THESIS APPROVAL
GRADUATE SCHOOL, KASETSART UNIVERSITY

Doctor of Philosophy (Bioscience)
DEGREE

Bioscience
FIELD

Interdisciplinary Graduate Program
PROGRAM

TITLE: Phylogenetic Analysis of Cycad and Sex Identification in *Cycas* using Molecular Markers

NAME: Miss Pattamon Sangin

THIS THESIS HAS BEEN ACCEPTED BY

(Associate Professor Mingkwan Mingmuang, Ph.D.)
THESIS ADVISOR

(Miss Chatuporn Kuleung, Ph.D.)
THESIS CO-ADVISOR

(Assistant Professor Vipa Hongtrakul, Ph.D.)
THESIS CO-ADVISOR

(Mr. Surapol Patharakorn, Ph.D.)
GRADUATE COMMITTEE CHAIRMAN

APPROVED BY THE GRADUATE SCHOOL ON .....

(Associate Professor Gunjana Theeragool, D.Agr.)
DEAN

Copyright by Kasetsart University All rights reserved

THESIS

PHYLOGENETIC ANALYSIS OF CYCAD AND SEX IDENTIFICATION IN *CYCAS* USING MOLECULAR MARKERS

PATTAMON SANGIN

A Thesis Submitted in Partial Fulfillment of
the Requirements for the Degree of
Doctor of Philosophy (Bioscience)
Graduate School, Kasetsart University
2010

Copyright by Kasetsart University All rights reserved

Pattamon Sangin 2010: Phylogenetic Analysis of Cycad and Sex Identification in *Cycas* using Molecular Markers. Doctor of Philosophy (Bioscience), Major Field: Bioscience, Interdisciplinary Graduate Program. Thesis Advisor: Associated Professor Mingkwan Mingmuang, Ph.D. 138 pages.

Phylogenetic relationships among 43 species within the family Zamiaceae and Stangeriaceae were examined using 4 non-coding regions of chloroplast DNA (*atp*B-*rbc*L, *psb*M-*trn*D, *trn*L-*trn*F and *trn*S- *trnf*M). Maximum likelihood (ML), Maximum parsimony (MP) and Neighbor- joining (NJ) analyses were performed on separated and combined data sets to generate phylogenetic trees. Zamiaceae and Stangeriaceae are not monophyletic entities based on these analyses with *Bowenia* and *Stangeria* embedded and separated within the former family. *Stangeria* was closely related with *Ceratozamia* in MP and NJ trees. *Dioon* was clearly showed as the most basal genus. *Encephalartos* and *Lepidozamia* were closer to each other than *Macrozamia* with these three genera forming a monophyletic group. Furthermore, in confirmation to recent publications, *Chigua* was found to be paraphyletic with *Zamia*. The *trn*L-*trn*F sequences were more informative than other regions in helping resolve relationships in the cycads. A new, simplified suprageneric classification for the extant cycads was presented whereby two families are recognized. The subfamily Dioonioideae was newly recognized. Further investigation on the infrageneric relationships among 27 species of five *Cycas* sections was made using the same methods. Three data sets were combined (*trn*S-*trn*G, *psb*M-*trn*D and *trn*L-*trn*F) because they did not differ significantly in structure. All three methods showed similar topology which divided *Cycas* into two main clades. The first clade consisted of two sections, *Cycas* and *Indosinenses*, while the other clade contained *Asiorientales*, *Wadeanae* and *Stangerioides*. Base substitution pattern, further revealed that subsection *Rumphiae* of the section *Cycas* could be separated into two groups (*C. rumphii* and *C. edentata*). In addition, *C. taitungensis* was closely related to *C. revotuta* while other species of the section *Wadeanae* (*C. wadei* and *C. curranii*) were closely related which agreed with the morphology. The *trn*S-*trn*G sequences were more informative than other regions in addressing phylogeny. Microsatellite and minisatellite detected in this region also indicated the high rate of evolution in *Cycas* species. The minisatellite identified in this study is the first report. RAPD and ISSR primers were tested on *Cycas* to identify sex-specific molecular markers. Sixty different RAPD primers and twenty-nine ISSR primers were screened out of which only two RAPD primers (OPB-8 and UBC 485) and three ISSR primers (001, 836 and 864) were found to be polymorphic. These sequences were converted into specific Sequence Characterized Amplified Region (SCAR) markers. However, none of SCAR markers showed sex specific fragment when testing using DNA of individual males and females *Cycas*.

/ /

Student's signature Thesis Advisor's signature

Copyright by Kasetsart University All rights reserved

ACKNOWLEDGEMENTS

I would like to express my deepest thanks to my advisor, Dr. Mingkwan Mingmuang for her kind, excellent supervision, valuable advice and great guidance throughout the period of the study.

I would also like to express appreciation to the other members of my committee, Dr. Chatuporn Kuleung and Dr.Vipa Hongtrakul, each member helped me define and focus my research. I would like to thank Dr. Amara Thongpan for her availability, valuable comments and partial correction of the thesis.

I am indebted to Dr. Goro Kokubugata and staff of Tsukuba Botanical Garden for their encouragement, kindness and helpful advices me during doing the research in Japan. I would like to thank Mr. Kampon Tansacha and Mr. Anders J. Lindstrom at Nong Nooch Tropical Botanical Garden for providing specimens.

Finally, I would like to express my deepest gratitude to my loved parents all member of my family for their love, care, understanding and encouragement extended throughout my study.

Pattamon Sangin

June, 2010

Copyright by Kasetsart University All rights reserved

TABLE OF CONTENTS

Page

Copyright by Kasetsart University All rights reserved

LIST OF TABLES

Copyright by Kasetsart University All rights reserved

LIST OF TABLES (Continued)

Copyright by Kasetsart University All rights reserved

LIST OF FIGURES

Copyright by Kasetsart University All rights reserved

LIST OF FIGURES (Continued)

Copyright by Kasetsart University All rights reserved

LIST OF FIGURES (Continued)

Copyright by Kasetsart University All rights reserved

LIST OF FIGURES (Continued)

Copyright by Kasetsart University All rights reserved

PHYLOGENETIC ANALYSIS OF CYCAD AND SEX IDENTIFICATION IN *CYCAS* USING MOLECULAR MARKERS

INTRODUCTION

Cycads are the most ancient now living of seed plants which their fossils record extents back to at least the early Paleozoic period, approximately 250 million years ago (MYA) (Zhifeng and Thomas, 1989; Schwendemann *et al*., 2009). They are a monophyletic group classified as a single order, Cycadales, which has been divided into two suborders with three families and approximately 300 species (Stevenson, 1992; Hill *et al*., 2004). The family Cycadaceae contains only a single genus - *Cycas*; the family Stangeriaceae consists of two genera, *Bowenia* and *Stangeria*; whereas the family Zamiaceae is the most diverse with the genera *Ceratozamia*, *Chigua, Dioon*, *Encephalartos*, *Lepidozamia*, *Macrozamia*, *Microcycas* and *Zamia* (Stevenson, 1992). However, the genus *Chigua* has recently been taxonomically merged with Zamia (Lindstrom, 2009). The relationship amongst these genera using only morphological characters remains unclear and there have been considerable differences of opinion as to the disposition of genera and within families. In particular, the continued recognition of the family Stangeriaceae as distinct from Zamiaceae (Stevenson, 1992; Norstog and Nicholls, 1997; Artabe *et al.*, 2005) can be viewed as controversial.

Recent phylogenetic studies have been based on various DNA regions of which chloroplast genes appear the most useful due to their relatively slow evolution in comparison with nuclear genes (Curtis and Clegg, 1984). There have been a number of previous studies of phylogenetic relationships of the cycads based upon molecular evolution of chloroplast DNA regions (Treutlein and Wink, 2002; Hill *et al*., 2003; Rai *et al*., 2003; Bogler and Francisco-Ortega, 2004; Chaw *et al*., 2005; Zgurski *et al*., 2008). All of these phylogenetic analyses indicated that the positions of *Bowenia*, *Stangeria* and *Dioon* were incongruent with the earlier classification based on morphology (Stevenson, 1992). The differing placement of three genera

Copyright by Kasetsart University All rights reserved

within these various studies has meant that a stable phylogenetic classification for the inter-generic relationships of the cycads has remained elusive.

