Chapter 5

Phylogeny of Poaceae Based on *mat*K Gene Sequences

5.1 Introduction

Phylogenetic reconstruction in the Poaceae began early in this century with proposed evolutionary hypotheses based on assessment of existing knowledge of grasses (e.g., Bew, 1929; Hubbard 1948; Prat, 1960; Stebbins, 1956, 1982; Clayton, 1981; Tsvelev, 1983). Imperical approaches to phylogenetic reconstruction of the Poaceae followed those initial hypotheses, starting with cladistic analyses of morphological and anatomical characters (Kellogg and Campbell, 1987; Baum, 1987; Kellogg and Watson, 1993). More recently, molecular information has provided the basis for phylogenetic hypotheses in grasses at the subfamily and tribe levels (Table 5.1). These molecular studies were based on information from chloroplast DNA (cpDNA) restriction sites and DNA sequencing of the *rbc*L, *ndh*F, *rps*4, 18S and 26S ribosomal DNA (rDNA), phytochrome genes, and the ITS region (Hamby and Zimmer, 1988; Doebley et al., 1990; Davis and Soreng, 1993; Cummings, King, and Kellogg, 1994; Hsiao et al., 1994; Nadot, Bajon, and Lejeune, 1994; Barker, Linder, and Harley, 1995; Clark, Zhang, and Wendel, 1995; Duvall and Morton, 1996; Liang and Hilu, 1996; Mathews and Sharrock, 1996).

Although these studies have refined our concept of grass evolution at the subfamily level and, to a certain degree, at the tribal level, major disagreements and questions remain to be addressed. Outstanding discrepancies at the subfamily level include: 1) Are the pooids, bambusoids senso lato, or herbaceous bamboos the most basal lineages in the family? 2) Do the herbaceous bamboos, such as Streptochaeteae and Anomochloeae, represent a distinct phylogenetic entity that deserves a subfamilial taxonomic treatment? 3) Is there an early major dichotomy in the family such as the proposed Panicoideae, Arundinoideae, Chloridoideae, and Centothecoideae (PACC group) and the bambusoid-pooid grasses (BOP group)? 4) Where do the centothecoid taxa belong in the phylogeny? 5) Do the oryzoid and bambusoid grasses belong to a single monophyletic lineage? 6) What is the phylogenetic position of tribes like the Aristideae and Ehrharteae.

TABLE 5.1. Summary of the phylogenetic studies of Poaceae using DNA sequences

Region	Total length	Variable sites		Informative Sites		Species	Authors
		Number	%	Number	%		
18s, 26s rRNA	1600	244	15.3%	85	5.3%	9	Hanmby & Zimmer 1988
5.8s rRNA & ITS	603	269	44.6%	118	19.6%	9	Hsiao et al. 1994
*rps*4	606	142	23.4%	62	10.2%	28	Nadot et al. 1994
*rpo*C2	543					11	Cummings et al 1994
*ndh*F	2124	749	35.3%	487	65.0%	45	Clark, Zhang, and Wendel 1995
*rbc*L	1425					9	Doebley et al. 1990
*rbc*L	1344			155	11.5%	34	Barker, Linder, and Harley 1995
*rbc*L						23	Duvall & Morton 1996
*mat*K	583	175	30.0%	87	14.9	18	Liang & Hilu (1996)

In this study, the *mat*K gene has been chosen to address these and other questions pertaining to grass systematics at higher taxonomic levels. Some of the previous nucleic acid studies on grasses had the shortcomings of using small sample size, choosing a very distant outgroup, and basing the analyses on too few informative sites to resolve clades. This latter case was due to either using small genes or genes that have low rates of substitutions. The *mat*K gene is ≈ 1500 base pairs (bp), located within the intron of the chloroplast gene *trn*K, and functionally might be involved in splicing group II introns (Neuhaus and Link, 1987; Ems et al., 1995). The effective application of this gene to grass systematics has already been documented (Liang and Hilu, 1996). The *mat*K gene is known to have relatively high rates of substitution compared with other genes used in grass systematics, possesses high proportions of transversion mutations, and the 3' section of its coding region has been proven quite useful for constructing phylogenies at the subfamily level in the Poaceae (Liang and Hilu, 1996).

5.2 Materials and Methods

5.2.1 Plant Materials and Nucleic Acid Methods

About 966 bps from the 3' end of the *mat*K gene were sequenced from grass species representing 48 genera from 21 tribes and seven subfamilies that are commonly recognized in the Poaceae (Table 5.2). The genera were selected to represent as adequately as possible the grass subfamilies and tribes. *Joinvillea plicata* (Joinvilleaceae) was used as an outgroup species. Total genomic DNA was isolated from leaves of either seedlings or fully mature plants. The procedures of DNA isolation, electrophoresis, PCR amplification, and DNA sequencing are those described in Liang and Hilu (1996). Primers used in sequencing and their position are reported in Table 5.3. Manual sequencing was used for all primers except for matK1176, where an automated sequencer was employed to cover a long stretch of the coding region . In the latter, the *trn*K region was amplified, electrophoresed on 1.0 agarose gels, and the DNA fragment excised and purified on QiaGen columns (QiaGen Inc.). The ABI Prism TM 377 Automated DNA

Sequencer with *Taq* polymerase, DyeDeoxy TM terminator cycle sequencing method was used.

TABLE 5.2. Forty-nine taxa and their respective tribes and subfamilies used in sequence analysis

Subfamilies & tribes	Species	Sources of material
Arundinoideae		
Aristideae	*Aristida adscensionis* L.	Hilu, KH5516
	Stipagrostis uniplumis (Licht. ex Roem. & Schult.) De Winter	USDA, PI365033
Arundineae	*Arundo donax* L.	Hilu, KH5546
	Cortaderia selloana (Schult. & Schult. f.) Asch. & Graebn.	USDA, PI490290
	Molinia caerulea (L.) Moench	L. Clark 1165
	Pharagmites australis (Cav.) Trin. ex Steud.	Gary P. Flemming, Assatague Island, VA
Bambusoideae		
Anomochloeae	*Chesquea circinata* Soderstr. & C. Calderon	Quail Botanical Garden
Bambuseae	*Arundinaria japonica* (Siebold & Zucc.) Makino	Liang, HL9508
	Phyllostachys aurea Riv.	Hilu, KH9418
	Sasa kurilensis Rupr.	Liang, HL9615
Phareae	*Pharus lappulaceus* Aubl.	L. Clark 1329
Streptochaeteae	*Streptochaeta angustifolia* Soderstr.	L. Clark 1304
Centothecoideae		
Centotheceae	*Zeugites pittieri* Hack.	L. Clark, 1171
Chloridoideae		
Chlorideae	*Cynodon dactylon* (L.) Pers.	USDA, PI224149
	Chloris distichophylla Lag.	Hilu. KH5538
	Cypholepis yemenica (Schweinf.) Chiov.	USDA, PI364502
	Dactyloctenium aegyptiacum (L.) Beauv.	USDA, PI271559
	Tragus berteronianus Schult.	Barker 1128, South Africa
Eragrosteae	*Eleusine indica* (L.) Gaertn.	USDA, PI408801
	Eragrostis capensis (Thunb,) Trin.	Hilu, KH5539
Pappophoreae	*Enneapogon cenchroides* (Licht. ex Roem. & Schult.) C. E. Hubb.	USDA, PI299902

