ABSTRACT

GUO, YUELONG. Molecular Systematics of *Philadelphus* and Molecular Evolution of *LFY* in the Core Eudicots. (Under the direction of Qiuyun (Jenny) Xiang).

Phylogenetic analysis is a powerful tool for elucidating evolutionary relationships of organisms and genes and for testing taxonomic and evolutionary hypotheses. I conducted phylogenetic analyses of DNA sequences from five gene regions to evaluate the classification scheme of *Philadelphus* and used phylogenetic analyses to provide a framework for examining molecular evolution of the LFY gene in plants. Results from my study suggested that *Philadelphus* is a paraphyletic group with the monotypic genus *Carpenteria* nested within. Three evolutionary distinct clades were identified in this large *Carpenteria*-*Philadelphus* complex, the subgenus *Deutzoides* clade, the genus *Carpenteria* clade, and the remaining Core-*Philadelphus* clade with each supported by a long branch and merited the recognition of a distinct genus. Our results indicated that Subg. *Philadelphus* and Subg. *Deutzioides* are not monophyletic. Subgenera *Gemmatus* and *Macrothyrsus* were nested within Subg. *Philadelphus*, for which the monophyly could not be evaluated due to sampling of only one species from each subgenus. *Philadelphus hirsutus* from Subg. *Duetzioides* is also nested within the large subg. *Philadelphus*, agreeing with evidence from leaf anatomy. Sectional and series classifications within Subg. *Philadelphus* were not supported by the results from phylogenetic analyses. Results from biogeographic analysis using the Statistical Bayes-DIVA method implemented in the S-DIVA and divergence time analysis accounting for phylogenetic uncertainty and molecular rate variation using BEAST suggested an

origin of *Philadelphus* s. l. in southwestern North America in the Oligocene and subsequently dispersals into other areas. Several intercontinental migrations among North America, Asia, and Europe occurred at different times of the later Tertiary to reach a worldwide distribution of the genus. For the molecular evolution study of the *LFY* gene, our results revealed that the evolution of *LFY* was under strong purifying selection due to functional constraint, with the C domain under the strongest selection force and the intervening domain being the most relaxed. Our study also showed that the detection of positive selection using the Branch Site Model was not affected by unbalanced sampling density among clade but is sensitive to alignment ambiguity. Our analyses under different conditions (different sampling density in Asterids and Brassicaceae where more sequences are available, alternative alignment in ambiguous regions, and removal of ambiguous sites) detected positive selection in Fabaceae, where *FLO/LFY* evolved a role of the *KNOX1* gene function in regulating compound leaf development. Under the best alignment, we detected adaptive selection at several sites in Asterales, Brassicaceae, and Fabaceae where gene duplication and/or novel function of *LFY* have been reported.

Molecular Systematics of *Philadelphus* and Molecular Evolution of *LFY* in the Core Eudicots

by
Yuelong Guo

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Master of Science

Plant Biology

Raleigh, North Carolina

2010

APPROVED BY:

____
Dr. Jenny (Qiu-Yun) Xiang
Chair of Advisory Committee

____
Dr. De-Yu Xie

____
Dr. Heike Winter-Sederoff

DEDICATION

To my family

BIOGRAPHY

Yuelong was born in Shenyang, a northeast city of China, where he spent 14 years before he was accepted by the College of Life Sciences, Zhejiang University. He was fascinated by the interesting biological experiments conducted there, and thus got involved in several laboratory experiences, including a field investigation in his first year of college, an Enzyme study during hissophomore year, a plant physiological project in his junior year, and a biochemistry study for his thesis. After he earned his bachelor's degree in 2006, he studied in a systematic and evolutionary biology lab in Zhejiang University, where he developed interests in the analysis of DNA sequences. Yuelong was accepted by North Carolina State University in 2007 for graduate study, and was lucky enough to meet his nice and great advisor Dr. Jenny Xiang. Under the direction of Dr. Xiang, Yuelong accomplished this thesis for his M.S. degree.

ACKNOWLEDGMENTS

I would like to acknowledge a great number of people who have helped me during my Master's study. First of all, I would like to express my gratefulness to my advisor, Dr. Jenny Xiang for her guidance and support on the research, for her patience, understanding, and encouragement, and for all her help in writing the thesis. I am also grateful to my committee members Dr. Heike Sederoff and Dr. Xie for their advice on the research and valuable comments on my thesis.

I am in debt to David Thomas and Xiang Liu, two knowledgeable lab managers, for teaching me many molecular lab techniques, Dr. Norman Douglas for great scholarly discussion and insightful suggestions on the research and data analyses, my lab mates Amanda Saville for the revision of my thesis, Chunmiao Feng for showing various analytic tools and software, Ashley Brooks, and Wade Wall for helpful discussion. I also thank Dr. Alan Weakley for sharing his knowledge on the morphology and geographic distribution of Philadelphus, Dr. Becky Boston and Dr. Sue Vitello for keeping me on the right track in the administrative process, the Department of Plant Biology and Department of Biology for providing the teaching assistantship, and all my fellow graduate students in Plant Biology for professional interactions.

My thanks further extend to JC Raulton Arboretum and Arnold Arboretum for providing fresh leaf material for the study, to Dr. David E Boufford, Dr. Michael

Dosmann, and Kathryn Richardson at the Arnold Arboretum for help in obtaining the leaf material, and to Herbaria at North Carolina State University, University of North Carolina at Chapel Hill, and Harvard University for providing specimens for the study of *Philadelphus*. Finally, I want to thank Fang Liu and Han-yu Cui for instructions on using the university High Performance Computing Cluster facility and for help in solving my computer and computing problems. The High Performance Computing Cluster facility at NCSU has been very helpful to my thesis study by offering an efficient platform for data analysis.

My graduate study would not be possible without my parents who have continuously supported me with love and encouragement.

TABLE OF CONTENTS

LIST OF TABLES IX

LIST OF TABLES X

MOLECULAR PHYLOGENETIC ANALYSIS OF *PHILADELPHUS* (HYDRANGEACEAE) --- NEW INSIGHTS INTO THE INFRAGENERIC CLASSIFICATION AND PHYLOGENETIC AFFINITY OF *CARPENTERIA* 1

Abstract 1

Introduction 2

Materials and Methods 5

Taxon Sampling and Molecular Data 5

Phylogenetic Analysis 7

Divergence Time Dating 9

Biogeographic Analysis 10

Results 12

Characteristics of the Molecular Markers 12

Phylogenetic Analysis 13

Chloroplast DNA Data: 13

Nuclear DNA Data: 15

Divergence Time Dating 17

Ancestral Distribution 18

Discussion 18

Phylogenetic Affinity and Taxonomic Status of Carpenteria 18

Infrageneric Classification Scheme of Hu 20

Biogeographic History of Philadelphus 20

Hybridization in Philadelphus 22

Conclusions ... 24
Acknowledgments ... 25
References ... 26
MOLECULAR EVOLUTION OF FLORICAULA/LEAFY-LIKE GENES IN THE CORE EUDICOTS – DETECTING POSITIVE SELECTION WITH THE BRANCH-SITE MODEL ... 58
Abstract ... 58
Introduction ... 60
Materials and Methods ... 62
Data Acquiring ... 62
Multiple Sequence Alignment ... 62
Reconstructing the Gene Genealogy Using Phylogenetic Analysis ... 63
Comparing Rate and Selection Force of Sequence Evolution ... 64
Pairwise Comparison Among Domains Using MEGA ... 64
Genealogy-Based Comparisons of Substitution Rates and Selection Mode and Force Among Branches and Sites using PAML ... 64
Testing Positive Selection on Specific Branch ... 66
Impact of sample size, sequence alignment, and ambiguity sites ... 68
Results ... 69
Structural Characteristics and Sequence Variation ... 69
Genealogy of LFY/FLO-like Gene of Core Eudicots ... 70
Pairwise Substitution Rate Comparison between Domains ... 70
Genealogy-Based Comparisons among Sites and Branches Using PAML 71
Test for Positive Selection Identified by FRM Using the Modified BSM. ... 73
Effects of Sample Size, Sequence Alignment and Ambiguity Sites ... 74
Discussion ... 75
Evolutionary Dynamics ... 75
Impaction of Sample Size, Sequence Alignment and Ambiguity Sites ... 80

LFY Duplication 80
Acknowledgments.. 81
References.. 82

LIST OF TABLES

Table 1-1. Chassification system, geographic distribution, and key morphological characters of *Philadelphus* from Hu (1954-1956) 34

Table 1-2. Species sampled for this study 39

Table 1-3. Primer information for pcr and sequencing 44

Table 1-4. Variable sites of *rpb2* gene. 45

Table 2-1. Species and sequence sourses included in the study 96

Table 2-2. Results from analysis of adaptive selection of *FLO/LFY*-like gene in the Eudicots using the site model 104

Table 2-3. Results from comparison of models using the likelihood ratio tests..... 105

Table 2-4. Results from analyses of testing positive selection using the modified branch-site model for the Brassicaceae branch, the Asterids branch, and the P+O branch.. 106

Table 2-5. Results from analyses of testing positive selection using the modified branch-site model for the fabaceae branches. 107

Table 2-6. Comparisons of results from analyses with different taxon sampling, alternative sequence alignments, and removal of ambiguity sites 108

Table 2-7. Comparison of the foreground and background translated amino acids from the positive selection sites in Brassicaceae, Asterids, and Fabaceae branch #4 detected from analyses of matrix 1 or matrix 2 at the significant levels 111

LIST OF FIGURES

Figure 1-1. One of the mrbayes trees resulting from bayesian analysis of the combined sequences of three cpdna regions 50

Figure 1-2. One of the mrbayes trees resulting from bayesian analysis of the combined sequences of two nuclear dna regions 52

Figure 1-3. One of the mrbayes trees resulting from bayesian analysis of the combined sequences of all five dna markers 54

Figure 1-4. The tree with the highest posterior probability resulting from beast analysis, showing divergence times from beast and geographic distributions at nodes from s-diva analysis 56

Figure 2-1 Schematic structure of *FLO/LFY*-like gene 112

Figure 2-2. Genealogy of the *FLO/LFY*-like gene in eudicots inferred from bayesian analysis using Mrbayes 3.1.2 program 113

Figure 2-3. Comparation of ds, dn, and dn/ds ratio among N-domain, C-domain and the I-domain. 115

Figure 2-4. Distribution of ds, dn, and dn/ds ratio among N-domain, C-domain, and the I-domains 116

CHAPTER 1

Molecular Phylogenetic Analysis of *Philadelphus* (Hydrangeaceae) --- New Insights into the Infrageneric Classification and Phylogenetic Affinity of *Carpenteria*

Abstract

Philadelphus, in the family Hydrangeaceae, has approximately 70 species distributed disjunctly in Eastern Asia, Eastern and Western North America, Central America and Europe. We conducted phylogenetic and biogeographic analyses for *Philadelphus* using both nuclear and chloroplast DNA markers. The results did not agree with the existing classification scheme. We found that *Philadelphus* is a paraphyletic group with the monotypic genus *Carpenteria* nested within. Three major lineages were identified in this Philadelphus-*Carpenteria* clade with each merited the recognition of a genus. Biogeographic analysis using the Statistical Bayes-DIVA method (implemented in the newly developed S-DIVA) and divergence time dating with the BEAST method resolved the origin of *Philadelphus* s. l. in southwestern North America in the Oligocene. The lineage subsequently spread into Asia and other areas. Several intercontinental migrations from North America to Asia and to Europe occurred at different times of the later Tertiary to reach a worldwide distribution of the genus.

Keywords:

Biogeography, *Carpenteria*, Divergence time, *Philadelphus*, Phylogeny

Introduction

Philadelphus, commonly known as Mock Orange, is a fragrant ornamental genus widely cultivated in gardens around the world. Plants in this genus have various uses including external treatment of eczema and infections, and production of soap and perfumes (Moerman, 1998).

The genus consists of approximately 70 species occurring in the north temperate regions, with its diversity centered in eastern Asia and western North America. The genus was first named by Linnaeus in the early 1700s. The most recent comprehensive treatment of the genus was published by Hu (1954-1956) in her monograph of the genus. Hu recognized approximately 70 species in *Philadelphus* and classified them into 4 subgenra: Subg. *Gemmatus*, Subg. *Euphiladelphus*, Subg. *Marcrothyrus*, and Subg. *Deutzioides*, where subgenera *Euphiladelphus* contained the type species, representing Subg. *Philadelphus* according to the rules of nomenclature (Weakley, 2002). Furthermore, Hu recognized a total of nine sections in the genus, and divided the largest section, Sect. *Stenostigma*, into eight series based on the morphological variation in position of bud, type of inflorescences, form of stigma, etc. (Table 1-1; Hu, 1954-1956). This classification scheme was evaluated by several subsequent studies (e.g. Styer and Stern, 1979;

Frazier, 1999; Weakley, 2008). In a study of leaf anatomy, Styler and Stern found support for subgenus *Deutzioides*, which differs from other species of the genus in having small, non-crystalliferous leaves, with the exception of *P.hirsutus* (Styer and Stern, 1979). Frazier (1999) conducted a taxonomic study of the New Mexican species of *Philadelphus*, and noted that Hu over emphasized some characteristics she observed from a small population of specimens. Frazer found that inflorescence type, position of placenta, and the presence and length of caudate of seed, etc., were not reliable characters for species delimitation due to their variation within populations. He reclassified the 10 species of Hu occurring in New Mexico into only two species. Hickman (1993) and Weakley (2008) also thought that there are fewer species of *Philadelphus* than that recognized in Hu's monograph and believed Hu over emphasized the character of pubescens in her classification.

At present, phylogenetic relationships within *Philadelphus* are poorly understood. Weakley (2002 – MS thesis) conducted a preliminary phylogenetic analysis for the genus using DNA sequences of the internal transcribed spacer (ITS) of the nuclear ribosomal gene for 20 *Philadelphus* species. The result supported Hu's grouping of subgenus *Deutzioides*, but did not support the monophyly of Subg. *Philadelphus*.

Carpenteria, the tree anemone, is a monotypic genus containing only *C. californica*, a rare evergreen shrub native to California. The species has long been recognized as a genus distinct from *Philadelphus* due to mainly its differences in floral merosity and chromosome numbers (Table 1-1; see Discussion for details). A

close relationship of *Carpenteria* to *Philadelphus* has been supported in previous phylogenetic analyses of Cornales and Hydrangeaceae, in which a few representative species of *Philadelphus* were included (Soltis et al. 1995; Xiang et al. 1998a, 2002, Fan and Xiang 2003). In these broad scale phylogenetic studies including all genera of Hydrangeaceae, *Carpenteria* was revealed to be the sister of *Philadelphus* with *Deutzia* sister to them. In the parsimony analysis of ITS sequences conducted by Weakley (2002) the genus appeared as one of the branches in the polytomy of the *Philadelphus* clade without a clear relationship if it was not used as an outgroup. This result raised a possibility that *Carpenteria* may be nested within *Philadelphus*.

Clearly, understanding the phylogenetic relationships within *Philadelphus* and between *Philadelphus* and *Carpenteria* is not only important to the classification of these taxa but also to the understanding of the biogeographic history and morphological evolution of these taxa. The distribution of *Philadelphus* represents a broader pattern of the well-known eastern Asia-eastern North American floristic disjunction (Wen 1999; Donoghue and Smith 2004; Xiang et al., 1998b; 2004. 2005) and has fossils dated back to the mid Tertiary (MacGinitie, 1953; Chaney, 1944). The biogeographic history of *Philadelphous* would contribute to a global understanding of phytogeography in the North Temperate regions. In the present study, I conducted a phylogenetic analysis of Philadelphpus using DNA sequences from five regions from both the chloroplast (*matK*, *rps16*, and *rpoB-trnC*) and nuclear

genomes (ITS, and *RPB2*). The major goal of the study is to evaluate the classification scheme of Hu (1954-56), particularly regarding the subgeneric division in the genus and the series delineation in the largest section of the largest Subg. *Philadelphus* (Sect. *Stenostigma*). Furthermore, we also attempted to test the hypothesis of *Carpenteria* nested in *Philadelphus* that emerged in Weakly's study. Additionally, we estimated divergence times and inferred the biogeographic history based on the phylogeny to gain insights into the timing of early diversification and migration events during the genus evolution.

Materials and Methods

Taxon Sampling and Molecular Data

Plant materials were obtained from plants growing in arboreta, botanical gardens, or from herbarium specimens. We included a total of 28 samples of 21 species of *Philadelphus*, the single species of *Carpenteria*, and two species of *Deutzia* in the study. *Deutzia* was found to be the sister of *Philadelphus* and *Carpenteria* in previous phylogenetic studies of Cornales (Xiang et al. 2002; Xiang et al., in review) and were used as the outgroup in this study (Table 1-1). This sampling represented all 4 subgenera, most sections and series recognized in Hu (1954-56) except sections *Poecilostigma* and *Microphyllus*. Five DNA regions were analyzed, including *matK*, *rpoB-trnC*, and *rps16* from chloroplast DNA (cpDNA) genome, and ITS and *RPB2* from the nuclear genome. Total genomic DNA were extracted from

fresh leaf samples or dried herbarium leaves using the CTAB method of Doyle and Doyle (1987) with modification described in Xiang et al. (1998c) or using the DNeasy Plant Mini kit (Qiagen, Inc., Valencia, California). Primers used for PCR in the study are *matK-1F*, *matK-3R*, *matK-3F*, *matK-1R* from Sang et al.(1997), *rpoB* and *trnC*-R from Shaw et al. (2005), *rpsF* and *rpsR2* from Oxelman et al. (1997), ITS5a from Stanford et al. (2000), ITS4 from White et al. (1990), and *RPB2-F1*, *RPB2-F2*, *RPB2-R2 RPB2-F2* designed for this study (Table 1-3). PCR amplification for the ITS region was performed in a 20 µL system containing 10.6 µL of sterile water, 1.6 µL of 2.5mmol/L dNTP, 2 µL of 10X Mg^{2+} free PCR buffer, 2.4 µL of 25mmol/L $MgCl_2$, 1 µL dimethyl sulfoxide (DMSO), 1 µL 10mg/ml BSA, 0.4 µL of each primer, 0.1 µL of 5U/L Taq DNA polymerase, and 0.5 µL of diluted DNA template (concentration ca. 20 ng/ µL). PCR amplification for all other regions was performed in a 25 µL system containing 15.7 µL of sterile water, 2 µL of 2.5mmol/L dNTP, 2.5 µL of 10X Mg^{2+} free PCR buffer, 3µL of 25mmol/L $MgCl_2$, 0.6 µL of each primer, 0.1 µL of 5U/L Taq DNA polymerase, and 0.5 µL of DNA template (diluted for cpDNA regions, ca. 20ng/ µL; undiluted for *RPB2* region, ca. 400 ng/ µL). For DNA amplification, the first cycle was 5 min at 94^0C for denaturation; followed by 35 cycle of 1min denaturation at 94^0C, 1min annealing at annealing temperature, 2.5 min extension at 72^0C; the last cycle was a 5 min final extension at 72^0C, the sample was then put on hold at 4^0C before taken out. Negative control was included in each set of amplification.

The PCR products were purified using an enzymatic clean-up method: 0.02 mL Exonuclease I (Amersham, Arlington Heights, Illinois), 0.2 mL shrimp alkaline phosphatase (Amersham), 0.25mL Buffer, and 2mL sterile water were added to 25 mL of PCR product. The mixture was incubated first for 15 min at 37^{0}C to degrade primers and dNTPs, and then for 15 min at 80^{0}C to denature the enzymes. Sequencing samples were prepared using cycle sequencing following Fan & Xiang (2001). Products of sequencing reactions were run on an ABI 3730xl sequencer at the Duke DNA Sequencing Facility.

Phylogenetic Analysis

The sequences were edited using CodenCode Aligner (http://www.codoncode.com/) and then aligned first using MUSCLE (Edgar, 2004). The resulting alignment was adjusted by eye manually on SE-AL (http://tree.bio.ed.ac.uk/software/seal/). The best-fit model for each DNA region was determined using ModelTest 3.7 (Posada and Crandall, 1998; Posada and Buckley, 2004). Phylogenetic analysis was conducted using the Maximum-likelihood and Bayesian methods implemented in GARLI v0.951 (Zwickl, 2006; www.bio.utexas.edu/faculty/antisense/garli/Garli.html) and MrBayes 3.1.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003), respectively. Gaps were treated as missing data. The Maximum-likelihood (ML) analysis was conducted for each individual molecular region, for the combined three cpDNA regions, the combined two nrDNA regions, and for the combined five DNA regions,

respectively. The analyses were performed using the best-fit model for each region, K81uf+G for *matK* and *rps16*, TIM+G for rpoB, GTR+I+G for ITS, HKY+I for *RPB2* according to the Modeltest result, with all parameters set to “estimated”. Each analysis started with a random seed and terminated if no new significantly better scoring topology was found in 10000 generation. The bootstrap analysis using the ML method was also performed using GARLI for 1000 replicates. The analyses for single molecular markers were conducted first. Although a couple of strongly supported terminal nodes were found to differ between cpDNA and nuclear DNA trees, we went ahead to combined the data for a “total evidence” analysis to see if support of basal nodes increased.

The Bayesian analyses were conducted for the combined dataset of the three cpDNA regions, of the combined data of the two nrDNA regions, and the combined data of all five regions. The following models were determined as the best-fit according to the Akaike information criterion (AIC): K81uf+G for *matK* and *rps16*, TIM+G for rpoB, GTR+I+G for ITS, HKY+I for *RPB2*. The Bayesian analyses for the combined data were conducted by partitioning the gene regions and specifying the best-fit model for each region. Two parallel runs with two different random seeds were conducted at the same time for 10 million generations to make sure the difference was less than 0.05, i.e. MrBayes was not trapped in a local optimum. Trees were sampled every 500 generations. Burn-in was set to one million generations, and was checked in Tracer 1.5 (Rambaut and Drummond, 2007).