In the present study, the *atp*B-*rbc*L, *psb*M-*trn*D, *trn*L-*trn*F and *trn*S-*trnf*M non-coding sequences of the chloroplast genome were used for reconstruction of phylogenetic trees and compared the phylogenies inferred from each region and then combined data to resolve intra and inter-relationship within the family Stangeriaceae and Zamiaceae to investigate utility of selected markers. In addition, the phylogeny of 27 species of *Cycas* was inferred using data of *trn*S-*trn*G, *psb*M-*trn*D, *trn*L-*trn*F and *trn*S-*trnf*M non-coding regions to investigate genetic relationships of *Cycas* and compared the utility of four non-coding regions for resolving intrageneric relationship within *Cycas*.

Cycas, the only genus in the Cycadaceae family, is composed of 98 species. *Cycas* are dioecious plant, male and female reproductive structures are borne on separate plants and reproduction is by seeds. The male reproductive structure is a determinate cone, comprising a central cone axis bearing numerous microsporophylls spirally attached. Female *Cycas* do not produce a compact but bear successive and repeated zones of vegetative leaves, cataphylls and megasporophylls (Zhang *et al*., 2002). *Cycas* are now cultivated as an ornamental landscape tree. The demand for female plants has increased in recent years because they are commercially valued for production of seed. Therefore, cultivation of female plant in such cases is preferred to the respective male plants. Although male and female cones are generally dissimilar in shape and size, the sex of cycad cannot be distinguished at the juvenile stage. Most of cycads have a very long life cycle; the time from the germination of a seed until that seedling grows into a reproductive plant is unknown for most species and would appear to be a very long period (Stevenson, 1990). Moreover, no morphological markers have been found which consistently identifies the sex before the first cone appears. The inability to determine the sex in an early stage is as major problem for plant breeding and plant improvement.

Copyright by Kasetsart University All rights reserved

Modern DNA technologies provide a variety of techniques to produce sex-linked molecular markers. Random Amplified Polymorphic DNA (RAPD) is a rapid and inexpensive method for studies the genotypic relationships and selection of traits of interest (Williams *et al*., 1990). Recently, a novel molecular technique for sex determination is Inter Simple Sequence Repeats (ISSR) which is semiarbitrary markers amplified by PCR in the presence of one primer complementary to a target microsatellite (Zietkiewicz *et al*., 1994; Danilova and Korlov, 2006). Moreover, they are widely used to identify the sex of several plants (Prakash and Staden, 2006; Hormanza *et al*.,1994; Urasaki *et al*.,2002; Di Stilio *et al*.,1998). However, these techniques can be difficult to reproduce because of its high sensitivity to reaction conditions. Therefore, sequence characterized amplified regions (SCAR) markers were developed by sequencing a single RAPD or ISSR bands and designing primers (approximately 20 bases) to amplify the band of specific size. By increasing the specificity, the results are less sensitive to changes in reaction conditions and are more reproducible. Therefore, attempts to identify the sex of *Cycas* before the seeding stage using these molecular marker techniques are also made.

Copyright by Kasetsart University All rights reserved

OBJECTIVES

1. To reconstruct phylogenetic trees and determine phylogenetic relationships within the family Stangeriaceae and Zamiaceae.

2. To evaluate the utility of four non-coding regions of chloroplast DNA (*atp*B-*rbc*L, *psb*M-*trn*D, *trn*L-*trn*F and *trn*S-*trnf*M) at the intergeneric level in the family Stangeriaceae and Zamiaceae.

3. To reconstruct phylogenetic trees and investigate phylogenetic relationships within the genus *Cycas*.

4. To compare the utility of four non-coding regions of chloroplast DNA (*trn*S-*trn*G, *psb*M-*trn*D, *trn*L-*trn*F and *trn*S- *trnf*M) for resolving intrageneric relationship within *Cycas*.

5. To develop molecular markers for identification of males and females *Cycas*.

Copyright by Kasetsart University All rights reserved

LITERATURE REVIEW

Cycads are ancient plants with a long continuity line of heredity. They appeared in the Pennsylvanian and have existed for approximately 250 million years. Cycads are collectively and commonly referred to as gymnosperm (Cycads, Ginkgo, Conifer, Gnetales). They are classified into 3 families, 11 genera (Stevenson, 1992) and 300 known species (Hill *et al*., 2004). Cycads are found in the tropical, subtropical and warm temperate regions of both the north and south hemisphere. Although they distribute themselves all over the world but are mainly found along the intertropical belt, i.e., Africa, India, Indonesia, and North Australia.

All cycad species were classified in to a single family, the Cycadaceae. Later they were separated into three families, Cycadaceae, Stangeriaceae and Zamiaceae (Johnson, 1959). Although in 1981 Stevenson created a fourth family, Boweniaceae, composed of the genus Bowenia, but further study placed this genus in subfamily Bowenioideae of the family Stangeriaceae. However, Stevenson (1992) has carried out a series of detail cladistic analyses on the cycads, based on the analysis of fifty-two characters resulted in a single cladogram. That classified is as follows.

Order Cycadales

Suborder Cycadineae

Family Cycadaceae

Genus *Cycas*

Suborder Zamiainea

Family Stangeriaceae

Subfamily Stangeriodideae

Genus *Stangeria*

Subfamily Bowenioideae

Genus *Bowenia*

Family Zamiaceae

Copyright by Kasetsart University All rights reserved

Subfamily Encephalartoideae
Tribe Diooeae
Genus *Dioon*
Tribe Encepharateae
Subtribe Encephalartinae
Genus *Encephalartos*
Subtribe Macrozamiinae
Genus *Macrozamia, Lepidozamia*
Subfamily Zamioideae
Tribe Ceratozamieae
Genus *Ceratozamia*
Tribe Zamieae
Subtribe Microcycadinae
Genus *Microcycas*
Subtribe Zamiinae
Genus *Zamia, Chigua*

1. Plant chloroplast DNA

About 50% of the chloroplast DNA ranging in size from 120 to 160 kb belongs to the protein-coding genes. The significant feature of the chloroplast DNA in most plants is the presence of a large inverted repeat (IR) which is separated by one large single-copy (LSC) and small single-copy (SSC) segments. In accordance with the chloroplast DNA pattern of *Cycas*, the complete chloroplast DNA of *Cycas* is 163,403 bp composed of 2 typical large inverted repeats (IR_A and IR_B) separated by LSC and SSC regions (Wu *et al*., 2007). The complete chloroplast genome of *Citrus sinensis* contains a pair of IRs and separated by a small and a large single-copy (Bausher *et al*., 2006). Like most of green algal chloroplast DNA, *Mesostigma*, *Chaetosphaeridium glabosum*, *Oltmannsiellopsis viridis* and *Nephroselmis olivacea* represent the same pattern of cpDNA (Turmel *et al*., 1999; Turmel *et al*., 2005 and Pombert *et al*., 2006). The chloroplast DNA of pea, alfalfa and pine do not show this pattern and lack IR. In addition, the chloroplast DNA of *Euglena gracilis* contains 3

Copyright by Kasetsart University All rights reserved

tandem repeats and rDNA gene cluster (Hallick *et al*., 1993). Therefore, the three group of chloroplast DNA are classified into chloroplast DNA lacking IRs (groupI), chloroplast DNA containing IRs (groupII) and chloroplast DNA having tandem repeats (groupIII).

2. Phylogenetic analysis using non-coding chloroplast DNA

In recent years, the most phylogenetic studies have successfully used chloroplast DNA for evolutionary relationships and molecular systematic of plants because chloroplast genes evolve slowly in comparison with nuclear gene (Curtis and Clegg, 1984). The conservation of chloroplast DNA suggests that any change in structure, arrangement, or content of chloroplast genome may have significant phylogenetic implication (Downie and Palmer, 1991). Several studies focused on protein-coding gene sequences such as *rbc*L, *ndh*B, *mat*K and *trn*L for investigation phylogenetic relationships among higher-level taxa (Suyama *et al*., 2000; Xiang *et al*., 2002 and Selvaraj *et al*., 2008). However, slow evolution of the coding regions of chloroplast DNA cannot resolve relationship among lower-level taxa but the variation of non-coding regions is higher than the coding regions and provides more informative characters in phylogenetic studies (Xu *et al*., 2000).

