Available online at www.sciencedirect.com
SCIENCE DIRECT®
Phytochemistry 65 (2004) 59–69
PHYTOCHEMISTRY
www.elsevier.com/locate/phytochem

Relationships in subtribe Diocleinae (Leguminosae; Papilionoideae) inferred from internal transcribed spacer sequences from nuclear ribosomal DNA

Eduardo S. Varela[a], João P.M.S. Lima[a], Alexsandro S. Galdino[a], Luciano da S. Pinto[a], Walderly M. Bezerra[a], Edson P. Nunes[b], Maria A.O. Alves[a], Thalles B. Grangeiro[a,*]

[a]*Laboratório de Citogenética e Genética Molecular, Departamento de Biologia, Bloco 906, Centro de Ciências, Universidade Federal do Ceará (UFC), Av. Humberto Monte, s/n. 60.451-970, Fortaleza-Ceará, Brazil*
[b]*Herbário Prisco Bezerra, Departmento de Biologia, Universidade Federal do Ceará (UFC), Av. Humberto Monte, s/n. 60.451-970, Fortaleza-Ceará, Brazil*

Received 16 May 2003; received in revised form 29 July 2003

Abstract

The complete sequences of nuclear ribosomal DNA (nrDNA) internal transcribed spacer regions (ITS/5.8S) were determined for species belonging to six genera from the subtribe Diocleinae as well as for the anomalous genera *Calopogonium* and *Pachyrhizus*. Phylogenetic trees constructed by distance matrix, maximum parsimony and maximum likelihood methods showed that *Calopogonium* and *Pachyrhizus* were outside the clade Diocleinae (*Canavalia*, *Camptosema*, *Cratylia*, *Dioclea*, *Cymbosema*, and *Galactia*). This finding supports previous morphological, phytochemical, and molecular evidence that *Calopogonium* and *Pachyrhizus* do not belong to the subtribe Diocleinae. Within the true Diocleinae clade, the clustering of genera and species were congruent with morphology-based classifications, suggesting that ITS/5.8S sequences can provide enough informative sites to allow resolution below the genus level. This is the first evidence of the phylogeny of subtribe Diocleinae based on nuclear DNA sequences.

Keywords: Subtribe Diocleinae; Leguminosae; Phylogenetic analysis; Nuclear ribosomal DNA; ITS/5.8S region

1. Introduction

The Leguminosae (Fabaceae) is the largest family of flowering plants after Compositae and Orchidaceae, with three subfamilies comprising some 650 genera and more than 18,000 species (Polhill and Raven, 1981). The tribe Phaseoleae (Papilionoideae) comprises the largest number of genera and it is the most economically important of any tribe of the Leguminosae. The tribe is currently divided into eight subtribes: Cajaninae, Phaseolinae, Clitoriinae, Ophrestiinae, Kennediinae, Erythrininae, Diocleinae and Glycininae (Polhill and Raven, 1981; Polhill, 1994).

The Diocleinae are mostly from the New World and possess a number of primitive characteristics: the habit is often woody or coarse, the frequently large flowers have prominent discs about the ovary, the hila are often long, and the non-protein amino acid canavanine is usually present (Lackey, 1981). Lackey recognized 13 genera of Diocleinae (*Dioclea*, *Cymbosema*, *Cleobulia*, *Canavalia*, *Pachyrhizus*, *Macropsychanthus*, *Luzonia*, *Camptosema*, *Cratylia*, *Collaea*, *Galactia*, *Calopogonium*, and *Herpyza*) and suggested that three, *Calopogonium*, *Herpyza* and *Pachyrhizus*, were likely to be improperly placed in the subtribe. Phylogenetic studies of Phaseoleae, based on chloroplast DNA (cpDNA) restriction-site mapping of the inverted repeated regions, showed that *Calopogonium* and *Pachyrhizus* arose within the Glycininae clade, outside Diocleinae (Doyle and Doyle, 1993). Based on these findings Polhill (1994) transferred *Calopogonium* and *Pachyrhizus* from the subtribe Diocleinae to Glycininae. More recently, a phylogenetic study of Glycininae using cpDNA rps16 intron sequences has also shown that Glycininae is monophyletic if

* Corresponding author. Tel.: +55-85-2889809; fax: +55-85-2889806.
E-mail address: thalles@ufc.br (T.B. Grangeiro).

0031-9422/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.phytochem.2003.08.005

Calopogonium and *Pachyrhizus* are included within Glycininae (Lee and Hymowitz, 2001). However, the study by Lee and Hymowitz did not include DNA sequence information from the true Diocleinae genera. On the other hand, it is known that phylogenetic hypotheses based on a single gene or character may not represent true organismal relationships. Therefore, it is crucial to evaluate additional genes for phylogeny reconstruction to test topologies based solely on chloroplast genes, to obtain additional resolution, and to elucidate relationships at a variety of taxonomic levels (Kuzoff et al., 1998). Although cpDNA sequences have been widely used in plant phylogeny, ribosomal RNA (rRNA) coding sequences (rDNA) have also been shown to be informative at distinct taxonomic levels, depending on the region investigated (Baldwin et al., 1995; Kuzoff et al., 1998). Thus, the nuclear ribosomal DNA (nrDNA) internal transcribed spacer region (ITS/5.8S) has been shown to provide informative sites for phylogenetic analysis in numerous plant families, including Leguminosae (Hu et al., 2002; Käss and Wink, 1997).

In the present work, we have sequenced the entire ITS/5.8S region of eight representative genera of Diocleinae sensu Lackey (1981). The goal of our work was to provide further evidence on the delimitation of the subtribe Diocleinae.

2. Results and discussion

The ITS region of nrDNA comprising both ITS sequences (ITS1 and ITS2), the 5.8S rDNA, the 3′ end of 18S rDNA and the 5′ end of 26S rDNA was amplified by PCR from 17 taxa belonging to *Canavalia*, *Camptosema*, *Cratylia*, *Dioclea*, *Cymbosema*, and *Galactia*; of the true Diocleinae as well as from the two pseudo Diocleinae genera, *Calopogonium* and *Pachyrhizus* (Table 1). The complete sequences of the PCR products generated in our laboratory were deposited in GenBank (accession numbers AY293832-AY293847, and AY343546). The DNA sequences from *Galactia striata* Urb. and *Canavalia brasiliensis* Mart. ex Benth. were retrieved from GenBank (accession numbers AF467049 and AF467034, respectively) and also included in the phylogenetic analyses. The length of the fully assembled ITS/5.8S region ranged from 613 bp (in *D. guianensis*) to 695 bp (in *C. caeruleum*), with an average length of 642 bp. ITS1 length varied from 229 bp (in *C. boliviana*) to 297 bp (in *C. caeruleum*), while ITS2 length ranged from 202 bp in *C. mucunoides* to 222 bp in *C. pedicellatum*. The above values agree with those reported for several plant species, including the Leguminosae (Baldwin et al., 1995). In all taxa examined, the 5.8S rDNA sequence had a constant length of 165 bp, which is similar to the values (162–164 bp) that have been reported in other plant species (Yokota et al., 1989). The multiple aligned sequences of the entire ITS/5.8S region had 750 characters, with 309 sites (41.2%) being conserved, 406 sites (54.1%) were variables, and 95 sites (12.7%) were unique to individual taxa. Within the whole region, ITS segments are highly variable, with only 12.8 and 28% of sites being conserved in ITS1 and ITS2, respectively. The pairwise genetic distances based on the Tamura–Nei substitution model were calculated, and upon inspection of the matrix (data not shown), one could recognize two main groups, one comprising the true Diocleinae, which has an overall mean distance of ca. 0.146, and a second one composed by the *Calopogonium* and *Pachyrhizus* species, with an overall average distance of ca. 0.103. However, the mean distance between the *Calopogonium–Pachyrhizus* complex and the true Diocleinae is ca. 0.383, which is ca. 1.5-fold the distance between the most divergent sequences in the true Diocleinae group. Although the 5.8S sequence was highly conserved among the taxa studied, a more careful inspection of its multiple alignments (Fig. 1) allowed us to identify two sites that were diagnostic for these two groups. Therefore, all true Diocleinae 5.8S sequences showed an adenine (A) at nucleotide 55 whereas in the *Calopogonium–Pachyrhizus* complex sequences, there was a guanine (G) in the same site. In a second site (nucleotide 136), a thymine (T) was found in all true Diocleinae sequences whereas a cytosine (C) was observed for the *Calopogonium-Pachyrhizus* complex sequences. Indeed the topology of the neighbour-joining (NJ) tree obtained from the Tamura–Nei genetic distances (Fig. 2) shows that these two major clades (the true Diocleinae and the *Calopogonium–Pachyrhizus* complex) are very well supported as monophyletic. Within the true Diocleinae clade, two very well supported monophyletic subclades are: a clade containing the *Canavalia* and another containing the *Dioclea*. A third, moderately supported subclade within the true Diocleinae clade contains the *Camptosema–Cratylia–Galactia* complex. Within the second major clade, *Calopogonium* and *Pachyrhizus* are highly supported as monophyletic.