Divergence Time Dating

The program BEAST 1.5.3 (Drummond and Rambaut, 2007; Drummond et al., 2006; Drummond et al., 2002) was used to date the divergence time using the molecular data. Fossil evidence of *Philadelphus* is scanty. Several leaf fossils were reported for the genus. MacGinitie (1953) described a leaf fossil *P. minutes* from the Oligocene (ca. 30 million years ago, Mya) of Colorado, and noted that the leaf fossil was similar to the modern species *P.microphyllus* (MacGinitie, 1953). Chaney et al. (1944) reported another *P. nevadensis* from the latest Miocene or early Pliocene of California that showed close affinity to the modern *P. lewisii* (Chaney et al. 1944). All of these relevant modern species were not sampled in the study due to lack of materials. Although the *matK* sequence of *P. lewisii* is available from previous study and a MrBayes analysis of the *matK* sequences of our data including *P. lewisii* suggested a close relationship of *P. lewisii* to *P.coronarius*, *P. kansuensis*, and *P. inodurus* clade (trees not shown), we did not use the Pliocene fossil as a calibration point because previous studies have shown that calibration using recent fossils resulted in shallower divergence times for deep nodes (e.g., Xiang et al. 2008). Because it was impossible to place the other two fossil leaves onto the existing phylogeny with any certainty, they were not useful for nodal calibration. We therefore applied indirect calibration points in the divergence time analysis. Based on divergence time dating for the Cornales with multiple fossil calibrations, the split between *Carpenteria* and the core *Philadelphus* was ~22 million years ago, and the

split between *Carpenteria-Philadelphus* and *Deutzia* was ~40 million years ago (Xiang et al. in review). Thus, in our analysis, the prior for the former node (split between *Deutzia* and *Philadelphus*) was set to follow a normal distribution with a mean of 40 Mya and a standard division of 8.6, according to the 95% credibility interval of the indirect calibration points (Xiang et al., in review). The prior for the split between *Philadelphus* and *Carpenteria* was set as a normal distribution with a mean of 22 Mya and a standard division of 6.1. The analysis was run for 10 million generations, with a burn-in of 1 million generations. Trees were sampled every 1000 generations. BEAST was first run using an empty alignment, or without data, to determine the effect of the calibration priors on the divergence time estimates, i.e. whether the interactions of priors resulted in major change of the priors. As the results showed no significant changes in divergence time at the calibrated nodes in the empty run, a complete run of the data was then conducted using these priors to estimate the divergence time of each branch. The program Tracer (Rambaut and Drummond, 2007) was used to determine the burn in (10%) when the analysis reached stationarity. All remaining 9000 trees were annotated by Tree annotator, and a best tree with the nodal divergence times was visualized using FigTree (http://beast.bio.ed.ac.uk/FigTree).

Biogeographic Analysis

To gain insights into the geographic origin and early dispersal of the genus, we reconstructed the ancestral distributions using the Statistical Dispersal-Vicariance

Analysis (Bayes-DIVA) recently developed from the DIVA method (Nylander et al., 2008; Harris and Xiang, 2009) implemented in the newly developed software S-DIVA (Yu et al., in press; http://mnh.scu.edu.cn/s-diva/),). The tree file containing 9000 trees resulting from the BEAST analysis was used as the input file for the analysis. Each species was coded for its presence or absence of the five geographic areas covering the distributional range of the ingroup and outgroup species: A) estern Asia; B) eastern North America (east of the Great Plain); C) western North America (West of the Great Plain, extending to Mexico); D) Central America; and E) Europe. Because eastern Asia is a large area, we further divided the region into two subregions, the western China and Himalyas (Awh) and the central and eastern China and northeastern Asia (Ace). The missing species were mostly from the same geographic areas of other species of the same section or series that were sampled in the phylogeny (In Hu's classification, species were grouped based on geographic distributions). The exception is Subg. *Gemmatus* Sect. *Poecilostigma*, occurring in Mexico to Panama, for which we did not have materials for the study. Thus, the geographic distributions of the missing species are not expected to impose major errors in developing a preliminary hypothesis regarding the early biogeographic history of the genus based on the results. Distribution of the outgroup has a great impact on the optimization using DIVA (see Xiang et al. 2006; Xiang and Thomas 2008). Due to the uncertainty of ancestral distribution range of *Deutzia*, we took into account the distribution of the entire genus in the analysis. The genus currently

occurs in eastern Asia, Europe, and Central America. S-DIVA estimated the probabilities of distributional ranges for each node based on the results of optimizations on all tree topologies using the parsimony method (Ronquist, 1997; 2001); thus the method takes into account phylogenetic uncertainty.

Results

Characteristics of the Molecular Markers

The lengths of each molecular marker were similar for all species. The length of *matK*, *rpoB-trnC*, *rps16*, ITS, *RPB2* was 1746bp, 1240bp, 877bp, 642bp, 795bp, respectively. For *matK* region, three indels were observed in *Deutzia* species, and were 6-bp, 3-bp, and 6-bp respectively. For *rpoB-trnC* region, a 178-bp indel was found in *Deutzia* species. In addition, a 7-bp and an 8-bp indel were found in *Deutzia* species, and a 7-bp indel was found in *P. coronarius*, *P. pubescens*, and *P. inodorus*. Furthermore, a 4-bp indel was found in *Deutzia*, *Carpentaria*, and *P. mearnsii.* For the *rps16* region, PCR failed for a few species. Among the available sequence data, several small indels of 1-4 bp were observed. For the ITS region, a 17-bp region was found to be difficult to align among species and was excluded from our matrix in the final analysis. For *RPB2* region, PCR for *Deutzia* species was unsuccessful probably due to primer divergence; the primer was designed based on a *Philadelphus* sequence. Several indels were found in *Carpenteria* species. A few species of *Philadelphus* are polymorphic at several sites of the *RPB2* gene, which

combined nucleotides of two other closely related species (Table 1-4). These are *P. kansuensis* combining sequences of *P. pubescens* and *P. inodorus2*, *P. coronarius* combing sequences of *P. pubescens* and *P. inodorus*, *P. triflorus2* combining sequences of *P. pubescens* and *P. inodorus*, and *P. pekinensis2* combining sequences of *P. coronarius* and *P. tenuifolius*.

The matrix combining all regions contains 5300bp, 28 accessions representing 21 *Philadelphus* species from the four subgenera of Hu's classification, one *Carpenteria* species and two *Deutzia* species. Of all the molecular markers, ITS yields the highest percentage of parsimony informative characters (16.20%), followed by *RPB2* (8.43%), *MatK* (5.50%), and *rpoB-trnC* (4.19%), and *rps16* as the least informative marker (1.14%).

Phylogenetic Analysis

Chloroplast DNA Data:

Phylogenetic analysis of individual chloroplast regions did not resolve much of the relationships within *Philadelphus* except those described below. Analysis of the *matK* sequences recognized a sister relationship between *P.texensis* and *P. mearnsii* (100% PP), which formed a trichotomy with *Carpenteria* species and the rest of the *Philadelphus* species (tree not shown). In addition, sister relationships between P*. coulteri* and *P.cordifolius*, between *P. pubescens* and *P. inodorus*, and between *P. kansuensis* and *P. inodorus2* were also recognized in the *matK* tree with high support (100%PP, 99%PP, and 100%PP respectively). Analysis of the *rpoB-*

trnC region similarly identified the *P. texensis-P.mearnsii* clade but included *Carpenteria* (77% PP) in the clade. This clade is sister to the rest of the *Philadelphus* species (100% PP). In the *rpoB-trnC* tree, a close relationship between *P. delavayi* and *P. purpurascens* was also identified (100%PP) (trees not shown). Analysis of the *rps16* sequences suggested that the European species *P. cancasicus* as the sister group to the rest of *Philadelphus* species (trees not shown). In this dataset, several species, including *P. texensis* and *P. mearnsii*, are missing.

Phylogenetic analysis of the combined data of the three chloroplast regions suggested similar phylogenetic backbones as that revealed by individual dataset (Fig. 1-1). *Carpenteria* was found to be nested within *Philadelphus* with strong support (100% PP). Within the *Philadelphus* clade, three major lineages were recognized: *Carpenteria*, *P. mearnsii-P. texensis*, and the core *Philadelphus* containing the remainder of the genus. *Carpenteria* was weakly supported (53% PP) to be the sister of the strongly supported *P. texensis-P.mearnsii* clade (100%PP) from Subg. *Deutzioides*. Within the core *Philadelphus* clade, *P. coulteri* from Subg. *Gemmatus* is closely related to *P. cordifolius* from Subg. *Marcrothyrsus* and *P. delavayi* from Subg. *Philadelphus* Sect. Stenostigma Ser. *Delavayani* is closely related to *P. purpurascens1 from series Pubescentes* of the same section and same subgenus, both with high support (100% and 100% respectively) (Fig.1-1). Several species of Sect. *Stenostigma* including *P. kansuensis* from *Ser. Sericanthi*, *P. coronarius* from *Ser. Coronarii*, and *P. pubescens* from *Ser.*

Pubescentes form a strongly monophyletic group with *P. inodorus* from Sect. *Pauciflorus* (100%).

Nuclear DNA Data:

Phylogenetic analyses of the nuclear data resolved more of the species relationships. The analysis of the ITS sequences recovered the same three major clades as those found by the cpDNA data, but resolved the relationships among them - with the *P. texensis-P.mearnsii* clade diverging out first, and *Carpenteria* sister to the core *Philadelphus* clade (tree not shown). The ITS tree further showed a close relationship 1) between *P. schrenkii* from Subg. *Philadelphus* Sect. *Stenostigma* Ser. *Sericanthi* and *P. hirsutus* from Subg. *Deutzioides*, 2) between *P. purpurascens* from Ser. *Delavayani* and *P. subcanus* from Ser. *Sericanthi* in Sect. *Stenostigma*, 3) among several species of Subg. *Philadelphus* Sect. *Stenostigma* (*P. coronarius* – Ser. *Coronarii*, *P. confuses* – Ser. *Gordoniani*, *P. tomentosus* – Ser. *Tomentosi*, *P. sericanthus* – Ser. *Sericonthi*, and *P. pubescens* – Ser. *Pubescentes*) plus *P. inodorus* from Sect. *Pauciflorus* and *P. cordifolius* from Subg. *Cordifolius* (100% PP, 99% PP, and 98% PP respectively). Analysis of the *RPB2* sequences confirmed ITS result on the placement of the three major clades identified by the cpDNA data, and further resolved a monophyletic group consisting of *P. kansuensus, P. pubescens, P. coronarius, and P. tomentosus*.

The results from analysis of the combined dataset of the two nuclear regions are similar to those supported by the ITS data, and strongly supported (100% PP) a sister relationship of *Carpenteria* to the core *Philadelphus* clade (Fig. 1-2). The results also supported that *P. schrenkii* is closely related to *P.hirsutus* (94%PP), a sister relationship between P. *kansuensis* and *P. pubescens* (92%) and between *P.purpurascens*1 and *P. subcanus* (90%).

A couple of well-supported conflicting relationships were observed between the chloroplast and nuclear DNA trees. In the cpDNA tree, *P. pubescens* is sister to *P. inodorus*, *P. purpurascens1* is sister to *P. delavayi*, whereas the former is sister to *P. kansuensis* and the latter sister to *P. subcanus* in the nuclear DNA tree (Figs. 1,2).

Analysis of the combined total data using the Bayesian method recovered the three major clades and the sister relationship between *Carpenteria* and the core *Philadelphus* clade (Fig. 1-3). The sister relationships between *P. pubescens* and P. *kansuensis* suggested by the nuclear data and between *P. purpurascens* and *P. delavayi* suggested by the cpDNA data were also recovered (Fig. 1-3). The branches at the base of the tree are much longer than the upper branches (all within the core *Philadelphus* clade). The internal branches in this clade are extremely short, resulting in little resolution at the base of the clade. The Bayesian consensus tree is congruent with that from the maximum likelihood analysis, with only slight difference on nodal support. Noticeably, the result shows that *Philadelphus* forms a paraphyletic group. The genus *Carpenteria* is nested within genus *Philadelphus* with

high support (100% PP, 62% bootstrap). Species of *P. texensis* and *P. mearnsii*, previously classified in subgenus *Deutzioides*, formed sisters and branched off first from the remainder of the *Carpenteria*-*Philadelphus* clade (100%PP, 62% bootstrap). The third species from *Deutzioides,* P*. hirsutus* occurring in southeastern North America, was grouped with a northeastern Asian species *P. schrenkii* (99%PP, 66% bootstrap) in the core-*Philadelphus* clade. Within the core *Philadelphus* clade, species *P. kansuensis*, *P. pubescens*, *P. coronarius*, *P.inodorus*, *P. inodorus2* formed a strongly supported subclade (99%PP, <50% bootstrap).

Divergence Time Dating

Divergence time analysis using BEAST resulted in estimates for all the nodes recognized in the tree with the highest posterior probability, but 95% credibility interval for only the nodes with >50% PP. The results showed that the *Philadelphus* split from the *Carpenteria*-*Philadelphus* clade during the Oligocene (~36 Mya; Fig. 1-4). The split of *Carpenteria* from the core *Philadelphus* occurred shortly after, around 35 Mya. The radiation of the core *Philadelphus* clade occurred around the middle Miocene (~ 13.32 Mya) after the European species *P.coronarius* had diverged in the early Miocene (~20.24 Mya; Fig. 1-4).

Ancestral Distribution

Results from S-DIVA showed that the ancestral distribution of the *Philadelphus-Carpenteria* clade was SW North America where it diverged into two lineages, *Carpenteria* and the ancestor of the core *Philadelphus* clade, which later spread into Asia and Europe. The ancestral distribution of the core *Philadelphus* crown clade was suggested to be most likely widespread in northeastern Asia, Europe and western North America (AceCE; P = 32.34%) (Fig. 1-4), or in Europe and western North America (CE; P= 26.91%). The ancestral distribution for the crown group of the core *Philadelphus* excluding the Caucasian species was shown to be most likely in eastern Asia and western North America. Dispersals from eastern Asia to eastern North America, from eastern China to western China and Himalayas, as well as exchanges between western China-Himalayas, Europe and eastern North America were implied based on S-DIVA results.

Discussion

Phylogenetic Affinity and Taxonomic Status of *Carpenteria*

A close relationship between *Philadelphus* and *Carpenteria* was recognized by many authors and supported by both morphological and molecular data (Andrea and Hufford, 1995; Hufford, 1995 and 1997; Hufford et al. 2001; Xiang et al. 2002, in review). *Carpenteria* was recognized as the sister of *Philadelphus* in previous phylogenetic analyses of Hydrangeaceae and Cornales that included one to a few

species of *Philadelphus* (Hufford, 1995 and 1997; Hufford et al. 2001; Soltis et al. 1995; Xiang et al. 2002, in review). The close affinity of *Carpenteria* to *Philadelphus* has not been debated. However, the finding of a position of the genus within *Philadelphus*, more derived than the subgenus *Deutzoides*, is new, although not surprising given the high morphological similarity between the two genera. In previous phylogenetic analyses of Hydrangeaceae and Cornales, although Subg. *Deutzoides* was sampled, the subgenus was represented by *P.hirsutus*. This species was revealed to be a member in the core *Philadelphus* clade (likely in Subg. *Philadelphus*, sister to an eastern Asian species) in our study; thus the relationship of *Carpenteria* to Subg. *Deutzoides* found in this study was missed in previous studies. *Carpenteria* is morphologically distinct from the *Philadelphus* clade largely by having numerous stamens, fused styles with distinct stigmas, a superior ovary, and the chromosome number, 2n=20 (Torrey, 1851). The *Deutzoides* clade is characterized by having few numbers of stamens, solitary flowers, inferior ovary, and ecaudate seeds, different from the other *Philadelphus* species. The clade is also distinct in wood anatomy and leaf morphology (Styer and Stern, 1979). Considering the substantial morphological differences, that there are long branches supporting each of these three lineages, and that *Carpenteria* has long been recognized as a genus, we propose to retain the generic status of *Carpenteria* and raise subgenus *Deutzoides* to the genus status (named as *Neophiladelphus*).

Infrageneric Classification Scheme of Hu

The results from phylogenetic analyses of the five DNA regions found no support for the monophyly of the following groups recognized in Hu's classification of the genus (Hu 1954-1956): Subg. *Deutzoides*, Subg. *Philadelphus*, sect. *Stenostigma* of Subg. *Philadelphus*, ser. *Gordoniani*, ser. *Coronarii*, ser. *Delavayani*, ser. *Pubescentes*, and ser. *Sericanthi* of sect. *Stenostigma.* The phylogenetic pattern revealed from our study support the separation of Subg. *Deutzoides* from the remainder of *Philadelphus* (Figs. 1-4). These results are different from the hypotheses of Hu (1954-56) which suggested that subgenus *Gemmatus* and section *Pauciflorus* were the most primitive in the genus and subgenus *Deutzoides* was derived from them (Hu, 1954-1956). Relationships among species in the core *Philadelphus* clade largely resolved but with low support except for the a few nodes noted above. The non-monophyly of Subg. *Deutzoides* found in this study is consistent with evidence from anatomy and leaf morphology (Styer and Stern, 1979), which similarly suggested a distance relationship of *P. hirsutus* to other species of the subgenus. Additional data and sampling of species are needed to clarity the species relationships in order to derive an infrageneric classification scheme for *Philadelphus* s. s.

Biogeographic History of *Philadelphus*

Weakley (2002) proposed that *Philadelphus* originated in southern North America, and later dispersed into Central America, Europe and Asia. This

hypothesis is congruent with the results of our S-DIVA analysis that accounted for phylogenetic uncertainty (Fig. 1-4). However, our results do not permit the evaluation of migration from southern North America to Central America because species from Central America was not included.

Our results further suggested that the origin and early divergence of the *Carpenteria*-*Philadelphus* clade occurred in the Oligocene likely during an early episode of the earth climatic cooling period (Berggren and Prothero 1992; Potts and Behrensmeyer 1992; Wolfe 1992). There were likely multiple intercontinental migrations between North American and Eurasia occurred at different geological times during the Miocene. Our S-DIVA analysis also suggested a migration from the Himalayan region to eastern North America and from eastern North America back to Europe and Himalayas in the mid Miocene, from eastern North America to western North America, and from east-central China to western China in the late Miocene. These inferred migration events, as well as the initial spread from western North America to Asia and Europe during the Oligocene to early Miocene, are more likely to have occurred via the Beringia Land Bridge because the North Atlantic Land Bridge was supposed to be broken in the early Eocene (Tiffney 1985; Tiffney and Manchester 2001). However, these inferred migration events in the mid Miocene during the diversification of the core *Philadelphus* clade need to tested with a more robust and complete phylogeny of the genus that requires additional data and species sampling.

In our present study, two intercontinental disjunct species pairs, *P. pubescens- P. kansuensis* and *P. hirsutus-P. schrenkii*, from eastern North America and eastern Asia were identified with strong support (Fig. 1-3). The disjunction events were estimated to be around the Pliocene of the late Tertiary, with an ancestor of the species pair in western China/Himalayas and eastern China, respectively. This suggested that either migration across the Beringia Land Bridge (broken around 3.5 Mya; Tiffney and Manchester 2001) or long distance dispersal played a role in the formation of the disjunction. Out of Asia migration has been found to be a dominant pattern in the Northern Hemisphere phytogeography (Donoghue and Smith 2004). The biogeographic history inferred from the present study suggested that the history of *Philadelphus* is probably complex, involving migrations out of America and out of Asia.

Hybridization in *Philadelphus*

A few accessions were found to carry two sequence alleles of *RPB2* each characterizing a different species (Table1- 4), suggesting a possible hybrid origin of the respective species. These putative hybrid species include *P. kansuensis, P. coronarius, P. tomentosus, and P. pekinensis2*. The *RBP2* sequence information and phylogenetic relationships suggested that *P. pubescens* and *P. inodorus* were probably the parental species or closely related to the unsampled parental species of *P. kansuensis*. This hybridization event must have occurred in the ancient time as

the putative parents are in North America while the hybrid species is in western China, unless the hybridization occurred in the garden. We consider the garden origin is unlikely. First, the leaf materials for these species although collected from plants growing in the Arnold Aboretum, the origin of which were either from the wild or from cuttings of plants grown from the wild materials in other gardens (Table 1-2). Second, the terminal branches of the hybrid and parental species are relatively long, showing more genetic divergence than expected from recent garden hybridization (the garden hybrid is expected to have identical genomes of the parents). The biogeographic history (Fig. 1-4) suggested that the ancestor of the *P. pubescens - P. confuses - P. inodorus* was most likely in eastern North America and later spread into western North America, then to Asia and Europe along with the divergence of these species. This permitted hybridization of these species in North America to give rise to the ancestor of *P. kansuensis* that was later spread into Asia. The hybrid species could be either a diploid or a polyploid.

Other putative hybrid species suggested from the sequence data include *P. coronarius, P. tomentosus, and P. pekinensis2.* Their putative parental species were less certain given the poor support of species relationships, but all probably involved *P. pubescens* or its close relatives (Table 1-4). Hybridization in *Philadelphus* was reported in an early survey of chromosome numbers (Sax, 1931). The basic chromosome number of *Philadelphus* was found to be 13 and no irregularity was observed by Bangham (1929) and Sax (1931). Furthermore, Sax noted that the

chromosome morphology were so similar among species that they could be completely paired between species. He also pointed out that different species were easily hybridized in gardens (Sax, 1931; Janaki Ammal, 1958). The rapid radiation of the core *Philadelphus* clade in the mid Miocene might have a link to some hybrid speciation in Asia (most putative hybrid species identified in this study are from Asia), or in other geographic regions, as well, if putative hybrid species from other areas were missed in the present sampling). The phylogenetic analyses with the putative hybrid species removed didn't result in significant changes of the phylogeny and the nodal support (data not shown), indicating that the polymorphisms in *RPB2* sequences of these taxa didn't cofound the phylogenetic analysis. It must be noted that sequence polymorphisms were not observed in the ITS region for these putative hybrid species, suggesting that concerted evolution of ribosomal genes toward one sequence type has likely completed. This inference of hybrid speciation in *Philadelphus* can be tested further with more data from other nuclear loci.

Conclusions

Based on our results from phylogenetic analyses, *Philadelphus* is a paraphyletic group, with the monotypic genus *Carpenteria* nested within. The *Philadelphus-Carpenteria* complex was found to consist of three clades: the core *Philadelphus* clade, the *Carpenteria* clade and the *P. texensis-P.mearnsii* clade. Biogeographic analysis and divergence time dating suggested that the *Philadelphus-Carpenteria*

clade evolved in SW North America in the Oligocene. Our results do not support the classification scheme proposed by Hu. Future phylogenetic study with more species sampling across the genus geographic distribution range would be necessary to provide a robust framework for classification of Core *Philadelphus* species and for a clearer understanding of the biogeographic history. Futher phylogeographic studies are also helpful to evaluate the current species delimitation.

Acknowledgments

I would like to especially thank Dr. Jenny (Qiu-Yun) Xiang for all her guidance on the project, for supporting the research, and for commenting and correcting the manuscript. I would like to thank my committee members Dr. De-Yu Xie and Dr. Heike Winter-Sederoff for advice. Special thanks to Arnold Arboretum for providing the leaf material and to Dr. Boufford and Dr. Richardson for the collection of leaf material from the arboretum. Many thanks to Dr. Alan Weakly for discussion on morphological characters and to Xiang Liu, Norm Douglas, and Chunmiao Feng for their comments on the thesis and help in the lab.

REFERENCES

Andrea, W.N., & Hufford, L. 1995. Androecial morphology and ontogeny of the sister genera *Carpenteria* and *Philadelphus* (Hydrangeaceae). *Amer. J. Bot.* 82(6): 171-172.

Bangham, W. 1929. The chromosomes of some species of the genus *Philadelphus. J. Arnold Arbor.* 10:167-169.

Bell, W.A. 1957. *Flora of the upper cretaceous Nanaimo group of Vancouver Island, British Columbia.* Geol. Surv. Canada, Memoir.

