Fall 2007

Phylogenetic Relationships within Coleeeae (Bignoniaceae Juss.)

Andreina Carvajal
Georgia Southern University

Follow this and additional works at: https://digitalcommons.georgiasouthern.edu/etd

Recommended Citation

This thesis (open access) is brought to you for free and open access by the Graduate Studies, Jack N. Averitt College of at Digital Commons@Georgia Southern. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Digital Commons@Georgia Southern. For more information, please contact digitalcommons@georgiasouthern.edu.
PHYLOGENETIC RELATIONSHIPS WITHIN COLEEAE (BIGNONIACEAE JUSS.)

by

ANDREINA FUENTES CARVAJAL

(Under the Direction of Michelle Zjhra)

ABSTRACT

On Madagascar, Bignoniaceae is represented by tribes Coleeeae (4 genera/58 species) and Tecomeae (4 genera/15 species). Species of Coleeeae occur in assemblages of sympatric, locally endemic species. The purpose of this study was to explore the phylogenetic relationships within Coleeeae. In order to do so, techniques such as PCR by using universal primers for waxy nuclear gene, agarose electrophoresis, cloning and sequencing were performed. A total of 26 taxa were successfully amplified. Sequences were analyzed using three different phylogenetic programs (Maximum parsimony, maximum likelihood and Bayesian analysis) resulting in various phylogenetic trees. Main conclusions drawn from this work are: 1. waxy is a good molecular marker to resolve relationships within Coleeeae; 2. ITS and waxy resolved main clades similarly showing monophyly of Colea and Rhodocolea. Since conservation requires reliable information on species identity, a good phylogenetic assessment is of great importance in determining taxonomic status of organisms.

INDEX WORDS: Bignoniaceae, Coleeeae, Phylogeny, GBSSI, waxy.
PHYLOGENETIC RELATIONSHIP WITHIN COLEEAE (BIGNONIACEAE JUSS.)

by

ANDREINA FUENTES CARVAJAL

B. S., Universidad de Oriente, Venezuela, 2003

A Thesis Submitted to the Graduate Faculty of Georgia Southern University in Partial
Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

STATESBORO, GA

2007
PHYLOGENETIC RELATIONSHIP WITHIN COLEEEAE (BIGNONIACEAE JUSS.)

by

ANDREINA FUENTES CARVAJAL

Major Professor: Michelle Zjhra
Committee: Lorenza Beati, James Hutcheon

Electronic Version Approved:
December 2007
DEDICATION

To my mother, Morella, my inspiration for everything I do.

To Michelle Zhara and Tatjana Good, for believing in me.

To my husband, Matthew, because of the happiness you have brought into my life.
ACKNOWLEDGMENTS

To my advisor, Dr. Michelle Zjhra

To my committee, Dr. James Hutcheon and Dr. Lorenza Beati,

To the Graduate students office,

To the Irene Boole foundation,

To Dr. Laura Regassa,

To Dr. Chris Cuttler,

To my friends Tatjana Good and Pete Way.

To Dr. Ray Chandler,

To Dr. Steve Vives, and

To Dr. Dick Diebolt
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>6</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>8</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>9</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>1 INTRODUCTION</td>
<td>10</td>
</tr>
<tr>
<td>2 MATERIAL AND METHODS</td>
<td>14</td>
</tr>
<tr>
<td>3 RESULTS</td>
<td>21</td>
</tr>
<tr>
<td>4 DISCUSSION</td>
<td>34</td>
</tr>
<tr>
<td>5 LITERATURE CITED</td>
<td>38</td>
</tr>
</tbody>
</table>
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1: PCR’s general conditions</td>
<td>19</td>
</tr>
<tr>
<td>Table 2: Primers used in PCR amplification</td>
<td>19</td>
</tr>
<tr>
<td>Table 3: Summarized results of modeltest</td>
<td>33</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Diagram of granule-bound synthase gene from <em>Solanum tuberosum</em></td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>Map of distribution of Coleeeae’s species</td>
<td>24</td>
</tr>
<tr>
<td>3</td>
<td>Phylogenetic trees from analyses of ITS</td>
<td>27</td>
</tr>
<tr>
<td>4</td>
<td>Phylogenetic trees from analysis of a subset of <em>waxy</em></td>
<td>28</td>
</tr>
<tr>
<td>5</td>
<td>Phylogenetic trees from analysis of <em>waxy</em></td>
<td>29</td>
</tr>
<tr>
<td>6</td>
<td>Phylogenetic trees from analysis of combined data 1 (ITS+<em>waxy</em>)</td>
<td>30</td>
</tr>
<tr>
<td>7</td>
<td>Phylogenetic trees from analysis of combined data 2 (ITS+subset <em>waxy</em>)</td>
<td>31</td>
</tr>
<tr>
<td>8</td>
<td>Bayesian phylogenetic trees</td>
<td>32</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

Taxonomy of Coleoeae.

