Systematic Botany (2008), 33(2): pp. 229–236

Phylogenetic Relationships of the Neotropical Genus *Dioon* (Cycadales, Zamiaceae) Based on Nuclear and Chloroplast DNA Sequence Data

Dolores González,[1] Andrew P. Vovides, and Cristina Bárcenas

Instituto de Ecología, A.C., Apartado Postal 63, Xalapa Veracruz, 91070 México
[1]Author for correspondence (dolores.gonzalez@inecol.edu.mx)

Communicating Editor: Lena Struwe

***Abstract*—**Previous hypotheses of relationships within *Dioon* Lindl. indicated the presence of two large clades within the genus. However, relationships among species still remained unresolved. In this study, molecular phylogenetic analyses were performed with individual and combined data sets from the ITS region of the ribosomal DNA and the *trnL–F* region of the chloroplast DNA. To explore whether indels were a source of phylogenetically informative characters, indels were analyzed by excluding them from the analyses (Coding A); including them as ambiguous data (Coding B); as multistate (Coding C) and as binary characters (Coding D). We found that the rate of mutation in the ITS region is appropriate to solve most relationships at the species level. This is in contrast with the *trnL–F* rgion, which showed little variation. Our results show that most clades obtained during analyses correspond with previously recognized species within *Dioon*. This phylogeny groups the genus into two main clades that show clear biogeographic relationships between the species occurring along the Pacific Sierra Madre Occidental and the Atlantic Sierra Madre Oriental.

***Keywords*—**Cycads, molecular phylogeny, *Dioon* biogeography, ITS, *trnL–F*.

The Neotropical cycad genus *Dioon* Lindl. (Zamiaceae) includes 12 species from Mexico and one from Honduras (Hill et al. 2004). *Dioon* is distributed mostly within mountain systems along the Sierra Madre Oriental and Sierra Madre Occidental. On the eastern seaboard, the range is from northeastern Nuevo León to southern Veracruz and northern Oaxaca. There is a vicariant *Dioon* in Honduras (*D. mejiae* Standl. & L.O. Williams) on the Caribbean coast. Along the Pacific seaboard, the range of *Dioon* extends from northwestern side of the Sierra Madre Occidental from central Sonora State to southern Chiapas (Fig. 1). The genus is commonly divided into two distinct morphological groups. One group contains species characterized generally by large fronds, large trunks, and massive cones, and the other group has species with generally shorter trunks, considerably shorter fronds, and smaller cones (Norstog and Nicholls 1997). These groups were also recovered in phylogenetic studies performed with morphology and chloroplast DNA (cpDNA) RFLPs (Moretti et al. 1993) and with sequence data (Bogler and Francisco-Ortega 2004). However, species relationships have not been clearly established because several morphological characters overlap among species. In the consensus tree based on morphology, all species of *Dioon* were divided into two major clades. The Spinulosum clade was fully resolved whereas the Edule clade collapsed in a polytomy from which only *D. edule* and *D. tomasellii* remained as separate clades (Moretti et al. 1993). In addition, there was lack of support obtained for some clades recovered by the phylogenetic analysis with cpDNA RFLPs. Discrepancies found between morphological analysis and proposed molecular phylogeny were interpreted by Moretti et al. (1993) as a fast series of vicariance effects caused by new orogenies thus not allowing sufficient time for the accumulation of synapomorphies. Their molecular phylogeny was congruent with a model for recent species radiation and vicariance biogeography (Moretti et al. 1993), thereby setting a working hypothesis for their two main clades dividing the *Dioon* taxa into two groups; the Spinulosum clade composed of *D. mejiae*, *D. spinulosum*, and *D. rzedowskii* of the Caribbean (northern Honduras) and Gulf of Mexico seaboard respectively and the Edule clade including the rest of the species. Cladogenesis being the result of vicariance where a putative common ancestor with continuous distribution was interrupted by the Laramide orogeny giving rise to species such as *D. edule* developing to the east and *D. tomaselli* to the west (Moretti et al. 1993).

DNA sequence-based studies have provided insights into phylogenetic relationships in a variety of cycads (e.g. De Luca et al. 1995; González and Vovides 2002; Treutlein and Wink 2002; Chaw et al. 2005). This study was conducted to explore the phylogenetic relationships within *Dioon* using nucleotide sequence data from the internal transcribed spacer (ITS) of the nuclear ribosomal DNA (nrDNA) and the *trnL–F* of cpDNA. Alignment of sequences of different length is a concern in phylogenetic analyses because gaps generated after alignment can be interpreted as artifacts or as phylogenetically informative characters. There are many studies where gaps are ignored as ambiguous characters or are simply deleted from the data sets arguing that they are uninformative (e.g. Olsen 1988; Olsen and Woese 1993). However, in other studies gaps are considered as putative insertion or deletion events (indels) and are incorporated in the analyses as a class of phylogenetic characters using different coding schemes. Some authors include multibase gaps as binary characters or as fifth state (e.g. Milinkovitch et al. 1994; Kropp et al. 1997; Coetzee et al. 2003). Others explore levels of variation among indels using a coding scheme to represent such variation (e.g. Hibbett et al. 1995; González 1996; Lutzoni et al. 2000; Simmons and Ochoterena 2000; Simmons et al. 2001; Grubisha et al. 2002; Aagesen et al. 2005; González et al. 2006; Müller 2006). Unfortunately, there is no consensus agreement as to how indels should be coded for phylogenetic analyses in spite of the big impact they have on the outcome of the analysis (e.g. Simons and Mayden 1997; González et al. 2006). In this study we also aim to test the potential utility of gaps, as characters, by assessing the effects of different coding schemes on clade support. In addition, we will explore the congruence between trees obtained in this study with previous phylogenetic hypothesis.

Materials and Methods

***Taxon Sampling and Molecular Techniques*—**Thirty-three exemplars representing all 13 recognized species of *Dioon* (twelve from Mexico and one from Honduras) and one exemplar of *Cycas rumphii* were used in this study. All of the exemplars are held in the Living National Cycad Collection of the Clavijero Botanic Garden at Xalapa (Appendix 1). The se-

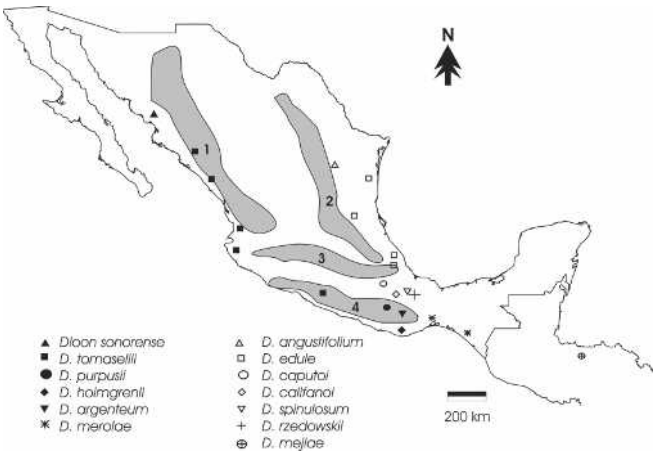

FIG. 1. Map of Mexico showing geographical distribution of *Dioon* species. Shaded areas: 1, Sierra Madre Occidental; 2, Sierra Madre Oriental; 3, Mexican Transverse volcanic range; 4, Sierra Madre del Sur.

