WILEY-BLACKWELL

Cladistics 27 (2011) 150–164

Cladistics

10.1111/j.1096-0031.2010.00321.x

A character-based approach in the Mexican cycads supports diverse multigene combinations for DNA barcoding

Fernando Nicolalde-Morejón[a,b,*], Francisco Vergara-Silva[c], Jorge González-Astorga[a], Dennis W. Stevenson[d], Andrew P. Vovides[e] and Victoria Sosa[f]

[a]*Laboratorio de Genética de Poblaciones, Biología Evolutiva. Instituto de Ecología, A. C., km 2.5 Antigua Carretera a Coatepec No. 351, Xalapa 91070, Veracruz, México;* [b]*Instituto de Investigaciones Biológicas, Universidad Veracruzana, Av. Luis Castelazo Ayala s/n, Col. Industrial Ánimas, Xalapa 91190, Veracruz, México;* [c]*Laboratorio de Sistemática Molecular (Jardín Botánico), Instituto de Biología, Universidad Nacional Autónoma de México, 3er Circuito Exterior, Ciudad Universitaria, Coyoacán 04510, México, D. F. México;* [d]*The New York Botanical Garden, Bronx, NY 10458-5120, USA;* [e]*Laboratorio de Biología Evolutiva de Cycadales, Biología Evolutiva. Instituto de Ecología, A. C., km 2.5 Antigua Carretera a Coatepec No. 351, Xalapa 91070, Veracruz, México;* [f]*Laboratorio Molecular, Biología Evolutiva. Instituto de Ecología, A. C., km 2.5 Antigua Carretera a Coatepec No. 351, Xalapa 91070, Veracruz, México*

Accepted 22 March 2010

Abstract

A DNA barcoding study was conducted to determine the optimal combination of loci needed for successful species-level molecular identification in three extant cycad genera—*Ceratozamia*, *Dioon*, and *Zamia*—that occur in Mexico. Based on conclusions of a previous multigene study in representative species of all genera in the Cycadales, we tested the DNA barcoding performance of seven chloroplast coding (*matK*, *rpoB*, *rpoC1*, and *rbcL*) and non-coding (*atpF/H*, *psbK/I*, and *trnH-psbA*) regions, plus sequences of the nuclear internal transcribed spacer. We analysed data under the assumptions of the "character attributes organization system" (CAOS), a character-based approach in which species are identified through the presence of 'DNA diagnostics'. In *Ceratozamia*, four chloroplast regions and one nuclear region were needed to achieve > 70% unique species identification. In contrast, the two-gene combination *atpF/H* + *psbK/I* and the four-gene combination *atpF/H* + *psbK/I* + *rpoC1* + *ITS2* were needed to reach 79% and 75% unique species identification in *Dioon* and *Zamia*, respectively. The combinations *atpF/H* + *psbK/I* and *atpF/H* + *psbK/I* + *rpoC1* + *ITS2* include loci previously considered by the international DNA barcoding community. However, none of the three combinations of potential DNA barcoding loci found to be optimal with a character-based approach in the Mexican cycads coincides with the 'core barcode' of chloroplast markers (*matK* + *rbcL*) recently proposed for universal use in the plant kingdom.

Holding approximately one-sixth of the total species number for the group in the Neotropics, Mexico is one of the three main centres of biological diversity in Cycadales, one of four groups of extant gymnosperms (Norstog and Nicholls, 1997; Vovides et al., 2007). As in other important seed plant groups, e.g. angiosperms (Mathews, 2009; and references therein), our understanding of the systematics of cycads has advanced greatly during the present decade, especially through the use of a large number of DNA sequences with proven value for the reconstruction of phylogenetic relationships (Treutlein and Wink, 2002; Hill et al., 2003; Rai et al., 2003; Bogler and Francisco-Ortega, 2004; Caputo et al., 2004; Chaw et al., 2005; Zgurski et al., 2008). These molecular data sets have added to a relatively limited number of morphological characters, traditionally established as the basis of intergeneric and/or intraspecific classification (Stevenson, 1990, 1992). New species of cycads from Mexico are still being described

*Corresponding author:
E-mail address: f_nicolalde@yahoo.com

(e.g. Schutzman and Vovides, 1998; Vovides et al., 2008a,b), but aside from a few exceptional cases (e.g. Nicolalde-Morejón et al., 2009a), the use of molecular information has not played a prominent role in the proposal of such taxonomic hypotheses.

In parallel with developments in the molecular systematics of cycads and seed plants in general (Mathews, 2009; and references therein), the use of DNA as a source of evidence in comparative biology has been recently extended outside the realm of phylogenetic inference *per se*. This extension has brought important changes in the disciplinary relationship between molecular biology/genomics, bioinformatics, and alpha-taxonomy (Miller, 2007). Crucial for these changes has been the explicit suggestion of the direct use of selected genomic regions as 'DNA barcodes', in close analogy to the way in which unique combinations of variable-width vertical marks work for the identification of industrialized goods in commercial and trade-related activities (Hebert et al., 2003a; Stoeckle and Hebert, 2008; for a survey of antecedents of this idea in molecular systematics, see Brower, 2006 and Meier, 2008). The proposal to employ molecular barcodes for large-scale, species-level identification is envisaged to be of utility in varied contexts beyond basic taxonomic research. These contexts include agricultural, clinical, ecological, forensic, illegal trade-related and even recreational applications, all of which would benefit a global community of users in need of reliable knowledge about biodiversity (Janzen, 2004; Hebert and Gregory, 2005; Stoeckle and Hebert, 2008).

A central biological claim backs up the current zoological implementation of a 'barcoding of life' international initiative (Hebert and Gregory, 2005; Ratnasingham and Hebert, 2007). According to this claim, certain regions in animal genomes are presumably identical or practically invariant within members (i.e. individuals) belonging to the same species, and simultaneously vary to a clearly detectable level when sampled between species (Hebert et al., 2003a,b, 2004). Confidence in the objective existence of such a 'barcoding gap' is derived from studies of nucleotide variation in a segment of the coding region of the mitochondrial cytochrome oxidase I (COI) gene, for which primers were designed with high PCR amplification success in a large number of animal species. The DNA barcoding initiative in animals quickly stabilized around phenetic (specifically, neighbor-joining-based) analyses of orthologues of this 'single-locus barcode' (e.g. Hebert et al., 2004; Barrett and Hebert, 2005; Ward et al., 2005; Hajibabaei et al., 2006; Smith et al., 2007, 2008). Nevertheless, use of genetic distance thresholds for animal COI sequences as a basis for DNA barcoding has been severely criticized by some molecular systematists (see, for instance, Moritz and Cicero, 2004; Meyer and Paulay, 2005; Brower, 2006; Rubinoff, 2006; Rubinoff et al., 2006a,b; Cognato and Sun, 2007; Wiemers and Fiedler, 2007). A common specific criticism in most of these papers is that there is an element of subjectivity in the definition of cutoffs of sequence similarity and dissimilarity—an old problem of phenetic approaches (DeSalle et al., 2005; Brower, 2006; DeSalle, 2007). This observation added to arguments in another set of articles critical of DNA barcoding; in those papers, political and/or sociological considerations have been appealed to in order to claim that the whole DNA barcoding enterprise is conceptually flawed (e.g. Ebach and Holdrege, 2005; Smith, 2005; Wheeler, 2004, 2005; Will and Rubinoff, 2004; Will et al., 2005; for some of the responses made to these statements, see Hebert and Gregory, 2005; Gregory, 2005; Packer et al., 2009; and for a list of the categories under which criticisms of DNA barcoding could be classified, see DeSalle et al., 2005, p. 1907).

Although additional criticisms of DNA barcoding as a research programme in plant taxonomy and systematics have also been raised (e.g. Seberg et al., 2003; Spooner, 2009), a sector of the botanical community has embraced the DNA barcoding initiative with clear optimism and sympathy (e.g. Chase et al., 2005; Cowan et al., 2006). As a result of a few preliminary studies and associated discussions—some of them held at international conferences—molecular biology-orientated plant systematists collectively decided that the chloroplast genome should provide a major proportion of plant DNA barcoding data. However, the specialists most heavily involved in selecting the 'definitive set' of plant DNA barcoding regions have had a difficult time agreeing upon which single locus, or combination of loci, might perform best in the largest number of groups (see Pennisi, 2007). Disagreements were evident in a number of proposals to settle the issue (e.g. Kress et al., 2005; Cowan et al., 2006; Chase et al., 2007; Kress and Erickson, 2007; Erickson et al., 2008; Fazekas et al., 2008; Lahaye et al., 2008; Ford et al., 2009). In keeping with their largely positive attitude towards DNA barcoding, however, the central point of discussion in these papers has been how much sequence data should be collected—and from how many chloroplast (or nuclear, in some special cases such as non-green mycoheterotrophs) genome regions—in order to successfully carry out rapid, cheap and reliable molecular identification of plant species across the widest possible plant diversity.

After an evaluation of the accumulated evidence, a consensus was reached in 2009 by a multinational assemblage of plant DNA barcoding researchers (CBOL Plant Working Group, 2009). The CBOL Plant Working Group settled for a two-locus 'standard' or 'core' barcode composed of two fragments of easy PCR amplification within the maturase K (*matK*) and the large subunit of the ribulose 1,5-bisphosphate carboxylase oxygenase (*rbcL*) loci, both of them chloroplast

coding regions with a long history of success in plant molecular systematics. According to the proposal of this group of botanists, the core plant DNA barcode could be supplemented by a group of three non-coding regions, also from the chloroplast genome—*atpF-atpH*, *psbK-psbI* and *trnH-psbA*—all of which had been considered (either separately or jointly) in several of the DNA barcoding studies preceding the consensus on the standard barcode.