Bignoniaceae (the Trumpet Creeper family) is predominately a tropical family of trees and lianas. The family is characterized by opposite, compound leaves; flowers that are five-merous, fused, and zygomorphic with four stamens; fruit that are usually elongate, septicidal to loculicidal capsules (De Jussie, 1789). Phylogenetic studies of the family have placed Bignoniaceae in order Lamiales, within Euasterids I of the core Eudicots (Olmstead et al., 1993, APG II, 2003).

The island of Madagascar is home to the world’s second greatest center of diversity for Bignoniaceae, exceeded only by the much larger neotropical region (Gentry, 1988). The Malagasy diversity of Bignoniaceae is represented by near endemic tribe Coleoeae (4 genera/58 species) and pantropical Tecomeae (4 genera/15 species) (Zjhra, 2006).

The first classification to consider Madagascar species was that of Bojer (1837) who divided the family into three tribes: Bignonieae, Crescentieae and Coleoeae. Further classifications considered Coleoeae as part of Crescentieae (De Candolle, 1838; 1845; Seeman 1860; Baillon 1887, 1888; Perrier De la Bathie, 1938) due predominately to the shared but unusual characteristics of indehiscent, fleshy, animal-dispersed fruit. Gentry (1976, 1980) circumscribed Bignoniaceae into 7 tribes: Coleoeae, Bignonieae, Crescentieae, Tecomeae, Eccremocarpeae, Tourrettiae, and Oroxyleae. Furthermore, Gentry (1980) suggested that similarities between Coleoeae and Crescentieae were due to derived position of Bignoniaceae among the angiosperms means that Bignoniaceae
arrived on Madagascar via long-distance dispersal. The late arrival of Coleeeae, compared to earlier Angiosperm groups, suggest that species of Coleeeae radiated rapidly and recently on Madagascar (Gentry, 1980). Coleeeae, therefore, make an ideal group for studies of diversification.

Species of Coleeeae occur in sympatric assemblages and are often local endemics (Zjhra, 1998, Good et al., 2006, Zjhra, in review). Zjhra (2003) investigated the radiation and diversification of traits using a nrDNA (ITS) derived phylogeny and found that traditional generic delimitations were not monophyletic. Yet cpDNA (ndhF, trnT-L, trnL-F spacer data) derived relationships suggested monophyletic generic clades (Zjhra et al., 2004). The latter could be possible due to the slower rates of evolution of chloroplast genome in comparisons to nuclear DNA, which have make them useful in resolving higher taxonomic levels than species (Sang, 2002).

**Molecular markers used to resolve phylogenies**

Plant molecular systematists predominately have relied upon chloroplast DNA (cpDNA) and nuclear highly repetitive rDNA arrays, using conserved genes to study ancient divergence events and spacer regions for lower taxonomic level (Mason-Gamer et al. 1998). However, cpDNA is generally inherited as a single unit and lacks recombination, thus phylogenies based on cpDNA effectively represent a single gene tree (Doyle, 1992) that traces the genealogy of one parent (Sang, 2002).

The internal transcribed spacer (ITS) is the most frequently sequenced nuclear region due to its comparatively easy amplification, compared to other nDNA markers, and sufficient variability to resolve phylogenetic relationship in different plants groups at the generic level or lower (Mort & Crawford, 2004). However, poor resolution at lower
taxonomic levels (Baldwin et al., 1995) is likely due to lack of sufficient informative characters in the short length of the spacer (Sang, 2002) in some groups.

The need for additional molecular markers for reconstructing speciation events is fundamentally a need for constructing organismal phylogenies rather than single gene trees (Cronn et al., 2003, Doyle, 1992). Low-copy genes, such as GBSSI, have attracted the attention of systematists since they offer several advantages such as limited concerted evolution (Álvarez & Wendel, 2003). Problems, however, about the use of low-copy loci include marker selection, primer design, duplications, and deletions, which can potentially lead to the reconstruction of gene duplication events, rather than speciation events (Sang, 2002).