lection of *Cycas rumphii* Miq. as an outgroup was based on previous phylogenetic analyses within Cycadales where *Cycas* was always placed as the basal group to the remaining genera. In contrast, the sister group of *Dioon* has not been clearly defined. *Zamia*, *Cycas*, and *Stangeria* have all been proposed as putative sister groups of *Dioon* (Caputo et al. 1991; Moretti et al. 1993; De Luca et al. 1995; González and Vovides 2002; Rai et al. 2003; Bogler and Francisco-Ortega 2004). We did not perform any analyses to find the sister group of *Dioon* because the alignment generated with sequences from the ITS region was ambiguous among all genera.

DNA extraction, amplification and sequencing were performed as described in González and Vovides (2002). Four new primers were designed to amplify and/or sequence the ITS spacers and the *trnL–F* spacer of cpDNA for all the exemplars of *Dioon*. For the ITS spacer, a larger product including the 3′ part of the 18S rDNA was amplified for most species using primers NS7 and ITS4 (White et al. 1990). Sequencing of the whole product was performed using the primers NS7, ITS4 and two internal *Dioon*-specific primers called Dioon-ITS/350 (5′-ATTTCGCTACGTTCTTCATCG-3′) and Dioon-ITS/180 (5′-GAATGCCGTGCGTATGCAA-3′). For the *trnL–F* spacer, primer f (Taberlet et al. 1991) and a *Dioon*-specific primer called Dioon-c 5′-CGGAATTGGTAGACGCTACG-3′ were used for amplification. Sequencing was performed with primers f and the Dioon-specific primers Dioon-c and Dioon-*trnL–F*/455 5′-TAGAGTCCAGTTCTACATGTC-3′. Difficulties were encountered in obtaining sequence data for the *trnL–F* regions in *Dioon spinulosum* A, *D. tomasellii* B, and *D. angustifolium* C. Consequently, the combined data matrix (ITS + *trnL–F*) including outgroup, consisted only of thirty-one exemplars (Appendix 1).

***Alignment of DNA Sequence Data*—**Sequences were aligned using the Clustal V program (Higgins et al. 1992) within the Megalign software package (Lasergene DNASTAR Inc). The alignment of the combined data set consisted of 31 OTUs. In total 2196 nucleotide positions from the ITS and *trnL–F* regions were used. A survey of primary homology assessment was performed on eight alignments with combinations of penalties of 10, 20, and 50 for introducing gaps and penalties of 5, 10, 20, and 50 for gap extension. Six gaps of one nucleotide and 16 of two or more contiguous nucleotides were required to align the studied taxa. Eight analyses were performed in total, utilizing single and combined data sets. Four analyses corresponded to the ITS sequence data alone and four to the combined data set (ITS + *trnL–F*). The *trnL–F* data matrix was not analyzed individually because there were only 13 informative characters. All matrices were deposited in TreeBASE (study number S2004).

To ascertain if gaps are useful phylogenetic characters in our data set, phylogenetic analyses were performed utilizing several different coding schemes: by excluding all regions with indels from the analysis (Coding A), by coding them as ambiguous characters using question marks (Coding B), as multistate (Coding C), and as binary characters (Coding D). Coding scheme C was performed with all indels of one nucleotide coded as a fifth character state, and recoding indels of two or more contiguous nucleotides (Fig. 2). First, we identified indels composed of two or more contiguous nucleotides perfectly delimited with conserved nucleotides. Second we removed all gaps. The nucleotides remained were recoded with a multistate code to describe their variation (e.g. Hibbett et al. 1995;

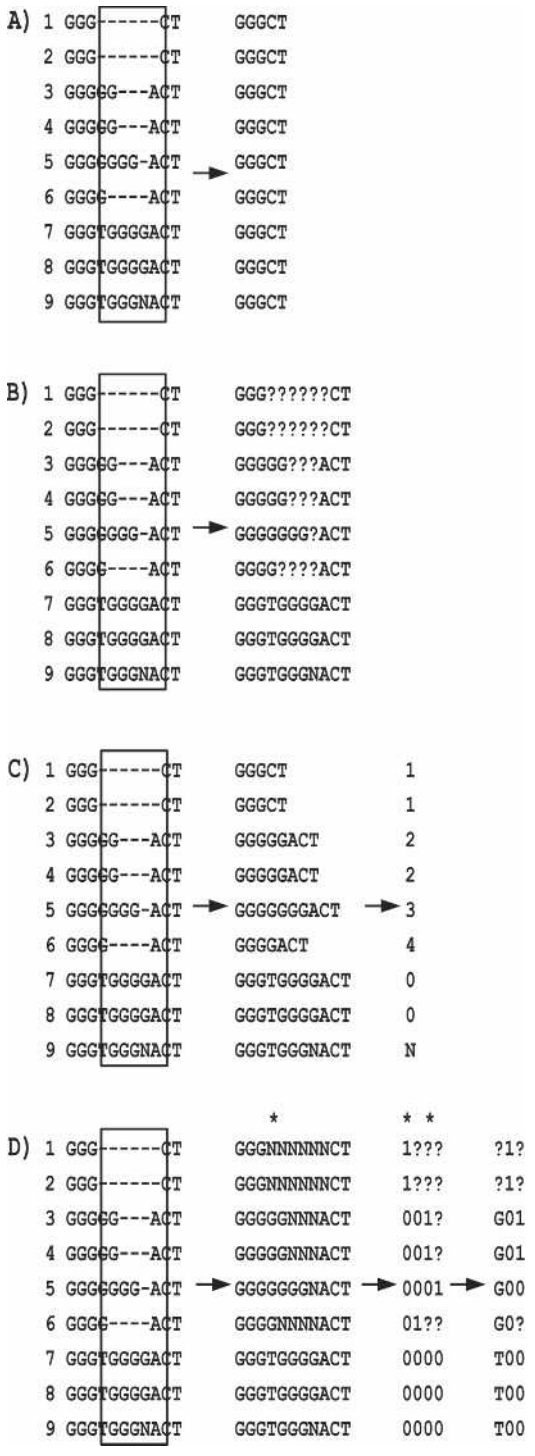

FIG. 2. Coding schemes used for indels of two or more contiguous nucleotides. A) excluded; B) included as ambiguous characters; C) included as multistate and D) coded as binary characters (asterisks indicate informative characters).

