CHASE, M. W. (1999) Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature,* 402, 402-403.

SORENG, R. J. & DAVIS, J. I. (1998) Phylogenetics and character evolution in the grass family (Poaceae): Simultaneous analysis of morphological and chloroplast DNA restriction site character sets. *Botanical Review,* 64, 1-85.

SPEARMAN, C. (1904) Proof and measurement of association between two things. *American Journal of Psychology,* 15, 72-101.

STACE, C. A. (1966) The use of epidermal characters in phylogenetic considerations. *New Phytologist,* 65, 304-318.

STEBBINS, G. L. & CRAMPTON, B. (1961) A suggested revision of the grass genera of temperate North America. *In* Recent Advances in Botany *(IX International Botaniucal Congress, Montreal, Canada),* Vol.1 Univ.Toronto Press, Toronto, Canada, 133-145.

STEBBINS, G. L. (1981) Coevolution of grasses and herbivores. *Annals of the Missouri Botanical Garden*, 68, 75-86.

STROMBERG, C. A. E. (2005) Decoupled taxonomic radiation and ecological expansion of open-habitat grasses in the Cenozoic of North America. *Proceedings of the National Academy of Science USA,* 102, 11980-11984.

STUCKY, R. (1990) Evolution of land mammal diversity in North America during the Cenozoic. *In* Genoways, H. H. (Ed.) *Current Mammology.* New York, USA, 375-432.

SWEET, A. R. (2001) Plants, a yardstick for measuring the environmental consequences of the Cretaceous-Tertiary boundary event. *Geoscience Canada,* 28, 127-138.

SWOFFORD, D. L. (2002) Phylogenetic Analysis Using Parsimony (*and Other Methods). Version4. *In* Sinauer Assoaciates (Ed.), Sunderland, Massachussetts, USA.

TABERLET, P., GIELLY, L., PAUTOU, G. & BOUVET, J. (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology*, 17, 1105-1109.

TAMURA, M. N., YAMASHITA, J., FUSE, S. & HARAGUSHI, M. (2004) Molecular phylogeny of monocotyledons inferred from combined analysis of plastid *matK* and *rbcL* gene sequences. *Journal of Plant Research*, 117, 109-120.

THEUNISSEN, J. D. (1994) A method for isolating and preparing silica bodies in grasses for scanning electron microscopy. *Biotechnic and Histochemistry*, 69, 291-294.

THOMPSON, J. N. (1982) *Interaction and coevolution,* New York, USA.

THOMPSON, J. N. (2005) *The geographic mosaic of coevolution,* Chicago, USA.

THOMPSON, J.D., HIGGINS, D.G. and GIBBSON, T.J. (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22, 14673-14680.

THORNE, J. L. (1998) Estimating the rate of evolution of the rate of molecular evolution. *Molecular Biology and Evolution,* 15, 1647-1657.

THORNE, J. L. & KISHINO, H. (2002) Divergence time and evolutionary rate estimation with multilocus data. *Systematic Biology,* 51, 689-702.

THUROW, T. L. (1991) Hydrology and erosion. *In* Stuth, H. (Ed.) *Grazing management : an Ecological Perspective.* Timber Press, Portland, OR, USA, 141-159.

TIFFNEY, B. H. & MANCHESTER, S. R. (2001) The use of geological and paleontological evidence in evaluating plant phylogeographic hypotheses in

the Northern Hemisphere Tertiary. *International Journal of Plant Science,* 162, S3-S17.

TOLIVIA, D. & TOLIVIA, J. (1987) Fasga: a new polychromatic method for simultaneous and differential staining of plant tissues. *Journal of Microscopy,* 148, 113-117.

TZVELEV, N. N. (1989) The system of grasses (Poaceae) and their evolution. *Botanical Review,* 55, 141-203.

VAN TIEGHEM, P. (1897) Morphologie de l'embryon et de la plantule chez les Graminees et les Cyperacees. *Annales des Sciences Naturelles Botanique,* 3, 259-309.

VERBOOM, G.A., NTSOHI, R. & BARKER, N.P. (2006) Molecular phylogeny of African Rytidosperma-affiliated danthonioid grasses reveals generic polyphyly and convergent evolution in spikelet morphology. *Taxon*, 55, 337-348.

VERMEIJ, G. J. (1999) Inequality and the directionality of history. *American Naturalist,* 153, 243-253.

VOS, R. A. & MOOERS, A. O. (2004) Reconstructing divergence times for supertrees. *In* Bininda-Emonds, O. R. P. (Ed.) *Phylogenetic supertrees: combining information to reveal the Tree of Life.* Kluwer Academic, Dordrecht, Netherlands, 281-299.

WANG, R. Z. (2004) Responses of *Leymus chinensis* (Poaceae) to long-term grazing disturbance in the Songnen grasslands of north-eastern China. *Grass and Forage Science,* 59, 191-195.

WATSON, L. & DALLWITZ, M. J. (1989) *Grass genera of the world: Illustrations of characters, Descriptions, Classification, Interactive identification, Information retrieval.* Booklet (45 p.) Microfice and computer program, privately distributed, Canberra, Australia.

WATSON, L. & DALLWITZ, M. J. (1992) *The Grass Genera of the World.* CAB International, Wallingford, UK.

WEBB, S. D. (1989) The fourth dimension in North American terrestrial mammalian communities. *In* Morris, D. W., Abramsky, Z., Willig, M.R. (Ed.) *Patterns in the structure of mammalian communities.* Texas Tech University, TEXAS, USA.

WIENS, J. J. (1998) The accuracy of methods for coding and sampling higher-level taxa for phylogenetic analysis: a simulation study. *Systematic Biology*, 47, 381-397.

