Kasetsart J. (Nat. Sci.) 44 : 544 - 557 (2010)

Phylogenetic Relationships Within Cycadaceae Inferred from Non-Coding Regions of Chloroplast DNA

Pattamon Sangin[1,2], Anders J. Lindstrom[3], Goro Kokubugata[4], Minta Chaiprasongsuk[5] and Mingkwan Mingmuang[2]*

ABSTRACT

The intrageneric relationships among 24 species of five *Cycas* sections based on the combined data of three non-coding regions of chloroplast DNA: *trn*S-*trn*G, *psb*M-*trn*D and *trn*L-*trn*F, were determined using Neighbor-joining, Maximum parsimony and Maximum likelihood methods. All three methods showed similar topology, which divided *Cycas* into several clades. The first clade consisted of two sections, *Cycas* and *Indosinenses*, while the other clades contained *Asiorientales*, *Wadeanae* and *Stangerioides*, which could not be resolved. This result indicated the polyphyletic group of *Cycas* sections. Base substitution patterns further revealed that the subsection *Rumphiae* of the section *Cycas* could be separated into two groups (*C. rumphii* and *C. edentata*). In addition, the high bootstrap demonstrated that *C. taitungensis* was closely related to *C. revotuta*, while two other species of the section *Wadeanae* (*C. wadei* and *C. curranii*) were closely related, which agreed with the morphology. Data analyses suggested that the *trn*S-*trn*G sequences were more informative than the *psb*M-*trn*D and *trn*L-*trn*F regions in addressing phylogeny. Short and moderate repeated sequences detected in this region also indicated the high rate of evolution in the *Cycas* species. The 14 bp tandem repeats identified in this study are the first to be reported on the non-coding chloroplast DNA of *Cycas*.

Keywords: Cycadaceae, non-coding chloroplast DNA, *trn*S-*trn*G, *psb*M-*trn*D, *trn*L-*trn*F

INTRODUCTION

Cycads are ancient plants with a long continuous heredity. They are classified into 3 families, 11 genera (Stevenson, 1992) and 303 known species (Hill *et al*., 2004). Although they distributed all over the world, they are mainly found along the intertropical belt, containing Africa, India, Indonesia and north Australia. *Cycas*, the only genus in the Cycadaceae family, is composed of 98 species and is widely distributed. In addition, morphological and molecular studies supported *Cycas* as a basal group of the remaining cycads (Treutlein and Wink, 2002; Rai *et al*., 2003). In a recent study, *Cycas* species were classified into five sections, *Asiorientales*, *Cycas*, *Wadeanae*, *Indosinenses* and *Stangerioides*, based on morphological characters

1 Bioscience program, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand.
2 Department of Earth Science, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand.
3 Nong Nooch Tropical Botanical Garden, 34/1 Sukhumvit Hgw, Najomtien, Sattahip, Chon Buri 20250, Thailand.
4 Department of Botany, National Museum of Nature and Science, 4-1-1 Amakubo, Tsukuba 305-0005, Japan.
5 Department of Botany, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand.
* Corresponding author, e-mail: fscimkm@ku.ac.th

Received date : 09/12/09 Accepted date : 12/03/10

(Hill, 2004). Various authors, on the other hand, have proposed different classifications of *Cycas* (Smitinand, 1971; De Laubenfels and Adema, 1998), but the genetic relationships among the *Cycas* sections are still unclear.

Recently, several studies have shown that chloroplast DNA analysis is a useful method for plant phylogeny and evolution. The chloroplast genome is uniparentally inherited in most plants. Although slow evolution of the coding regions of chloroplast DNA cannot resolve relationships among lower-level taxa, the variation of non-coding regions is higher than in coding regions and provides more informative characters in phylogenetic studies (Xu *et al*., 2000). Non-coding regions of the chloroplast genome have been used to elucidate phylogenetic relationships at lower taxonomic levels, specifically intergeneric, interspecific and intraspecific, in different plant groups (Kim *et al*., 1999; Small *et al*., 2005; Mort *et al*., 2007). To date, only one study has focused on intrageneric classification in *Cycas* using molecular and morphological data (Hill, 2004). Therefore, in this study, the phylogeny of 24 species of *Cycas* was inferred using the combined data of *trn*S-*trn*G, *psb*M-*trn*D and *trn*L-*trn*F non-coding regions to investigate the genetic relationships of *Cycas*. The study also compared the utility of three non-coding regions for resolving intrageneric relationships within *Cycas*.

MATERIALS AND METHODS

Plant materials and DNA extraction

Total genomic DNA of 24 *Cycas* species was extracted from plants cultivated in the Nong Nooch Tropical Botanical Garden, Thailand. Two *Dioon* species, *D. edule* and *D. spinolusum*, were included in the analysis as an outgroup. Samples collected in this study are shown in Table 1, with the section for *Cycas* classification following that described by Hill (2004).