Although fully recognizing that the *matK* + *rbcL* pair will now be recognized as the standard, universal barcode in land plants, in the present plant DNA barcoding study we have decided to address the discussion topic that was at the origin of the objectives pursued by the CBOL Plant Working Group. In the language chosen by the group, this topic amounts to the definition of three straightforward criteria: (i) universality, (ii) sequence quality and coverage, and (iii) discrimination. Our study focused on the three genera of cycads that occur in Mexico, *Ceratozamia* Brongn., *Dioon* Lindl., and *Zamia* L., and involved comparison of the performance of several different genome regions, all of them potentially useful for unique species identification in plants by CBOL standards. In a manner similar to the multinational group, we have explicitly considered the proposals resulting from the Second International Barcode of Life Conference (held in Taipei in 2007; see Pennisi, 2007) as well as relevant previous work on plant DNA barcoding, including taxonomically restricted studies in selected plant genera and/or families (Kress et al., 2005; Cowan et al., 2006; Chase et al., 2007; Kress and Erickson, 2007; Little and Stevenson, 2007; Fazekas et al., 2008; Lahaye et al., 2008; Ford et al., 2009; Seberg and Petersen, 2009). Importantly, we have also taken into account results of the pioneer DNA barcoding study conducted by Sass et al. (2007) on a set of selected cycad species representative from all biogeographical centres of diversity. As a result of these considerations, the definitive group of regions assayed here includes four chloroplast coding loci (the genes *matK*, *rpoC1*, *rpoB*, and *rbcL*), three non-coding intergenic spacer regions from the same plastid genomic compartment (*atpF/H*, *psbK/I*, and *trnH-psbA*), and finally a non-plastid genome region, the nuclear ribosomal internal transcribed spacer (*ITS*).

Initial processing of the resulting data sets included standard phenetic tests (e.g. neighbor-joining phenogram construction) following previously published DNA barcoding work (e.g. Hebert et al., 2003a,b, 2004). However, for advanced data analysis, we have employed the recently proposed ''character attributes organization system'' approach (CAOS; Sarkar et al., 2008), a character-based approximation to DNA barcoding implemented in the software package of the same name. To our knowledge, this is the first report in which a plant DNA barcoding data set has been explicitly analysed under these methodological assumptions (for examples of CAOS analyses with animal DNA barcoding data, see Kelly et al., 2007; Rach et al., 2008; Naro-Maciel et al., 2010). In the context of our empirical results, we discuss the reasons underlying our selection of this analytical tool. We also interpret our findings in the light of a recently proposed 'intrinsic limit' to resolution (i.e. percentage of unequivocal species identification) in plant DNA barcoding efforts. Finally, we take an explicit stance with respect to the future use of DNA data for species identification and species discovery.

Materials and methods

Sampling of biological materials

We collected leaf samples from all Mexican cycad species known to date from the three genera that occur in Mexico—*Ceratozamia*, *Dioon*, and *Zamia*—as published in the World Cycad List by Hill et al. (2007), plus three new species recently published (Vovides et al., 2008a,b; Nicolalde-Morejón et al., 2009a). We also collected leaf material from at least one individual from each of the non-Mexican cycad genera (see Table 2 below, Fig. 1). All materials were obtained from living plants included in the National Cycad Collection at the Jardín Botánico 'Francisco Javier Clavijero' (JBC), which houses materials from the Mexican species as well as specimens in cultivation from several countries, covering the entire order of the Cycadales. We also performed *ex profeso* field collections to complement the set of sampled materials for the present study. Leaf tissue from *Chigua restrepoi* D. W. Stev., *Zamia standleyi* Schutzman, *Z. tuerchkeimii* Donn. Sm., and *Z. prasina* W. Bull was kindly donated by the Montgomery Botanical Center (MBC).

Leaf genomic DNA extraction and PCR amplification (including DNA sequencing)

With the exception of the leaf samples transported from the field to the lab, fresh materials were always used for the total leaf genomic DNA extractions of material collected at the greenhouses of the JBC. For the extractions, we used either the DNAeasy Plant Mini kit (Qiagen) or a user-tailored protocol based on a widely employed CTAB DNA extraction procedure (Doyle and Doyle, 1987). PCR amplification and automated sequencing included all loci proposed at the Second International Barcode of Life conference (see Pennisi, 2007; Fazekas et al., 2008; CBOL, 2009; Ford et al., 2009). We then evaluated nucleotidic variability in four chloroplast coding regions (the genes *matK*, *rpoC1*, *rpoB* and *rbcL*), three non-coding intergenic spacer regions from the same plastid genomic compartment (*atpF-atpH*, *psbK-psbI* and *trnH-psbA*) and the nuclear

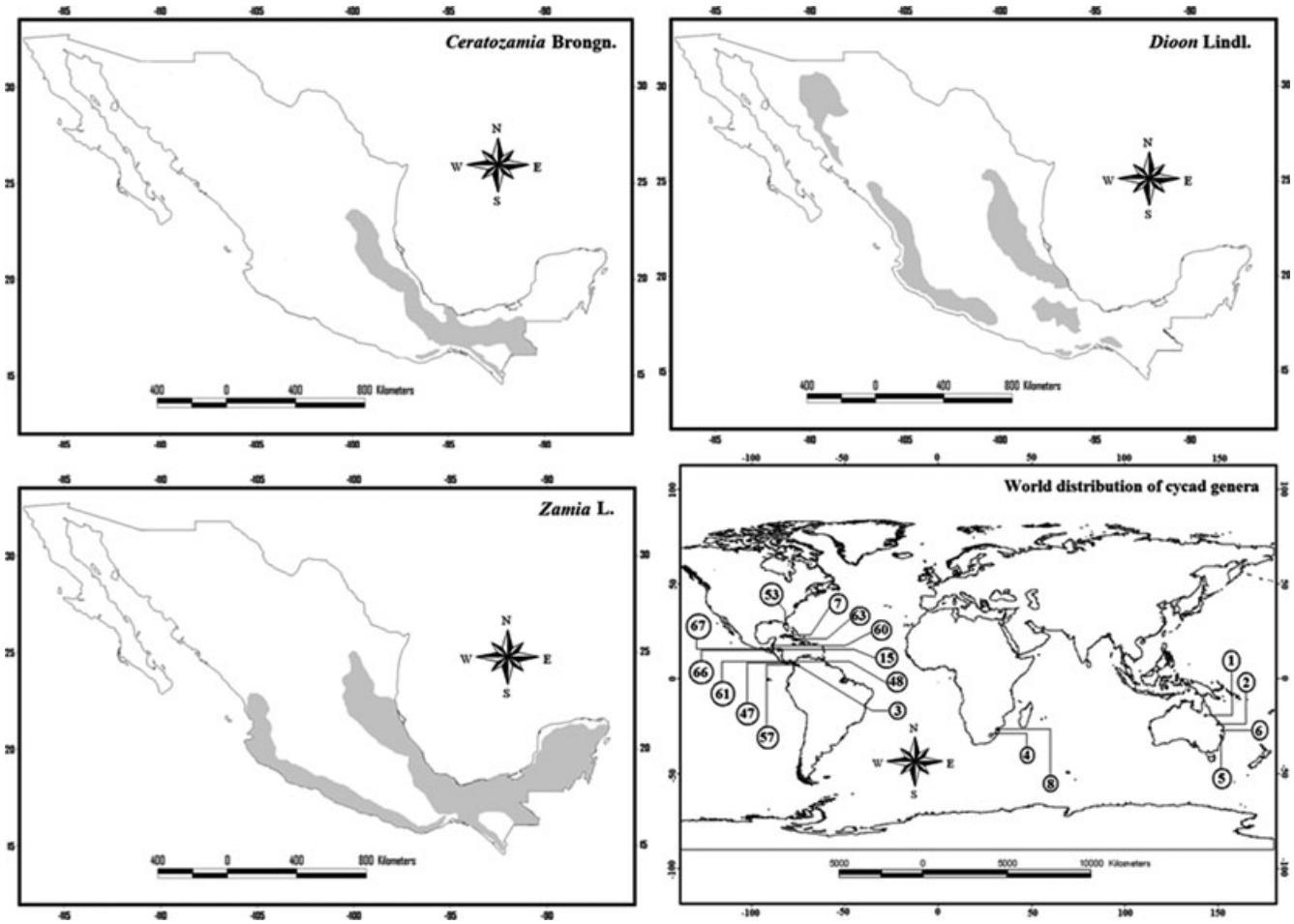

Fig. 1. Distribution of the cycads of Mexico and world distribution of representative species of the remaining genera in the order Cycadales.

ribosomal *ITS* as a complement non-chloroplast locus (see Table 1 for primers used in each type of PCR reaction, Table 2 for an overview of amplification success for each gene assayed for each species, and Table 3 for GenBank accession numbers).

PCR amplification experiments were performed as reported in recent plant DNA barcoding publications (e.g. Sass et al., 2007). Amplification products were visualized through gel electrophoresis in 1% agarose gels stained with ethidium bromide. In all cases where single bands were clearly detected, PCR products were directly purified using the QIAquick® PCR Purification Kit (Qiagen). Automated sequencing was carried out in Macrogen (South Korea; http://dna.macrogen.com).