WIENS, J. J. (2003) Missing data, incomplete taxa, and phylogenetic accuracy. *Systematic Biology*, 52, 528-538.

WIENS, J. J. (2005) Can incomplete taxa rescue phylogenetic analyses from long-branch attraction? *Systematic Biology*, 54, 731-742.

WING, S. L. (1987) Eocene and Oligocene floras and vegetation of the Rocky Mountains. *Annals of the Missouri Botanical Garden,* 74, 748-784.

WING, S. L. (1998) Terrestrial carnivores, ungulates and ungulatelike mammals. *In* Janis C.M. (Ed.) *Evolution of Tertiary Mammals in North America.* Cambridge University Press, Cambridge, UK, 37-60.

WOLFE, J. A. (1992) Climatic, floristic, and vegetational changes near the Eocene/Oligocene boundary in North America. *In* Berggren, D. R. (Ed.) *Eocene-Oligocene climatic and biotic evolution.* Princeton Univ. Press, Princeton, NJ, USA, 421-436.

YANG, Z. (2000) *Phylogenetic analysis by maximum likelihood (PAML)3.0.,* London, UK.

YANG, Z. H. (1994) Maximum-likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular Evolution,* 39, 306-314.

YEATES, D. K. (1995) Groundplans and exemplars: path to the Tree of Life. *Cladistics,* 11, 343-357.

YODER, A. D. & YANG, Z. H. (2000) Estimation of primate speciation dates using local molecular clocks. *Molecular Biology and Evolution,* 17, 1081-1090.

YULE, G. U. (1924) A mathematical theory of evolution, based on the conclusions of Dr. J.C. Willis. *Philosophical Transactions of the Royal Society, B,* 213, 21-87.

ZUCKERLANDL, E. & PAULING, L. (1965) Evolutionary Divergence and Convergence in Proteins. *In* Vernon Bryson and HenryVogel, J. (Eds), *Evolving Genes and Proteins: a symposium held at the Institute of Microbiology of Rutgers, with support from the National Science Foundation,* New York Academic Press, NY, USA, 97-166.

Appendices

Appendix 2.1 *List of accessions sampled with GenBank numbers for rbcL, matK and trnL-F sequences. Voucher numbers are provided for sequences newly determined in this study. Subfamilial and tribal classification follows Watson and Dallwitz (1992) except for Chloridoideae (Clayton and Renvoize 1986). Accession numbers in bold represent newly submitted sequences.*

Subfamily	Tribe	Genus species	Voucher (present study)	Database Accession Numbers		
				rbcL	*matK*	*trnL-F*
Anomochlooideae	Anomochloeae	*Anomochloa marantoidea*		AF021875	AF164381	N/A
	Streptochaetae	*Streptochaeta spicata*		SSP419949	AF164383	N/A
		Streptochaeta sodiroana	*Hodkinson574TCD*	N/A	N/A	**EF137515**
Aristidoideae	Aristideae	*Aristida congesta*		ACU31359	N/A	N/A
		Aristida adscensionis		N/A	AF164412	N/A
		Stipagrostis zeyheri		SZU31378	N/A	N/A
Arundinoideae	Amphipogoneae	*Amphipogon strictus*		ASU88403	N/A	N/A
		Amphipogon caricinus		N/A	AF312794	N/A
	Arundineae	*Arundo donax*	*Hodkinson131TCD*	ADU13226	AF164408	**EF137516**
		Phragmites australis	*Hodkinson.n.TCD*	U29900	AF144575	**EF137518**
		Spartochloa scirpoidea	*s.n.*	N/A	N/A	**EF137519**
	Cyperochloeae	*Cyperochloa hirsuta*		CHU88404	N/A	N/A
	Bambuseae	*Arthrostylidium sp.*	*Stapleton1133TCD*	AJ746254	N/A	**EF137521**
		Arundinaria tecta	*Chase1992K*	AJ746179	**EF125165**	N/A
		Arundinaria gigantea	*Chase1995K*	N/A	N/A	**EF137522**
		Aulonemia longiaristata	*Clark & Aslmb.1389QCA*	N/A	N/A	**EF137523**
		Bambusa multiplex	*Chase2000K*	BAMCPRBCL	**EF125166**	N/A
		Bambusa vulgaris	*Chase1963K*	N/A	N/A	**EF137524**

Borinda emeryi	*Kew1992-0401*	EF125079	EF125167	EF137525
Chimonobambusa marmorea	*Chase1982K*	AJ746176	EF125168	EF137526
Chusquea circinata	*Stapleton1126TCD*	CCU13227	N/A	N/A
Chusquea coronalis		N/A	AF164389	EF137527
Dendrocalamus barbatus	*Xia345*	AJ746173	EF125169	EF137528
Drepanostachyum falcatum	*Chase1987K*	AJ746265	EF125170	EF137529
Fargesia dracocephala	*Hodkinson63BTCD*	AJ746266	N/A	EF137530
Fargesia mutielae	*Chase1964K*	N/A	EF125171	N/A
Gaoligongshania megalothyrsa	*Hodkinson50BTCD*	N/A	N/A	EF137531
Gigantochloa verticillata	*Chase1966K*	N/A	N/A	EF137532
Greslania circinata	*Pillon et al. 47NOVP.K*	EF125080	N/A	N/A
Hibanobambusa tranquillans	*Hodkinson149BTCD*	AJ746267	N/A	EF137533
Himalayacalamus cupreus	*Stapletons.n.TCD*	EF125081	N/A	EF137534
Himalayacalamus asper	*Chase2004K*	N/A	EF125172	N/A
Indocalamus latifolius	*Chase1994K*	AJ746177	EF125173	EF137535
Melocanna baccifera	*Xia3831BSC*	EF125082	EF125174	EF137536
Neurolepis elata	*Clark etal. 1409QCA*	EF125083	N/A	EF137537
Olmeca recta	*Hodkinson139BTCD*	AJ746270	N/A	EF137538
Otatea acuminata	*Chase1986K*	AJ746269	EF137435	EF137539
Oreobambos sp.	*Hodkinson195BTCD*	N/A	N/A	EF137541