DNA was extracted using the CTAB method described by Doyle and Doyle (1990), with slight modification. Approximately, 0.3 g of young leaves was ground to fine powder in liquid nitrogen. One mL of preheated (65°C) CTAB isolation buffer (4% CTAB, 2.8 M NaCl, 20 mM EDTA, 200 mM Tris-HCl, pH 8.0, and 10 mM 2-mercaptoethanol) was added to the sample. The homogenate was incubated at 65°C for 1 h and then extracted using equal volumes of chloroform: isoamyl alcohol (24:1). The mixture was centrifuged at 10,000 g for 10 min at room temperature. The aqueous phase was collected and mixed with $^1/_5$ volume of 5X CTAB (5% CTAB and 0.7 M NaCl) and $^2/_3$ volume of isopropanol. The nucleic acid pellet was air-dried and resuspended in 100 µL TE buffer (10 mM Tris-HCl, 1mM EDTA, pH 7.0). RNaseA was added to the sample at the final concentration of 10 ng/µL. After incubating at 37°C for 30 min, the sample was extracted with phenol: chloroform (1:1). Absolute ethanol was added at equal volume of DNA to form a precipitate and centrifuged at 12,000 g for 10 min. Then, 70 % ethanol was added at equal volume to wash the DNA pellets. Finally, the DNA pellets were air-dried and resuspended in 30 µL TE buffer. The DNA concentration was measured on a 1% agarose gel using a UV spectrophotometer.

Polymerase chain reaction (PCR) amplification and DNA sequencing

The *trn*S^{GUC}-*trn*G ^{UCC} intergenic spacer region was amplified using primers *trn*S^{GUC} 5′ - GCCGCTTTAGTCCACTCAGC -3′ and *trn*G ^{UCC} 5′ - GAACGAATCACACTTTTACCAC -3′ (Hamilton, 1999), while that of *psb*M-*trn*D was amplified using primers *psb*M 5′ - AGCAAT AAATGCRAGAATATTTACTTCCAT -3′ and *trn*D 5′ - GGGATTGTAGYTCAATTGGT -3′ as carried out by Shaw *et al*. (2005). As for the *trn*L^{UAA}-*trn*F^{GAA} intergenic spacer region, it was amplified using primers “c” and “f” for the *trn*L intron and the *trn*L-*trn*F intergenic spacer. The

specific primer "c" was *trn*L^{UAA} 5′ - CGAAATCG TAGACGCTACG -3′ and primer "f" was *trn*F^{GAA} 5′ - ATTTGAACTGGTGACACGAG -3′ as previously used by Taberlet *et al.* (1991).

PCR amplification was performed in 50 µL of reaction mixture consisting of 5µL of 10X PCR buffer (TaKaRa), 2.5 mM $MgCl_2$, 200 µM dNTP, 1.25 unit *taq* DNA polymerase (TaKaRa), 10 µM of each primer and 2 µL of genomic DNA (20-30 ng). Amplifications were undertaken on a Perkin Elmer 9700 Thermo Cycler with an initial denaturation of 5 min at 94°C followed by 35 cycles each with 30 s at 94°C, 45 s at 52°C (for the *psb*M-*trn*D), 55°C (for the *trn*L-*trn*F) and 60°C (for the *trn*S–*trn*G), 1.30 min at 72°C and a final extension of 7 min at 72°C. PCR products were fractionated in 1% agarose gel and DNA bands were visualized using ethidium bromide staining. The PCR products were purified using the ExoSAP-IT kit (United States Biochemical) prior to sequencing with the ABI Big Dye Terminator Cycle Sequencing Kit V3.1 using a 3130X/ Genetic Analyzer.

Data analysis

Sequences were edited and assembled using the program ATGC var. 4 (GENETYX Co., Tokyo, Japan). The assembled contigs of species for the combined data set were initially aligned using Clustal X multiple sequence alignment software (Thompson *et al.*, 1997). The data were imported to GENEDOC 2.6 (Nicholas *et al.*, 1997) and manually adjusted. All DNA sequences have been deposited in the Genbank database (Table 1).

Before the analysis of combined data, the congruence between different data sets was tested by performing incongruence length difference tests (ILD) to evaluate the conflict that can occur between sets of characters from different data sources (Farris *et al.*, 1994). PAUP software was then used for ILD tests using 1,000 replicates and 99,999 taxon additions.

The Kimura 2-parameter (Kimura, 1980) method implemented in the software program MEGA 4 (Tamura *et al.*, 2007) was used to calculate the transitions:transversions ratio (Table 2) and the number of base substitutions per site (Tables 3 and 4). All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (pairwise deletion option).

Maximum likelihood (ML) and Maximum parsimony (MP) were constructed using PAUP 4.0 beta10 (Swofford, 2002). MP was conducted using heuristic search methods under the equal weighted criteria with a tree bisection reconnection (TBR) branch-swapping algorithm and 100 replicates of random taxon addition with gaps treated as missing data. Bootstrap analysis of 1,000 replicates using a heuristic search under equal weighted criteria was also used to assess the internal support for clades. ML and Neighbor-joining (NJ) were performed by the best evolutionary model and parameter values estimated by the hierarchical likelihood ratio tests (hLRTs) determination using the MODELTEST 3.7 program (Posada and Crandall, 1998). MODELTEST indicated that the best fitting model of evolution for the combined three non-coding chloroplast DNA regions was K81uf+G (also called K3P) with a base frequency of A =0.2975, C=0.1852, G= 0.1939, T=0.3234**;** base substitution of A-C=1.000, A-G=2.4072, A-T=0.5472, C-G= 0.5472, C-T=2.4072, G-T=1.0000 and gamma distribution shape parameter = 0.4871. ML heuristic searches used 10 random addition sequences and TBR swapping. Support for braches in the ML trees was tested by bootstrap analysis of 1,000 replicates. NJ was constructed using the MEGA 4 program. Clustering of species of each tree was confirmed with a bootstrap value of 10,000 replicates.

Table 1 Sections and species of *Cycas* and GenBank accession numbers of nucleotide sequences of the three regions.