Sequence analysis

Electropherograms were edited and contigs were assembled using the software program Sequencher 4.8 (Gene Codes Corp., Ann Arbor, MI, USA). Sequences were aligned in BioEdit 7.0.9 (Hall, 1999), through its implementation of the Clustal X (Thompson et al., 1997) multiple alignment mode. Alignments were imported into MacClade (Sinauer Associates, Sunderland, MA, USA) and further edited manually, after visual inspection. Indel regions resulting from the automated alignment were checked and accepted in the final edition. The resulting matrices were saved in Nexus format for ulterior character analysis, and are available from the corresponding author upon request.

Character-based analysis of chloroplast gene sequences from Mexican cycad species: identification of 'DNA diagnostics'

Neighbor-joining (NJ) phenograms were obtained for each matrix/amplified locus, using a Kimura two-parameter distance model in PAUP ver. 4.0b10 (Swofford, 2002). Although our ultimate purpose was not to carry out phenetic analyses with our matrices, this preliminary manipulation of the data and the selection of that particular evolutionary model allowed initial estimation of the performance of the analytical strategy that is currently more favoured by DNA barcoding

Table 1
Genes, primers and protocols used in the present plant DNA barcoding project

Gene	Primer	Sequence 5′–3′	Source of reaction conditions
rpoB	2	ATGCAACGTCAAGCAGTTCC	http://www.kew.org/barcoding/
	3	CCGTATGTGAAAAGAAGTATA	
rpoC1	1	GTGGATACACTTCTTGATAATGG	http://www.kew.org/barcoding/
	4	CCATAAGCATATCTTGAGTTGG	
matK	f	ATACCCCATTTTATTCATCC	http://www.kew.org/barcoding/
	r	GTACTTTTATGTTTACGAGC	
matK	2.1	CCTATCCATCTGGAAATCTTAG	http://www.kew.org/barcoding/
	2.1a	ATCCATCTGGAAATCTTAGTTC	
	5	GTTCTAGCACAAGAAAGTCG	http://www.kew.org/barcoding/
rbcL	3.2	CTTCCTCTGTAAAGAATTC	
	f	ATGTCACCACAAACAGAGACTAAAGC	http://www.kew.org/barcoding/
	R	GAAACGGTCTCTCCAACGCAT	
trnH/psbA	H	CGCGCATGGTGGATTCACAATCC	Shaw et al. (2005)
	A	GTTATGCATGAACGTAATGCTC	
atpF/H	F	ACTCGCACACACTCCCTTTCC	Lahaye et al. (2008)
	H	GCTTTTTATGGAAGCTTTAACAAT	
psbK/I	K	TTAGCCTTTGTTTGGCAA G	Lahaye et al. (2008)
	I	AGA GTTTGAGAGTAAGCAT	
nrITS	5a	CCTTATCATTTAGAGGAAGGAG	Sass et al. (2007)
	4 rev	TCCTCCGCTTATTGATATGC	

workers, especially in animals. Phenograms for each of the Mexican cycad genera and for the remaining sets of cycad species analysed here were stored in Nexus format and edited in MacClade, to be used as 'guide trees' for further processing with the CAOS software package (Sarkar et al., 2008). As stated by Rach et al. (2008, p. 238), guide trees provided to CAOS can be "generated using any number of tree building methods" (see also Sarkar et al., 2002). Subsequently, the most inclusive tree topology selected and used in these character-based analyses corresponds to Stevenson's (1992) phylogenetic hypothesis for the Cycadales.

Program P-Gnome from the CAOS software was executed according to the authors' instructions (Sarkar et al., 2008). For this purpose, we also consulted the Material and Methods section in the Odonata character-based DNA barcoding study of Rach et al. (2008). Actual determination of DNA diagnostics involved the manual revision of the "CAOS-attribute file" and "CAOS-group file" archives generated by P-Gnome. Only characters ('attributes') with confidence value of 1.00 were selected. Corroboration of these attributes was performed by visual inspection, and involved comparing the information of the "CAOS-group file" archives with the original, MacClade-edited matrices.

Results

trnH-psbA

Given that the *trnH-psbA* region was the first chloroplast locus to be suggested as a universal DNA barcode in plants (Kress et al., 2005; Kress and Erickson, 2007), we were particularly interested in the degree of nucleotidic variability at the species level that this chloroplast intergenic spacer could show in the three Mexican cycad genera. Sass et al. (2007) had already found that, with the exception of *Cycas*, all amplifications of cycad genomic DNA with the primers suggested by Kress et al. (2005) for this region yielded two distinct bands, even when the annealing temperature is raised to 62°C. We replicated this result for all the leaf genomic DNA samples tested. Through visual inspection of sequences from the smallest of these bands in all cases where amplification was successful (all but one *Dioon* species, several *Ceratozamia* and *Zamia* species, and all outgroups; see Table 2), we have unequivocally observed that *trnH-psbA* is not variable between species in either *Ceratozamia*, *Dioon*, or *Zamia*. The small *trnH-psbA* band was not analysed by Sass et al. (2007), but our result is compatible with the conclusion reached by these authors on the basis of the species-level variability of the larger band in at least one species per genus in the Cycadales. When comparing the *trnH-psbA* small band sequences of the entire set of species studied, we also noticed that the distribution of four indels varies consistently between genera (data not shown) and, therefore, might be useful for systematic purposes at the genus level. Characters corresponding to indels in the *trnH-psbA* region sequenced here might provide information to clarify the phylogenetic position of *Dioon*, possibly the most contentious subject in current cycad molecular systematics (Hill et al., 2003; Rai et al., 2003; Bogler and Francisco-Ortega, 2004; Chaw et al., 2005; Zgurski et al., 2008). However, our results do not lend support to the status of a 'potential' core DNA barcode that this non-coding chloroplast region still had

Table 2
Comparative performance—in terms of amplification success—of the six coding (*matK*, *rpoB*, and *rpoC1*) and non-coding (*atpF/H*, *psbK/I* and *trnH-psbA*) chloroplast regions, plus the two versions of the nuclear internal transcribed spacer region (*ITS* and *ITS2*), used in the present study*

			Genes							
No.	Taxon	Distribution	*psbI/K*	*atpF/H*	*rpoC1*	*trnH-psbA*	*matK*	*rpoB*	*ITS*	*ITS2*
1	*Cycas couttsiana*	Australia	+	+	+	+				
2	*Bowenia serrulata*	Australia	+	+	+	+				
3	*Chigua restrepoi*	Colombia	+	+	+	+				
4	*Encephalartos natalensis*	South Africa	+	+	+	+	+			
5	*Lepidozamia peroffskyana*	Australia	+	+	+	+				
6	*Macrozamia fawcettii*	Australia	+	+	+	+				
7	*Microcycas calocoma*	Cuba	+	+	+	+				
8	*Stangeria eriopus*	South Africa	+	+	+	+				
9	*Dioon angustifolium*	Mexico	+	+	+	+		+		
10	*Dioon argenteum*	Mexico	+	+	+	+		+		
11	*Dioon califanoi*	Mexico	+	+	+	+		+		
12	*Dioon caputoi*	Mexico	+	+	+	+		+		
13	*Dioon edule*	Mexico	+	+	+	+		+		
14	*Dioon holmgrenii*	Mexico	+	+	+	+		+		
15	*Dioon mejiae*	Honduras	+	+	+	+		+		
16	*Dioon merolae*	Mexico	+	+	+	+		+		
17	*Dioon purpusii*	Mexico	+	+	+	+		+		
18	*Dioon rzedowskii*	Mexico	+	+	+			+		
19	*Dioon sonorense*	Mexico	+	+	+	+		+		
20	*Dioon spinulosum*	Mexico	+	+	+	+		+		
21	*Dioon tomasellii*	Mexico	+	+	+	+		+		
22	*Dioon stevensonii*	Mexico	+	+	+	+		+		
23	*Ceratozamia alvarezii*	Mexico	+	+	+	+	+		+	
24	*Ceratozamia becerrae*	Mexico	+	+	+	+	+		+	
25	*Ceratozamia chimalapensis*	Mexico	+	+	+		+		+	
26	*Ceratozamia decumbens*	Mexico	+	+	+		+		+	
27	*Ceratozamia euryphyllidia*	Mexico	+	+	+	+	+		+	
28	*Ceratozamia hildae*	Mexico	+	+	+	+	+		+	
29	*Ceratozamia huastecorum*	Mexico	+	+	+	+	+		+	
30	*Ceratozamia kuesteriana*	Mexico	+	+	+		+		+	
31	*Ceratozamia latifolia*	Mexico	+	+	+	+	+		+	
32	*Ceratozamia matudae*	Mexico	+	+	+	+	+		+	
33	*Ceratozamia mexicana*	Mexico	+	+	+	+	+		+	
34	*Ceratozamia microstrobila*	Mexico	+	+	+		+		+	
35	*Ceratozamia miqueliana*	Mexico	+	+	+		+		+	
36	*Ceratozamia mirandae*	Mexico	+	+	+		+		+	
37	*Ceratozamia mixeorum*	Mexico	+	+	+		+		+	
38	*Ceratozamia morettii*	Mexico	+	+	+	+	+		+	
39	*Ceratozamia norstogii*	Mexico	+	+	+		+		+	
40	*Ceratozamia robusta*	Mexico	+	+	+		+		+	
41	*Ceratozamia sabatoi*	Mexico	+	+	+		+		+	
42	*Ceratozamia vovidesii*	Mexico	+	+	+	+	+		+	
43	*Ceratozamia whitelockiana*	Mexico	+	+	+		+		+	
44	*Ceratozamia zaragozae*	Mexico	+	+	+		+		+	
45	*Ceratozamia zoquorum*	Mexico	+	+	+		+		+	
46	*Zamia cremnophila*	Mexico	+	+	+	+	+			+
47	*Zamia cunaria*	Panama	+	+	+					+
48	*Zamia elegantissima*	Panama	+		+					+
49	*Zamia fischeri*	Mexico	+	+	+	+				+
50	*Zamia furfuracea*	Mexico	+	+	+	+	+			+
51	*Zamia herrerae*	Mexico	+	+	+	+	+			+
52	*Zamia inermis*	Mexico	+	+	+	+	+			+
53	*Zamia integrifolia*	USA	+	+	+					+
54	*Zamia katzeriana*	Mexico	+	+	+	+				+
55	*Zamia lacandona*	Mexico	+	+	+	+				+
56	*Zamia loddigesii*	Mexico	+	+	+	+				+
57	*Zamia manicata*	Colombia	+	+	+					+
58	*Zamia paucijuga*	Mexico	+	+	+	+	+			+
59	*Zamia polymorpha*	Mexico	+	+	+	+	+			+