Berggren, W.A., & Prothero, D.R. 1992. Eocene-Oligocene climatic and biotic evolution: an overview. Pp. 1-28 in: Prothero, D.R. and Berggren, W.A. (eds.) *Eocene-Oligocene climatic and biotic evolution.* Princeton: Princeton Univ. Press

Chaney, R.W., Condit, C., Axelrod, D.I. 1944. *Pliocene floras of California and Oregon.* Carnegie Inst. Washington, Publ.

Donoghue, M.J., & Smith, S.A. 2004. Patterns in the assembly of temperate forests around the northern hemisphere. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 29: 1633-1644

Doyle, J.J. & Doyle, J.L. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19: 11-15.

Drummond, A.J. & Rambaut, A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7: 214-221

Drummond, A.J., Ho, S.Y.W., Phillips, M.J. & Rambaut, A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology* 4: e88.

Drummond, A.J., Nicholls, G.K., Rodrigo, A.G. & Solomon, W. 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* 161: 1307-1320.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res.* 32: 1792-1797.

Fan, C.Z. & Xiang, Q.Y. 2001. Phylogenetic relationships within *Cornus* (Cornaceae) based on 26S rDNA sequences. *Amer. J. Bot.* 88: 1131-1138.

Fan, C.Z. & Xiang, Q.Y. 2003. Phylogenetic analyses of Cornales based on *26S* rRNA and combined *26 S* rDNA-*matK*-*rbcL* sequence data. *Amer. J. Bot.* 90: 1357-1372.

Frazier, C.Z. 1999. A taxonomic study of *Philadelphus* (Hydrangeaceae) as it occurs in New Mexico. *New Mexico Botanist* 13:1-6.

Harris, A.J., Xiang, Q.Y. 2009. Estimating ancestral distributions of lineages with uncertain sister groups: a statistical approach to dispersal–vicariance analysis and a case using *Aesculus* L. (Sapindaceae) including fossils. *J. Syst. Evol.* 47: 349-368

Hickman, J.C. 1993. *The Jepson manual: higher plants of California*. University of California Press.

Hu, S.Y. 1954-1956. A monograph of the genus *Philadelphus*. *J. Arnold Arbor.* 35:275-333; 36:52-109, 325-368; 37:15-90.

Huelsenbeck, J.P., Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* 17: 754-755.

Hufford, L. 1995. Seed morphology of Hydrangeaceae and ITS phylogenetics implications. *Int. J. Plant Sci.* 156: 555-580.

Hufford, L. 1997. A phylogenetic analysis of Hydrangeaceae based on morphological data. *Int. J. Plant Sci.* 158: 652-672.

Hufford, L., Moody, M.L., & Soltis, D.E. 2001. A phylogenetic analysis of Hydrangeaceae based on sequences of the plastid gene *matK* and their combination with *rbcL* and morphological data. *Int. J. Plant Sci.* 162: 835-846.

Janaki-Ammal, E.K. 1958. Iso-chromosomes and the origin of triploidy in hybrids between old and new world species of *Philadelphus*. *Proceedings: Plant Sciences.* 48: 251-258.

MacGinite, H.D. 1953. *Fossil plants of the Florissant beds, Colorado.* Publ. Carnegie Inst. Wash.

Moerman, D. 1998. *Native American Ethnobotany.* Oregon: Timber Press.

Nylander, J.A.A., Olsson, U., Alström, P., Sanmartin, I. 2008. Accounting for phylogenetic uncertainty in biogeography: a Bayesian approach to dispersal-vicariance analysis of the thrushes (Aves: *Turdus*). *Syst. Biol.* 57: 257-268

Posada, D. & Buckley, TR. 2004. Model selection and model averaging in phylogenetics: advantages of the AIC and Bayesian approaches over likelihood ratio tests. *Syst. Biol.* 53: 793-808

Posada, D. & Crandall, K.A. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14: 817-818.

Potts, R., and Behrensmeyer, A. K. 1992. Late Cenozoic terrestrial ecosystems. Pp. 419–541 in: Behrensmeyer, A. K., Damuth, J. D., DiMichele, W. A., Potts, R., Sues, H.-D., and Wing, S. L., (eds.) *Terrestrial ecosystems through time: evolutionary paleoecology of terrestrial plants and animals.* Chicago: Univ. of Chicago Press.

Rambaut , A. & Drummond, A.J. 2007. Tracer v1.4, Available at http://beast.bio.ed.ac.uk/Tracer

Ronquist, F. 1997. Dispersal-vicariance analysis: A new approach to the quantification of historical biogeography. *Syst. Biol.* 46: 195–203.

Ronquist, F. 2001. DIVAversion 1.2. Computer program for MacOS and Win32. Evolutionary Biology Centre, Uppsala University. Available at http://www.ebc.uu.se/systzoo/research/diva/diva.html.

Ronquist, F., Huelsenbeck, J.P. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572-1574.

Sang, T., Crawford, D.J., & Stuessy, T.F. 1997. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of paeonia (paeoniaceae). *Amer. J. Bot.* 84: 1120-1136

Sax, K. 1931. Chromosome numbers in the ligneous Saxifragaceae. *J. Arnold Arbor.* 12: 198-206.

Shaw, J., Lickey, E.B., Beck, J.T., Farmer, S.B., Liu, W., Miller, J., Siripun, K.C., Winder, C.T., Shilling, E.E., & Small, R.L. 2005. The tortoise and the hare II: Relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Amer. J. Bot.* 92: 142-166

Soltis, D.E., Xiang, Q.Y., & Hufford, L. 1995.Relationships and evolution of Hydrangeaceae based on *rbcL* sequence data. *Amer. J. Bot.* 82:504-514

Stanford, A.M., Harden, R. and Parks, C.R. 2000. Phylogeny and biogeography of *Juglans* (Juglandaceae) based on *matK* and ITS sequence data. *Amer. J. Bot.* 87:872-882.

Styer, C.H. & Stern, W.L. 1979. Comparative anatomy and systematic of woody Saxifragaceae. *Philadelphus. Bot. J. Linn.Soc* 79:267-289.

Tiffney, B.H. 1985. The Eocene North Atlantic land bridge: ITS importance in Tertiary and modern phytogeography of the northern hemisphere. *J. Arnold Arbor.* 66: 243-73.

Tiffney, B.H. & Manchester, S.R. 2001. The use of geological and paleontological evidence in evaluating plant phylogeographic hypothesis in the northern hemisphere tertiary. *Int. J Plant Sci.* 162: S3-S17

Oxelman, B, Liden, M, & Berglund, D. 1997. Chloroplast *rps16* intron phylogeny of the tribe *Sileneae* (Caryophyllaceae). *Plant Syst. Evol.* 206: 393-410.

Weakley, A.S. 2008. *Flora of the Carolinas, Virginia, and Georgia and Surrounding Areas.* Working Draft of Aprial 2008. University of North Carolina Herbarium, North Carolina Botanical Garden, Chapel Hill, NC.

Weakley, A.E. 2002. Evolutionary Relationships within the genus *Philadelphus* (Hydrangeaceae): A molecular phylogenetic and biogeographic analysis. Master's Thesis, Chapel Hill: University of North Carolina.

Wen, J. 1999. Evolution of eastern Asian and eastern North American disjunct distributions in flowering plants. *Annu. Rev. Ecol. Syst.* 30: 421-455.

White, T. J., Bruns T., Lee S., & Taylor J.. 1990. Amplifications and direct sequencing of fungal ribosomal RNA genes for phylogenetics.Pp. 315-322 in: Innis M., Gelfand D., Sninsky J., and White T. (eds.), *PCR protocols: a guide to methods and applications*, San Diego: Academic Press.

Wolfe, J. A. 1992. Climatic, floristic, and vegetational changes near the Eocene/Oligocene boundary in North America. Pp. 131–436 in: Prothero, D. R. & Berggren, W. A., (eds.) *Eocene-Oligocene climatic and biotic evolution.* Princeton: Princeton Univ. Press.

Xiang, Q.Y., Crawford, D.J., Wolfe, A. D., Tang, Y. C. & dePamphilis, C. W. 1998c. Origin and biogeography of *Aesculus* L. (Hippocastanaceae): a molecular phylogenetic perspective. *Evolution* 52: 988-997.

Xiang, Q.Y., Manchester, S.R., Thomas, D.T., Zhang, W.H., & Fan C.Z. 2005. Phylogeny, biogeography, and molecular dating of cornelian cherries (*Cornus*, Cornaceae): tracking tertiary plant migration. *Evolution* 59: 1685-1700.

Xiang, Q.Y., Moody, M.L., Soltis, D.E., Fan C.Z. &Soltis, P.S. 2002. Relationships within Cornales and circumscription of Cornaceae – *matK* and *rbcL* sequence data and effects of outgroups and long branches. *Mol. Phyl. Evol.* 24: 35-57.

Xiang, Q.Y. Soltis, D.E. & Soltis, P.S. 1998a. Phylogenetic relationships of Cornaceae and close relatives inferred from *matK* and rbcL sequences. *Amer. J. Bot.* 85: 285-297.

Xiang, Q.Y., Soltis, D.E. Soltis, P.S. 1998b. The eastern Asian and eastern and western North American floristic disjunction: congruent phylogenetic patterns in seven diverse genera. *Mol. Phyl. Evo.* 10:178-190.

Xiang, Q.Y., & Thomas, D.T. 2008. Tracking character evolution and biogeographic history through time in Cornaceae – Does choice of methods matter? *J. Syst. Evol.* 46: 349-374.

Xiang, Q.Y., Thomas, D.T., Zhang, W.H., Manchester, S.R., & Murrell, Z. 2006. Species level phylogeny of the Dogwood genus *Cornus* (Cornaceae) based on

molecular and morphological evidence- implication in taxonomy and Tertiary intercontinental migration. *Taxon* 55:9-30.

Xiang, Q.Y., Thorne, J.L., Seo, T.K., Zhang, W.H., Thomas, D.T., & Ricklefs, R.E. 2008. Rates of nucleotide substitution in Cornaceae (Cornales) – Pattern of variation and underlying causal factors. *Mol. Phyl. Evo.* 49: 327-342.

Xiang, Q.Y., Zhang, W.H., Ricklefs, R.E., Qian, H., Chen, Z.D., Wen, J., & Li, J. H. 2004. Regional differences in rates of plant speciation and molecular evolution: a comparison between eastern Asia and eastern North America. *Evolution* 58: 2175-2184.

Zwickl, D.J., 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation, The University of Texas at Austin.

Table 1-1. Classification system, geographic distribution, and the key morphological characters of *Philadelphus* from Hu (1954-1956). Subg. *Euphiladelphus* contained the type species, representing Subg. *Philadelphus*.

	Geographic distribution	key morphology
Subg. I.*Gemmatus*	Two sections, 14 spp.	Bud exposed; depauperate paniculate inflorescences, stamens 40-60; semi-inferior ovary, enlarged and often cristate, separated stigmas; fruits obovoid-ellipsoid or subglobose, with circumferential sepals; seed caudate long.
Sect. 1. *Coulterianus*	7 spp.; Mexico to Panama	Inflorescences compound, stamens 40-50; stigmas elongated and cristate; fruits obovoid-ellipsoid
Sect. 2. *Poecilostigma*	7 spp.; NE Mexico	Inflorescences with solitary flowers, rarely 3-flowered; stamens 30-44; stigmas elongated, not cristate; fruit subglobose
Subg. II *Euphiladelphus*	Three sections, 41 spp.	Bud enclosed; determinate-racemose or pauciflorous inflorescences, large (60-90) or medium (25-50) number of stamens; clavate or oar-shaped or linear stigmas; seed caudate long or short.

Table 1-1 cont.

Sect. 1. *Pauciflorus*	2 spp.; S Appalachian region of the United States	Inflorescences true dichasial cymes, often reduced to one or three flowers; stamens 60-90; stigmas oar shaped; fruit ellipsoid with circumferential sepals; seed caudate long, with pointed lobes on the crown.
Sect. 2. *Stenostigma*	Eight series, 33 spp.;	Inflorescences determinate-racemose; stamens 25-35, rarely over 40; stigmas clavate or linear; fruit ellipsoid with circumferential or subapical persistent sepals; seed caudate short, with pointed or rounded lobes on the crown
Ser. a. *Gordoniani*	4 spp.; NW US adjacent Canada and S US	Hypanthia glabrous; adaxial stigmatic surface longer than the abaxial; seed long caudate, embryo half as long as caudate; leaves glabrous or glabrescent
Ser. b. *Tomentosi*	3 spp.; W Himalayan region	Hypanthia glabrous; adaxial stigmatic surface longer than the abaxial; seed long caudate, embryo equal to or longer than caudate; leaves tomentose beneath

Table 1-1 cont.

Ser. c. *Pekinenses*	3 spp.; N China uplands	Hypanthia glabrous; adaxial stigmatic surface longer than the abaxial; seed short caudate, embryo longer than the caudate
Ser. d. *Coronarii*	3 spp.; European and Caucasus	Hypanthia glabrous; adaxial stigmatic surface longer than the abaxial; seed medium caudate, embryo equal to the caudate
Ser. e. *Delavayani*	2 spp.; SW China	Hypanthia glabrous; adaxial stigmatic surface shorter than the abaxial;
Ser. f. *Pubescentes*	2 spp.; Nashville Basin	Hypanthia uniformly and thickly pubescent; large shrubs up to 4 or 5 m high; lower surface of the leaves, the hypanthia and sepal uniformly long villose
Ser. g. *Sericanthi*	11 spp.; W. China	Hypanthia uniformly and thickly pubescent; medium or low shrub 1-3m high; lower surface of the leaves, the hypanthia and sepal strigose or scabrid; Seeds short-caudate; hair on the hypanthia strigose, scabrid or villose

Table 1-1 cont.

Ser. h. *Satsumani*	3 spp.; Japan	Hypanthia uniformly and thickly pubescent; medium or low shrub 1-3m high; lower surface of the leaves, the hypanthia and sepal strigose or scabrid; Seeds long- or medium-caudate; hair on the hypanthia weak, short, sparse
Sect. 3. *Microphyllus*	11 spp.; S Rocky Mountain to N Mexico	Inflorescences pauciflorous; stamens 25-40, rarely up to 50, stigmas linear; fruit ellipsoid with circumferential or subapical persistent sepals; seed caudate short, with rounded lobes on the crown
Subg. III. *Marcrothyrsus* Sect. 1. *Californicus*	One section, 3 spp; W flank of California	Bud exposed; paniculate inflorescences; medium number (25-37) of stamens; clavate stigmas; fruit ellipsoid with supermedian persistent sepals; seed caudate short.
Subg. IV *Deutzioides*	Three section, 8 spp.	Bud exposed; inflorescences with solitary, rarely ternate flowers; stamens 13-35; inferior ovary, columnar or subcapitate 4-grooved stigmas; fruit turbinate or subglobose, with apical persistent sepals; ecaudate seed.
Sect. 1. *Hirsutus*	2 spp.; S Appalachian	Flowers ternate; mesophytic plants; leaves serrate, hirsute, hairs all straight, style 4mm long

Table 1-1 cont.

Sect. 2. *Pseudoserpyllifolius*	4 spp.; Texas to Mexico	Flowers solitary; xerophytic, dwarf plants, leaves entire, strigose or strigose-villose, hairs all straight; style up to 1 mm.
Sect. 3. *Serpyllifolius*	2 spp.; Texas to Mexico	Flowers solitary; xerophytic, dwarf plants, leaves entire, lanate and hirsute beneath, trichomes dimorphous
Carpenteria	1 spp.; California	Leaves persistent, leathery; petals 5-8, stamens many (ca. 200); ovary superior; style 1, branches 5-7 above middle; stigma linear along style branch. Seeds fusiform. Chromosome 2n=20

Table 1-2 Species sampled for this study. List of species follows the classfication of *Philadelphus* by Hu (1954-56). Information on species number, geographic distribution, and GeneBank accession number of sequences used in the study follows the name of the section or series. Arn. Arb. = Arnold Arboretum. Bot. Gdn. = Botanical Garden. Question marks indicate missing data.

	Species sampled	Voucher	GenBank Acc. #				
Ingroup			*matK*	*rpoB-trnC*	*Rps16*	ITS	*RPB2*
Subg. I. *Gemmatus*							
Sect. 1. *Coulterianus*	*P. coulteri* Watson	Arn. Arb., 132-2000; Nat. Bot. Gdn., Ireland.	To be submitted	To be submitted	To be submitted	???	To be submitted
Sect. 2. *Poecilostigma*	Not sampled in this study						
Subg. II *Philadelphus*							
Sect. 1. *Pauciflorus*	*P. inodorus* Linn.	Arn. Arb., 320-79; North Carolina, US	To be submitted	To be submitted	To be submitted	To be submitted	To be submitted
		Arn. Arb., 449-94; R.B.G., Kew, England	To be submitted	To be submitted	To be submitted	To be submitted	To be submitted

Table 1-2 cont.

Sect. 2. Stenostigma								
Ser. a. *Gordoniani*	*P. lewisii* Pursh		Xiang et al., in review	???	???	???	???
Ser. b. *Tomentosi*	*P. tomentosus* Wall.	Arn. Arb., 853-80; Ludwig Spaeth, Germany	To be submitted	To be submitted	To be submitted	To be submitted	To be submitted
Ser. c. *Pekinenses*	*P. pekinensis* Rupr.	Arn. Arb., 14-95; Beijing, China	To be submitted	To be submitted	To be submitted	To be submitted	To be submitted
		Arn. Arb., 64-90; Cholla Pukto, Korea	To be submitted	To be submitted	To be submitted	To be submitted	To be submitted
Ser. d. *Coronarii*	*P. coronarius* L.	Arn. Arb., 537-76; North Carolina, US	To be submitted	To be submitted	???	To be submitted	To be submitted
		Arn. Arb., 214-77; Caucasus	To be submitted	To be submitted	To be submitted	To be submitted	To be submitted
Ser. e. *Delavayani*	*P. delavayi* L.Henry	Arn. Arb., 295-85; Sikang, China	To be submitted	To be submitted	???	To be submitted	???
	P. purpurascens Rehder	Arn. Arb., 127-95; Sichuan, China	To be submitted	To be submitted	To be submitted	To be submitted	To be submitted
		Arn. Arb., 296-85; Sikang, China	To be submitted	To be submitted	???	To be submitted	To be submitted
Ser. f. *Pubescentes*	*P. pubescens* Loisel.	Arn. Arb., 702-80; North Carolina, US	To be submitted	To be submitted	To be submitted	To be submitted	To be submitted

Table 1-2 cont.

	P. pubescens Loisel. *var. intectus* *A*.H.Moore	Arn. Arb., 4454; North Carolina, US	To be submitted	To be submitted	To be submitted	To be submitted	To be submitted
	P. incanus Koehne	Arn. Arb., 1842-80; Shennongjia, China	To be submitted	To be submitted	To be submitted	To be submitted	To be submitted
Ser. g. *Sericanthi*	*P. sericanthus* Koehne	Arn. Arb., 602-84; Zhejiang, China	To be submitted	To be submitted	To be submitted	To be submitted	To be submitted
		Arn. Arb., 217-77; U. Uppsala Bot. Gdn., Sweden	To be submitted	To be submitted	To be submitted	To be submitted	To be submitted
		Arn. Arb., 563-88; W. Zhejiang, China	To be submitted	To be submitted	To be submitted	To be submitted	To be submitted
	P. tenuifolius Rupr.	Arn. Arb., 418-89; Cheju-do Island., Korea	To be submitted	To be submitted	To be submitted	To be submitted	To be submitted
		Arn. Arb., 612-88; Mt. Mudung, Korea	To be submitted	To be submitted	To be submitted	To be submitted	To be submitted
	P. subcanus Koehne	Arn. Arb., 218-77; Sikang, China	To be submitted	To be submitted	To be submitted	To be submitted	To be submitted
	P. kansuensis (Rehder) S.Y.Hu	Arn. Arb., 815-80; Tatsuto Kadjaku, China	To be submitted	To be submitted	To be submitted	To be submitted	To be submitted

Table 1-2 cont.

	P. schrenkii Rupr.	Arn. Arb., 306-97; Jilin, China	To be submitted	To be submitted	To be submitted	To be submitted	To be submitted
		Arn. Arb., 838-69; Mt. Sorak, Korea	To be submitted	To be submitted	???	To be submitted	???
		Arn. Arb., 1717-77; Gyebang-san, Korea	To be submitted	To be submitted	???	To be submitted	To be submitted
Ser. h. *Satsumani*	*P. satsumi* Sieb.	Arn. Arb., 112-2001; Gunma Prefecture, Japan	To be submitted	To be submitted	To be submitted	To be submitted	To be submitted
Sect. 3. *Microphyllus*	Not sampled in this study						
Subg. III. *Marcrothyrsus*							
Sect. 1. *Californicus*	*P. cordifolius* Lange	Arn. Arb., 340-92; Virginia, US	To be submitted	To be submitted	???	To be submitted	To be submitted
Subg. IV *Deutzioides*							
Sect. 1. *Hirsutus*	*P. hirsutus* Nutt.	Arn. Arb., 18346; Arkansas, US	To be submitted	To be submitted	???	To be submitted	To be submitted
Sect. 2. *Pseudoserpyllifolius*	*P. mearnsii* W.H.Evans ex Rydb.	Harvard U. herbaria; By A.L. Hershey #3063	To be submitted	To be submitted	???	???	To be submitted
Sect. 3. *Serpyllifolius*	*P. texensis* S.Y.Hu var. *coryanus* S.Y.Hu	Arn. Arb., 517-90; Texas, US	To be submitted	To be submitted	???	To be submitted	To be submitted

Table 1-2 cont.

Outgroup						
1	*Carpenteria californica* Torr.	To be submitted	To be submitted	???	To be submitted	To be submitted
2	*Deutzia sp.*	To be submitted	To be submitted	???	To be submitted	???
3	*Deutzia sp.*	To be submitted	To be submitted	To be submitted	To be submitted	???

Table 1-3. Primer information for PCR and Sequencing.

Gene Region	Primer	5’ Sequence 3’	Source
matK	*matK*-1F	ACTGTATCGCACTATGTATCA	Sang et al., 1997
	matK-3F	AAGATGCCTCTTCTTTGCAT	Sang et al., 1997
	matK-3R	GATCCGCTGTGATAATGAGA	Sang et al., 1997
	matK-1R	GAACTAGTCGGATGGAGTAG	Sang et al., 1997
rpoB-trnC	*rpoB*	CKACAAAAYCCYTCRAATTG	Shaw et al., 2005
	trnC-R	CACCCRGATTYGAACTGGGG	Shaw et al., 2005
rps16	rpsF	GTGGTAGAAAGCAACGTGCGACTT	Oxelman et al., 1997
	rpsR2	TCGGGATCGAACATCAATTGCAAC	Oxelman et al., 1997
ITS	ITS5a	CCTTATCATTTAGAGGAAGGAG	Stanford et al. 2000
	ITS4	TCCTCCGCTTATTGATATGC	White et al. 1990
RPB2	*RPB2*-F1	GACAATTGGTCAGCTTATTGAGTGC	This Study
	RPB2-F2	GGAAGGAGATGCKACTCCCTTTAC	This Study
	RPB2-R2	TCAAACCCACGCATYTGATACCC	This Study
	RPB2-R1	TTTGGTAGTAAGTGGGACCCARGAA	This Study

Table 1-4. Variable sites of *RPB2* gene. First column shows the site number of the *RPB2* gene in the original alignment. Each row represents one site. Nucleotides that are same as the first column are shown as dots, missing data are shown as question marks, and indels are shown as dashes.