Bambu		*Otatea acuminata*	*Chase1989K*	AJ746271	EF137436 EF137542
		Phyllostachys bambusoides		PBU13230	N/A N/A
		Phyllostachys aurea		N/A	AF164390 N/A
		Phyllostachys nigra	*Hodkinson134BTCD*	N/A	N/A EF137544
		Pseudosasa japonica	*Chase1985K*	N/A	EF137438 N/A
		Pseudosasa amabilis	*Hodkinson62BTCD*	AJ726273	N/A EF137545
		Pseudostachyum polymorphum	*Chase1974K*	N/A	N/A EF137546
		Racemobambos hepburnii	*WSmith & Everard147K*	N/A	N/A EF137547
		Rhipidocladum harmonicum	*Clark et al.1103QCA*	N/A	N/A EF137549
		Sasa palmata	*Chase1991K*	AJ746278	N/A EF137550
		Sasa kurilensis		N/A	AF164391 N/A
		Schizostachyum funghomii	*Xia351 IBSC*	EF125084	EF137440 EF137551
		Shibataea chinensis	*Hodkinson2BTCD*	EF125085	N/A EF137552
		Shibataea kumasaca	*Chase1976 K*	N/A	EF137441 N/A
		Sinobambusa tootsik	*Chase1996K*	EF125086	EF137442 EF137553
		Thamnocalamus spathiflorus	*Chase1978K/Stapleton1388K*	EF125087	EF137443 EF137554
		Thyrsostachys siamensis	*Xia384 K*	EF125088	EF137444 EF137555
		Yushania maculata	*Chase1989K*	N/A	EF137445 N/A
		Yushania maling	*Hodkinson24BTCD*	AJ746277	N/A EF137556
Olyreae		*Buergersiochloa bambusoides*		AY632362	N/A N/A
		Cryptochloa strictiflora	*Hodkinson554TCD*	EF125089	EF137434 N/A
		Eremitis sp.		AY622892	N/A N/A

		Lithachne humilis		LHU13231	N/A	N/A
		Lithachne pauciflora		N/A	AF164385	N/A
		Oliva latifolia	*A.M. de Carvalho4394 CEPEC*	**EF125090**	AF164386	**EF137540**
		Pariana parvisaica	*Hodkinson528TCD*	**EF125091**	N/A	**EF137543**
		Pariana radiciflora		N/A	AF164387	N/A
		Raddia brasiliensis	*Chase2003K*	AJ746275	**EF137439**	**EF137548**
Centotheceae	Centotheceae	*Centotheca lappacea*	*Hodkinson235TCD*	**EF125092**	**EF137446**	**EF137557**
		Chasmanthium latifolium	*Renvoize s.n.*	CLU31101	N/A	**EF137558**
		Chasmanthium laxum		N/A	AF164414	N/A
		Danthoniopsis petiolata		AY618661	N/A	N/A
		Leptatherum gracilis	*Chant. & Parnell 99/194 TCD*	**EF125093**	AF164415	N/A
		Zeugites pittieri		AY632374	AF144576	N/A
		Orthoclada laxa	*Hodkinson522TCD*	**EF125094**	AF164416	N/A
	Thysanolaeneae	*Thysanolaena maxima*	*G.Sanchez-Ken220310K / HodkinsonK10TCD*	TMU31380	**EF137433**	**EF137520**
		Aegopogon cenchroides		N/A	AF312324	N/A
	Chlorideae	*Aeluropus littoralis*	*Ferguson634TCD*	**EF125095**	AF144597	N/A
		Brachyachne ciliaris		N/A	AF312327	N/A
		Buchloe dactyloides		N/A	AF312325	N/A
		Chloris virgata	*Chase19521K/Chase9262K*	**EF125096**	N/A	**EF137560**
		Chloris truncata		N/A	AF312330	N/A
		Coelachyrum yemenicum		N/A	AF144581	N/A

	Enteropogon dolichostachyus			N/A	AF312332	N/A
	Gymnopogon brevifolius			N/A	AF312333	N/A
	Kengia serotina	*Chase19527K*		**EF125097**	N/A	N/A
	Kengia sonderica			N/A	AF164428	N/A
	Oxychloris scariosa			N/A	AF312334	N/A
	Perotis rara	*Hodkinson274TCD*		**EF125098**	AF144590	**EF137567**
	Schedonnardus paniculatus			N/A	AF312335	N/A
	Tetrapogon tenellus			N/A	AF312336	N/A
Cynodonteae	*Astrebla lappacea*			N/A	AF144589	AY576672
	Bouteloua gracilis	*Kew 1974-3996*		AJ784823	N/A	N/A
	Bouteloua curtipendula			N/A	AF144578	N/A
	Cynodon transvaalensis	*Chase9265K*		**EF125099**	AF312331	**EF137562**
	Eustachys petraea	*Hodkinson546TCD*		**EF125100**	N/A	N/A
	Eustachys mutica			N/A	AF144587	N/A
	Harpochloa sp.	*Smook6674TCD*		**EF125101**	**EF137448**	N/A
	Microchloa caffra	*Smook6636TCD*		**EF125102**	AF164425	N/A
	Pleuraphis jamesii			N/A	AF144579	N/A
	Spartina pectinata	*Chase9274TCD*		AJ784821	AF312353	**EF137568**
	Tragus racemosus	*Chase19533K/Chase9277K*		**EF125103**	N/A	**EF137570**
	Tragus berteronianus			N/A	AF144591	N/A
	Trichloris crinita			N/A	AF144588	N/A