	Pappophorum bicolor Fourn.	Hilu, KH5542
Sporoboleae	*Sporobolus fimbriatus* (Trin.) Nees	Hilu, KH5545
Zoysieae	*Perotis patens* Grand.	USDA, PI364995
Oryzoideae		
	Ehrharta erecta Lam.	Hilu, KH9417
	Oryza sativa L.*	
	Zizania aquatica L.	Hilu, KH9423
Panicoideae		
Andropogoneae	*Andropogon gerardi* Vitman	USDA, PI215661
	Hyparrhenia hirta (L.) Stapf	Barker 1134, South Africa
	Sorghum bicolor (L.) Moench	Hilu, KH9408
	Zea deploperenis Iltis, Dobley & Guzman	Hugh Iltis, Univ. of Wisconsin
Arundinelleae	*Tristachya biseriate* Stapf	Barker 1126, South Africa
Paniceae	*Echinochloa utilis* Ohwi & Yabuno	Hilu, KH4040
	Panicum capillare L.	USDA, PI220025
	Digitaria sanguinalis (L.) Scop.	Sharp Brothers Seed Co.,
	Paspalum almum Chase	USDA, PI303958
Pooideae		
Agrostideae	*Agrostis scabra* Willd.	USDA, PI234681
	Phleum pratense L.	USDA, PI303130
Aveneae	*Avena sativa* L.	Hilu, KH9406
	Phalaris angusta Nees ex Trin.	USDA, PI207928
Bromeae	*Bromus inermis* Leyss	USDA, PI173645
Poeae	*Briza erecta* Lam.	USDA, PI282880
	Poa pratensis L.	USDA, PI620724
Stipeae	*Stipa armeniaca* P. Smirn.	Hilu, KH5508
Triticeae	*Brachypodium distachyon* (L.) P. Beauv.	USDA, PI250647
	Hordeum vulgare L.*	
	Leymus angustus (Trin.) Pilg.	USDA, PI547359
	Triticum aestivum “Massey”	Startling, VA 1988
Joinvilleaceae	*Joinvillea plicata* (Hooker f.) Newell & Stone	C. S. Campbell

* Sequence available in GenBank.

5.2.2 Data Analysis

The entire DNA sequences were aligned with the Clustal W 1.6 (Thompson, Higgins, and Gibson, 1994) computer program. Because of the lack of indels, alignment of the *mat*K coding region was not difficult. Sequence statistics such as nucleotide frequencies, transition/transversion (tr/tv) ratio and variability in different regions of the sequences were computed by MEGA 1.01 (Kumar, Tamura, and Nei, 1993) and MacClade 3.0 (Maddison and Maddison, 1992). The Relative Apparent Synapomorphy Analysis (RASA 2.1; Lyons-Weiler, 1997) program was used to evaluate the phylogenetic signal in the data set. The aligned sequences were used as input data for PAUP 3.1.1 (Swofford, 1991) and MEGA. The data set was transformed into a NEXUS format using MacClade 3.0 and then analyzed by the Wagner parsimony method of PAUP using *Joinvillea* as an outgroup. The parsimony analyses were performed using a heuristic search with MULPARS, tree-bisection-reconnection (TBR) branch swapping, and CLOSEST addition to estimate relationships and tree topology. Strict consensus trees were generated by the CONSENSUS option based on the most equally parsimonious trees produced by heuristic search. Bootstrapping (100 replicate; Felsenstein, 1985) and decay analyses (Bremer, 1988; Donoghue et al., 1992) were performed with PAUP to determine relative support for the clades.

TABLE 5.3. Structure and location of primers used in sequencing

Primers	Sequence	Location
matK5	5’ CGA TCC TTT CAT GCA TT 3’	915
matK5B	5’ CGA TCC TCT TAT GCA TT TG 3’	915
matK7	5’ GTA TTA GGG CAT CCC ATT 3’	1152
matK7B	5’ GAT TTA TCA (G) GAT TGG GAT 3’	1188
matK8	5’ CTT CGA CTT TCT TGT GCT 3’	1399
matK1176B	5’ CTA TTC ATT CA[A/C] TAT TTC CTT 3’	488

Locations indicate the positions of the *mat*K gene in rice (*Oryza sativa*) starting from the 5’ region.

To test the adequacy of the number of informative sites obtained from the 966 nucleotides, phylogenetic reconstruction was tested by sequential increments of informative sites starting from the internal end of the sequences and moving downstream the coding region. The phylogeny of the 49 species was first tested with 100 informative sites, followed by increments of 50 informative sites each until the topology and bootstrap values matched the consensus tree obtained from using the whole data set.

5.3 Results

5.3.1 Sequence Comparison

The sequence data from 49 species (including one outgroup) resulted in an aligned sequence length of 966 bp. No indels were found in this section of the *mat*K. The aligned sequences of the 49 taxa showed that among the 966 nucleotides, 495 (51%) were variable and 390 (36%) were phylogenetically informative. The distribution of the number of variable sites along the *mat*K 3' section was calculated with the MEGA program by dividing the sequences into ten sectors. Variable sites were distributed nearly evenly among the 10 sectors.

Transition-transversion ratios ranged from around unity (1.03 between pooids and panicoid species) to 1.45 (within the Bambusoideae and between the centothecoid *Zeugites pittieri* and the Arundinoideae taxa). The mean rs/rv ratio for the entire data set was 1.26. This is similar to ratios obtained for the *mat*K from sequences of the Apiaceae (1.0; Plunkett, Soltis, and Soltis, 1996) and Saxifragaceae (1.3; Johnson and Soltis, 1994). The 1.26 ratio is, however, lower than the 1.79 obtained in our previous study of the Poaceae (Liang and Hilu, 1996). The reasons for the incongruance are disscussed below.

