2007

Large Trees, Supertrees, and Diversification of the Grass Family

Trevor R. Hodkinson
Trinity College, Dublin, Ireland

Nicolas Salamin
University of Lausanne, Lausanne, Switzerland

Mark W. Chase
Royal Botanic Gardens, Kew, UK

Yanis Bouchenak-Khelladi
Trinity College, Dublin, Ireland

Stephen A. Renvoize
Royal Botanic Gardens, Kew, UK

See next page for additional authors

Large Trees, Supertrees, and Diversification of the Grass Family

Authors

Trevor R. Hodkinson, Nicolas Salamin, Mark W. Chase, Yanis Bouchenak-Khelladi, Stephen A. Renvoize, and Vincent Savolainen

LARGE TREES, SUPERTREES, AND DIVERSIFICATION OF THE GRASS FAMILY

Trevor R. Hodkinson,[1,5] Nicolas Salamin,[2] Mark W. Chase,[3] Yanis Bouchenak-Khelladi,[1,3] Stephen A. Renvoize,[4] and Vincent Savolainen[3]

[1]*Department of Botany, School of Natural Sciences, University of Dublin, Trinity College, Dublin 2, Ireland;*
[2]*Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland;* [3]*Molecular Systematics Section, Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3DS, UK;* [4]*Herbarium, Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AB, UK*
[5]*Corresponding author (trevor.hodkinson@tcd.ie)*

ABSTRACT

Phylogenetic studies of grasses (Poaceae) are advanced in comparison with most other angiosperm families. However, few studies have attempted to build large phylogenetic trees of the family and use these for evaluating patterns of diversification or other macroevolutionary hypotheses. Two contrasting approaches can be used to generate large trees: supermatrix analyses and supertrees. In this paper, we evaluated the suitability of each of these methods for the study of patterns and processes of evolution in the grasses. We collected data from DDBJ/EMBL/GenBank to determine sequence availability and asked how far we are from a complete generic-level phylogenetic tree of the grasses. We generated almost complete tribal-level supertrees (39 tribes) with over 400 genera using MRP methods, described their major clades, assessed their accuracy, and used them for the study of diversification. We generated a proportional supertree, by modifying the original supertree, to remove sampling bias associated with the original supertree that may affect diversification statistics. We used methods that incorporate information on the topological distribution of taxon diversity from all internal nodes of the phylogenetic tree to show that the grasses have experienced significant variations in diversification rates (M statistic P-values <0.001 for the original tree; <0.002 for the proportional tree) and show where on the trees significant shifts in diversification have occurred (seven shifts for the original tree; four shifts for the proportional tree). Such tests have not previously been attempted for the grasses, and we discuss future research directions in this area.

Key words: diversification, grasses, large trees, macroevolution, phylogenetics, Poaceae, species richness, supertrees.

INTRODUCTION

Comprehensive phylogenetic trees are valuable tools for the study of evolutionary patterns and processes. The most ambitious phylogenetic tree, the tree of life, would encompass all life on earth and include a total number of species in the range of 4–100 million (Baldauf 2003; Eisen and Fraser 2003; Mace et al. 2003; Pennisi 2003; Hodkinson and Parnell in press *a, b*). Even for the grass family (Poaceae), the task of producing a complete phylogenetic tree is still immense (Salamin et al. 2005; Hodkinson et al. in press) and would need to include ca. 10,000 species (Clayton and Renvoize 1986; Watson and Dallwitz 1992).

Advances in grass phylogenetics have occurred more rapidly than in most groups of higher plants because of their socioeconomic and ecological importance. They cover, chiefly as grasslands or bamboo forests, more than one-third of the terrestrial surface (Archibold 1995) and provide, among other things, staple cereals, sugar crops, forage, and reeds. Grass classification has been influenced heavily by recent phylogenetic studies. Widely used systems based largely on gross morphology and anatomy, such as Clayton and Renvoize (1986), Renvoize and Clayton (1992), and Watson and Dallwitz (1992), are being revised by those based on additional molecular evidence (for a historical account of grass classifications, see Clark et al. 1995 or the Grass Phylogeny Working Group [GPWG] 2001). The first molecular phylogenetic trees of the grass family were produced by Hamby and Zimmer (1988) using nuclear ribosomal DNA and Doebley et al. (1990) using the plastid *rbc*L gene. Both supported the monophyly of a group containing Panicoideae, Arundinoideae, Centothecoideae, and Chloridoideae (the PACC clade). Davis and Soreng (1993) then used restriction site variation to study plastid DNA on a sample of taxa that included all subfamilies of Clayton and Renvoize (1986), and the results also supported the PACC clade. A large number of single-region analyses from all genomes have since been produced, but these are dominated by studies of the plastid and, to a lesser extent, nuclear genomes. These include, from the plastid DNA (*mat*K: Liang and Hilu 1995, Hilu et al. 1999; *ndh*F: Clark et al. 1995; *rbc*L: Barker et al. 1995, Duvall and Morton 1996; *rpl*16: Zhang 2000; *rpo*C2: Barker et al. 1999; *rps*4: Nadot et al. 1994; *trn*L: Briggs et al. 2000, Gómez-Martínez and Culham 2000), nuclear DNA (*gbss*I [waxy]: Mason-Gamer et al. 1998; ITS: Hsiao et al. 1999; *PHY*B: Mathews and Sharrock 1996, Mathews et al. 2000), and limited information from mitochondrial DNA (*atp*A: Michelangeli et al. 2003).

Among these single-region analyses was a comprehensive study by Clark et al. (1995) using the plastid gene *ndh*F. The study included a representative sampling of grass diversity and many previously poorly sampled bambusoid and ehrhartoid taxa. They recovered a tree with two major groups, the PACC clade and BEP clade (Bambusoideae, Ehrhartoideae, and an expanded Pooideae). The study also showed

that *Anomochloa* Brongn. and *Streptochaeta* Schrad. ex Nees (Anomochlooideae) were sister to the rest of the grasses (the earliest-diverging lineage relative to the rest of the grasses). These are broad-leaved, forest understory genera from the Neotropics. The next diverging lineage was *Pharus* P. Browne (Pharoideae).

A smaller number of combined analyses or multigene studies have been reported such as the combined *ndh*F, *phy*B, and *rbc*L data set of Clark et al. (2000) and morphological, chromosomal, biochemical, and plastid DNA character set of Soreng and Davis (1998). However, most combined analyses have concentrated on smaller taxonomic units than the whole grass family (Hodkinson et al. 2002*a, b*). The most significant combined data analysis of the whole family included 62 grasses, sampling ca. 8% of the genera (GPWG 2001). Matrices of DNA sequences (plastid: *ndh*F, *rbc*L, *rpo*C2; nuclear: *gbss*I, ITS2, *PHY*B), plastid restriction site data, morphological and anatomical data were analyzed alone and in combination. A relatively well-resolved and supported topology was obtained including a well-supported PACCAD group (PACC, plus Aristidoideae and Danthonioideae).

Despite these advances in grass phylogenetics, no large phylogenetic trees of the family have been produced, except for the supertree of Salamin et al. (2002). For the purpose of this paper, we define large trees as those representing at least half of the genera within the family. Comprehensive phylogenetic trees can be constructed in different ways and categorized into either supermatrix or supertree methods. Supermatrix analyses rely on the sampling of many taxa for one or more regions. When sparse sampling of sequence data limits combination of sequence information into large multigene analyses, a supermatrix approach might not be optimal. Some form of "divide and conquer" strategy must be used instead, in which trees are built from individual data matrices and later assembled into a supertree on the basis of taxonomic overlap with other such trees (Sanderson and Driskell 2003). These are sometimes known as meta-analysis techniques (Bininda-Emonds et al. 2002; Salamin et al. 2002; Bininda-Emonds 2003).

Comprehensive and, therefore, large phylogenetic trees can be used for the examination of patterns and processes in evolution. They are important for identifying major clades of organisms, assessing their interrelationships, and helping to construct taxonomic classifications. They are also useful for pinpointing areas of sampling deficiency and highlighting where future DNA sequencing efforts should lie. Furthermore, they are valuable for assessing differential diversification rates of clades and exploring the processes that lead to these patterns (Barraclough and Nee 2001; Chan and Moore 2002; Moore et al. 2004).