Chloridoideae	Eragrostideae	*Zoysia sp*	*Hodkinson833TCD*	AY632375	N/A	N/A
		Zoysia macrostachya		N/A	AF164426	N/A
		Zoysia japonica		N/A	N/A	**EF137571**
		Acrachne racemosa		N/A	N/A	AY574637
		Calamovilfa longifolia	*Hodkinson617TCD/Chase9261K*	**EF125104**	N/A	**EF137559**
		Calamovilfa gigantea		N/A	AF312354	N/A
		Cladoraphis spinosa		AM235054	N/A	N/A
		Crypsis schaenoides	*Chase19523K/ Chase9264K*	**EF125105**	**EF137447**	**EF137561**
		Dactyloctenium aegyptium	*Hodkinson619TCD*	**EF125106**	N/A	N/A
		Dactyloctenium radulans		N/A	AF312338	N/A
		Dinebra retroflexa	*Crank8176TCD*	**EF125107**	AF144584	AY576674
		Diplachne malabarica		N/A	AF312345	N/A
		Eleusine indica	*Hodkinson126TCD*	**EF125108**	AF144580	**EF137563**
		Eragrostiella brachyphylla		N/A	AF312339	N/A
		Eragrostis caespis	*Chase9270K*	ECU31104	N/A	N/A
		Eragrostis bicolor		N/A	AF312341	N/A
		Eragrostis mexicana ssp virescens		N/A	N/A	**EF137565**
		Fingerhuthia africana		AM235059	N/A	N/A
		Fingerhuthia sesleriiformis		N/A	AF144606	N/A
		Leptochloa dubia		N/A	AF312344	N/A
		Monanthochloe littoralis		N/A	AF312349	N/A
		Monodia stipoides		N/A	AF144602	N/A
		Muhlenbergia racemosa		AJ784836	N/A	N/A

	Muhlenbergia wrightii		N/A	AF312356	N/A
	Nevraudia reynaudiana	*Hodkinson203TCD*	**EF125109**	**EF137449**	N/A
	Plectrachne pungens		N/A	AF144603	N/A
	Sporobolus sp.		AM235073	N/A	N/A
	Sporobolus indicus	*Chase9275K*	N/A	AF144601	**EF137569**
	Tetrachne dregei		N/A	AF312363	N/A
	Trichoneura grandiglumis		N/A	AF144595	N/A
	Tridens brasiliensis		N/A	AF144596	N/A
	Triodia scariosa		N/A	AF312358	N/A
	Triraphis schlechteri		N/A	AF312347	N/A
	Uniola pittieri	*Hodkinson584TCD*	**EF125110**	N/A	N/A
	Uniola paniculata		N/A	AF144607	N/A
	Vaseyochloa multinervosa		N/A	AF312348	AY576675
Leptureae	*Lepturus repens*	*Hodkinson272TCD*	**EF125111**	AF144598	**EF137566**
Orcuttieae	*Neostapfia colusana*		N/A	AF312351	N/A
	Orcuttia californica		N/A	AF144599	N/A
	Tuctoria greenei		N/A	AF312352	N/A
Pappophoreae	*Cottea pappophoroides*		N/A	AF312359	N/A
	Enneapogon polyphyllus	*Chase9267K*	N/A	N/A	**EF137564**
	Enneapogon scaber		ESU31103	N/A	N/A
	Enneapogon glaber		N/A	AF312360	N/A

		Pappophorum bicolor		N/A	AF144604	N/A
		Schmidtia pappophoroides		N/A	AF312362	N/A
	Incertae Sedis	*Centropodia glauca*		CGU31100	AF164410	N/A
		Heteranthe aborting		N/A	AF312343	N/A
		Lintonia nutans		N/A	AF312337	N/A
		Melanocenchris abyssinica		N/A	AF312326	N/A
		Merxmuellera sp.		AM235065	N/A	N/A
		Chaetobromus dregeanus		AM235053	N/A	N/A
	Danthonieae	*Cortaderia richardii*	*Hodkinson1587CD*	AJ784830	EF137450	EF137572
		Danthonia spicata		DSU31102	AF164409	N/A
		Danthonia sp.	*Chase9254K*	N/A	N/A	EF137573
		Elytrophorus globularis		EGU88405	N/A	N/A
		Gynerium sagittatum	*G.Sanchez-Ken22003K*	GSU31105	EF137431	N/A
		Hakonechloa macra	*Chase9255K*	N/A	N/A	EF137574
		Karroochloa purpurea		KPU31437	N/A	N/A
Danthonioideae		*Merxmuellera macowanii*	*G.Sanchez-Ken22021K/Chase9256K*	MMU31438	EF137451	EF137575
		Molinia littoralis	*Chase9257K*	N/A	N/A	EF137517
		Molinia caerulea		AJ746295	AF164411	N/A
		Moliniopsis japonica(syn. Molinia japonica)		MJU31439	N/A	N/A
		Monachather paradoxus		MPU31379	N/A	N/A