Site Number	P.hirsutus	P.schrenkii1	P.satsumi	P.tenuifolius1	P.tenuifolius2	P.sericanthus1	P.schrenkii3	P.subcanus	P.purpurascens1	P.cordifolius	P.sericanthus3	P.kansuensis	P.pubescens	P.coronarius	P.tomentosus	P.inodorus	P.pekinensis2	P.coulteri	P.inodorus2	P.sericanthus2	P.incanus1	P.pekinensis1	P.purpurascens2	P.coronarius2	Carpenteria	P.mearnsii	P.texensis
16	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.
19	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.
22	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.
23	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.
25	A	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.
36	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	G	G
42	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.
46	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	C
47	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A
48	A	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T
57	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T
71	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A
74	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.
75	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	G	G
76	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	G	G
84	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.
90	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	C	C
91	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T
101	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	G	G
128	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	T	T
129	G	.	.	.	.	.	.	.	.	.	.	R	A	R	R	.	.	.	.	.	.	.	.	.	.	.	.
133	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.
136	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	G
141	G	.	.	S	.	.	.	.	.	.	.	.	.	.	.	.	S	.	.	.	.	.	.	C	.	C	C

Table 1-4 cont.

145	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.
146	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.
157	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.
159	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.
163	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T
169	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	G
178	A	.	.	.	.	.	.	.	.	.	.	G	G	R	R	.	.	.	.	.	.	.	.	.	.	.	.
179	C	.	T	.	.	.	.	M	.	K	.	.	.	.	Y	.	.	.	.	.	.	.	.	.	.	.	.
180	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.
181	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T
190	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.
192	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.
194	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.
203	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T
205	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A
213	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.
220	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	C
235	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	A	-	.	.
238	A	.	.	.	.	.	.	.	.	.	.	T	T	T	T	.	.	.	.	?	.	.	.	.	-	.	.
246	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	-	A	A
249	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	R	.	.	.	A	A	A
269	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	A	A	A
276	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	.	C	.
277	A	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	?	.	.	.	.	.	.	.
284	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	.	A	A
296	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	G	.	.
300	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	G	.	.
303	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	.	A	A
312	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	G	G	G
317	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	C	.	.
321	G	.	.	W	T	.	.	.	.	R	.	A	A	A	A	A	R	.	A	?	.	.	.	A	A	A	A
325	T	.	.	W	.	.	.	.	.	W	.	A	A	A	A	A	W	.	A	?	.	.	.	A	A	A	A
326	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	C	.	.
332	G	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	?	.	.	.	.	.	.	.
333	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	A	.	.
340	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	T	T	T

Table 1-4 cont.

3 5 0	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	A	.	.
3 5 3	G	A	A	A	A	A	R	A	A	A	A	A	A	A	A	A	A	A	A	?	A	A	A	A	A	A	A
3 5 5	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	T	.	.
3 5 7	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	Y	T	A	A
3 6 9	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	T	.	.
3 7 1	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	A	A	A
3 7 4	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	C	.	.
3 7 5	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	.	G	G
3 7 8	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	C	C	C
3 8 2	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	T	-	-
3 9 6	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	T	T	T
4 0 4	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	A	.	.
4 0 5	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	R	?	.	.	.	.	.	.	.
4 0 6	C	.	.	.	.	.	.	.	.	.	.	Y	T	Y	Y	.	.	.	Y	?	.	.	.	.	T	T	T
4 0 9	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	G	.	.
4 1 1	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	A	.	.
4 1 4	A	.	.	.	.	.	.	.	.	.	.	R	G	R	R	.	.	.	R	?	.	.	.	.	G	G	G
4 1 7	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	.	C	C
4 1 8	C	.	.	.	.	.	.	.	S	S	.	.	.	.	.	.	.	G	.	?	.	.	.	.	.	.	.
4 1 9	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	G	G	G
4 2 4	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	C	C	C
4 3 2	G	.	.	.	.	.	.	.	.	.	.	.	.	R	R	A	.	.	?	?	.	.	.	.	.	.	.
4 3 3	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	?	?	.	.	.	.	.	.	.
4 4 6	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	?	.	.	.	.	.	T	T
4 5 1	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	?	.	.	.	.	A	.	.
4 5 4	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	?	.	.	.	.	.	G	G
4 5 6	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	?	.	.	.	.	T	.	.
4 5 8	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	?	.	.	.	.	T	.	.
4 6 0	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	?	.	.	.	.	A	.	.
4 6 2	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	?	.	.	.	.	.	T	T
4 6 3	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	?	.	.	.	.	C	.	.
4 6 4	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	?	.	.	G	.	.	.	.
4 7 4	A	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	?	?	.	.	?	.	C	.	.
4 7 7	T	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	?	?	.	.	?	.	C	.	.
4 7 8	G	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	?	?	.	.	?	.	A	.	.
4 8 2	C	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	?	?	.	.	?	.	A	.	.

Table 1-4 cont.

484	T	.	.	.	.	.	.	.	.	.	.	Y	C	?	Y	.	.	.	?	?	.	.	?	.	.	.	.
491	C	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	?	?	.	.	?	.	G	.	.
493	C	.	G	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	?	?	.	.	?	.	.	.	.
502	G	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	?	?	.	.	?	.	.	T	T
519	T	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	?	?	.	.	?	.	G	.	.
523	C	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	?	?	.	.	?	.	A	.	.
527	T	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	?	?	.	.	?	.	G	.	.
530	G	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	?	?	.	.	?	.	.	T	T
533	G	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	?	?	.	.	?	.	A	.	.
536	C	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	?	?	.	.	?	.	.	G	G
566	T	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	?	?	.	.	?	.	A	C	C
567	G	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	?	?	.	.	?	.	T	.	.
576	A	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	?	?	.	.	?	.	.	G	G
578	G	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	?	?	.	.	?	.	C	C	C
582	T	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	?	?	.	.	?	.	A	.	.
585	T	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	?	?	.	.	?	.	G	.	.
594	T	.	.	.	.	.	.	.	.	?	.	.	.	?	.	.	.	.	?	?	.	.	.	.	.	C	C
597	C	.	.	.	.	.	.	.	.	?	.	.	.	?	.	.	.	.	?	?	.	.	.	.	.	.	T
614	A	.	.	.	.	.	.	.	.	?	.	.	.	?	.	C	.	.	?	?	.	.	.	.	.	.	.
626	A	.	.	.	.	.	.	.	.	?	.	.	.	?	.	.	.	.	?	?	.	.	.	.	C	T	T
628	T	.	.	.	.	.	.	.	.	?	.	.	.	?	.	.	.	.	?	?	.	.	.	.	G	.	.
629	C	.	.	.	.	.	.	.	.	?	.	.	.	?	.	.	.	.	?	?	.	.	.	.	G	.	.
652	T	.	.	.	.	.	.	.	.	?	.	.	.	?	.	.	.	.	?	?	.	.	C	.	C	C	C
656	C	.	.	.	.	.	.	.	.	?	.	.	.	?	.	.	.	.	?	?	.	.	.	.	T	.	.
657	A	.	.	.	.	.	.	.	.	?	G	.	.	?	.	.	.	.	?	?	.	.	.	.	.	.	.
658	T	.	.	.	.	.	.	.	.	?	.	.	.	?	.	.	.	.	?	?	.	.	.	.	C	.	.
660	A	.	.	.	.	.	.	.	.	?	.	.	.	?	.	.	.	.	?	?	.	.	.	.	G	.	.
661	C	.	.	.	.	.	.	.	.	?	.	.	G	?	.	.	.	.	?	?	.	.	.	.	.	.	.
662	T	.	.	.	.	.	.	.	.	?	.	.	.	?	.	A	.	.	?	?	.	.	.	.	C	.	.
667	A	.	.	.	.	.	.	.	.	?	.	.	.	?	.	.	.	.	?	?	.	.	.	.	.	T	T
673	C	T	.	.	.	.	.	.	.	?	.	.	.	?	.	.	.	.	?	?	.	.	.	.	.	.	.
674	G	.	.	.	.	.	.	.	.	?	.	.	.	?	.	.	.	.	?	?	.	.	.	.	T	.	.
675	A	.	.	.	.	.	.	.	.	?	.	.	.	?	.	.	.	.	?	?	.	.	.	.	G	.	.
679	T	.	.	.	.	.	.	.	.	?	.	C	C	?	.	.	.	.	?	?	.	.	.	.	.	.	.
682	T	.	.	.	.	.	.	.	.	?	.	.	.	?	.	C	.	.	?	?	.	.	.	.	.	.	.
683	T	.	.	.	.	.	.	.	.	?	.	.	.	?	.	.	.	.	?	?	.	.	.	.	C	.	.

Table 1-4 cont.

692	C	.	.	.	.	.	.	.	.	?	.	.	.	?	.	.	.	.	?	?	.	.	.	.	.	T	T
695	G	.	.	.	.	.	.	.	.	?	.	.	.	?	.	-	.	.	?	?	.	.	.	.	.	A	.
697	G	.	.	.	.	.	.	.	.	?	.	.	.	?	A	.	.	.	?	.	.	.	.	.	A	.	.
698	T	.	.	.	.	.	.	.	.	?	.	.	.	?	.	.	.	.	?	.	.	.	.	.	A	G	G
708	A	?	.	?	?	.	.	.	.	.	.	?	?	?	?	.	?	?	?	.	?	?	.	?	C	C	C
716	T	?	C	?	?	.	.	K	.	.	.	?	?	?	?	.	?	?	?	.	?	?	.	?	-	.	.
720	G	?	.	?	?	R	.	.	.	.	?	?	?	?	?	.	?	?	?	.	?	?	.	?	A	.	.
722	T	?	.	?	?	.	.	.	.	.	?	?	?	?	?	.	?	?	?	.	?	?	.	?	A	.	.
727	C	?	.	?	?	.	.	.	.	.	?	?	?	?	?	.	?	?	?	.	?	?	.	?	G	T	T
728	A	?	.	?	?	.	.	.	.	.	?	?	?	?	?	.	?	?	?	.	?	?	.	?	T	C	C
730	-	?	.	?	?	.	.	.	.	.	?	?	?	?	?	.	?	?	?	.	?	?	.	?	T	C	C
731	-	?	.	?	?	.	.	.	.	.	?	?	?	?	?	.	?	?	?	.	?	?	.	?	A	T	T
733	-	?	.	?	?	.	.	.	.	.	?	?	?	?	?	.	?	?	?	.	?	?	.	?	T	C	C
734	C	?	.	?	?	.	.	.	.	.	?	?	?	?	?	.	?	?	?	.	?	?	.	?	.	T	T
740	T	?	.	?	?	.	.	.	.	.	?	?	?	?	?	C	?	?	?	.	?	?	.	?	G	G	G
751	T	?	.	?	?	.	.	.	.	.	?	?	?	?	?	.	?	?	?	.	?	?	A	?	.	.	.
755	T	?	.	?	?	.	.	.	.	.	?	?	?	?	?	.	?	?	?	.	?	?	.	?	.	A	A
761	G	?	.	?	?	.	.	.	.	.	.	?	?	?	?	.	?	?	?	.	?	?	.	?	A	.	.
763	T	?	.	?	?	.	.	.	.	.	.	?	?	?	?	.	?	?	?	.	?	?	.	?	C	.	.
767	T	?	.	?	?	.	.	.	.	A	.	?	?	?	?	A	?	?	?	.	?	?	.	?	A	A	A
772	C	?	.	?	?	.	.	.	.	.	.	?	?	?	?	T	?	?	?	.	?	?	.	?	.	.	.
774	C	?	.	?	?	.	.	.	.	.	.	?	?	?	?	.	?	?	?	.	?	?	.	?	T	T	T
787	G	?	.	?	?	.	.	.	.	.	.	?	?	?	?	.	?	?	?	.	?	?	.	?	T	.	.
793	G	?	.	?	?	.	.	.	.	.	.	?	?	?	?	.	?	?	?	.	?	?	.	?	A	.	.

Fig.1-1. One of the MrBayes trees resulting from Bayesian analysis of the combined sequences of three cpDNA regions. Long branches were broken for a better view of shorter branches. Numbers below the branches are branch length and numbers above the branches are Bayesian posterior probabilities. Branches without numbers represent those with posterior probability values < 50%. The classification scheme of Hu (1954-56) are indicated in the parentheses next to the species name, following Subg., Sect., Ser.

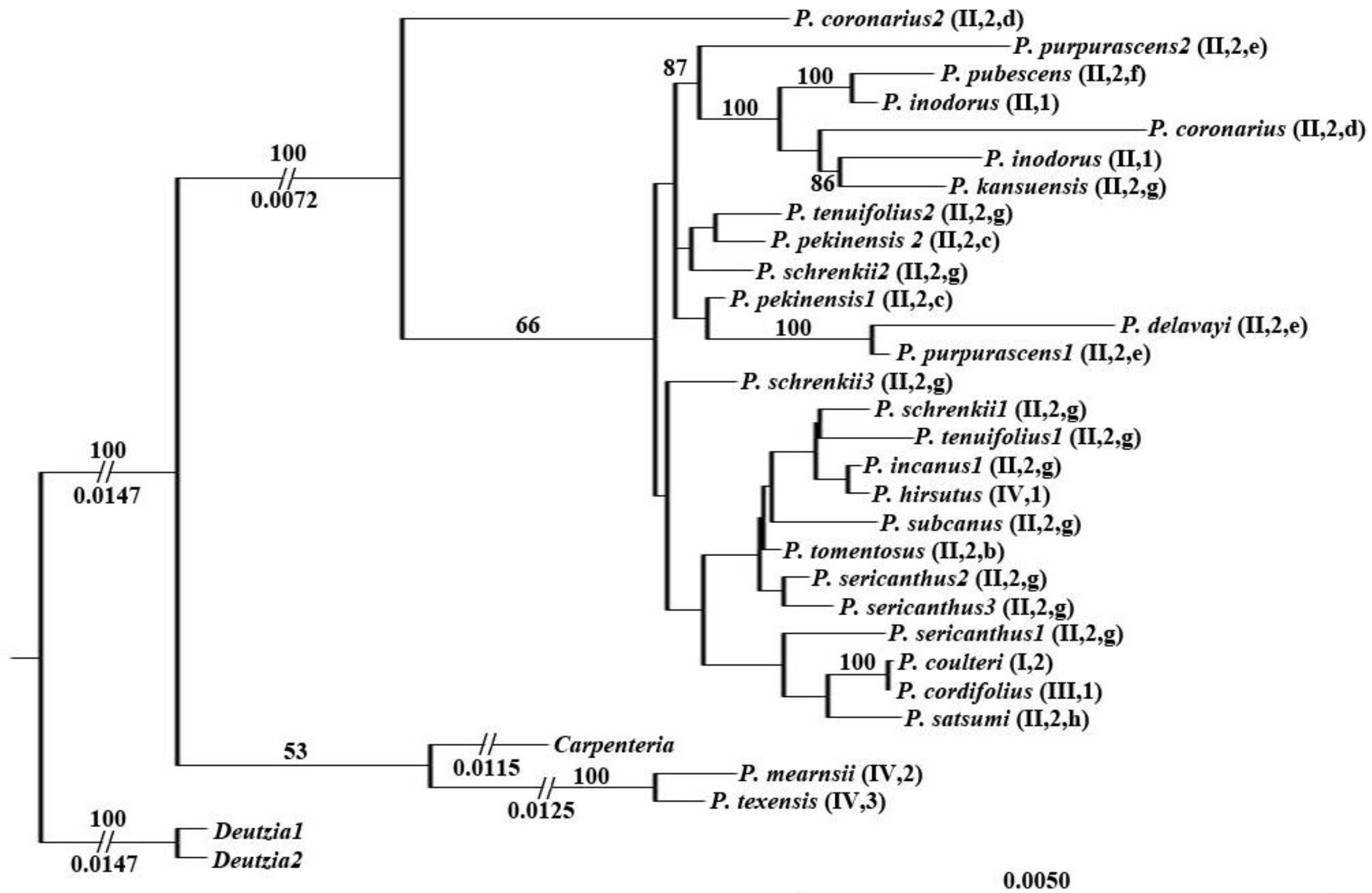

P. coronarius2 (II,2,d)
P. purpurascens2 (II,2,e)
87
100
P. pubescens (II,2,f)
100
P. inodorus (II,1)
P. coronarius (II,2,d)
100
0.0072
P. inodorus (II,1)
86
P. kansuensis (II,2,g)
P. tenuifolius2 (II,2,g)
P. pekinensis 2 (II,2,c)
P. schrenkii2 (II,2,g)
P. pekinensis1 (II,2,c)
66
100
P. delavayi (II,2,e)
P. purpurascens1 (II,2,e)
P. schrenkii3 (II,2,g)
P. schrenkii1 (II,2,g)
P. tenuifolius1 (II,2,g)
100
0.0147
P. incanus1 (II,2,g)
P. hirsutus (IV,1)
P. subcanus (II,2,g)
P. tomentosus (II,2,b)
P. sericanthus2 (II,2,g)
P. sericanthus3 (II,2,g)
P. sericanthus1 (II,2,g)
100
P. coulteri (I,2)
P. cordifolius (III,1)
P. satsumi (II,2,h)
53
Carpenteria
0.0115
100
P. mearnsii (IV,2)
0.0125
P. texensis (IV,3)
100
Deutzia1
0.0147
Deutzia2
0.0050

Fig. 1-2 One of the MrBayes trees resulting from Bayesian analysis of the combined sequences of the two nuclear DNA regions. Long branches were broken to show a better view of the shorter branches. Numbers below the branches are branch length and numbers above the branches are Bayesian posterior probabilities. Branches without numbers represent those with posterior probability values < 50%. The classification scheme of Hu (1954-56) are indicated in the parentheses next to the species name, following Subg., Sect., Ser.