Lutzoni et al. 2000; González et al. 2006; Fig. 3). In this analyses, a dash symbol (-) was added into the Nexus file in PAUP*, ver. 4.0b8 (Swofford 2001) in order to incorporate indels of one nucleotide as a different character state in the analyses. Nine character states were needed to score one large indel from the *trnL–F* gene that corresponded to a poly-A segment of 22 positions. Indels that presented one or more unidentified nucleotide (N) in the sequence (regardless of sequence length) were maintained as N for the multistate code (Fig. 2). In addition, we incorporate indels as binary characters (Coding D) into phylogenetic analyses according to the simple indel coding procedure implemented in programs such as

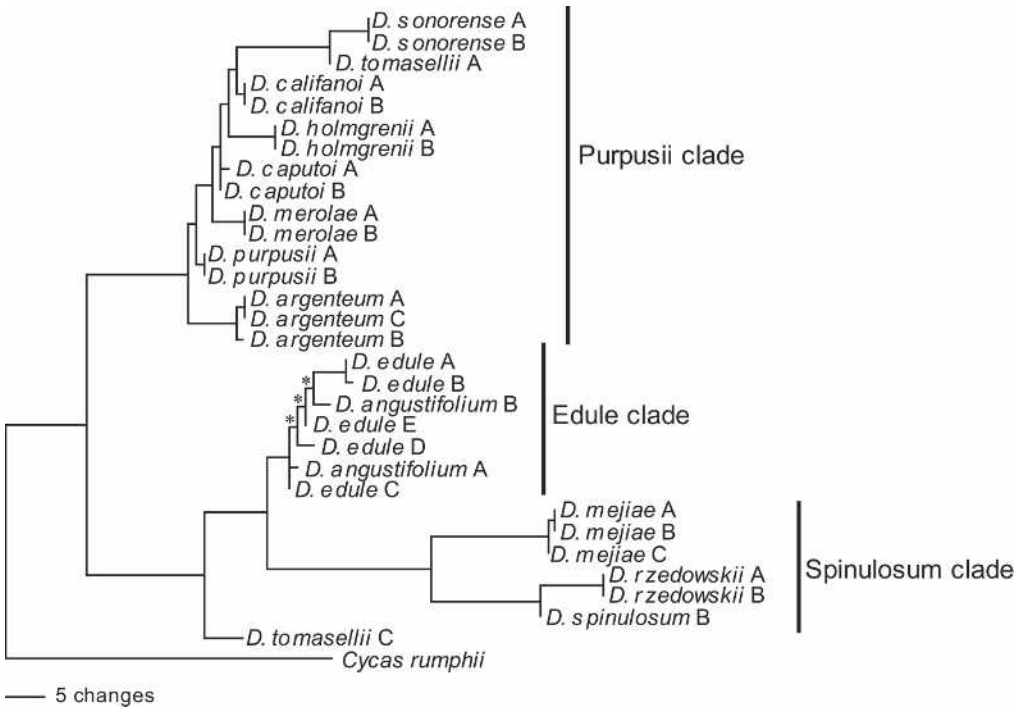

FIG. 3. One of 28 equally parsimonious trees (Length = 212, CI = 0.7453, RI = 0.9179), resulting from combined analyses of ITS and *trnL–F* gene and including indels recoded as multistate (Coding C). Nodes that collapsed in the strict consensus tree are marked with an asterisk above the branch.

GapCoder or Gaprecoder (Simmons and Ochoterena 2000; Young and Healy 2003).

Phylogenetic Analyses—All analyses were performed using maximum parsimony (PAUP*, ver. 4.0b4 Swofford 2001). We executed branch and bound searches with furthest addition of sequences. Branch support was assessed by the Bremer support index and bootstrap analyses based on 1,000 replicates and 10 heuristic searches per replicate with MAXTREES set to 10,000. We used the following arbitrary categories of bootstrap support: < 50% unsupported, 50–70% weakly supported, 71–85% moderately supported, 86–100% highly supported.

RESULTS

Sequence Analyses—Sequences of the ITS for all 33 exemplars of *Dioon* were easily aligned with minor visual adjustments. All gaps generated were clearly located. However, the sequences of *Cycas rumphii* or any other genus of cycads tested as putative outgroup were very different in the ITS region (González and Vovides 2002). In contrast, sequences of the *trnL–F* for the 31 exemplars including outgroup were almost alike. Among the eight alignments generated, we chose the alignment with a gap penalty of 40 and a gap extension of 15 because it maintained the 5.8S region of the rDNA well aligned in all taxa. The combined alignment generated six gaps of one nucleotide and 16 of two or more contiguous nucleotides. The ITS had four gaps of one nucleotide each and 11 gaps of two or more nucleotides. The *trnL–F* region had two of one nucleotide each and five of two or more. From the 22 gaps generated after alignment only five from the ITS and one from the *trnL–F* were informative.

After alignment, the combined ITS + *trnL–F* data set had 2196 characters. From these, 275 nucleotides (13%) were removed from the data matrix for all taxa because some had incomplete sequences from the beginning or end of one or both genes. From the remaining 1921 nucleotides, 1813 (94%) were uninformative. The removal of indels (Coding A) produced 103 (5%) informative characters. Coding indels as ambiguous (Coding B) increased characters to 108 (6%). Coding indels as multistate characters (Coding C) produced 109 (6%) informative characters while coding them as binary characters (Coding D) produced 125 informative characters (7%).

Nucleotide composition for the combined data matrix varied little among taxa (A = 0.226, C = 0.256, G = 0.271, T = 0.247), with an average A + T composition of 47%. However, the ITS alone had an average A + T composition of 34%, while the *trnL–F* gene had an average of 62%. The nucleotide composition for the combined data set was homogeneous based on a χ^2 test of homogeneity of base frequencies across taxa. The value obtained for the combined data matrix and all characters included was 12.816 (df = 90, p = 1.000).

Phylogenetic Analysis—A comparison of general features of the most parsimonious trees (MPTs) found in the eight analyses performed is summarized in Table 1. Two levels of resolution for phylogenetic relationships were reconstructed

TABLE 1. Results of six analyses of the internal transcribed spacer (ITS), and the *trnL–F* gene under different combinations of data sets and with indels excluded (A), included as ambiguous characters (B), coded as multistate (C) and coded as binary characters (D). We report the number of phylogenetically informative sites (PIS); most parsimonious trees (MPT) found with branch and bound searches; length of the MPT, consistency index (CI) and retention index (RI).

Gene partition	# of PIS	# of MPT	Length of MPT	CI	RI
ITS					
A	94	4932	166	0.7349	0.9330
B	98	4932	172	0.7442	0.9365
C	99	36	185	0.7568	0.9379
D	108	108	187	0.7380	0.9328
ITS + *trnL–F*					
A	103	4212	178	0.7368	0.9202
B	108	4212	185	0.7459	0.9241
C	109	28	212	0.7453	0.9179
D	125	1782	217	0.7143	0.9104

in our analyses with the combined data set and with the ITS data set alone. In all analyses performed, with different coding schemes and with single and combined data sets the consensus trees were alike and only differ in resolution (Fig. 4).