		Plinthanthesis paradoxa	*Chase19306K*	PPU31440	N/A	N/A
		Rytilasperma macowanii		**EF125112**	**EF137432**	N/A
		Schismus barbatus		AM235071	N/A	N/A
		Stypoctochloa synopiosca		SGU88406	N/A	N/A
Ehrhartoideae	Ehrharteae	*Microlaena avenacea*	*Forrestal574NBG/HodkinsonG25TCD*	AY691633	N/A	N/A
		Ehrharta calycina		AM235057	N/A	**EF137576**
		Ehrharta longifolia		N/A	AF164392	N/A
	Oryzeae	*Chikusichloa aquatica*		N/A	AF489912	N/A
		Hygroryza aristata		N/A	AF489913	N/A
		Leersia oryzoides		LOU13228	N/A	N/A
		Leersia perrieri		N/A	AF148677	N/A
		Luciola leiocarpa		N/A	AF489911	N/A
		Oryza sativa	*Hodkinson46TCD*	RICCPRBCL	AF148650	**EF137577**
		Potamophila parviflora		N/A	AF489914	N/A
		Prosphytochloa prehensilis		N/A	AF489916	N/A
		Rhynchoryza subulata		N/A	AF148675	N/A
		Zizaniopsis villanuevis		N/A	AF489917	N/A
		Zizania texana		ZIZCPRBCL	N/A	N/A
		Zizania latifolia		N/A	AY092064	N/A

Incertae Sedis	Eriachneae	*Eriachne_triodioides*		AF352580	N/A	N/A
	Micraireae	*Micraira_subulifolia*		AY632366	N/A	N/A
	Streptogyneae	*Streptogyna_americana*	s.n.	EF125113	N/A	N/A
	Andropogoneae	*Andropogon_gerardii*	*Kew 1969-19004*	AJ784818	AF144577	AY116263
		Arthraxon sp.	*Hodkinson133TCD*	EF125114	EF137454	EF137578
		Bothriochloa_ischaemum	*Kew 1975-1575*	EF125115	EF137456	N/A
		Coix_lacryma-jobi	*Hodkinson124TCD*	EF125116	EF137458	EF137580
		Cymbopogon_citratus	*Hodkinson42TCD*	EF125117	EF137459	EF137581
		Eulalia_irritans		N/A	N/A	AY116242
		Heteropogon_contortus		AM235063	N/A	N/A
		Hyparrhenia_hirta		HHU31436	AF164417	N/A
		Imperata_cylindrica	*Chase19300K/Kew 1996-2520*	AJ784826	EF137462	AY116262
		Miscanthidium_junceus	s.n.	N/A	N/A	EF137582
		Miscanthus_sinensis	*Hodkinson5TCD*	EF125118	N/A	AJ426571
		Miscanthus_sinensis	s.n.	N/A	EF137466	N/A
		Pogonatherum sp	*Hodkinson21TCD*	EF125119	EF137468	N/A
		Saccharum_officinarum	*HodkinsonK104TCD*	EF125120	EF137470	AY116253
		Schizachyrium_scoparium	*Kew 1985-1663*	N/A	N/A	EF137584
		Sorghastrum_nutans	*Chase 19531 K*	EF125121	EF137473	N/A
		Sorghum_halepense	*Kew 1966-54209*	EF125122	N/A	AY116244

		Sorghum_bicolor		N/A	AF164418	N/A
		Spodiopogon_sibiricus	*Kew 1997-5606*	AJ784833	**EF137474**	**EF137586**
		Themeda_triandra		N/A	N/A	AY116261
		Trachypogon_spicatus	*Hodkinson579TCD*	**EF125123**	**EF137476**	N/A
		Tripsacum_dactyloides	*Kew 1974-4027*	**EF125124**	N/A	N/A
		Zea_mays	*Kew 1956-5307*	ZMA86563	ZMA86563	N/A
		Zea_diploperrenis		N/A	N/A	AY116260
	Arundinelleae	*Loudetiopsis_chrysothrix*		N/A	AF164419	N/A
		Tristachya_biseriata		TBU31381	AF164420	N/A
	Isachneae	*Isachne_arundinacea*		AY618662	N/A	N/A
Panicoideae	Paniceae	*Acroceras sp.*	*Hodkinson264TCD*	**EF125125**	**EF137452**	N/A
		Alloteropsis sp.	*Hodkinson230TCD*	**EF125126**	**EF137453**	N/A
		Axonopus_compressus	*Hodkinson509TCD*	**EF125127**	**EF137455**	N/A
		Brachiaria sp.	*Forestetal 778NBG*	AM235052	N/A	N/A
		Cenchrus_incertus	*Chase9279 K*	N/A	N/A	**EF137579**
		Cenchrus_setiger us	*Hodkinson447TCD*	CEHRBCL	EF137457	N/A
		Digitaria_sanguinalis	*Hodkinson110TCD*	AJ746264	AF164421	AY116268
		Echinochloa_esculenta	*Chase19524K*	**EF125128**	N/A	N/A
		Echinochloa_crus-galli		N/A	N/A	AY116269
		Echinochloa_utilis		N/A	AF164422	N/A
		Homopholis_proluta		N/A	N/A	AY342713

		Hymenachne amplexicaulis	*Hodkinson633TCD*	EF125129	EF137460	N/A
		Ichnanthus nemorosus	*Hodkinson505TCD*	EF125130	EF137461	N/A
		Isophorus unisetus	*Hodkinson520TCD*	EF125131	EF137463	N/A
		Lasiacis nigra	*Hodkinson561TCD*	EF125132	EF137464	N/A
		Melinis repens	*Hodkinson216TCD*	EF125133	EF137465	N/A
		Neurachne tenuifolia	*s.n.*	EF125134	N/A	N/A
		Oplismenus sp.		AM235067	N/A	N/A
		Panicum virgatum	*Kew 1969-19104*	EF125135	N/A	AY116267
		Panicum capillare		N/A	AF164423	N/A
		Paspalum dilatatum	*Kew 1995-1370*	N/A	N/A	EF137583
		Pennisetum glaucum		PENCARBOXL	N/A	N/A
		Pennisetum macrostrum	*Hodkinson 137TCD*	N/A	EF137467	AY116266
		Rhynchelytrum repens	*Cronk56TCD*	EF125136	EF137469	N/A
		Sacciolepis indica	*Hodkinson533TCD*	EF124137	EF137471	N/A
		Setaria italica	*Hodkinson227TCD*	EF125138	N/A	N/A
		Setaria pumila	*Hodkinson1217TCD*	N/A	EF137472	N/A
		Setaria viridis	*Kew 1995-2149*	N/A		EF137585
		Stenotaphrum secundatum	*Hodkinson570TCD*	EF125139	EF137475	N/A
		Urochloa sp.	*Hodkinson213TCD*	EF125140	EF137477	N/A
≥	Phareae	*Pharus latifolius*	*Hodkinson514TCD*	AY357724	AF164388	EF137587