The nucleic acids were translated into amino acids to assess the rates of substitution at the protein level. Out of the total 320 amino acids, 243 (76%) were variable. To explain the high variability found in amino acids for the *mat*K gene, substitutions at the first, second, and their codon positions were computed and found to account for 29%, 28%, and 43%, respectively.

5.3.2 Phylogenetic Analysis

Parsimony analyses of the 390 informative sites in the 49 species provided 6 most parsimonious trees. The strict consensus tree was 791 steps in length, and had 0.65 CI and 0.85 RI values. The major lineages in the tree were supported by high bootstrap values that ranged from 73% to 100%. The Relative Apparent Synapomorphy Analysis (RASA) indicated that highly significant phylogenetic signals exist in this data set. The tRASA value is 23.4, which is almost ten-fold the theoretical value of tRASA = 2.70 estimated at $\alpha = 0.05$ level (Lyons-Weiler, 1997).

To address the adequacy of the number of informative sites, the effect of cumulative addition of informative sites starting at the internal end of the sequences was tested. The study revealed that at 300 sites, tree topology and bootstrap values matched those of the consensus tree based on the entire sequence. The 300 sites represent ca. 700 bp of the 966 sequences used. The Power and Effect test of the RASA was used to determine if signal can be improved by adding more characters, i.e. informative sites. The analysis showed that after 130 informative sites, the signal reaches a plateau (Fig. 5.1).

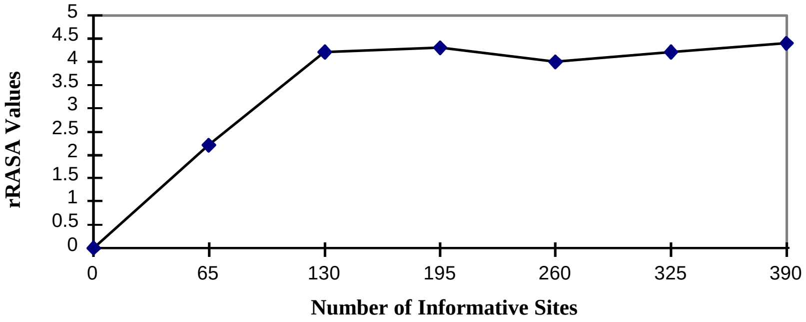

Fig. 5.1. A tRASA Power Plot based on *mat*K sequences from the 49 taxa showing the amount of phylogenetic signal (tRASA) with the accumulative addition of informative sites. The addition starts from the internal end of the data set and progresses toward the 3' end of the coding region.

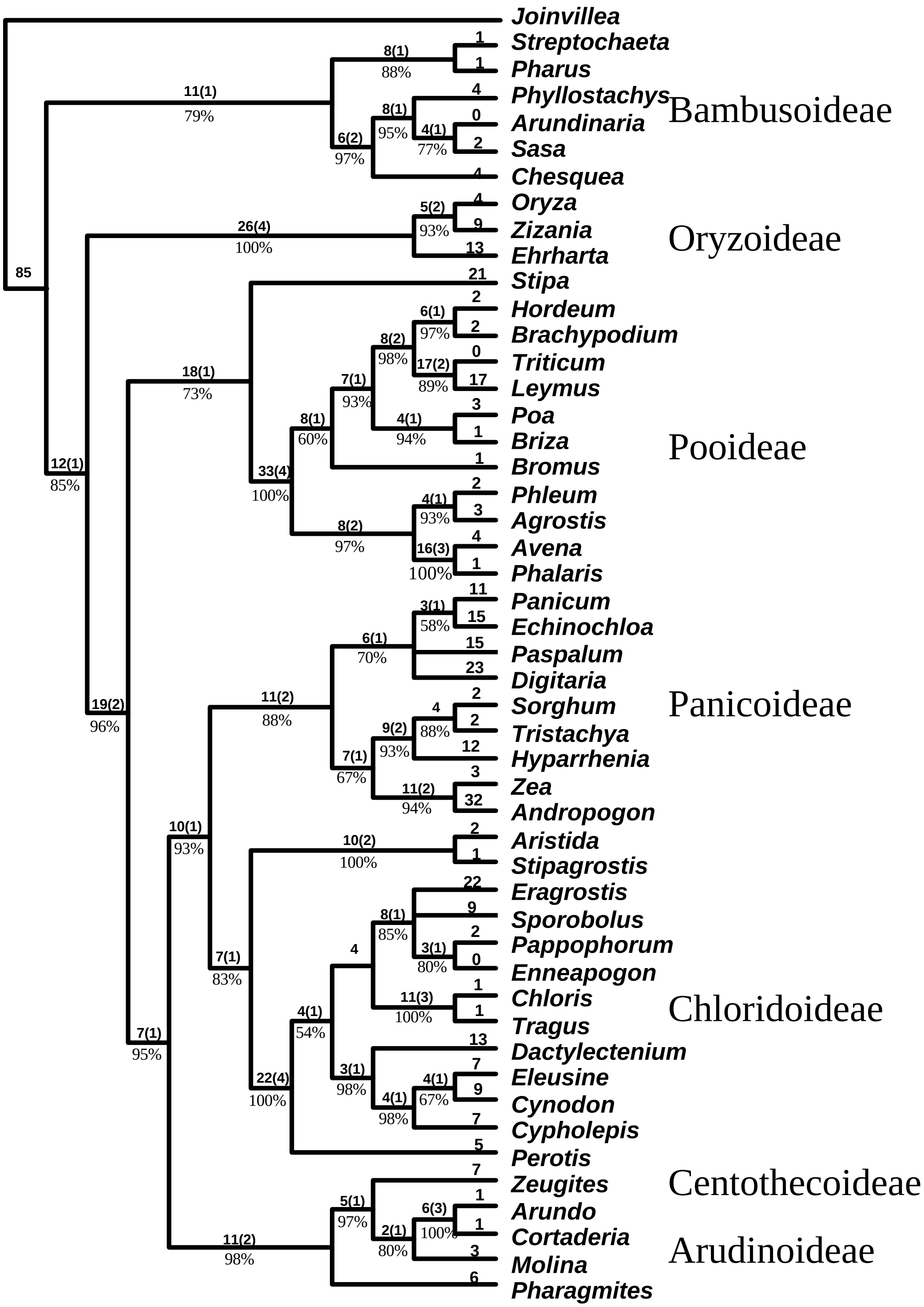

Fig. 5.2. The strict consensus tree of six most parsimonious trees derived from *mat*K sequence analysis for 48 grass species and the outgroup *Joinvillea*. Numbers above each branch indicate base substitutions and decay index (in parentheses). The bootstrapping support is reported below branches as percentages based on 100 bootstrap replications.

