Chapter 4

Characterization of the matK Gene in the Poaceae

4.1 Introduction

The matK gene is named according to its possible maturase function and its location within the trnK gene encoding the tRNA$_{Lys}$ (UUU). A 509 open reading frame was found to reside in an intron when the trnK gene was sequenced in tobacco (*Nicotiana tabacum*) (Sugita et al., 1985). Later, a complete sequence of the liverwort *Marchantia polymorpha* confirmed the existence of this open reading frame in the non-vascular plant, although it is short compared to those in other plants (Ohyama et al., 1986). Neuhaus and Link (1987) suggested the possible maturase function of the matK gene in mustard (*Sinapis alba*), based upon their homology search result. A segment near the carboxyl terminus of the derived mustard trnK polypeptide is structurally related to the portions of maturase-like polypeptides of introns of the mitochondrial cytochrome c oxidase subunit I gene (COXI) of yeast (*Saccharomyces cerevisiae*) and ascomycete (*Podospora anserina*). The maturase MatK presumably helps fold the intron RNA into the catalytically-active structure. Thereafter, the matK gene was sequenced from various plants, such as rice (*Oryza sativa*), two pine species (*Pinus contorta* and *P. thunbergii*), and beechdrop (*Epifagus virginiana*) (Hiratsuka et al., 1989; Lidholm and Gustafsson, 1991; Wolfe et al., 1992). In beechdrop, this open frame appeared as a free-standing gene with neither the trnK exons nor the interrupting introns present.

Recently, the 3’ end of the matK was identified to contain a conserved region of about 100 amino acids and this region was named domain X (Mohr et al., 1993, Liang and Hilu 1996). Comparison of group II introns (Mohr et al., 1993) found that three domains, reverse
transcriptase, X and Zn-finger-like, existed in the ancestral open reading frame of all group II introns. During the evolutionary process, the reverse transcriptase and Zn-finger-like domains were lost in some cases. The retention of domain X suggests its essential function in RNA splicing.

In plant molecular systematics and evolution, the matK gene is emerging as another valuable gene to study because of its reasonable size, high substitution rate, evenly distributed codon position variation, low transition and transversion ratio, and the easiness of amplification due to its two flanking coding trnK exons. Since its high substitution rate has a potential of providing more informative sites, most application of matK in phylogenetic reconstruction has been at subfamily and tribe levels, such as in Saxifragaceae (Johnson and Soltis, 1994, 1995; Johnson et al., 1996), Polemoniaceae (Steele and Vilgalys, 1994; Johnson and Soltis, 1995), Poaceae (Liang and Hilu, 1996), Orchidaceae tribe Vandeae (Jarrell and Clegg, 1995), Myrtaceae (Gadek et al., 1996), and Apiaceae (Plunkett et al., 1996). The use of the matK gene at higher taxonomic levels appears very promising (Hilu and Liang, 1997). Based upon the study of species representing different major plant groups, the conservative 3’ region of the matK gene was found to contain more phylogenetic information than the high variable 5’ region (Hilu and Liang, 1997).

In order to use the matK gene sequence at various taxonomic levels, more detailed information on its pattern of variation, phylogenetic signal distribution, and transition and transversion ratios will be helpful. This chapter attempts to characterize the matK gene based upon entire gene sequences from across the grass family (Poaceae). The rate, type and pattern of nucleotide substitution of the matK gene in Poaceae were examined. The distribution of transition/transversion ratios along the different parts of the matK gene in Poaceae and their relationships to the taxonomic hierarchy were also studied. Relative Apparent Synapomorphy Analysis (RASA) was used to pinpoint the regions of the phylogenetic signals. The phylogeny of the Poaceae based on the entire coding region sequence data was also examined.
4.2 Materials And Methods

4.2.1 Plant Material and DNA Sequence Methods

The entire matK genes were sequenced from 11 grass genera, representing 7 subfamilies and 11 tribes, and one outgroup (Joinvillea in Joinvillaceae) (Table 4.1). Two additional entire sequences: rice (Oryza sativa) and barley (Hordeum vulgare), from GenBank were included in data analysis. 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). The new primer matK1210R (5’- GTA GTT GAG AAA GAA TCG C -3’) was designed to cover the starting region of the matK gene using an automated sequencer. For this sequencing approach, the trnK region was amplified and electrophoresed on 1.0 agarose gels. The amplified fragment was excised and purified on QiaGen columns (QiaGen Inc.) prior to sequencing. The ABI Prism TM 377 Automated DNA Sequencer with Taq polymerase, DyeDeoxy TM terminator cycle sequencing method was used.