The granule-bound starch synthase (GBSSI) gene, also known as waxy (Echt & Schwartz, 1981), encodes an enzyme involved in starch synthesis. This starch synthesis enzyme adds glucose residue from ADP glucose to the non-reducing end of a growing glucan chain and is the only one within the starch synthases that is required for amylose synthesis, although there is some evidence of glucose incorporation into amyllopectin in starch granules (Denyer et al., 2001). The waxy gene cloned in potato (van der Leij et al., 1991) comprised 4663 bp; the translated portion encompassed 2961, had 13 translated exons, and 13 introns, without the one in the leader. It encoded a 58.2 kilodalton mature protein with 540 amino acids. Waxy is a low-copy nuclear gene, existing as a single copy in nearly all plants examined (Mason & Gamer et al., 1998). The exceptions to date, however, are duplications in Rosaceae and Rhamnaceae (Evans et al.; 2000), as well as Viburnum (Adoxaceae) (Winkworth & Donoghue, 2004). The phylogenetic utility of the nuclear gene waxy have been demonstrated across a wide range of taxonomic levels in
different groups of plants, such as Poaceae (Yang et al., 2007; Ingram and Doyle, 2003, Mason-Gamer et al. 1998), Malvaceae (Small, 2004), Solanaceae (Levin et al., 2005; Peralta & Spooner, 2001; van der Leij et al., 1991), and Convolvulaceae (Miller et al., 1999). Waxy exhibited sufficient variability to resolve relationships including between distantly related species (e.g., wide hybridization between distantly related species within subfamily Maloideae; Evans & Campbell, 2002) and among species that gave rise to the allopolyploid Geinae (Colurieae: Rosaceae) (Smedmark et al., 2003; Smedmark et al.; 2005).

**Importance of a good phylogenetic framework**

Madagascar is one of the eight “hottest” biodiversity hotspots based on richness and endemism of plants, and on habitat loss (estimated at >90 per cent) (Ganzhorn et al., 2001). The plant diversity on Madagascar is seriously threatened: 54% of the species of Coleoeae are threatened with extinction (Good et al., 2006)). Since conservation requires reliable information on species identity, a good phylogenetic assessment is of great importance in determining taxonomic status of organisms.

Additionally, a robust phylogenetic framework is useful in assessing the evolution of traits important to diversification (Donogue, 1989). Coleoeae provide a rich variety of traits including floral, fruit, and leaf characters that provide clues to the astounding recent diversification of this group.

The purpose of this study is to determine the relationships within Coleoeae using waxy, comparing and contrasting waxy with the existing ITS phylogeny, and evaluating the combined data phylogeny.
CHAPTER 2
MATERIALS AND METHODS

Study organisms, sampling and outgroup choice.

Materials were available for approximately 65 taxa. Two outgroup taxa were chosen: *Kigelia africana*, an Africa genus placed by Gentry (1976) in Coleeeae, and *Tabebuia*, a member of Central American tribe Crescentiae that originally included Coleeeae (Gentry, 1980). Outgroup taxa were chosen due to their close relationship with Coleeeae (Spangler & Olmstead, 1999). Both taxa were used to root the tree in all the analyses.

DNA extraction

Total DNA was extracted from ~0.1 g of herbarium or silica gel preserved leaves. The silica gel material was collected in Madagascar between 1994 and 1998. DNA was extracted using either a modified CTAB (cetyltrimethylammonium bromide) procedure (Weigel & Glazebroo, 2002) or CTAB DNA mini-prep with 6% CTAB (Murray & Thompson, 1980; Saghai-Maroof *et al.*, 1986). In both methods, DNA was resuspended in 50 µl of distilled-deionized water.

Amplification of nDNA gene region

To amplify GBSSI, 2µl of total DNA was used in a total volume of 50µl reaction. Optimal concentrations for PCR were determined (Table 1). The first pair of primers employed were waxy13R 5’GGAGTGGCRACGTTTTCCTT3’, waxy10F 5’ACTGCTGGNGCTGATTTTATG3’, and 7F 5’GYYTTSTGCATCCACAACATTGC3’ (Olmstead, pers. comm.). With 10F and 13R a 650-750 bp length was expected while for 7F and 13R was of 1500 bp. The annealing
temperature was optimized for the combination of primers 10F & 13R, resulting in *Rhodoclea* amplifying at 61°C, and *Phyllarthron* and *Colea* at 60°. A hot start and touchdown procedure was employed frequently as follows: 94°C for 50 sec, 59°C for 1 min, 72°C for 1:30 min (ten cycles), 94°C for 50 sec, 56°C for 1:00 min, 72°C for 1:30 (ten cycles), 94°C for 50 sec, 54°C for 1:30 min, 72°C for 1:30 (20 cycles), and a final extension step at 72°C for 7 min. PCR products were checked on 1.2% agarose gels stained with 5µl of 1:10 ethidium bromide and viewed with ultraviolet light. Concentrations were estimated by visual comparison with bands containing known amounts of DNA.