This paper assesses the impact of phylogenetic trees on the study of patterns and processes of evolution in the grasses. We compile species and genus numbers for the subfamilies in two grass classifications (Clayton and Renvoize 1986; GPWG 2001) to gain a rough idea of diversification patterns in the family. We data-mine the DNA Data Bank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL), and GenBank (at the National Center for Biotechnology Information, USA) to determine DNA sequence availability for the production of large phylogenetic trees in the grasses and to see how these sequences are distributed

across the subfamilies. We ask how far we are from a complete generic-level phylogenetic tree of the grasses, and evaluate the approaches most suitable or practical for achieving one.

Having discovered that sequence information is limited for the generation of large phylogenetic trees in the grasses (without the incorporation of a high proportion of missing data), we take a meta-analysis approach to generate a sufficiently large phylogenetic tree, incorporating 426 genera and 39 tribes, for subsequent diversification rate studies. This represents an almost complete tribal-level tree of the grasses and includes approximately two-thirds of the grass genera. We describe the major clades in this tree, assess its accuracy, and use the tree for the study of diversification. We use methods that incorporate information on the topological distribution of taxon diversity from all internal nodes of the phylogenetic tree to determine if the grasses have experienced significant variations in diversification rates between sister clades and where on the tree major shifts of diversification have occurred. We also investigate the influence of sampling bias and phylogenetic uncertainty of tree topology on these diversification estimations. The methods help to distinguish chance variation in diversification patterns from patterns that require deterministic explanations. We also discuss what may have lead to differential diversification rates between lineages and highlight future research directions for the use of large phylogenetic trees in the study of diversification, species richness, and other aspects of macroevolution in the grasses.

MATERIALS AND METHODS

Data Collection

Species and genera numbers for taxa found in the subfamily classifications of Clayton and Renvoize (1986) and the GPWG (2001) were compiled and used to generate summary histograms. These statistics are found in Clayton and Renvoize (1986) but not in the GPWG (2001). The two systems differ largely in the subfamily classification, but the tribes vary less. It was therefore possible to compile species and genera numbers per subfamily of the GPWG (2001) by using the tribal data from Clayton and Renvoize (1986) and the tribes included within the subfamily revision of the GPWG. Then, all grass sequences from DDBJ/EMBL/GenBank were extracted (Jan 2004) and summary charts produced for the number of sequences within each grass subfamily of the GPWG (2001). We also summarized information for the "top ten or more gene regions." These are large multi-region analyses that have been sequenced most with respect to species and genus numbers.

Supertree Reconstruction

A supertree was constructed using the matrix representation with parsimony (MRP) method with the Baum/Ragan coding scheme and additional bootstrap percentage weighting following Salamin et al. (2002) using the software SuperTree0.85b (Salamin et al. 2002; *www.tcd.ie/Botany/dnabank/software*). The character states derived from the source trees were considered irreversible (see Salamin et al. 2002). Supertree reconstruction requires an overlap of taxon

sampling between source trees. However, few species are in common between published source trees so we considered only generic names for the production of the supertree (following Salamin et al. 2002). We were able to sample 39 of the 42 included tribes of the GPWG (2001) and 38 of the 41 included tribes in Clayton and Renvoize (1986). The absent tribes (Brylkinieae, Hubbardieae, Steyermarkochloeae) are all monogeneric and therefore only represent a small proportion of diversity within the family. Sampling also increased from 395 genera, in the previous largest supertree of the grasses (Salamin et al. 2002), to 426 genera by incorporating a large *rbc*L tree (Y. Bouchenak-Khelladi et al. in prep.) that contained 190 genera. The tree therefore combined 62 topologies; see Salamin et al. (2002) for a full list of other topologies used. Heuristic searches under maximum parsimony were performed using PAUP* vers. 4.08b (Swofford 2000) with 1000 replicates of random stepwise addition, using the nearest-neighbor interchange swapping algorithm and keeping only 100 trees at each replicate. The outgroups for the supertrees were created by the supertree method. Because all source trees have potentially a different real outgroup, an additional one is added to all source trees before producing the supertree.

Diversification Assessments

Shifts of diversification within the grass supertree were detected using SYMMETREE software (Chan and Moore 2004). Assuming the equal-rates Markov (ERM) random-branching model as a null model of diversification, it is relatively straightforward to test whether a descendant of a particular node has a significantly higher species richness than its sister clade (Slowinski and Guyer 1993). However, a test on the whole tree is more difficult to implement because each nodal statistic has to be assumed to be independent and able to realize any probability value between 0 and 1, a condition clearly violated in a tree-like hierarchy (Chan and Moore 2002). To avoid this problem, Monte Carlo approximations of the distribution of whole-tree statistics are used in SYMMETREE. Six M statistics were used: M_{Π}, M^*_{Π}, M_{Σ}, M^*_{Σ}, and M_R (Chan and Moore 2002; Moore et al. 2004) as well as the Ic (Colless 1982) imbalance measure, which have different sensitivities depending on where the asymmetry is located in the tree. For example, the M_R statistic is sensitive to large asymmetry, such as the ones found near the base of the tree. In contrast, M_{Σ} will be less sensitive to these basal differences in species richness.

Having established that the branches of the grass phylogenetic supertree diversified at significantly different rates, we also wanted to detect in which clades significant shifts in diversification have occurred. The approach used here, developed by Moore et al. (2004) and implemented in SYMMETREE, still assumed an underlying ERM model of branching but calculated two likelihood ratios for each three-taxon subtree defined by the internal nodes of the tree. The ratios test the likelihood of a one-rate homogeneous model of speciation against a two-rate heterogeneous model in which the two daughter lineages have different branching rates. One likelihood ratio is estimated at the root and one at the ingroup node of the three-taxon subtree (Moore et al. 2004). When comparing two sister clades' species numbers,
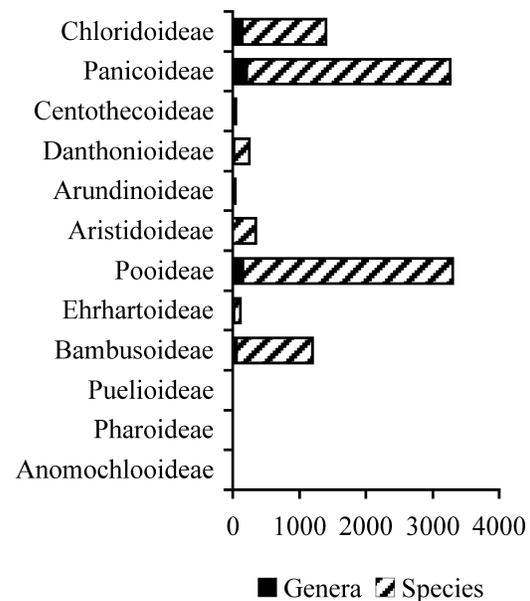

Fig. 1.   Distribution of species and genera in subfamilies of the grasses according to the GPWG (2001).

a significant difference could be due to a shift of diversification within one of the sister clades. The three-taxon subtree comparison of likelihood ratios condition the identification of a significant shift at the root node by the likelihood of a rate shift within the ingroup, therefore avoiding dubious significant shifts located deeper in the tree (Moore et al. 2004). The Δ_1 shift statistic was used because it was shown to have relatively low bias and outperformed other statistics in simulation (Moore et al. 2004). *P*-values for these statistics are obtained by numerical approximation using a Monte Carlo approach, specifically on each node of the tree. The value of the statistic itself is therefore not given here but only its significance level.

Diversification statistics will be influenced by taxon sampling, unless an entire generic tree of the family is used. The supertree has approximately two-thirds of the genera recognized in the grass family. A second supertree (the proportional tree) was therefore produced to investigate the effect of sample bias on the diversification statistics. The new tree had genera in proportion to the number of genera found in the tribes recognized by the GPWG (2001). Some genera needed to be removed from the original supertree, such as from Triticeae that were overrepresented, and were selected randomly within each tribe. Other genera needed to be added, such as to Cynodonteae and Paniceae, which were underrepresented in the supertree, and they were randomly connected to terminal taxa within the tribe until proportionality was reached. The procedure was repeated 100 times to average out the phylogenetic uncertainty introduced by adding taxa at specific places. Modifications at finer taxonomic scales than this were not attempted. Diversification statistics were then calculated for this new ''proportional tree.''