4.2.2 Data analysis

The entire DNA sequences were aligned with the Clustal W 1.6 (Thompson et al., 1994) computer program. 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 et al., 1993) and MacClade 3.0 (Maddison and Maddison, 1992). The data sets were 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 conducted 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 equally most 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 generated in the parsimony analysis.

### Table 4.1 The fourteen entire sequences of *matK*

<table>
<thead>
<tr>
<th>Subfamilies &amp; tribes</th>
<th>Species</th>
<th>Sources of material</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Arundinoideae</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aristideae</td>
<td><em>Aristida adscensionis</em> L.</td>
<td>Hilu, KH5516</td>
</tr>
<tr>
<td>Arundineae</td>
<td><em>Arundo donax</em> L.</td>
<td>Hilu, KH5546</td>
</tr>
<tr>
<td></td>
<td><em>Phragmites australis</em> (Cav.) Trin. ex Steud.</td>
<td>Gary P. Flemming, Assatague Island, VA</td>
</tr>
<tr>
<td><strong>Bambusoideae</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bambuseae</td>
<td><em>Sasa kurilensis</em> Rupe.</td>
<td>Liang, HL9615</td>
</tr>
<tr>
<td><strong>Centotheoideae</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Centothecae</td>
<td><em>Zeugites pittieri</em> Hack.</td>
<td>L. Clark, 1171</td>
</tr>
<tr>
<td><strong>Chloridoideae</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chlorideae</td>
<td><em>Cypholepis yemenica</em> (Schweinf.) Chiov.</td>
<td>USDA, PI364502</td>
</tr>
<tr>
<td><strong>Eragrostideae</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eragrosteeae</td>
<td><em>Eleusine indica</em> (L.) Gaertn.</td>
<td>USDA, PI408801</td>
</tr>
<tr>
<td><strong>Oryzoideae</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oryzoideae</td>
<td><em>Oryza sativa</em> L.*</td>
<td>USDA, PI220025</td>
</tr>
<tr>
<td></td>
<td><em>Zizania aquatica</em> L.</td>
<td>Hilu, KH9423</td>
</tr>
<tr>
<td><strong>Panicoideae</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Andropogoneae</td>
<td><em>Zea deploperenis</em> Itlis, Doble &amp; Guzman</td>
<td>Hugh Itlis, Univ. of Wisconsin</td>
</tr>
<tr>
<td>Paniceae</td>
<td><em>Panicum capillare</em> L.</td>
<td>USDA, PI220025</td>
</tr>
<tr>
<td>Pooideae</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aveneae</td>
<td><em>Avena sativa</em> L.</td>
<td>Hilu, KH9406</td>
</tr>
<tr>
<td>Triticeae</td>
<td><em>Hordeum vulgare</em> L.*</td>
<td>C. S. Campbell</td>
</tr>
<tr>
<td>Joinvelliaceae</td>
<td><em>Joinvella plicata</em> (Hooker f.) Newell &amp; Stone</td>
<td></td>
</tr>
</tbody>
</table>

* Sequence available in GenBank.

To investigate the variation pattern of the tr/tv ratio along the entire gene, the aligned data sets were divided into five sectors, and the total variation, tr/tv ratio and AT
content were counted for each sector. The tr/tv ratio fluctuation along the taxonomic hierarchy was also traced through the following approach. A phylogenetic tree was constructed from all informative sites using the method mentioned above. Then, the program Trace Character in MacClade was used to map each informative site on the phylogenetic tree. Numbers of transitions and transversions were counted for each node of the phylogenetic tree and tr/tv ratios were calculated for each node, as well. The sites with uncertain ancestral states were excluded and only unambiguous sites were used for counting.

The Relative Apparent Synapomorphy Analysis (RASA 2.1; Lyons-Weiler, 1997) program was used to evaluate the phylogenetic signal in the data set. In order to pinpoint the exact location where these major signals come from, the entire data set was split in two ways: two halves in one experiment and five sectors in the other. The RASA analyses were performed on each of those sections. The additive effect of the phylogenetic signal in the data set was conducted by the Power and Effect program in RASA twice, one starting from the 5’ end and the other from the 3’ end of the matK gene.