Analyses of the ITS using coding schemes A and B, had low resolution due to 4932 competing topologies obtained in both analyses. In contrast, the analysis of coding scheme C recovered only 36 trees of 185 steps with good resolution in terminal and basal branches. Analyses with coding scheme D recovered 108 trees of 187 steps. The consensus trees of these two analyses were indels were coded as multistate or as binary characters, were the same (Fig. 4A2–3).

Analyses of the two data partitions combined better re-

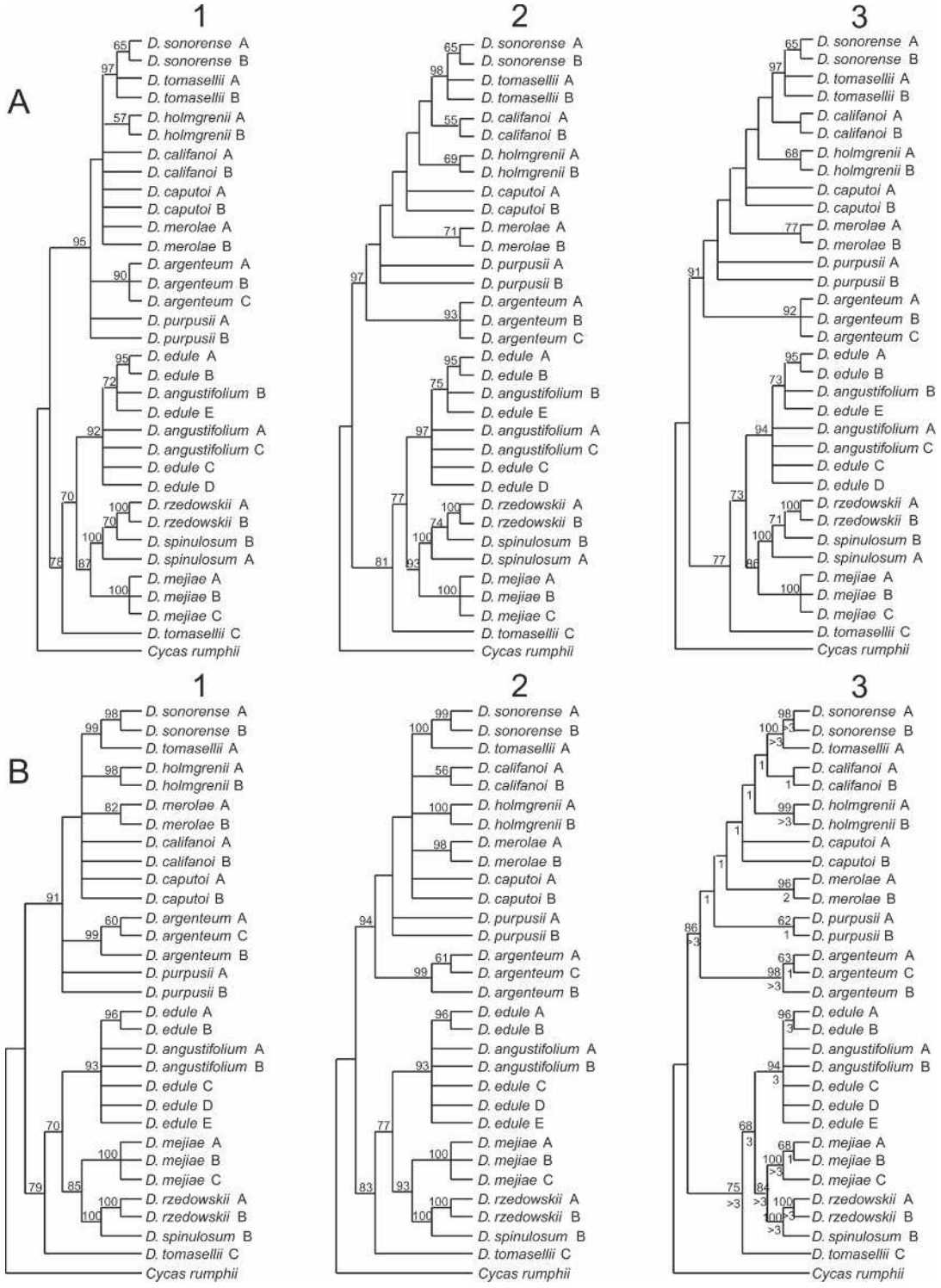

FIG. 4. Strict consensus trees resulting from cladistic analyses of the ITS sequences alone and combined with the *trnL–F*. A) Analysis of the ITS with (1) the indels excluded or coded as ambiguous data, (2) included recoded as binary characters and (3) included as multistate. B) Analysis of combined data set with (1) the indels excluded or coded as ambiguous data, (2) included recoded as binary characters and (3) included as multistate. Bootstrap frequency values above 50% and Bremer support values (in tree B3) are indicated.

solved topologies than with the ITS alone. Although the consensus of 4212 trees found when indels were excluded or coded as ambiguous data (Coding A and B) had low resolution, exemplars of *D. merolae* were recovered in one clade and the clade formed by three exemplars of *D. argenteum* was better resolved (Fig. 4B1). The consensus of 1782 trees obtained with indels coded as binary characters (Coding D) was better resolved in basal branches (Fig. 4B2). Two additional clades were recovered one conformed by exemplars of *D. argenteum* and other by exemplars of *D. califanoi*. In contrast, the consensus of 28 trees obtained when indels were recoded as multistate (Coding C) was well resolved and supported in most branches (Fig. 4B3). We adopted this topology as the presumed species tree. This topology included two major clades that received from moderate to high support in all analyses. One of this clade contained exemplar of eight species of *Dioon* (*D. sonorense*, *D. califanoi*, *D. holmgrenii, D. caputoi*, *D. merolae, D. purpusii*, *D. argenteum,* and *D. tomasellii* exemplar A) while the other contained exemplars from six species (*D. edule, D. angustifolium, D. mejiae, D. rzedowskii, D. spinulosum,* and *D. tomasellii* exemplar C). Three species (*D. edule, D. angustifolium,* and *D. tomasellii*) did not resolve as monophyletic taxa.

Despite only thirteen informative characters were found in the *trnL–F* gene, three additional clades were recovered in the combined data set analysis within exemplars of *D. mejiae*, *D. argenteum,* and *D. purpusii*. Clades obtained in combined analyses including indels coded as multistate were, in general, better supported than with the ITS region alone (Table 2). However, levels of support when indels are recoded as multistate for some clades were slightly less effective than coding indels as binary characters in the combined analysis. This is particularly evident in the larger clades (Table 2). Six clades with only exemplars of the same species (*D. merolae, D. argenteum, D. holmgrenii, D. sonorense, D. rzedowskii,* and *D. mejiae*) were always recovered with moderate to high bootstrap support indicating they are good phylogenetic species (Table 2). In contrast, exemplars of *D. tomasellii* are mixed and dispersed across the entire cladogram in each consensus tree obtained with the ITS alone and combined with the *trnL–F*.