Various Taq polymerases were used for most routine procedures. For cloning, however, *ExTaq* buffer (TaKaRa) and 0.25 µl *ExTaq* (Takara) were used to assure the 5’ and 3’ terminal deoxyadenine overhang. PCR were performed in an Eppendorf Thermal Cycler. When more than one product was obtained, stringency was applied by increasing the annealing temperature to yield a simple PCR product. PCR products were purified using QIAquik PCR purification kit (Qiagen). Final elution was done with buffer EB provided with the kit. Amplification products were checked on 1.2% agarose gels stained with 5µl of 1:10 ethidium bromide and viewed with ultraviolet light for presence of inserts. Concentrations were estimated by visual comparison with bands containing known amounts of DNA.

Some samples were directly sent to the department of Biochemistry and Molecular Biology at Oklahoma State University for sequencing along with downstream and upstream primers in concentrations of 10picomoles/µl. Other samples were used in cloning. Oklahoma State University used the Applied Biosystems Big Dye Terminator
1.1 cycle sequencing kit. Reactions were purified using Edge Biosystems Dye Terminator Removal System. Purified reactions were analyzed on an Applied Biosystems 3730 DNA analyzer.

**Cloning**

After PCR product was purified, the samples were cloned with the TOPO TA cloning kit for sequencing (Invitrogen, Carlsbad, California, USA). Ligation, transformation, and plating were carried out following the manufacturer's instructions, except that ligation and transformation volumes were halved. Ligation incubated for 30 minutes on a thermocycler at 22°C. The nutritive media for plating and growing the bacteria was Luria Bertani agar and Luria Bertani broth (Fisher, Fair Lawn, New Jersey) with 100 µg/ml of kanamycin to grew colonies with the inserted plasmid. Incubation was done in 1ml per colony and left for approximately 48 hours with vertical shaking. The DNA plasmid was purified using QiaPrep Spin Miniprep kit (Qiagen). Plasmids with inserts were screened by PCR using the M13 reverse and forward primers that come with the cloning kit.

**Alignment**

Sequence chromatograms were checked and edited manually in Genes Code Corporation’s Sequencher, version 4.1.2 for Machintosh (Ann Arbor Michigan, USA). With cloned sequences, the vector was trimmed by imputing the flanked regions around the insert. Sometimes manual trimming was necessary. Once edited, a consensus sequence was obtained by assembling the upstream and downstream DNA sequences for each taxon. Consensus sequences were exported as text to McClade, 4.03 and further
manually adjusted. Consensus sequences were subjected to a BLAST search (Altschul et al., 1997) in GenBank to test for GBSSI identity.

**Primer design**

From the few initial sequences obtained with the two universal pairs of primers, new primers were designed (Table 2). Once aligned, conserved regions at the beginning and end of the sequences were selected and primers were ordered from Invitrogen. Primers were designed with a similar melting point, as recommended (Strand et al., 1997). Initially, four primers were designed without considering exon or intron position in order to amplify as large a portion of the gene as possible (Mort & Crawford, 2004). Primers “waxy8downs,” “waxy12ups” were designed to align with exon positions 8 and 12 of Potato *waxy* gene (Figure 1) (GenBank x58453) (van der Leij et al., 1991).

**Phylogenetic analyses**

Phylogenetic trees were inferred using maximum parsimony (MP), and maximum likelihood (ML) optimality criterion in PAUP*4.0b10 (Swofford, 2002). In addition, data was analyzed using a Bayesian analysis approach with Mr. Bayes (v3.0; Huelsenbeck & Ronquist, 2001).

MP was set for heuristic search with “tree-bisection-reconnection” (TBR) branch swapping, collapsed branches if maximum length was zero, and all characters equally weighted and unordered. Gaps were treated as missing data. Multistate taxa were interpreted as uncertainty. A strict consensus tree was obtained. Relative measures of support for clades were estimated using Bootstrap analysis (Felsenstein, 1985), and are represented by numbers over the branches. MaxTrees was set to 2000 replicates with full heuristic search.
For both ML and Bayesian analyses, Modeltest version 3.7 (Posada & Crandall, 1999) was used in combination with PAUP* to select for the best evolutionary model for each data based on the Akaike information criterion (AIC; Akaike, 1974). ML analyses were set to a maximum of 2000 trees. TBR branch-swapping and the MulTrees options were in effect. Starting trees were obtained from the MP analysis.

Bayesian analyses were conducted with Mr. Bayes version 3.1.2. Each analysis was initiated from a random starting tree and the program was set to run four Markov chain Monte Carlo iterations (three heated and one cold) simultaneously for 50,000 generations and a tree was saved every 100 generations. The posterior probabilities for clades were estimated and are represented by numbers above the branches.