Three of the most parsimonious trees were generated by maximum parsimony (MP) analysis of the ITS/5.8S region sequences, and the bootstrap consensus tree is shown in Fig. 3. This analysis included 526 parsimony-informative sites and each generated tree has a length of 630 steps, a consistency index (CI) of 0.679, a retention index (RI) of 0.785, and a rescaled consistency index (RC) of 0.533. The three most parsimonious MP trees differed basically in the relative positions of *Canavalia* species belonging to subgenus *Canavalia*. The overall topology of the consensus MP tree (Fig. 3) is essentially similar to that produced by NJ analysis (Fig. 2). This same general topology was also observed in the maximum-likelihood (ML) phylogenetic tree (Fig. 4). Differences between NJ,

Table 1
Origin of plant samples used in the ITS/5.8S sequencing study

Taxon	Origin of samples	Voucher specimen numbers[a]
Calopogonium caeruleum	Parque Nacional de Ubajara, Ceará	EAC 28475
Calopogonium mucunoides	Campus do Pici-UFC, Fortaleza, Ceará	EAC 31697
Camptosema pedicellatum	Chapada do Araripe, Crato, Ceará	EAC 29770
Canavalia boliviana	Arquipélago de Anavilhasnas, Amazonas	EAC 24200
Canavalia gladiata	Campus do Pici-UFC, Fortaleza, Ceará	EAC 32586
Canavalia grandiflora	São Benedito, Ceará	EAC 25326
Canavalia bonariensis	Pelotas, Rio Grande do Sul	EAC 13393
Cymbosema roseum	Arquipélago de Anavilhasnas, Amazonas	EAC 24199
Cratylia argentea	Campus do Pici-UFC, Fortaleza, Ceará	EAC 27356
Dioclea sp.	Mucambo, Ceará	EAC 30896
Dioclea guianensis	Campus do Pici-UFC, Fortaleza, Ceará	EAC 30076
Dioclea bicolor	Chapada do Araripe, Crato, Ceará	EAC 28286
Dioclea megacarpa	Campus do Pici-UFC, Fortaleza, Ceará	EAC 29277
Dioclea virgata	Viçosa, Ceará	EAC 27783
Galactia latisiliqua	Campus da UENF, Campos, Rio de Janeiro	EAC 32587
Pachyrhizus erosus	INPA, Manaus, Amazonas	EAC 32588
Pachyrhizus tuberosus	INPA, Manaus, Amazonas	EAC 32589

[a] Voucher specimens were deposited in the Herbarium Prisco Bezerra-UFC, Fortaleza-Ceará, Brasil.

MP and ML trees concerned mainly the bootstrap values at some nodes, which showed slight variations from tree to tree. For instance, the clade comprising the *Camptosema–Cratylia–Galactia* complex is supported by bootstrap values of 81, 83 and 93 in the NJ, MP and ML trees, respectively (Figs. 2–4). Felsenstein (1985) suggested that a bootstrap $\geqslant 95\%$ should be taken as robust support for monophyletic groupings. However, Hillis and Bull (1993) found, using empirical data and computer simulations, that almost every internal branch with a bootstrap value 80% defined a true clade. Therefore, taking into account these findings, the *Camptosema–Cratylia–Galactia* cluster is equally supported by the three methods used for the phylogenetic inference.

According to Lackey (1981), although Diocleinae is presumably natural, the subtribe is difficult to define except in a negative way. He noted that most characters do not give clear-cut indication of internal structure or some generic limits. However, one could recognize a general trend, from the primitive *Dioclea* and allies, which show a woody or coarse habit, through a series of intermediate genera terminating with *Galactia*, that has a more delicate habit, with smaller seeds with short hila. The NJ, MP and ML trees (Figs. 2–4) constructed using the ITS/5.8S region sequences of most Diocleinae genera (six out of ten) shows that species belonging to the same genus are consistently clustered together. Therefore, the ITS/5.8S sequences have enough informative sites to be useful in defining intrageneric limits of Diocleinae.

For *Dioclea* and *Canavalia*, which were both represented by species belonging to distinct subgenera, the ITS/5.8S sequences also provided resolution at the infrageneric level (i.e., below the genus level). Maxwell (1969) had recognized three subgenera for *Dioclea*: Pachylobium, Plathylobium and Dioclea, based on characteristics of inflorescence, seeds and habitat. In our analyses, we have representatives for these three subgenera as follows: *D. megacarpa* (subgenus Pachylobium), *D. bicolor* (subgenus Plathylobium) and *D. guianensis* and *D. virgata* (subgenus Dioclea). In the NJ, MP and ML trees (Figs. 2–4) the clustering of *Dioclea* ITS/5.8S region sequences are congruent with the morphologically defined subgenera (Maxwell, 1969), being well supported by high bootstrap values. In addition, *Dioclea* sp. clustered together with *D. megacarpa* (subgenus Pachylobium), which is in complete agreement with the seed morphology showed by the species without identification.