Discussion

All phylogenetic analyses concord with previously recognized relationships within *Dioon*. Our analyses also supported two major clades and the close relationship among *D. mejiae*, *D. spinulosum,* and *D. rzedowskii* found by Moretti et al. (1993) and Bogler and Francisco-Ortega (2004). However, *D. edule* was placed in a separate clade more related to these three species than with the rest of the species as found by Moretti et al. (1993). The inclusion in our study of all known species of *Dioon* determined to date and multiple exemplars of each species resolved better the relationships within the genus. In this study *D. holmgrenii* and *D. merolae* did not result as sister taxa, as found by Moretti et al. (1993) with cp DNA RFLPs, but agreed with their morphological analysis.

Cladistic analyses of individual and combined data sets from the ITS and the *trnL–F* gene confirmed that multiple samples for most species within *Dioon* formed monophyletic groups. However, multiple accessions of *D. tomasellii*, *D. edule,* and *D. angustifolium* were not recovered as monophyletic entities. Both *D. edule* and *D. tomasellii* should be considered as species complexes, since both taxa have a very large distribution range, with some populations of each species showing morphological differences. Therefore, more exemplars of these species and perhaps another region of the genome should be examined to better circumscribe these taxa and define more precisely their phylogenetic relationships.

All phylogenetic analyses with one or two gene partitions and indels excluded, included as ambiguous characters or recoded as multistate recovered the same topology but with different levels of support. This suggests that there are phylogenetically informative characters in both genes, and clades are stable to the different strategies used for including the

Table 2. Effect of indels in clade support. Bootstrap percentages for the subset of the clades supported by bootstrap values above 50% in at least three analyses for the ITS alone and in combination with the *trnL–F*, when positions with indels are excluded (A), included as ambiguous characters (B), recoded as multistate (C) or recoded as binary characters (D). Empty boxes indicate bootstrap values lower than 50%. * indicates exemplars not included in combined analyses due incomplete sequence for the *trnL–F*

	ITS				ITS + *trnL–F*			
CLADE	A	B	C	D	A	B	C	D
D. purpusii A,B; *D. argenteum* A,B,C; *D. merolae* A,B; *D. caputoi* A,B; *D. califanoi* A,B; *D. holmgrenii* A,B; *D. tomasellii* A,B*; *D. sonorense* A,B	94	95	91	97	91	94	86	94
D. tomasellii C; *D. mejiae* A,B,C; *D. spinulosum* A*,B; *D. rzedowskii* A,B; *D. edule* A,B,C,D,E; *D. angustifolium* A,B,C*	76	78	77	81	79	81	75	83
D. mejiae A,B,C; *D. spinulosum* A*,B; *D. rzedowskii* A,B; *D. edule* A,B,C,D,E; *D. angustifolium* A,B,C*	69	70	73	77	70	74	68	77
D. edule A,B,C,D,E; *D. angustifolium* A,B,C*	92	92	94	97	93	93	94	93
D. edule A,B,E; *D. angustifolium* B	73	72	73	75				
D. edule A,B	95	95	95	95	96	96	96	96
D. mejiae A,B,C; *D. spinulosum* A*,B; *D. rzedowskii* A,B	85	87	86	93	85	90	84	93
D. mejiae A,B,C	100	100	100	100	100	100	100	100
D. spinulosum A*,B; *D. rzedowskii* A,B	100	100	100	100				
D. spinulosum B; *D. rzedowskii* A,B	71	70	71	74	100	100	100	100
D. rzedowskii A,B	100	100	100	100	100	100	100	100
D. tomasellii A,B*; *D. sonorense* A,B	96	97	97	98	99	99	100	100
D. sonorense A,B	64	65	65	65	98	98	98	99
D. holmgrenii A,B	59	57	68	69	98	99	99	100
D. argenteum A,B,C	90	90	92	93	99	99	98	99
D. merolae A,B			77	71	82	93	96	98
D. argenteum A,C					60	61	63	61

indels in the analyses. The best resolved cladogram was obtained with the combined gene partitions and indels recoded as multistate. This may indicate that this code scheme best recovered the phylogenetic information of these characters. Resolution and most bootstrap support are the highest when this scoring scheme was used. However, coding indels as binary characters gave some bootstrap values slightly higher than coding indels as multistate for some clades. One possible explanation is that the program used for coding indels as binary characters is also taking into account the substitutions present in the region of the indel in addition to coding the presence/absence of the gap. While coding as multistate, the region with the indel is coded only once (Fig. 2).

Two major clades with moderate to high support were always recovered from our analyses. The Purpusii clade and another combined Edule - Spinulosum clades (Fig. 3). These clades partially agree with those previously suggested by morphology and cpDNA RFLP data (Moretti et al. 1993). The cladogram found in our study placed *D. edule* in a separate clade with the related *D. angustifolium*, which we refer to as the Edule clade (Fig. 3). This clade is congruent with the Gulf of Mexico seaboard distribution (Fig. 1) and is more related to the Spinulosum clade than to the Edule clade of Moretti et al. (1993). In that study *D. edule* and *D. tomasellii* resulted sister taxa; in our analyses exemplars of *D. tomasellii* were dispersed amongst the two major clades. In one of these clades, which we refer to as the Purpusii clade exemplars A and B of *D. tomasellii,* resulted the sister group of *D. sonorense* with a high bootstrap support (Figs. 3, 4). The close relationship between these two species has been previously recognized. Until few years ago *D. sonorense* was known as *D. tomasellii* var. *sonorense* De Luca et al. (1984), but important morphological characters enabled Chemnick et al. (1997) to assign this taxon species status. *Dioon tomasellii* exemplar C, resulted sister group of a large clade formed by the Edule and Spinulosum clades with 13–15 exemplars of *Dioon* amongst five species (Figs. 3, 4). *Dioon tomasellii* exemplar C may be a different entity and it is evident from our results that considerable work is still needed within *D. tomasellii* at the population level to find out how many species are there and to fully solve the phylogenetic relationships between them. The exemplars of *D. califanoi*, *D. caputoi*, and *D. purpusii*, were not resolved as a single clade or were not supported by bootstrap analyses.

Both exemplars of *D. angustifolium* used in this study were included in a highly supported clade mixed in with four exemplars of *D. edule*. *Dioon angustifolium* was previously recognized as *D. edule* var. *angustifolium* (De Luca et al. 1982). Moretti et al. (1993) used two exemplars of *D. edule* (*D. edule* Lindley var. *angustifolium* Miq., and *D. edule* Lindley var. *edule*) for their morphological and molecular analyses. In those analyses both exemplars formed a highly supported clade. González-Astorga et al. (2003a,b; 2005) recognized the species status of *D. angustifolium* based on variation of leaflet morphology, and genetic variation of fourteen allozymic loci in populations of *D. edule* from its total distribution range and by sampling all known populations. However, data presented here from sequence data from the ITS and the *trnL–F* places several exemplars of *D. edule* and *D. angustifolium* intermixed in the same place. These results may indicate either that allozymic variation is overestimated or our sequence data is not enough to solve this relationship.