Table 2
(*Continued*)

No.	Taxon	Distribution	Genes							
			psbI/K	*atpF/H*	*rpoC1*	*trnH-psbA*	*matK*	*rpoB*	*ITS*	*ITS2*
60	*Zamia prasina*	Belize	+	+	+	+	+			+
61	*Zamia pseudoparasitica*	Panama	+	+	+					+
62	*Zamia purpurea*	Mexico	+	+	+	+				+
63	*Zamia pygmea*	Cuba	+		+		+			+
64	*Zamia soconuscensis*	Mexico	+	+	+	+				+
65	*Zamia spartea*	Mexico	+	+	+	+	+			+
66	*Zamia standleyi*	Honduras	+	+	+	+				+
67	*Zamia tuerckheimii*	Guatemala	+	+	+					+
68	*Zamia variegata*	Mexico	+	+	+	+				+
69	*Zamia vazquezii*	Mexico	+	+	+	+				+

*A "plus" sign indicates successful amplification.

The order of columns additionally reflects the information content (under CAOS assumptions) in each locus. For the chloroplast regions, notice the following peculiarities: (i) *trnH-psbA* was in general successfully amplified in this study, but primers failed in a few species in each Mexican genus, (ii) *matK* was successfully amplified in all species of *Ceratozamia,* (iii) despite working for all species of *Dioon*, *rpoB* could not be amplified in either *Ceratozamia* or *Zamia*, and (iv) *matK* was not successfully amplified in any species of *Dioon*, and only in a few *Zamia* species.

prior to the selection of the two-locus standard plant DNA barcode (CBOL Plant Working Group, 2009, p. 12795).

matK and rbcL

The chloroplast coding regions for the *matK* and *rbcL* genes, both of them well established as important sources of characters for molecular systematics in angiosperms, are two loci whose variability we also decided to explore in detail. Our decision was based on recent claims of the relative superiority of *matK* as a DNA barcode in plants (Lahaye et al., 2008), the suggestion that *rbcL* might complement *trnH-psbA* in a two-locus global plant DNA barcode (Kress and Erickson, 2007), and the ultimate selection of both *matK* and *rbcL* as the two core DNA barcoding regions (CBOL Plant Working Group). It is important to note, however, that *rbcL* had been discarded as a DNA barcoding region by Sass et al. (2007) as it did not comply with basic reproducibility criteria. With an interest in checking if that negative result could be reversed, a random sample of five *Ceratozamia*, six *Dioon*, and five *Zamia* species was selected, for which we obtained complete amplification success. This result was, however, again linked to an absolute lack of variability at the nucleotide level within genera (data not shown). This result confirmed the conclusion of Sass et al. that *rbcL* is unsuitable as a DNA barcoding region in cycads.

In contrast to *rbcL*, Sass et al. (2007) did not completely eliminate *matK* from their selection of loci in their cycad DNA barcoding study. However, the region was not considered beyond Step 2 (i.e. testing of selected primer pairs) of their optimized flowchart. The elimination of *matK* for further testing in that study was due to failure of amplification in selected species of eight cycad genera [*Ceratozamia*, *Chigua*, *Dioon*, *Encephalartos* Lehm., *Lepidozamia* Regel, *Macrozamia* Miq., *Microcycas* (Miq.) A. DC. and *Stangeria* T. Moore] and to the fact that in the remaining two genera—*Cycas* L. and *Zamia* L.—amplification was only partially successful (products were not obtained in *Cycas platyphylla* K. D. Hill and *Zamia variegata* Warsz., one of three species tested for each genus). In their exploration of the performance of a series of clustering and similarity methods for DNA barcoding, Little and Stevenson (2007, p. 15) had also found that '*matK* barcodes' for *Dioon* and *Zamia* are "not positively differentiable from other Cycadopsida". It the present study, it was also impossible to obtain good quality amplifications of *matK* in the genus *Dioon*, and only a few *Zamia* samples behaved successfully as templates in the PCR reactions (Table 2). In contrast, leaf genomic DNA from all the species of *Ceratozamia*, tested with the set of primers that were not used by Sass et al. (2007), supported sufficient product amplification with high associated quality of sequences.

Despite the partial amplification success, *matK* data from *Zamia* species were not variable enough to be further considered useful for DNA barcoding purposes (data not shown). On the other hand, analysis of this *Ceratozamia* matrix with our preferred analytical regime, CAOS (Sarkar et al., 2008), retrieved 'DNA diagnostics' for only five of 24 species. However, our results do not necessarily rule out the use of *matK* in DNA barcoding because it remains to be seen if the use of different universal primers for this locus, which amplify larger segments of the gene (the primers that were successful in the present study yield an approximately 800-bp-long, N-terminal fragment of this coding region), is more suitable for obtaining better amplification/sequencing success. Also, improved

Table 3
GenBank accession numbers for the chloroplast gene sequences used in this study

Taxon	*psbI/K*	*atpF/H*	*rpoC1*	*trnH-psbA*	*matK*	*rpoB*	*ITS*	*ITS2*
Cycas couttsiana	GU807180	GU807114	GU807251					
Bowenia serrulata	GU807181	GU807115	GU807252					
Chigua restrepoi	GU807182	GU807116	GU807253	GU807394				
Encephalartos natalensis	GU807183	GU807117	GU807254		GU807320			
Lepidozamia peroffskyana	GU807184	GU807118	GU807255					
Macrozamia fawcettii	GU807185	GU807119	GU807256					
Microcycas calocoma	GU807186	GU807120	GU807257					
Stangeria eriopus	GU807187		GU807258					
Dioon angustifolium	GU807188	GU807121	GU807259	GU807405		GU807240		
Dioon argenteum		GU807122	GU807260	GU807406		GU807241		
Dioon califanoi	GU807189	GU807123	GU807261			GU807242		
Dioon caputoi	GU807190	GU807124	GU807262	GU807407		GU807243		
Dioon edule		GU807125	GU807263	GU807408		GU807244		
Dioon holmgrenii		GU807126	GU807264	GU807409		GU807245		
Dioon mejiae		GU807127	GU807265	GU807410				
Dioon merolae	GU807191	GU807128	GU807266	GU807411		GU807246		
Dioon purpusii		GU807129	GU807267	GU807412		GU807247		
Dioon rzedowskii		GU807130	GU807268					
Dioon sonorense	GU807192	GU807131	GU807269	GU807413				
Dioon spinulosum	GU807193	GU807132	GU807270	GU807414		GU807248		
Dioon tomasellii	GU807194	GU807133	GU807271	GU807416		GU807250		
Dioon stevensonii	GU807195	GU807134	GU807272	GU807415		GU807249		
Ceratozamia alvarezii	GU807196	GU807135	GU807273	GU807395	GU807321		GU807372	
Ceratozamia becerrae		GU807136	GU807274	GU807396	GU807322		GU807373	
Ceratozamia chimalapensis	GU807197	GU807137	GU807275		GU807323		GU807374	
Ceratozamia decumbens	GU807198	GU807138	GU807276		GU807324		GU807375	
Ceratozamia euryphyllidia		GU807139	GU807277	GU807397	GU807325		GU807376	
Ceratozamia hildae	GU807199	GU807140	GU807278	GU807398	GU807326		GU807377	
Ceratozamia huastecorum	GU807200	GU807141	GU807279	GU807399	GU807327		GU807378	
Ceratozamia kuesteriana	GU807201	GU807142	GU807280		GU807328		GU807379	
Ceratozamia latifolia	GU807202	GU807143	GU807281	GU807400	GU807329		GU807380	
Ceratozamia matudae	GU807203	GU807144	GU807282	GU807401	GU807330		GU807381	
Ceratozamia mexicana	GU807204	GU807145	GU807283	GU807402	GU807331		GU807382	
Ceratozamia microstrobila	GU807205	GU807146	GU807284		GU807332		GU807383	
Ceratozamia miqueliana	GU807206	GU807147	GU807285		GU807333		GU807384	
Ceratozamia mirandae	GU807207	GU807148	GU807286		GU807334		GU807385	
Ceratozamia mixeorum	GU807208	GU807149	GU807287		GU807335		GU807386	
Ceratozamia morettii	GU807209	GU807150	GU807288	GU807403	GU807336		GU807387	
Ceratozamia norstogii	GU807210	GU807151	GU807289		GU807337		GU807388	
Ceratozamia robusta	GU807211	GU807152	GU807290		GU807338		GU807389	
Ceratozamia sabatoi	GU807212	GU807153	GU807291		GU807339		GU807390	
Ceratozamia vovidesii	GU807213	GU807154	GU807292	GU807404	GU807340		GU807391	
Ceratozamia whitelockiana	GU807214	GU807155	GU807293		GU807341			
Ceratozamia zaragozae	GU807215	GU807156	GU807294		GU807342		GU807392	
Ceratozamia zoquorum	GU807216	GU807157	GU807295		GU807343		GU807393	
Zamia cremnophila	GU807217	GU807158	GU807296		GU807344			GU807352
Zamia cunaria	GU807235	GU807159	GU807297					GU807353
Zamia elegantissima			GU807298					
Zamia fischeri	GU807218	GU807160	GU807299	GU807417				GU807354
Zamia furfuracea	GU807219	GU807161	GU807300	GU807418	GU807345			GU807355
Zamia herrerae	GU807220	GU807162	GU807301	GU807419	GU807346			GU807356
Zamia inermis	GU807221	GU807163	GU807302		GU807347			
Zamia integrifolia	GU807239	GU807164	GU807303					GU807357
Zamia katzeriana	GU807222	GU807165	GU807304	GU807420				GU807358
Zamia lacandona	GU807223	GU807166	GU807305	GU807421				GU807359
Zamia loddigesii	GU807224	GU807167	GU807306	GU807422				GU807360
Zamia manicata	GU807236	GU807168	GU807307					GU807361
Zamia paucijuga	GU807225	GU807169	GU807308	GU807423	GU807348			GU807362
Zamia polymorpha	GU807226	GU807170	GU807309		GU807349			GU807363
Zamia prasina	GU807227	GU807171	GU807310					GU807364
Zamia pseudoparasitica	GU807237	GU807172	GU807311					
Zamia purpurea	GU807228	GU807173	GU807312	GU807424	GU807350			GU807365