In the case of *Canavalia*, Sauer (1964) grouped the species in four subgenera: Canavalia, Catadonia, Maunaloa, and Wenderothia. In the present work, species beloging to three of these subgenera were included: *C. bonariensis* (subgenus Catadonia), *C. grandiflora* (subgenus Wenderothia), and *C. brasiliensis*, *C. gladiata*, *C. boliviana* (subgenus Canavalia). As observed for *Dioclea*, the clustering of *Canavalia* ITS/5.8S sequences in the NJ, MP and ML trees (Figs. 2–4) correlates very well with the morphologically recognized subgenera, and these topologies are supported. The basal position of *C. grandiflora* (Wenderothia) in the *Canavalia*'s subclade (Figs. 2–4) is supported and agrees with the assumption that it represents the most primitive subgenus of *Canavalia* (Sauer, 1964). On the other hand, Piper (1925) had considered subgenus Wenderothia as generically distinct, an assumption that is not supported by the ITS/5.8S data. The ITS/5.8S molecular results

```
                                                                                                                       110
[1]     -TCGATGCCC CCACAAACAG TTCGACCTGT GAATT--TGT TTAACCATGC ---------- AGGGTAAT-T TGAGA----- ---TGTTCAA CA-CATC--- ----------
[2]     -TCGATGCCC --ACAAACAG TTCGACCTGT GAATT--TGT TTAACCATGC ---------- AGGGTAAT-T TGAGA----- ---TGTTCAA CA-CATC--- ----------
[3]     -TCGATGCCT C-ACAAACAG TTCGACCCGT GAATT--TGT TTAACCGTGT ---------- AAGGTAAT-T TGAGA----- ---TGTTCAA C---ATC--- ----------
[4]     -TCGATGCCT ---CAATCAG TTCGACCCGT GAATT---GT TTAATAACGT ---------- AGGGTAAG-T TGAGG----- ---TGTTCAT CATC------ ----------
[5]     -TCGATGCCT ---CAATCAG TTCGACCCGT GAATT---GT TTAATAACGT ---------- AGG-TAAT-T TGAGG----- ---TGTTCAT CATCATC--- ----------
[6]     -TCGATGCCT ---CAATCAG TTCGACCTGT GAATT--TGT TTAATAATGT ---------- AGGGCAGG-T TGAGA----- ---TGTTCAC CATC------ ----------
[7]     --TCGATGCC TCATAACCAG TTCAACTTGT GAATT-TGTT TTACT--ATG C--------A GGGTTGATTG GA-------- ---TGTTCAA CATCTT---- ----------
[8]     --TCCCAGCG AAGTG--CGA AACACCTTGT GGACTGTAGT CTACTCGATA C--------C ATGTTCATTT GGAGT----- ---TGTTCAA CATCTC---- ----------
[9]     ---TATAGC- TCACGATCAG TTTGACCTGT GAATA--TGT TTCATCATTA ---------- GGGATGA--C TAAGA----- ---TGTTGAT CATCT----- ----------
[10]    -TCGATGCC- TCACGATCAA TTTGACCTGT GAATA--TGT TTCATCATTA ---------- GGGTTGA--C TAAGA----- ---TGTTAAT CATCT----- ----------
[11]    -TCGATGCC- TTACGATCAA TTTGACCTGC GAATT--TGT TTCATTGTTA ---------- GGGTAGAGAT GAAGA----- ---TGTTGAA CATCT----- ----------
[12]    -TCGATGCCC TCGCTATCAG TTTGACCTGC GAATT--TGT TTCATCAATA ---------- GGGTTGA--T GTAGA----- ---TGTTGAA CATCT----- ----------
[13]    -TCGATGCCT C-ACAATCAG TTCAACTCGC GAACC--TGT TTAACCGCTT ---------- AGGGTTGGCT GGA------- ---TGTCTAA CATCT----- ----------
[14]    -TCGATGCCT C--AAATCAG ATTGACCGTT GAATTCGTGT TATCCACCGA G-----GGGT TGGCTGGCCT GGGAT----- ---CGTGCAA CGTCACGTCT ----------
[15]    -TCGATGCCT C--AAATCAG ATTGACCGTT GAATTCGTGT TATCCACCGA G-----GGGT TGGCTGGCCT GGGAT----- ---CGTGCAA CGTCACGTCT ----------
[16]    -TCGATGCCT C--AAATCAG ATTGACCGGC GAATTCGTGT TATCAACCGA GCAGGGGGGC AGGCTGGGC- GAGACTCACC GTGCGTGCAG TGT------- ----------
[17]    TCTGATGCCT C--AAATCAG ATAAACCGGC GAATTCGTGT TATACACCAA GCAAGGGGGC AGGTTGGGCC GAGACAGACT GTGCGTGCAG TGCAATATCT CACCACACGA
                *         *        *     * *      *                         *                       **  *

                                                                                                                       220
[1]     ------TCAT CTTTCCCCTT A--TGTTGGG AGAAA-GACA TGCCATGCG- ----TGG--- -ATTTCTTCC A----ACAAA A--CATAAAC CCCGACGCAA CATGCGTCAA
[2]     ------TCAT CTTTCCCCTT A--TGTTGGG AGAAA-GACA TGCCATGCG- ----TGG--- -ATTTCTTCC A----ACAAA A--CATAAAC CCCGACGCAA CATGCGTCAA
[3]     ------TCAT CTTTCCTCTT A--TGTTGGG AGCAA-GACA TGCCATGCA- ----TGG--- -ATTTCTTCC A----ACAAA A--CATAAAC CCCGACGCAG CATGCGTCAA
[4]     ------TCAT CTTTCCTCTT A--CGTTGGG AGCAAAGACA TGCCGTGCA- ----TGG--- -CTTTCTTCC A----ACCAA A-ACATAAAC CCCGGCGCAA CATGCGTCAA
[5]     ------TCAT CTTTCCTCTT A--TGTTGGT AGCAAAGACA TGCCGTGCA- ----TGG--- -CTTTCTTCC A----ACCAA ATACATAAAC CCCGGCGCAA CATGCGTCAA
[6]     ------TCAT CTTTCCTTTT A--TGTTGGG AGCAA-GGCA TGCCATGCA- ----TGA--- -CTCTCTCCC A----ACCAA A-ACATAAAC CCCGGCGCAA CATGCGTCAA
[7]     ------AGC- ---TTCCCTT A--TGTTGGG GGTAG-GTT- TATG--CGAT -GCGTGG--- -CCGCCTTCC A----GCAAA A-ACACAAAC CCCGACGCAA CATGCGTCAA
[8]     ------TGAC CTATCCCGTT G--TGCAGGG GGTAG-GTTG TATTGCCGAT TGCGTTGT-- GCTTCGTTCC A----AGAAA A-ACACAACA CCCGAGGCAT CATGCTTCTG
[9]     ------TAAC ATC---CCAT G--AGTTGGA AGTAG-GTTA TGCAATGCA- ----TGG--- -CTTCCTTCC A-----GCGA A-ACTCAAAC CCCGGCGCAA CATGCGTCAA
[10]    ------TAAC ATC---CCTT G--GGTTGGA AGTAG-GTTA TGCAATGCA- ----TGG--- -CTTCCTTCC A-----GCAA A-ACTCAAAC CCCGGCGCAA CATGCGTCAA
[11]    ------TAAC TTCTTCCCTT G--TGGTAGG AGTAG-GTTA TGCAATGCA- ----TCG--- -CTTACTTCC AACCCAGCAA A-ACACAAAC CCCGGCGCAA CATGCGTCAA
[12]    ------TAAC ATC---CCTT C--TGTTGGG AGTAG-GTCA TGCAATGCAA TGCATGG--- -CTGCCCTTC T-TCCAGCAA A-ACACAAAC CCCGACGCAA CATGCGTCAA
[13]    ------TAGC TTC---CCTT G--TGTTGTG GACGGGCTTG TGCTCTGCA- -----CA--- -CCCTCCTTC A-----ACAA ATAGACAAAC CCCGGCGCAA CATGCGTCAA
[14]    ------CGGC CAACTGCCTC C---CCTTGC GTTGG-GCAG CGGCCTCGGC CC-------- --------TT G-CTCGACAA CCACA-AAAC CCCGGCGCTT CGTGTGCCAA
[15]    ------CGGC CAACTGCCTC C---CCTTGC GTTGG-GCAG CGGCCTCGGC CC-------- --------TT G-CTCGACAA CCACA-AAAC CCCGGCGCTT CGTGTGCCAA
[16]    ------CAAC CTCCCCGCCC C-TTGCTTGC GTCAG-GCAG CGGAGGCGGC CTTGCTGCCT ACCTTCCTCT G-CCCGACAG CCACACAAAC CCCGGCGCTT CGTGCGCCAA
[17]    CGTTTCCAAC CTCCCTACCC CCTTGCTTGC ATCGA-GCAG CGGTGGCGGC CT--CTGT-T GCCTACTTCT G-CTCGACAA CCACACAAAC CCCGGCGCTT CGTGCGCCAA
                                                                                                      **    ****  **    * **   *
```