2.1 *atp*B-*rbc*L region of chloroplast DNA

The non-coding region between *atp*B and *rbc*L genes is located in the large single-copy region. This region has been used in phylogenetic studies within some plants. Universal primers for amplifying and sequencing the *atp*B-*rbc*L region were designed by comparing sequences of liverwort, fern, tobacco and rice. The *atp*B-*rbc*L primers were used to amplify genomic DNA from two mosses, one fern, one gymnosperm, three monocots and three dicots and the results indicated that the evolution was increased from liverwort, through mosses, to vascular plants. These sequences showed variation of chloroplast DNA (base substitutions, insertion, deletion and length mutation) (Chiang *et al*., 1998). In addition, Chiang and Schall

Copyright by Kasetsart University All rights reserved

(2000) found the nucleotide variation of the *atp*B-*rbc*L region and used it to estimate phylogeny of 11 species of true mosses. The sequences were highly conserved within mosses and between mosses and liverwort.

The sequence of the *atp*B-*rbc*L region and *rbc*L gene in 15 species of tribe Rubieae (Rubiaceae) were phylogenetically reconstructed and compared the evolution between two regions. The results showed that the rate of nucleotide substitution at synonymous sites of *rbc*L was higher than the *atp*B-*rbc*L region. It could be explained that *rbc*L sites were changing at higher rate in these plants and the *atp*B-*rbc*L region contained several unknown sequences but important regulatory elements. However, the *atp*B-*rbc*L phylogenetic tree was more congruent with morphological data than the *rbc*L (Manen and Natali, 1995). Huang *et al*. (2005) successfully used this region to study phylogeography and genetic conservation of *Hygraphila pogonocalyx* (Acanthaceae) and two geographical groups of western and eastern were identified in NJ tree and the minimum-spanning network. These two groups may be separated from each other since the formation of the Central Mountain Range about 5 million year ago which was consistent with the rate estimates based on molecular clock of chloroplast DNA. Furthermore, sequence variations of the *atp*B-*rbc*L region were employed to examine the population genetic structure and phylogeographical pattern of eight relict *Alsophila podophylla* populations (Su *et al*., 2005). The phylogeographic pattern of *Cycas taitugensis* was also investigated base on the *atp*B-*rbc*L region, ribosomal DNA (rDNA) and internal transcribed spacer (ITS) of mitochondria (Huang *et al*., 2001).

2.2 The *trn*S-*trn*G region

The *trn*S-*trn*G is located in the large single-copy. Several studies showed that this region is highly variable. Xu *et al*. (2000) showed that the *trn*S-*trn*G region of *Glycine* provided highly informative among nine non-coding chloroplast DNA regions within two closely related subgenera of *Glycine* (*Glycine* and *Soja*). Shaw *et al*. (2005) evaluated the relative level of variability among 21 non-coding chloroplast

Copyright by Kasetsart University All rights reserved

DNA regions in seed plants and divided them into three tiers. The *trn*S-*trn*G region (tier 1) provided the greatest numbers of Potentially Informative Character (PICs). Moreover, the phylogenetic relationships of disjunctly specimens of *Tristicha trifaria* were analyzed using *mat*K, *rbc*L, *trn*K intron, *trn*S-*trn*G, *trn*T-*trn*L-*trn*F, *trn*L intron and nuclear ribosomal ITS regions which also showed that the *trn*S-*trn*G region was the most extensive sequence divergence. In addition, Yuan and Olmstead (2008) used *trn*S-*trn*G region combined with *trn*G, *trn*T-*trn*L, *trn*L, *trn*L-*trn*F, *trn*S-*trn*fM and *trn*D-*trn*T to study intergeneric chloroplast transfers of *Verbena* complex.

2.3 The *trn*L-*trn*F region

This region is located in the large single-copy region of chloroplast genome and is composed of the *trn*L intron and the *trn*L-*trn*F intergenic spacer (IGS). It is usually used in phylogeny studies of closely related species. Sequences from the *trn*L-*trn*F elucidated the phylogenetic relationships in the tribe Diseae (Orchidoideae: Orchidaceae) and the results showed that the utility of the *trn*L-*trn*F region was used to resolve the phylogenetic relationships within the Disinae (Bellsted *et al*., 2001). Hao *et al*. (2008) investigated the evolution of the *trn*L-*trn*F region from all taxa of Taxaceae and Cephalotaxaceae which was variable and contained tandem repeats. Small *et al*. (2005) screened 30 lycophyte and monilophyte species to determine the potential utility of PCR amplification primers for 18 non-coding chloroplast DNA regions as the *trn*L-*trn*F intergenic spacer provided the highest percentage of variable and parsimony-informative sites. In addition, the *trn*L-*trn*F and *rbc*L regions were used to calculate molecular-based age estimates of *Equisetum* which are an ancient lineage of seed-free vascular plant. These molecular-base age estimates agreed with the fossil record (Des Marais *et al*., 2003). The interspecific relationships in Osmorhiza (Apiaceae: Apioidae) were evaluated using ITS region, *ndh*F gene and *trn*F-*trn*L region. The *trn*F-*trn*L sequences exhibited a wide range of size from 878 to 960 bp among species of *Osmorhiza* and contained a large deletion. This region provided fast nucleotide substitution rates and successfully employed at interspecific level (Yoo *et al*., 2002). Stoneberg Holt *et al*. (2004) found the insertion/deletion (indel) patterns from sequences of the *trn*F-*trn*L region in the genus *Pao* (Poaceae).

Copyright by Kasetsart University All rights reserved

Indels provided a meaningful infrageneric classification criterion for *Pao* and has proven to be useful tools in examining relationships within this genus. Although the *trn*L-*trn*F region presented a few PICs by Shaw *et al*. (2005), these studies indicated that the *trn*F-*trn*L region was more informative in helping resolve relationships in intra and interspecific taxa.

2.4 The *psb*M-*trn*D region

The *psb*M-*trn*D region is located in the large single-copy region. Shaw *et al*. (2005) indicated that the *psb*M-*trn*D region was identified as potentially useful to elucidated phylogenetic relationships. Lee and Wen (2004) amplified the *psb*M-*trn*D region along with *trn*C-*pet*N intergenic spacer, the *pet*N gene, the *pet*N-*psb*M intergenic spacer and the *psb*M gene to evaluate the utility of these regions at the interspecific level among species in *Panax*. The *psb*M-*trn*D region ranging from 1144 to 1174 bp in length had 7 deletions and 5 insertions. The phylogeny of *Panax* inferred from these regions was congruent with morphology. Yue *et al*. (2009) studied the phylogeny of *Solms-Laubachia* (Brassicaceae) s.l., based on the data from two nuclear genes (*LEAFY* and *G3pdh*) and two chloroplast intergenic spacers (*pet*N-*psb*M and *psb*M-*trn*D) as the *psb*M-*trn*D region showing high percentage of variable sites. In addition, the *psb*M-*trn*D region of the phlox family was reported to be varied in length from 320-1,123 bp and were used to combine with nuclear ITS, chloroplast matK, *trn*L intron plus *trn*L-*trn*F, *trn*D-*trn*T, *trn*S-*trn*G and *psb*M-*trn*D to assess relationships within the phlox family (Johnson *et al*., 2008).

2.5 The *trn*S- *trnf*M region

The *trn*S- *trnf*M region is located in the large single-copy region. Shaw *et al*. (2005) identified that the *trnS-trnf*M region could provide the greatest number of Potentially Informative Character (PIC) from 21 non-coding regions. The inter- and intraspecific differences in the genus *Citrullus* were analyzed using PCR-RFLP of *trn*S- *trnf*M region combined with *ndh*A and *trn*C-*trn*D regions (Dane *et al*., 2004). Minami *et al*. (2009) showed that DNA polymorphisms in the *trn*S- *trnf*M region

Copyright by Kasetsart University All rights reserved

could be used to identify curcuma plants. In addition, the number of AT repeats in this region could detect the curcumin content in the rhizome of *Curcuma longa*. Plant molecular systematic of closely related data was investigated using the sequence data from *trn*S- *trnf*M, *rpl*16, *psb*M-*trn*D, *rpo*B-*trn*C, *rps*16, *trn*D-*trn*T, *trn*S-*trn*G, *trn*T-*trn*L and *ycf*6-*psb*M and nuclear ITS (ITS1, 5.8S and ITS2). This *trn*S- *trnf*M region has shown to provide the greatest number of informative characters in several plants (Mort *et al*., 2007). Yuan and Olmstead (2008) suggested that *trn*S- *trnf*M could give the highest number of PIC compared to the other regions from the phylogenetic study of *Verberna* complex (Verbenaceae).