Taxa	*trn*S-*trn*G	*psb*M-*trn*D	*trn*L-*trn*F
Sect. *Asiorientales*			
Cycas revoluta Thunb.	AB434465	GQ273612	GQ273656
Cycas taitungensis C.F. Shen, K.D. Hill, C.H. Tsou & C.J. Chen	GU250456	GU250479	GU250502
Sect. *Wadeanae*			
Cycas curranii (J. Schust.) K.D. Hill	GU250457	GU250480	GU250503
Cycas wadei Merrill	GU250458	GU250481	GU250504
Sect. *Indosinenses*			
Cycas chamaoensis K.D. Hill	GU250459	GU250482	GU250505
Cycas clivicola K.D. Hill	GU250460	GU250483	GU250506
Cycas condaoensis K.D. Hill, Hiäp & S.L.Yang	GU250461	GU250484	GU250507
Cycas elephantipes A. Lindstrˆm & K.D. Hill	GU250462	GU250485	GU250508
Cycas lindstromii S.L.Yang, K.D. Hill & Hiäp	GU250463	GU250486	GU250509
Cycas nongnoochiae K.D. Hill	GU250464	GU250487	GU250510
Cycas siamensis Miq.	GU250465	GU250488	GU250511
Cycas tansachana K.D.Hill & S.L.Yang	GU250466	GU250489	GU250512
Sect. *Cycas*			
Cycas edentata de Laub.	GU250467	GU250490	GU250513
Cycas pranburiensis S.L. Yang, W. Tang, K.D. Hill & Vatcharakorn	GU250468	GU250491	GU250514
Cycas rumphii Miq.	GU250469	GU250492	GU250515
Cycas seemannii A. Braun	GU250470	GU250493	GU250516
Cycas thouarsii R. Br. Ex Gaudich.	GU250471	GU250494	GU250517
Cycas zeylanica (Schuster) K D. Hill & A. Lindström	GU250472	GU250495	GU250518
Sect. *Stangerioides*			
Cycas balansae Warb.	GU250473	GU250496	GU250519
Cycas changjiangensis N. Liu	GU250474	GU250497	GU250520
Cycas hainanensis C.J. Chen	GU250475	GU250498	GU250521
Cycas hoabinhensis K.D. Hill, Hiäp & P.K. Loc	GU250476	GU250499	GU250522
Cycas segmentifida D. Yue Wang & C.Y. Deng	GU250477	GU250500	GU250523
Cycas sexseminifera F.N. Wei	GU250478	GU250501	GU250524
Outgroup			
Dioon edule Lindl.	AB434427	GQ273609	GQ273655
Dioon spinulosum Dyer	AB434428	GQ273610	GQ273654

Table 2 Characterization of non-coding region of chloroplast DNA in *Cycas* taxa.

	*trn*S-*trn*G	*psb*M-*trn*D	*trn*L-*trn*F	Combined data
No. of taxa	26	26	26	26
Aligned sequences length	960	1001	976	2896
No. variation sites	210	182	187	599
No. of parsimony informative characters	151	146	111	436
% ofparsimony informative characters	15.72	14.58	11.37	15.05
A+T%	63.2	62.1	62.7	62.7
G+C%	36.8	37.9	37.3	37.3
Transition/Transversion	0.814	1.322	1.296	1.346
Nucleotide diversity	0.03181	0.02660	0.02432	0.02930
Average number of substitutions per site	0.0381	0.0304	0.0294	0.0352

Table 3 Number of base substitutions per site averaged from combined data sequences pairs. All results are based on the pairwise analysis of five sections.

Section	*Stangerioides*	*Asiorientaless*	*Wadeanae*	*Cycas*	*Indosinenses*
Stangerioides					
Asiorientales	0.0262				
Wadeanae	0.0068	0.0264			
Cycas	0.0085	0.0280	0.0076		
Indosinenses	0.0091	0.0288	0.0082	0.0050	

Table 4 Estimates of average evolutionary divergence over sequence pairs within sections.

Section	
Stangerioides	0.0075
Asiorientales	0.0036
Wadeanae	0.0033
Cycas	0.0034
Indosinenses	0.0051

RESULTS

Sequence characteristics

The length of the *trn*S-*trn*G non-coding region varied from 896 bp (*C. wadei*) to 931 bp (*C. nongnoochiae*) with an average of 918 bp. The sequence alignment contained 960 bp, of which 210 were variable and 151 (15.72%) were parsimony informative sites. The GC content was 36.8% and AT was 63.2%. In addition, 14 bp insertions, GTAAGAATCTATAA, were observed at position 212-225 in the section *Indosinenses*, whereas the section *Cycas* (*C.edentata*, *C. seemannii, C. thouarsii, C.zeylanica* and *C. rumphii*) showed ATAAGAATCTATAA at the same position. Interestingly, the difference in the insertions between the sections *Indosinenses* and *Cycas* was found in only one base transition, namely, G⇔A at position 212, except for *C. pranburiensis*, whose sequence was exactly the same as those of the section *Indosinenses***.** The *trn*S-*trn*G non-coding sequences of the sections *Cycas* and *Indosinenses* carried 14 bp repeats (TCTATAAGTAAGAA) (Figure 4). The section *Asiorientales* (*C. taitungensis* and *C. revolute*), on the other hand, shared 5 bp repeats (TTTAC) and had 5 bp insertions at positions 161-165. The *trn*S-*trn*G sequence of this section showed 42 positions

of base substitution, 2 deletions and 1 insertion, which indicated the high genetic diversity in the section *Asiorientales*.