Pharmaceutical						
	Ampelodesmeae	*Ampelodesmos mauritanica*	*Chase19299K/Chase3523K*	AJ746172	EF137483	N/A
	Aveneae	*Arrtagrostis latifolia*		N/A	N/A	AY237904
		Agrostis capillaris	*Hodkinson35TCD*	N/A	EF137479	N/A
		Agrostis stolonifera		AJ746280	N/A	N/A
		Alopecurus pratensis	*Hodkinson30TCD*	EF125141	EF137481	EF137589
		Ammophila brevitigulata	*Chase19296K*	AJ784835	EF137482	N/A
		Amphibromus fluitans		AY691631	N/A	N/A
		Anthoxanthum odoratum	*Chase19298K/Hodkinson27TCD*	AJ746282	EF137484	EF137590
		Arrhenatherum elatius	*Hodkinson27TCD*	AJ784823	EF137486	EF137591
		Avena sativa		ASTCPRBCL	AF164395	N/A
		Avena fatua	*Hodkinson31TCD*	N/A	N/A	EF137592
		Beckmannia syzigachne	*Hodkinson51TCD*	EF125142	EF137487	N/A
		Calamagrostis epigejos	*Chase19295K*	AJ784820	EF137489	N/A
		X Calammophila baltica	*Chase19304K*	N/A	EF137490	N/A
		Corynephorus canescens	*HodkinsonK47 TCD*	N/A	N/A	EF137598
		Gaudinia fragilis	*Scannel 1992 TCD*	EF125143	EF137499	N/A
		Helictotrichon pubescens	*T.C.G.Rich145HNMW*	N/A	EF137502	N/A
		Helictotrichon reuseri	*Kew 1995-1350*	EF125144	N/A	N/A
		Helictotrichon sp.	*HodkinsonB167CD*	N/A	N/A	EF137604

	Hierochloe odorata	*Chase19307K/Kew 387-51.38701*	AJ784828	EF137503	EF137605
	Holcus lanatus	*Hodkinson25TCD*	AJ746279	EF137504	EF137606
	Koeleria pyramidata	*Hodkinson10TCD*	AJ784825	EF137505	EF137607
	Lagurus ovatus	*Hodkinson6TCD*	N/A	N/A	EF137608
	Milium effusum	*Kew 1985-1440*	N/A	N/A	EF137612
	Phalaris arundinacea	*Hodkinson20TCD*	AJ784827	AF164396	EF137615
	Phleum pratense		AJ784832	AF164397	N/A
	Polypogon litoralis	*Chase19529K*	EF125145	EF137510	N/A
	Trisetum flavescens	*D.Gowing960K*	AJ746278	N/A	N/A
	Trisetum sp.	*Hodkinson18TCD*	N/A	EF137513	EF137619
Brachyelytreae	*Brachyelytrum erectum*		AY622888	AF164384	N/A
Brachypodieae	*Brachypodium sylvaticum*	*Hodkinson22TCD*	AJ746258	AF164400	EF137593
Bromeae	*Bromus ramosus*	*Hodkinson41TCD*	N/A	N/A	EF137595
	Bromus inermis		BICHRBPCX	AF164398	N/A
	Boissiera saxatilis	*Chase19519K*	N/A	EF137488	N/A
Diarrheneae	*Diarrhena obovata*		AY622890	N/A	N/A
Lygeae	*Lygeum spartum*	*HodkinsonSp18TCD*	EF125146	EF137507	N/A
Meliceae	*Glyceria fluitans*	*Chase19776K*	AJ746290	EF137500	N/A
	Glyceria maxima	*Hodkinson19 TCD*	N/A	N/A	EF137603
	Melica uniflora	*Hodkinson44TCD*	AJ746263	N/A	EF137611

	Melica altissima		N/A	AF164399	N/A
Nardeae	*Nardus stricta*	*M.C.Sheahan007K*	AJ746296	AF164394	N/A
Phaenospermateae	*Phaenosperma globosa*	*Chase19294K*	AY632370	EF137437	N/A
Poeae	*Aira praecox*	*M.C.Sheahan004K*	AJ746255	EF137480	EF137588
	Apera interrupta	*Chase19518K*	EF125147	EF137485	N/A
	Arctophila fulva	*Jokela1962TCD*	EF125148	N/A	AY237900
	Briza media *Briza erecta*	*M.E.Fay143K*	AJ746285 N/A	N/A AF164401	EF137594 N/A
	Castellia tuberculosa	*Chase19520K/Kew1995-4283*	EF125149	EF137492	EF137596
	Catapodium rigidum	*Hodkinson23TCD*	EF125150	EF137491	EF137597
	Crepsis delicatula		N/A	N/A	AF478537
	Cutandia maritima		N/A	N/A	AF487618
	Cynosurus cristatus	*Hodkinson14TCD*	EF125151	N/A	EF137599
	Dactylis glomerata	*Hodkinson26TCD*	AY395535	EF137494	EF137600
	Deschampsia caespitosa	*D.Gowing663K*	EF125152	EF137495	EF137601
	Dupontia fisheri		N/A	N/A	AY237897
	Festuca rubra	*Hodkinson15TCD*	AJ746261	EF137498	EF137602
	Moerma cylindrica	*Chase19526K*	EF125153	EF137501	N/A
	Helleria fragilis		N/A	N/A	AF533059
	Lamarckia aurea		AJ784834	N/A	AF533029