Fig. 1. Multiple alignments of the ITS1, 5.8S and ITS2 nuclear ribosomal DNA sequences from Diocleinae species. The segment comprising the 5.8S rDNA sequence is underlined. The species are: [1] *Dioclea megacarpa*, [2] *Dioclea* sp, [3] *D. bicolor*, [4] *D. virgata*, [5] *D. guianensis*, [6] *Cymbosema roseum*, [7] *Cratylia argentea*, [8] *Camptosema pedicellatum*, [9] *Canavalia boliviana*, [10] *C. gladiata*, [11] *C. bonariensis*, [12] *C. grandiflora*, [13] *Galactia latisiliqua*, [14] *Pachyrhizus tuberosus*, [15] *P. erosus*, [16] *Calopogonium mucunoides*, [17] *C. caeruleum*. Conserved nucleotides are denoted by asterisks (*). Diagnostic nucleotides within the 5.8S sequence that separate *Calopogonium* and *Pachyrhizus* from the true Diocleinae are in bold face. All sequences are deposited in GenBank (accession numbers AY293832-AY293847, and AY343546).

```
                                                                                                                       330
[1]  GGAA-CCTGA AGTTATTTAT GATATGCACT TG--CAGACC C-GGAGACGG -TGCTTCTAC AAGCG----- GTGTGTCGTG AAAACATGT- -----CTCAA AATGACTCTC
[2]  GGAA-CCTGA AGTTATTTAT GATATGCACT TG--CAGACC C-GGAGACGG -TGCTTCTAC AAGCG----- GTGTGTCGTG AAAACATGT- -----CTCAA AATGACTCTC
[3]  GGAA-CTTGA AGTTGTTTAT GATATGCACT TG--TAGACC C-GGAGACGG -TGCTTCTTC AAGTG----- GTGTCGTG-A AAAACATGT- -----CTCAA AATGACTCTC
[4]  GGAA-ATTGA AACTGTTTAT GAT--GCACT TG--CAGGCC C-GGAGACGG -TGCTTCCGC AAGTG----- GTGTCGTG-- -AAACATGT- -----CTTAA AACGACTCTC
[5]  GGAA-ATTGA AACTGTTTAT GAT--GCACT TG--CAGACT C-GGAGACGG -TGCTTCCGC AAGTG----- GTGTCGTG-- -AAACATGT- -----CTTAA AACGACTCTC
[6]  GGAA-ACTTA AGTTGTTTAT GATATGCACT TG--TAGACC C-GGAGACGG -TGCTTCCAC AAGTG----- ATGTTGTG-- -AAACATGT- -----CTTAA AATGACTCTC
[7]  GGAA-CTTGA AATTGTT--T GAT--ACGCT TGTGTAGACC C-GGAGACGG G-GTTTCAAC AAGTG----- ATGTCGTT-- -CAACATGT- -----CTCAA AATGACTCTC
[8]  TCTA-CTTAC AATCGTT--T GAT--ACACT TG-ATTGACC C-GGATACGG GTGTTTCTAC AAGTG----- ATGTCGTT-- -CAACATGT- -----CTCAA AATGACTCTC
[9]  GGAA-CTTGA AATTGTTAAC GAT--GAGCT TG--TAGGTC C-GGAGACGG -TTGCTCTAC ATGTG----- GTGTCGTG-- -AAACATGT- -----CCCAA AATGACTCTC
[10] GGAA-CTTGA AATTGTTAAC GAT--GAGCT TG--TAGATC C-GGAGACGG -TTCCTCTAC ATGTG----- GTGTCGTG-- -AAACATGT- -----CCCAA AATGACTCTC
[11] GGAAACCTCA TATTGTTAAT GAT--GCCCT TG--TAGATT C-GGAGACGG -TGCCTTTAC ATGTG----- GTGTTGTG-- -AAACATGT- -----CCTAA AATGACTCTC
[12] GGAA-CTTGA AATTGTTAAT GAT--ACACT TG--TAGGTC C-GGAGACGG -TTCC--TGC ATGTG----- GTGTCGTG-- -AAACATGT- -----CCCAA AATGACTCTC
[13] GGAA-CCTGG AATTGTT--T GAT--GCGCC TG--TAGACC CCGGAGACGG -TGTTTCTGC AGGTG----- ATGTCGTT-- -CAACATGT- -----ATCAA AATGACTCTC
[14] GGAATC-GAA ACATGTTTGT GAAGTGCGAT TCTCGTGGGC TCGGAGACGA -TGTCCCCAC GAGCGGTCGT TCTTCACGAT ACAATT-GTA TAC-ATACAA AATGACTCTC
[15] GGAATC-GAA ACATGTTTGT GAAGTGCGAT TCTCGTGGGC TCGGAGACGA -TGTCCCCAC GAGCGGTCGT TCTTCACGAT ACAATT-GTA TAC-ATACAA AATGACTCTC
[16] GGAATTTGAA ACATGTTTGT CGCGTGCGAC TCTCGCGGGC CCGGAGACGG -TGTCCC-GC GAGCG----- TCGTCACGAC GCATTGAGTA TACTATGTAA AATGACTCTC
[17] GGAACT-GAA ACATGTTTGT TACGTGCGAT TCTAGCGGGC CCGGAGACGG -TGTCCC-GC GAGCG----- TCGTCACGAC ACACTGAGT- -----TATAA AATGACTCTC
        *            **               *     *      *** ***           *   * *         *         *    **          ** ** *******

                                                                                                                       440
[1]  GGCAACGGAT ATCTCGGCTC TTGCATCGAT GAAGAACGTA GCAAAATGCG ATACTTGGTG TGAATTGCAG AATCCCGTGA ACCATCGAGT TTTTGAACGC AAGTTGCGCC
[2]  GGCAACGGAT ATCTCGGCTC TTGCATCGAT GAAGAACGTA GCAAAATGCG ATACTTGGTG TGAATTGCAG AATCCCGTGA ACCATCGAGT TTTTGAACGC AAGTTGCGCC
[3]  GGCAACGGAT ATCTCGGCTC TTGCATCGAT GAAGAACGTA GCAAAATGCG ATACTTGGTG TGAATTGCAG AATCCCGTGA ACCATCGAGT TTTTGAACGC AAGTTGCGCC
[4]  GGCAACGGAT ATCTCGGCTC TTGCATCGAT GAAGAACGTA GCAAAATGCG ATACTTGGTG TGAATTGCAG AATCCCGTGA ACCATCGAGT TTTTGAACGC AAGTTGCGCC
[5]  GGCAACGGAT ATCTCGGCTC TTGCATCGAT GAAGAACGTA GCAAAATGCG ATACTTGGTG TGAATTGCAG AATCCCGTGA ACCATCGAGT TTTTGAACGC AAGTTGCGCC
[6]  GGCAACGGAT ATCTCGGCTC TTGCATCGAT GAAGAACGTA GCAAAATGCG ATACTTGGTG TGAATTGCAG AATCCCGTGA ACCATCGAGT TTTTGAACGC AAGTTGCGCC
[7]  GGCAACGGAT ATCTCGGCTC TTGCATCGAT GAAGAACGTA GCAAAATGCG ATACTTGGTG TGAATTGCAG AATCCCGTGA ACCATCGAGT CTTTGAACGC AAGTTGCGCC
[8]  GGCCACGGAT ATCTCGGCTC TTGCATCGAT GAAGCACGTA GCAAAATGCG ATACTTGGTG TGAATTGCAG GATCCCGTGG ACCATCGAGT CTTTGAACGC AAGTTGCGCC
[9]  GGCAACGGAT ATCTCGGCTC TTGCATCGAT GAAGAACGTA GCAAAATGCG ATACTTGGTG TGAATTGCAG AATCCCGTGA ACCATCGAGT CTTTGAACGC AAGTTGCGCC
[10] GGCAACGGAT ATCTCGGCTC TTGCATCGAT GAAGAACGTA GCAAAATGCG ATACTTGGTG TGAATTGCAG AATCCCGTGA ACCATCGAGT CTTTGAACGC AAGTTGCGCC
[11] GGCAACGGAT ATCTCGGCTC TTGCATCGAT GAAGAACGTA GCAAAATGCG ATACTTGGTG TGAATTGCAG AATCCCGTGA ACCATCGAGT CTTTGAACGC AAGTTGCGCC
[12] GGCAACGGAT ATCTCGGCTC TTGCATCGAT GAAGAACGTA GCAAAATGCG ATACTTGGTG TGAATTGCAG AATCCCGTGA ACCATCGAGT CTTTGAACGC AAGTTGCGCC
[13] GGCAACGGAT ATCTCGGCTC TTGCATCGAT GAAGAACGTA GCAAAATGCG ATACTTGGTG TGAATTGCAG AATCCCGTGA ACCATCGAGT CTTTGAACGC AAGTTGCGCC
[14] GGCAACGGAT ATCTCGGCTC TTGCATCGAT GAAGAACGTA GCGAAATGCG ATACTTGGTG TGAATTGCAG AATCCCGTGA ACCATCGAGT CTTTGAACGC AAGTTGCGCC
[15] GGCAACGGAT ATCTCGGCTC TTGCATCGAT GAAGAACGTA GCGAAATGCG ATACTTGGTG TGAATTGCAG AATCCCGTGA ACCATCGAGT CTTTGAACGC AAGTTGCGCC
[16] GGCAACGGAT ATCTCGGCTC TTGCATCGAT GAAGAACGTA GCGAAATGCG ATACTTGGTG TGAATTGCAG AATCCCGTGA ACCATCGAGT CTTTGAACGC AAGTTGCGCC
[17] GGCAACGGAT ATCTCGGCTC TTGCATCGAT GAAGAACGTA GCGAAATGCG ATACTTGGTG TGAATTGCAG AATCCCGTGA ACCATCGAGT CTTTGAACGC AAGTTGCGCC
     *** ****** ********** ********** **** ***** ** ******* ********** **********  ********  **********  ********* **********
```