3. Simple sequence repeats (SSR) of chloroplast

Simple sequence repeats (SSR) are short tandem repeats and are classified into two groups as microsatellite and minisatellite depending on the number of base in the repeat sequence. Microsatellite characterized by short (<6 bp) and minisatellite characterized by long (>6 bp), are present in coding and non-coding regions of genomes. Length variable microsatellites are the result of microstructural mutations. Microstructure change such as insertions/deletions (indels) and inversions in chloroplast genome can be useful tool for resolving phylogenetic relationship in several plants. Microsatellites are abundant and spread throughout the chloroplast genome and are used as genetic markers. Furthermore, microsatellite and minisatellite of non-coding regions reveal high polymorphism and have been used to population and systematic studies.

Deguilloux *et al*. (2004) successfully used chloroplast microsatellite to test differentiate oak population in two economically tree species: *Quercus petraca* and *Q. robur*. Furthermore, Pakkad *et al*. (2008) studied the level of genetic diversity and differentiation of 10 populations of *Quercus semiserrata* Roxb. in northern Thailand. The variation at eight nuclear (nSSR) and nine chloroplast (cpSSR) microsatellite loci was detected which indicated the highest genetic diversity of these populations. Tesfaye *et al*. (2007) detected regions in the chloroplast genome within infraspecific variation in *Coffea*. Seven variable of microsatellite were characterized in *Coffea* and

Copyright by Kasetsart University All rights reserved

all microsatellites were mononucleotide repeats of $(A/T)_n$. This study was supported by Arroyo-Garcia *et al*. (2002) who tested 10 pairs of consensus primers in *Vitis* species. Three chloroplast microsatellite loci were found to be polymorphism due to the variable number of A and T residues in the amplified regions. Polymorphisms were used to examine the maternal inheritance of chloroplast in *V. vinifera*. PCR markers were developed from SSR loci in the chloroplast DNA of *Nicotiana tabacum* and applied to potato and tobacco analysis. These markers showed high levels of intra- and interspecific diversity suggesting that chloroplast SSR could be use in population genetic, germplasm management, evolutionary and phylogenetic studies (Bryan *et al*., 1999).

A variable minisatellite sequence in the chloroplast of *Sorbus* L. (Rosaceae: Maloideae) was presented by PCR-RFLP analysis of *trn*M-*rbc*L region. The most mechanism for the evolution of the *Sorbus* minisatellite was slipped-strand misparing (King and Ferris, 2002). Tian *et al*. (2008) successfully studied the genetic diversity and structure of *Pinus kwangtungensis* (Pinaceae) by minisatellite of chloroplast DNA. Plastid minisatellites have been proven to be useful marker in population genetics. Cozzolino *et al*. (2003) studied the molecular evolution of a chloroplast minisatellite locus in *Anacamptis palustris* from the $tRNA^{LUE}$ intron which the chloroplast microsatellite locus analyzed in *A. palustris* revealed high levels of polymorphism in the two populations.

4. Phylogenic relationships of Cycad

There have been a number of previous studies of phylogenetic relationships of the cycads based upon molecular evolution of chloroplast DNA regions. Analyses based on 17 chloroplast genes and associated with non-coding regions (*atp*B, *rbc*L, *psb*B, *psb*T, *psb*N, *psb*H, *psb*D & *psb*C, *psb*E, *psb*F, *psb*L, *psb*J, *ndh*F, *rpl*2, 3'- *rpl*12, *rpl*7, *ndh*B and *trn*L) suggested that the genus *Stangeria* was the sister group of subfamily Zamioideae (*Ceratozamia*, *Chigua*, *Zamia* and *Microcycas*), that *Bowenia* was basal to most genera traditionally allocated to the family Zamiaceae, and that *Dioon* was basal to all cycads with the exception of *Cycas* (Zgurski *et al*., 2008). In

Copyright by Kasetsart University All rights reserved

contrast both Bogler and Francisco-Ortega (2004) and Chaw *et al*. (2005) using combined datasets of chloroplast DNA and nuclear DNA concluded that *Stangeria* was sister to the tribe Zamieae which comprises only *Microcycas* and *Zamia*. Combined data from ITS, 26S rRNA, *rbc*L and *trn*F-*trn*L regions implicated that *Stangeria* was the most basal group to all cycad genera with the exception of *Cycas* and that *Dioon* was sister to the tribe Encephalarteae (*Encephalartos, Macrozamia*, *Lepidozamia*) and *Bowenia* (Hill *et al*., 2003). In addition, Treutlein and Wink, (2002) used *rbc*L sequences to reconstruct phylogenetic tree of cycad. *Stangeria* clustered within Zamiaceae while *Bowenia* was a basal group. All of these phylogenetic analyses indicated that the positions of *Bowenia*, *Stangeria* and *Dioon* were incongruent with the earlier classification based on morphology (Stevenson, 1992). The differing placement of these three genera within these various studies has meant that a stable phylogenetic classification for the inter-generic relationships of the cycads has remained elusive. These unclear phylogenetic relationships must be analyzed using the other DNA sequences in further studies.

5. Sex determination of plant

Sex determination is the developmental decision that occurs during the plant life cycle and leads to the differentiation of the two organs or cells producing two gametes (Juarez and Banks, 1998). Sex determination in the plant kingdom can be separated into 2 groups. Sexually monomorphic species are hermaphrodite and considered a large group of plants. The term cosexual is used when individual plants have both sex functions either present within the same flower (hermaphrodite) or in separate male and female flowers (monoecious). A minority of plants species are sexually polymorphic (dioecious), in which male and female flowers appear on separate individuals (Charlesworth, 2002).

Dioecious plants are thought to be the most evolved member of the plant kingdom in terms of sex differentiation and heteromorphic sex chromosome. The small groups of dioecious plants have an XY system, in which male are heterogametic (XY) and female are homogametic (XX). There are two types of sex chromosome;

Copyright by Kasetsart University All rights reserved

homomorphic sex chromosome, in which the sex chromosome are morphologically indistinguishable from autosomes, and heteromorphic sex chromosome, which can be discriminated in cytological analyses (Stehilk and Blattner, 2004). In *Cycas*, The males and females plant of *C. revoluta* have 2n=22 chromosome number. The karyotype of *C. revoluta* showed that the short telocentric chromosome was presented in all male plants. The shortest telocentric chromosome was Y chromosome and one of the other telocentric chromosomes was X chromosome. *In situ* hybridization (ISH) and fluorescence *in situ* hybridization (FISH) were used to examine the structure change of this heteromorphic pair. These signals were appeared as large segment at the proximal region of all telocentric chromosomes and they did not show in the shortest telocentric chromosome in the male plants. It is concluded that this chromosome pair of male *C. revoluta* was heterozygous and in the female plant was homozygous (Hizume *et al*., 1998).

Although dioecious plants have a XX/XY chromosome arrangement similar to mammals, the Y chromosome of *Silence*, *Rumex* and *Cannabis* are much larger than the X chromosomes, in contrast to mammal (Sakamoto *et al*., 2005). However, some dioecious plants have homomorphic (equal length) sex chromosome such as papaya and kiwi fruit. Hence, the sex-determining genes of these plants seem to map at small region of autosomes (Harvery *et al*., 1997 and Ming *et al*., 2007). Irish and Nelson (1986) described that genetic determination systems can be controlled by a single locus on autosome, multiple loci on autosomes, or a gene on heteromorphic sex chromosome.

6. Molecular markers for sex identification

A full understanding of the biology of any sexually reproducing species is dependent on being able to recognize male from female because sex determination is an important developmental event in the life cycle of all sexually reproducing plants. Moreover, sex identification is important to the design of breeding program, whether for industrial or conservation purpose (Griffiths and Tiwari, 1993). The sex of many species cannot be separated from morphology, male and female plants cannot be

Copyright by Kasetsart University All rights reserved

distinguished at the seedling stage. Moreover, this problem usually increased in species where the sex of an individual is revealed only after flowering which may take few months to several years (Parasnis *et al*., 2000). Thus the development of a rapid molecular technique for sex identification is necessary.