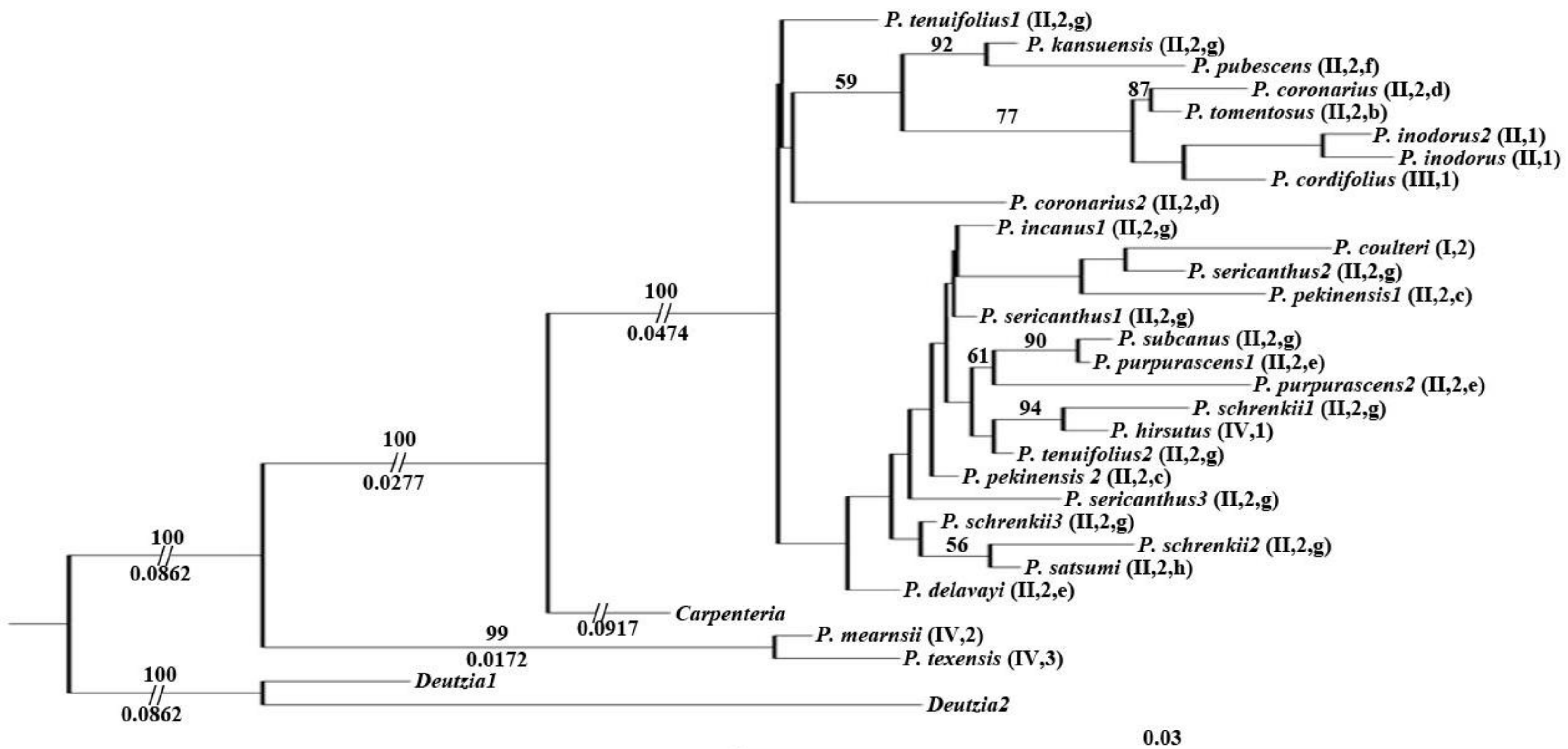
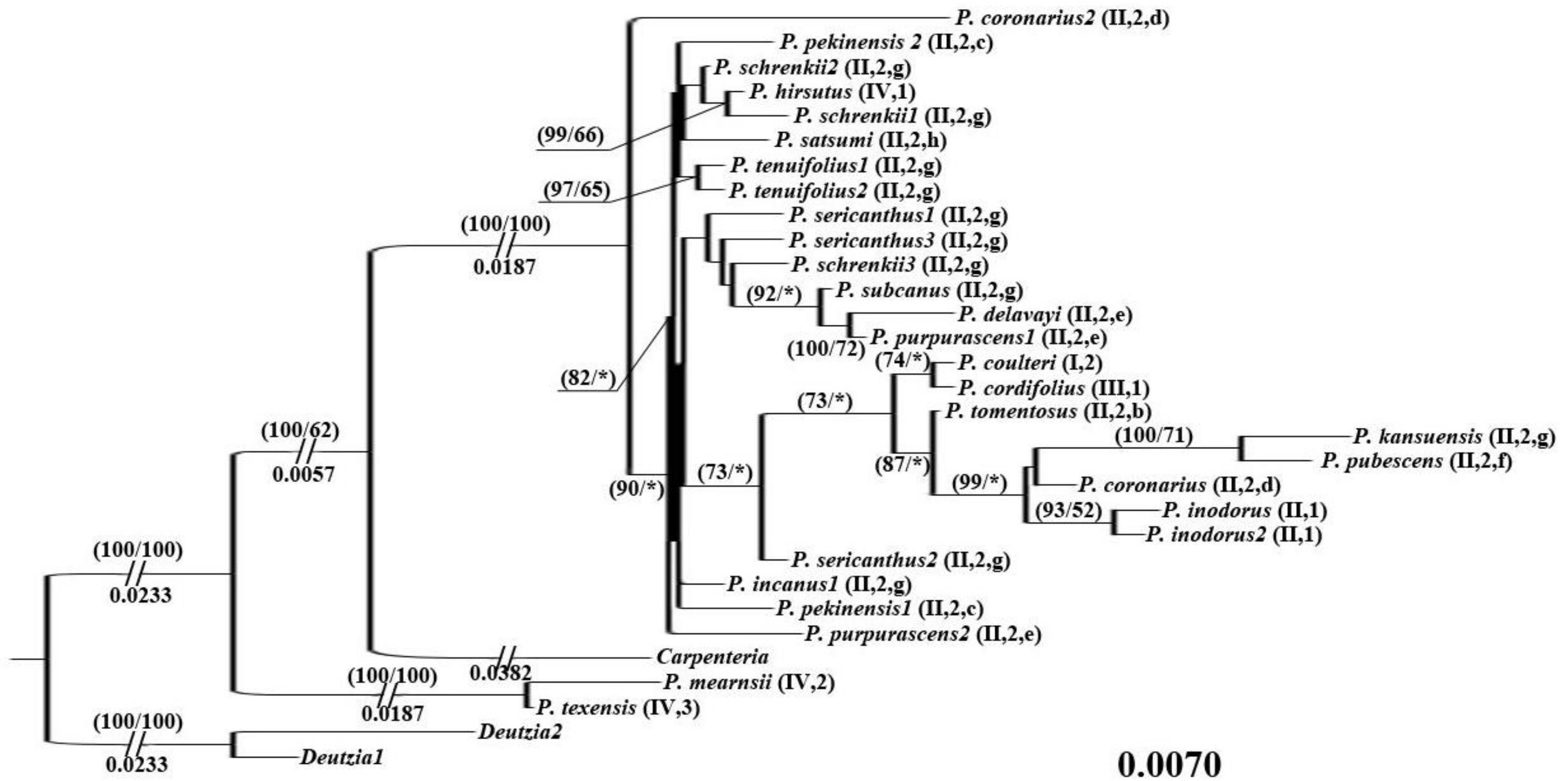
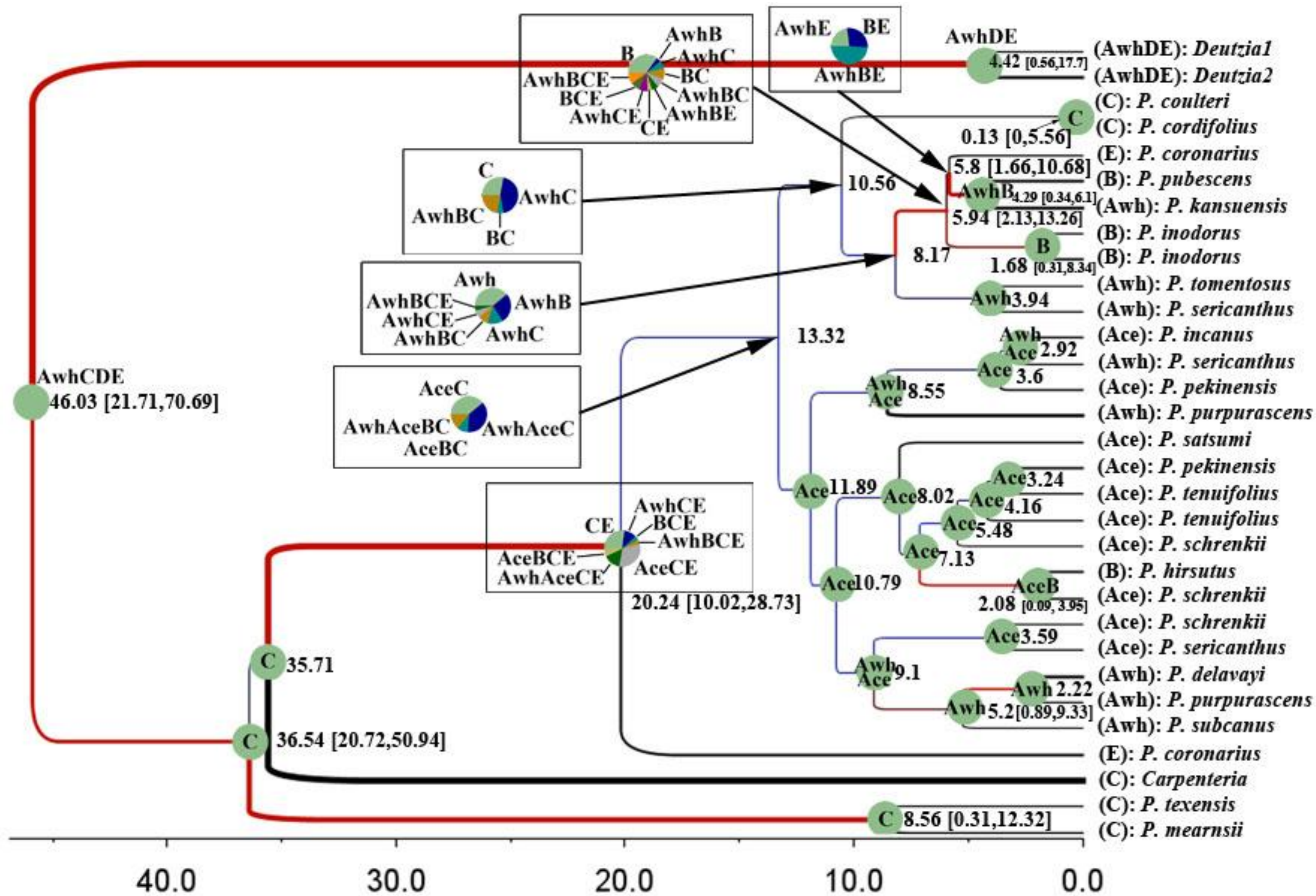

P. tenuifolius1 (II,2,g)
92
P. kansuensis (II,2,g)
P. pubescens (II,2,f)
59
87
P. coronarius (II,2,d)
77
P. tomentosus (II,2,b)
P. inodorus2 (II,1)
P. inodorus (II,1)
P. cordifolius (III,1)
P. coronarius2 (II,2,d)
P. incanus1 (II,2,g)
P. coulteri (I,2)
P. sericanthus2 (II,2,g)
P. pekinensis1 (II,2,c)
100
0.0474
P. sericanthus1 (II,2,g)
90
P. subcanus (II,2,g)
61
P. purpurascens1 (II,2,e)
P. purpurascens2 (II,2,e)
94
P. schrenkii1 (II,2,g)
100
P. hirsutus (IV,1)
0.0277
P. tenuifolius2 (II,2,g)
P. pekinensis 2 (II,2,c)
P. sericanthus3 (II,2,g)
100
P. schrenkii3 (II,2,g)
56
P. schrenkii2 (II,2,g)
0.0862
P. satsumi (II,2,h)
P. delavayi (II,2,e)
Carpenteria
0.0917
99
P. mearnsii (IV,2)
0.0172
P. texensis (IV,3)
100
Deutzia1
0.0862
Deutzia2
0.03

Fig.1- 3 One of the MrBayes trees resulting from Bayesian analysis of the combined sequences of all five DNA markers. Numbers below the branches are branch length and numbers above the branches are Bayesian posterior probabilities followed by bootstrap value from maximum likelihood analysis. An asterisk indicates the branch has <50% bootstrap value. Branches without numbers represent those with posterior probability values < 50%. The classification scheme of Hu (1954-56) are indicated in the parentheses next to the species name, following Subg., Sect., Ser.



P. coronarius2 (II,2,d)
P. pekinensis 2 (II,2,c)
P. schrenkii2 (II,2,g)
P. hirsutus (IV,1)
P. schrenkii1 (II,2,g)
P. satsumi (II,2,h)
P. tenuifolius1 (II,2,g)
P. tenuifolius2 (II,2,g)
P. sericanthus1 (II,2,g)
P. sericanthus3 (II,2,g)
P. schrenkii3 (II,2,g)
P. subcanus (II,2,g)
P. delavayi (II,2,e)
P. purpurascens1 (II,2,e)
P. coulteri (I,2)
P. cordifolius (III,1)
P. tomentosus (II,2,b)
P. kansuensis (II,2,g)
P. pubescens (II,2,f)
P. coronarius (II,2,d)
P. inodorus (II,1)
P. inodorus2 (II,1)
P. sericanthus2 (II,2,g)
P. incanus1 (II,2,g)
P. pekinensis1 (II,2,c)
P. purpurascens2 (II,2,e)
Carpenteria
P. mearnsii (IV,2)
P. texensis (IV,3)
Deutzia2
Deutzia1
(99/66)
(97/65)
(100/100)
0.0187
(92/*)
(100/72)
(82/*)
(74/*)
(73/*)
(100/62)
0.0057
(87/*)
(100/71)
(73/*)
(90/*)
(99/*)
(93/52)
(100/100)
0.0233
0.0382
(100/100)
0.0187
(100/100)
0.0233
0.0070

Fig.1-4. The tree with the highest posterior probability resulting from BEAST analysis, showing divergence times from BEAST and geographic distributions at nodes from S-DIVA analysis. The numbers next to nodes are divergence time (Million years ago, Mya) and 95% confidence interval. The pie chart at each node indicates the probability of alternative ranges for the node. Branch color gradient corresponds to values of posterior probability, from blue (lowest) to red highest); Branch width corresponds to the rate of molecular evolution, thicker branches represent higher rate.


(AwhDE): Deutzia1
(AwhDE): Deutzia2
(C): P. coulteri
(C): P. cordifolius
(E): P. coronarius
(B): P. pubescens
(Awh): P. kansuensis
(B): P. inodorus
(B): P. inodorus
(Awh): P. tomentosus
(Awh): P. sericanthus
(Ace): P. incanus
(Awh): P. sericanthus
(Ace): P. pekinensis
(Awh): P. purpurascens
(Ace): P. satsumi
(Ace): P. pekinensis
(Ace): P. tenuifolius
(Ace): P. tenuifolius
(Ace): P. schrenkii
(B): P. hirsutus
(Ace): P. schrenkii
(Ace): P. schrenkii
(Ace): P. sericanthus
(Awh): P. delavayi
(Awh): P. purpurascens
(Awh): P. subcanus
(E): P. coronarius
(C): Carpenteria
(C): P. texensis
(C): P. mearnsii
AwhCDE
46.03 [21.71,70.69]
36.54 [20.72,50.94]
35.71
20.24 [10.02,28.73]
13.32
10.56
8.17
AwhDE
4.42 [0.56,17.7]
0.13 [0,5.56]
5.8 [1.66,10.68]
5.94 [2.13,13.26]
1.68 [0.31,8.34]
8.56 [0.31,12.32]
40.0
30.0
20.0
10.0
0.0

CHAPTER 2

Molecular Evolution of *FLORICAULA/LEAFY*-Like Genes in the Core Eudicots – Detecting Positive Selection with the Branch-Site Model

Abstract

FLORICAULA/LEAFY-like (*FLO/LFY*-like) gene is an important transcription factor in the plant kingdom. The present study examined the dynamics of molecular evolution of *FLO/LFY*-like gene in the Core Eudicots. We compared the rates of nucleotide substitutions at synonymous sites (dS), nonsysnonymous sites (dN), and the dN/dS ratio among domains. We also conducted a series of analyses to determine if positive selection is driving the gene evolution in some lineages that have duplication and/or functional divergence of *LFY*, and if so, at what sites. Our results revealed that the evolution of *LFY* was in general under purifying selection due to strong functional constraint, with the C domain under the strongest selection force and the intervening domain being the most relaxed. Our study also showed that the detection of positive selection using the Branch Site Model was robust to taxon sampling density, but sensitive to sequence length and alignment ambiguity. Our analyses under various conditions consistently detected positive selection in Fabaceae, where *FLO/LFY* evolved a role of the *KNOX1* gene function in regulating compound leaf development. Under the best alignment, we detected positive selection at several sites in Asterales, Brassicaceae, and Fabaceae where gene duplication and/or novel function of *LFY* have been reported. The positive selection

sites involved substations of amino acids with different chemical properties. These sites are potential candidates for future analyses of functional divergence of LFY.

Keywords

FLO/LFY-like gene, Molecular evolution, nonsynonymous substitution rate, positive selection, purifying selection, synonymous substitution rate

Introduction

Evolutionary analysis of transcription factor gene sequences is a powerful way to uncover potential biochemical and structural changes at molecular level that may be important to morphological evolution. Molecular evolutionary analysis of transcription factors can not only reveal the changes underlying gene functional divergence, but also identify the driving forces governing the molecular evolution. *LFY* is a plant specific transcript factor regulating inflorescence and floral development (Baum et al., 2005; Parcy, 2005; Schultz and Haughn, 1991, Weigel et al., 1992, Huala and Sussex, 1992, Blázquez et al., 1997). Unlike many plant transcription factors, *LFY* is known to exist as a single copy in most flowering plants and consists of three exons and two introns (Himi et al., 2001). In the model plant *Arabidopsis thaliana*, *LFY* controls the transition from vegetative to reproductive stage and determining floral meristem identity (Carpenter and Coen, 1990; Coen et al., 1990; Haughn and Somerville, 1988; Schultz and Haughn, 1991; Weigel et al., 1992). This function of *LFY* has been found to be largely conserved in all monocot and dicot plants that have been investigated (see review Benlloch et al., 2007). Duplication and functional divergence of *LFY* were reported in a few diploid lineages including Fabaceae, Brassicaceae, and Lamiales (Hofer et al., 1997; DeMason et al., 2001; Sliwinski et al., 2007; Aagaard et al., 2006). In Fabaceae, *LFY* plays a role overlapping with KNOX1 in compound leaf development (Sliwinski et al, 2007; Champagne et al., 2007). In Brassicaceae duplicated *LFY* shifted its function to

control rosette flowering development (Baum et al., 2005), while in Lamiales, duplications of *LFY* were reported but functional studies of paralogs have not been conducted (Sliwinski et al., 2007; Aagaard et al., 2006).

Maizel et al (2005) compared the *LFY* sequences among major lineages of land plants and identified two putative conserved domains (N- and C-domain). They demonstrated, by using transformation and comparative sequence analyses, that the function of *LFY* in controlling floral initiation in angiosperms evolved by nucleotide mutations in the DNA binding domain (specifically in the C domain). The conserved function of *LFY* in controlling floral meristem initiation suggests a slow rate of sequence evolution in the gene, especially in the C domain, driven by strong purifying selection, in order to preserve its function. However, the great variation in the pattern and tempo of floral initiation present in angiosperms suggests differences in the spatiotemporal patterns of *LFY* expression among lineages, possibly controlled by changes in the cis-regulatory element or the cis-coding region. Furthermore, duplication and functional divergence of the gene suggest possible adaptive selections on *LFY* in those lineages that experienced gene duplication and functional divergence. We attempted to gain insights into some of these predictions by examining the evolutionary history, substitution rate, and selection force of the gene. Specifically, we are interested in the following questions: Does *LFY* sequence evolve at the lower rate in the C-domain than in other domains? Are nucleotide substitutions in *LFY* extremely rare at the nonsynonymous sites as expected from its

conserved function? Is there adaptive evolution following gene duplication events or associated with functional divergence? What might be the evolutionary driving force governing the sequence and functional divergence?

Materials and Methods

Data Acquiring

We downloaded all available *FLO/LFY*-like genes and their orthologs from Genbank using a protein BLAST search (BLASTp) against the Non-redundant protein sequences database on NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The BLASTp search was conducted using the *LFY* protein sequences of *Arabidopsis thaliana* (NCBI gi: 18424517) as the query (word size = 3, matrix = PAM 70). Candidate sequences with a Score >=450 and E-value <= 1e-80 were retained and their DNA sequences were used for the analyses. The BLASTp hits filtered out by these criteria were short cDNA sequences and the flanking introns generated for phylogenetic reconstruction in previous studies. The Complete cDNA sequence of *LFY* in *Arabidopsis thaliana* is 1263 base pairs and was used as the reference for aligning the retrieved sequences. All sequences retained were reported as LFY homologs.

Multiple Sequence Alignment

A preliminary full-length multiple sequence alignment was conducted using the MUSCLE program (Edgar, 2004). The alignment was further adjusted by eye based

on codons and then translated into amino acid sequences (applying the universal codon) using the software Se-Al (http://tree.bio.ed.ac.uk/software/seal/). The protein sequence matrix containing a few angiosperms and other land plant species from Maizel et al. (2005) was used as a reference to re-check against our amino acid sequence alignment, which was then used as the guide for our cDNA alignment. The sequence between the N and C domains is too divergent to align when including the earlier diverging angiosperms. Due to this reason, the early diverging angiosperms were removed and only those sequences from the core-Eudicots were retained. The sequence matrix contains 95 species and 1773 base pairs, representing Asterids, Rosids, Caryophyllids, and an outgroup species from early diverging Eudicots.

Reconstructing the Gene Genealogy Using Phylogenetic Analysis

Phylogenetic analysis of the aligned data matrix was performed for the cDNA sequences to reconstruct the gene genealogy. The analysis was performed using a Bayesian Metropolis-Hastings coupled Markov chain Monte-Carlo (MCMC) in MrBayes 3.1.2 (Huelsenbeck et al., 2001; Ronquist et al., 2003). ModelTest 3.7 (Posada et al., 1998, 2004) was used to estimate parameters in the MrBayes analysis, which was run prior to the MrBayes analysis. Two parallel runs started from a random tree and simulated for 5,000,000 generations with an output of trees for every 100 generations. The program Tracer 1.4 (Rambaut and Drummond, 2007)

was then used to determine the approximate burn-in, which is the former 10% for every dataset. Trees from No. 5001 to No.50000, a total of 45000 trees, were imported to PAUP 4.0b10 (Swofford, 2002) to compute a 50% major-rule consensus tree. Consensus trees were saved and compared among duplicate runs. The genealogy reconstructed by MrBayes was used for substitution rate analysis.

Comparing Rate and Selection Force of Sequence Evolution

Pairwise Comparison Among Domains Using MEGA

We compared the substitution rates and selecting forces among the two conserved domains, the N-domain and C-domain, and the domain in between (named the I-domain, meaning the intervening domain). We used the method of Nei and Gojobori (Nei and Gojobori, 1986) implemented in MEGA 3.1(Kumar et al., 2004) to estimate the pairwise synonymous substitution rates (dS) and nonsynonymous substitution rates (dN). The selection force [dN/dS (ω) values] and its mean and standard deviation were calculated for the three regions using Microsoft Excel.

Genealogy-Based Comparisons of Substitution Rates and Selection Mode and Force Among Branches and Sites using PAML

We estimated the dN and dS of the gene as a whole for all the branches of the gene genealogy to examine the heterogeneity of these substitution rates and their ratios among clades and gene copies. Selection mode and force on a gene is

commonly determined by the ω value (the dN/dS ratio). A value of $\omega < 1$ is generally interpreted as evidence of purifying (or negative) selection (to minimize changes in protein sequences); the smaller the ω value, the stronger the selection force. In contrast, a value of $\omega > 1$ is commonly used as evidence of positive/diversifying selection favoring changes in protein sequences; the greater the ω value, the stronger the selection force, whereas $\omega = 1$ is evidence of neutral evolution. Although there were some criticisms on using the ω value to predict positive selection (Nazawa et al., 2009a), this criticism is refuted (Yang et al. 2009; Nazawa et al., 2009b). A value of $\omega > 1$ for a gene sequence as a whole or at a site of the gene sequence indicates the presence of excess nonsynonymous substitutions in a population than expected from neutral evolution (i.e., equal rates of synonymous and nonsynonymous substitutions). Thus, the observation of sites with omega > 1 predicts diversifying selection and identification of positive selection sites based on ω values provides a small list of sites to guide gene functional analyses to test functional divergence, as argued by Yang et al. (2009). Here we applied a genealogy-based maximum likelihood method implemented in ‘codeml’ of PAML 4.1 (Yang, 1997, 2007) to estimate the dN, dS, and ω values along each branch on the gene genealogy in order to gain insights into the selection mode and force acting on *LFY* in the Core Eudicots. The “free ratio model” (FRM) was used to estimate the ω value for each branch (by setting “model=1”). This model assumes an independent

ω for each branch. The tree topology resulting from MrBayes was used as the genealogy framework.

Furthermore, we used the likelihood-based Site Model (SM) in PAML 4.1 to determine if any specific sites of *LFY* are under positive selection based on the ω values. In the analysis, the codeml control file was set with parameters "model=0 Nsite=0 1 2 3 7 8", so that the ω values among branches was constant but vary among sites. A total of six site models [M0 (one-ratio model), M1 (neutral model), M2 (selection model), M3 (discrete model), M7 (β model), and M8 (β and ω model)] (Yang et al., 2000) were compared. Likelihood Ratio Test (LRT) were conducted for the following pairs of models, M0 vs. M3, M1 vs. M2, and M7 vs. M8 to determine which model fit the data better.

Testing Positive Selection on Specific Branch

The Free Ratio Model (FRM) in which the ω ratio is estimated for each branch is apparently very parameter-rich. The program authors recommended caution in using the ω values derived from the FRM to infer positive selection on branches. On the other hand, the Site Model may fail to detect positive selection confining to a particular lineage because the model assumes the ω of a site is constant on all the branches and positive selection signals present in a particular lineage cannot be detected. Therefore, we further tested the positive selection detected by the FRM and SM using the modified Branch-Site Model (BSM) (Yang, 2002; Zhang et al.,

2005). In other words, the branches revealed to be under positive selection from the FRM were further tested for positive selection using the BSM followed by detection of positive selection sites using the Bayes Empirical Bayes (BEB) (Yang et al., 2005) as described below.

