The Phylogenetic Reconstruction of the Grass Family (Poaceae) Using *mat*K Gene Sequences

by

Hongping Liang

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in Biology

APPROVED:

Khidir W. Hilu, Chairman

Brent D. Opell

Duncan M. Porter

Saghai Maroof

Brenda W. Shirley

September, 1997
Blacksburg, Virginia
Keywords: Poaceae, Grass, DNA sequence, Phylogeny, Evolution, Taxonomy

The Phylogenetic Reconstruction of the Grass Family (Poaceae) Using *mat*K Gene Sequences

by
Hongping Liang

Committee Chairman: Khidir W. Hilu
Department of Biology

Abstract

Comparative DNA sequencing of *mat*K, a maturase-encoding gene located within the intron of the chloroplast *trn*K gene, was evaluated for phylogenetic utility above the family level and within the grass family (Poaceae). There are three major objectives in the research. The first one is to study the utility of the *mat*K gene in plant evolution. The second objective is to characterize the *mat*K gene in the grass family. The last major goal is to address the phylogenetic questions in the Poaceae using the *mat*K sequences from representatives of different grass groups.

In order to study the potential application of *mat*K to plant systematics above the family level, eleven complete sequences from GenBank representing seed plants and liverworts and nine partial sequences generated for genera representing the monocot families Poaceae, Joinvilleaceae, Cyperaceae, and Smilacaceae were analyzed. The study underscored the following useful properties of the *mat*K gene for phylogenetic reconstruction: reasonable size (1500 bp), high rate of substitution, large proportion of variation at the first and the second codon positions, low transition-transversion ratio, and the presence of mutationally-conserved sectors. The use of different sectors of the gene and the cumulative inclusion of informative sites showed that the 3' region was the most useful in resolving phylogeny, and that the topology and robustness of the tree reached a plateau after the inclusion of 100 informative sites. The presence of a relatively conserved 3' region and the less conserved 5' region provides two sets of characters that can be used at different taxonomic levels from the tribal to the division levels. It also has demonstrated the potential of partial sequencing in resolving systematic relationships from the tribe to the division level.

The *mat*K gene in the Poaceae was characterized with complete sequences from 11 grass genera, representing 7 subfamilies and 11 tribes, and one outgroup (*Joinvillea plicata*, Joinvillaceae). The alignment of 1632 base pairs from 14 species yielded a data set of 601 (36.8)% variable sites and 246 (15.1%) informative sites. The variations at nucleic and amino acid levels evenly distributed throughout the entire gene, and the 5' region appears to have more variation than the 3' region. The changes at the third codon position are very low as compared to the total of the first and second positions. This has led to a similar variation pattern at nucleic and at amino acid levels. The average tr/tv ratio

generated from 14 entire *mat*K sequences is 1.29. It is intriguing to find that the tr/tv ratios were regionally related. RASA analysis of the alignment data indicated a relatively high phylogenetic signal in the data set of 14 taxa. In the two half analyses, while the tRASA of the 5' half of the *mat*K gene (0.43) is not significant, the 3' of the *mat*K gene showed a significant phylogenetic signal. Among the 5 sections of the 14 entire *mat*K sequences, only the fourth sector contains a statistically significant phylogenetic signal. These results indicate that *mat*K is a phylogenetically valuable gene and that the 3' region of the *mat*K gene contains strong phylogenetic information. A single most parsimonious tree was obtained from the 246 informative sites of the 14 entire *mat*K sequences. Seven major groups were well resolved on the most parsimonious tree, corresponding to the seven commonly recognized subfamilies: Aruninoideae, Bambusoideae, Centothecoideae, Chloridoideae, Panicoideae, Pooideae and Oryzoideae.

Approximately 960 base pairs of the *mat*K gene were sequenced from grass species representing 48 genera, 21 tribes, and seven subfamilies to reconstruct a phylogeny for the Poaceae. *Joinvillea plicata* (Joinvilleaceae) was used as an outgroup species. The aligned sequences showed that 495 nucleotides (51%) were variable and 390 (36%) were phylogenetically informative. RASA indicated that very significant phylogenetic signals exist in this data set. The cumulative addition of informative sites starting at the internal end of the sequences revealed that at 300 sites, tree topology and bootstrap values matched those of the consensus tree based on the entire sequence. Parsimony analyses using PAUP resulted in six most parsimonious trees and a strict consensus tree showing major lineages supported by high bootstrap values. These lineages corresponded to six subfamilies: Bambusoideae, Oryzoideae, Pooideae, Chloridoideae, Panicoideae, and Arundinoideae. The Bambusoideae, including woody and herbaceous taxa, diverged as the most basal lineage, and the monophyletic oryzoid species formed a sister group. The Chloridoideae, Panicoideae, Arundinoideae, and the centothecoid *Zeugitis* (PACC group) emerged as a monophyletic assemblage with 95% bootstrap support. The Aristideae branched off as a monophyletic line basal to the chloridoid clade. Stipeae appeared as a sister taxon to the Pooideae. The *mat*K-based phylogeny did not reveal a major dichotomy in the family. The *mat*K gene has provided sequence information sufficient for good resolution of the major grass lineages.

Acknowledgments

The author wishes to recognize several individuals for the contributions and help they have provided toward the completion of this dissertation.

First and foremost, I would like to thank Dr. Khidir W. Hilu for his encouragement, guidance and patience in his role as committee chairman of this study throughout the past years.

Thanks to Dr. Brent D. Opell, Dr. Duncan M. Porter, Dr. Saghai Maroof, and Dr. Brenda W. Shirley for serving on my committee, and for their advice and assistance. Special thanks to Dr. Opell for his help on statistical analyses.

I would like to thank Department of Biology, Graduate Student Assembly, and Sigma Xi for their support of the research.

Thanks to Nigel P. Barker, Christopher Campbell, Lynn Clark, Gary P. Fleming, Thomas Wieboldt, and Weiping Zhang for supplying DNA or plant samples. Seed material for some accessions were kindly provided by the U. S. Department of Agriculture Southern Regional and the ARS Plant Introduction Stations.

Special thanks is extended to William Speer and Weiping Zhang for the friendship we have been sharing in the Lab.

Finally, I thank my parents for their support and encouragement. Special thanks to my wife, Xiaoyun Li and my son, Jingyuan Liang

Table of Contents

List of Figures

List of Tables