Poaid		*Lolium perenne*	*D.Gowins089K*	AJ746293	EF137586 EF137619
		Micropyropsis tuberosa		N/A	N/A AF533037
		Micropyrum patens		N/A	N/A AF495885
		Nardurodes salzmanii		N/A	N/A AF478535
		Parafestuca albida		N/A	N/A AF533022
		Paratholis incurva	*Chase19528K*	EF125154	EF137508 AF533036
		Poa trivialis	*D.Gowins066K*	AJ746301	N/A EF137616
		Poa pratensis		N/A	AF164402 N/A
		Psilurus incurvus	*Chase19530K*	EF125155	N/A AF478533
		Puccinellia distans		PUZRBCL	N/A AF533024
		Sclerochloa dura		N/A	N/A AF533023
		Schedonorus pratensis (syn. Festuca pratensis)		AY395536	N/A N/A
		Sesleria caerulea	*Hoskinson45TCD*	EF125156	EF137511 EF137617
		Simplicia laxa		AY691641	N/A N/A
		Sphenopus divaricatus		N/A	N/A AF533033
		Vulpia ciliata	*Kew 1986-852*	EF125157	N/A N/A
		Vulpia myuros		N/A	AF164403 AY118103
		Wangenheimia lima		N/A	N/A AF478536
	Stipeae	*Anemanthele lessoniana*	*Chase 19369K*	EF125158	N/A N/A

	Aeluropopon avenaceus		AY622886	N/A	N/A
	Nassella trichotoma	*Kew119-78.01227*	EF125159	N/A	EF137613
	Nassella tenuis		N/A	AF164406	N/A
	Oryzopsis sp	*Hodkinson65TCD*	N/A	N/A	EF137614
	Piptatherum miliaceum		AY622898	N/A	N/A
	Piptochaetium bicolor	*Chase19302K*	N/A	EF137509	N/A
	Stipa dregeana var. dregeana		SDU31442	N/A	N/A
	Stipa offneri		N/A	AF164407	N/A
	Stipa sp.	*HodkinsonK79TCD*	N/A	N/A	EF137618
			CHACHSRG	N/A	N/A
Triticeae	*Aegilops crassa*		N/A	N/A	AF519112
	Aegilops speltoides		EF125160	EF137478	N/A
	Agropyron sp.	*Hodkinson62TCD*	N/A	N/A	AF519117
	Agrotyron mongolicum		AY691636	N/A	N/A
	Australopyrum calcis subsp. Calcis		N/A	N/A	AF519118
	Australopyrum retrofra	*Chase19522K*	EF125161	EF137493	N/A
	Crithopsis delileana		EGCHRBPCX	EF137496	N/A
	Elymus glaucescens	*Chase19301K*	N/A	N/A	AF519144
	Elymus virginicus		EECHRBPCX	N/A	N/A
	Eremium erianthum		EF125162	EF137497	N/A
	Eremopyrum bonaepartis	*Chase19525K*	N/A	N/A	AF519151
	Eremopyrum orientale		AY836163	N/A	AF519129
	Haynaldia villosa				

		Henrardia persica	*Laeuard 32 TCD*	AY836167	N/A	AF519152
		Hordelymus europaeus		**EF125163**	N/A	N/A
		Hordeum vulgare		AY137456	X64129	N/A
		Hordeum murinum		N/A	N/A	AF519126
		Leymus chinensis	*Hodkinson7 TCD*	LCCHRBPCX	N/A	N/A
		Leymus angustus		N/A	AF164404	N/A
		Leymus sp.		N/A	N/A	**EF137609**
		Peridictyon sanctum		PSCHRBPCX	N/A	AF519154
		Psathyrostachys fragilis		AY137461	N/A	N/A
		Psathyrostachys juncea		N/A	AB078140	AF519170
		Pseudoroegneria spicata		PSPCHRBPC	N/A	AF519160
		Secale cereale		AY836181	N/A	AF478501
		Stenostachys gracilis		AY691639	N/A	N/A
		Taeniatherum caput-medusae	*Chase19532K*	**EF125164**	**EF137512**	AF519164
		Thinopyrum distichum	*Forestetal.658NBG*	AM235075	N/A	N/A
		Thinopyrum scirpeum		N/A	N/A	AF519167
		X Triticosecale sp.	*Chase19308K*	N/A	**EF137514**	N/A
		Triticum aestivum		AY328025	AB042240	AF148757
2	Psoreae	*Psoralea silicata*		AF164780	N/A	N/A

OUTGROUPS	Psychoidea			
	Eleriasp.	L32675	N/A	N/A
	Elegia_squamosa	N/A	AF881526	N/A
	Elegia_cuspe	N/A	N/A	AF148735
	Joinvillea_alicata	L01471	N/A	N/A
	Joinvillea_ascendens	N/A	AF164380	N/A

Appendix 4.1 *Sampling list of the herbarium specimens used for silica density measurements*