4.3 Results

4.3.1 Sequence Variation

The length of the matK gene varies from 1616 bp (Eleusine indica) to 1626 bp (Joinvillea plicata). Only two indels were found in the entire data set. A nine base pair insertion was found at the 5’ region of the outgroup Joinvillea plicata and was not shared by the grass species. The other indel was a insertion unique to the 5’ region of wild rice (Zizania aquatica).

The alignment of 14 species and 1632 base pairs yielded a data set of 601 (36.8%) variable sites and 246 (15.1%) informative sites. The informative site data were used in further phylogenetic analyses. In order to study the variation at amino acid level, the aligned sequences were translated into amino acids. Out of the deduced 544 codons, 290 were variable and accounted for 54% of the open reading frame. Fig. 4.1 summarizes the
variation pattern within the *matK* gene at nucleic acid and at amino acid levels. The variation at both levels was evenly distributed throughout the entire gene. The 5’ region appears to have more variation than the 3’ region, which could be related to its function domain. This pattern of variation is similar to that reported in a previous study based on *matK* genes from various plant groups (Hilu and Liang, 1997) and the preliminary study on the *matK* in Poaceae (Liang and Hilu, 1996). The variations at nucleic level and at amino acid level were similar in their patterns (Fig. 4.1). Of the 246 informative sites, the number at the first, second, and third codon positions are 70 (29%), 65 (26%), and 111 (45%), respectively. Unlike *rbcL* and *ndhF* genes, the changes at the third codon position for the *matK* gene are very low as compared to the total (55%) of the first and second position total. This result explains why the variation pattern at the nucleotide level is similar to the one at the amino acid level (Fig. 4.1).

![Fig. 4.1. Variability of the matK gene in the grass family at nucleotide and amino acid levels. The sections start from the 5’ region of the matK.](image)

**4.3.2 The tr/tv Ratio**

The average tr/tv ratio generated from 14 entire *matK* sequences is 1.29, which is similar to those cited in the previous studies (Johnson and Soltis, 1994; Plunkett et al.,
Figure 4.2 shows the tr/tv ratios, the A + T content and the variable sites for each of the five sectors of the *matK* gene. It is intriguing to find that the tr/tv ratios were regionally related. The last sector has the highest ratio (2.13) and the first and third sectors have low ratios (1.14 and 0.94, respectively). The twofold difference between the last and the first sector indicated that the 3’ region has more transition and less transversion than the 5’ region. The ratio tends to become higher toward the 3’ end of the *matK* gene. Correlation analysis with Spearman Correlation Coefficients using the SAS program confirmed that the tr/tv ratios were correlated neither with the A + T content nor with the variable sites (p = 0.90, 0.66, respectively).

![Graph](image)

**Fig. 4.2.** The tr/tv ratio, A + T content and variable sites at different sections of the entire *matK* genes. The tr/tv is scaled at the left Y axis, and the percentage of A + T content and variable sites at the right Y axis.

Using the Trace Characters of the MacClade program, all 246 informative sites of 14 taxa were mapped onto the parsimonious tree generated from PAUP and transition and transversion were count for each node (Fig. 4.3). The result indicated that the tr/tv ratios were not related to the hierarchy of the parsimonious tree of Poaceae, and no particular
trends were observed for the tr/tv ratio at the basal and terminal nodes. Spearman correlation analysis demonstrated that the tv/tr ratio and the taxonomic hierarchy show no significant correlation (correlation = 0.04 and p = 0.85).

![Diagram of tr/tv ratios on the most parsimonious tree.](image)

**Fig 4.3.** Mapping of the tr/tv ratios on the most parsimonious tree. The tr/tv ratios are indicated for each node and the G+C contents are provided following the taxa labels.