The *psb*M-*trn*D intergenic spacer sequence was in the range 978 to 984 bp for the ingroup of the *Cycas* species. The sequence alignment was 1,001 bp, of which 182 were variable sites, where 146 (14.58%) were parsimony informative. This spacer was AT rich (62.1%), having GC content only at 37.9% with several base substitutions in this region. The average length of the *trn*L-*trn*F intergenic spacer was 896 bp. The *trn*L-*trn*F alignment was 976 bp, of which 187 were variation sites and only 111 (11.37%) were parsimony informative. This sequence contained 62.7% AT and 37.3% GC.

The ILD test revealed that the data from the three regions (*trn*S-*trn*G, *psb*M-*trn*D and *trn*L-*trn*F) were not significantly different (P= 0.5620). Therefore, the three non-coding regions were combined, with an alignment length of 2,896 bp, of which 599 were variable and 436 (15.05%) were parsimony informative. Three regions (*trn*S-*trn*G, *psb*M-*trn*D and *trn*L-*trn*F) of the 24 *Cycas* species showed 11 positions of base substitution (Table 5), which distinguished the *Cycas* into two groups.

Table 5 Base substitution sites in three non-coding regions of chloroplast DNA of 24 *Cycas* species.

Species	*trn*S-*trn*G (position)					*psb*M-*trn*D (position)			*trn*L-*trn*F (position)		
	1 6 9	2 1 2	2 9 1	4 6 2	5 2 4	6 5 4	6 8 7	9 6 5	2 8 6	2 9 1	3 9 2
Cycas chamaoensis	T	G	G	G	T	C	C	A	A	T	A
Cycas clivicola	T	G	G	G	T	C	C	A	A	T	A
Cycas condaoensis	T	G	G	G	T	C	C	A	A	T	A
Cycas elephantipes	T	G	G	G	T	C	C	A	A	T	A
Cycas lindstromii	T	G	G	G	T	C	C	A	A	T	A
Cycas nongnoochiae	T	G	G	G	T	C	C	A	A	T	A
Cycas siamensis	T	G	G	G	T	C	C	A	A	T	A
Cycas tansachana	T	G	G	G	T	C	C	A	A	T	A
Cycas pranburiensis	T	G	G	G	T	C	C	A	A	T	A
Cycas edentata	T	A	A	A	T	C	C	A	C	G	A
Cycas zeylanica	T	A	A	A	T	C	C	A	C	G	A
Cycas thouarsii	T	A	A	A	T	C	C	A	C	G	A
Cycas seemannii	T	A	A	A	T	C	C	A	A	T	A
Cycas rumphii	T	A	A	A	T	C	C	A	A	T	G
Cycas balansae	G	-	G	G	C	T	T	G	A	T	G
Cycas changjiangensis	G	-	G	G	C	T	T	G	A	T	G
Cycas hainanensis	G	-	G	G	C	T	T	G	A	T	G
Cycas hoabinhensis	G	-	G	G	C	T	T	G	A	T	G
Cycas segmentifida	G	-	G	G	C	T	T	G	A	T	G
Cycas sexseminifera	G	-	G	G	C	T	T	G	A	T	G
Cycas revoluta	G	-	G	G	C	T	T	G	A	T	G
Cycas taitungensis	G	-	G	G	C	T	T	G	A	T	G
Cycas curranii	G	-	G	G	C	T	T	G	A	T	G
Cycas wadei	G	-	G	G	C	T	T	G	A	T	G

The first group was the sections *Cycas* and *Indosinenses*, while the sections *Asiorientales*, *Wadeanae* and *Stangerioides* formed several groups.

Phylogenetic analyses

NJ analysis was performed using the combined data with the K81uf+G method and the gamma shape parameter equal to 0.4871 (Figure 1). The NJ tree could be separated into several clades. The first clade contained the sections *Indosinenses* and *Cycas*, while the other clades comprised the sections *Stangerioides*, *Asiorientales* and *Wadeanae*. The grouping of the first clade indicated that *C. edentata*, *C. thouarsii* and *C. zeylanica* were closely related with 67% bootstrap support (BP), while *C. rumphii* and *C. seemannii* were a sister group. The remaining nine species; *C. clivicola, C. chamaoensis, C. nongnoochiea*, *C. lindstromii, C.elephantipes, C.*