7. Randomly Amplified Polymorphic DNA (RAPD)

In 1990 Williams *et al*. developed a new PCR-based genetic assay namely Randomly Amplified Polymorphic DNA (RAPD). This procedure detects nucleotide sequence polymorphisms in DNA by a single primer of arbitrary nucleotide sequence. In this reaction, a single species of primer anneals to the genomic DNA at two different sites on complementary strands of DNA template. If these priming sites are within an amplifiable range of each other, a discrete DNA product is formed through thermocyclic amplification. On an average, each primer amplifies several discrete loci in the genome making the assay useful for efficient screening of nucleotide sequence polymorphism between individuals. RAPD assay has been used by several groups as efficient tools for identifying of markers linked to agronomically important traits, which are introgressed during the development of near isogenic lines (Yu *et al*., 2005).

Michelmore *et al*. (1991) described an application of RAPD termed bulked segregation analysis (BSA) to identify molecular markers linked to a trait of interest. The method involved comparing two pooled DNA samples of individuals from a segregating population originating from a single cross. Within each pool or bulk, the individuals have an identical phenotype for trait but are arbitrary for other traits. Two pools contrasting for the phenotype (e.g., male and female plant) are analyzed to identify markers that can distinguish them. Variation not associated with sex determination should be common to both bulks, while any polymorphisms between the pooled samples should be linked to sex determination. Because RAPDs are dominant markers whose inheritance can be traced within families, markers can be confirmed by mapping in families. The frequency of sex-linked markers is related to

Copyright by Kasetsart University All rights reserved

chromosome number, total size of genome and relative size of the segment which determines sex (Harvey *et al.*, 1997).

A study of sex identification of *Cycas circinslis*, family Cycadaceae, found that OPB-01 and OPB-05 RAPD primers appeared one-male specific band and another female specific band, respectively. Sequencing of a male-specific RAPD band revealed homology with putative retro elements of diverse plants, probably indicating its use in the detection of male *C. circinslis*. On the other hand, the female-specific DNA fragment did not result in any significant match (Gangopadhyay *et al*., 2007). In addition, another species of Cycad, *Encephalartoes natalensis*, found that only OPD-20 primer generated a specific band in female DNA and an absent band in male DNA (Prakash and Staden 2006). In addition, RAPD technique was used to identify the sex determination element on the Y chromosome of *Silene latifolia*. Thirty-one male-specific DNA fragments were detected and some of them showed high specificity to the male genome. The most specific clone, which contains a long open reading frame (ORF), was successfully determined to exist as a single copy on the Y chromosome (Nakao *et al*., 2002).

Many researchers have been studied the marker linked to sex determination in plants such as *Silene latifolia*, dioecious plant with heteromorphic sex chromosomes. The study used RAPD method and bulked segregation analysis for sex determination. Of 60 primers, only 4 primer produced specific fragments in male DNA but absent in female (Mulcahy *et al*.,1992). Hormoza *et al*. (1994) detected a single marker associated with female and absent in males in *Pistacia vera*. This marker could be used in a breeding program to screen the gender of pistachio plants before they reach reproductive maturity, resulting in considerable saving of time and economic resources. Alstrom-Rapoport *et al*. (1998) used bulked segregation analysis of RAPD products to identify markers linked to sex determination in *Salis viminalis*. The result showed that only a single band of 1080 RAPD bands was linked to a sex determination locus. Furthermore, RAPD can be used to indentify a male-specific DNA fragment in *Atriplex garrettii*, perennial diploid dioecious species. Even though

Copyright by Kasetsart University All rights reserved

species of the genus *Atriplex* are diverse in both ploidy and sex determination, RAPD marker can be used to detecte sex of this species as well (Ruas *et al*., 1998).

8. Inter Simple Sequence Repeat (ISSR)

ISSR technique was developed by Zietkiewicz *et al*. (1994). ISSRs are semi-arbitrary markers amplified by PCR in the presence of one primer complementary to the target microsatellite-primed PCR. Amplification does not require genome sequence information and lead to multilocus and highly polymorphous patterns. Each band corresponds to a DNA sequence delimited by two inverted microsatellites. Like RAPDs, ISSR markers are quick and easy to handle, but they seem to have the reproducibility of simple sequence repeat (SSR) marker because of the longer length of their primers (Bornet and Branchard, 2001).

ISSR method has been used to detect a sex specific molecular marker in hop (*Humulus lupulus* L.). Two ISSR primers revealed fragments specific to male plants of hop. The result revealed that these sequences were found high homology to express sequence from EST plants of EMBL database, most of which code cell wall glycoprotein (Danilova and Karlov, 2006). The same method also identified sex of papaya, an angiosperm, in pre-flowering stage. The result showed that one female specific band generated from ISSR profiled using primer $(GACA)_4$ was detected in papaya, which seems to have importance from agricultural point of view (Gangopadhyay *et al*., 2007). Sharma *et al*. (2008) identified male and female plants of jojoba (*Simmondsia chinensis*) using ISSR marker. Of the 42 primers analysis, only one primer (UBC-807) produced a unique 1,200 bp fragment in the male DNA. This was a first report of the use of ISSR markers to investigate sex in *S. chinensis* plants.

9. Sequence characterized amplified regions (SCAR)

SCAR is a DNA fragment amplified by polymerase chain reaction (PCR) using specific 15-30 bp primers, designed from a nucleotide sequence established

Copyright by Kasetsart University All rights reserved

from cloning RAPD of ISSR fragments linked to a trait of interest. The presence or absence of a band indicates variation in sequence. These are better reproducible than RAPDs (Paran and Michelmore, 1993). In comparison with arbitrary primers, SCARs exhibit several advantages in mapping studies (codominant SCARs are informative for genetic mapping than dominant RAPDs), map-based cloning as they can be used to screen pooled genomic libraries by PCR (Dioh *et al*.,1997), physical mapping (Didion *et al*., 2000), locus specificity (Arnedo-Andres, 2002). Besides, SCARs allow comparative mapping or homology studies among related species (Huaracha *et al*., 2004).

Recently, SCAR markers were developed to determine sex in many dioecious plants. This technique was used to identify the sex of papaya (*Carica papaya* L.). Papaya is a polygamous diploid plant species with three sex type, i.e. male, female and hermaphrodite (Moore, 2004). For the first time, there was a report that PCR-based Seeding Sex Diagnostic Assay (SSDA) specially designed for early sexing of papaya seedling and developed a male-specific SCAR marker in papaya (Parasnis *et al*., 2000). Then, Urasaki *et al*. (2002) identified sex of papaya and the result showed that a 450 bp fragment exists in all male and hermaphrodite plants but not all females. SCAR marker was developed from this band to determine the sex of papaya and this marker was located on chromosome segment which was specific both for male and the hermatophrodite. Chaves-Bedoya and Nunez (2007) reported a new SCAR marker specific to male and hematophrodite papaya plants but the marker could not be used to amplify the female plants.

Annual dioecious *Mercurialis annua* does not have sex chromosome which is similar to that of papaya. It has a multi-loci mechanism for sex determination and there are three unlinked loci A, B1 and B2 that involved in gender determination. However, male-specific RAPD markers were identified and SCAR primers were amplified as male specific marker (Khadka *et al*., 2002). In addition, SCAR markers were used to map molecular marker to sex locus M of *Asparagus offcinalis* L. (asparagus). Therefore, its sex determination is controlled by a single dominant gene

Copyright by Kasetsart University All rights reserved

(M). The male genotype is dominant (M_) while the female is homozygote recessive (mm) (Jiang and Sink, 1997).