The consensus tree (Fig. 5.2) resolved clades that correspond to six recognizable subfamilies: Bambusoideae, Oryzoideae, Pooideae, Chloridoideae, Panicoideae, and Arundinoideae. The Centothecoideae, represented here by *Zeugites pittieri*, appeared within the arundinoid clade. The Bambusoideae, including woody and herbaceous bamboos, diverged as the most basal lineage. This clade is supported by a bootstrapping value of 79%. The oryzoid species formed a monophyletic group supported by 100% bootstrapping and appeared as a sister group to the Bambusoideae. The oryzoid clade included *Ehrharta* as a basal taxon. The Chloridoideae, Panicoideae, Arundinoideae, and Centothecoideae (PACC group) emerged as a monophyletic lineage with 95% bootstrap support. Within the PACC clade, the Arundinoideae was basal and the Chloridoideae was terminal. The two genera of the Aristedeae (*Aristida* and *Stipagrostis*) branched off before the chloridoid clade as a monophyletic line supported by a 100% bootstrap value and 10 mutations (Fig. 5.2). Stipeae (represented by *Stipa*), a tribe of disputable taxonomic position, appeared basal to the Pooideae.

5.4 Discussion

Analyses of the 966 bp sequenced for the 49 species demonstrated again the high rate of substitution in the *mat*K gene. The 495 bp (51%) variation in the sequences, of which 390 (36%) were phylogenetically informative, has contributed a sizable number of characters for resolving the phylogeny of the Poaceae. This data set is considerably larger than the one used in the exploratory *mat*K study of Liang and Hilu (1996) in which 17 species were analyzed with 583 nucleotides, of which 87 were phylogenetically informative. Therefore, the phylogenetic resolution is higher in this study, as is evident in the near lack of polytomy and the strong support for the majority of the clades (Fig. 5.2). The overall mean for the tr/tv was 1.26. This ratio is lower than the 1.79 value obtained by Liang and Hilu (1996) for the 583 bp sequenced from the extreme end of the 3' region for seventeen species of grasses, but is in line with the 1.01 tr/tv ratio calculated by Hilu and Liang (1997) for the *mat*K gene in various plant groups. The relatively larger number of transversion mutations detected in this study appears to be a reflection of differential rates of transversion mutations in different sectors of the gene (Hilu and Liang, in prep).

The present study was based on coding region sequences that extended further upstream of those used in Liang and Hilu (1996).

The presence of 76% amino acid substitutions is remarkably high, pointing to the possibility of relatively low selection constraints on this gene. This type of substitution at the protein level is a characteristic trend for the *mat*K gene (Hilu and Liang 1997). The amount of non-synonymous amino acid substitution in the *mat*K is considerably higher (59%) than that observed in the *rbc*L gene (5%) in the Saxifragaceae (Johnson and Soltis, 1995). The percentages of substitutions in the *mat*K at the first, second, and third codon positions were 29%, 28%, and 43%, respectively. Higher than normal rates of substitution at the first and second codon positions account for the predominance of non-synonymous mutations in this gene since, theoretically, the first, second, and third codon positions are translated to 96%, 100%, and 31% non-synonymous substitutions, respectively (Li and Graur, 1991).

5.4.1 Grass Phylogeny

The consensus tree of 791 steps, CI = 0.649, and RI = O.85 (Fig. 5.2) displayed subfamily clades that are highly supported by bootstrap values of 71-100%. In fact, the mean bootstrap value for the major clades was 91%, indicating very strong support. The 23.4 tRASA value, being almost ten times the theoretic value of tRASA=2.70 estimated at α=0.05 level (Lyons-Weiler, 1997), provides another strong statistical support of the phylogeny constructed from the *mat*K.

Phylogenetic construction for the 49 species with the cumulative addition of nucleotides unequivocally points to the presence of sufficient information for the phylogenetic analysis. The stabilization of the topology and the strong support for the clades at 300 informative sites (bootstrap 47% to 100%) clearly underscore the adequacy of the number of informative characters used in this study. In addition, the Power and Effect test of the RASA program provided yet further statistical support since the curve reached a plateau after using 130 informative sites (Fig. 5.1).

Basal lineages in Poaceae--Rooted with *Joinvillea* (Joinvilleaceae), the consensus tree showed the Bambusoideae as the most basal lineage in grasses. The identity of the most primitive grasses has been debated. Traditional studies and recent phylogenetic analyses disagree on whether or not the Bambusoideae is a basal lineage. The Bambusoideae was traditionally considered the most primitive group of grasses based on the presence of ancestral reproductive characters such as the bracteate, indeterminate inflorescences, "pseudospikelets," and flowers with trimerous parts (Stebbins, 1956; Soderstrom and Calderon, 1979). At the same time, the presence of derived anatomical, vegetative and some reproductive characters led Soderstrom (1981) to state that the Bambusoideae is a specialized group and not to be regarded as the precursor of the other groups of grasses. In a cladistic study of morphological and anatomical characters, Kellogg and Watson (1993) maintained that the Bambusoideae can not be both basal and monophyletic. Phylogenies based on cpDNA restriction sites (Davis and Soreng, 1993) could not resolve the position of the Bambusoideae. Sequence data based on 63 informative sites from the *rps*4 gene (Nadot, Bajon, and Lejeune, 1994) showed a split in the family into two lineages: Pooideae and the rest of the family, including the Bambusoideae. Cummings, King, and Kellogg (1994), using sequence data from 430 bp of the *rpo*C2 gene, presented a phylogeny rooted with spinach and tobacco in which the Arundinoideae was the most basal lineage.

This study also shows that the Bambusoideae are both monophyletic and basal in the Poaceae. The Bambusoideae lineage is supported by a 79% bootstrap value and 11 mutations. Phylogenies based on sequence data from rDNA (Hamby and Zimmer, 1988) and *rbc*L (Doebley et al., 1990; Barker, Linder, and Harley, 1995; Duvall and Morton, 1996) support the basal position of the Bambusoideae.