RESULTS

Summary Statistics and Sequence Availability

Figures 1 and 2 show the distribution of species and genera in grass subfamilies according to the classifications of
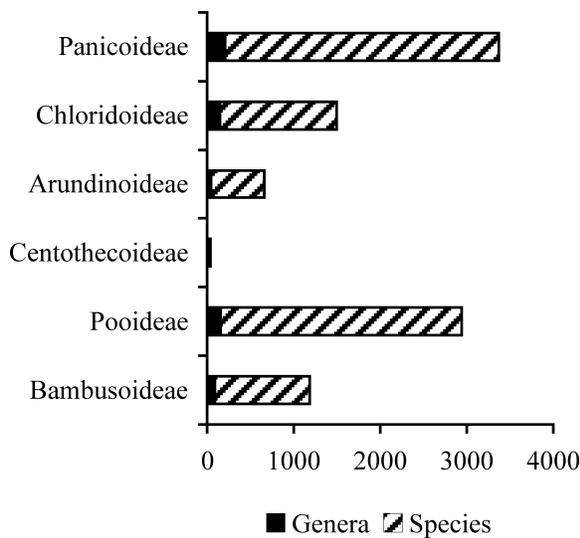
Fig. 2. Distribution of species and genera in subfamilies of the grasses according to Clayton and Renvoize (1986).

Fig. 3. Subfamily distribution of grass sequences deposited in DDBJ/EMBL/GenBank.

the GPWG (2001) and Clayton and Renvoize (1986), respectively. The GPWG (2001) system has 12 subfamilies and that of Clayton and Renvoize (1986) six. Panicoideae, Pooideae, Chloridoideae, and Bambusoideae dominate both systems in terms of genera and species numbers. However, Arundinoideae sensu Clayton and Renvoize (1986) have been reduced considerably in size by the GPWG (2001), with three subfamilies recognized or created (in the case of Danthonioideae) from its members (Aristidoideae, Arundinoideae, and Danthonioideae). Likewise, Bambusoideae sensu Clayton and Renvoize (1986) have been divided by the GPWG (2001), with five subfamilies recognized from its members (Anomochlooideae, Bambusoideae, Ehrhartoideae, Pharoideae, and Puelioideae). Pooideae have been expanded by the GPWG (2001).

Sequence availability in DDBJ/EMBL/GenBank for the production of large single-region or multi-region analyses is shown in Fig. 3, 4. Figure 3 shows the distribution of sequences across subfamilies. Ehrhartoideae, Panicoideae, and Pooideae dominate the available data. Arundinoideae, Bambusoideae, and Chloridoideae, despite being large subfamilies, are poorly represented. Figure 4 shows the number of grass species and genera sequenced for each of the top ten most frequent regions. ITS and *ndh*F were the most frequently sequenced regions with 370 and 351 species, respectively.

Supertrees

Figure 5 shows the strict consensus tree for the proportional supertree reconstruction. It has the same clades as the original supertree (not shown) and has been labeled to show the subfamilies and tribes. Visible printing of terminal taxa in the supertree is not possible at this size, and the original tree is available from the authors by request or from TreeBASE (study accession no. SN3054). Subfamilies, according to the GPWG (2001), are well defined in the supertree, and the topology of these clades is broadly congruent with most other major phylogenetic analyses of the grasses. The PACCAD clade is resolved and positioned in a polytomy with mainly Pooideae taxa (PACCAD-P). The panicoids
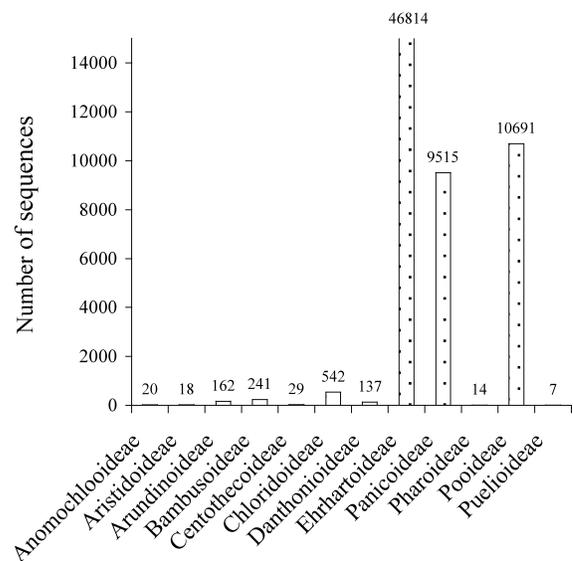
are found in a trichotomy with a centothecoid lineage (excluding *Centotheca* P. Beauv.) and a *Thysanolaena* Nees (Thysanolaeneae) lineage. The sister group of the panicoids is therefore not resolved. Sister to PACCAD-P is a bambusoid-ehrhartoid (B-E) group.

Within the PACCAD-P, B-E group, tribes are generally well defined (Fig. 5), but some are not monophyletic such as Arundinelleae, Aveneae, and Poeae. Some taxa are considered misplaced or unresolved in comparison to Clayton and Renvoize (1986) or the GPWG (2001), including *Hakonechloa* Makino ex Honda (arundinoid), *Luziola* Juss. (ehrhartoid), and *Sehima* Forssk. (panicoid) that group with the chloridoids. A number of other taxa with uncertain positions are also found in a polytomy with the PACCAD-P group, including *Craspedorhachis* Benth., *Farrago* Clayton, *Lopholepis* Decne., *Neurolepis* Meisn., *Pseudozoysia* Chiov., *Rhipidocladum* McClure, and *Zizaniopsis* Döll & Asch. Perhaps the strangest placements are those of *Cottea* Kunth (chloridoid) with the panicoids and *Poganatherum* P. Beauv. (panicoid) with the chloridoids.

Relationships beyond the PACCAD-P, B-E group are not fully resolved (Fig. 5). A unigeneric ehrhartoid lineage (*Microlaena* R. Br.) and a clade including *Puelia* Franch. (Puelioideae), *Henrardia* C. E. Hubb. (Pooideae), and two apparently misplaced panicoids (*Digitaria* Haller and *Oplismenus* P. Beauv.) form a trichotomy with PACCAD-P, B-E. Pharoideae and two lineages containing Anomochlooideae taxa (*Anomochloa* and *Streptochaeta*) diverge from the core grasses in that order. Some Bambuseae taxa (*Colanthelia* McClure & E. W. Sm., *Greslania* Balansa, *Melocalamus* Benth., *Merostachys* Spreng., *Oreobambos* K. Schum.), not generally considered as early diverging lineages, group with *Streptochaeta. Danthonidium* C. E. Hubb. (danthonioid) and *Pheidochloa* S. T. Blake (Eriachneae) also split unexpectedly from deep nodes in the tree.

Diversification Rate Variation

Both the original and proportional supertrees showed significant diversification rate variation at the 1% nominal level
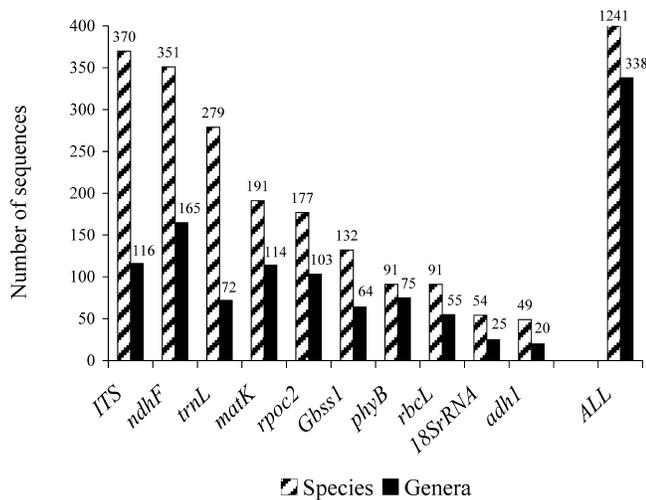
Fig. 4. Number of grass species and genera sequenced for each of the top ten most frequent regions in DDBJ/EMBL/GenBank.

for all six whole-tree test statistics used. The simulations under the equal-rate Markov process resulted, for each statistic in the trees, in less imbalance in generic number within clades than the observed supertrees. One whole-tree statistic, M^*_Σ, was the only test resulting in a slightly smaller P-value for the original supertree (<0.001 and 0.0012, respectively). For all the other five whole-tree tests, the significance values were identical or smaller than 0.001.

However, the number of shifts in diversification rate differed between the original and proportional supertrees (Fig. 5). P-values between 0.05 and 0.06 also were included because they highlighted groups showing different rates of diversification, although not significantly different at the 5% nominal level. Four shifts were significant (P-values of 0.033, 0.050, 0.046, and 0.025; Fig. 5), and two were just outside the 5% nominal value (P-values of 0.058 and 0.054; Fig. 5) if the number of genera in each tribe was proportional to the described numbers. On the original supertree, seven shifts were detected as significant (P-values of 0.025, 0.047, 0.026, 0.020, 0.043, 0.016, and 0.022; Fig. 5), whereas, four were just outside the 5% nominal value (P-values of 0.058, 0.052, 0.058, and 0.054; Fig. 5). Six shifts were in common to both trees (Fig. 5), the deepest representing the grasses, *Joinvillea* Gaudich. ex Brongn. & Gris (Joinvilleaceae), *Elegia* L. (Restionaceae), and *Flagellaria* L. (Flagellariaceae). Of these six shifts, the deepest significant shift within the grasses represents the spikelet clade, the next deepest the

majority of the chloridoids, and the others representing more derived clades (Fig. 5).