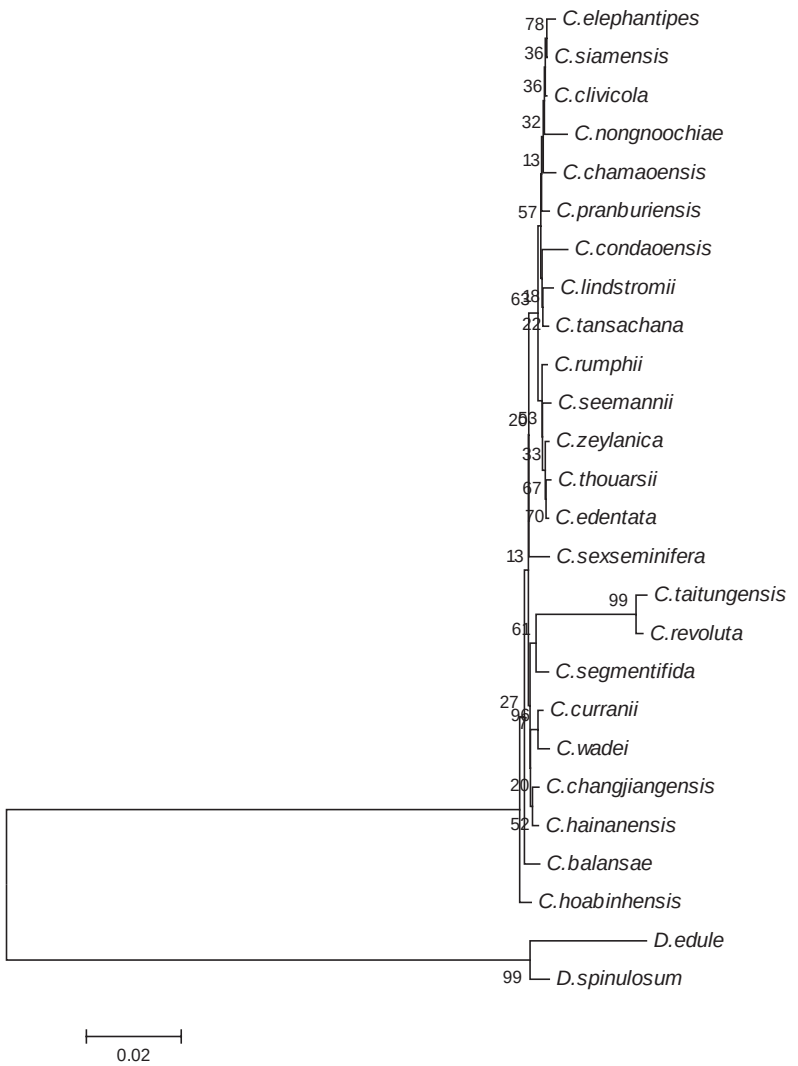

Figure 1 NJ tree obtained from MEGA for combined data set of *trn*S-*trn*G, *psb*M-*trn*D and *trn*L-*trn*F non-coding regions. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test is shown next to the branches.

siamensis, C. condaoensis, *C. tansachana* and *C. pranburiensis* were grouped together. Within this sub-clade, only *C. elephantipes* was grouped with *C. siamensis* with 78% bootstrap support. In the remaining clades, the high bootstrap support showed that *C. curranii* was closely related to *C. wadei* based on morphological taxonomy. On the other hand, *C. revoluta* and *C. taitungensis* were placed in the same clade with 99% bootstrap support, while *C. segmentifida* was a sister group. Although the bootstrap support levels for *C. changjiangensis* and *C. hainanensis* were slightly lower (52%), they were placed in the same section of *Stangerioides*. The number of nucleotide substitutions per site among the five sections of *Cycas* varied from 0.0050 to 0.0288, with an average of 0.0352 (Table 3). The number of nucleotide substitutions per site between the sections *Asiorientales* (*C. revoluta* and *C. taitungensis*) and other sections showed that the section *Asiorientales* contained the highest variation sequences (Table 3). In comparison to other sections, section *Stangerioides* was found to have the highest sequence divergence (Table 4).

Maximum parsimony (MP) analysis resulted in 763 trees with equal parsimony, 789 steps long with a consistency index (CI) of 0.8897and a retention index (RI) of 0.8540. The strict consensus tree based on the combined data is shown in Figure 2. Maximum likelihood (ML) (Figure 3) resulted in a tree with score -ln 7826.18 that generated consistent topology with NJ and MP trees. The MP and ML trees showed that the section *Indosinenses* and the section *Cycas* formed a clade with 85% bootstrap support in MP and 71 % in ML, which made these two sections a polyphyletic group. In contrast, the section *Wadeanae* (*C. curranii* and *C.wadei*) was a monophyletic group and gave high bootstrap support in both MP and ML**.** Furthermore, the high bootstrap (100%) suggested a close relationship between *C. revoluta* and *C. taitungensis*. The difference between MP and ML was that *C. changjiangensis* formed a clade with *C. hainanensis* (51% bootstrap support) in MP. Within the *Indosinenses-Cycas* clade, the relationships of the subsection *Rumphii* were resolved. One sub-clade comprised of *C. edentata* clustered with *C. zeylanica* and *C. thouarsii*, while *C. rumphii* and *C. seemannii* were a sister group.

DISCUSSION

Sequence variability in the family Cycadaceae

The results from the present study indicated that the utility of the *trn*S-*trn*G region resolved the phylogenetic relationship in some *Cycas* sections and provided more parsimony informative characters than using the *psb*M-*trn*D and *trn*L-*trn*F regions. These results agreed with Shaw *et al.* (2005), who evaluated the relative level of variability among 21 non-coding chloroplast DNA regions in seed plants and divided them into three tiers. The *trn*S-*trn*G region (tier 1) provided the greatest number of potential informative characters (PICs). The *psb*M-*trn*D region (tier 2), on the other hand, was identified as potentially useful, while the *trn*L-*trn*F (tier 3) provided the fewest PICs. Moreover, several earlier studies suggested that the *trn*S-*trn*G region showed high divergence and variability to resolve relationships among the closely related taxa (Small *et al*., 2005).

The *trn*S-*trn*G region of the section *Indosinenses* and the section *Cycas* contained 14 bp repeats, which revealed high levels of polymorphism in the two sections. This result strongly indicated that the two sections were closely related and implied that they may have evolved from a common ancestor. Furthermore, the results showed that the *trn*S-*trn*G region contained 5 bp repeats in *C. taitungensis* and *C. revoluta* and they were placed in the same clade, which is consistent with the morphological classification. Previous studies in other plants also showed that the chloroplast genome contained

tandem repeat sequences, which provide polymorphic DNA markers and can be used for studies in plant ecology and evolution (Provan *et al*., 2001; Matsuoka *et al*., 2002).