Table 3
(*Continued*)

Taxon	*psbI/K*	*atpF/H*	*rpoC1*	*trnH-psbA*	*matK*	*rpoB*	*ITS*	*ITS2*
Zamia pygmea	GU807238		GU807313					
Zamia soconuscensis	GU807229	GU807174	GU807314	GU807425				GU807366
Zamia spartea	GU807231	GU807175	GU807315	GU807426	GU807351			GU807367
Zamia standleyi	GU807232	GU807176	GU807316					GU807368
Zamia tuerckheimii	GU807230	GU807177	GU807317					GU807369
Zamia variegata	GU807233	GU807178	GU807318	GU807427				GU807370
Zamia vazquezii	GU807234	GU807179	GU807319	GU807428				GU807371

taxon sampling could result in higher concentration of CAOS-based DNA diagnostics in other regions of *matK*. However, because this caveat could hardly apply to *rbcL* data, we conclude that using the CBOL Plant Working Group core DNA barcode as the main source of information for DNA barcoding in cycads would be inadequate.

psbK-psbI

The locus *psbK/I* is a chloroplast region that has only been included recently as part of two of the sets of candidate loci for DNA barcoding (according to the abstract by Kim et al. at the 2007 DNA barcoding Taipei conference; see Pennisi, 2007). We selected this locus as the fourth region of interest in the present DNA barcoding study mainly because no information was previously available for it, either in gymnosperms or in cycads in particular, and because this locus was included in the set of non-coding regions supplementary to the core DNA barcode by the CBOL Plant Working Group (2009). Pragmatic criteria for the selection of a good DNA barcoding region in the flowchart in Sass et al. (2007) were easily fulfilled, starting with amplification success (100%; Table 2). In view of its levels of variability for species in the three genera tested here and under the assumptions of the CAOS approach (see Fig. 2, and Table 4), we propose—*contra* Fazekas et al. (2008, p. e2802), who argued against its use due to a "higher failure rate in amplification and sequencing"—that the *psbK/I* locus should be seriously considered as a candidate for inclusion in any final DNA barcoding gene combination used in cycads. Presumably, performance of this chloroplast genome region for DNA barcoding in other gymnosperms might also be acceptable under character-based analytical regimes.

By itself, *psbK/I* allows variable and not very high (i.e. < 50%) levels of unique species identification in the three Mexican cycad genera (Table 4). However, in *Dioon* and *Zamia*, these levels are the highest for any single gene (Fig. 3), reaching 57% (8/14 species) in the former and 50% (12/24 species) in the latter. Only in *Ceratozamia* does the single-gene percentage of species

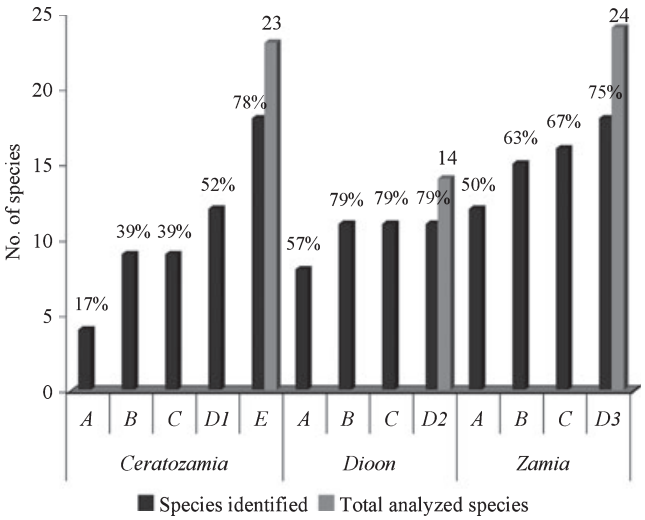

Fig. 2. Percentages of unique species identification using diverse combinations of candidate loci for DNA barcoding in *Ceratozamia*, *Dioon*, and *Zamia*, the three cycad genera occurring in Mexico. Each individual combination is identified with a letter, according to the following key: A, *psbK-psbI*; B, *psbK-psbI* + *atpF-atpH*; C, *psbK-psbI* + *atpF-atpH* + *rpoC1*; D1, *psbK-psbI* + *atpF-atpH* + *rpoC1* + *matK*; D2, *psbK-psbI* + *atpF-atpH* + *rpoC1* + *rpoB*; D3, *psbK-psbI* + *atpF-atpH* + *rpoC1* + *ITS2*; E, *psbK-psbI* + *atpF-atpH* + *rpoC1* + *matK* + *ITS*.

identification have a really low value for this region (17%, or 4/23 species). On the basis of the CAOS results for *psbK/I*, the remaining of our character-based DNA barcoding analysis was performed in order to quantify the increase of unique species identification percentages for loci added to this chloroplast non-coding region. The obvious choices for this purpose were *atpF-H* and *rpoC1*—another intergenic spacer and a coding chloroplast locus, respectively—given the perfect record (i.e. 100% success) of amplification that they showed for our study taxa (Table 2). It is worth noting that *atpF-H* and *rpoC1* had not been jointly considered before as candidates for low loci number plant DNA barcoding (Pennisi, 2007), and only the former was included in the set of supplementary DNA barcoding loci (CBOL Plant Working Group, 2009).

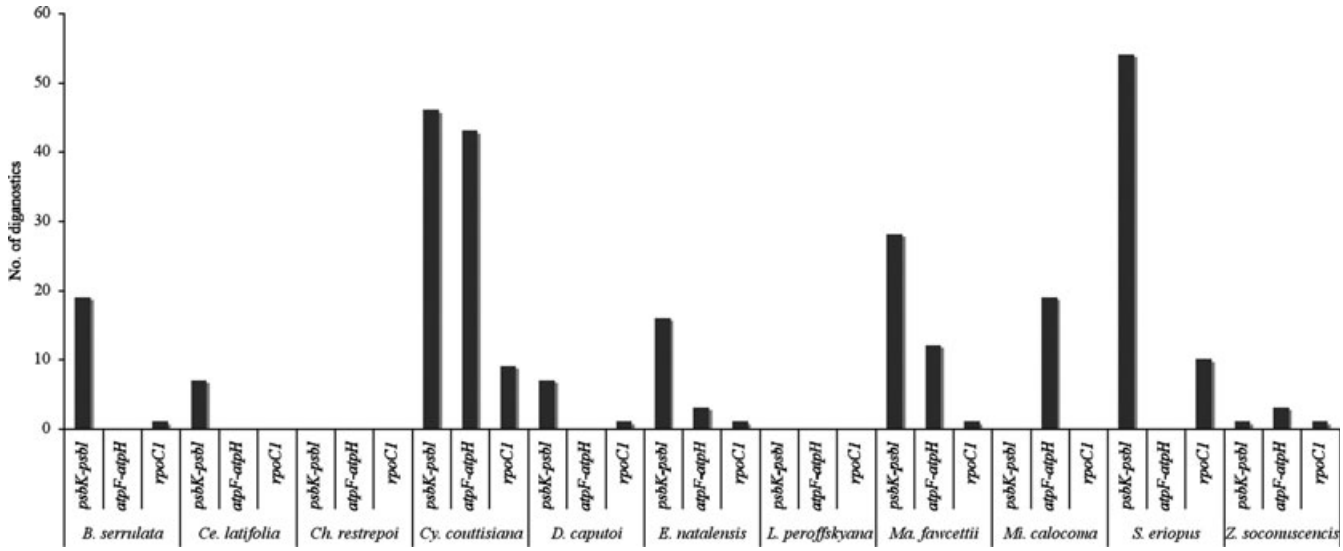

Fig. 3. DNA diagnostic sites under CAOS assumptions (Sarkar et al., 2008), obtained after the analyses performed in the software package of the same name, for one species in each valid cycad genus. The species set is ordered alphabetically. Species corresponding to the three Mexican genera—*Ceratozamia latifolia*, *Dioon caputoi*, and *Zamia soconuscensis*—were chosen at random.