<center>DISCUSSION</center>

Data Availability

Comparison of the subfamily treatment of the GPWG (2001; Fig. 1) to that of Clayton and Renvoize (1986; Fig. 2) reveals contrasting broad-scale patterns of diversification. The new GPWG system, based on combined molecular, anatomical, and morphological data, has 12 subfamilies, which is double those recognized by Clayton and Renvoize (1986). The additional subfamilies were recognized in order to accommodate the non-monophyly of Arundinoideae and Bambusoideae sensu Clayton and Renvoize (1986). This classification has a more unbalanced distribution of species and genera among subfamilies than that of Clayton and Renvoize (1986) but reflects robust monophyletic groupings. It is more likely to represent broad diversification patterns in the family (major clades).

Sequencing of DNA regions for phylogenetics is becoming easier, and sequence data are accumulating at a rapid rate, especially for the grass family. A first examination of sequence availability in the family looks promising (Fig. 3) because there are 68,190 sequences deposited in DDBJ/ EMBL/GenBank (Jan 2004). However, examination of the subfamily distribution of these sequences reveals that most (over 98%) have been produced for Ehrhartoideae (68.6%), Pooideae (15.6%), and Panicoideae (13.9%). This is not surprising since they contain the most important cereal crops, with the sequencing of ca. 400 mega base pairs of the *Oryza* L. (Ehrhartoideae) genome complete (Adam 2000; Goff et al. 2002; Yu et al. 2002). Note that these figures are the number of entries in DDBJ/EMBL/GenBank and will therefore be an overestimation of what can be used in a phylogenetic analysis.

Supermatrices

For maximum-sized multigene analyses (maximum combined data sets) of the grasses, it is desirable to include data from some combination of the most frequently sequenced regions, with good taxonomic sampling across the family. The prime candidate regions for combination are therefore from the ''top ten'' most-sequenced regions. The number of grass species and genera sequenced for each of these regions (Jan 2004) is given in Fig. 4. The greatest number of species and genera sequenced for any particular DNA region is 370

<div align="right">→</div>

Fig. 5. Proportional supertree of the grasses indicating positions of shifts in diversification rate. A semi-strict consensus of 1000 trees is shown. *P*-values are for the proportional supertree except for those in parentheses that are from the original supertree; *P*-values above 0.05 have been shown in some cases as these are close to the arbitrary 0.05 significance limit. Subfamilies and tribes follow the GPWG (2001) except Aveneae and Parianeae (sensu Clayton and Renvoize 1986), which are treated as a synonyms of Poeae and Olyreae, respectively, by the GPWG (2001). Early diverging lineages include elements of Anomochlooideae, Bambuseae, and Pharoideae. The asterisk represents unresolved taxa (see text). The tribes are labeled as follows: Amp = Ampelodesmeae; And = Andropogoneae; Ano = Anomochloeae; Ari = Aristideae; Arundinel = Arundinelleae; Aru = Arundineae; Ave = Aveneae; Bam = Bambuseae; Bra = Brachyelytreae; Bro = Bromeae; Cen = Centotheceae; Cyn = Cynodonteae; Dan = Danthonieae; Dia = Diarrheneae; Ehr = Ehrharteae; Era = Eragrostideae; Eri = Eriachneae; Gua = Guaduelleae; Hai = Hainardieae; Isa = Isachneae; Lep = Leptureae; Lyg = Lygeeae; Mel = Meliceae; Mic = Micraireae; Nar = Nardeae; Oly = Olyreae; Orc = Orcuttieae; Ory = Oryzeae; Pap = Pappophoreae; Pan = Paniceae; Par = Parianeae; Phae = Phaenospermatideae; Pha = Phareae; Phy = Phyllorachideae; Poe = Poeae; Pue = Puelieae; Sti = Stipeae; Str = Streptogyneae; Strep = Streptochaeteae; Thy = Thysanolaeneae; Tri = Triticeae.
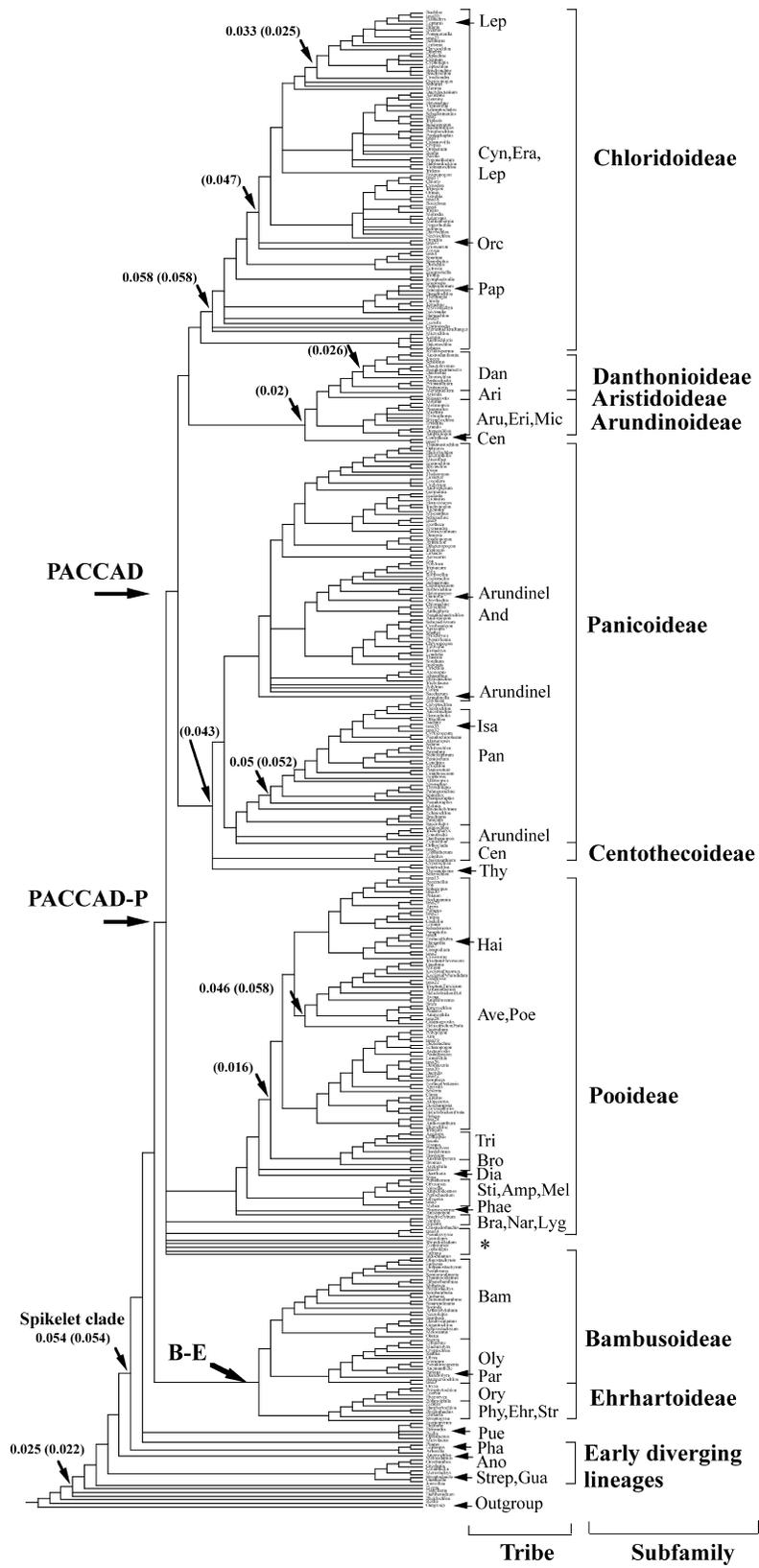
and 116, respectively, for nuclear ribosomal ITS. The plastid gene *ndh*F is the second best-represented region (351 species, 165 genera). Assuming complete overlap of taxa, the data set for the two regions would be limited to 351 species and 116 genera. The reality is far worse than this scenario as there is often poor overlap of taxa between regions. For example, representative species for *trn*L (exon and intron) are not present in six out of the 12 grass subfamilies. Combined data sets will therefore have to accommodate missing data, and targeted sequencing needs to be conducted to maximize future data set content.