Fig. 1 (*continued*)

```
                                                                                                                         550
[1]    TGAAACCATT AGGTTGAGGG CACGCCTGCC TGGGTGTCAC ACAATGTTTC CC-TATGCCC CAACATC--A ----AATTGG CGGTGAGA-A GGGTGAATGT TGGCTTCCTG
[2]    TGAAACCATT AGGTTGAGGG CACGCCTGCC TGGGTGTCAC ACAATGTTTC CC-TATGCCC CAACATC--A ----AATTGG CGGTGAGA-A GGGTGAATGT TGGCTTCCTG
[3]    TGAAACCATT AGGTTGAGGG CACGCCTGCC TGGGTGTCAC ACAACGTTTC CC-TATACC- -AATGTC--A ----AATTGG TGTTGAGA-A GGGTGAATGT TGGCTTCCCG
[4]    TGAAACCATT AGGTTGAGGG CACGCCTGCC TGGGTGTCAC ACAATGTTTC CC-TGCACCA ACACCTTTGA ----AATTGG TGTTGAGC-A GGGTGAACGC TGGCTTCCCG
[5]    TGAAACCATT AGGTTGAGGG CACGCCTGCC TGGGTGTCAC ACAATGTTTC CC-TGCACCA ACGCCTTTGA ----AATTGG TGTCGAGC-A GGGTGAACGC TGGCTTCCCG
[6]    TGAAACCGTT AGGTTGAGGG CACGTCTGCC TGGGTGTCAC ACGATGTTTC CC-TGCTCCA ACACCTTTGA ----AGTTGG CGTTGAGC-A GGGTGAATTT TGGCTTCCCG
[7]    TGAAACCATT AGGTTGAGGG CACGTCTGCC TGGGTGTCAC ACAATGTTGT CC-TATGCCA ATGCCTTTGA -ATAATC-GG TGTTGAGC-A AGGTGAATGT TGGCTTCCCA
[8]    TGAAACCATT AGGTTGAGGG CACGTCTGCC TGGGTGTCAC ACAATGTTGC CC-TATGCCA ATTCCTTTGA -ATTATC-GG TGTTGAGC-A GGGCGAATGT TGGCTTCCCA
[9]    CGAAGCCATT AGGTTGAGGG CACGCCTGCC TGGGTGTCAC ATAATGTTCC CC-TATACCT ATGCCTTTTC CA--ATAAGG TATTTTGT-G GGGTGAAGGT TGGCTTCCCA
[10]   CGAAGCCATT AGGTTGAGGG CACGCCTGCC TGGGTGTCAC ACAATGTTCC CCCTACACCT ATGCCTTTTC CA--ATAAGG TATTTTGT-G GGGTGAAGGT TGGCTTCCCA
[11]   CGAAGCCATT AGGTTGAGGG CACGCCTGCC TGGGTGTCAC ACAATGCTCC CC-TATACCT ATGCCTTTTA CA--ATTAGG TGTTTTGT-A GGGTGTATGT TGGCTTCCCA
[12]   CGAAGCCATT AGGTTGAGGG CACGCCTGCC TGGGTGTCAC ACAATGTTCC CC-TATACCT ATACCTTTTA TATAATAAGG TATCTTGTTG GGGTAAATGT TGGCTTCCCA
[13]   TGAAACCATT AGGTTGAGGG CACGTCTGCC TGGGTGTCAC ACAATGTTGC CC-AACGCCA ACGCCGTTTC CACGATC-GG CGTTTTAGCG GGGCGAAAGT TGGCTTCCCA
[14]   CAAAGCCATT AGGCTGAGGG CACGCCTGCC TGGGTGTCAC ACATCGTCAC CC-TCCACAC GCACACACAC --------TG TGCAGGAGGA GGGTGTATGC TGGCTTCCCG
[15]   CAAAGCCATT AGGCTGAGGG CACGCCTGCC TGGGTGTCAC ACATCGTCAC CC-TCCACAC GCACACACAC --------TG TGCAGGAGGA GGGTGTATGC TGGCTTCCCG
[16]   CAAAGCCATT AGGCTGAGGG CACGCCTGCC TGGGTGTCAC ACATCGTTAC CC-TC-ACGC AAAC-TGTGC --------CG TGCTG----- -GGTGGACGC TGACTTCCCG
[17]   CAAAGCCATT AGGCCGAGGG CACGCCTGCC TGGGTGTCAC ACATCGTTAC CC-TC-ACGC AAAAACGTGC --------CG TGCTG----- -GGTGGACGC TGACTTCCCG
         ** ** ** ***  ***** **** ***** ********** *    *     **     *                       *             **   *    ** *****

                                                                                                                         660
[1]    CGAGC-ATGG TCTCGTGGTT GGTTGAAAAA TAAGTCCTCG GCGGAATGCT CCATGATAA- ATGGTAGTTG AGTAA--TGC TCGAGACCA- ATCATGTGTG GCTTCCT-C-
[2]    CGAGC-ATGG TCTCGTGGTT GGTTGAAAAA TAAGTCCTCG GCGGAATGCT CCATGATAA- ATGGTAGTTG AGTAA--TGC TCGAGACCA- ATCATGTGTG GCTTCCT-C-
[3]    CGAGC-ACGG TCTCGTGGTT GGTTGAAAAA TGAGTCCTCG GCGGAATGCT TCATGATAA- ATGGTAGTTG AGTGA--TGC TCGAGACCA- ATCATGTGTG GCTTCCT-C-
[4]    CGAGC-GTGG CCTCGCGGTT GGTTGAAAAA TGAGTCCTCG GTGGAATGCT CCATGACGA- ATGGTAGTTG AGTGA--TGC TCGAGACCA- ATCATGGTTT GCTTCCT-T-
[5]    CGAGC-GTGG CCTCGCGGTT GGTTGAAAAA TGAGTCCTCG GTGGAATGCT CCATGACGA- ATGGTAGTTG AGTGA--TGC TCGAGACCA- ATCATGGTTT GCTTCCT-T-
[6]    CAAGC-ATGG TCTTGTGGTT GGTTGAAAAA TGAGTCCTTG GTGGAATGCT CCATGATGA- ATGGTAGTTG AGTGA--TGC TCGAGACCA- ATCATGTGTT GCTTCCCAT-
[7]    CGAGC-ATCG TCTCGTGGTT GGTTGAAAAA TGAGTCCTCG ATGGAATGCT TCATGATAA- ATGGTGGTTG AGTCA--TGC TCGAGGCCA- AACATGTGTG GCTTCCTTC-
[8]    CGAGC-ATCG TCTCGTGGTT GGTTGAAAAA TGAGTCCTCG GCGGAATGCT CCATGATAAG ATGGTGCTTG AGTGA--TGC TCGAGATCT- ATCATGCTTG GCGTCCTGCG
[9]    TAAGC-ATTG CCTTGTGGTT GGTTGAAATA TGAGTCCTTG GTGGAATGCC CCATGATAA- ATGGTAGTTG AGTGA--TCC TCGAGGCCA- ATCATGCATG GTTTCCTCT-
[10]   TAAGC-ATTG CCTTGTGGTT GGTTGAAATA TGAGTCCTTG GTGGAATGCC CCATGATAA- ATGGTAGTTG AGTGA--TCC TCGAGGCCA- ATCATGCATG GCTTCCTCT-
[11]   CAAGC-ATTG TCTTGTGGTT GGTTGAAAAA TGAGTCCTTG GTGGAATGCC CCATGATAA- ATGGTAGTTG AGTGA--TGC TCGAGACCA- ATCATGCATG GCTTCCTCT-
[12]   CAAGTTGTTG TCTTGTGGTT GGCTGAAAAA TGAGTCCTTG GTGGAATGCC CCATGATAA- ATGGTGGTTG AGTGA--TCC TCGAGACCA- ATCATGCATG GCTTCCACT-
[13]   CGAGC-TCTG TCCCGTGGTT GGTTGAAAAA TGAGTCCTCG GTGGAATGCT CCATGAGAA- ATGGTGGTTG AGTGA--TGC TCGAGACCA- AACATGCGTG GCTTCCCCC-
[14]   CGAGC-ACCG ACTCGCGGTT GGTTGAAAAT CGAGTTCGCG ACAGTCTGCG TCGTGATAAA ATGGTGGATG GGCAACATGC TCGAGACCG- ATCACACGCG ACATGTC---
[15]   CGAGC-ACCG ACTCGCGGTT GGTTGAAAAT CGAGTTCGCG ACAGTCTGCG TCGTGATAAA ATGGTGGATG GGCAACATGC TCGAGACCG- ATCACACGCG ACATGTC---
[16]   TGAGC-ACCG ACTCGCGGTT GGTTGAAAAC TTAGTCCACG GCCGTCTGCG TCGTGATCAA ACGGTGGATG GGCAACACGC TCGAGACCA- ATCACACGCG ACGCGCC---
[17]   CGAGC-ACCG ACTCGCGGTT GGTTGAAAAC TTAGTCCACG GCTGTCTGCG TCGTGATCAA ATGGTGGATG GGCAACATGC TCGAGACCAG ATCACACGCG ACACGTC---
         **     *  *  * **** ** *****     *** *  *    *  ***   * ***  *  *   ***   **  *  *    * *****  *   * **
```