The phylogeny depicted in Fig. 3 reveals that the first dicotomy separates the ingroup in two major clades. The Purpusii clade with species from xeric environments, and the combined Edule - Spinulosum clades that included *D. tomaselli* exemplar C, *D. edule*, *D. angustifolium*, *D. mejiae*, *D. rzedowskii,* and *D. spinulosum* with species from xeric and humid environments. The Edule clade consists of *D. edule* and *D. angustifolium*. The Spinulosum clade groups, *D. spinulosum, D. rzedowskii,* and *D. mejiae*, which are species with a more humid habitat. *Dioon spinulosum* and *D. rzedowskii* are distributed along the more humid northern mountains of Oaxaca facing the Gulf of Mexico and *D. mejiae* in northern Honduras along the Caribbean seaboard. They have wide leaflets and relatively open stomata, whereas the rest of the genus has narrow leaflets and partially occluded stomata and are well adapted to xeric conditions (Moretti et al. 1993). Moretti et al. (1993) suggested that the lack of xeric adaptations in *Dioon spinulosum, D. rzedowskii,* and *D. mejiae* may have developed in response to more benign climatic conditions, and that their common ancestor originated in a relatively humid habitat. Moreover, the habitat of *D. rzedowskii* and *D. mejiae* has not shown any great change in climate since the proposed origin of the genus in the late Cretaceous (Moretti et al. 1993). Even after the Pleistocene glaciations, which caused a severe decrease of tropical rain forest habitats, tropical species survived in refugium areas in southeastern Mexico (Toledo 1982). Some of these refugia are close to the present range of *D. spinulosum* and *D. rzedowskii* (Moretti et al. 1993). Our analyses also revealed that clades with high bootstrap and Bremer support indices correlate with geographical origin (Fig. 1). The Spinulosum clade includes species distributed southeast of the Transverse Volcanic Mountain Range. The Edule clade comprises species located along the Eastern Mountain Ranges of the Gulf of Mexico seaboard north of the Transverse Volcanic Mountain Range (Sierra Madre Oriental). The Purpusii clade has species found along the Pacific seaboard Mountain Ranges (Sierra Madre Occidental and Sierra Madre del Sur).

In the genus *Dioon* there was no reliable knowledge of a putative sister group, therefore we decided to use one species of *Cycas* to polarize the cladograms with the single criterion that this genus has always been considered basal to all cycads (Johnson 1959; Stevenson 1990). However, it is necessary to do more extensive analyses to better address suitable outgroup taxa for *Dioon* and to improve our knowledge of relationships among all genera within Cycadales.

The results of this study did not fully resolve the relationships for *D. tomasellii*, *D. angustifolium,* and *D. edule*. However, congruence with the hypothesis of recent species radiation and vicariance biogeography for the genus as suggested by Moretti et al. (1993) appears to have been confirmed. Increasing species richness with decreasing latitude appears to be the case for *Dioon* along the Pacific seaboard, with maximum species richness in Oaxaca in southern Mexico (Vovides et al. 2003) as illustrated in the Purpusii clade (Figs. 1, 3). This distribution appears congruent with the amelioration of climates during post Pleistocene times favoring the spread of taxa from floristic refugia (Toledo 1982). This also supports a pattern of diversification similar to *Bursera* spp. (Becerra 2005) that share the same vegetation types and distribution with most *Dioon* species, the floristic refuge hypothesis and cycad distribution in Mexico has been discussed by González and Vovides (2002) and Vovides et al. (2003). The species in

the Edule and Spinulosum clades are congruent with distribution in northeast and southeast Mexico to northern Honduras. Our data so far presented appears to be in agreement with data available for biogeography and distribution of *Dioon* species by Balduzzi et al. (1981), Vovides et al. (1983), and Sabato and De Luca (1985; see Fig. 1), and is in accord with the suggestion of recent species radiation and vicariance biogeography by Moretti et al. (1993). Sabato and De Luca (1985) suggested that the genus *Dioon* could have evolved during late Jurassic to early Cretaceous and, from fossil evidence, during Eocene times; *Dioon* later contracting its range by early Oligocene following cooling trends (Moretti et al. 1993). Further work is being undertaken to elucidate biogeographical relationships of *Dioon* in the light of molecular data as well as morphological.

In conclusion, this study has attempted to resolve phylogenetic relationships in the genus *Dioon* as well as to test the potential utility of gaps as characters. The proposed phylogeny is almost completely resolved and is congruent, with few exceptions, with the one found by Moretti et al. (1993) with cp DNA RFLP. In addition, the topology found is consistent with biogeographic relationships between the species occurring along the Pacific seaboard on the Sierra Madre Occidental and the Gulf of Mexico seaboard on the Sierra Madre Oriental. Although the precise strategy for coding indels for phylogenetic studies still remains to be established our results indicated that including indels as multistate best recovered the phylogenetic information. No doubt the distribution and characteristics of indels in our data matrix allowed us to evaluate them with a multistate code, but this may not be the case for more complex data matrices.

ACKNOWLEDGMENTS. This research was supported by a grant from CONACYT- SEMARNAT-2002-C01- 0183. The authors also thank the Jardín Botánico Francisco Javier Clavijero for access to the National Mexican Cycad collection where living material was obtained for DNA extraction and to Victor Luna for assistance in collecting this material and assistance with Fig. 1.

LITERATURE CITED

Aagesen, L., G. Petersen, and O. Seberg. 2005. Sequence length variation, indel costs, and congruence in sensitivity analysis. *Cladistics* 21: 15–30.

Balduzzi, A., P. De Luca, and S. Sabato. 1981–82. A phytogeographical approach to the New World Cycads. *Delpinoa, n.s.* 23–24: 185–202.

Becerra, J. X. 2005. Timing the origin and expansion of the Mexican tropical dry forest. *Proceedings of the National Academy of Sciences USA* 102: 10919–10923.

Bogler, D. J. and J. Francisco-Ortega. 2004. Molecular systematic studies in cycads: Evidence from *trnL* intron and ITS2 rDNA squences. *Botanical Review* 70: 260–273.

Caputo, P., D. W. Stevenson, and E. T. Wurtzel. 1991. A phylogenetic analysis of American Zamiaceae (Cycadales) using chloroplast DNA restriction fragment length polymorphisms. *Brittonia* 43: 135–145.

Chaw, S.-M., T. W. Walters, C.-C. Chang, S.-H. Hu, and S.-H. Chen. 2005. A phylogeny of cycads (Cycadales) inferred from chloroplast *matK* gene, *trnK* intron, and nuclear rDNA ITS region. *Molecular Phylogenetics and Evolution* 37: 214–234.