Therefore, despite the great accumulation of sequence data for the grasses and the advances that have been made in determining the phylogenetic patterns among grass lineages (and the timing of their radiations), we are still a long way from a complete species-level or even generic-level tree of the family based on a supermatrix approach (single or multiple regions) without the incorporation of many missing data. These trees will have, among other things, great utility for macroevolutionary studies of grasses including diversification studies. A concerted targeted sequencing effort should be made that focuses on two aspects: (1) filling in gaps in existing data, and (2) improving taxonomic sampling across the family.

Supertrees

When sparse sampling of sequence data limits either single-region analyses or combination of sequence information into large multigene analyses, some form of ''divide and conquer'' strategy must be used (Sanderson and Driskell 2003; Wilkinson et al. 2005). Trees are built from individual data matrices and later assembled on the basis of taxonomic overlap with other such trees (Bininda-Emonds et al. 2002; Salamin et al. 2002; Bininda-Emonds 2003; Sanderson and Driskell 2003). From a theoretical stance, supertrees may be less favorable than large multigene analyses for phylogenetic reconstruction, but they may offer an adequate solution to the problem of constructing large trees when we consider current data availability.

Our supertree (Fig. 5) is broadly congruent with most published phylogenetic studies of the grass family. Poorly supported nodes from previous phylogenetic analyses account for most of the incongruence between the supertree and other published topologies (soft incongruence). The supertree contains nearly all grass tribes and 426 genera. To our knowledge it is the most complete phylogenetic tree of the grasses in terms of representation of genera. Some taxa have to be treated as unresolved or misplaced in the supertree because they cannot be grouped with other taxa or they appear to have erroneous placements. For example, *Hakonechloa* (arundinoid) and *Sehima* (panicoid) group with a broadly defined Chloridoideae sensu the GPWG (2001). A number of other genera with uncertain positions are also found in a polytomy with the PACCAD-P group (indicated by an asterisk in Fig. 5), including four chloridoids (*Craspedorhachis, Farrago, Lopholepis, Pseudozoysia*), *Zizaniopsis* (Oryzeae), and two Bambuseae (*Neurolepis, Rhipidocladum*). The positions of these Oryzeae and Bambuseae genera are highly unlikely because most other studies have found them grouping with their respective tribes and subfamilies (Clark

et al. 1995; GPWG 2001). Supertree reconstruction is sensitive to sampling overlap of taxa among source trees, and these problematic genera are typically found in only one or two of the source trees used.

Although weighting binary characters by bootstrap percentages has been shown to give more reliable supertrees (Salamin et al. 2002), it can have an effect on such taxa if the only binary characters available to place them are further down-weighted because of their low support in the source trees. Supertrees, like all trees, remain hypotheses of evolutionary relationships, and uncertainty in some taxon placements is inevitable. The use of a strict or semi-strict consensus tree as the final supertree produced by MRP can help highlight where those uncertainties are. The resulting polytomies can then be used in a topological model of diversification by iterating all possible resolutions of these polytomies and comparing their impact on diversification. This approach has been implemented in SYMMETREE, for example, and was used on the present grass supertree.

Anomochloa and *Streptochaeta* (both Anomochlooideae) are sister to the majority of grasses but are not grouped together (Fig. 5). Pharoideae are also resolved as an early diverging lineage. Although the order of early diverging lineages is consistent with Clark et al. (1995) and the GPWG (2001), bambusoid genera are intermixed among Anomochlooideae, Pharoideae, and Puelioideae. Within the clade comprising the rest of the grasses, the largest subgroup resolved is the PACCAD clade. The PACCAD clade is strongly supported in most phylogenetic studies of the grasses (GPWG 2001). An Arundinelleae-Paniceae group is sister to Andropogoneae. However, Andropogoneae contain some Paniceae genera sensu Clayton and Renvoize (1986); these are *Acroceras* Stapf, *Anthephora* Schreb., *Lasiacis* (Griseb.) Hitchc., *Pseudochaetochloa* Hitchc., *Thuarea* Pers., *Tricholaena* Schrad. ex. Schult. & Schult. f., and *Xerochloa* R. Br. The non-monophyly of Paniceae was also shown by Giussani et al. (2001) with *ndh*F sequences, and Duval et al. (2001) with *rpo*C2 sequences. Arundinelleae are not monophyletic, and this finding is consistent with Spangler et al. (1999) using *trn*L–F sequences (see also Kellogg 2000*b*). The sister group of the panicoids was not resolved. Centothecoideae were found to be positioned sister to a panicoid-*Gynerium* Willd. ex P. Beauv. group by the GPWG (2001) and *Thysanolaena* was included in the centothecoids. Consistent with the GPWG (2001), we found Centothecoideae to be closely related to Panicoideae, but their position was not fully resolved as they were found in a trichotomy with Panicoideae and a clade comprising *Cyperochloa* Lazarides & L. Watson, *Spartochloa* C. E. Hubb., and *Thysanolaena* (Thysanolaeneae). Centothecoideae are not monophyletic because *Centotheca* is positioned sister to a Danthonioideae, Aristidoideae, and Arundinoideae group and not with the other centothecoids. Chloridoideae are resolved, but their two major tribes, Cynodonteae and Eragrostideae sensu Clayton and Renvoize (1986), are not monophyletic. This is consistent with the *mat*K data of Hilu et al. (1999) that also conflicts with the tribal classification of the subfamily. The group of danthonioids, aristidoids, arundinoids, and the oddly positioned *Centotheca* is collectively sister to the chloridoids. The GPWG (2001) placed the arundinoids as a poorly supported sister group to the chloridoids. Eriachneae were

treated as incertae sedis by the GPWG (2001) but are embedded within the arundinoids in the supertree presented here.

Pooideae (P) are not monophyletic in the supertree (Fig. 5), but there is also no contrary evidence (the outlying pooid taxa form a polytomy with the PACCAD clade and the remaining pooids). The pooids are well supported in most other analyses (Soreng and Davis 2000; GPWG 2001). A Bambusoideae-Ehrhartoideae (B-E) group resolves as sister to the PACCAD-P group. The clade defined as BEP (sensu Clark et al. 1995; GPWG 2001) was therefore not resolved. Analyses of *rbc*L (Barker et al. 1995; Duvall and Morton 1996), ITS (Hsiao et al. 1999), plastid restriction sites (Davis and Soreng 1993), and morphology (GPWG 2001) placed the pooids as sister to the PACCAD clade. In contrast, *ndh*F (Clark et al. 1995; GPWG 2001) and *PHY*B (Mathews et al. 2000) resolved the BEP clade with moderate and high support, respectively. It is therefore more conservative to recognize a PACCAD, B-E, P clade instead of a PACCAD-BEP clade because single-region analyses conflicted in the placement of Bambusoideae, Ehrhartoideae, and Pooideae relative to the PACCAD clade. It has been demonstrated that many of these single-region analyses show random and systematic error (Salamin 2001). Salamin (2001) showed that in most regions, this error can be removed by the addition of more characters indicating random error; but in some regions inconsistency was detected (that is, increasing the number of characters compounds the error). However, in most cases our simulation studies have shown that if taxon sampling is improved by judiciously breaking long branches, then even this error can be reduced (N. Salamin et al. in prep.).

The reliability of supertrees has been questioned (Gatesy et al. 2002) and debated at length (Bininda-Emonds et al. 2002, 2003; Wilkinson et al. 2005). Many of the data reliability issues, including duplication, poor quality, and accountability, have been addressed (Bininda-Emonds et al. 2003; Wilkinson et al. 2005). Several empirical studies are adding support to the validity of the supertree approach. Salamin et al. (2002) used the trees from eight character partitions of the GPWG (2001) to produce supertrees. These trees were broadly congruent with the combined analysis of these data sets (the primary data of the GPWG analysis). The source trees in Salamin et al. (2002) and the current paper do incorporate some duplicated data (the source data are not totally independent). We recognize this limitation and are currently working on generating new supertrees that remove this problem. To do this we are generating optimum single-region trees from all available sequence data (primary data analyses of single regions) and then combining the trees using meta-analysis methods into supertrees (M. S. Kinney et al. in prep.). This removes the problem of duplication and non-independence of data. We are also evaluating how these supertrees compare with trees generated using the supermatrix approach (for the same data).