Copyright by Kasetsart University All rights reserved

MATERIALS AND METHODS

1. Plant materials

Cycad leaves of 45 species from 11 genera (Table 1) were collected from plants cultivated in Tsukuba Botanical Garden, Ibaraki, Japan and Nongnooch Botanical Tropical Garden, Chonburi, Thailand. All species were used in the phylogenetic relationships within the family Zamiaceae and Stangeriaceae, while two *Cycas* species, *C. revoluta* and *C. elephantipes*, were included in the analysis as an outgroup. Young cycad leaves from 27 species of *Cycas* and 2 species of *Dioon* were employed for examining the phylogenetic relationship within the family Cycadaceae having *Dioon* as an outgroup (Table 2). For sex identification study, the plant materials were collected from Nongnooch Tropical Garden, Chonburi, Thailand. Young cycad leaves were collected from three male and three female plants from each species (*C. clivicola*, *C. edentata and C. chamaoensis*), male and female *C. elephantipes* and *C. rumphii*, female *C. noognoochiae* and male *C. pectinata* and *C. siamensis*.

Copyright by Kasetsart University All rights reserved

Table 1 Forty-five cycad species used in phylogenetic relationships within the family Zamiaceae and Stangeriaceae**.**

Family	Subfamily	Tribe	Subtribe	Genus	Species
Stangeriaceae	Stangerioideae			*Stangeria*	*Stangeria eriopus* (Kunze) Baill.
	Bowenioideae			*Bowenia*	*Bowenia serrulata* (W.Bull) Chamb.
					Bowenia spectabilis Hook. Ex Hook. F.
Zamiaceae	Encephalartoideae	Diooeae		*Dioon*	*Dioon edule* Lindl.
					Dioon spinulosum Dyer
		Encephalarteae	Encephalartinae	*Encephalartos*	*Encephalartos altensteinii* Lehm.
					Encephalartos arenarius R.A.Dyer
					Encephalartos barteri Carruth. Ex Miq.
					Encephalartos ferox Bertol. F.
					Encephalartos friderici-guilielmi Lehm.
					Encephalartos hildebrandtii A. Braum & Bouche
					Encephalartos lehmannii Lehm.
					Encephalartos longifolius (Jacq.) Lehm.
					Encephalartos manikensis (Gilliland)
					Encephalartos natalensis R.A.Dyer et I.Verd.
					Encephalartos paucidentatus Stapf et Burtt Davy
					Encephalartos trispinosus (Hook.) R.A.Dyer
					Encephalartos villosus Lem.
			Macrozamiinae	*Macrozamia*	*Macrozamia communis* L.A.S.Johnson
					Macrozamia fawcettii C.Moore
					Macrozamia macdonnellii (F.Muell. ex Miq.) A.DC.
					Macrozamia miquelii (F.Muell. ex Miq.) A.DC.
					Macrozamia moorei F.Muell.
					Macrozamia pauli-guilielmi W.Hill et F.Muell.

Copyright by Kasetsart University. All rights reserved

Table 1 (Continued)

Family	Subfamily	Tribe	Subtribe	Genus	Species
			Macrozamiinae	*Macrozamia*	*Macrozamia platyrhachis* F.Muell.
					Macrozamia reducta K.D.Hill et D.L.Jones
					Macrozamia riedlei (Gaudich.) C.A.Gardner
					Macrozamia secunda C.Moore
					Macrozamia spiralis (Salisb.) Miq.
				Lepidozamia	*Lepidozamia hopei* Regel
					Lepidozamia peroffskyana Regel
	Zamioideae	Ceratozamioideae		*Ceratozamia*	*Ceratozamia hildae* G.P.Landry et M.C.Wilson
					Ceratozamia kuesteriana Regel
					Ceratozamia mexicana Brongn.
					Ceratozamia microstrobila Vovides et J.D.Rees
					Ceratozamia mirandae P-Farrera et al.
					Ceratozamia norstogii D.W.Stev
		Zamieae	Microcycadinae	*Microcycas*	*Microcycas calocoma* (Miq.) A.DC.
			Zamiinae	*Zamia*	*Zamia vasquesii* D.W.Stev.
					Zamia integrifolia L. f.
					Zamia manicata Linden ex Regel
					Zamia pygmaea Sims
				Chigua	*Chigua restrepoi* D. W. Stev.
Cycadaceae				*Cycas*	*Cycas elephantipes* A. Lindström & K. D. Hill
					Cycas revoluta Thunb.

Copyright by Kasetsart University All rights reserved

Table 2 Twenty-seven *Cycas* species and two species of *Dioon* used in phylogenetic relationships within the family Cycadaceae.

Family	Genus	subsections	Species
Cycadaceae	*Cycas*	*Asiorientales*	*Cycas revoluta* Thunb.
			Cycas taitungensis C.F. Shen, K.D. Hill, C.H. Tsou & C.J. Chen
		Wadeanae	*Cycas curranii* (J. Schust.) K.D. Hill
			Cycas wadei Merrill
		Indosinenses	*Cycas chamaoensis* K.D. Hill
			Cycas clivicola K.D. Hill
			Cycas condaoensis K.D. Hill, Hiêp & S.L.Yang
			Cycas elephantipes A. Lindström & K.D. Hill
			Cycas lindstromii S.L.Yang, K.D. Hill & Hiêp
			Cycas nongnoochiae K.D. Hill
			Cycas siamensis Miq.
			Cycas tansachana K.D.Hill & S.L.Yang
		Cycas	*Cycas bougainvilleana* K.D. Hill
			Cycas edentata de Laub.
			Cycas media R. Brown
			Cycas pranburiensis S.L. Yang, W. Tang, K.D. Hill& Vatcharakorn
			Cycas rumphii Miq.
			Cycas seemannii A. Braun
			Cycas semota K.D. Hill
			Cycas thouarsii R. Br. Ex Gaudich.
			Cycas zeylanica (Schuster) K D. Hill & A. Lindström
		Stangerioides	*Cycas balansae* Warb.
			Cycas changjiangensis N. Liu
			Cycas hainanensis C.J. Chen
			Cycas hoabinhensis K.D. Hill, Hiêp & P.K. Loc
			Cycas segmentifida D. Yue Wang & C.Y. Deng

Copyright by Kasetsart University All rights reserved

Table 2 (Continued)

Family	Genus	subsections	Species
Zamiaceae	*Dioon*		*Dioon edule* Lindl. *Dioon spinulosum* Dyer

Copyright by Kasetsart University All rights reserved

2. DNA extraction

DNA was extracted using 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 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 volume 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, 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 DNA pellets. Finally, the DNA pellets were air-dried and resuspended in 30 µl TE buffer. DNA concentration was measured on a 1% agarose gel using UV spectrophotometer.

3. PCR amplification for non-coding regions of chloroplast DNA

PCR amplification was performed in 50 µl of reaction mixture consisting of 5 µl of 10X PCR buffer (TaKaRa), 2.5 mM $MgCl_2$, 200 µM dNTP, 1.25 unit of *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 thermocycler 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, *atp*B-*rbc*L and *trn*S- *trnf*M), 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 non-coding chloroplast DNA primers are shown in Table A1 (Appendix).

Copyright by Kasetsart University All rights reserved

4. Sequencing analysis

4.1 Pre-sequencing preparation

PCR products were purified prior to sequencing using Exosap-IT kit (United States Biochemical). An aliquot of 2 µl of PCR product was mixed with 0.3 µl of Exosap-IT. The mixtures were incubated at 37°C for 15 min to degrade remaining primers and nucleotides. Then, the mixtures were incubated at 80°C for 15 min to inactivate Exosap-IT.

4.2 Sequencing

After PCR products were purified, the purified PCR products were sequenced by a modified and automated dideoxynucleotide chain-reaction method (Sanger *et al*., 1977). The reaction used the BigDye™ Terminator v3.1 Cycle Sequencing Kit (Applied Biosystem, USA). The reaction mixture contained 1 µl of the purified PCR products, 1 µl of BigDye ready reaction mix, 0.5 µl of BigDye sequencing buffer, 3.2 pmol of each primer and 1.7 µl of deionized water. The thermal cycle sequencing reaction was run on a Perkin Elmer 9700 thermocycler with the initial denaturation of 1 min at 96 °C, followed by 25 cycles each with 10 s at 96°C, 5 s at 50°C, 4 min at 60°C and hold on 4°C. The products were precipitated and purified by ethanol (EtOH)/sodium acetate (NaOAc) precipitation method. The products were added with 12.5 µl of 95% EtOH and 0.5 µl of 3M NaOAc, pH 4.6 and mixed using vortex. The products were incubated at room temperature for 15 min and centrifuged at 15,000 g for 15 min. After centrifugation, supernatant was discarded. The pellets were washed with 70% EtOH and centrifuged at 15,000 g for 15 min. The pellets were dried in vacuum for 20 min and dissolved in 15 µl of DiHi formamide solution and heated at 95°C for 2 min. The products were run on ABI PRISM® 3130X/ Genetic Analyzer using KB_3130_POP7_BDV3 protocol.