ITS and \textit{waxy} were analyzed separately to construct phylogenetic trees: ITS alone with a total length of 720 bp (using data from Zjhra, 1998); \textit{waxy} with a total length of 1198 bp and a subset of \textit{waxy} that minimized missing data with 439 bp. Two combined dataset with ITS and \textit{waxy} were analyzed: \textit{waxy} with 1198 bp; \textit{waxy} with 439 bp. Analyses were conducted with and without \textit{P. cauliflorum} since it was variously placed in the \textit{Colea} clade in the \textit{waxy} data set. \textit{Kigelia} and \textit{Tabebuia} were used in all cases as outgroups. The different genera are indicated by different colors on the phylogenetic trees.
Table 1. PCR’s general conditions.

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Start concentration</th>
<th>End concentration</th>
<th>Volume µL/50µL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buffer</td>
<td>10X</td>
<td>1X</td>
<td>5</td>
</tr>
<tr>
<td>MgCl</td>
<td>25mM</td>
<td>2.5 mM</td>
<td>5</td>
</tr>
<tr>
<td>dNTP</td>
<td>10µM</td>
<td>0.4µM</td>
<td>2</td>
</tr>
<tr>
<td>Upstream Primer</td>
<td>20µM</td>
<td>0.8µM</td>
<td>1</td>
</tr>
<tr>
<td>Downstream Primer</td>
<td>20µM</td>
<td>0.8µM</td>
<td>1</td>
</tr>
<tr>
<td>BSA</td>
<td>100%</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Water</td>
<td></td>
<td></td>
<td>to volume</td>
</tr>
</tbody>
</table>

Table 2. Primers used in PCR amplification and sequencing of *waxy*.

<table>
<thead>
<tr>
<th>Pair of primers</th>
<th>Primer sequence (with forward and reverse)</th>
<th>Sequence length product expected (bp.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WaxyFColeeae, WaxyRColeeae</td>
<td>5’gccatgccccacaggctcacc 3’ 5’tggtctcatccccaatgtgatcctg 3’</td>
<td>650</td>
</tr>
<tr>
<td>Coleeaeupstream1, Coleeaedownstream1</td>
<td>5’atcgccccacaggctcacc 3’ 5’ccaatgtgatcctg 3’</td>
<td>No product</td>
</tr>
<tr>
<td>Coleeaeupstream2, Coleeaedownstream2</td>
<td>5’ctgcttgaatcaggtacctatctg 3’ 5’gttctcaatgactaccttccaagagatc 3’</td>
<td>520</td>
</tr>
<tr>
<td>GBSS1coleeaependown, GBSS1coleeaandup</td>
<td>5’gtgagatatgcacgattagttgag 3’ 5’gttctcaatgactaccttccaagagatc 3’</td>
<td>800</td>
</tr>
<tr>
<td>Waxy8downs, waxy12ups</td>
<td>5’ctggatgmacggctggaatcctgtag 3’ 5’gttctcaatgactaccttccaagagatc 3’</td>
<td>780</td>
</tr>
</tbody>
</table>
Figure 2. Diagram of granule-bound synthase gene from *Solanum tuberosum*. Arrows indicate locations and directions of waxy8ups and waxy12downs. (Diagram adapted from van der Leij *et al.*, 1991)
CHAPTER 3

RESULTS

Amplification and alignment:

A total of 24 ingroup taxa were successfully amplified (Fig. 2) which represent approximately 36% of the total species described, and encompasses three of the four genera. Despite intensive efforts, *Ophiocolea* failed to amplify. The alignment of the sequences of *Colea, Rhodocolea*, and *Phyllartron* were straightforward. Outgroup alignment of *Tabebuia* and *Kigelia* was straightforward up to exon 12, after which variability made the alignment ambiguous. This last ambiguous portion was eliminated from the analysis; this variability was potentially a result of low sequence signal.

Phylogenetic analysis.

Maximum Parsimony, Maximum Likelihood, and Bayesian analyses were performed on 3’ portion of *waxy* obtained in this study. The maximum bp length encompassed 1198 bp ("*waxy" data set) from exon 8. Missing data resulted from using different combinations of primers for successful amplification. To explore the impact of missing data, a data subset that minimized missing data was also analyzed. This data set was 439 bp long ("subset of *waxy*"). Previous analysis with *waxy* under MP showed *P. cauliflorum* as a member of *Colea* clade. To test the impact of *P. cauliflorum* on the analysis, analyses were conducted with and without it. Trees obtained with *P. cauliflorum* are not show since placement of the taxa was uncertain.

MP vs. ML vs. bayesian analysis.