Clark, Zhang, and Wendel (1995) raised another question related to the identity of the basal taxa in grass phylogeny. Their *ndh*F sequence study showed the herbaceous bamboos *Anomochloa*, *Streptochatae* and *Pharus* forming two most basal lineages. This observation is not supported by the *mat*K data, as the herbaceous and woody Bambusoideae formed a monophyletic group. Within this lineage, the herbaceous genera

Streptochaeta and *Pharus* constituted a single clade (88% bootstrap and 8 mutations) and the four woody genera appeared in a monophyletic lineage (97% bootstrap and 6 mutations). It is, however, worth mentioning that during the construction of phylogenies based on cumulative numbers of informative sites, the two herbaceous bamboo genera appeared in a clade that was distinct and basal. This phylogenetic position for the herbaceous bamboos emerged only when 100 and 150 informative sites were used, but the clade merged with the rest of the Bambusoideae when the number of informative sites was increased. The numbers of most parsimonious trees obtained from the 100 and 150 informative sites were 2535 and 225, respectively, and the consensus trees had very poor resolution at the base, failing to resolve the major lineages and several of the tribes.

Proposed early bifurcation in grass phylogeny--Neither the consensus tree nor the six most parsimonious trees based on the *mat*K gene sequences showed a basal bifurcation in the phylogeny of the Poaceae. Dichotomies of various taxonomic compositions have been reported in the Poaceae. Clark, Zhang, and Wendel (1995) resolved a bifurcation in the family following the herbacious bamboos; The two lineages: 1) Panicoideae, Arundinoideae, Chloridoideae, and Centothecoideae, and 2) Pooideae and Bambusoideae (excluding herbacious taxa and including the oryzoids). This split led them to coin the term BOP for the second group as opposed to the PACC acronym for the other group. The appearance of the subfamilies Chloridoideae, Panicoideae, Arundinoideae, and Centothecoideae in one major group was first demonstrated by Hilu and Wright (1982) on the bases of morphological-anatomical characters. Further support for this grouping was provided by the protein and immunological study of Hilu and Esen (1988) and Esen and Hilu (1989). The monophyly of the group has since been substantiated by molecular information from nucleic acid studies (Davis and Soreng, 1993; Barker, Linder, and Harley, 1995; Clark, Zhang, and Wendel, 1995; Duvall and Morton, 1996; Liang and Hilu, 1996). Mathews and Sharrock (1996) reported a similar bifurcation on the bases of 174 informative sites from combined phytochrome nucleotide sequence data. The Bambusoideae and Oryzoideae were represented by *Bambusa* and *Oryza*, respectively.

A split of grasses into two lineages has also been reported by Davis and Soreng (1993) and Cummings, King, and Kellogg (1994), but the taxonomic composition of the groups varied in all cases. The phylogeny based on cpDNA restriction sites (Davis and Soreng, 1993) depicted two lineages, one corresponded to the Pooideae and the other included all remaining subfamilies. The alliance of the Bambusoideae-Oryzoideae clade with the PACC group in the second major lineage was unstable and the overall topology of the tree changed considerably after the exclusion of one restriction site (Davis and Soreng, 1993; Fig. 2-3). Cummings, King, and Kellogg (1994) reported a phylogeny based on the *rpo*C2 gene, showing two major grass lineages:1) Pooideae- Arundinoideae, 2) Oryzoideae-Panicoideae-Arundinoideae. The *rpo*C2 gene is small (430 bp), and the tree was rooted with spinach and tobacco. The phylogeny showed the oryzoids to be polyphyletic, with the panicoid taxa nested within them, and allocated some arundinoid species to the pooid clades. These evolutionary trends are not supported by current information on grasses.

Systematic status of the oryzoids and Ehrharteae--The Oryzoideae genera *Oryza* and *Zizania* formed a monophyletic lineage that included *Ehrharta* as a basal taxon. The lineage was sister to the Bambusoideae. The subfamilial status of the Oryzoideae has been controversial. The group either has been placed as a distinct subfamily or included within the Bambusoideae (see Hilu and Wright 1982). Analyses of morpho-anatomical data have been inconsistent in terms of proposed taxonomic status of the oryzoids (Hilu and Wright, 1982; Baum, 1987; Campbell and Kellogg, 1987; Kellogg and Watson, 1993). The presence of 10-18 kDa (kilodalton) prolamins emphasized the affinities between the two groups (Hilu and Esen, 1988), but the low immunological cross-reactivity between them clearly demonstrated high divergence (Esen and Hilu, 1989). Not all DNA sequence studies on the Poaceae included representatives of both bambusoid and oryzoid groups (Doebley et al., 1990; Hsiao et al., 1994; Cummings, King, and Kellogg, 1994; Nadot, Bajon, and Lejeune, 1994). Information from 18S and 26S rRNA (Hambey and Zimmer, 1988) and *rbc*L (Duvall and Morton, 1996) showed the oryzoids as a distinct entity and sister to the Bambusoideae.

Two recent studies based on the *ndh*F and *rbc*L (Clark, Zhang, and Wendel, 1995; Barker, Linder, and Harley, 1995) resolved the oryzoids as a monophyletic group in a major clade that included both the oryzoids and bambusoid grasses. Clark, Zhang, and Wendel (1995; Fig. 5.2) reported one of two equally-most parsimonious trees; the bambusoid-oryzoid clade had a zero decay index and was supported by two mutations. The strict consensus tree of the two most parsimonious trees showed the oryzoid, bambusoid, and pooid lineages as an unresolved trichomy. Therefore, the *ndh*F sequence data, although they substantiated the monophyly of the oryzoids and underscored their affinities to the bambusoids, did not provide suffecient reolution for the systematic status of the Oryzoideae. Barker, Linder, and Harley (1995; Fig. 1) presented one of 26 equally-most parsimonious trees, without showing the consensus tree. The tree displayed *Bambusa* nesting between the oryzoids *Oryza* and *Zizania*. This oryzoid-bambusoid line was supported by the bootstrap analysis (51%). The small sample size for the Bambusoideae and Oryzoideae is due to the focus of the study on the Arundinoideae. Therefore, considering the small sample size, the relatively large number of equally-most parsimonious trees they obtained, and the 51% bootstrap value for that clade, a conclusive statement on the systematic status of the oryzoid grasses can not be drawn. Barker, Linder, and Harley (1995) also analyzed their data with the Neighbor Joining method; the phylogeny obtained depicted the bambusoids as a basal taxon in the Poaceae and the oryzoids as a monophyletic lineage sister to the bambusoids.