Copyright by Kasetsart University All rights reserved

5. Data analysis

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

Before the combined data was analyzed, the congruence between different data sets was tested by performing incongruence length difference test (ILD) to evaluate the conflict that can occur between sets of characters from different data sources (Farris *et al*., 1994). PAUP were 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 and the number of base substitution per site. 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 beta 10 (Swofford, 2002). MP were conducted using heuristic search methods under the equal weighted criteria with 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 MODELTEST 3.7 program (Posada and Crandall, 1998). ML heuristic searches were performed using 10 random addition sequences and TBR swapping. Support for braches in the ML trees was tested by bootstrap analysis of

Copyright by Kasetsart University All rights reserved

1,000 replicates. NJ was constructed using the MEGA 4 program. Clustering of species of each tree was confirmed with bootstrap value of 10,000.

6. Bulk segregant analysis for sex identification

DNA concentration of an individual samples was estimated with spectrophotometer and adjusted to 100 ng/µl. Two bulked samples were made from an equal amount of either male *Cycas* or female *Cycas* (*C. clivicola*, *C. edentata* and *C. chamaoensis*).

7. RAPD analysis

The protocol for RAPD analysis was adapted from that of Williams (1990). The volume of the final reaction (25 µl) consisted of 1 X PCR buffer (10 mM Tris-HCl pH 8.0 and 50 mM KCl), 3.0 mM $MgCl_2$, 1.25 unit of *Taq* DNA polymerase, 200 µM dNTP, 10 µM random primer (OPA, OPB, OPK from Operon Tecnologies and UBC 480-500), 100 ng of genomic DNA. Amplifications were made in a Perkin Elmer 9600 thermocycler with an initial denaturing step of 3 min at 94°C, followed by 45 cycles of 1 min at 94°C, 1 min at 36°C, 2 min at 72°C and a final extension of 5 min at 72°C. PCR products were subjected to 1% agarose gel in TBE buffer (89 mM Tris-HCl, 89 mM boric acid and 2.5 mM EDTA pH 8.0), electrophoresis run at 90 V and DNA bands were visualized by ethidium bromide staining.

9. Inter Simple Sequence Repeat (ISSR) analysis

Twenty-nine ISSR primers (Biotechnology Laboratory, University of British Columbia, Vancouver, Canada, http://www.ubc.ca) were used to screen sex specific marker of male and female *Cycas*. All ISSR primers used are shown in Table A2 (Appendix). Reactions were carried out in a total volume of 20 µl consisting of 20 ng of template DNA, 1X PCR buffer, 2.5 mM $MgCl_2$, 0.1 mM dNTPs, 200 nM primers, 1.0 unit of *Taq* polymerase and sterile water. Amplifications were made in a Perkin

Copyright by Kasetsart University. All rights reserved

Elmer 9600 thermocycler with an initial denaturing step of 3 min at 94 °C, followed by 35 cycles of 30 s at 94°C, 45 s at 52°C, 1.30 min at 72°C and a final extension of 5 min at 72°C. PCR products were separated by electrophoresis on 1.6% agarose gels in TBE buffer and visualized using ethidium bromide staining.

10. Sequence Characterized Amplified Region (SCAR) analysis

10.1 DNA elution from agarose gel

Candidate sex-specific RAPD and ISSR bands were excised from the agarose gel using a sterile razor blade. DNA was extracted from the agarose gel using the QIAquick Gel Extraction Kit (QIAGEN) according to the protocol recommended by the manufacturer. The excised gel was transferred into a microcentrifuge tube. Three volume of GQ Buffer to one volume of gel (100 mg ~100 µl) was added to gel slice after measuring the weight of the excised gel band. The gel slice was incubated at 50°C for 10 min until it was completely dissolved. The solution was mixed by vortex every 2-3 min during the incubation to help dissolving the gel. After the gel slice was completely dissolved, the color of the mixture was determined for its pH. If the color was orange or violet, 10 µl of 3 M sodium acetate (pH 5.0) was added and mixed again until it turned to yellow. Equal volume of isopropanol was added to the sample and mixed for DNA precipitation. Then, the solution was pipetted to the QIAquick spin column which was placed in a collection tube and centrifuged at 13,000 g for 1 min. Since the maximum volume of the column reservoir was 800 µl**,** the solution volume of more than 800 µl was added later to the QIAquick spin column and centrifuged again. The flow- through was discarded and the QIAquick spin column was placed back in the same collection tube. Then, 0.75 ml of PE buffer was added to the QIAquick spin column and centrifuged at 13,000 g for 1 min to wash DNA pellet. The flow-through was discarded and the QIAquick spin column was centrifuged at 13,000 g for an additional of 1 min to get rid of the remaining ethanol. The QIAquick spin column was placed into a clean 1.5 ml microcentrifuge tube and 50 µl of EB buffer or sterile distilled water was added to the center of the QIAquick spin column membrane. The QIAquick spin column was incubated at room

Copyright by Kasetsart University All rights reserved

temperature for 2-5 min to increase DNA concentration and centrifuged at 13,000 g for 1 min. The elute sample was stored at -20 °C until use.

10.2 Cloning

10.2.1 Preparation of competent *E. coli* cells (Sambrook *et al*.,1989)

A single colony of *E. coli* (DH5α) was collected from a freshly grown plate of 16-20 h at 37ºC and transferred into 100 ml of LB broth (1% Bacto tryptone, 0.5% Bacto yeast and 0.5% NaCl) in a flask. Then, the cell culture was incubated at 37ºC for 3-5 h in a rotary shaker of 300 cycles/min. To monitor the growth OD_{600} of the cell culture, the solution was determined every 1 h (OD_{600} = 0.2-0.4). The cell culture was transferred to 50 ml polypropylene tube. The cell culture was incubated on ice for 15 min and centrifuged at 3,500 g at 4 ºC for 15 min. The cell pellet was resuspended with 5 ml of RF1 (10 mM KCl, 50 mM $MnCl_2.4H_20$, 30 mM CH_3COOK, 10 mM $CaCl_2$ and 15% glycerol, pH 5.8, sterilized by filtration and stored at 4°C) and incubated on ice for 15 min. The cell culture was centrifuged at 3,500 g at 4 ºC for 15 min. The cell pellets were resuspended with 2 ml of RF2 (10 mM MOPs, 10 mM KCl, 75 mM $CaCl_2$ and 15% (V/V) glycerol, pH 6.8, sterilized by filtration and stored at 4°C). Then, 100 µl of each suspension competent cell was transferred to a 1.5 ml microcentrifuge tube and stored at -80°C.

10.2.2 Ligation (GeneJET^{TM}PCR Cloning Kit)

The ligation reaction was set up at the total volume of 10 µl containing 1 µl of the purified DNA fragment, 1 µl of DNA blunting enzyme and 6 µl of water (nuclease-free). The ligation reaction was mixed by brief vortex and centrifuged at 2,500 g for 3-5 s. The ligation reaction was incubated at 70°C for 5 min and chilled on ice for 5 s. Then, the ligation reaction was added with 1 µl of pJET1/blunt cloning vector (50 ng/µl) and 1 µl of T4 DNA ligase (5U/µl). The ligation reaction was mixed by brief vortex, centrifuged at 2,500 g for 3-5 s and incubated at room temperature for 5 min.

Copyright by Kasetsart University All rights reserved

10.2.3 Transformation (Heat-Shock)

The competent *E.coli* cells kept at –80°C were defrozen on ice for 5 min. For transformation, 2 µl of the ligation reaction was added to the competent *E.coli* cells, incubated on ice for 10 min, and the reaction mixture tube was placed in the water bath at 42°C for 90 s without shaking. The reaction mixture tube was transferred on ice for 2 min to reduce damage to the *E.coli* cells. The reaction mixture was added with 900 µl of SOC medium (2% Bacto tryptone, 0.5% Bacto yeast extract, 10 mM $NaCl_2$, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$ and 20 mM glucose) and incubated at 37°C for 1 h. Thereafter, 100 µl of the culture was spread onto LB-Amplicillin agar plate (without IPTG and X-gal) and incubated overnight at 37°C.