Phylogeny relationships within family Cycadaceae

The NJ, MP and ML analysis using the combined data set of 24 *Cycas* species produced similar tree topologies. These results confirmed that *Cycas* was monotypic. The three analyses showed clearly that the sections *Cycas* and *Indosinenses* were placed in the same clade, while the other clades contained *Asiorientales*, *Wadeanae* and *Stangerioides*.

This result was incongruent with the earlier classification based on morphology (De Laubenfels and Adema, 1998; Hill, 2004). The *Cycas-Indosinenses* clade showed that *C. edentata*, *C. seemannii* and *C. thouarsii* were grouped together, while *C. rumphii* and *C. zeylanica* formed a sister group based on base pair substitutions and

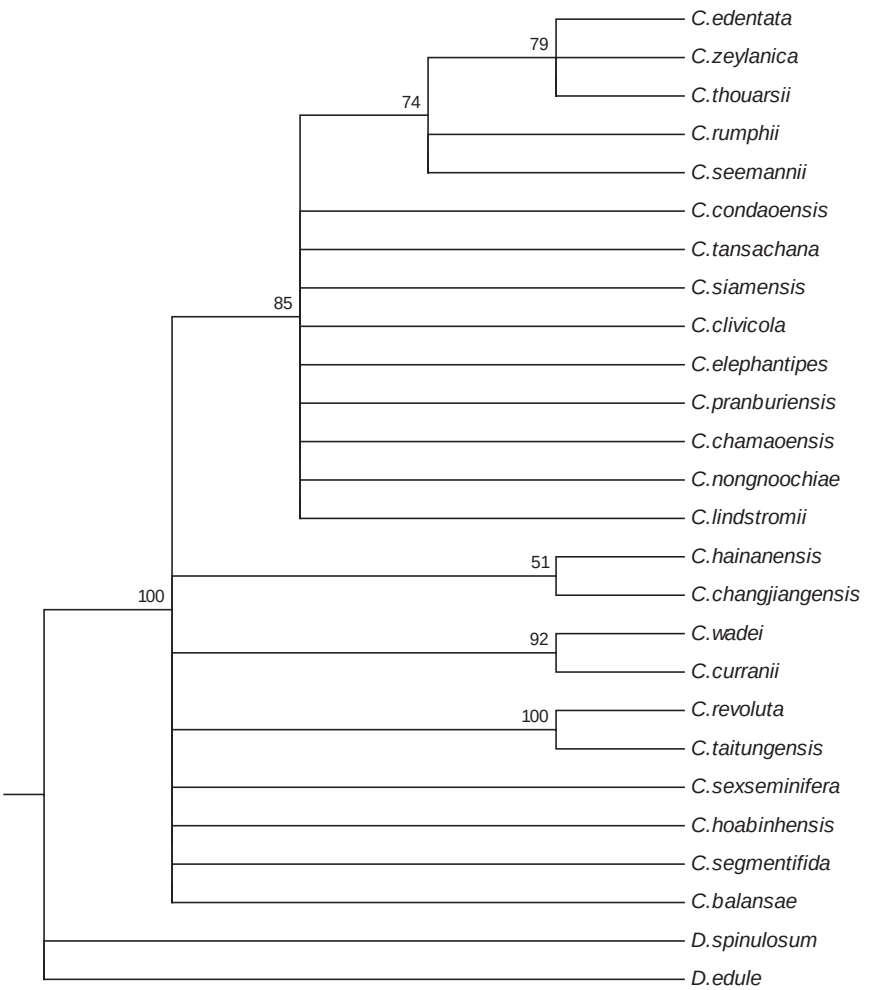

Figure 2 MP strict consensus tree obtained from PAUP for combined data set of *trn*S- *trn*G, *psb*M-*trn*D and *trn*L-*trn*F non-coding regions (CI = 0.8897and RI = 0.8540). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (>50%) is shown next to the branches.

they belonged to the same subsection *Rumphiae* of the section *Cycas*. This was in agreement with the study by Keppel *et al*. (2008), who investigated relationships within the subsection *Rumphiae* of the section *Cycas* using morphological and allozyme data. Although members of the subsection *Rumphiae* are widely distributed from Malaysia to East Africa and the Western Pacific Ocean (Hill, 1994), the current molecular study suggested that there was low genetic variation within the subsection *Rumphiae*. These results supported a long-distance seed dispersal mechanism, especially that which usually occurs in *Cycas*. It is known that *Cycas* species produce seeds with a special adaptation, a spongy layer in the seed, which give positive buoyancy in seawater (Dehgan and Yuen 1983) and hence, assists their wide distribution while maintaining the genetic composition.