### 4.3.3 RASA Test

RASA analysis of the alignment data indicated a relatively high phylogenetic signal in the data set of 14 taxa. The tRASA (6.60) is significantly higher than the theoretic value at $\alpha = 0.05$ level (2.02) (Lyons-Weiler, 1997). In the two half analysis, while the
tRASA of the 5’ half of the *matK* gene is low (0.43) and is not significant at $\alpha = 0.05$ level (2.02), the 3’ of the *matK* gene indicated a significant phylogenetic signal (tRASA = 2.83). Among the 5 sections of the 14 entire *matK* sequences, only the fourth sector contains a statistically significant phylogenetic signal and the other four sectors were not statistically significant (Table 4.2). Interestingly, this sector is the functional domain of the *matK* gene. The first sector is the lowest with a tRASA of -0.57. The other three sectors indicated some signal, but it was not very strong. The fifth sector has a signal which is just below the significant level at $\alpha = 0.1$ (1.11 vs. 1.31). These results support the view that the *matK* gene is phylogenetically valuable and the 3’ region of the *matK* gene, especially the functional fourth sector, contains strong phylogenetic information.

<table>
<thead>
<tr>
<th>Statistical Value</th>
<th>Sector 1</th>
<th>Sector 2</th>
<th>Sector 3</th>
<th>Sector 4</th>
<th>Sector 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed slope</td>
<td>1.51</td>
<td>1.87</td>
<td>1.99</td>
<td>3.69</td>
<td>2.00</td>
</tr>
<tr>
<td>Null slope</td>
<td>1.59</td>
<td>1.68</td>
<td>1.94</td>
<td>2.06</td>
<td>1.76</td>
</tr>
<tr>
<td>tRASA</td>
<td>-0.57</td>
<td>0.92</td>
<td>0.25</td>
<td>5.60</td>
<td>1.11</td>
</tr>
<tr>
<td>Significance</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>***</td>
<td>X</td>
</tr>
</tbody>
</table>

*** Significant at $\alpha=0.05$

The additive phylogenetic signals generated by Power and Effect in RASA starting from both directions are summarized in Fig. 4.4. From the 5’ end downstream to the 3’ end, although the signal rises very quickly, it decreases very rapidly as well. No stable plateau exists, and the phylogenetic signal does not increase constantly with the addition of the characters. However, the tRASA curve from the 3’ end upstream to the 5’ end of
the *matK* gene shows a stable plateau, and strong phylogenetic signal was contained in these regions (Fig. 4.4).

![Graph showing tRASA values for different number of informative sites. The x-axis is the number of the informative sites with the 5' region at left and the 3' region at right.]

**Fig. 4.4** The Power and Effect tests starting from both directions of the entire *matK* gene data set. The x-axis is the number of the informative sites with the 5' region at left and the 3' region at right.

### 4.3.4 Phylogenetic Analysis

A single most parsimonious tree was obtained from the 246 informative sites of the 14 entire *matK* sequences (Fig 4.5). The consistency index (CI) and retention index (RI) of the tree were 0.649 and 0.652, respectively. Seven major groups were well resolved on the most parsimonious tree, and they correspond to the seven commonly recognized subfamilies: Arundinoideae, Bambusoideae, Centothecoideae, Chloridoideae, Panicoideae, Pooideae, and Oryzoideae. The Bambusoid species *Sasa kurilensis* appears at the base of
the tree. The two oryzoid species are grouped together and stand out as a separate lineage. Both oryzoid and pooid clades were supported by the bootstrapping value of 100%. The species of the PACC group (Panicoideae, Arundinoideae, Centothecoideae, and Chloridoideae) appear to be a well resolved single lineage and are supported by the bootstrapping value of 100%.

![Tree Diagram](image)

**Fig. 4.5.** The single most parsimonious tree generated from *matK* sequence analysis for 13 grass genera and the outgroup *Joinvillea*. The bootstrapping support is reported below the branches as percentages based on 100 bootstrap replications. Numbers above each branch indicate base substitutions, and the decay index is cited in parentheses.
The entire *matK* sequence data were also split into two halves and each half was used to generate phylogenies of Poaceae using *Joinvillea* as an outgroup (Fig. 4.6). While the PAUP analysis of the 3’ half data provided only a single most parsimonious tree, five equally most parsimonious trees were obtained when the 5’ half data were used. The tree from the 5’ half shows very low resolution with two polytomies, one comprised the bambusoid, pooid, and PACC group species, while the other represented unresolved subfamilies of PACC group. Bamsoid *Sasa* did not appear as the basal group and was not grouped with the oryzoid species. However, the topology of the most parsimonious tree based on the 3’ half of the *matK* gene was identical to the tree generated based on the entire *matK* sequence data. No polytomy was observed in the tree based on the 3’ half of the *matK* gene. The bootstrapping value and decay index in the 3’ half tree are comparable to those based on the entire data set of 14 taxa.