Differential Diversification Rates

Identification of shifts in diversification rates during the evolutionary history of grasses is an essential step to understanding what has caused their evolutionary and ecological success. Phylogenetic trees are extremely useful tools in such investigations because of the hierarchical structure they represent.

The genera in the supertree (Fig. 5) are represented in proportion to tribe size, so the shape of the tree provides a better visual indication of diversification patterns in the family than the original supertree. However, to discern whether differences in diversification rates at any particular point in the tree have occurred, significance tests are needed. We applied the *M* stats (Chan and Moore 2002, 2004) for the first time in the grasses and found that there has been significant diversification rate variation among lineages. These are variations in diversification not accountable by the stochastic nature of the process. Significant diversification rate variation was found in both the original supertree (diversification at the generic rank) and the proportional supertree ($P < 0.002$ in all *M* stats). Some clades contain more genera in the original supertree because they are better sampled than the others. Running the diversity statistics on the proportional tree was an attempt to remove this bias.

Having established that the tree was asymmetrical, we went on to identify the nodes where significant shifts in diversification have occurred. In other words, we identified clades that are more or less diverse than would be expected via random variation. More significant shifts (7) were found in the original tree than the proportional tree (4), and we can assume that at least some of these differences must be attributed to sampling bias. This shows that we have to be highly careful with sampling when using topology-based estimates of diversification. Large trees are favored over small ones, and ideally a full tree is required (in this case a full generic tree).

The nodes where significant shifts in diversification have occurred are spread relatively evenly across the tree (Fig. 5). A shift in diversification rate variation is found at a deep node including all the grasses and the earlier diverging lineages Joinvilleaceae, Restionaceae, and Flagellariaceae (in that order). Although significant, this shift should be interpreted with caution until expanded analyses are included that adequately sample the non-grass Poales. There are trends toward shifts in diversification (*P*-values of 0.054 for both the original and proportional tree) above the earliest-diverging lineages of grass including elements of Anomochlooideae, Pharoideae, Puelioideae, and all other grasses. This corresponds to the ''spikelet clade,'' a group of grasses with ''typical'' grass spikelets and structures homologous to glumes, lemmas, paleas, and lodicules. Two other shifts at relatively deep nodes correspond to the majority of chloridoids (*P*-values of 0.058 in both trees) and a mainly panicoid-centothecoid group (*P*-value of 0.043 in the original tree only). The panicoids and chloridoids represent the largest and third-largest subfamilies of grasses, respectively (Fig. 1). There are also significant, or marginally significant, rate shifts within the chloridoids (*P*-values of 0.033 and 0.025) in the proportional and original trees, respectively, corresponding to the clade including a mixture of Cynodonteae, Eragrostideae, and Leptureae, and the panicoids (*P*-values of 0.050, 0.052), corresponding to a large clade within Paniceae.

In the original tree, a significant shift (*P*-value of 0.016) is recorded corresponding to a pooid clade (including Aveneae, Bromeae, Hainardieae, Poeae, and Triticeae) but ex-

cluding the other pooid tribes. This shift was not found in the proportional tree, indicating that it may have resulted from sampling bias. However, a later shift (more recent in time) is recorded within the pooids in both trees (*P*-values of 0.046, 0.058), including some Aveneae and Poeae genera, indicating that diversification rate variation in the pooids occurred only after the ancestors of all their currently recognized tribes had evolved. Aveneae and Poeae are two highly diverse and phylogenetically poorly defined tribes that dominate pooids in terms of genus and species numbers (Clayton and Renvoize 1986; Soreng and Davis 2000). The factors promoting diversification will be better understood once a stable phylogenetic classification emerges for this subfamily.

Future Directions

Factors influencing diversification rate variation include prolific cladogenesis, adaptive radiation, key innovations, and mass extinctions (Chan and Moore 2002). The relative importance of each of these factors as causal agents for the diversification patterns we have detected is unknown. Further study is therefore required and will need to incorporate branch length information so that dates can be applied to the diversification rate shifts. This will also facilitate the examination of correlations between diversity rate variation and past environmental and global change factors such as temperature, aridity, and CO_2 levels. This study used genera and not species because of practical data/tree availability limitation. However, we envisage future studies that attempt to investigate species-level variation. A small number of genera including *Agrostis* L., *Digitaria* Haller, *Eragrostis* Wolf, *Festuca* L., *Panicum* L., *Paspalum* L., *Poa* L., and *Stipa* L. account for a high percentage of all Poaceae species. It is not known if these genera are monophyletic or what key innovations or circumstances have led to such high diversification rates in these genera.

It has been proposed that certain traits, or key innovations, might influence the rate of evolution and production of new species (Burger 1981; Maynard Smith and Szathmary 1995). The observed differences in species richness between certain clades would then be correlated with the presence of particular key innovations. However, to identify correlates of species richness, the hierarchical nature of evolutionary history has to be taken into account to avoid erroneous inferences (Felsenstein 1985; Harvey and Pagel 1991). Comprehensive phylogenetic trees are required for such studies to have a more complete view of the process (Pagel 1999; Salamin and Davies 2004). One approach to help study the causes (correlations) of diversification and species richness is to use sister clade comparison tests (Slowinski and Guyer 1993; Purvis 1996). In these tests, comparisons are made between sister taxa, one possessing a trait/factor and the other not. For example, Salamin and Davies (2004) mapped traits from Watson and Dallwitz (1992) onto a supertree and identified all sister clades with contrasting traits for the grasses (e.g., annual vs. perennial life form). A comparison was then made between number of species in each sister clade against the null hypothesis of equal speciation rates using the methods of Slowinski and Guyer (1993) and Goudet (1999). They found that an herbaceous habit and an annual life cycle have a significant correlation, at the 5% level, with species rich-

ness in grasses. This supports the hypothesis that annuals might be better able to fit new niches and become more species rich (Bousquet et al. 1992). Annuals clearly also have a short generation time that will facilitate increased genetic recombination and change. Woodiness is also linked to generation time in grasses. Many of the woody bamboos have long generation times, with some species only flowering every decade or more (Clayton and Renvoize 1986). Although the link between speciation rates and nucleotide substitution rates has been established in other taxonomic groups (Barraclough et al. 1996; Savolainen and Goudet 1998; Barraclough and Savolainen 2001), the evidence is inconclusive in the grasses (Gaut et al. 1997). The analysis by Gaut et al. (1997) was restricted to a small fraction of grass diversity, and extending the sampling could change its outcome. However, the results of Salamin and Davies (2004) indicated that generation time is a factor influencing species richness in grasses. The results of the analyses failed to show any link between a number of other characters and speciation rate, including the ability to resist drought, ability to tolerate salty environments, open vs. forest habitats, and bisexual vs. monoecious breeding systems. The study was not exhaustive, and many other factors require investigation. For example, Kellogg (2000*a*) identified C_4 metabolism as a possible key innovation influencing species richness (by simply comparing lineages with known species numbers).

This study has generated a near-complete tribal-level tree of the grasses, but we are still a long way from a complete and accurate phylogenetic tree of the grasses at the genus and species levels. However, the large trees generated have facilitated the study of macroevolution by allowing, for example, statistical tests regarding patterns of diversification-rate variation and causes of such variation to be made. We are currently producing additional supertrees and large trees using supermatrix methods to help study diversity in the family. For example, we are generating improved supertrees by removing any non-independent data, and we are evaluating their accuracy and utility in comparison to supermatrix methods. We are also incorporating diversification tests that allow the temporal nature of phylogenetic trees (branch lengths) to be included in our trees and are further investigating the effects of phylogenetic uncertainty on our findings.

LITERATURE CITED

ADAM, D. 2000. Now for the hard ones. *Nature* **408**: 792–793.
ARCHIBOLD, O. I. V. 1995. Ecology of world vegetation. Chapman and Hall, London, UK. 510 p.
BALDAUF, S. L. 2003. The deep roots of eukaryotes. *Science* **300**: 1703–1706.
BARKER, N. P., H. P. LINDER, AND E. H. HARLEY. 1995. Polyphyly of Arundinoideae (Poaceae): evidence from *rbc*L sequence data. *Syst. Bot.* **20**: 423–435.
———, ———, AND ———. 1999. Sequences of the grass-specific

insert in the chloroplast *rpo*C2 gene elucidate generic relationships of the Arundinoideae (Poaceae). *Syst. Bot.* **23**: 327–350.

BARRACLOUGH, T. G., P. H. HARVEY, AND S. NEE. 1996. Rate of *rbc*L gene sequence evolution and species diversification in flowering plants (angiosperms). *Proc. Roy. Soc. Lond., Ser. B, Biol. Sci.* **263**: 589–591.