In the NJ analysis, C. *elephantipes* and *C. siamensis* were grouped together, which was in agreement with the geographic distribution in Thailand (Hill and Yang, 1999; Lindstrom and Hill, 2002), as these two species belong to the section

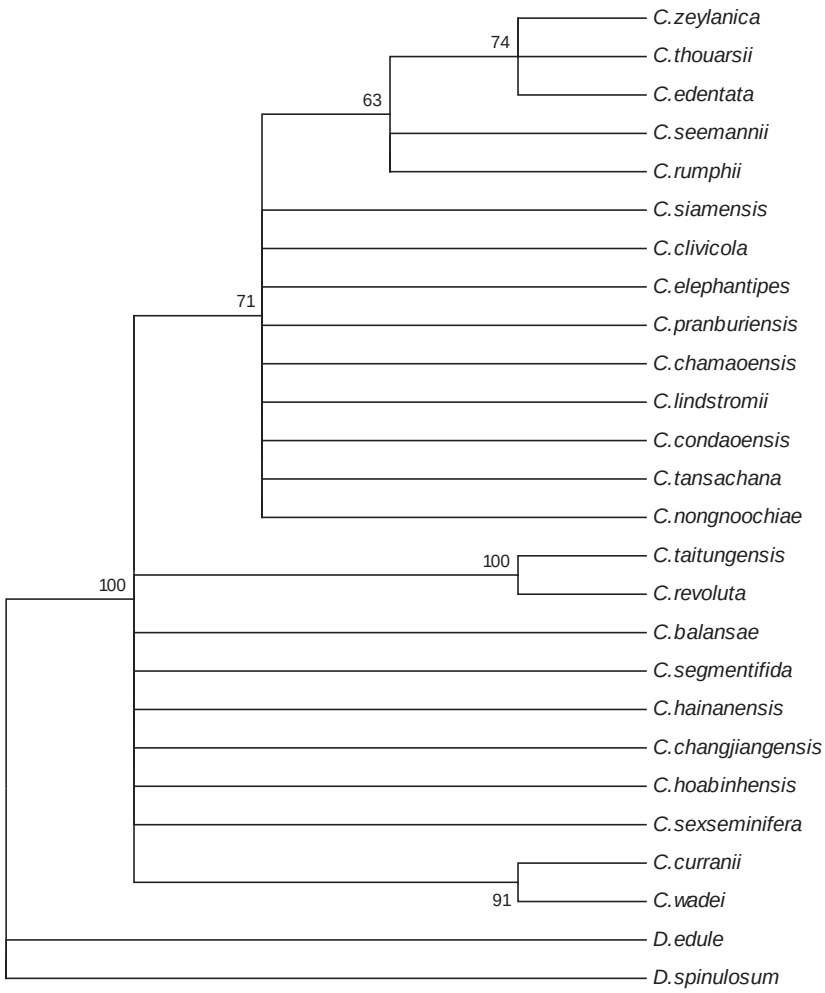

Figure 3 ML generated under the K81uf+G model using combined data. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (>50%) is shown next to the branches.

Figure 4 Partial alignment of chloroplast *trn*S-*trn*G region from 24 *Cycas* species. The boxed region show 14 bp repeated sequences in the sections *Cycas* and *Indosinenses* and 5 bp repeated sequences in the section *Asiorientales*.

Indosinenses. Furthermore, the current study found that the two species of the section *Wadeanae* (*C. wadei* and *C. curranii*) formed a clade that was consistent with morphological taxonomy and they are also endemic to the Philippines (Lindstrom *et al*., 2008). *C. revoluta* is found in the southern part of Kyushu, on Ryukyu Island in Japan and in Fukien province in China (Norstog and Nichols, 1997), while *C. taitungensis* is an endemic species to Taiwan (Shen *et al*., 1994). However, the high bootstrap support implied that *C. revoluta* and *C. taitungensis* were closely related and that they shared identical base substitutions and indel positions.

Within the section *Stangerioides*, *C. changjiangensis* and *C. hainanensis* were more closely related to each other than to any other species. Morphological data indicated that *C. changjiangensis* was similar to *C. hainanensis*, but *C. changjiangensis* has a different dwarf, largely subterranean habit, and a smaller megasporophyll lamina and smaller seed than *C. hainanensis*, with the two species being endemic to Hainan Island, China (Hill, 2008). The relationships among the Chinese species of *C. balansae*, C. *sexseminifera* and *C. segmentifida*, and the Vietnamese species, *C. hoabinhensis*, in the section *Stangerioides* were unresolved. However, the number of base substitutions per site within this section indicated high sequence divergence, which was consistent with the results shown by Pu and Chiu (1999) in their study of genetic variation in the section *Stangerioides* using anchored microsatellite primers, which showed high genetic variation in this section.

CONCLUSION

Sequence data from the three non-coding regions of chloroplast DNA (*trn*S-*trn*G, *psb*M-*trn*D and *trn*L-*trn*F) were used to elucidate phylogenetic relationships within the family Cycadaceae. The study provided the first valuable phylogeny of 24 *Cycas* species. The results showed clearly that the sections *Cycas* and *Indosinenses* were placed in the same clade, while the other clades contained *Asiorientales*, *Wadeanae* and *Stangerioides*. In addition, the data analyses suggested that *trn*S-*trn*G sequences were more informative than other regions in helping resolve relationships in *Cycas*. This work was the first report on the 14 bp repeats detection in *trn*S-*trn*G region. Although the *Cycas* species are a closely related family, non-coding chloroplast DNA sequences were able to resolve the relationships within *Cycas*.

ACKNOWLEDGEMENTS

The study was financially supported by the Bilateral Research Cooperation (BRC) Scholarship and the Science Thesis and Research Fund (ScTRF), Faculty of Science, Kasetsart University, Thailand and a project titled "Integration of Systematics and Molecular Phylogenetics in All Groups of Organisms" handled by the National Museum of Nature and Science, Japan. The authors would like to thank Mr Kampon Tansacha at the Nong Nooch Tropical Botanical Garden for providing *Cycas* specimens.

LITERATURE CITED

Dehgan, B. and C.K.K. Yuen. 1983. Seed morphology in relation to dispersal, evolution, and propagation of *Cycas* L. **Bot. Ga.** 144: 412-418.