Chemnick, J., T. I. Gregory, and S. Salas. 1997. A revision of *Dioon tomasellii* (Zamiaceae) from western Mexico, a range extension of *D. merolae*, and clarification of *D. purpusii. Phytologia* 83: 1–6.

Coetzee, M. P. A., P. Bloomer, G. S. Ridley, and M. J. Wingfield. 2003. Molecular identification and phylogeny of *Armillaria* isolates from South America and Indo-Malaysia. *Mycologia* 95: 285–293.

De Luca, P., S. Sabato, and M. Vazquez. 1982. Distribution and variation of *Dion edule* (Zamiaceae). *Brittonia* 34: 355–362.

De Luca, P., S. Sabato, and M. Vazquez. 1984. *Dioon tomasellii* (Zamiaceae), a new species with two varieties from western Mexico. *Brittonia* 36: 223–227.

De Luca, P., A. Moretti, G. Siniscalco Gigliano, P. Caputo, S. Cozzolino, L. Gaudio, D. W. Stevenson, E. T. Wuertzel, and R. Osborne. 1995. Molecular systematics of cycads. Pp. 131–137 in *Proceedings of the Third International Conference on Cycad Biology 1993*, eds. P. Vorster. Pretoria: Cycad Society of South Africa, Stellenbosch.

González, D. 1996. Codificación de las inserciones-deleciones en el análisis filogenético de secuencias génicas. *Boletín de la Sociedad Botánica de México* 59: 115–129.

González, D. and A. P. Vovides. 2002. Low intralineage divergence in *Ceratozamia* sp. (Zamiaceae) detected with nuclear ribosomal DNA ITS and chloroplast DNA *trnL–F* non-Coding region. *Systematic Botany* 27: 654–661.

González, D., M. A. Cubeta, and R. Vilgalys. 2006. Phylogenetic utility of indels within ribosomal DNA and β-tubulin sequences from fungi in the *Rhizoctonia solani* species complex. *Molecular Phylogenetics and Evolution* 40: 459–470.

González-Astorga, J., A. P. Vovides, and C. Iglesias. 2003a. Morphological and geographical variation of the cycad *Dioon edule* Lindl. (Zamiaceae): ecological and evolutionary implications. *Botanical Journal of the Linnean Society* 141: 465–470.

González-Astorga, J., A. P. Vovides, M. M. Ferrer, and C. Iglesias. 2003b. Population genetics of *Dioon edule* Lindl. (Zamiaceae, Cycadales): biogeographical and evolutionary implications. *Biological Journal of the Linnean Society* 80: 457–467.

González-Astorga, J., A. P. Vovides, A. Cruz-Angón, P. Octavio-Aguilar, and C. Iglesias. 2005. Allozyme variation in three extant populations of the narrowly endemic cycad *Dioon angustifolium* Miq. (Zamiaceae) from north-eastern Mexico. *Annals of Botany* 95: 999–1007.

Grubisha, L. C., J. M. Trappe, R. Molina, and J. W. Spatafora. 2002. Biology of the ectomycorrhizal genus *Rhizopogon*. VI. Re-examination of infrageneric relationships inferred from phylogenetic analyses of ITS sequences. *Mycologia* 94: 607–619.

Hibbett, D. S., Y. Fukumasa-Nakai, A. Tsuneda, and M. J. Donoghue. 1995. Phylogenetic diversity in shiitake inferred from nuclear ribosomal DNA sequences. *Mycologia* 87: 618–638.

Higgins, D. G., A. J. Bleasby, and R. Fuchs. 1992. CLUSTAL V: Improved software for multiple sequence alignment. *CABIOS* 8: 189–191.

Hill, K., D. W. Stevenson, and R. Osborne. 2004. The world list of cycads. Pp 219–235 in *Cycad classification: concepts and recommendations*, eds. T. Walters and R. Osborne. Massachusetts: CABI Publishing.

Johnson, L. A. S. 1959. The families of cycads and the Zamiaceae of Australia. *Proceedings of the Linnean Society of New South Wales* 81: 64–117.

Kropp, B. R., D. R. Hansen, P. G. Wolf, K. M. Flint, and S. V. Thomson. 1997. A study on the phylogeny of Dyer's woad rust fungus and other species of *Puccinia* from crucifers. *Phytopathology* 87: 565–571.

Lutzoni, F., P. Wagner, V. Reeb, and S. Zoller. 2000. Integrating ambiguously aligned regions of DNA sequences in phylogenetic analyses without violating positional homology. *Systematic Biology* 49: 628–651.

Milinkovitch, M. C., A. Meyer, and J. R. Powell. 1994. Phylogeny of all major groups of cetaceans based on DNA sequences from three mitochondrial genes. *Molecular Biology and Evolution* 11: 939–948.

Moretti, A., P. Caputo, S. Cozzolino, P. De Luca, L. Gaudio, G. Siniscalco Gigliano, and D. W. Stevenson. 1993. A Phylogenetic analysis of *Dioon* (Zamiaceae). *American Journal of Botany* 80: 204–214.

Müller, K. 2006. Incorporating information from length-mutational events into phylogenetic analysis. *Molecular Phylogenetics and Evolution* 38: 667–676.

Norstog, K. J. and T. J. Nicholls. 1997. *The biology of the cycads*. Ithaca: Cornell University Press.

Olsen, G. J. 1988. Phylogenetic analysis using ribosomal RNA. *Methods in Enzymology* 164: 793–838.

Olsen, G. J. and C. R. Woese. 1993. Ribosomal RNA: a key to phylogeny. *The FASEB Journal* 7: 113–123.

Rai, H. S., H. E. O'Brien, P. A. Reeves, R. G. Olmstead, and S. W. Graham. 2003. Inference of higher-order relationships in the cycads from a large chloroplast data set. *Molecular Phylogenetics and Evolution* 29: 350–359.

Sabato, S. and P. De Luca. 1985. Evolutionary trends in *Dioon* (Zamiaceae). *American Journal of Botany* 72: 1353–1363.

Simmons, M. P. and H. Ochoterena. 2000. Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* 49: 369–381.

Simmons, M. P., H. Ochoterena, and T. Carr. 2001. Incorporation, relative homoplasy, and effect of gap characters in sequence-based phylogenetic analyses. *Systematic Biology* 50: 454–462.

Simons, A. M. and R. Mayden. 1997. Phylogenetic Relationships of the Creek Chubs and the Spine-Fins: an Enigmatic Group of North American Cyprinid Fishes (Actinopterygii: Cyprinidae). *Cladistics* 13: 187–205.

Stevenson, D. 1990. Morphology and systematics of cycadales. *Memoirs of the New York Botanical Garden* 57: 8–55.

Swofford, D. L. 2001. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods), version 4.0b8. Sunderland: Sinauer Associates.

Taberlet, P., L. Gielly, G. Pautou, and J. Bouvet. 1991. Universal primers for amplification of three non-Coding regions of chloroplast DNA. *Plant Molecular Biology* 17: 1105–1109.