Fig. 1 (*continued*)

```
                                                                    713
[1]   CTAAACGTGG GCC-----TT GACCCTTGCA TTTTCT--TA AGAATGCTCA TAT
[2]   CTAAACGTGG GCC-----TT GACCCTTGCA TTTTCT--TA AGA-TGCTCA TAT
[3]   CTAAATGTGG GCT-----TT GACCCTTGCG TTTTCT--TG AGAACGCTCA TAT
[4]   CCAAAGGAGG GCC-----TT GACCCTTGCG TTTTTTG-CG AAAACGCTCA TAT
[5]   CTAAAGGAGG GCC-----T- GACACTTGCG TTTTTT--CG AAAACGCTCA TAT
[6]   CTAAATGTGG GCC-----TT GACCCTTACG TTTTCT--TG AAAACGTTCA TAT
[7]   -TAAATGTGG GAC---CTTT GACCCTTGCG TTCTCTTAC- AAC--GCTCA CAA
[8]   CTGTATCCGG GTCGCTCTTT ACCGCCCGCG TTCTCCTAC- AATCTGTTCA CTA
[9]   -CACATTTGG ACC----CTG ATCCCTTGAG TTTTCTCT-- AGAACACTCA TAA
[10]  -CACATTTGG ACC----TTG ACCCCTTGAG TTTTCTTT-- AGA-CACTCA TAA
[11]  -TACAATTGG ACC----TTG ACCCTGTGTT CTCTCTTTCG AGAACACTCA CAA
[12]  -TAAATGTGG ACT----TTT GACCCTTGTG TTTTT----- AGAACACTCG ATA
[13]  --TAAGGTGG ACG---TTCG ACCCCTTGCG TCCTAAAAC- -GACGTCTCA CAA
[14]  --GATTTCGG ACT---CTTT GATCCCTGTG CGTCT----G TGGACGCTCT TCA
[15]  --GATTTCGG ACT---CTTT GATCCCTGTG CGTCT----G TGGACGCTCT CA-
[16]  --GCTCCTGG ACT---TGCT GATCCCCACG CGTCTT--TG TGCACGCTCT CAA
[17]  --GTTTCCGG ACT---CACT GATCCCCACG CGTCTT--TG TGCACGCTCT CAA
            **                                        **
```

Fig. 1 (*continued*)

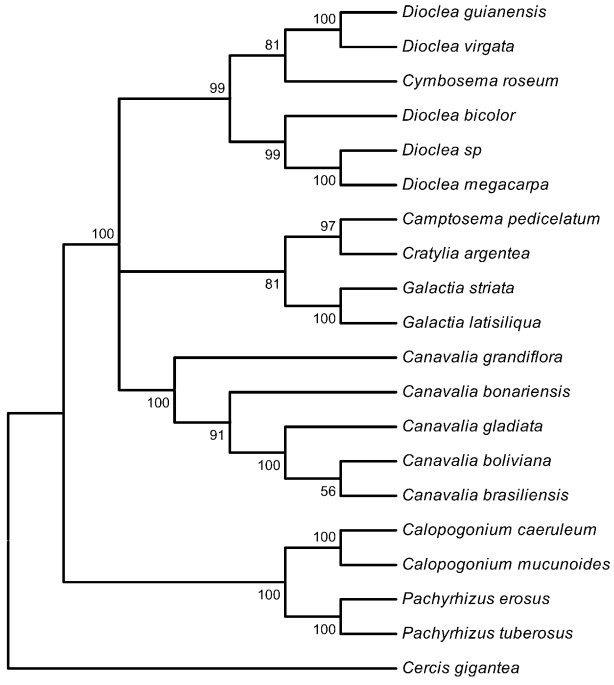

Fig. 2. Neighbour-joining (NJ) tree based on the Tamura–Nei distance algorithm from ITS/5.8S sequences from Diocleinae and the anomalous *Calopogonium* and *Pachyrhizus* genera. *Cercis gigantea* (Caesalpinioideae) was used as the outgroup. Bootstrap values (500 replicates) are shown at each node.

Fig. 3. Consensus maximum parsimony (MP) tree derived from ITS/5.8S sequences from Diocleinae members and the anomalous *Calopogonium* and *Pachyrhizus* genera. *Cercis gigantea* (Caesalpinioideae) was used as the outgroup. Bootstrap values (500 replicates) are shown at each node.

support previous leaflet anatomy studies on Phaseoleae that showed that all *Canavalia* (including Wenderothia) consistently share the presence of epidermal crystals which are distributed in a distinctive paired arrangement (Lackey, 1978). This particular anatomic trait has been considered a unifying trend for this genus (Lackey, 1981).

Lackey (1977a) had questioned the maintenance of the monotypic Brazilian genus *Cymbosema* as a distinct genus from *Dioclea*. In his words, "*Cymbosema* differs from *Dioclea* by the free vexillary stamen". Pollen morphology analyses placed *Cymbosema* with the *Dioclea* subgenus (Kavanagh and Ferguson, 1981). In the NJ, MP and ML trees obtained using ITS/5.8S sequences (Figs. 2–4), *Cymbosema* arose as a sister taxon of subgenus Dioclea, inside the *Dioclea* clade. This shows that *Dioclea* is not monophyletic. Therefore to maintain the monophyly of *Dioclea*, *C. roseum* should be treated as a *Dioclea* species. To our best knowledge, this is the first molecular sequence that shows the close relationship

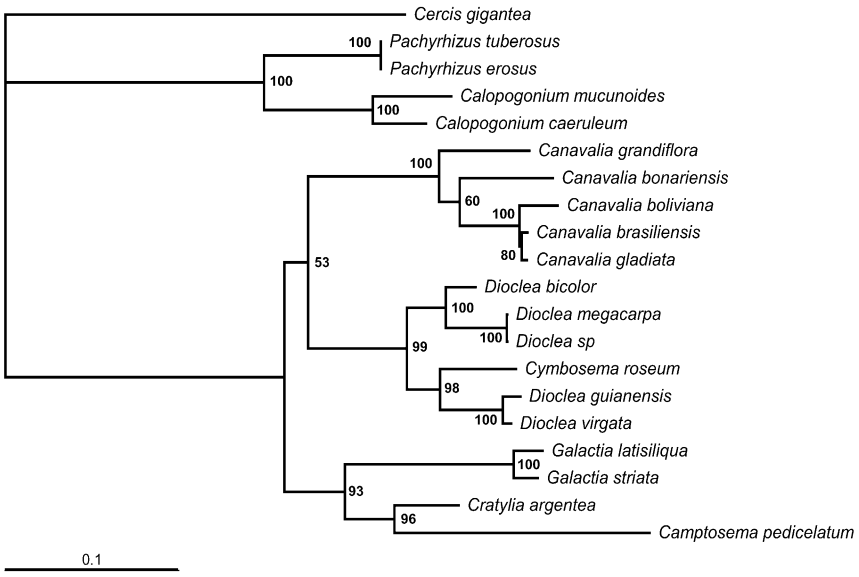

Fig. 4. Maximum likelihood (ML) tree derived from ITS/5.8S sequences from Diocleinae members and the anomalous *Calopogonium* and *Pachyrhizus* genera. *Cercis gigantea* (Caesalpinioideae) was used as the outgroup. Bootstrap values (100 replicates) are shown at each node. The scale bar represents the branch lengths (0.1 nucleotide substitutions per site).

between *Cymbosema* and *Dioclea*, providing additional support for the previous morphological studies (Lackey, 1977a; Kavanagh and Ferguson, 1981). It is apparent that a revision about the status of *Cymbosema roseum* is needed in the light of these data.