Each analysis was conducted by defining the branch of interest as the foreground branch and the rest of the branches on the genealogy as the background branch. By setting "model=2", the branches were divided into 2 groups. The foreground group was marked out in the tree file to be distinguished from the background group. To test whether the chosen foreground groups are under positive selection, a test hypothesis was compared to a null hypothesis. In the null hypothesis, we defined "fix_omega=1" for the foreground group, to allow the greatest ω estimated by the software to be less than or equal to one, i.e. no positive selection, while in the test hypothesis, the largest ω is allowed to be greater than one, i.e. positive selection exists. In both null and test hypotheses, the background branch was allowed to have omega values that are less than one. An LRT test was then performed to compare the test hypothesis vs. the null hypothesis. If the null hypothesis was rejected at a significant level, the foreground group was confirmed to be under positive selection. Bayes Empirical Bayes (BEB) analysis implemented in the BSM was then employed to identify the positive selection sites in the branches confirmed with positive selection site (Yang et al., 2005).

The FRM revealed three of the branches are under positive selection (see below). We tested each of these branches separately using the Modified BSM. In addition, we also tested for positive selection in three other clades: the Fabaceae clade (with functional divergence of duplicated *LFY*), the Rosales clade, and the Lamiales clade (both have duplication of *LFY*).

Impact of sample size, sequence alignment, and ambiguity sites

Due to the fact that the number of *FLO/LFY*-like gene sequences available from GenBank vary among lineages, we questioned whether the detection of positive selection depends on the sampling density of taxa in the foreground and background branches as it is well known that node density effect influences the estimation of substitution rates (Hugall and Lee, 2007; Xiang et al., 2008). In our matrix, we had more sequences from Brassicaceae and Asterids than from other lineages. These two lineages were confirmed to have positive selection in the analysis using BSM model. Therefore, we tested whether the positive selection detected on these two lineages was due to their larger sample size. Because the estimation of substitution rates is also affected by sequence alignment, we tested how changes in alignments of ambiguous regions affect the detection of positive selection. Thus, we compared the results from four matrices that differed either in taxon sampling size or in the alignment of the most variable regions to determine if the detection of positive selection using the modified BSM is sensitive to these factors. The four matrices are: Matrix 1, containing 1773 base pairs and 95 core Eudicot species + an outgroup,

Eschscholzia californica, from Ranunculaceae, an early diverging Eudicot; Matrix 2, containing 62 species and 1773 base pairs; In this matrix we reduced the sampling from Brassicaceae and Asterids by randomly removing terminal species to reach a sampling size more or less equal to their sister clade. This resulted in 11 and 23 species from each of these clades, respectively. Matrix 3 is Matrix 1 with alternative alignment in the most variable portions; and Matrix 4 is Matrix 1 with the most variable portions that have ambiguity in alignment removed (132 base pairs flanking the 5' end of the N-domain and 162 Base pairs in the I-domain).

Results

Structural Characteristics and Sequence Variation

We retained a total of 95 sequences of the Core Eudicots, including 30 sequences from Asterids, 61 from Rosids, and four from Caryophyllales in the study (Table 2-1). All of these *FLO/LFY*-like cDNA sequences are approximately the same length (ca. 1300 base pairs). The two conserved domains, the N-domain and the C-domain defined by Maizel, et al., (2005) were identified in our alignment. In *Arabidopsis,* the N domain is 228 base pairs and located in the first exon, while the C domain is 477 base pairs and located in the second and the third exons (Fig. 2-1). There were many indels in the aligned sequences, and most of them are short (3 -15 Base pairs) and are present in the I-domain, or after the C-domain. Two long indels

were observed in Caryophyllales species; one of them is located at the 3' end of 1st exon, in the I-domain, ranging from 33 to 45 base pairs among the four Caryophyllales species, and the other is located in the 2nd exon, also in the I-domain, ranging from 69 to 93 bps (Fig.2-1). These two indels were not present in other clades nor the outgroup sequence.

Genealogy of *LFY/FLO*-like Gene of Core Eudicots

The 50% Majority rule consensus tree resulting from the MrBayes analysis (Fig. 2-2) shows the major clades and their relationships generally congruent with the angiosperm phylogeny. Gene duplication was indicated for Rosales, Brassicaceae and Lamiales. In Rosales, each copy formed a monophyletic group sister to each other. The duplication event occurred at the base of Rosales clade. The same pattern was observed for the Lamiales clade. However, in Brassicaceae, the duplicated copies from the same species were closely related to each other, indicating recent gene duplication events occurred independently in different species of the clade.

Pairwise Substitution Rate Comparison between Domains

Pairwise substitution rate estimated from MEGA showed that the N-domain has an average of 0.090 for nonsysnonymous substitutions (dN), an average of 0.706 for synonymous substitutions (dS) and an average of 0.130 for the dN/dS ratio. These

dN and dN/dS values are higher than those observed for the C-domain (0.055 for dN and 0.077 for dN/dS ratio) while the dS values are similar between the two domains (0.706 vs. 0.715). In the intervening non-conserved region, both average dN (0.263) and average dN/dS ratio (0.390) are greater than values for the N and C domains while the average dS (0.675) remains similar (Fig. 2-3). The distribution of the estimates of dN, dS and their ratios for each region are shown in Figure 2-4. The 95% confidence intervals of dN/dS range from 0.127 to 0.133) for the N-domain, 0.076 to 0.078 for the C-domain, and 0.384 to 0.395 for the between-domain region (Fig. 2-4). These values show significant difference of selection force among the three gene regions. The 95% confidence intervals of dN and dS for each region are similarly in a relatively small range (for dN: N-domain, 0.089 - 0.092, C-domain, 0.055 - 0.057, and I-domain, 0.259 - 0.267; for dS: N-domain, 0.698 - 0.714, C-domain, 0.708 - 0.721, and the I -domain, 0.668 - 0.681). While there is much variation in the dN among regions, there is little difference of dS among the three regions.

Genealogy-Based Comparisons among Sites and Branches Using PAML

Analyses using the Site Model (SM) detected positive selection at site 104 (on the 5' end before the N-domain) by the discrete model (M3) and the $\beta + \omega$ model (M8), but not in the selection model (M2) (Table 2-2). Further LRT tests showed that the likehood score differences between the neutral model (Null model M1) and selection model (M2), and between the β model (Null model, M7) and the $\beta + \omega$

model (M8) are not statistically significant, but the difference of likelihoods between the one-ratio model (Null model, M0) and the discrete model (M3) is significant (Table 2-3). The discrete model allows the ω to vary among sites, which is more realistic for the *FLO/LFY* gene, thus having a greater likelihood score.

The phylogeny-based estimation of dN, dS, and their ratios using the FRM implemented in PAML showed that most branches have dN/dS ratios much smaller than 1, but three branches have dN/dS ratios greater than 1, suggesting possible diversifying selections. These branches include the branch supporting the Brassicaceae clade, the branch supporting the Asterids clade, and a branch within the Lamiales (marked by an asterisk in Fig. 2-2).

The estimates of dN, dS, and omega from the FRM showed a slight increase in omega values for several Brassicaceae rosette-flowering species (ω = 0.20 vs. 0.11 for the sister and outgroup clades), but it was not elevated for branches associated with *LFY* gene duplication events in the family. The FRM analysis also revealed elevated dN and dS values for two branches of Fabaceae that did not result in an increase of the omega values. There are branches supporting the entire Fabaceae clade (dN = 0.07 vs. 0.015 for the sister and their outgroup clade), and that supported the cpDNA Inverted Repeat-lacking clade (IRLC, dN = 0.03 vs. 0.015) where functional divergence of *LFY* was reported (Hofer et al., 1997; Demason et al., 2001; Champagne et al., 2007). Furthermore, increase of ω was observed in one copy, copy B, of the Lamiales clade in comparing to the other copy (ω = 0.20 vs.

0.04). However, no elevated dN, dS, or omega values was observed for any of the Rosales branches, where gene duplication also occurred.

Test for Positive Selection Identified by FRM Using the Modified BSM.

The analyses using the FRM revealed positive selection at several sites on three branches. These are the Brassicaceae clade in Rosids, the Asterids clade, and the Phrymaceae+Orobanchaceae (P+O) clade in Lamiales. However, when comparing model A (positive selection model, Zhang et al., 2005) with the Null model (no positive selection), the Null hypothesis was rejected at a significant level only for Brassicaceae and Asterids clades, but not for the P+O clade (Table 2-4). Analyses using the Modified BSM confirmed positive selection for the Brassicaceae and Asterids clades.

The results from Modified BSM analyses also revealed positive selection at significant levels on the branch connecting node #4 in Febaceae at two sites (Table 2-5) and one positive selection site for the branch connecting node #2 of Fabaceae at non-significant level (i.e., the selection model was not significantly better than the null model that assumes no positive selection) (Table 2-5). No positive selection was detected for the branches connecting the Lamiales node or the Rosales node (data not shown).

Effects of Sample Size, Sequence Alignment and Ambiguity Sites

The Modified BSM analysis for the original data matrix (Matrix 1) detected positive selection for the Asterids, Brassicaceae, and a node in Fabaceae (as described above). These results were recovered in the analyses of Matrix 2 with reduced sampling from Asterids and Brassicaceae. The positive selection in Asterids, Brassicaceae, and Fabaceae #4 clade (Table 2-6a) in the two analyses differed slightly at the particular sites. Most of the positive sites were detected in both analyses although the significant level varied. For example, for the branch supporting the Asterids clade, sites 8,302,303,381,578, and 579 were identified in analyses of both the original dataset and the reduced dataset; positive selection at site 381 was not statistically significant in the original dataset, but was significant in the reduced dataset (P>99%), whereas the reverse for site 579 (statistically significant in the original dataset at P>95%, but insignificant in the reduced dataset). Analysis of the original dataset with alterted alignment in the variable regions (Matrix 3) identified positive selection only in the Brassicaceae clade (Table 2-6, b). Most of the positive selection sites (e.g., 9, 14, 123, 250, 267, 291, and 564) were detected by analyses of both alignments, but a few positive selection sites were identified from the alternative alignment and not from the original alignment. The original alignments is significantly better in the Likelihood scores of the best tree, lnL=-40250 for the original alignment vs -41121.130 for the alternative alignment.

The phylogenetic analysis of Matrix 4 (removing the regions with uncertain alignment) failed to recover the Asterids clade, thus test for positive selection could not be conducted for the Asterids branch. For the other two branches (Brassicaceae, Fabaceae), a number of positive selection sites were detected for Brassicaceae and a few for two clades of Fabaceae (Table 2-6). The positive sites revealed in the original analysis were recovered except those in the ambiguous regions, which were removed from the matrix.

In our dataset, most of the ambiguity in alignment was between Asterids and Rosids species. We, therefore, performed a separate analysis for the sequence data containing only the Rosids species, focusing on the Brassicaceae clade and Fabaceae clade, in which no ambiguity in alignment was present. The result similarly showed positive selection within the Fabaceae clade, but not within Brassicaceae (because positive selection for Brassicaceae was detected at the root of the clade) (data not shown).

Discussion

Evolutionary Dynamics

Morphological evolution is believed to occur largely through changes in cis-regulatory elements (i.e., through the gain and loss of transcription factor binding sites) while transcription factor proteins are largely evolutionary constrained (Prud'homme et al., 2007; Wray, 2007; Wray et al., 2003). Our evolutionary analysis

of the plant specific transcription factor *LFY* revealed very low rate of nonsysnonymous substitutions (dN), but much higher (one magnitude greater) sysnonymous substitution rates in all three domains of the gene, supporting that the evolution of *LFY* coding regions is constrained for conservation of its function. The low values of the dN/dS ratios further indicate that the evolution of the gene is governed by strong purifying selection. The DNA binding domain (C-domain) of *LFY* was indeed found to have the lowest values of dN and dN/dS ratio, suggesting this domain is under stronger functional constraint than other domains. These results in general agree with the expectation from the cis-regulatory model. Our study, however, further revealed that the evolution of the gene is dynamic, varying among domains and lineages, and sites. The purifying selection force in the N-domain is lower than in the C-domain and much more relaxed in the I-domain (Fig. 2-4). A number of sites in the Asterids, Fabaceae, and Brassicaceae were found to be under positive selections (Table 2-6a). Two of the lineages under positive selection (Brassicaceae and Fabaceae) are associated with functional divergence of *LFY.* The positive selection sites detected in these lineages involve replacements of amino acids of different chemical properties that could have potentially played roles in the functional divergence of the gene (Table 2-6a), e g., directional selection for fixation of S at site 9 (involving changes from non-polar/hydrophobic to polar/hydrophilic amino acid), and G at site 206 and I at site 312 (involving changes from polar to nonpolar and aliphatic amino acid) in Fabaceae (Table 2-7), G(non-

polar/hydrophobic) and R (polar/hydrophilic) at sites 123; G,V (non-polar/hydrophobic) and N,S (polar/hydrophilic) at 234; and C (ionizable, forms covalent cross-link), F,L(non-polar/hydrophobic), and S (polar/hydrophilic) for site 533 in Brassicaceae. These amino acid changes could have substantially influenced the protein structures. These positive selection sites provide a list of target sites for gene functional study of *LFY*.

Heterogeneity in evolutionary rate and selection force among lineages and sites have also been documented in other transcription factor genes, e.g., the myc-like anthocyanin regulatory gene (Fan et al., 2004), the petal organ identity gene *PISTILLATA* (Zhang et al., 2008), the floral regulatory genes *APETALA3* and *APETALA1*, the photosynthetic structural gene *CHLOROPHYLL A/B BINGDING PROTEIN9* (Barrier et al, 2001) and the *MADS-box* gene family (Martinez-Castilla and Alvarez-Buylla, 2003). Some of these studies similarly revealed that although a majority of the protein sequences of transcription factors are conserved by purifying selection, frequent changes at small number of sites in the protein sequence occur in some lineages. Some of the changes may result in new functions that are favored by natural selection. These adaptive changes leave positive selection signals in the sequences and can be detected by molecular evolutionary analyses.

Our results are congruent with previous studies of *LFY* evolution at the smaller scales (Aagaard et al., 2006, Baum et al., 2005, Bomblies et al., 2005). The molecular evolution of *LFY* was recently examined in Brassicaceae and Poaceae

(Baum et al., 2005; Bomblies and Doebly, 2005). These studies similarly found that *LFY* are functionally constrained at both nonsynonymous and synonymous sites. In Brassicaceae, heterogeneous selection force (dN/dS ratios) was detected among lineages, with apparent in duplicated copies and in the lineages showing divergent function of *LFY* (i.e., rosette-flowering) (Sliwinski et al., 2007). However, none of these previous studies used the BSM to detect directional selection in *LFY*. Our analyses with the FRM and Modified BSM revealed that the evolution of *LFY* in Brassicaceae was probably driven by directional selection at one or more sites (indicated by omega >1) at the root of the clade (Table 2-6a). Some of these sites involve amino acid changes in polarity and/or charges of amino acids, suggesting a potential link to functional divergence. The positive selection sites detected in the Brassicaceae clade are evenly distributed in the gene, with site 9 at the 5' flanking region of the N-domain, site 123 in the N-domain, sites 234, 273, 290, and 401 in the I-domain, sites 533 and 564 in the conserved C-domain, and the rest in the 3' flanking region of the C-domain (Table 2-7). It is unknown which of these are truly linked with the rosette-flowering function of LFY or any other functional divergence of the gene. We did not detect positive selection at the root of the rosette-flowering clade.

FLO/LFY ortholog in Fabaceae was well studied for its function in regulating compound leaf development (Hofer et al., 1997; Demason et al., 2001; Champagne et al., 2007). A difference in promoter activities between *LFY* and its ortholog *SGL1*

was detected in these studies (Wang et al., 2008). Evidence from these studies suggested that *FLO/LFY* orthologs in Fabaceae play a role, partially overlapping with other genes to control compound leaf development in the clade lacking the inverted repeat in the chloroplast genome (IRLC) (Champagne et al., 2007). Our analyses detected positive selection for the IRLC branch (#4, Fig. 2-2), a signature of adaptive evolution possibly linked to the new function found in those Legume species. The positive sites involved changes of amino acids in the I-domain with different polarity and charges.

Presence of positive selection in the Asterid branch cannot be certain because the signals disappeared in analysis with alternative alignment or removal of the ambiguous region. Two of the positive sites detected with the best alignment are located in the I-domain, close to the 5' end of C-domain, and the other two sites are located at the 3'end close to the C-domain. This signal suggests possible gene duplication and/or functional diversification of *LFY* in the clade that has not yet been observed. Duplication of *LFY* occurred in Lamiales and Rosales (Aagaard et al. 2006; Fig. 2-2), but our analyses did not detect positive selection for these two clades (data not shown). It is possible that the duplicated *LFY* genes in Lamiales and Rosales are taking an evolutionary trajectory different from functional diversification, e.g., subfunctionalization.

Impaction of Sample Size, Sequence Alignment and Ambiguity Sites

Our results showed that sequence alignment has substantial impact on the result of PAML Modified BSM analysis. This result highlighted the importance of caution in applying the method to identify positive selection. An alteration of alignment may result in changes in the branches, sites, or both that were identified to be under positive selection. Authors are recommended to conduct rigorous analyses of the data including alteration of alignment unless a single best alignment is present without ambiguity. Schneider et al. (2010) similarly reported that estimates of positive selection genes in human were inflated in previous publications by error in sequencing, annotation, and ambiguity in alignment. In our study, after removal of the ambiguity regions, most positive selection sites located outside of the removed region were recovered, suggesting that the method seems to be robust to sequence length.

LFY Duplication

LFY duplication was recognized in Rosales, Lamiales, and Brassicaceae branches. For Rosales and Lamiales, although the duplication event occurred long before present, no positive selection was detected to drive the evolution of the duplicated copies. This suggested that the evolution of duplicated *LFY* copies in Rosales and Lamiales may not be driven by diversifying selection, but the evolution in Brassicaceae where functional divergence of paralogous gene copies have

occurred is likely so. The correlation of positive selection with functional divergence of *LFY* in Brassicaceae and Fabaceae suggests that presence of positive selection of a gene in a lineage may be a good predictor for functional divergence. However, analysis of positive selection must be conducted carefully with consideration of all factors that affect the results.

Acknowledgments

I would like to thank my advisor, Dr. Jenny Xiang for all the guidance on the project, for supporting the research, and for commenting and correcting the manuscript, to my committee members Dr. Heike Saderoff, and Dr. Xie for advice. to Chunmiao Feng for helpful discussion, to Fang Liu and Han-yu Cui for the technical support on computer and computing issue, and the High Performance Computing Cluster facility at NCSU for offering an efficient platform for data analysis.

REFERENCES

Aagaard J.E., Willis J.H., Phillips P.C., 2006. Relaxed Selection Among Duplicate Floral Regulatory Genes in Lamiales. *Journal of Molecular Evolution.* **63**:493–503

Aagaard J.E., Olmstead R.G., Willis J.H. and Phillips P.C. 2005. Duplication of floral regulatory genes in the Lamiales. *American Journal of Botany* **92:** 1284-1293.

Adkins J.A., Williamson J.D. and Werner D.J. Cloning and expression analysis of *FLORICAULA/LEAFY* homologs in Buddleja davidii. Unpublished.

Allnutt,G.V., Rogers,H.J., Francis,D. and Herbert,R.J., 2007. A *LEAFY-like* gene in the long-day plant, *Silene coeli-rosa* is dramatically up-regulated in evoked shoot apical meristems but does not complement the Arabidopsis *LFY* mutant. *Journal of Experimental Botany* **58**:2249-2259

Barrier M., Robichaux R.H., Purugganan M.D., 2001. Accelerated regulatory gene evolution in an adaptive radiation. *Proceedings of the national academy of sciences* **98:** 10208–10213.

Baum D.A., Yoon H.S., Oldham R.L., 2005. Molecular evolution of the transcription factor LEAFY in Brassicaceae. *Molecular phylogenetics and evolution*. **37**:1–14

Benlloch R., Berbel A., Serrano-Mislata A., and Madueño F., 2007. Floral Initiation and inflorescence architecture: A comparative view. *Annals of botany* **100:** 659-676

Blazquez A.M., 2000. Flower development pathways. *Journal of Cell Science*. **113**:3547-3548

Blázquez A. M., Soowal N. L., Lee I. and Weigel D. 1997. *LEAFY* expression and flower initiation in Arabidopsis. *Development* **124**:3835-3844.

Bomblies K. , Doebley J.F., 2005. Molecular Evolution of FLORICAULA/LEAFY Orthologs in the Andropogoneae (Poaceae). *Molecular Biology and Evolution*. **22**:1082-1094.

Bradley D., RatcliVe O., Vincent C., Carpenter R., Coen, E., 1997. Inflorescence commitment and architecture in Arabidopsis. *Science* **275:**80–83.

Bradley D., Carpenter R., Copsey L., Vincent C., Rothstein S., Coen E., 1996a. Control of inflorescence architecture in Antirrhinum. *Nature* **379**:791–797.

Bradley D., Vincent C., Carpenter R., Coen E., 1996b. Pathways for inflorescence and floral induction in Antirrhinum. *Development* **122:**1535–1544.

Busch A.G. and Gleissberg,S. 2003. *EcFLO*, a *FLORICAULA-like* gene from Eschscholzia californica is expressed during organogenesis at the vegetative shoot apex. *Planta* **217**(6):841-848

Carpenter R and Coen ES. 1990. Floral homeotic mutations produced by transposon-mutagenesis in Antirrhinum majus. *Genes and Development* **4:**1483-1493.

Champagnea C., Golibera T., Wojciechowskib M., Meia R., Townsleya B., Wang K., Pazc M., Geetad R. and Sinhaa N.(2007) Compound Leaf Development and Evolution in the Legumes. *The Plant Cell* **19:**3369-3378

Chen Y.L., and Chen Y.K., Cloning and characterization of the LEAFY homolog from longan. Unpublished.

Coen E.S., Romero J.M., Doyle S., Elliott R., Murphy G., Carpenter R., 1990. *Floricaula* – a homeotic gene required for flower development in *Antirrhinum majus. Cell* **63:** 1311-1322

Deb Roy S., Saxena M., Bhalla-Sarin N. Brassica juncea *BJLFY* mRNA for LEAFY. Unpublished.

DeMason D., Schmidt R., 2001. Roles of the Unigene in Shoot and Leaf Development of Pea (Pisum sativum): Phenotypic Characterization and Leaf Development in the uni and uni-tac Mutants. *International Journal of Plant Sciences* **162:**1033-1051

Dong Z.C., Zhao Z., Liu C.W., Luo J.H., Yang J., Huang W.H.,Hu X.H., Wang T.L. and Luo D., 2005. Floral patterning in Lotus japonicas. *Plant physiology* **137:** 1272-1282.

Dornelas M.C., Amaral W. and Rodriguez A.P., 2004. *EgLFY*, the Eucalyptus grandis homolog of the Arabidopsis gene LEAFY is expressed in reproductive and vegetative tissues, *Brazilian Journal of Plant Physiology* **16:** 105-114

Dornelas M.C., and Rodriguez A.P., 2005. The rubber tree (Hevea brasiliensis Muell. Arg.) homologue of the LEAFY/FLORICAULA gene is preferentially expressed in both male and female floral meristems. *Journal of experimental botany* **56:** 1965-1974

Dornelas M.C., and Rodriguez A.P., 2006. The tropical cedar tree (*Cedrela fissilis* Vell., Meliaceae) homolog of the Arabidopsis LEAFY gene is expressed in reproductive tissues and can complement Arabidopsis *leafy* mutants. *Planta* **223**: 306-314.