———, AND S. NEE. 2001. Phylogenetics and speciation. *Trends Ecol. Evol.* **16**: 391–399.

———, AND V. SAVOLAINEN. 2001. Evolutionary rates and species diversity in flowering plants. *Evolution* **55**: 677–683.

BININDA-EMONDS, O. R. P. 2003. Novel versus unsupported clades: assessing the qualitative support for clades in MRP supertrees. *Syst. Biol.* **52**: 839–848.

———, J. L. GITTLEMAN, AND M. A. STEEL. 2002. The (super)tree of life: procedures, problems, and prospects. *Ann. Rev. Ecol. Syst.* **33**: 265–289.

———, K. E. JONES, S. A. PRICE, R. GRENYER, M. CARDILLO, M. HABIB, A. PURVIS, AND J. L. GITTLEMAN. 2003. Supertrees are a necessary not-so-evil: a comment on Gatesy et al. *Syst. Biol.* **52**: 724–729.

BOUSQUET, J., S. H. STRAUS, A. H. DOERKENSEN, AND R. A. PRICE. 1992. Extensive variation in evolutionary rate of *rbc*L gene-sequences among seed plants. *Proc. Natl. Acad. Sci. U.S.A.* **89**: 7844–7848.

BRIGGS, B. G., A. D. MARCHANT, S. GILMORE, AND C. L. PORTER. 2000. A molecular phylogeny of Restionaceae and allies, pp. 661–671. *In* S. W. L. Jacobs and J. E. Everett [eds.], Grasses: systematics and evolution. CSIRO Publishing, Collingwood, Victoria, Australia.

BURGER, W. C. 1981. Why are there so many kinds of flowering plants? *Bioscience* **31**: 577–581.

CHAN, K. M. A., AND B. R. MOORE. 2002. Whole-tree methods for detecting differential diversification rates. *Syst. Biol.* **51**: 855–865.

———, AND ———. 2004. SYMMETREE: whole-tree analysis of differential diversification rates. *Bioinformatics* **21**: 1709–1710.

CLARK, L. G., M. KOBAYASHI, S. MATHEWS, R. E. SPANGLER, AND E. A. KELLOGG. 2000. The Puelioideae, a new subfamily of Poaceae. *Syst. Bot.* **25**: 181–187.

———, W. ZHANG, AND J. F. WENDEL. 1995. A phylogeny of the grass family (Poaceae) based on *ndh*F sequence data. *Syst. Bot.* **20**: 436–460.

CLAYTON, W. D., AND S. A. RENVOIZE. 1986. Genera graminum: grasses of the world. *Kew Bull., Addit. Ser.* **13**: 1–389.

COLLESS, D. H. 1982. Review of *Phylogenetics: theory and practice of phylogenetic systematics,* by E. O. Wiley. *Syst. Zool.* **31**: 100–104.

DAVIS, J. I, AND R. J. SORENG. 1993. Phylogenetic structure in the grass family (Poaceae) as inferred from chloroplast DNA restriction site variation. *Amer. J. Bot.* **80**: 1444–1454.

DOEBLEY, J., M. DURBIN, E. M. GOLENBERG, M. T. CLEGG, AND D. P. MA. 1990. Evolutionary analysis of the large subunit of carboxylase (*rbc*L) nucleotide sequence among the grasses (Gramineae). *Evolution* **44**: 1097–1108.

DUVALL, M. R., AND B. R. MORTON. 1996. Molecular phylogenetics of Poaceae: an expanded analysis of *rbc*L sequence data. *Molec. Phylogenet. Evol.* **5**: 353–358.

———, J. D. NOLL, AND A. H. MINN. 2001. Phylogenetics of Paniceae (Poaceae). *Amer. J. Bot.* **88**: 1988–1992.

EISEN, J. A., AND C. M. FRASER. 2003. Phylogenomics: intersection of evolution and genomics. *Science* **300**: 1706–1707.

FELSENSTEIN, J. 1985. Phylogenies and the comparative method. *Amer. Nat.* **125**: 1–12.

GATESY, J., C. MATTHEE, R. DESALLE, AND C. HAYASHI. 2002. Resolution of a supertree/supermatrix paradox. *Syst. Biol.* **51**: 652–664.

GAUT, B. S., L. G. CLARK, J. F. WENDEL, AND S. V. MUSE. 1997. Comparisons of the molecular evolutionary process at *rbc*L and *ndh*F in the grass family (Poaceae). *Molec. Biol. Evol.* **14**: 769–777.

GIUSSANI, L. M., J. H. COTA-SÁNCHEZ, F. O. ZULOAGA, AND E. A. KELLOGG. 2001. A molecular phylogeny of the grass subfamily Panicoideae (Poaceae) shows multiple origins of C_4 photosynthesis. *Amer. J. Bot.* **88**: 1993–2012.

GOFF, S. A., D. RICKE, T.-H. LAN, G. PRESTING, R. WANG, M. DUNN, J. GLAZEBROOK, A. SESSIONS, P. OELLER, H. VARMA, D. HADLEY, D. HUTCHISON, C. MARTIN, F. KATAGIRI, B. M. LANGE, T. MOUGHAMER, Y. XIA, P. BUDWORTH, J. ZHONG, T. MIGUEL, U. PASZKOWSKI, S. ZHANG, M. COLBERT, W.-L. SUN, L. CHEN, B. COOPER, S. PARK, T. C. WOOD, L. MAO, P. QUAIL, R. WING, R. DEAN, Y. YU, A. ZHARKIKH, R. SHEN, S. SAHASRABUDHE, A. THOMAS, R. CANNINGS, A. GUTIN, D. PRUSS, J. REID, S. TAVTIGIAN, J. MITCHELL, G. ELDREDGE, T. SCHOLL, R. M. MILLER, S. BHATNAGAR, N. ADEY, T. RUBANO, N. TUSNEEM, R. ROBINSON, J. FELDHAUS, T. MACALMA, A. OLIPHANT, AND S. BRIGGS. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* **296**: 79–92.

GÓMEZ-MARTÍNEZ, R., AND A. CULHAM. 2000. Phylogeny of the subfamily Panicoideae with emphasis on the tribe Paniceae: evidence from the *trn*L–F cpDNA region, pp. 136–140. *In* S. W. L. Jacobs and J. E. Everett [eds.], Grasses: systematics and evolution. CSIRO Publishing, Collingwood, Victoria, Australia.

GOUDET, J. 1999. An improved procedure for testing the effects of key innovations on rate of speciation. *Amer. Nat.* **153**: 549–555.

GRASS PHYLOGENY WORKING GROUP [GPWG]. 2001. Phylogeny and subfamilial classification of the grasses (Poaceae). *Ann. Missouri Bot. Gard.* **88**: 373–457.

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

HARVEY, P. H., AND M. D. PAGEL. 1991. The comparative method in evolutionary biology. Oxford University Press, Oxford, UK. 239 p.

HILU, K. W., L. A. ALICE, AND H. P. LIANG. 1999. Phylogeny of Poaceae inferred from *mat*K sequences. *Ann. Missouri Bot. Gard.* **86**: 835–851.

HODKINSON, T. R., M. W. CHASE, M. D. LLEDO, N. SALAMIN, AND S. A. RENVOIZE. 2002*a*. Phylogenetics of *Miscanthus, Saccharum* and related genera (Saccharinae, Andropogoneae, Poaceae) based on DNA sequences from ITS nuclear ribosomal DNA and plastid *trn*L intron and *trn*L–F intergenic spacers. *J. Plant Res.* **115**: 381–392.

———, ———, C. TAKAHASHI, I. J. LEITCH, M. D. BENNETT, AND S. A. RENVOIZE. 2002*b*. The use of DNA sequencing (ITS and *trn*L–F), AFLP, and fluorescent in situ hybridization to study allopolyploid *Miscanthus* (Poaceae). *Amer. J. Bot.* **89**: 279–286.

———, AND J. A. N. PARNELL (editors). In press *a*. Reconstructing the tree of life: taxonomy and systematics of species rich groups. Taylor and Francis, CRC Press, Florida, USA.

———, AND ———. In press *b*. Introduction to the systematics of species rich groups. *In* T. R. Hodkinson and J. A. N. Parnell [eds.], Reconstructing the tree of life: taxonomy and systematics of species rich groups. Taylor and Francis, CRC Press, Florida, USA.