In the NJ, MP and ML trees (Figs. 2–4), the two *Galactia* species form a cluster, which arises as a sister group of a subclade comprising *Camptosema* and *Cratylia*. The *Camptosema–Cratylia–Galactia* complex is also supported suggesting a common ancestor. Existing taxonomic evidence shows that *Camptosema*, *Cratylia* and most of the *Galactia* species have similar pollen types (Kavanagh and Ferguson, 1981). It is also noteworthy that some species that were described before as belonging to *Camptosema* have been transferred to either *Cratylia* or *Galactia* genera (Queiroz, 1999).

Lackey (1981) realized that the subtribe Diocleinae was difficult to define except in a negative way. Although he had included *Calopogonium* and *Pachyrhizus* in Diocleinae (Lackey, 1977b), he questioned whether these genera might be improperly placed in this subtribe on the basis of some anomalous morphological characters, such as the peculiar stigma-style structure of *Pachyrhizus* and the anomalous somatic chromosome number ($2n = 36$) of *Calopogonium*. In addition, because he could not recognize any distinctive features that could be used to place them elsewhere, they were therefore retained in the subtribe. The phylogeny derived from ITS/5.8S region sequences reported here does support the exclusion of *Calopogonium* and *Pachyrhizus* from Diocleinae sensu Lackey (1981) by Polhill (1994). The segregation of *Calopogonium* and *Pachyrhizus* from Diocleinae in the ITS/5.8S phylogeny is congruent both with morphological and phytochemical data. Kavanagh and Ferguson (1981) examined the pollen morphology of the 13 genera included in the subtribe Diocleinae. They found that pollen type clearly separates the genera *Calopogonium* and *Pachyrhizus* from the rest of the subtribe. Ingham (1990) carried out a survey of phytoalexins within the tribe Phaseoleae, and reported that *Calopogonium* and *Pachyrhizus* produce phytoalexins sufficiently distinct from other Diocleinae. *Calopogonium* and *Pachyrhizus* produce complex isoflavonoids of various types that are also found in many representatives of the Phaseoleae subtribes Glycininae, Kennediinae and Phaseolinae. Indeed the inclusion of *Calopogonium* and *Pachyrhizus* in subtribe Glycininae has been supported by a restriction-site mapping study of the cpDNA inverted repeat region (Doyle and Doyle, 1993) as well as by a phylogenetic study based on cpDNA rps 16 intron sequences (Lee and Hymowitz, 2001). To test this hypothesis using the nuclear sequence information, a MP tree was constructed including ITS/5.8S sequences from Glycininae species available in GenBank at the time of this study. Sequences of only three Glycininae genera were available (*Amphicarpaea*, *Glycine* and *Pueraria*), and representative species for each genus were randomly chosen (GenBank accession numbers AF338215, AF417013, AF417017, GMU60551 and AF023444). To give the hypothesis on the phylogenetic affiliation of *Calopogonium* and *Pachyrhizus* a stronger support, ITS/5.8S sequences from the subtribe Phaseolinae were also included in the ingroups (GenBank accession numbers AF115135, AF115137-AF115139,

AF069120, AF069114-AF069117, AF115239, AF009125, AF069123, AF069122). Twenty-four most equally parsimonious trees were generated, and the bootstrap consensus tree is shown in Fig. 5. Each generated tree has a length of 806 steps, a CI of 0.481, a RI of 0.785, and a RC of 0.378. The main finding of this analysis is that the *Calopogonium–Pachyrhizus* complex arises as a sister group of the Glycininae genera, and the Glycininae clade (including *Calopogonium* and *Pachyrhizus*) is very well supported. Therefore, the phylogeny produced by the nuclear ITS/5.8S sequence (herein) is congruent to that obtained using a cpDNA gene sequence (Lee and Hymowitz, 2001).

In conclusion, the ITS/5.8S phylogeny reported in this work unambiguously supports the hypothesis that *Calopogonium* and *Pachyrhizus* do not belong to subtribe Diocleinae. The nuclear sequence data suggest that these genera are more closely related to Glycininae species. In addition, the ITS/5.8S sequences from the true Diocleinae also provide informative sites to resolve relationships below the subtribe level.

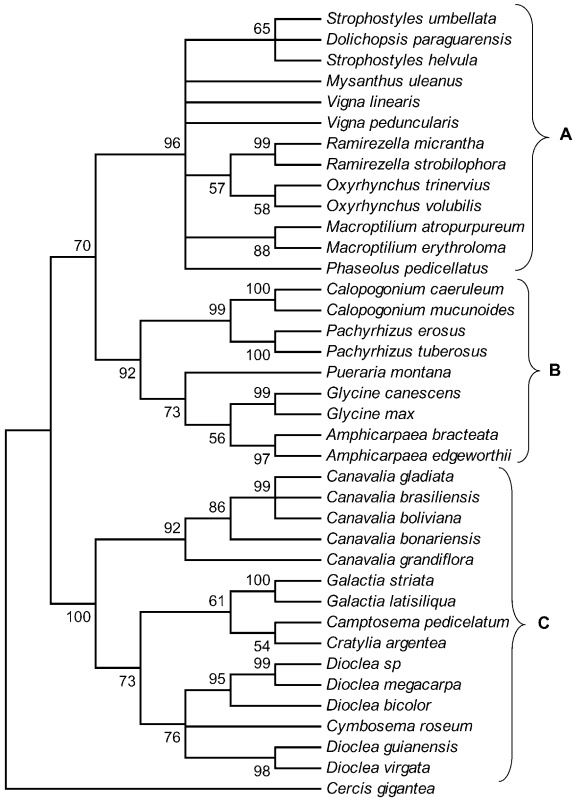

Fig. 5. Consensus maximum parsimony (MP) tree derived from ITS/5.8S sequences showing the close relationship of *Calopogonium* and *Pachyrhizus* genera with some Glycininae members (clade B). Sequences from subtribe Phaseolinae (clade A) were also included. The subtribe Diocleinae clade in indicated by C. *Cercis gigantea* (Caesalpinioideae) was used as the outgroup. Bootstrap values (500 replicates) are shown at each node.

3. Experimental

3.1. Plant material

Mature seeds were harvested from plants growing wild in different localities of the states of Ceará, Amazonas and Rio Grande do Sul, in Brasil. Seeds were stored at 4 °C until used. The species were: *Calopogonium caeruleum* Hemsl., *C. mucunoides* Desv., *Camptosema pedicellatum* Benth., *Canavalia gladiata* DC., *C. grandiflora* Benth., *C. bonariensis* Lindl., *C. boliviana* Piper, *Cratylia argentea* (Desv.) Kuntze, *Cymbosema roseum* Benth., *Dioclea* sp., *Dioclea bicolor* Benth., *D. guianensis* Benth., *D. megacarpa* Rolfe, *D. virgata* (L.C.Rich.) Amshoff, *Galactia latisiliqua* Desv., *Pachyrhizus erosus* (L.) Urb., and *P. tuberosus* Spreng. Voucher specimens are deposited in the Herbarium Prisco Bezerra, Universidade Federal do Ceará, Fortaleza-Ceará, Brasil. Voucher numbers are shown in Table 1.

3.2. DNA purification

Genomic DNA was purified from fresh leaves of 1 week-old seedlings grown from mature seeds. The protocol for DNA purification uses the CTAB (cetyl triethylammonium bromide) reagent, as described by Foster and Twell (1996). The concentration of DNA in the various samples was determined by measuring the absorbance at 260 nm (A_{260}) of a ten-fold dilution of each sample. The quality of all DNA preparations was checked by 0.8% agarose gel electrophoresis according to Sambrook et al. (1989).

3.3. PCR amplification and DNA sequencing

Amplification reactions were performed in a final volume of 25 μl containing 500–800 ng of genomic DNA (template), 20 mM Tris–HCl, pH 8.4, 50 mM KCl, 1.5 mM $MgCl_2$, 100 μM of each dATP, dCTP, dGTP and dTTP (Amersham Biosciences, Sweden), 5 pmoles of each primer and 0.5 units of Taq DNA Polymerase (Amersham Biosciences, Sweden). PCR reactions were carried out in a MJ-Research Inc. (Watertown, MD) PTC-100 thermocycler programmed for an initial denaturation step (3 min at 94 °C) followed by 45 cycles of 1 min at 94 °C, 1 min at 55 °C, and 2 min at 72 °C. The last cycle was followed by a final incubation of 10 min at 72 °C. The samples were then stored at 4 °C until used. Amplified fragments were analyzed by standard horizontal electrophoresis on 1.0% agarose gels in TBE buffer (10 mM Tris-borate, 1 mM EDTA, pH 8.0) at 100 V. The DNA bands were stained with 0.5 μg/ml ethidium bromide as described before (Sambrook et al., 1989). Control samples containing all reaction components except DNA were always used to test that no self-amplification or DNA contamination occurred. The

primers used for amplification were ITS4 (TCCTCCGCTTATTGATATGC) and ITS5 (GCAAGTAAAAGTCGTAACAAGG), as suggested by Becerra and Venable (1999). These primers are complementary to the end of the 18S rDNA and to the beginning of the 26S rDNA, therefore they amplify a fragment of nrDNA containing ITS1, 5.8S rDNA and ITS2.