Table 4
Comparative performance—in terms of number of species uniquely identified—of the various combinations of loci used in the present study, after character-based DNA barcoding analyses with the CAOS software (Sarkar et al., 2008)

Combination of loci	*Ceratozamia* Brongn.	*Dioon* Lindl.	*Zamia* L.
psbK-psbI	4/23	8/14	12/24
psbK-psbI + *atpF-atpH*	9/23	11/14	15/24
psbK-psbI + *atpF-atpH* + *rpoC1*	9/23	11/14	16/24
psbK-psbI + *atpF-atpH* + *rpoC1* + *rpoB*	–	11/14	–
psbK-psbI + *atpF-atpH* + *rpoC1* + *matK*	12/23	–	–
psbK-psbI + *atpF-atpH* + *rpoC1* + *matK* + ITS	18/23	–	–
psbK-psbI + *atpF-atpH* + *rpoC1* + ITS2	–	–	18/24

Table 5
Number of DNA diagnostic sites per species exemplar for each genera in the Cycadales, including the three target genera in the study, i.e. *Ceratozamia*, *Dioon*, and *Zamia*, after character-based analyses with the CAOS software (Sarkar et al., 2008) for the three chloroplast regions with 100% amplification success

No.	Species exemplar for each genera in the order Cycadales	*psbK-psbI*	*atpF-atpH*	*rpoC1*
1	*Bowenia serrulata* (W. Bull) Chamb.	19	0	1
2	*Ceratozamia latifolia* Miq.	7	0	0
3	*Chigua restrepoi* D. W. Stev.	0	0	0
4	*Cycas couttsiana* K. D. Hill	46	43	9
5	*Dioon caputoi* De Luca, Sabato & Vázq. Torres	7	0	1
6	*Encephalartos natalensis* R.A. Dyer & I. Verd.	16	3	1
7	*Lepidozamia peroffskyana* Regel	0	0	0
8	*Macrozamia fawcetii* C. Moore	28	12	1
9	*Microcycas calocoma* (Miq.) A. DC.	0	19	0
10	*Stangeria eriopus* (Kunze) Baill	54	–	10
11	*Zamia soconuscensis* Schutzman, Vovides & Deghan	1	3	1

The non-coding region *psbK-psbI* provided the highest number of sites, only failing to provide information in *Chigua*, *Lepidozamia*, and *Microcycas*, genera which were considered as reference taxa in this work.

Other loci: atpF-atpH, rpoC1, and ITSs

In *Ceratozamia*, *Dioon*, and *Zamia*, the addition of *atpF-H* to *psbK/I* increased the percentage of unique species identification to 39%, 79% and 63%, respectively, under CAOS assumptions (Fig. 2). However, in the first two genera, CAOS analyses with matrices where *rpoC1* was further added did not increase the percentage

of molecular identification, indicating that although the *rpoC1* primers used here are highly efficient for amplification with cycad genomic DNA, nucleotide variability is close to zero for cycad homologues of this gene (data not shown). Additional differences in the additive performance of a fourth region were observed for each genus: in *Ceratozamia*, the addition of *matK* improved species-level identification by 13% (reaching a global value of 52%), while in *Zamia* the addition of the *ITS2* increased the proportion to 75% (given a 67% correct species identification using the three-gene combination *atpF/H* + *psbK/I* + *rpoC1*). In contrast, for *Dioon* addition of a fourth gene *rpoB* did not contribute to any DNA diagnostics in discriminating unique species in the genus (Fig. 2). Finally, to achieve the highest percentage of identification in *Ceratozamia*, it was necessary to jointly analyse four chloroplast genome loci (*atpF-atpH* + *matK* + *psbK-psbI* + *rpoC1*) and one nuclear genome region (*ITS*).

Discussion

A DNA barcode for land plants: difficult roads toward a consensus

Although still unexplored in its full extent for a wide array of taxa, analyses of molecular evolutionary processes taking place in certain plant mitochondrial genomes indicate that the *COI* coding region, which has been adopted by consensus as the 'universal DNA barcode' in animals (Hebert and Gregory, 2005; Hebert et al., 2003a,b, 2004; Stoeckle and Hebert, 2008), is not suitable for analogous use in plants (Chase et al., 2005; Kress et al., 2005; Kress and Erickson, 2007; Spooner, 2009). A central interest in the international effort to develop DNA barcoding in plants has therefore been directed to find a different individual locus, or a combination of loci, that could fulfil the set of pragmatic criteria that during different stages of this international research initiative have been proposed as the mark of an adequate DNA barcode. A concise description of these criteria has been provided in a cycad DNA barcoding study (Sass et al., 2007) whose results form the basis for the present study.

The comparative performance of several different combinations of loci in plant genomes—particularly from the chloroplast compartment—has received special attention from botanists sympathetic to DNA barcoding. The selection of the best combination among these loci has been discussed in recent meetings specifically devoted to plant DNA barcoding initiatives, as well as in recent publications derived from these meetings (Kress et al., 2005; Cowan et al., 2006; Newmaster et al., 2006; Chase et al., 2007; Kress and Erickson, 2007; Erickson et al., 2008; Fazekas et al., 2008; Lahaye et al., 2008; Ford et al., 2009; see also Pennisi, 2007). Only recently, an international consensus on a standard plant DNA barcode has been reached by a group of highly renowned botanical experts (CBOL Plant Working Group, 2009). It is worth noting that some prominent plant DNA barcoding studies initially promoted the idea that empirical evidence was enough to support either *psbA-trnH*, a non-coding chloroplast region of about 400 bp (Kress et al., 2005; Kress and Erickson, 2007), or an N-terminal approximately 800-bp-long segment of the plastid coding region for the maturase K gene (*matK*; Lahaye et al., 2008), as sufficiently good candidates to achieve the status of individual, standard plant DNA barcodes. However, the fact that the final decision on a 'core' and a supplementary set of DNA barcodes ultimately involved a total of five (two coding and three non-coding) regions from the chloroplast genome clearly shows that the early hopes to reach a botanical analogue of the 'single-locus' zoological DNA barcode scheme could not be realized.

DNA barcoding in the cycads redux: reaching the 'Seberg–Petersen limit' with CAOS

In the present study, we have not ignored the international consensus on plant DNA barcoding arrived at by the Plant Working Group of the Consortium for the Barcoding of Life. However, guided mainly by the results of a previous study which dealt specifically with our taxon of interest (Sass et al., 2007), we decided to take a step back from that compromise. This decision allowed us to consider a whole range of reasonable possibilities for selecting DNA barcoding regions in the Mexican cycads. Assuming that the processing of these data under a methodology with robust theoretical foundations could only improve our DNA barcoding exercise, we selected the CAOS (Sarkar et al., 2008; see also Sarkar et al., 2002) as our analytical tool. In contrast to phenetic approaches to DNA barcoding, which are based on genetic distance thresholds, CAOS rests "on the fundamental concept that members of a given taxonomic group share attributes (…) that are absent from comparable groups." (Rach et al., 2008). The theoretical basis that justifies the use of these 'character attributes' or 'DNA diagnostics' for DNA barcoding has been stated clearly by DeSalle (2007, p. 1289; see also Bergmann et al., 2009): "in the case of DNA sequence diagnostics, there is no need to discuss arbitrary cutoffs for what is significant, or what is not, (as there is with phenetic approaches) because diagnostics are either there or they aren't." In summary, our stance in this paper has been to employ a character-based approach, without taking an *a priori* decision on which chloroplast genome regions are 'best' for DNA barcoding in the Mexican cycads. We consider that such a stance might further enable us to re-analyse the

balances intrinsic to the criteria that the CBOL Plant Working Group itself considered central in their selection of the standard DNA barcode for land plants, namely (i) universality, (ii) sequence quality and coverage, and (iii) discrimination (CBOL Plant Working Group, 2009, p. 12794).

In this context, we think it is relevant to retrieve a portion of one of the recently published, pre-CBOL Plant Working Group consensus plant DNA barcoding papers, where equality of performance was supported for multiple multilocus combinations. The authors of that work stated that "from the perspective of species resolution, the identity of the regions used is less important than the number" (Fazekas et al., 2008, p. e2802). We agree completely with this statement; in fact, we further suggest that for each of the three cycad genera studied in detail here a *different* combination of chloroplast regions clearly works 'best' as a DNA barcoding set. Judging from the distribution of nucleotide variation we observed in the present study, we estimate that if our research team gathered more sequence data from several additional chloroplast and/or nuclear regions from Mexican cycads, the percentages of unique (i.e. correct) species identification would probably approach a higher value, simply because more DNA diagnostics would probably accumulate. This could certainly be the case in *Zamia* species from Megamexico, a set of cycad taxa for which we have an increasingly better understanding of basic taxonomy and systematics (Nicolalde-Morejón et al., 2009b). Moreover, we consider that after such effort the construction of a fairly good DNA barcoding reference library would be feasible for the Mexican cycads as a whole, given the relatively low number of species in each of the three genera vis à vis the excellent sampling that we have of their diversity, down to the population level. Aiming for large amounts of genomic information in order to obtain good levels of unique species recognition had been shown in some publications previous to the consensus (for example, see Ford et al., 2009), but the fact that surpassing a certain threshold of 'sequencing volume' would render the approach too expensive and, ultimately, unmanageable was simultaneously acknowledged by the plant DNA barcoding community. This crucial point was actually reflected in the CBOL Plant Working Group decision to choose a two-locus rather than three-locus 'core' barcode (CBOL Plant Working Group, 2009, p. 12795). In line with this point of view, we conclude that, given our current knowledge, the 'best' overall combination of chloroplast for DNA barcoding in the Mexican cycads is the minimal two-locus set that worked for *Dioon*. At the same time, we recognize that such a two-locus barcode is insufficient to attain very high levels of molecular species identification.