———, V. SAVOLAINEN, S. W. L. JACOBS, Y. BOUCHENAK-KHELLADI, M. S. KINNEY, AND N. SALAMIN. In press. Supersizing: progress in documenting and understanding grass species richness. *In* T. R. Hodkinson and J. A. N. Parnell [eds.], Reconstructing the tree of life: taxonomy and systematics of species rich taxa. Taylor and Francis, CRC Press, Florida, USA.

HSIAO, C., S. W. L. JACOBS, N. J. CHATTERTON, AND K. H. ASAY. 1999. A molecular phylogeny of the grass family (Poaceae) based on the sequences of nuclear ribosomal DNA (ITS). *Austral. Syst. Bot.* **11**: 667–688.

KELLOGG, E. A. 2000*a*. The grasses: a case study in macroevolution. *Ann. Rev. Ecol. Syst.* **31**: 217–238.

———. 2000*b*. Molecular and morphological evolution in the Andropogoneae, pp.149–158. *In* S. W. L. Jacobs and J. E. Everett

[eds.], Grasses: systematics and evolution. CSIRO Publishing, Collingwood, Victoria, Australia.

LIANG, H., AND K. W. HILU. 1995. Application of the *mat*K gene sequences to grass systematics. *Canad. J. Bot.* **74**: 125–134.

MACE, G. M., J. L. GITTLEMAN, AND A. PURVIS. 2003. Preserving the tree of life. *Science* **300**: 1707–1709.

MASON-GAMER, R. J., C. F. WEIL, AND E. A. KELLOGG. 1998. Granule-bound starch synthase: structure, function, and phylogenetic utility. *Molec. Biol. Evol.* **15**: 1658–1673.

MATHEWS, S., AND R. A. SHARROCK. 1996. The phytochrome gene family in grasses (Poaceae): a phylogeny and evidence that grasses have a subset of the loci found in dicot angiosperms. *Molec. Biol. Evol.* **13**: 1141–1150.

————, R. C. TSAI, AND E. A. KELLOGG. 2000. Phylogenetic structure in the grass family (Poaceae): evidence from the nuclear gene phytochrome B. *Amer. J. Bot.* **87**: 96–107.

MAYNARD SMITH, J., AND E. SZATHMARY. 1995. The major transitions in evolution. W. H. Freeman, Oxford, UK. 346 p.

MICHELANGELI, F. A., J. I DAVIS, AND D. W. STEVENSON. 2003. Phylogenetic relationships among Poaceae and related families as inferred from morphology, inversions in the plastid genome, and sequence data from the mitochondrial and plastid genomes. *Amer. J. Bot.* **90**: 93–106.

MOORE, B. R., K. M. A. CHAN, AND M. J. DONOGHUE. 2004. Detecting diversification rate variation in supertrees, pp. 487–533. *In* O. R. P. Bininda-Emonds [ed.], Phylogenetic supertrees: combining information to reveal the tree of life. Kluwer Academic, Dordrecht, The Netherlands.

NADOT, S., R. BAJON, AND B. LEJEUNE. 1994. The chloroplast gene *rps*4 as a tool for the study of Poaceae phylogeny. *Pl. Syst. Evol.* **191**: 27–38.

PAGEL, M. 1999. Inferring the historical patterns of biological evolution. *Nature* **401**: 877–884.

PENNISI, E. 2003. Modernizing the tree of life. *Science* **300**: 1692.

PURVIS, A. 1996. Using interspecies phylogenies to test macroevolutionary hypotheses, pp. 153–168. *In* P. H. Harvey, A. J. L. Brown, J. M. Smith, and S. Nee [eds.], New uses for new phylogenies. Oxford University Press, Oxford, UK.

RENVOIZE, S., AND W. CLAYTON. 1992. Classification and evolution of the grasses, pp. 3–37. *In* G. Chapman [ed.], Grass evolution and domestication. Cambridge University Press, Cambridge, UK.

SALAMIN, N. 2001. Large trees, supertrees and the grass phylogeny. Ph.D. thesis, University of Dublin, Trinity College, Dublin, Ireland.

————, AND T. J. DAVIES. 2004. Using supertrees to investigate species richness in grasses and flowering plants, pp. 461–486. *In* O. R. P. Bininda-Emonds [ed.], Phylogenetic supertrees: combining information to reveal the tree of life. Kluwer Academic Publishers, Dordrecht, The Netherlands.

————, T. R. HODKINSON, AND V. SAVOLAINEN. 2002. Building su-

pertrees: an empirical assessment using the grass family (Poaceae). *Syst. Biol.* **51**: 136–150.

————, ————, AND ————. 2005. Towards building the tree of life: a simulation study for all angiosperm genera. *Syst. Biol.* **54**: 183–196.

SANDERSON, M. J., AND A. C. DRISKELL. 2003. The challenge of constructing large phylogenetic trees. *Trends Pl. Sci.* **8**: 374–379.

SAVOLAINEN, V., AND J. GOUDET. 1998. Rate of gene sequence evolution and species diversification in flowering plants: a re-evaluation. *Proc. Roy. Soc. London, Ser. B, Biol. Sci.* **265**: 603–607.

SLOWINSKI, J. B., AND C. G. GUYER. 1993. Testing whether certain traits have caused amplified diversification: an improved method based on a model of random speciation and extinction. *Amer. Naturalist* **142**: 1019–1024.

SORENG, R. J., AND J. I. DAVIS. 1998. Phylogenetics and character evolution in the grass family (Poaceae): simultaneous analysis of morphological and chloroplast DNA restriction site character sets. *Bot. Rev. (Lancaster)* **64**: 1–85.

————, AND ————. 2000. Phylogenetic structure in Poaceae subfamily Pooideae as inferred from molecular and morphological characters: misclassification vs. reticulation, pp. 61–74. *In* S. W. L. Jacobs and J. E. Everett [eds.], Grasses: systematics and evolution. CSIRO Publishing, Collingwood, Victoria, Australia.

SPANGLER, R., B. ZAITCHIK, E. RUSSO, AND E. KELLOGG. 1999. Andropogoneae evolution and generic limits in *Sorghum* (Poaceae) using *ndh*F sequences. *Syst. Bot.* **24**: 267–281.

SWOFFORD, D. L. 2000. PAUP*: phylogenetic analysis using parsimony (*and other methods), vers. 4.08b. Sinauer Associates, Inc., Sunderland, Massachusetts, USA.

WATSON, L., AND M. J. DALLWITZ. 1992. The grass genera of the world. CAB International, Wallingford, Oxfordshire, UK. 1038 p.

WILKINSON, M., J. COTTON, C. CREEVEY, O. EULENSTEIN, S. HARRIS, F. LAPOINTE, C. LEVASSEUR, J. MCINERNEY, D. PISANI, AND J. THORLEY. 2005. The shape of supertrees to come: tree shape related properties of fourteen supertree methods. *Syst. Biol.* **54**: 419–431.

YU, J., S. HU, J. WANG, G. K.-S. WONG, S. LI, B. LIU, Y. DENG, L. DAI, Y. ZHOU, X. ZHANG, M. CAO, J. LIU, J. SUN, J. TANG, Y. CHEN, X. HUANG, W. LIN, C. YE, W. TONG, L. CONG, J. GENG, Y. HAN, L. LI, W. LI, G. HU, X. HUANG, W. LI, J. LI, Z. LIU, L. LI, J. LIU, Q. QI, J. LIU, L. LI, T. LI, X. WANG, H. LU, T. WU, M. ZHU, P. NI, H. HAN, W. DONG, X. REN, X. FENG, P. CUI, X. LI, H. WANG, X. XU, W. ZHAI, Z. XU, J. ZHANG, S. HE, J. ZHANG, J. XU, K. ZHANG, X. ZHENG, J. DONG, W. ZENG, L. TAO, J. YE, J. TAN, X. REN, X. CHEN, J. HE, D. LIU, W. TIAN, C. TIAN, H. XIA, Q. BAO, G. LI, H. GAO, T. CAO, J. WANG, W. ZHAO, P. LI, W. CHEN, X. WANG, Y. ZHANG, J. HU, J. WANG, S. LIU, J. YANG, G. ZHANG, Y. XIONG, Z. LI, L. MAO, C. ZHOU, Z. ZHU, R. CHEN, B. HAO, W. ZHENG, S. CHEN, W. GUO, G. LI, S. LIU, M. TAO, J. WANG, L. ZHU, L. YUAN, AND H. YANG. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* **296**: 79–92.

ZHANG, W. P. 2000. Phylogeny of the grass family (Poaceae) from *rpl*16 intron sequence data. *Molec. Phylogenet. Evol.* **15**: 135–146.