In this study, the oryzoid clade was strongly supported by a 100% bootstrap, 26 apomorphies and a decay index of 4. This molecular information, thus, strongly supports the treatment of the Oryzoideae as a distinct subfamily.

The taxonomic position of *Ehrharta* (Ehrharteae) has been disputed. The genus has been placed within the Bambusoideae (Renvoize, 1986), Arundinoideae (Ellis, 1987), or Oryzoideae (see Hilu and Wright 1982). In a review of the Ehrharteae, Tateoka (1963) concluded that the tribe could be placed within or near the Oryzoideae or near the "arundigform" grasses. The strong support for the *Ehrharta*-Oryzoideae clade clearly demonstrates the phylogenetic affinity of *Ehrharta* to the oryzoid grasses (Fig. 5.2). This phylogenetic position for *Ehrharta* is in agreement with the *ndh*F-based phylogeny of Clark, Zhang, and Wendel (1995).

Pooideae and Stipeae--The monophyly of the Pooideae was supported by a 73% bootstrap value and 18 synapomorphies. *Stipa* emerges as a basal taxon in this subfamily. The Stipeae has been inconsistently considered as a member of the Bambusoideae, Pooideae, Arundinoideae, or as a subfamily with affiliation to the Bambusoideae (discussed in Barkworth and Everett, 1987). The sister position of *Stipa* to the Pooideae is in agreement with the *rbc*L study of Barker, Linder, and Harley (1995), the *ndh*F results of Clark, Zhang, and Wendel (1995), and the phytochrome study of Mathews and Sharrock (1996).

The remaining eleven pooid genera formed four lineages. The Triticeae clade included *Brachypodium,* an intriguing phylogenetic relationship since the genus has been inconsistently placed in the Triticeae, Bromeae or Brachypodieae (Bor, 1970; Harz, 1980; Hilu and Wright, 1982; MacFfarlane, 1987; Clayton and Renvoize, 1986). The taxonomic uncertainties are due to the presence in *Brachypodium* of floret characteristics that are intermediate between these tribes, its possession of smaller chromosomes than those found in the Triticeae, and basic chromosome numbers of 7, 9 and 10 (see Hilu and Wright, 1982). The Triticeae-Brachypodium clade is supported by a 98% bootstrap and eight apomorphies. Therefore, the *mat*k-based phylogeny supports a subtribal position for *Brachypodium* within the Triticeae (Brachypodiinae) as proposed by Hilu and Wright (1982).

Bromus appeared as a distinct clade basal to the Triticeae and Poeae lineages. The tribal position for *Bromus* is generally recognized, and Clayton (1978) considered it as a link between the Poeae and Triticeae. The remaining four pooid genera formed two lineages:1) the *Phleum* and *Agrostis* clade and 2) the *Avena* and *Phalaris* clade. These two clades represent the traditional tribes Agrostideae and Aveneae senso Hitchcock (1951) but do not reflect Clayton and Renvoizes's (1986) tribal treatments. The sample size in this study is too small to address the systematic questions at the tribal level in the Pooideae.

Structure of the PACC clade--The Arundinoideae are known to be a taxonomically problematic group. Renvoize (1981) and Conert (1987) indicated that the Arundinoideae include genera and tribes that do not fit well in any other grass subfamilies. Ellis (1987) stated that most of the genera lack sharp diagnostic characteristics and are only slightly less similar to each other than to taxa in other grass subfamilies. Kellogg and Campbell (1987) indicated that there is no single character that unites the subfamily. The polyphyletic or paraphyletic natures of the arundinoid grasses have been proposed on the bases of morpho-anatomical characters (Campbell and Kellogg, 1987) and *rbc*L and *ndh*F sequence data (Barker, Linder, and Harley 1995; Clark, Zhang, and Wendel 1995). The Arundinoideae sample in Barker, Linder, and Harley (1995) is one of the largest among these studies. They presented one of 26 most-parsimonious trees showing the arundinoid taxa split between two major clades supported by bootstrap values of 33% and 47%. The large number of most parsimonious trees and the low bootstrap values make it difficult to assess the phylogeny of the arundinoid taxa and the topology of the consensus tree. A more recent study based on the *rbc*L gene (Duvall and Morton, 1996) reflected the monophyly of the arundinoids. However, the study included only two arundinoid taxa (*Arundo* and *Phragmites*).

This study shows the Arundinoideae to be monophyletic and basal in the PACC clade. The monophyly of the arundinoids is supported by a 98% bootstrap, 11 mutations and a decay index of 2 (Fig. 5.2). Although the arundinoid assemblage is shown to be strongly monophyletic, a more detailed study of the subfamily with the *mat*K gene is planned to ascertain this phylogenetic status. *Phragmites* emerged as a basal lineage in the Arundinoideae in this study (97% bootstrap and 5 mutations). Renvoize (1986) and Ellis (1987) excluded *Phragmites* from the core genera of the Arundinoideae on the basis of distinctive anatomical characteristics such as the presence of arm cells. This isolated position was supported by prolamin size and immunological data (Hilu and Esen, 1990).

The Panicoideae appeared monophyletic and a sister group to the Arundinoideae. This is in contrast with the apparent polyphyly of the subfamily reported in our previous study (Liang and Hilu, 1996). The present study included more species and utilized a larger number of sequences (966 vs. 583 bp and 390 vs. 87 informative sites). The polyphyly of the Panicoideae in the previous study was due to the presence of the *Aristida* lineage within the panicoid group; however, the *Aristida* clade was only weakly supported (31% bootstrap and 2 mutations). The support here for the monophyly of the Panicoideae is very convincing (88% bootstrap and 11 mutations) in contrast with weak support for the panicoid clades (25% bootstrap and 3 mutations; Liang and Hilu, 1996).

Tristachya (Arundinelleae) is nested within the Andropogoneae clade of *Sorghum* and *Hyparrhenia* (Fig. 5.2). The degree of support for the inclusion of *Tristachya* within the andropogonoid clade was very high (93% bootstrap and 9 apomorphies). *Tristachya* also appeared nested within the Andropogoneae in the *rbc*L study of Barker, Linder, and Harley (1995). The Arundinelleae shares some spikelet features with the Andropogoneae and has been proposed as an ancestral group to the latter (Clayton, 1981). The affinities of *Tristachya* to the Andropogoneae is supported by the *mat*K study, but the ancestral position was not.