Sequences of PCR products were determined using the DYEnamic ET terminators sequencing kit (Amersham Biosciences, Upsala, Sweden) following the protocol supplied by the manufacturer. Sequencing reactions were then analysed in a MegaBACE 1000 automatic sequencer (Amersham). Each PCR product was sequenced four times in both directions using the same primers (ITS4 and ITS5) previously described.

3.4. Sequence alignment and phylogenetic analyses

The quality of DNA sequences was checked and overlapping fragments were assembled using Phred/Phrap/Consed package (http://www.phrap.org). The sequence boundaries between the two spacers (ITS1 and ITS2) and the three coding regions (18S, 5.8S and 26S) of nrDNA were determined by comparison with a published sequence from *Vicia faba* (Yokota et al., 1989). Assembled sequences with high quality (phred $>$ 20) were aligned using CLUSTALX (Thompson et al., 1997), with default gap penalties. The corresponding ITS/5.8S sequence from *Cercis gigantea* (Caesalpinioideae) was used as outgroup (GenBank accession number AF390194). The multiple alignments were then used to construct phylogenetic trees using distance matrix (neighbour-joining) and maximum parsimony methods programs of MEGA (Molecular Evolutionary Genetics Analysis) software (Kumar et al., 2001). In the neighbour-joining (NJ) analyses genetic distances were calculated based on the Tamura–Nei algorithm. The choice of the most suitable substitution model to be used for the data set in the NJ analyses was previously calculated using ModelTest program (Posada and Crandall, 1998). Clades stability in NJ and MP trees was assessed by bootstrap analysis (Felsenstein, 1985) with 500 replicates. Maximum likelihood analysis was done with the DNAML program of PHYLIP package (Felsenstein, 1993). A bootstrap analysis of these data was carried out using 100 resampled data sets, which were generated using the SEQBOOT program. PHYLIP's CONSENSE program was then implemented in order to construct a strict consensus tree. Trees generated using PHYLIP were visualized by TreeView (Page, 1996).

Acknowledgements

This work was supported by grants from Conselho Nacional de Desenvolvimento Cientifico e Tecnologico (CNPq), and Fundação Cearense de Apoio ao Desenvolvimento Cientifico e Tecnologico (FUNCAP). We are grateful to Drs Benildo S. Cavada (Dept. Bioquimica e Biologia Molecular-UFC) and Danilo F. da Silva Filho (INPA) for providing us some plant material used in this work. The authors are also indebted to Dr Afranio G. Fernandes (Herbarium Prisco Bezerra-UFC) for his invaluable contribution for the correct identification of all specimens studied. The corresponding author (T. B. Grangeiro) has a research fellowship (PQ-2B) from CNPq.

References

Baldwin, B.G., Sanderson, M.J., Porter, J.M., Wojciechowski, M.F., Campbell, C.S., Donoghue, M.J., 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. Ann. Missouri Bot. Gard. 82, 247–277.

Becerra, J.X., Venable, D.L., 1999. Nuclear ribosomal DNA Phylogeny and its implications for evolutionary trends in Mexican *Burserea* (Burseraceae). Am. J. Bot. 86, 1047–1057.

Doyle, J.J., Doyle, J.L., 1993. Chloroplast DNA phylogeny of the Papilionoid legume tribe Phaseoleae. Syst. Bot. 18, 309–327.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Felsenstein, J., 1993. PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.

Foster, G.D., Twell, D., 1996. Plant Gene Isolation. Principles and practice. John Wiley & Sons, England.

Hillis, D.M., Bull, J.J., 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42, 182–192.

Hu, J., Lavin, M., Wojciechowski, M.F., Sanderson, M.J., 2002. Phylogenetic analysis of nuclear ribosomal ITS/5.8S sequences in the tribe Millettieae (Fabaceae): *Poecilanthe–Cyclolobium*, the core Millettieae, and the *Callerya* group. Syst. Bot. 27, 722–733.

Ingham, J.L., 1990. Systematic aspects of phytoalexin formation within tribe Phaseoleae of the Leguminosae (Subfamily Papilionoideae). Biochem. Syst. Ecol. 18, 329–343.

Käss, E., Wink, M., 1997. Phylogenetics relationships in the Papilionoideae (Family Leguminosae) based on nucleotide sequences of cpDNA (rbcL) and ncDNA (ITS 1 e ITS 2). Mol. Phylogenet. Evol. 8, 65. - 88.

Kavanagh, T.A., Ferguson, I.K., 1981. Pollen morphology and taxonomy of the subtribe Diocleinae (Leguminosae, Papilionoideae, Phaseoleae). Rev. Palaeobot. Palynol. 32, 317–367.

Kumar, S., Tamura, K., Jakobsen, I.B., Nei, M., 2001. MEGA2: Molecular Evolutionary Genetics Analysis software. Arizona State University, Tempe, AZ.

Kuzoff, R.K., Sweere, J.A., Soltis, D.E., Soltis, P.S., Zimmer, E.A., 1998. The phylogenetic potential of entire 26S rDNA sequences in plants. Mol. Biol. Evol. 15, 251–263.

Lackey, J.A., 1977a. A synopsis of Phaseoleae (Leguminosae, Papilionoideae). PhD thesis, Iowa State University, Ames, IA.

Lackey, J.A., 1977. A revised classification of the tribe Phaseoleae (Leguminosae: Papilionoideae), and its relation to canavanine distribution. Bot. J. Linn. Soc. 74, 163–178.

Lackey, J.A., 1978. Leaflet anatomy of Phaseoleae (Leguminosae, Papilionoideae) and its relation to taxonomy. Bot. Gaz. 139, 436–446.

Lackey, J.A., 1981. Tribe 19. Phaseoleae DC. (1825). In: Polhill, R.M., Raven, P.H. (Eds.), Advances in Legume Systematics, vol. 1. Royal Botanic Gardens, Kew, UK, pp. 301–327.

Lee, J., Hymowitz, T., 2001. A molecular phylogenetic study of the

subtribe glycininae (leguminosae) derived from chloroplast DNA rps16 intron sequences. Am. J. Bot. 88, 2064–2073.
Maxwell, R.H., 1969. The Genus Dioclea (Fabaceae) in the New World. PhD thesis. Southern Illinois University, IL.
Page, R.D.M., 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. 12, 357–358.
Piper, C.V., 1925. The American species of *Canavalia* and *Wenderothia*. Contrib. U.S. Nat. Herb. 20, 555–588.
Polhill, R.M., Raven, 1981. Advances in Legume systematics, Vol. 1.. Royal Botanical Gardens, Kew.
Polhill, R.M., 1994. Classification of the Leguminosae. In: Bisby, F.A., Buckingham, J., Harborne, J.B. (Eds.), Phytochemical Dictionary of the Leguminosae. Chapman & Hall, New York, pp. 45–47.
Posada, D., Crandall, K.A., 1998. ModelTest: testing the model of DNA substitution. Bioinformatics 14, 817–818.
Queiroz, L.P., 1999. Sistemática e filogenia do gênero *Camptosema* W.J.Hook & Arn. (Leguminosae: Papilionoideae: Phaseoleae). Thesis, Universidade de Sao Paulo, Sao Paulo.
Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning: a Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
Sauer, J., 1964. Revision of *Canavalia*. Britonia 16, 106–181.
Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 24, 4876. - 4882.
Yokota, Y., Kawata, T., Iida, Y., Kato, A., Tanifuji, S., 1989. Nucleotide sequences of the 5.8S rRNA gene and internal transcribed spacer regions in carrot and broad bean ribosomal DNA. J. Mol. Evol. 29, 294–301.