Toledo, V. M. 1982. Pleistocene changes of vegetation in tropical Mexico. Pp. 93–111 in *Biological diversification in the tropics. Proceedings of the 5th International Symposium of the Association for Tropical Biology 1979*, eds. G. T. Prance. New York: Columbia University Press.

Treutlein, J. and M. Wink. 2002. Molecular phylogeny of cycads inferred from *rbc*L sequences. *Die Naturwissenschaften* 89: 221–225.

Vovides, A. P., J. D. Rees, and M. Vázquez Torres. 1983. Zamiaceae. *Flora de Veracruz* fasc. 26. eds. A. Gómez-Pompa, and V. Sosa. Xalapa: INIREB. 31pp.

Vovides, A. P., M. A. Pérez-Farrera, J. González-Astorga, D. González, T. Gregory, J. Chemnick, C. Iglesias, P. Octavio-Aguilar, S. Avendaño, C. Bárcenas, and S. Salas-Morales. 2003. An outline of our current knowledge on Mexican cycads (Zamiaceae, Cycadales). *Current Topics in Plant Biology* 4: 159–174.

White, T. J., T. Bruns, S. Lee, and J. Taylor. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Pp. 315–322 in *PCR Protocols a guide to methods and applications*, eds. M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White. San Diego: Academic Press.

Young, N. D. and J. Healy. 2003. GapCoder automates the use of indel characters in phylogenetic analysis. *BMC Bioinformatics* 4: 1–6.

Appendix 1. Exemplars used for cladistic analysis of ribosomal DNA internal transcribed spacer (ITS 1 and 2), and *trnL–F* cpDNA non-coding region. Multiple specimens from the same species are identified by capital letters. Voucher details are listed in the following sequences: collection locality, collector name and number, and herbaria at which the voucher is deposited (parentheses). The first accession number is for ITS, the second for the *trnL–F*. Abbreviations correspond to: VL (Víctor E. Luna Monterrojo), MAC (Michael Calonje), JLH (Jody Haynes), XAL (Herbario del Instituto de Ecología, A.C., Xalapa), TEFH (Herbario Nacional de la Universidad Nacional Autónoma de Honduras), FTG (Fairchild Tropical Garden). Asterisks indicate incomplete sequence data.

Cycas rumphii, Asia, 1990-032 D, *VL-1573* (XAL), AF407283, AF407314. *Dioon califanoi* A, Teotitlan, Oax, 1992-179.05, *VL-1574* (XAL), DQ926722, DQ926755. *D. califanoi* B, Teotitlan Oax, 1992-179.03, *VL-1575* (XAL), DQ926723, DQ926756. *D. caputoi* A, Caltepec, Pue, 1986-175 A, *VL-1576* (XAL), DQ926724, DQ926757. *D. caputoi* B, Caltepec, Pue, 1986-175 B, *VL-1577* (XAL), DQ926725, Q926758. *D. edule* A, Monte Oscuro, Ver, 1991-224.07, *VL-1578* (XAL), DQ926726, DQ926759. *D. edule* B, Monte Oscuro, Ver, 2002-026 B, *VL-1579* (XAL), DQ926727, DQ926760. *D. edule* C, Mpio. Casas, Tamps, 2001-151 A, *VL-1580* (XAL), DQ926728, DQ926761. *D. edule* D, Tamasopo, SLP, 2001-067 A, *VL-1581* (XAL), DQ926729, DQ926762. *D. edule* E, Actopan, Veracruz, 1987-152-A, *VL-1582* (XAL), DQ926730, DQ926763. *D. holmgrenii* A, S. Gabriel Mixtepec, Oax, 1981-698-.02, *VL-1583* (XAL), DQ926732, DQ926765. *D. holmgrenii* B, S. Gabriel Mixtepec, Oax, 1985-015-01, *VL-1584* (XAL), DQ926733, DQ926766. *D. mejiae* A, Gualaco Honduras, 20030872, *MAC-07-011* (FTG), DQ926734, DQ926767. *D. mejiae* B, Esquipulas, Honduras, 20030882, *JLH-03-032A* (TEFH), DQ926735, DQ926768. *D. mejiae* C, Olanchito, Honduras, 76967, *76-967* (FTG), DQ926736, DQ926769. *D. merolae* A, Jiquipilas, Chis, 1992-180.01, *VL-1585* (XAL), DQ926737, DQ926770. *D. merolae* B, Jiquipilas, Chis, 1992-180.04, *VL-1586* (XAL), DQ926738, DQ926771. *D. purpusii* A, Cuicatlán, Oax, 1979-128.01, *VL-1587* (XAL), DQ926739, DQ926772. *D. purpusii* B, Cuicatlán, Oax, 1979-128.03, *VL-1588* (XAL), DQ926740, DQ926773. *D. rzedowskii* A, S. Bartolomé L., Oax, 1993-129.02, *VL-1589* (XAL), DQ926741, DQ926774. *D. rzedowskii* B, S. Bartolomé L., Oax, 1993-129.03, *VL-1590* (XAL), DQ926742, DQ926775. *D. sonorense* A, Alamos, Son, 2000-066.01, *VL-1591* (XAL), DQ926743, DQ926776. *D. sonorense* B, Alamos, Son, 2000-068.01, *VL-1592* (XAL), DQ926744, DQ926777. *D. spinulosum* A, Acatlán, Oax, 1976-016.01, *VL-1593* (XAL), DQ926745, *. *D. spinulosum* B, S.Juan Bautista, La Mina Oax, 1980-208.02, *VL-1594* (XAL), DQ926746, DQ926778. *D. tomaselli,* A, Cabo Corrientes, Jal, 2001-059.02, *VL-1595* (XAL), DQ926747, DQ926779. *D. tomasellii* B, Cabo Corrientes, Jal, 2001-059.04, *VL-1596* (XAL), DQ926748, *. *D. tomasellii* C, Zirandaro, Guerrero, 2001-115 A, *VL-1597* (XAL), DQ926749, DQ926780. *D. argenteum* A, S. Pedro Yolox, Oax, 2000-007.01, *VL-1598* (XAL), DQ926719, DQ926752. *D. argenteum* B, S. Pedro Yolox, Oax, 2000-007.06, *VL-1599* (XAL), DQ926720, DQ926753. *D. argenteum* C, S. Pedro Yolox, Oax, 2000-007.07, *VL-1600* (XAL), DQ926721, DQ926754. *D. angustifolium* A, Rancho Muralla, Linares NL, 2001-147.01, *VL-1601* (XAL), Q926716, DQ926750. *D. angustifolium* B, Rancho Muralla, Linares NL, 2001-149.01, *VL-1602* (XAL), DQ926717, DQ926751. *D. angustifolium* C, Mpio. San Carlos, Tamps, 1999-056 A, *VL-1603* (XAL), DQ926718, *.