**ITS:** this data set included 21 ingroup taxa and 2 outgroup. The ITS sequences were 720 bp in length that included ITS 1, 5.8 s and ITS 2 from nuclear ribosomal DNA
and included 106 (11.08%) parsimony-informative character-sites. The score of the best tree found was L = 286 and 23 trees were retained from which a strict consensus tree was obtained (Fig. 3A). In this tree, two subclades of *Phyllarthron* showed up: *Phyllarthron* I is sister to *Phyllarthron* II and sister to the rest. Taxa that belong to different populations but the same species came out as more closely related except for *P. madagascariense* and *R. nobilis*. Support was over 70% for most of the branches, except *Colea* and some branches in *Phyllarthron* I. The ML tree (Fig 3B) is more resolved than MP, showing *C. cava* as sister of the rest of *Colea*, a monophyletic *Rhodocolea*, and two clades of *Phyllarthron*. The best tree score was -ln L 2606.09118. Topology of the Bayesian analyses tree (Fig.3C) is similar to the one obtained by Maximum likelihood analysis, with a minor difference in the placement of the two accessions of *Rhodocolea nobilis*.

**Subset of waxy**: No missing data was present except for *Kigelia* and *Tabebuia*. Sixty-one characters (13.8 %) were parsimony-informative. The best tree score was= 219. Two trees were retained and a strict consensus obtained (Fig. 4A). As in ITS, *Colea* and *Rhodocolea* were monophyletic under ML and Bayesian analyses (Fig. 4B, C). However, *waxy* results in monophyletic *Colea* and *Rhodocolea* under MP (Fig. 4A) whereas ITS does not. The best tree score was -ln L 1759.15410. Bayesian tree (Fig. 4C) has the same topology as the ML tree.

**3’ waxy**: A total of 23 ingroup taxa were analyzed for *waxy* with 1198 characters, from which 196 were parsimony-informative. Score of best tree was 391 and number of trees retained was equal to 45. The topology of the MP tree is similar to the one obtained from the subset data analysis (Fig. 5A), differing only in slight variations in the relationships within *Phyllarthron* II. Branch support is higher for 3’ *waxy*, in general,
than for subset \textit{waxy}. Topology is consistent among the three analyses of the 3’ \textit{waxy} (Fig. 5A, B, C). The best tree score found was -ln L 3807.66189 for ML.

**Combined data 1**: This data set included 20 taxa with both ITS and \textit{waxy} sequences, using 3’ \textit{waxy} (1198 bp). A total of 18 ingroup taxa were included. Out of 1907 characters, there were 161 parsimony-informative characters. The best tree found was 636 and number of trees retained was 10. The combined data analyses trees (Fig. 6A, B, C) are similar to separate analyses, but with the two \textit{Phyllarthron} clades adjacent and \textit{Rhodocolea} within \textit{Colea}. ML tree score for best tree found was -ln L  6173.42045.

**Combined data set 2**: This dataset included the subset of \textit{waxy} plus ITS for the 18 included ingroup taxa. There were 125 informative characters. This data set included 441 bp length from \textit{waxy} and 720 of ITS data. The best parsimony tree was 475. The number of trees retained was 31. Results are similar to the combined data set 1, but with more resolution within clades and monophyly of \textit{Colea} in the Bayesian tree (Fig. 7A, B, C). The best tree score found for ML analysis was -ln L  4219.36415.

**Congruence between individual gene trees vs. combined data**: Tendencies are mostly the same when data is combined in comparison with individual data (Fig. 8). However, in the combined data, \textit{Colea} is paraphyletic.
A) Colea

C. cava

C. gentry

C. muricata

C. ramiflora

C. systmae
B) Phyllarthron

- P. articulatum
- P. bernieranum
- P. madagascariense
- P. nocturnum
- P. sahamalense
- P. yokoamensis
Figure 2. Map of distribution of Coleeeae’s species. Shaded areas: 2000 forest cover; dark green=rain forest, yellow= spiny desert. Pale green= mangrove forest, blue= cloud cover. Blue dots= species localities for A) Colea, B) Phyllarthron, and C) Rhodocolea. (Maps by Dr. Good)
Figure 3. Phylogenetic trees from analyses of ITS. A) Strict consensus tree of 23 trees (L=286) from MP analyses. Bootstrap values are numbers above the branches. B) ML tree (-ln L 2606.09118). C) Bayesian tree showing posterior clade probabilities.
Figure 4. Phylogenetic trees from analysis of a subset of waxy. A) Strict consensus tree of 2 trees (L=219) from MP analysis. Bootstrap values are numbers above the branches. B) ML tree (-ln L1759.15410). C) Bayesian tree showing posterior clade probabilities.
Figure 5. Phylogenetic trees from analysis of waxy (1198 bp). A) Strict consensus tree of 45 trees (L=391) from MP analyses. Bootstrap values are numbers above the branches B) ML tree (-ln L 3807.66189). C) Bayesian tree showing posterior clade probabilities.
Figure 6. Phylogenetic trees from analysis of combined data 1 (ITS+waxy). A) Strict consensus tree of 10 trees (L= 636) from MP analyses. Bootstrap values are numbers above the branches. B) ML tree (-ln L  6173.42045). C) Bayesian tree showing posterior clade probabilities.
Figure 7. Phylogenetic trees from analysis of combined data 2 (ITS+subset waxy). A) Strict consensus tree of 31 trees (L=475) from MP analyses of the combined data 2. Bootstrap values are numbers above the branches. B) ML tree (-ln L 4219.3615). C) Bayesian tree showing posterior clades probabilities.
Figure 8. Bayesian phylogenetic trees of A) ITS, B) waxy, C) combined data, and D) phylogenetic representation of Gentry’s (1980) taxonomic relationships within Coleeeae.
Table 3. Summarized results of modeltest:

<table>
<thead>
<tr>
<th>Data set</th>
<th>Model selected</th>
<th>Substitution model</th>
<th>Bases frequencies</th>
<th>Proportion of invariable sites</th>
<th>Gamma distribution shape</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITS</td>
<td>GTR+I+G</td>
<td>R (a) [A-C] = 1.1082, R (b) [A-G] = 1.6632, R(c) [A-T] = 0.9372, R (d) [C-G] = 0.5916, R (e) [C-T] = 4.2120, R (f) [G-T] = 1.0000</td>
<td>A= 0.2369; C= 0.3059; T= 0.1791; G= 0.2782.</td>
<td>0.3700</td>
<td>0.7983.</td>
</tr>
<tr>
<td>Subset de waxy</td>
<td>TVM+G</td>
<td>R(a) [A-C] = 1.2551, R(b) [A-G] = 3.1914, R(c) [A-T] = 0.8781, R(d) [C-G] = 2.1382, R(e) [C-T] = 3.1914, R(f) [G-T] = 1.0000.</td>
<td>A= 0.3031; C= 0.1325; T= 0.3615; G= 0.2030.</td>
<td>0</td>
<td>0.9577</td>
</tr>
<tr>
<td>Waxy</td>
<td>TrN+G</td>
<td>:R(a) [A-C] = 1.0000, R(b) [A-G] = 2.6295, R(c) [A-T] = 1.0000, R(d) [C-G] = 1.0000, R(e) [C-T] = 3.8432, R(f) [G-T] = 1.0000.</td>
<td>A= 0.2953; C= 0.1600; T= 0.3282; G= 0.2165.</td>
<td>0</td>
<td>0.4761</td>
</tr>
<tr>
<td>Combined data 1</td>
<td>TrN+G</td>
<td>R(a) [A-C] = 1.0000, R(b) [A-G] = 2.4952, R(c) [A-T] = 1.0000, R(d) [C-G] = 1.0000, R(e) [C-T] = 4.0940, R(f) [G-T] = 1.0000.</td>
<td>A= 0.2691; C= 0.2183; T= 0.2733; G= 0.2392.</td>
<td>0.3774</td>
<td>0.8258</td>
</tr>
<tr>
<td>Combined data 2</td>
<td>TrNeF+I+G</td>
<td>R(a) [A-C] = 1.0000, R(b) [A-G] = 2.4787, R(c) [A-T] = 1.0000, R(d) [C-G] = 1.0000, R(e) [C-T] = 3.5945, R(f) [G-T] = 1.0000.</td>
<td>Equal frequencies</td>
<td>0.3338</td>
<td>0.7905.</td>
</tr>
</tbody>
</table>
CHAPTER 4

DISCUSSION

Effect of sampling: differences in resolution among the main clades as well as within clades probably derived from the differences in both the number and species sampled in ITS and waxy data sets. Complete absence of members of genus Ophiocolea may also account for the relationships observed. In Zjhra’s (2003) analysis of ITS with 65 taxa, Colea was sister to Ophiocolea. In the larger dataset of ITS, Phyllarthron is also paraphyletic. This can be due to the paucity of sampling within Phyllarthron itself, since phylogenies constructed with chloroplast DNA including broad representation of Phyllarthron taxa showed Phyllarthron as monophyletic (Zjhra, 2004).