![Fig. 4.6. Consensus trees generated from the 5’ half (left) and 3’ half (right) of the 13 grass genera and the outgroup *Joinvillea*. Numbers above each branch indicate base substitutions, and the decay index is cited in parentheses. The bootstrapping support is reported below the branches as percentages based on 100 bootstrap replications.](image)
The five consensus trees generated from each of the five sectors further confirmed the RASA test results (Fig. 4.7). All but the fourth sector generated phylogenies with some unresolved polytomies. The fourth sector, which was identified to contain significant phylogenetic signal by RASA (Table 4.2), provided three equally most parsimonious trees and the consensus tree of these three trees showed an almost identical topology to the tree generated from the entire data set. The only unresolved clade is the node joining the Arundinoideae to the other subfamilies of the PACC group. The bootstrapping values are also quite comparable to those based on the entire data set (Fig. 4.7).

4.4 Discussions

4.4.1 The tr/tv Ratio

Transition and transversion and their weighting are very important during phylogenetic reconstruction in plant systematics. Transitions are generally more frequent than transversions, although there are more ways for transversion to take place (8 vs. 4). Furthermore, chemically, spontaneous tranversion is more difficult to achieve because of the mispairing of purine and pyrimidine with different structures. During phylogenetic reconstruction, since transition substitutions occur more frequently than transversions, it is possible that transitions quickly degenerate into noise and should be ignored in some cases (Swofford et al., 1996). Therefore, transversion is usually considered as the more reliable type of mutation in constructing phylogenies (Quicke, 1993). Consequently, some researchers have assigned more weight to transversion in phylogenetic analyses, or based the analyses on transversion alone, resulting in what is called transversion parsimony (Lake, 1987; Quicke, 1993). Therefore, it will be quite helpful to find the underlying causes that derive a substitution toward transition via a transversion. What is the factors that causes the bias of transition and transversion?
Fig. 4.7. The strict consensus trees generated from each of the five sectors of the entire matK gene. Sectors 1 through 5 were based on 8, 5, 5, 3, and 171 equally most parsimonious trees, respectively. Numbers above each branch indicate base substitutions. The bootstrapping support is reported below the branches as percentages based on 100 bootstrap replications.
There have been several speculations regarding this question in the grass family. The neighboring base composition might be a factor, and the tr/tv ratio has been found to be correlated with its regional A + T content (Morton and Clegg, 1995). The underlying mechanism is that the neighboring bases affect the conformation during DNA pairing. The likelihood of transversion was also found to be influenced by the G + C content of a gene, and the tr/tv ratios and G + C content show a ranking order correlation for four different chloroplast genes (Liang and Hilu, 1996). Based upon the results of the 14 entire matK sequence data, it is quite possible that transition and transversion are regionally related in the grass family. Different regions of the matK gene show different tr/tv ratios. It appears that the 3’ functional region of the matK gene has a higher tr/tv ratio, which means more transitions and fewer transversions. The tr/tv ratios were not correlated with the A + T content or the number of variable sites in Poaceae, which was further supported by correlation analysis in the SAS program (Fig. 4.2).

Based on the results of tracing 246 informative sites on the phylogenetic tree, it was found that the tr/tv ratios were not related to the taxonomic hierarchy of Poaceae. No particular trends were observed for the tr/tv ratio at the basal subfamily hierarchy or at the terminal genus branch. If transversion is more conservative than transition, it is expected that more transversions occur at the base of the tree and more transitions at the terminal nodes. The tr/tv ratios should increase from the basal nodes to the terminal nodes. However, this was not the case in our study. The tr/tv ratios are not in an obvious order from basal subfamily level to terminal genus level. Correlation analysis further confirmed that the tv/tr ratios and the taxonomic hierarchy are not correlated. In addition, the variation within the taxonomically-related terminal nodes is very high. Some subfamilies have similar ratios, while others show a wide range of variation (Fig. 4.2). For example, on the one hand, the two oryzoid species, Oryza sativa and Zizania aquatica have almost the same ratio (1.28 and 1.25, respectively). A small difference (0.03) was observed for two arundonoid species, Arundo donax and Phragmites australis. On the other hand, two pooid species, Hordeum vulgare and Avena sativa, have tr/tv ratios of
0.56 and 3.00, respectively. A large difference was also obtained in two chloridoid species, *Eleusine indica* and *Cypholepis yemenica* (1.25 and 5.00, respectively).