Our study might illuminate another aspect of greater generality for plant DNA barcoding. Having identified the tension between the need to sequence multiple loci and the convenience of not exceeding a certain threshold of sequencing, we recall that Fazekas et al. (2008) also mentioned that "fundamental upper limits" exist to "what is possible for any current plant DNA barcoding approach". Interestingly, in their taxonomically restricted study of DNA barcoding in the genus *Crocus*, Seberg and Petersen (2009, p. e4598) took a step forward in the direction indicated by Fazekas et al., when they affirmed that "in a taxonomic setting and with a reasonable effort *it is unlikely that barcoding will enable us to identify more than around 70–75% of the known species*—in some instances less, in some instances more" (emphasis added). Again, this boundary of resolution was acknowledged in the international consensus: "the unique identification to species level of 72% of cases and to 'species groups' in the remainder will be useful for many applications of DNA barcoding (…)" (CBOL Plant Working Group, 2009, p. 12796). Using a relatively low number of combinations of chloroplast genes in our study of the Mexican cycads, we might have reached the Seberg and Petersen limit. The key point here is that, for each genus, we arrived at this limit using a *different* combination of loci.

Conclusions: gene quantity versus universality and the relevance of analytical methods in plant DNA barcoding

Considering the clear-cut unsuitability of both *matK* and *rbcL* as the main DNA barcoding loci in cycads, we think that an unavoidable conclusion from the cycad work presented here in fact echoes the idea expressed by Fazekas et al. (2008) discussed above. This notion was implicitly adopted by the CBOL Plant Working Group as one of their directives: in DNA barcoding, what should matter is not so much gene identity (i.e. which regions are used), but gene quantity (i.e. a quantity that is *low* must be preferred). Although we seem to have reached the 70–75% 'limit of resolution' mentioned by Seberg and Petersen (2009), at the same time we suggest that the substitution of *rpoC1* (a gene region which had 100% amplification success, but zero DNA diagnostics) with another locus with suboptimal amplification success—but higher nucleotide variability—could improve the percentages of unique species identification in some of the Mexican cycad genera. This still unknown coding or non-coding region might lie in the chloroplast, but it might well have been overlooked because the initial plant DNA barcoding pilot studies did not include a thorough sampling of cycads. In this regard, future projects orientated towards the sequencing of complete chloroplast genomes in selected species from all cycad genera could be useful in the search for those alternative regions.

On the basis of the results of our character-based analysis of potential DNA barcoding regions in the

Mexican cycads (Table 5), we suggest that the future of plant DNA barcoding might lie in being flexible with respect to the seemingly unattainable ideal of a unique set of universal molecular barcodes. This does not mean that we question the merits entailed by the current consensus around a standard DNA barcode for the land plants (CBOL Plant Working Group, 2009), and the utility that the loci selected there might have in several plant taxa. However, we think that a serious consideration of the use of 'local' combinations of plant DNA barcodes, that work best for relatively restricted taxonomic–phylogenetic assemblages, should not be discarded yet. This relaxation of constraints in the international plant DNA barcoding initiative will ultimately facilitate the construction of *optimal* although not 'perfect' (see Chase and Fay, 2009) reference libraries for DNA barcoding, with applications in either floristic, conservation biology, or strictly taxonomic–nomenclatural research contexts. Given current anthropogenic pressures on charismatic but rare plant groups that inhabit the Neotropical regions of the planet, this might prove important in the short term. At the same time, however, we call attention to the fact that many plant DNA barcoding studies are based on analytical approaches that do not necessarily constitute the best available options, on theoretical grounds. As the number of available methodologies for analysing DNA barcoding increases, it is obvious that the issue of which strategy is 'best' will become more pressing. In this regard, we suggest that molecular biology-minded botanists should not forget that the primary use of DNA barcodes is in species identification (DeSalle, 2006), and that the employment of DNA data in species discovery is theoretically justified only in an integrative taxonomic context (*sensu* DeSalle et al., 2005).

Acknowledgements

This research was supported by CONABIO grant no. GE004 and NSF Grants BSR-8607049 and EF-0629817 to D.W.S. The first author thanks the Red Latinoamericana de Botánica for the award of a PhD fellowship (RLB-06-D2; Systematics Program, Instituto de Ecología, A.C., Xalapa, Mexico), and especially Javier Simonetti for additional support during the final phase of this project. Thanks to Raúl Jiménez-Rosenberg and Martha Gual for their encouragement at the planning stages. Thanks to Alejandro Espinosa de los Monteros for data analysis support. Our sincere thanks are expressed to Julia Hernández Villa for technical assistance in the laboratory. We are also grateful to the staff of the Jardín Botánico Fco. J. Clavijero at the Instituto de Ecología, A.C. (Xalapa) for access to living specimens of the Mexican National Cycad Collection (Colección de Cycadas Mexicanas), and to the Montgomery Botanical Center, Miami, Florida, for supplying samples of some species. Finally, we thank the two anonymous reviewers who made very thoughtful constructive criticisms.

References

Barrett, R.D.H., Hebert, P.D.N., 2005. Identifying spiders through DNA barcodes. Can. J. Zool. 83, 481–491.

Bergmann, T., Hadrys, H., Breves, G., Schierwater, B., 2009. Character-based DNA barcoding: a superior tool for species identification. Berl. Munch. Tierarztl. Wochenschr. 122, 446–450.

Bogler, D.J., Francisco-Ortega, J., 2004. Molecular systematic studies in cycads: evidence from *trn*L intron and ITS2 rDNA sequences. Bot. Rev. 70, 260–273.

Brower, A.V.Z., 2006. Problems with DNA barcodes for species delimitation: 'ten species' of *Astraptes fulgerator* reassessed (Lepidopetera: Hesperiidae). Syst. Biodivers 4, 127–132.

Caputo, P., Cozzolino, S., De Luca, P., Moretti, A., Stevenson, D.W., 2004. Molecular phylogeny of *Zamia* (Zamiaceae). In: Walters, T., Osborne, R. (Eds.), Cycad Classification: Concepts and Recommendations. CABI Publishing, Wallingford, UK, pp. 149–157.

CBOL Plant Working Group (2009) A DNA barcode for land plants. Proc. Natl Acad. Sci. USA 106, 12794–12797.

Chase, M.W., Fay, M.F., 2009. Barcoding of plants and fungi. Science 325, 682–683.

Chase, M.W., Salamin, N., Wilkinson, M., Dunwell, J.M., Kesanakurth, R.P., Haidar, N., Savolainen, V., 2005. Land plants and DNA barcodes: short-term and long-term goals. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 360, 1889–1895.

Chase, M.W., Cowan, R.S., Hollingsworth, P.M., van der Berg, C., Madriñán, S., Petersen, G., Seberg, O., Jørgsensen, T., Cameron, K.M., Carine, M., Pedersen, N., Hedderson, T.A.J., Conrad, F., Salazar, G.A., Richardson, J.E., Hollingsworth, M.L., Barraclough, T.G., Kelly, L., Wilkinson, M., 2007. A proposal for a standardized protocol to barcode all land plants. Taxon 56, 295–299.

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

Cognato, A.I., Sun, J.H., 2007. DNA based cladograms augment the discovery of a new *Ips* species from China (Coleoptera: Curculionidae: Scolytinae). Cladistics 23, 539–551.

Cowan, R.S., Chase, M.W., Kress, W.J., Savolainen, V., 2006. 300,000 species to identity: problems, progress, and prospects in DNA barcoding of land plants. Taxon 55, 611–616.

DeSalle, R., 2006. Species discovery versus species identification in DNA barcoding efforts: response to Rubinoff. Conserv. Biol. 20, 1545–1547.

DeSalle, R., 2007. Phenetic and DNA taxonomy; a comment on Waugh. Bioessays 29, 1289–1290.

DeSalle, R., Egan, M.G., Siddall, M., 2005. The unholy trinity: taxonomy, species delimitation and DNA barcoding. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 360, 1905–1916.

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

Ebach, M., Holdrege, C., 2005. DNA barcoding is no substitute for taxonomy. Nature 434, 697.

Erickson, D.L., Spouge, J., Resch, A., Weigt, L.A., Kress, W.J., 2008. DNA barcoding in land plants: developing standards to quantify and maximize success. Taxon 57, 1304–1316.

Fazekas, A.J., Burgess, K.S., Kesanakurti, P.R., Graham, S.W., Newmaster, S.G., Husband, B.C., Percy, D.M., Hajibabaei, M.,

Barrett, S.C.H., 2008. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS ONE 3, e2802.

Ford, C.S., Ayres, K.L., Toomey, N., Haider, N., van Alphen Stahl, J., Kelly, L.J., Wikström, N., Hollingsworth, P.M., Duff, R.J., Hoot, S.B., Cowan, R.S., Chase, M.W., Wilkinson, M.J., 2009. Selection of candidate coding DNA barcoding regions for use on land plants. Bot. J. Linn. Soc. 159, 1–11.

Gregory, T.R., 2005. Synergy between sequence and size in large-scale genomics. Nat. Rev. Genet. 6, 699–708.

Hajibabaei, M., Janzen, D.H., Burns, J.M., Hallwachs, W., Hebert, P.D.N., 2006. DNA barcodes distinguish species of tropical Lepidoptera. Proc. Natl Acad. Sci. USA 103, 968–971.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence aligment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95–98.

Hebert, P.D.N., Gregory, T.R., 2005. The promise of DNA barcoding for taxonomy. Syst. Biol. 54, 852–859.

Hebert, P.D.N., Cywinska, A., Ball, S.L., deWaard, J.R., 2003a. Biological identifications through DNA barcodes. Proc. R. Soc. Lond. B 270, 313–321.