Effect of missing data: When analyzing waxy data with a subset or the total length of the data set, better resolution for most of the branches came out with the total length of the sequences than for a subset, but phylogenetic relationships stayed the same for most of the branches. Cracraft and Helm-Bychowski (1991) analyzed primate relationships using mitochondrial DNA and found that different subsets of the data resolved consistently with few exceptions, and although the tendency is to improve the resolution with the addition of data, this is not always true. This was true for the Coleeeae waxy data, where the analyses of a subset showed better resolution for Colea, whereas within Phyllarthron the resolution improved with adding more data. This can be due to few phylogenetically informative characters in the original short Colea sequences. Missing data can be an obstacle when reconstructing phylogenetic relationships (Donogue et al., 1989), leading to multiple shortest trees and poorly resolved consensus trees (Gauthier, 1986). Although missing data can result in inaccurate phylogenies, the
mechanisms still remain unclear. Huelsenbeck (1991), for instance, proposed that this is due to the percentage of equivocally resolved ancestral characters. By computer simulations, Wiens (2003) found support for the hypothesis that the lack of resolution when using incomplete taxa is primarily one of including too few characters rather than including too many missing data cells.

**Utility of waxy vs. ITS.** The internal transcriber spacer of the nuclear rDNA has been used to study relationships among closely related species, but frequently is not variable enough to distinguish species in some groups (Mason-Gamer *et al.*, 1998). In *Hibiscus* Sect. *Muenchhusia* (Small, 2004) ITS data contained 17 nucleotide substitutions out of a total of 683 bp aligned nucleotides, with only 2 potentially phylogenetically informative characters. GBSSI, on the other hand, consisted of a total of 1972 bp with 28 phylogenetically informative. However, in other cases, ITS have shown more resolving power than *waxy*: in *Ipomoea* (Miller *et al*. 1999) a total length of 573 nucleotides had 163 phylogenetically informative sites, contrasting with *waxy* which showed just 86 potentially informative sites in a final alignment of 651 bp length. In Coleeeae, ITS shows enough variability to resolve relationships among species, even with fewer taxa. *Waxy* was also useful in resolving phylogenetic relationships within species, even with a very short subset of data. The informativeness of ITS (14.68%) and *waxy* (13.89%) was equivalent. However, GBSSI is a longer region and therefore potentially provides a greater number of phylogenetically informative sites.

Common terminal branch relationships among ITS data and *waxy* include *P. vokoanina* 709 and *P. madagascariense* CapEst (*P. madagascariense* 967 in *waxy*), which is also supported by previous analyses of ITS (Zjhra, 2003). Finally, monophyly of
Colea and Rhodocolea are supported by both ITS and waxy as well as the presence of two subclades of Phyllarthron.

**Combined data:** In general, there was less resolution in the combined data sets compared to the individual data sets. This can be due to differences in sampling between the waxy and ITS data sets. Only in taxa better represented, such as Phyllarthron, the resolution is consistent between the combined and separate data analyses. The combined data sets showed Colea as paraphyletic and Rhodocolea unresolved. Also, in both combined data sets, Colea and Rhodocolea came out as a single clade. This finding conflict with results derived from three cpDNA data sets where Rhodocolea is sister to the rest of the tribe (Zjhra et al. 2004). When taxa don’t have strong support, they can appear anywhere in the tree; Colea lacked strong support in both of the individual data analyses.

**Phylogeny methods:** Bayesian analysis generally resulted in better resolution than ML and MP. Posterior probabilities tend to be higher in all the cases than bootstrap values. However, Bayesian posterior probabilities determine the strength of the data in supporting particular nodes, whereas bootstrap values indicate areas where additional data is needed to resolve relationships (Miller et al. 2004).

**Phylogenetic relationships:** The main phylogenetic relationships observed are monophyly of Colea and Rhodocolea, and two subclades of Phyllarthron. These relationships are also supported in the broader phylogenetic analysis of ITS (Zjhra, 2003) with 61 ingroup taxa. When comparing phylogenetic analyses of ITS data with waxy data, the major differences are that Rhodocolea is sister to Phyllarthron 1 which is sister of Colea and Phyllarthron 2 (ITS data), whereas Phyllarthron 2 is sister to a
monophyletic *Colea* and *Rhodocolea* (*waxy*). In the 61 taxa ITS phylogeny (Zjhra, 2003), *Rhodocolea* was sister of the rest of the tribe. The results from *waxy* place *Rhodocolea* in a derived position, whereas Gentry (1988) placed *Rhodocolea* in a primitive position within the tribe (Figure 8).

Finally, phylogenetic relationship were surprisingly robust to the effects of missing data, with branch support the major difference between analyses of the entire vs. subset of *waxy*. ITS shows enough variability to resolve relationships among species, even with fewer taxa. *Waxy* was also useful in resolving phylogenetic relationships within species, even with a very short subset of data. Monophyly of *Colea* and *Rhodocolea* are supported by both ITS and *waxy* as well as the presence of two subclades of *Phyllarthron*. Less resolution in most of the branches can be observed in combined data sets when compare with individuals data.