### 4.4.2 Phylogenetic Signal in *matK*

The gene is not a consistent unit, and different regions of a gene might evolve at different rates because of their unique functional constraints. During phylogenetic reconstruction in plant systematics, careful selection of a given region of a gene to generate a phylogeny is almost as important as the selection of a gene. It has been demonstrated that the different regions of the *matK* gene provide quite different phylogenies (Hilu and Liang, 1997). This finding was further supported by the results of the entire *matK* gene for the grass family (Poaceae) in this study.

RASA tests and phylogenetic reconstruction of the different regions of the gene indicated the variability within the *matK* gene in the Poaceae. In the two half analyses, the phylogenies based on each half of the *matK* gene were consistent with the result revealed by the RASA tests (Fig. 4.6). The 3’ region of the gene, identified to contain significant phylogenetic information by RASA, provided a single parsimonious tree. The topology of the tree based on the 3’ half tree was identical to that based on the entire gene. The bootstrapping values and decay index are comparable. The only difference was that the 3’ based tree showed fewer substitutions than the one based on the entire gene (Fig. 4.5-4.6). The oryzoid and chlroidoid clades have a 100% bootstrapping value in both trees and the other major clades showed similar values. In contrast, the 5’ region has weak phylogenetic signal revealed by RASA analysis and the phylogenies based upon this half were not stable. Five equally most parsimonious trees were generated when the 5’ half data were used. The strict consensus tree of these five parsimonious trees indicates two unresolved polytomies (Fig. 4.6). The bambusoid species *Sasa kurilensis* occupied different positions on these five parsimonious trees. It either stood out as the basal group of the tree or was grouped with oryzoid or pooid species. Two arundinoid species, *Arundo donax* and *Phragmites australis*, were not always grouped together and were separated by the chloridoid and panicoid lineage in some parsimonious trees. The
centothecoid species *Zeugites pittieri* grouped with panicoid species in some trees and with arundinoid species in others.

In the analysis of the five sectors of the *matK* genes, the Power and Effect analyses of RASA pinpointed the fourth sector as the one having significant phylogenetic signal (Table 4.2; Fig. 4.7). The phylogeny based on this sector has the highest resolution with only one polytomy, and its topology is almost identical to the one based on the entire data set (Fig. 4.7). Other regions did not display high tRASA values at a significant level, and their phylogenies contained many unresolved clades (Table 4.2). Interestingly, the fourth sector is the region which was suggested to be “X domain”, the functional part of the *matK* gene (Mohr et al., 1993, Liang and Hilu 1996). Therefore, it is very promising that the fourth sector of the *matK* provides strong phylogenetic signal, and future expanded studies should focus on this region to resolve the phylogeny of Poaceae at the subfamily or tribe levels.

### 4.4.3 Grass Phylogeny

The 14 entire *matK* sequences provided information to resolve the seven subfamilies that are most commonly recognized in the Poaceae. The basal position of the bambsoid species was further supported by the tree based on the entire gene and the tree based on the 3’ region. The PACC group (Panicoideae, Arundinoideae, Chloridoideae, and Centothecoideae) was a well resolved lineage, even the consensus tree based on the 5’ region of *matK* supports this group with a bootstrapping value of 78%, as well. Arundinoideae was the basal clade for the PACC group, and the centothecoid species *Zeugites pittieri* was grouped with Arundinoideae. Two oryzoid species, *Oryza sativa* and *Zizania aquatica*, are always grouped together and have an unique position from other subfamilies, which supports subfamily treatment as indicated in the previous numerical and immunological studies (Hilu and Wright, 1982; Esen and Hilu, 1989). Further expanded study with more representatives from major groups have a potential of providing more detailed relationships at the subfamily and tribe levels.
4.5 Literature Cited