Hebert, P.D.N., Ratnasingham, S., deWaard, J.R., 2003b. Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. Proc. Biol. Sci. 270(Suppl. 1), S96–S99.

Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H., Hallwachs, W., 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proc. Natl Acad. Sci. USA 101, 14812–14817.

Hill, K.D., Chase, M.W., Stevenson, D.W., Hills, H.G., Schutzman, B., 2003. The families and genera of cycads: a molecular phylogenetic analysis of Cycadophyta based on nuclear and plastid DNA sequences. Int. J. Plant Sci. 164, 933–948.

Hill, K.D., Stevenson, D.W., Osborne, R., 2007. The world list of cycads/La lista mundial de cícadas. Mem. N. Y. Bot. Gard. 97, 454–483.

Janzen, D.H., 2004. Now is the time. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 359, 731–732.

Kelly, R.P., Sarkar, I.N., Eernisse, D.J., DeSalle, R., 2007. DNA barcoding using chitons (genus *Mopalia*). Mol. Ecol. Notes 7, 177–183.

Kress, W.J., Erickson, D.L., 2007. A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. PLoS ONE 2, e508.

Kress, W.J., Wurdack, K.J., Zimmer, E.A., Weigt, L.A., Janzen, D.H., 2005. Use of DNA barcodes to identify flowering plants. Proc. Natl Acad. Sci. USA 102, 8369–8374.

Lahaye, R., van der Bank, M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G., Maurin, O., Duthoit, S., Barraclough, T.G., Savolainen, V., 2008. DNA barcoding the floras of biodiversity hotspots. Proc. Natl Acad. Sci. USA 105, 2923–2928.

Little, D.P., Stevenson, D.W., 2007. A comparison of algorithms for the identification of species using DNA barcodes: examples for gymnosperms. Cladistics 23, 1–21.

Mathews, S., 2009. Phylogenetic relationships among seed plants: persistent questions and the limits of molecular data. Am. J. Bot. 96, 228–236.

Meier, R., 2008. DNA sequences in taxonomy: opportunities and challenges. In: Wheeler, Q.D. (Ed.), The New Taxonomy. CRC Press, Boca Raton, Florida, pp. 95–127.

Meyer, C.P., Paulay, G., 2005. DNA barcoding: error rates based on comprehensive sampling. PLoS Biol. 3, e422.

Miller, S.E., 2007. DNA barcoding and the renaissance of taxonomy. Proc. Natl Acad. Sci. USA 104, 4775–4776.

Mortiz, C., Cicero, C., 2004. DNA barcoding: promise and pitfalls. PLoS Biology 3, e354.

Naro-Maciel, E., Reid, B., Fitsimmons, N.N., Lee, M., DeSalle, R., Amato, G., 2010. DNA barcodes for globally threatened marine turtles: a registry approach to documenting biodiversity. Mol. Ecol. Resour. 10, 252–263.

Newmaster, S.G., Fazekas, A.J., Ragupathy, S., 2006. DNA barcoding in land plants: evaluation of rbcL in a multigene tiered approach. Can. J. Bot. 84, 335–341.

Nicolalde-Morejón, F., Vergara-Silva, F., González-Astorga, J., Vovides, A.P., Espinosa de los Monteros, A., 2009a. Reciprocal illumination of morphological characters upon a molecular hypothesis supports the proposal of a new species of cycad from Mexico. Syst. Biodivers. 7, 73–79.

Nicolalde-Morejón, F., Vovides, A.P., Stevenson, D.W., 2009b. Taxonomic revision of *Zamia* in Mega-Mexico. Brittonia 61, 301–335.

Norstog, K., Nicholls, T.J., 1997. The Biology of the Cycads. Cornell University Press, Ithaca, New York.

Packer, L., Gibbs, J., Sheffield, C., Hanner, R., 2009. DNA barcoding and the mediocrity of morphology. Mol. Ecol. Resour. 9, 42–50.

Pennisi, E., 2007. Wanted: a barcode for plants. Science 318, 190–191.

Rach, J., DeSalle, R., Sarkar, I.N., Schierwater, B., Hadrys, H., 2008. Character-based DNA barcoding allows discrimination of genera, species and populations in Odonata. Proc. Biol. Sci. 275, 237–247.

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

Ratnasingham, S., Hebert, P.D.N., 2007. BOLD: the barcode of life data system (http://www.barcodinglife.org). Mol. Ecol. Notes 7, 355–364.

Rubinoff, D., 2006. DNA barcoding evolves into the familiar. Conserv. Biol. 20, 1548–1549.

Rubinoff, D., Cameron, S., Will, K., 2006a. Are plant DNA barcodes a search for the Holy Grail? Trends Ecol. Evol. 21, 1–2.

Rubinoff, D., Cameron, S., Will, K., 2006b. A genomic perspective on the shortcomings of mitochondrial DNA for "barcoding" identification. J. Hered. 97, 581–594.

Sarkar, I.N., Thornton, J.W., Planet, P.J., Figurski, D.H., Schierwater, B., DeSalle, R., 2002. An automated phylogenetic key for classifying homeoboxes. Mol. Phylogenet. Evol. 24, 388–399.

Sarkar, I.N., Planet, P.J., DeSalle, R., 2008. CAOS software for use in character-based DNA barcoding. Mol. Ecol. Resour. 8, 1256–1259.

Sass, Ch., Little, D.P., Stevenson, D.W., Specht, Ch.D., 2007. DNA barcoding in the Cycadales: testing the potential of proposed barcoding markers for species Identification of cycads. PLoS ONE 11, e1154.

Schutzman, B., Vovides, A.P., 1998. A new *Zamia* (Zamiaceae-Cycadales) from Eastern Chiapas, Mexico. Novon 8, 441–446.

Seberg, O., Petersen, G., 2009. How many loci does it take to DNA barcode a *Crocus*? PLoS ONE 42, e4598.

Seberg, O., Humphries, J.Ch., Knapp, S., Stevenson, D.W., Petersen, G., Scharff, N., Andersen, N.M., 2003. Shortcuts in systematics? A commentary on DNA-based taxonomy Trends Ecol. Evol. 18, 63–65.

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

Smith, V.S., 2005. DNA barcoding: perspectives from a "Partnerships for Enhancing Expertise in Taxonomy". Syst. Biol. 54, 841–844.

Smith, M.A., Wood, D.M., Janzen, D.L., Hallwachs, W., Hebert, P.D.N., 2007. DNA barcodes affirm that 16 species of apparently generalist tropical parasitoid flies (Diptera, Tachinidae) are not all generalists. Proc. Natl Acad. Sci. USA 104, 4967–4972.

Smith, M.A., Poyarkov, N.A., Hebert, P.D.N., 2008. CO1 DNA barcoding amphibians: take the chance, meet the challenge. Mol. Ecol. Resour. 8, 235–246.

Spooner, D.M., 2009. DNA barcoding will frecuently fail in complicated groups: an example in wild potatoes. Am. J. Bot. 96, 1177–1189.

Stevenson, D.W., 1990. Morphology and systematics of the Cycadales. Mem. N. Y. Bot. Gard. 57, 8–55.

Stevenson, D.W., 1992. A formal classification of the extant cycads. Brittonia 44, 220–223.

Stoeckle, M.Y., Hebert, P.D.N., 2008. Barcode of life. Sci. Am. 299, 82–88.

Swofford, D.L. 2002. PAUP*. Phylogenetic Analysis Using Parsimony. Sinauer Associates, Sunderland, MA.

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 tool. Nucleic Acids Res. 25, 4876–4882.

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

Vovides, A.P., González-Astorga, J., Pérez-Farrera, M.A., González, D., Bárcenas, C., Iglesias, C., 2007. The cycads of Mexico: 25 years of research and conservation. Mem. N. Y. Bot. Gard. 97, 611–641.

Vovides, A.P., Avendaño, S., Pérez-Farrera, M.A., González-Astorga, J., 2008a. A new species of *Ceratozamia* (Cycadales, Zamiaceae) from Veracruz, Mexico. Novon 18, 109–114.

Vovides, A.P., Pérez-Farrera, M.A., González-Astorga, J., Iglesias, C., 2008b. A new species of *Ceratozamia* (Zamiaceae) from Oaxaca, Mexico with comments on habitat and relationships. Bot. J. Linn. Soc. 157, 169–175.

Ward, R.D., Zemlak, T.S., Innes, B.H., Last, P.R., Hebert, P.D.N., 2005. DNA barcoding Australia's fish species. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 360, 1847–1857.

Wheeler, Q., 2004. Taxonomic triage and the poverty of phylogeny. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 359, 571–583.

Wheeler, Q., 2005. Losing the plot: DNA ''barcodes'' and taxonomy. Cladistics 21, 405–407.

Wiemers, M., Fiedler, K., 2007. Does the DNA barcoding gap exist? – a case study in blue butterflies (Lepidoptera: Lycaenidae). Front. Zool. 4, e8.

Will, K.W., Rubinoff, D., 2004. Myth of the molecule: DNA barcodes cannot replace morphology for identification and classification. Cladistics 20, 47–55.

Will, K.W., Mishler, B.D., Wheeler, Q.D., 2005. The perils of DNA barcoding and the need for integrative taxonomy. Syst. Biol. 54, 844–851.

Zgurski, J.M., Rai, H.S., Fai, Q.M., Bogler, D.J., Francisco-Ortega, J., Graham, S.W., 2008. How well do we understand the overall backbone of cycad phylogeny? New insights from a large, multigene plastid data set. Mol. Phylogent. Evol. 47, 1232–1237.