Edgar R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Research* **32**:1792-97.

Esumi T., Tao R., and Yonemori K., 2005. Isolation of *LEAFY* and *TERMINAL FLOWER 1* homologues from six fruit tree species in the subfamily Maloideae of the Rosaceae. *Sexual plant reproduction* **17:** 277-287.

Fares M.A., Bezemer D., Moya A., Marin I., 2003. Selection on coding regions determined Hox7 genes evolution. *Molecular biology and evolution* **20:** 2104-2112

Fan C.Z., Purugganan M.D., Thomas D.T., Wiegmann B.M., and Xiang Q.Y., 2004. Heterogeneous evolution of the *Myc-like* Anthocyanin regulatory gene and its phylogenetic utility in *Cornus* L. (Cornaceae). *Molecular Phylogenetics and Evolution*, **33:** 580-594.

Fan C.Z., Xiang Q.Y., Remington D.L., Purugganan M.D., Wiegmann B.M., 2007. Evolutionary patterns in the antR-Cor gene in the dwarf dogwood complex (Cornus, Cornacese). *Genetica* **130:** 19-34.

Fernando D.D., and Zhang S., 2005. Constitutive expression of the *SAP1* gene from willow (*Salix discolor*) causes early flowering in*Arabidopsis thaliana*. *Development genes and evolution* **216:** 19-28.

Gilad Y., Oshlack A., Smyth G.K., Speed T.P., and White P., 2006. Expression profiling in primates reveals a rapid evolution of human transcription factors. *Nature* **440:** 242-245.

Guirao-Rico S., Aguade M., 2009. Positive selection has driven the evolution of the *Drosophila* Insulin-like receptor (*InR*) at different timescales. *Molecular biology and evolution* **26:** 1723-1732

Guo C.L., Wang C.W. and Wang H.Z., Molecular cloning and sequencing of leafy gene in peach. Unpublished.

Guo J.L., and Yang Q. Solanum tuberosum LEAFY/FLORICAULA (*LFY*) genomic sequence. Unpublished.

Haughn G.W., Somerville C.R., 1988. Genetic control of morphogenesis in Arabidopsis. *Developmental genetics* **9 :**73-89

He X.H., Guo Y.Z., Li Y.R., Study on *LFY* homolog genes of several fruit trees. Unpublished

Himi S., Sano R., Nishiyama T., Tanahashi T., Kato M., Ueda K., Hasebe M., 2001. Evolution of *MADS-Box* gene Induced by *FLO/LFY* gene. *Journal of Molecular Evolution* **53**: 387-393.

Hofer J., Turner L., Hellens R., Ambrose M., Matthews P., Michael A., Ellis N., 1997 *UNIFOLIATA* regulates leaf and flower morphogenesis in pea. *Current Biology* **7:** 581–587

Huala E. and Sussex I. M., 1992. *LEAFY* interacts with floral homeotic genes to regulate Arabidopsis floral development. *Plant Cell* **4**: 901-913.

Huang J.Q., Wang Z.J., Zheng B.S., Huang Y.J., and Zeng Y.R., Direct submission, gi: 114229060

Huelsenbeck J.P. and Ronquist F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755.

Hugall, A.F., Lee, M.S.Y., 2007. The likelihood node density effect and consequences for evolutionary studies of molecular rates. *Evolution* **61:** 2293–2307.

Kelly A.J., Bonnlander M.B. and Meeks-Wagner D.R.1995. *NFL*, the tobacco homolog of *FLORICAULA* and *LEAFY*, is transcriptionally expressed in both vegetative and floral meristems. *Plant Cell* **7:** 225-234

Kim and Kim, Pepper ortholog of *FLO/LFY*. Unpublished

Kumar S., Tamura K., Nei M., 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics* **5:**150-163.

Liu F.Q., Zhu G.L., Luo D., Wu X.Y. and Xu Z.H., 1999. Cloning and analysis of CFL-a *LFY*-like gene from cucumber. *Acta Botanica sinica* **41:** 813-819.

Liu T., and Hu Y., A leafy-like clone (*Cmlfy*) from Castanea mollissima. Unpublished.

Loennig W.E., Stueber K., Saedler H. and Kim J., 2007. Biodiversity and Dollo's law: to what extend can the phenotypic differences between Antirrhinum majus and Misopates orontium be bridged by mutagenesis? *Bioremediation, Biodiversity and Bioavailability* **1:** 1-30.

Ma Y., Dai S., Fang X., Chen F. and Shen Y. *FLO/LFY* homolog gene in *Dendranthema lavandulifolium*. Unpublished

Maizel A., Busch M.A., Tanahashi T., Perkovic J., KatoM. , Hasebe M.,Weigel D., 2005. The Floral Regulator *LEAFY* Evolves by Substitutions in the DNA Binding Domain. *Science*. **308**: 260-263.

Martinez-Castilla L.P. and Alvarez-Buylla E.R., 2003. Adaptive evolution in the Arabidopsis *MADS-box* gene family inferred from its complete resolved phylogeny. *Proceeding of the national academy of sciences of the United States of America* **100:** 13407-13412.

Molinero-Rosales N., Jamilena M., Zurita S., Gomez P., Capel J. and Lozano R. 1999. *FALSIFLORA*, the tomato orthologue of *FLORICAULA* and *LEAFY*, controls flowering time and floral meristem identity. *Plant Journal* **20:** 685-693.

Nazawa M., Suzuki Y., and Nei, M. 2009a. Reliabilities of identifying positive selection by the branch-site and the site-prediction methods. *Proceedings of the National Academy of Sciences of the United States of America.* **106:** 6700-6705

Nazawa M., Suzuki Y., and Nei, M. 2009b. Response to Yang et al.: Problems with Bayesian methods of detecting positive selection at the DNA sequence level. *Proceedings of the National Academy of Sciences of the United States of America.* **106:** E96

Nei M., Gojobori T. 1986 Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular biology and evolution* **3:** 418-426

Nielsen R., and Yang Z., 1998 Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148:**929–936.

Nilsson O., Lee I., Blázquez A. M., and Weigel D., 1998 Flowering-Time Genes Modulate the Response to *LEAFY* Activity. *Genetics* **150**: 403-410.

Nishimoto Y., Ohnishi O. and Hasegawa M. 2003. Topological incongruence between nuclear and chloroplast DNA trees suggesting hybridization in the

urophyllum group of the genus *Fagopyrum* (Polygonaceae). *Genes & Genetic system* **78:** 139-153

Olsen K.M., Womack A., Garrett A.R. Suddith J.I. and Purugganan M.D., Contrasting evolutionary forces in the Arabidopsis thaliana floral developmental pathway. Unpublished.

Ordidge M., Chiurugwi T., Tooke F. and Battey N.H., 2005. *LEAFY, TERMINAL FLOWER1* and *AGAMOUS* are functionally conserved but do not regulate terminal flowering and floral determinacy in Impatiens balsamina. *Plant Journal* **44:** 985-1000

Oshima S., and Nomura K., 2008. *RsLFY*, a *LEAFY* homologue gene in radish (Raphanus sativus), is continuously expressed in vegetative, reproductive and seed development. *Plant biotechnology journal* **25:** 579-582

Parcy F., 2005 Flowering: a time for integration. *International Journal of Developmental Biology.* **49:**585-593

Pillitteri L.J., Walling I.L., and Lovatt C.J., 2004. Isolation and characterization of LEAFY and APETALA1 homologues from Citrus sinensis L. Osbeck 'Washington'. *Journal of the American society for horticultural science* **129:** 846-856

Ponniah K. *AML*, the *LEAFY* homolog of Acacia mangium. Unpublished

Ponniah K., and Kulaveerasingam H. *AML*, the *LEAFY* homolog of Acacia mangium. Unpublished

Posada D. and Buckley T.R., 2004. Model selection and model averaging in phylogenetics: advantages of the AIC and Bayesian approaches over likelihood ratio tests. *Systematic Biology* **53**: 793-808

Posada D. and Crandall K.A. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**:817-818.

Prud'homme B., Gompel N., Carroll S.B. 2007. Emerging principles of regulatory evolution. *Proceedings of the national academy of sciences of the United States of America* **104:** 8605-8612

Prusinska J.M., Kukla I., Chab D., Storchova H., and Wagner E., Identification and characterization of Chenopodium rubrum CrFL gene involved in the photoperiodic flower induction. Unpublished

Rambaut A., Drummond A.J., 2007. Tracer v1.4, Available from http://beast.bio.ed.ac.uk/Tracer.

Rockman M.V., Hahn M.W., Soranzo N., Goldstein D.B., and Wray G.A., 2003. Positive selection on a human-specific transcription factor binding site regulating *IL4* Expression. *Current Biology* **13:** 2118-2123.

Ronquist F., Huelsenbeck J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.

Rosales M. N., Jamilena M., Zurita S., Gómez P., Capel J. and Lozano R., 1999. *FALSIFLORA*, the tomato orthologue of *FLORICAULA* and *LEAFY*, controls flowering time and floral meristem identity. *The Plant Journal* **20**: 685-693.

Schultz E. A. and Haughn G. W., 1991. *LEAFY*, a homeotic gene that regulates inflorescence development in Arabidopsis. *Plant Cell* **3**: 771-781.

Shu G., Amaral W., Hileman L.C., Baum D.A., 2000. *LEAFY* and the evolution of rosette flowering in violet cress (*Jonopsidium acaule,*Brassicaceae). *American Journal of Botany* **87:** 634-641.

Silva C.N., Sanchez A.M. and Oliveira M.M., PrdFL, an almond putative homolog of *FLORICAULA/LEAFY* genes and its role in flowering. Unpublished.

Sliwinski M., Bosch J., Yoon H., Balthazar M. and Baum D., 2007 The role of two LEAFY paralogs from Idahoa scapigera (Brassicaceae) in the evolution of a derived plant architecture. *The Plant Journal* **51:** 211–219

Souer E., van der Krol A., Kloos D., Spelt C., Bliek M., Mol J. and Koes R. 1998. Genetic control of branching pattern and floral identity during Petunia inflorescence development. *Development* 125: 733-742

Suzuki, Y., and T. Gojobori. 1999. A method for detecting positive selection at single amino acid sites. *Molecular Biology Evolution* **16:**1315–1328.

Swofford D.L., 2002. PAUP*: phylogenetic analysis using parsimony (*and other methods). version 4.10b. Sinauer Associates, Sunderland, Mass.

Van de Peer Y., Taylor J.S., Braasch I., Meyer A., 2001. The ghost of selection past: rates of evolution and functional divergence of anciently duplicated genes. *Journal of Molecular Evolution* **53:** 436-446.

Wada M., Cao Q.F., Kotoda N., Soejima J. and Masuda T., 2002. Apple has two orthologues of FLORICAULA/LEAFY involved in flowering. *Plant molecular biology* **49:** 567-577

Wang C.N., Moller M., and Cronk Q.C., 2004. Altered expression of *GFLO*, the Gesneriaceae homologue of *FLORICAULA/LEAFY*, is associated with the transition to bulbil formation in *Titanotrichum oldhamii*. *Development genes and evolution* **214:** 122-127.

Wang,H., Chen,J., Wen,J., Tadege,M., Li,G., Liu,Y., Mysore,K.S.,Ratet,P. and Chen,R., 2008. Control of Compound Leaf Development by *FLORICAULA/LEAFY* Ortholog *SINGLE LEAFLET1* in *Medicago truncatula*. *Plant Physiology*,**146:** 1759–1772

Wang Y., Ballard H.E., and Wyatt S.E., direct submission. Gi:157649149

Wanger G.P., Lynch V.J., 2008. The gene regulatory logic of transcription factor evolution. *Trends in ecology and evolution* **23:** 377-385.

Wei Q. and Hu B., *CBFLO1*, a new member of *FLO/LFY* homologous genes in Coleus blumei. Unpublished.

Weigel D., Alvarez J., Smyth D. R., Yanofsky M. F. and Meyerowitz E.M., 1992. *LEAFY* controls floral meristem identity in Arabidopsis. *Cell* **69**: 843-859.

Wray G.A., 2007. The evolutionary significance of cis-regulatory mutations. *Nature reviews genetics* **8:** 206-216.

Wray G.A., Hahn M.W., Abouheif E., Balhoff J.P., Pizer M., Rockman M.V., Romano L.A. 2003. The evolution of transcriptional regulation in eukaryotes. *Molecular biology and evolution* **20:** 1377-1419

Xiang Q.Y., Thorne J.L., Seo T.K., Zhang W.H., Thomas D.T. and Ricklefs R.E., 2008. Rates of nucleotide substitution in Cornaceae (Cornales)—Pattern of variation and underlying causal factors. *Molecular phylogenetics and evolution.* **49:** 327-342.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood *Computer Applications in BioSciences* **13**:555-556

Yang Z. 2000. Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza birus A. *Journal of Molecular Evolution* **51:**423-432.

Yang Z. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Molecular biology and evolution* **19:** 908-917

Yang Z. 2007. PAML: a program package for phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24**:1586-1591

Yang Z, Nielsen R, and Goldman N. 2009 In defense of statistical methods for detecting positive selection. *Proceedings of the National Academy of Sciences of the United States of America.* **106:** E95

Yang Z., Wong W., Nielsen R., 2005. Bayes empirical bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution* **22:**1107–1118

Yoon H.S., and Baum D.A., 2004. Transgenic study of parallelism in plant morphological evolution. *Proceedings of the national academy of sciences* **101:** 6524–6529.

Yu Q., Moore P.H., Albert H.H., Roader A.H.K. and Ming R., 2005. Cloning and characterization of a *FLORICAULA/LEAFY* ortholog, *PFL*, in polygamous papaya. *Cell Research* **15:** 576-584.

Zhang W.H., Xiang Q.Y.,Thomas D.T., Wiegmann B.M., Frohlich M.W., Soltis D.E., 2008. Molecular evolution of *PISTILLATA*-like genes in the dogwood genus *Cornus* (Cornaceae). *Molecular Phylogenetics and Evolution* **47**:175–195

Zhang J., Nielsen R., and Yang Z., 2005. Evaluation of an Improved Modefied Branch-Stie Likelihood Method for Detecting Positive Selection at the Molecular Level. *Molecular Biology and Evolution* **22:**2472–2479

Table 2-1 Species and sequence sources included in the study.

Order or Clade	Family	Species	Accession	Author
Ranunculales				
	Papaveraceae	*Eschscholzia californica*	**gi:30313798**	Busch and Gleissberg 2003
Caryophyllales				
	Caryophyllaceae	*Silene coeli-rosa*	**gi:20067656**	Allnutt et al. 2007
		Chenopodium rubrum	**gi:166851670**	Prusinska et al. unpublished
	Polygonaceae	*Fagopyrum gilesii*	**gi:31088148**	Nishimoto et al. 2003
		Fagopyrum rubifolium	**gi:31088146**	Nishimoto et al. 2003
Asterids				
	Asteraceae	*Chrysanthemum lavandulifolium*	**gi:49072849**	Ma et al. unpublished
	Balsaminaceae	*Impatiens balsamina*	**gi:82775185**	Ordidge et al. 2005

Table 2-1 cont.

Gesneriaceae	*Titanotrichum oldhamii*	**gi:46404424**	Wang et al. 2004
Lamiaceae	*Salvia coccinea*	**gi:42795322**	Aagaard et al. 2005
	Solenostemon scutellarioides	**gi:145581370**	Wei and Hu, unpublished
Phrymaceae	*Mazus reptans*	**gi:42795264**	Aagaard et al. 2006
Oleaceae	*Syringa vulgaris*	**gi:42795334**	Aagaard et al. 2005
Orobanchaceae	*Pedicularis groenlandica*	**gi:42795338**	Aagaard et al. 2005
Paulowniaceae	*Paulownia tomentosa*	**gi:42795328**	Aagaard et al. 2005
	Paulownia tomentosa	**gi:42795340**	Aagaard et al. 2005
Phrymaceae	*Leucocarpus perfoliatus*	**gi:42795268**	Aagaard et al. 2006
	Leucocarpus perfoliatus	**gi:42795272**	Aagaard et al. 2006
	Mimulus guttatus	**gi:42795332**	Aagaard et al. 2005
	Mimulus guttatus	**gi:42795344**	Aagaard et al. 2005
	Mimulus lewisii	**gi:42795330**	Aagaard et al. 2005
	Mimulus lewisii	**gi:42795342**	Aagaard et al. 2005
	Mimulus kelloggii	**gi:42795270**	Aagaard et al. 2005

Table 2-1 cont.

		Mimulus ringens	**gi:42795266**	Aagaard et al. 2005
	Plantaginaceae	*Antirrhinum majus*	**gi:166429**	Coen et al. 1990
		Chelone glabra	**gi:42795318**	Aagaard et al. 2005
		Misopates orontium	**gi:82734196**	Loennig et al. 2007
	Scrophulariaceae	*Buddleja davidii*	**gi:76495760**	Adkins et al. unpublished
	Solanaceae	*Capsicum annuum*	**gi:151933913**	Kim and Kim, unpublished
		Lycopersicon esculentum	**gi:7658232**	Molinero-Rosales et al. 1999
		Lycopersicon esculentum	**gi:7658236**	Molinero-Rosales et al. 1999
		Nicotiana tabacum	**gi:561687**	Kelly et al. 1995
		Petunia x hybrida	**gi:2625049**	Souer et al. 1998
		Solanum tuberosum	**gi:166012290**	Guo and Yang, unpublished
	Verbenaceae	*Verbena officinalis*	**gi:42795320**	Aagaard et al. 2005
		Verbena officinalis	**gi:42795336**	Aagaard et al. 2005
Rosids				
	Brassicaceae	*Arabidopsis lyrata*	**gi:20563244**	Olsen et al. unpublished

Table 2-1 cont.

	Arabidopsis thaliana	**gi:18424517**	
	Barbarea vulgaris	**gi:58003768**	Baum et al. 2005
	Boechera stricta	**gi:58339136**	Baum et al. 2005
	Brassica juncea	**gi:52222044**	Deb Roy et al. unpublished
	Brassica juncea	**gi:93278160**	Deb Roy et al. unpublished
	Capsella bursa-pastoris	**gi:58003766**	Baum et al. 2005
	Cochlearia officinalis	**gi:58081326**	Baum et al. 2005
	Cochlearia officinalis	**gi:58081328**	Baum et al. 2005
	Iberis amara	**gi:75915139**	Baum et al. 2005
	Idahoa scapigera	**gi:40388193**	Yoon and Baum, 2004
	Idahoa scapigera	**gi:40388194**	Yoon and Baum, 2004
	Ionopsidium abulense	**gi:75915151**	Baum et al. 2005
	Ionopsidium acaule	**gi:37788345**	Yoon and Baum, 2004
	Ionopsidium acaule	**gi:6003581**	Shu et al. 2000
	Ionopsidium acaule	**gi:6003579**	Shu et al. 2000

Table 2-1 cont.

	Leavenworthia crassa	**gi:37788347**	Yoon and Baum, 2004
	Raphanus sativus	**gi:219565988**	Oshima and Nomura, 2008
	Raphanus sativus	**gi:219565990**	Oshima and Nomura, 2008
	Selenia aurea	**gi:58339138**	Baum et al. 2005
	Stanleya pinnata	**gi:58339141**	Baum et al. 2005
	Streptanthus glandulosus	**gi:75915130**	Baum et al. 2005
	Streptanthus glandulosus	**gi:75915132**	Baum et al. 2005
Caricaceae	*Carica papaya*	**gi:66864714**	Yu et al. 2005
Cucurbitaceae	*Cucumis sativus*	**gi:3776288**	Liu et al. 1999
Euphorbiaceae	*Hevea brasiliensis*	**gi:49183025**	Dornelas and Rodriguez, 2005
Fabaceae	*Acacia auriculiformis*	**gi:29150721**	Ponniah unpublished
	Acacia mangium	**gi:29150719**	Ponniah and Kulaveerasingam, unpublished
	Glycine max	**gi:145700461**	Champagne et al. 2007
	Glycine max	**gi:145700459**	Champagne et al. 2007

Table 2-1 cont.

	Lotus japonicus	**gi:60100333**	Dong et al. 2005
	Medicago truncatula	**gi:60502224**	Wang et al. 2008
	Pisum sativum	**gi:9864838**	Hofer et al. 1997
Fagaceae	*Castanea mollissima*	**gi:82561364**	Liu and Hu, unpublished
Juglandaceae	*Carya cathayensis*	**gi:114229060**	Huang et al. unpublished
Meliaceae	*Cedrela fissilis*	**gi:48762787**	Dornelas and Rodriguez, 2006
Myrtaceae	*Eucalyptus grandis*	**gi:49615780**	Dornelas et al. 2004
Rosaceae	*Chaenomeles sinensis*	**gi:42491287**	Esumi et al. 2005
	Chaenomeles sinensis	**gi:42491299**	Esumi et al. 2005
	Cydonia oblonga	**gi:42491297**	Esumi et al. 2005
	Cydonia oblonga	**gi:42491285**	Esumi et al. 2005
	Eriobotrya japonica	**gi:42491301**	Esumi et al. 2005
	Eriobotrya japonica	**gi:42491289**	Esumi et al. 2005
	Malus X domestica	**gi:18147634**	Wada et al. 2002
	Malus X domestica	**gi:18147636**	Wada et al. 2002

Table 2-1 cont.

	Malus X domestica	**gi:42491279**	Esumi et al. 2005
	Malus X domestica	**gi:42491291**	Esumi et al. 2005
	Prunus dulcis	**gi:63094574**	Silva et al. unpublished
	Prunus persica	**gi:166025438**	Guo et al. unpublished
	Pyrus communis	**gi:42491283**	Esumi et al. 2005
	Pyrus communis	**gi:42491295**	Esumi et al. 2005
	Pyrus pyrifolia	**gi:42491293**	Esumi et al. 2005
	Pyrus pyrifolia	**gi:42491281**	Esumi et al. 2005
Rutaceae	*Citrus maxima*	**gi:116282361**	He et al. unpublished
	Citrus reticulate	**gi:116282363**	He et al. unpublished
	Citrus sinensis	**gi:37703727**	Pillitteri et al. 2004
	Citrus unshiu	**gi:116282359**	He et al. unpublished
	Clausena lansium	**gi:113968339**	He et al. unpublished
	Fortunella crassifolia	**gi:113968341**	He et al. unpublished
Salicaceae	*Salix discolor*	**gi:29423801**	Fernando and Zhang, 2005

Table 2-1 cont.

	Sapindaceae	*Dimocarpus longan*	**gi:76152079**	Chen and Chen, unpublished
	Violaceae	*Viola pubescens*	**gi:157649149**	Wang et al. unpublished

Table 2-2 Results from analysis of adaptive selection of *FLO/LFY*-like gene in the Eudicots using the Site Model.