The Chloridoideae emerged as a monophyletic lineage and a sister group to the Aristideae (Fig. 5.2). The monophyletic origin of the Chloridoideae has been disputed. Kellogg and Campbell (1987) indicated that the Chloridoideae may be monophyletic but the evidence is less conclusive than that for the other subfamilies. They suggested that the inflated distal microhair cell defining the subfamily was one of the weakest synapomorphies in their analysis. Jacobs (1987) stated that it would be wise not to discard the possibility of a polyphyletic origin of the Chloridoideae, as is currently accepted. On the other hand, Clayton and Renvoize (1986) believe that the subfamily can be considered as a monophyletic unit whose adoption of efficient C_4 photosynthesis has led to the successful proliferation in the tropics.

The taxonomic position of the Aristideae in the Poaceae is disputable because of unique anatomical features, among other characters. The tribe has been placed either in the Chloridoideae or Arundinoideae (reviewed in Hilu and Wright, 1982). Caroline and Jacobs (see Jacobs, 1987) found different degrees of differentiation of the two Kranz sheaths in *Aristida* species from different ecological habitats, leading Jacobs (1987) to believe that *Aristida* is not very distant from the chloridoid grasses. Based on prolamin polypeptide size and immunological similarities, the tribe appeared intermediate between the Chloridoideae and Arundinoideae (Hilu and Esen, 1990, 1993; Esen and Hilu 1991). The systematic affinities of the Aristideae to the Chloridoideae are supported by the *rbc*L and *ndh*F sequence studies (Barker, Linder, and Harley 1995; Clark, Zhang, and Wendel 1995).

Taxonomic status of the Centothecoideae--The Centothecoideae was segregated from the Arundinoideae as a monotribal subfamily (Clayton, 1978). This taxonomic treatment gains support from the P+PP centostecoid embryo type (Reeder, 1962). Systematic affinities of the Centothecoideae to the Arundinoideae, Panicoideae, and Bambusoideae have been proposed (Renvoize, 1981; Kellogg and Campbell, 1987; Watson, 1990). Based on the *rbc*L study of Barker, Linder, and Harley (1995), the centothecoid *Chasmanthium* appeared within the Arundinoideae in a clade that included *Thysanolaena*, and was separated from the Panicoideae lineage by the arundinoid *Gynerium*. However, the position of *Gynerium* was described as “equivocal”. In the *ndh*F study of Clark, Zhang, and Wendel (1995), the centothecoids *Zugites* and *Chasmanthium* appeared in a clade within the Panicoideae assemblage; the clade encompassed *Danthoniopsis* (Arundinilleae, Panicoideae) and *Thysanolaena*. The *mat*K study reflects the phylogenetic affinities of the Centothecoideae and Arundinoideae and questions the subfamilial status of the Centothecoideae. In contrast, the *ndh*F study points to panicoid affinities of the centothecoid grasses, while the *rbc*L study reflects both arundinoid and panicoid affinities. A more representative sample of the Centothecoideae and the inclusion of *Thysalonaena* are important prerequisites for resolving the taxonomic status and affinities of the centothecoid taxa.

The *mat*K gene has provided sequence information sufficient for good resolution of the major grass lineages. The cumulative addition of informative sites demonstrated the achievement of a stable and robust phylogeny with this data set. A more detailed study of some of the subfamilies with the *mat*K gene is planned, and the accumulating data sets should provide more insight into grass systematics and evolution.

5.5 Literature Cited

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

BARKWORTH, M. E., AND J. EVERETT. 1987. Evolution in the Stipeae: identification and relationships of its monophyletic taxa. *In* T. R. Soderstrom, K. W. Hilu, C. S. Campbell, and M. A. Barkworkth [eds.], Grass systematics and evolution, 251-264. Smithsonian Institution Press, Washington, DC.

BAUM, B. R. 1987. Numerical taxonomic analyses of the Poaceae. *In* T. R. Soderstrom, K. W. Hilu, C. S. Campbell, and M. A. Barkworkth [eds.], Grass systematics and evolution, 334-342. Smithsonian Institution Press, Washington, DC.

BEW, E. A. 1929. The world's grass: their differentiation, distribution, economics and ecology. Longmans, Green and Co., London, New York

BOR, N. L. 1970. Gramineae. *In* K. H. Rechinger [ed.], Flora iranica. No. 70. Akademische Druck M. Verlagsanstalk, Graz.

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

CAMPBELL, C. S., AND E. A . KELLOGG. 1987. Sister group relationships of the Poaceae. *In* T. R. Soderstrom, K. W. Hilu, C. S. Campbell, and M. A. Barkworkth [eds.], Grass systematics and evolution, 217-224. Smithsonian Institution Press, Washington, DC.

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

CLAYTON, W. D. 1978. Gramineae. *In* V. H., Heywood, [eds.], Flowering Plants of the World, 285-290. Mayflower Books, New York

-------- 1981. Evolution and distribution of grasses. *Annals of the Mossouri Botanic Garden* 68: 5-25.

-------- AND S. A. RENVOIZE. 1986. Genera Graminum. HMSO publications, London:

CONERT, H. J. 1987. Current concepts in the systematics of the Arundinoideae. *In* T. R. Soderstrom, K. W. Hilu, C. S. Campbell, and M. A. Barkworkth [eds.], Grass systematics and evolution, 239-250. Smithsonian Institution Press, Washington, DC.

CUMMINGS, M. P., L. M. KING, AND E. A. KELLOGG. 1994. Slipped-strand mispairing in a plastid gene: *rpo*C2 in grasses (Poaceae). *Molecular Biology and Evolution* 11: 1-8.

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

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.

DONOGHUE, M. J., R. G. OLMSTEAD, J. F. SMITH, AND J. D. PALMER. 1992. Phylogenetic relationships of Dipsacales based on *rbc*L sequences. *Annals of the Missouri Botanic Garden* 79: 333-345.

DUVALL, M. R., AND B. R. MORTON. 1996. Molecular phylogenetics of Poaceae: an expanded analysis of *rbc*L sequence data. *Molecular Phylogenetics and Evolution* 5: 352-358.

ELLIS, R. E. 1987. A review of comparative leaf blade anatomy in the systematics of the Poaceae: the past twenty-five years. *In* T. R. Soderstrom, K. W. Hilu, C. S. Campbell, and M. A. Barkworkth [eds.], Grass systematics and evolution, 3-10. Smithsonian Institution Press, Washington, DC.