Genus_species	Herbarium	Country	Collector	Coll. Number	Year
Anomochloa_marantoidea	K	Brazil	Santos & Silva	3236	1978
Streptochaeta_spicata	K	Costa Rica	Chacon	227	2001
Aristida_congesta	K	Ethiopia	Mooney	8015	1909
Stipagrostis_zeyheri	K	South Africa	Smith	4224	1999
Arundo_donax	K	Spain	Spenner	s.n.	??
Molinia_caerulea	K	Sweden	Lillieroth	s.n.	1934
Phragmites_australis	K	Australia	P.I.Forster	PIF13172	1993
Thysanolaena_maxima	K	Thailand	Geesink & Santisuk	4971	1973
Arundinaria_tecta	K	USA	McClure	s.n.	1960
Bambusa_multiplex	K	USA	McClure	21327	1945
Borinda_emeryi	K	Nepal	Merlyn Edwards	205	1994
Chimonobambusa_marmorea	K	Japan	Townsend & Bridger	104	1999
Chusquea_circinata	K	Mexico	Wendt, Ishiki & Maya	5048	1985
Dendrocalamus_barbatus	TCD	Thailand	SS & AF	124	2004
Drepanostachyum_falcatum	K	British Isles	Souster	s.n.	1954
Himalayacalamus_cupreus	K	Nepal	Merlyn Edwards	208a	1994
Indocalamus_sinicus	TCD	Hong Kong	H.F. Hance	1945	1863
Melocanna_baccifera	K	Hong Kong	Nan Zhu	2822	1980
Olmeca_recta	K	Mexico	Nee & Taylor	26693	1983
Olyra_latifolia	TCD	Thailand	Hodkinson	614	1997
Otatea_acuminata	K	Mexico	Davidse et al.	30076	1984
Pariana_parvispica	TCD	Thailand	Hodkinson	528	1997
Pseudosasa_japonica	K	British Isles	Pohl	14523	1983
Raddia_brasiliensis	K	Brazil	Marquete et al.	688	1992
Sinobambusa_tootsik	K	USA	McClure	21065	1942
Thamnocalamus_spathiflorus	K	Nepal	Merlyn Edwards	203	1994
Thyrsostachys_siamensis	K	India	D.Prain	s.n.	1903
Yushania_maling	K	India	Soderstrom	2643	1982
Calamovilfa_longifolia	TCD		Hodkinson	61	1997

Chloris_virgata	K	Kenya	NMK	13	2001
Crypsis_schoenoides	K	Iran	Dobson	284	1968
Cynodon_transvaalensis	TCD		Hodkinson	116	1997
Dinebra_retroflexa	TCD	Senegal	s.n	8181	s.n
Eleusine_indica	TCD		Hodkinson	126	1997
Enneapogon_scaber	K	South Africa	Smook	8008	1909
Eragrostis_capensis	K	South Africa	Ellis	242	1970
Spartina_pectinata	K	USA	Pase	738	1956
Sporobolus_capensis	K	England	Ryves	s.n.	1974
Tragus_racemosus	K	Italy	Prior	s.n.	1903
Zoysia_japonica	K	China	Steward	2506	1924
Centotheca_lapasia	TCD		Hodkinson	235	1997
Danthonia_spicata	K	USA	Ovrebo & Sladewski	WO222	1989
Merxmuellera_macrowanii	BOL	South Africa	N.Barker	1012	?
Oryza_sativa	K	Russia	von Minkwitz	1202	1930
Andropogon_gerardii	K	USA	Gould	11488	1965
Arthraxon_quartinianus	TCD	Taiwan	M.T.Kao	9972	1983
Cenchrus_setigerus	K	Pakistan	Lamond	739	1965
Coix_lacryma-jobi	TCD	Mauritius	Cronk et al.	95	1985
Cymbopogon_citratus	TCD		Hodkinson	42	1997
Digitaria_sanguinalis	K	USA	Stevens	965	1913
Echinochloa_utilis	K	Belgium	Fasseaux	28	1909
Imperata_cylindrica	K	Greece	H.G. Tedd	410	1930
Miscanthus_sinensis	K	Japan	Furuse	7120	1974
Panicum_virgatum	TCD		Hodkinson	120	1997
Pennisetum_setaceum	K	USA	Brummitt	19810	1998
Saccharum_officinarum	K	Hawai	Oppenheimer	H10134	2001

Pharus_latifolius	K	Costa Rica	Pohl	12639	1971
Agropyron sp.	TCD		Hockinson	62	1997
Arrhenatherum_elatius	TCD	Denmark	J.Nielsen et al.	832	1979
Avena_sativa subsp sativa	K	France	Hepper	9185	1990
Brachypodium_sylvaticum subsp sylvaticum	K	Iran	Limmerton	3447	1958
Briza_media	K	Bhutan	Sinclair & Long	5187	1984
Bromus_inermis	K	Turkey	Davis	47325	1966
Castellia_tuberculosa	K	Saudi Arabia	Collenette	4023	1983
Catapodium_rigidum	TCD		Hockinson	23	1997
Cynosurus_cristatus	TCD		Hockinson	14	1997
Dactylis glomerata	TCD	France	Martin & Kerguelen	10973	1981
Deschampsia_caespitosa subsp caespitosa	K	Yugoslavia	Cook et al.	181	1959
Elymus_glaucus	K	USA	Hermann	7840	1936
Eremopyrum_orientale	TCD	Russia	Bellanina and Matsanko	209	1974
Festuca_littoralis	TCD	s.n	Drummond	583	s.n
Glyceria_fluitans	TCD	Russia	G.M.Proskuriakova	116	1969
Hierochloe_borealis	TCD	UK	Oakes	s.n.	s.n
Holcus_mollis	TCD	Denmark	J.Nielsen et al.	369	1967
Hordeum_leporinum	TCD	USA	A.B.West	22	1966
Koeleria_wallesiana	TCD	Spain	Chater & Moore	277	1966
Lolium_rigidum	TCD	Greece	J.R.Akeroyd	390	1983

Trisetum_paniceum	TCD	Spain	Rodriguez & Dorja	s.n.	1963
Triticum_aestivum	TCD	Chile	Cuming	244	1909
Vulpia_sicula	TCD	Italy	D.A.Webb	8254	1984