De Laubenfels, D.J. and F. Adema. 1998. A taxonomic revision of the genera *Cycas* and *Epicycas* gen. nov. (Cycadaceae). **Blumea** 43: 351-400.

Doyle, J.J. and J.L. Doyle. 1990. Isolation of plant DNA from fresh tissue. **Focus** 12: 12-15.

Farris, J.S., M. Källersjö, A.G. Kluge and C. Bult. 1994. Testing significance of incongruence. **Cladistics** 10: 315-319.

Hamilton, M.B. 1999. Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. **Mol. Ecol.** 8: 521-523.

Hill, K.D. 1994. The *Cycas rumphii* complex (Cycadaceae) in New Guinea and the Western Pacific. **Aust. Syst. Bot.** 7: 543- 567.

_____ and S.L. Yang. 1999. The genus Cycas (Cycadaceae) in Thailand. **Brittonia** 51: 48-73.

_____ , D.W. Stevenson and R. Osborne. 2004. Appendix 1: The world list of cycads, pp. 219-235. *In* T. Walters and R. Osborne (eds.). **Cycad Classification Concepts and Recommendations.** CABI Pub, Wallingford.

_____ . 2004. Character evolution, species recognition and classification concepts in the Cycadaceae. pp. 23-44. *In* T. Walters and R. Osborne (eds.). **Cycad Classification Concepts and Recommendations.** CABI Pub, Wallingford.

_____ . 2008. The genus *Cycas* (Cycadaceae) in China. **Telopea** 12: 71-118.

Keppel, G., P.D. Hodgskiss and G.M. Plunkett. 2008. Cycads in the insular South-west Pacific: dispersal or vicariance? **J. Biogeogr.** 35: 1004-1015.

Kim, S.C., D.J. Crawford, R.K. Jansen and A. Santos-Guerra. 1999. The use of a non-coding region of chloroplast DNA in phylogenetic studies of the subtribe *Sonchinae* (*Asteraceae*: *Lactuceae*). **Plant Syst. Evol.** 215: 85-99.

Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. **J. Mol. Evol.** 16: 111-120.

Lindstrom, A.J. and K.D. Hill. 2002. New species and a new record of Cycas (Cycadaceae) from Thailand. **Britonia** 54: 298-304.

Lindstrom, A.J., K.D. Hilland and L.C. Stanberg. 2008. The genus *Cycas* (Cycadaceae) in the Philippines. **Telopea** 12: 119-145.

Matsuoka, Y., S.E. Mitchell, S. Kresovich, M. Goodman and J. Doebley. 2002. Microsatellite in Zea-variability, patterns of mutations, and use for evolutionary studies. **Theor. Appl. Genet.** 104: 436-450.

Mort, M.E., J.K. Archibald, C.P. Randle, N.D. Levsen, T.R. O'Leary, K. Topalov, C.M. Wiegand and D.J. Crawford. 2007. Inferring phylogeny at low taxonomic levels: utility of rapidly evolving cpDNA and nuclear ITS loci. **Am. J. Bot.** 94: 173-183.

Nicholas, K.B., H.B.J. Nicholas and D.W. Deerfield. 1997. GeneDoc: analysis and visualization of genetic variation. **EMBNEW NEWS** 4: 14.

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

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

Provan, W. Powell and P.M. Hollingsworth. 2001. Chloroplast microsatellites: new tools for studies in plant ecology and evolution. **Trends Ecol. Evol.** 16: 142-147.

Pu, H. and W.L. Chiu. 1999. Genetic variation in *Cycas* Section *Stangierioides* using anchored microsatellite primers, pp. 334-344. *In* C.J. Chen (ed.). **Biology and Conservation of Cycads. Proceedings of the Fourth International Conference on Cycad Biology.** International Academic Publishers, Beijing.

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. **Mol. Phylogenet. Evol.** 29: 350-359.

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

Shen, C.F., K.D. Hill, C.H. Tsou and C.J. Chen. 1994. *Cycas taitungensis* sp. nov. (Cycadaceae), a new name for the widely known cycad species endemic in Taiwan. **Bot. Bull. Acad. Sin.** 35: 133-140.

Small, R.L., E.B. Lickey, J. Shaw and W.D. Hauk. 2005. Amplification of noncoding chloroplast DNA for phylogenetic studies in lycophytes and monilophytes with a comparative example of relative phylogenetic utility from Ophioglossaceae. **Mol. Phylogenet. Evol.** 36: 509-522.

Smitinand, T. 1971. The genus *Cycas* Linn. (Cycadaceae) in Thailand. **Nat. Hist. Bull. Siam Soc.** 24: 163-175.

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

Swofford, D.L. 2002. **PAUP* Phylogenetic Analysis Using Parsimony (*and other methods), Version 4.0b10.** Sinauer associates, Sunderland, MA.

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

Tamura, K., J. Dudley, M. Nei and S. Kumar. 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software Version 4.0. **Mol. Biol. Evol.** 24: 1596-1599.

Thompson, J.D., T.J. Gibson, F. Plewniak, F. Jeanmougin and D.G. Higgins. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. **Nucleic Acids Res.** 24: 4876-4882.

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

Xu, D.H., J. Abe, M. Sakai, A. Kanazawa and Y. Shimamoto. 2000. Sequence variation of non-coding regions of chloroplast DNA of soybean and related wild species and its implications for the evolution of different chloroplast haplotypes. **Theor. Appl. Genet.** 101: 724-732.