EMS, S. C., C. W. MORDEN, C. K. DIXON, K. H. WOLFE, C. W. DEPAMPHILIS, AND J. D. PALMER. 1995. Transcription, splicing and editing of plastid RNAs in the nonphotosynthetic plant *Epifagus virginiana*. *Plant Molecular Biology* 29: 721-733.

ESEN, A., AND K. W. HILU. 1989. Immunological affinities among subfamilies of the Poaceae. *American Journal of Botany* 76: 196-203.

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

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

HARZ, C. O. 1980. Beitrage zur systematik der Gramineen. *Linnaea* 43: 1-48.

HILU, K. W., AND A. ESEN. 1988. Prolamin size diversity in the Poaceae. *Biochemical Systematics and Ecology* 16:457-465.

-------, AND --------. 1991. Prolamin and immunological studies in the Poaceae: III. Subfamily Chloridoideae. *American Journal of Botany* 80: 104-113.

-------, AND H. LIANG. 1997. The *mat*K gene: sequence variation and application in plant systematics. *American Journal of Botany* 84: 830-839.

-------, AND K. WRIGHT. 1982. Systematics of Gramineae: A cluster analysis study. *Taxon* 31: 9-36.

HITCHCOCK, A. S. 1951. Manual of the grasses of the United States. Second edition, revised by A. Chase. *United States Department of Agriculture Miscellaneous Publication* 200:1051.

HSIAO, C., N. J. CHATTERTON, K. H. ASAY, AND K. B. JENSEN. 1994. Phylogenetic relationships of 10 grass species: an assessment of phylogenetic utility of the internal transcribed spacer region in nuclear ribosomal DNA in monocots. *Genome* 37: 112-120.

HUBBARD, C. E. 1948. Gramineae. *In*. J. HUTCHINSON [ed.], British Flowering Plants. P. R. Gawthorn Lt., London.

JACOBS, S. W. L. 1987. Systematics of the chloridoid grasses. *In* T. R. Soderstrom, K. W. Hilu, C. S. Campbell, and M. A. Barkworkth [eds.], Grass systematics and evolution, 310-322. Smithsonian. Institution Press, Washington, DC.

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

______. 1995. Phylogenetic inference in Saxifragaceae sensu stricto and Gilia (Polemoniaceae) using *mat*K sequences. *Annals of the Missouri Botanic Garden* 82: 149-175.

KELLOGG, E. A., AND C. S. CAMPBELL. 1987. Phylogenetic analysis of the Gramineae. *In* T. R. Soderstrom, K. W. Hilu, C. S. Campbell, and M. A. Barkworkth [eds.], Grass systematics and evolution, 310-322. Smithsonian. Institution Press, Washington, DC.

-------, AND L. WATSON. 1993. Phylogenetic studies of a large data set. I. Bambusoideae, Andropogonodae, and Pooideae (Gramineae). *Botanical Review* 59: 273-320.

KUMAR, S., K.TAMURA, AND M. NEI. 1993. MEGA: molecular evolutionary genetics analysis. Version 1.01. Pennsylvania State University, University Park, Pennsylvania.

LI, W. -H., AND D. GRAUR. 1991. Fundamentals of molecular evolution. Sinauer, Sunderland, MA.

LIANG, H., AND K. W. HILU. 1996 . Application of the *mat*K gene sequences to grass systematics. *Canadian Journal of Botany* 74: 125-134.

LYONS-WEILER, J. F. 1997. RASA 2.1 Software and documentation for the Mac. http://loco.biology.unr.edu/archives/rasa/rasa.html (updated 9/18/96).

MACFFARLANE, T. D. 1987. Poaceae subfamily Pooideae. *In* T. R. Soderstrom, K. W. Hilu, C. S. Campbell, and M. A. Barkworkth [eds.], Grass systematics and evolution, 265-276. Smithsonian Institution Press, Washington, DC.

MADDISON, W. P., AND D. R. MADDISON. 1992. MacClade: analysis of phylogeny and character evolution. Version 3.0. Sinauer Associate Inc., Sunderland, MA.

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

NADOT, S. R. BAJON, AND B. LEJEUNE. 1994. The chloroplast gene *rps*4 as a tool for the study of Poaceae phylogeny. *Plant Systematics and Evolution* 191: 27-38.

NEUHAUS, H., AND G. LINK. 1987. The chloroplast tRNA^{LYS}(UUU)gene from mustard (*Sinapis alba*) contains a class II intron potentially coding for a maturase-related polypeptide. *Current Genetics* 11: 251-257.

PLUNKETT, G. M., D. E. SOLTIS, AND P. S. SOLTIS. 1996. Evolutionary pattern in Apiaceae: inferences based on *mat*K sequence data. *Systematic Botany* 21: 477-495.

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

REEDER, J. R. 1962. The bambusoid embryo: a reappraisal. *American Journal of Botany* 44: 756-769.

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

---------. 1986. A survey of leaf-blade anatomy in grasses: VIII. Arundinoideae. *Kew Bulletin*. 41: 323-342.

SODERSTROM, T. R. 1981. Some evolutionary trends in the Bambusoideae (Poaceae). *Annals of the Missouri Botanic Garden* 68: 15-47.

--------- AND C. E. CALDERON. 1979. A commentary on the bamboos (Poaceae: Bambusoideae.). *Biotropica* 11: 161-172.

STEBBINS, G. L. 1956. Cytogenetics and evolution of the grass family. *American Journal of Botany* 43: 890-905.

------. 1982. Major trends of evolution in the Poaceae and their possible significance. IN J. R. Estes, R. J. Tyrl, and J. N. Brunken [Eds.]. Grasses and grasslands: systematics and Ecology. 3-36. University of Oklahoma Press, Norman.

SWOFFORD, D. L. 1991. PAUP: phylogenetic analysis using parsimony. Version 3.1.1. Illinois Natural History Survey, Champaign, IL.

TATEOKA, T. 1963. Notes on some grasses. XIII. Relationship between Oryzeae and Ehrharteae, with special reference to leaf anatomy and histology. *Botanical Gazzette* 124: 264-270.

THOMPSON, J. D., HIGGINS, D. G., AND GIBSON, T. J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22: 4673-4680.

TSVELEV, N. N. 1983. Grasses of the Soviet Union. Part I. Translated by B. R. Sharma, Amerind Publ. Co. Pvt. Ltd., New Delhi.

WATSON, L. 1990. The grass family, Poaceae. *In* G. P. Chapman [ed.]. Reproductive versatility in the grasses, 1-31. Cambridge Univ. Press, Cambridge