Model	p[a]	lnL[b]	dN/dS	Estimates of parameters[c]	Positively selected sites[d]
M0: one-ratio	1	-32050.929	0.0310	ω=0.0897	None
M1: neutral (K=2)	1	-31627.658	0.2947	P_0=0.756,P_1=0.244 (ω_1=1.000)	None
M2: selection (K=3)	3	-31627.658	0.2947	P_0=0.756,P_1=0.151, P_2=0.093 (ω_1= ω_2=1.000)	None
M3:discrete (K=10)	5	-30830.858	0.1469	P_9=0.0034, ω_9=1.770	104*
M7:β	2	-30837.511	0.1456	P=0.433 q=2.472	None
M8:β and ω	4	-30837.125		P_0=0.997, P=0.437, q=2.545 P_1=0.0024, ω=1.828	104*

a Number of parameters in the model

b Log-likelihood scores

c P_i denotes proportion of site falling in site class ω_i

d Sites potentially under positive selection identified with a posterior probability

*>95%

Table 2-3 Results from comparison of models (see Table 2-2) using the likelihood ratio tests (LRTs)

LRT	Degrees of freedom	χ^2 critical value(5%)	χ^2 critical value(1%)	2 × Δl
M0 vs. M3	4	9.49	13.28	2440.142**
M1 vs. M2	2	5.99	9.21	0
M7 vs. M8	2	5.99	9.21	0.772

Table 2-4 Results from analyses of testing positive selection using the modified Branch-Site Model for the Brassicaceae branch, the Asterids branch, and the P+O branch. Asterisks indicate that the test model significantly differs from the null model at p = 0.05 (one asterisk) and p = 0.01 (two asterisks) respectively, based on likelihood ratio tests. "Δl" indicates the difference of likelihood scores between the two models. χ^2 critical value is 5.99 for α=0.05, and 9.21 for α=0.01.

Branch	**model**	**lnL**	**Positive Selection Sites**	**2 × Δl**
Brassicaceae	Model A	-31604.310	9*,123*,234*,273*,290*,401*,533**,564*,579*	7.73*
	Null Model	-31608.175		
Asterids	Model A	-31614.113	302(0.949),381(0.996**),579(0.946)	16.85**
	Null Model	-31622.538		
P+O	Model A	-31621.599	534(0.973*),535(0.974*).542(0.972*)	0.072
	Null Model	-31621.635		

Table 2-5. Results from analyses of testing positive selection using the modified Branch-Site Model for the Fabaceae branches. Asterisks indicate that the test model significantly differs from the null model at P > 95% (one asterisk) and P > 99% (two asterisks) respectively, based on likelihood ratio tests. "Δl" indicates the difference of likelihood scores between the two models. When degree of freedom = 2, χ^2 critical value is 5.99 for α=0.05, and 9.21 for α=0.01.

branch	model	lnL	Positive Selection Sites	2 × Δl
#1	Model A	-31627.658	None	0
	Null Model	-31627.658		
#2	Model A	-31627.658	452 (0.976*)	-4.44
	Null Model	-31625.438		
#3	Model A	-31627.658	None	-0.076
	Null Model	-31627.620		
#4	Model A	-31622.430	312 (0.990*)	6.812*
	Null Model	-31625.836		

Table 2-6. Comparisons of results from analyses with different taxon sampling, alternative sequence alignments, and removal of ambiguity sites. The values of Δl indicates differences in the likelihood scores between model A (i.e. model allows positive selection present in foreground branches) and the Null model (i.e. omega less than or equal to one for foreground branches). Sites in bold faces are those detected in both analyses. Sites indicated with one asterisk are positive selection sites at a significance level with a P > 95% and sites indicated with two asterisks are positive selection sites at a significance level with a P > 99%.

a) Impact of sample size

Branch	2*Δl	Positive selection Sites (Matrix I - Original)	2*Δl	Positive selection Sites (Matrix 2 -Reduced sample size)
Asterids	15.3**	**8**;84;**302**;**303**;369**;**381**;567;**578**;**579****;586;587**;614	16.85**	**8**;99;260;299;**302**;**303**;323;325;374;**381****;382;388; **578**;**579**
Brassicaceae	10.0**	5;**9***;**14**;98;**123**;129;188;212;**234**;**250**;**251**;**267**;**273***;**290***;**291**;331;373;389;**401***;476;519;**533**;**564***;**578**;**579**;**585**;590*593;**607***611*;613;614	7.73*	**9***;**14**;**123***;**234***;237;**250**;**251**;**267**;**273***;**290***;**291**;**401***;452;530;**533****;**564***;**578**;**579***;**585**;595;598;**607**;
Fabaceae 2	1.9	**198**;**262**;**452***;**502**	3.86	**198**;**262**;**452***;**502**;588
Fabaceae 4	7.1*	**206**,**312****	6.81*	**206***;**312****

Table 2-6 cont.

b) Impact of sequence alignment

Branch	2*Δl	Positive selection Sites (Matrix 1 - Original)	2*Δl	Positive selection Sites (Matrix 3 - Different alignment)	Corresponding sites in Matrix 1 of sites detected from Matrix 3
Asterids	15.33**	**8**;84;302;303;369**;381;**567**;578;579**;586;587**;614	0	8;145;259;260;262*;263*;298;551	**8**;146;267;268;270*;271*;300;**567**
Brassicaceae	10.02**	5;**9*;14**;98;**123**;129;188;212;234;**250**;251;**267**;273*;290*;**291**;331;373;389;401*;476;519;533;**564***;578;579;585;590*593;607*611*;613;614	10.27**	9*;14;50*;110;122**;237;259;284*;432;548*	**9*;14**;56*;109;**123****;**250**;**267**;**291***;452;**564***
Fabaceae 2	1.94	**198**;**262**;**452***;**502**	6.19*	196;254;432**;483	**198**;**262**;**452****;**502**
Fabaceae 4	7.10*	**206**,312**	1.35	204*	**206***

Table 2-6 cont.

c) Impact of ambiguity sites

Branch	2*Δl	Positive selection Sites (Matrix 1 - Original)	2*Δl	Positive selection Sites (Matrix 4 - Ambiguity sites removed)	Corresponding sites in Matrix 1 of sites detected from Matrix 4
Brassicaceae	10.02**	5;**9***;**14**;98;**123**;129;188;212;234;**250**;**251**;267;**273***;290*;291;331;373;389;401*;476;519;**533**;**564***;578;579;585;590*593;607*611*;613;614	3.17	9;14;29*;156;157;179;319*;350	**9**;**14**;**123***;**250**;**251**;**273**;**533***;**564**
Fabaceae 2	1.94	**198;262;452*;502**	4.11	104;168;238*;288;	**198;262;452*;502**
Fabaceae 4	7.10*	**206**,312**	2.54	112*	**206***

Table 2-7. Comparison of the foreground and background translated amino acids from the positive selection sites in Brassicaceae, Asterids, and Fabaceae branch #4 detected from analyses of Matrix 1 (original) or Matrix 2 (with reduced sampling) at the significant levels. Bold sites indicates those were detected in both analyses at significant levels

Branch	Sites	Foreground amino acid	Background amino acid
Brassicaceae	**9**	**S**	**V,T,A**
	123	G,R	E,Q,H
	234	G,V,N,S	T,N,A,S,I
	273	**N,I,G,A**	**G,T,D**
	290	**T,S**	**V,A,M**
	401	**L,M**	**G**
	533	C,F,L,S	A,G,S,V
	564	**L,M**	**A,V,T,S**
	579	A,T	S,A,T,G,N
	590	I	A,G,S,V
	607	E,G,H	H,Y,Q
	611	S,L	L,M,V
Asterids	*369*	A,C,S,L,T,M	D,N,E,H
	381	T,S,G,A	D
	579	A,G,T	A,T,S
	587	S,T,P	G,S,T
Fabaceae branch #4	*206*	G	H,N,R,L,Y,Q
	312	**I**	**Q,R,D,G,P**

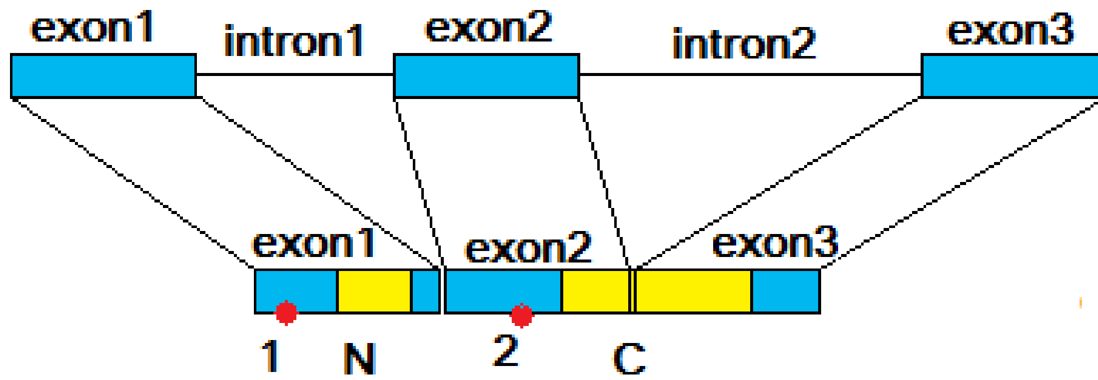
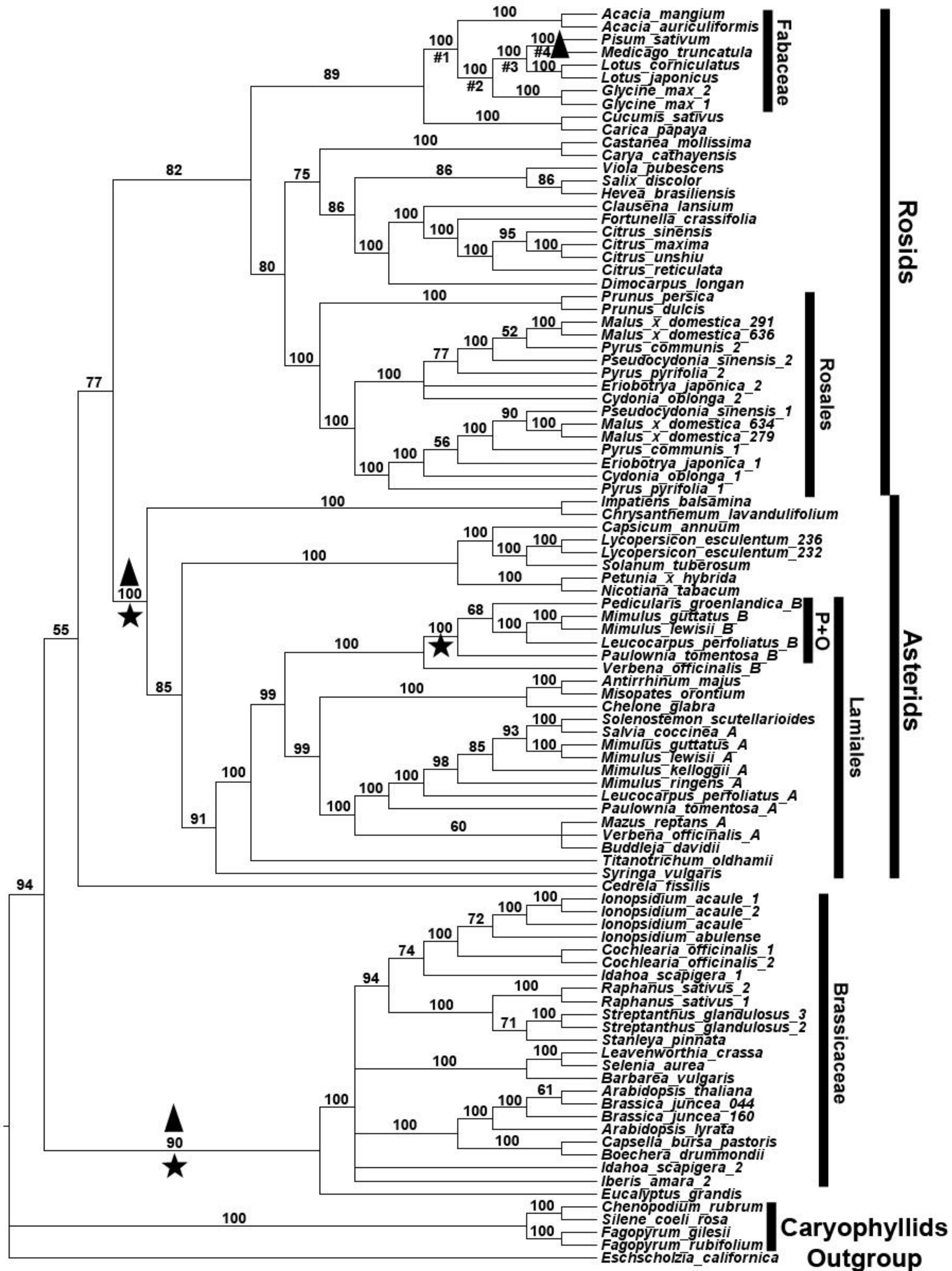


Figure 2-1 Schematic structure of *FLO/LFY*-like gene. The two conserved domains were represented by the yellow block. The red dots mark the location of the two long insertions found in Caryophyllales species.

Fig.2-2 Genealogy of the *FLO/LFY*-like gene in Eudicots inferred from Bayesian analysis using MrBayes 3.1.2 Program. Numbers above branches are Bayesian posterior probabilities. Asterisk labeled the positive selection clade detected using FRM model. Solid triangles on the branches labeled the ones that were confirmed with analyses using the modified BSM model. The #1, #2, #3, and #4 marked for Fabaceae species represents the 4 branches that was tested for positive selection using the modified BSM.


Acacia_mangium
Acacia_auriculiformis
Pisum_sativum
Medicago_truncatula
Lotus_corniculatus
Lotus_japonicus
Glycine_max_2
Glycine_max_1
Cucumis_sativus
Carica_papaya
Castanea_mollissima
Carya_cathayensis
Viola_pubescens
Salix_discolor
Hevea_brasiliensis
Clausena_lansium
Fortunella_crassifolia
Citrus_sinensis
Citrus_maxima
Citrus_unshiu
Citrus_reticulata
Dimocarpus_longan
Prunus_persica
Prunus_dulcis
Malus_x_domestica_291
Malus_x_domestica_636
Pyrus_communis_2
Pseudocydonia_sinensis_2
Pyrus_pyrifolia_2
Eriobotrya_japonica_2
Cydonia_oblonga_2
Pseudocydonia_sinensis_1
Malus_x_domestica_634
Malus_x_domestica_279
Pyrus_communis_1
Eriobotrya_japonica_1
Cydonia_oblonga_1
Pyrus_pyrifolia_1
Impatiens_balsamina
Chrysanthemum_lavandulifolium
Capsicum_annuum
Lycopersicon_esculentum_236
Lycopersicon_esculentum_232
Solanum_tuberosum
Petunia_x_hybrida
Nicotiana_tabacum
Pedicularis_groenlandica_B
Mimulus_guttatus_B
Mimulus_lewisii_B
Leucocarpus_perfoliatus_B
Paulownia_tomentosa_B
Verbena_officinalis_B
Antirrhinum_majus
Misopates_orontium
Chelone_glabra
Solenostemon_scutellarioides
Salvia_coccinea_A
Mimulus_guttatus_A
Mimulus_lewisii_A
Mimulus_kelloggii_A
Mimulus_ringens_A
Leucocarpus_perfoliatus_A
Paulownia_tomentosa_A
Mazus_reptans_A
Verbena_officinalis_A
Buddleja_davidii
Titanotrichum_oldhamii
Syringa_vulgaris
Cedrela_fissilis
Ionopsidium_acaule_1
Ionopsidium_acaule_2
Ionopsidium_acaule
Ionopsidium_abulense
Cochlearia_officinalis_1
Cochlearia_officinalis_2
Idahoa_scapigera_1
Raphanus_sativus_2
Raphanus_sativus_1
Streptanthus_glandulosus_3
Streptanthus_glandulosus_2
Stanleya_pinnata
Leavenworthia_crassa
Selenia_aurea
Barbarea_vulgaris
Arabidopsis_thaliana
Brassica_juncea_044
Brassica_juncea_160
Arabidopsis_lyrata
Capsella_bursa_pastoris
Boechera_drummondii
Idahoa_scapigera_2
Iberis_amara_2
Eucalyptus_grandis
Chenopodium_rubrum
Silene_coeli_rosa
Fagopyrum_gilesii
Fagopyrum_rubifolium
Eschscholzia_californica
Fabaceae
Rosales
Rosids
P+O
Lamiales
Asterids
Brassicaceae
Caryophyllids
Outgroup
#1
#2
#3
#4
100
89
82
75
86
95
80
52
77
90
56
55
94
85
91
99
68
93
98
60
74
72
71
61

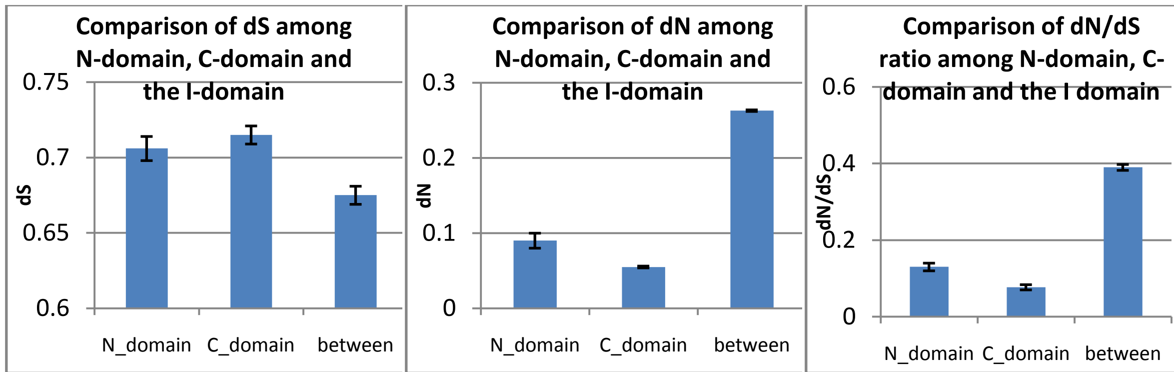
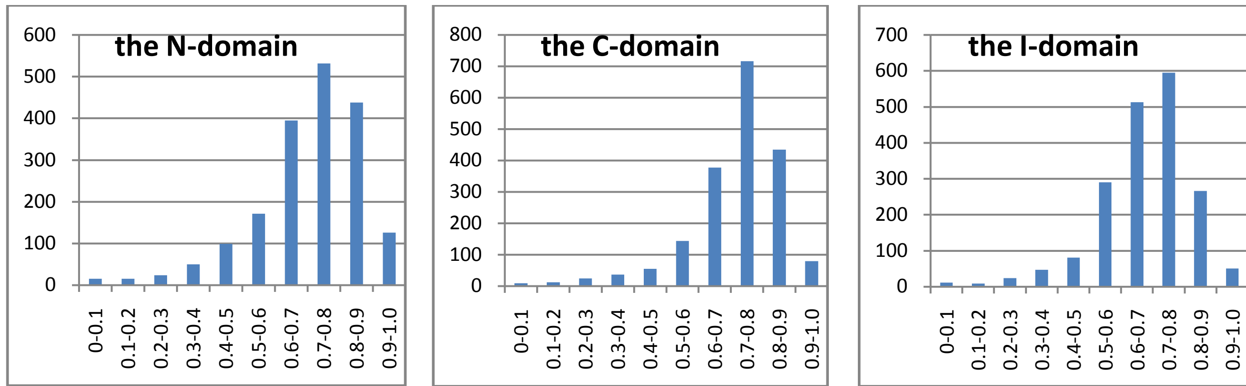
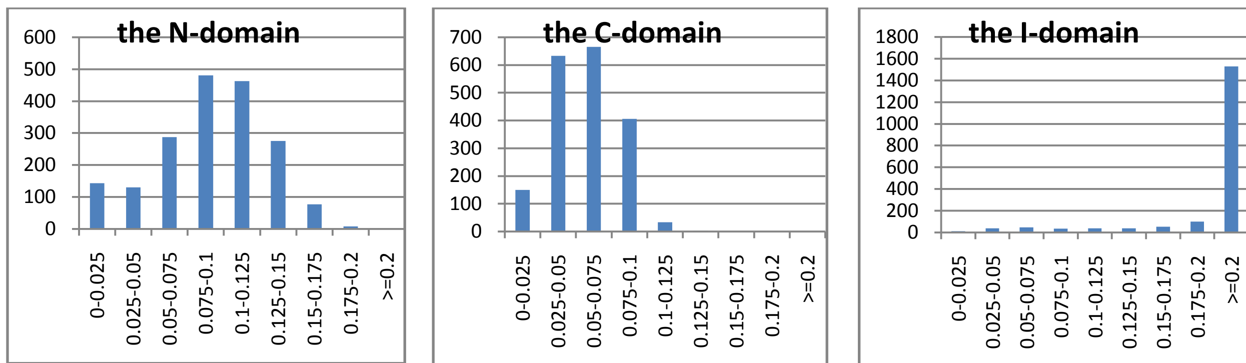


Fig.2-3 Comparation of dS, dN, and dN/dS ratio among N-domain, C-domain and the I-domain. Column height shows the average value of each domain and the error bar shows the 95% confidence interval.

dS



dN



dN/dS

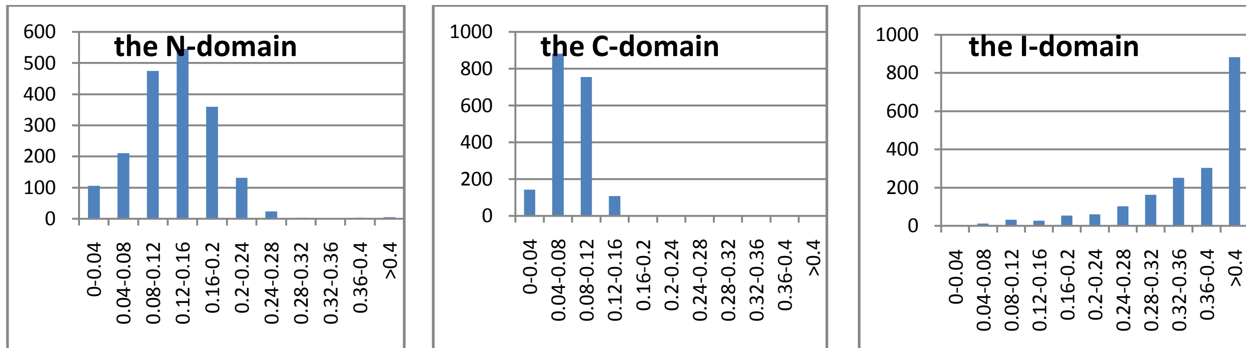


Fig.2-4 Distribution of dS, dN, and dN/dS ratio among N-domain, C-domain and the I-domains. Column height shows the number of pairwise distances that fall into each distribution range.