10.2.4 Plasmid extraction

Plasmid was extracted from *E.coli* cells using GeneJET™ Plasmid Miniprep Kit according to the protocol recommended by the manufacturer. A single colony was collected from a freshly streaked selective plate and cultured in 3 ml of LB broth containing 100 µg/ml of Ampicillin and then incubated for 12-16 h at 37°C with shaking at 200-250 rpm. The cell culture was transferred to a 1.5 ml microcentrifuge tube and centrifuged at 8000 g for 2 min at room temperature. The pellet was resuspened with 250 µl of the resuspension solution. The cells were resuspended by vortexing or pipetting up and down until cell clumps were completely dispersed. Then, 250 µl of the lysis solution was added to the cells solution and thoroughly mixed by inverting the tube 4-6 times until the solution became clear to lyses bacteria cell. The cells solution was added with 350 µl of the neutralization solution and mixed immediately and thoroughly by inverting the tube 4-6 times. The cells solution was centrifuged at 12,000 g for 5 min. The supernatant was transferred to the supplied GeneJET™ spin column by decanting or pipetting. Then, 500 µl of the washing solution was added to the GeneJET™ spin column and centrifuged at 12,000 g for 30-60 s and the flow-through was discarded. The GeneJET™ spin column was placed back into the same collection tube and

Copyright by Kasetsart University All rights reserved

centrifuged at 12,000g for 1 min. The GeneJET™ spin column was transferred into a new 1.5 ml microcentrifuge tube. Fifty µl of the elution buffer was added to the center of the GeneJET™spin column membrane and incubated for 2 min at room temperature and then centrifuged at 12,000 g for 2 min. The purified plasmid DNA was kept at -20 °C.

11. Sequencing

To analyze the DNA sequence, the purified plasmid DNA was determined by Bioservice, NSTDA, Thailand. BLAST program was used to identify similarity sequence from the other species as well as to analyze the pairwise alignment. Sequence analysis and alignment was performed using ClustalX sequence program.

12. Primer design for SCAR markers

PCR specific primers were designed from the results of RAPD and ISSR analysis using Primer3 program. The selected criteria for primer designing was the 18-25 bases long, the melting temperature of 55-65°C, random base distribution to avoid polypurine and polypyrimidine and less than 5°C differences of the melting temperature of a primer pair.

14. SCAR marker testing

The male and female *Cycas* were used for testing SCAR primer. The PCR reaction was performed in 25 µl reaction mixture containing 1 X PCR buffer, 2.5 mM $MgCl_2$, 1.25 unit of *Taq* DNA polymerase, 200 µM dNTP, 10 µM SCAR primers, 100 ng of genomic DNA. Amplifications were made in a Perkin Elmer 9600 thermocycler. PCR products were subjected to 1% agarose gel in TBE buffer and DNA bands was visualized by ethidium bromide staining.

Copyright by Kasetsart University All rights reserved

RESULTS

1. Phylogenetic relationships within the family Zamiaceae and Stangeriaceae

1.1 Nucleotide sequences and variation

1.1.1 *atp*B-*rbc*L spacer

The average length of the *atp*B-*rbc*L spacer was 645 base pairs (bp) varying in length from 637 (*M. pauli-guilielmi*) to 668 bp (*Mi. calocoma*). This spacer was AT rich (64.6%) while GC content was found at 35.4% (Table 3). The sequence alignment was 689 bp and showed variation among the cycad genera. A variable site was 187 and 119 (14.9%) which were informative sites. This sequence contained 13 indels (insertion /deletion). Furthermore, the average number of substitution per site was found at 0.0419. Although, *Ceratozamia* and *Stangeria* were placed in the different family, they shared 2 sites of 1 base substitution site at the position 403 and 541 and three base insertions (GGA) at the position 266-269. *Zamia*, *Chigua* and *Microcycas* shared 15 bp insertions (CCTTAGATTCATTAA) at the position 459-473. *Dioon* contained 4 bp insertions at the position 232-235 and *Microcycas* had 4 bp insertions at the position 192-195. However, *Macrozamia* and *Dioon* showed 5 bp deletions at the position 588-592. Genetic distances matrix based on the *atp*B-*rbc*L sequences was estimated using the formula Kimura-2 parameter (Table 4). Genetic distance for outgroup and ingroup ranged from 0.1041(*Cycas-Encephalartos*) to 0.1406 (*Cycas-Ceratozamia*) and 0.0080 (*Encephalartos-Lepidozamia*) to 0.0858 (*Chiqua-Dioon*), respectively.

1.1.2 *psb*M-*trn*D spacer

The *psb*M-*trn*D non-coding region of chloroplast DNA varied in length from 662 (*CZ. maxicana*) to 974 bp (*E. hildebrandtii*) and its average length was 890 bp. It showed 62.4% AT and 37.6% GC. The sequence alignment was 1162

Copyright by Kasetsart University All rights reserved

bp and the average number of substitutions per site was 0.1182. The *psb*M-*trn*D non-coding region of these cycads was highly variable at 26 indels. Large deletions (301 bp) were found in *Ceratozamia* resolving in the shortest sequences in the alignment data. Furthermore, *Ceratozamia* showed 1 bp insertion at the position 507, 3 bp insertions at the positions 206-208 and 689-691, 7 bp insertions at the positions 102-108 and 392-398 and 8 bp insertions at the position 632-639. *Lepidozamia* contained 1 bp insertion at the position 596, 2 bp insertions at the position 912-913 and 5 bp insertion at the position 915-919. *Encephalartos* showed 5 bp insertions at the position 745-749 and only three species, *E. manikensis*, *E. villosus* and *E. altensteinii*, appeared a unique 19 bp insertions at the position 571-589. In addition, *Lepidozamia* showed 17 bp insertions at the position 997-1013 and share 5 bp insertions with *Encephalartos*. *Zamia* and *Chigua* shared 1 bp insertion at the position 89. Genetic distances matrix based on the *psb*M-*trn*D sequences was estimated using the formula Kimura-2 parameter (Table 5). Genetic distance for outgroup and ingroup ranged from 0.1360 (*Cycas-Dioon*) to 0.3018 (*Cycas-Microcycas*) and 0.0374 (*Encephalartos-Lepidozamia*) to 0.2996 (*Chiqua-Ceratozamia*), respectively.

1.1.3 *trn*L-*trn*F spacer

This region was composed of the *trn*L intron and the *trn*L-*trn*F intergenic spacer IGS). It ranged from 889 (*S. eriopus*) to 1018 bp (*E. longifolius*) and the average length was 959 bp. This sequence comprised of 63.2% (AT) and 36.8% (GC) with the average number of substitutions per site of 0.0803. The alignment length was 1128 bp with 17 indels. Genetic distances matrix based on the *trn*L-*trn*F sequences was estimated using the formula of Kimura-2 parameter (Table 6). Genetic distance for outgroup and ingroup ranged from 0.1549 (*Cycas-Ceratozamia*) to 0.2158 (*Cycas-Microcycas*) and 0.0131 (*Zamia-Chigua*) to 0.1638 (*Chiqua-Microcycas*), respectively.

Interestingly, tandem repeats were found in some genera. *Chigua* and *Zamia* showed 2 repeats with 12 bp (CTCTATCTAGAT) and 2 repeats of 12 bp (ATCTAGATAGAG) were presented in *Bowenia* and *Macrozamia* (Figure 1). Large tandem repeats appeared in *Microcycas* as 2 repeats of 40 bp long

Copyright by Kasetsart University All rights reserved

(ATCTAGATAGAGAT CTAGATAT ATCTGTATGG ATAT CTCT). In addition, this region also presented $(A/T)_n$ mononucleotide repeats and microsatellite repeats. *Zamia* and *Chigua* had one tandem repeat $(TG(A)_{10-12})$ and *Microcycas* had 2 tandem repeats - $TG(A)_{22}$ and $TCC(T)_{13}$. *Ceratozamia* showed a unique tandem repeat $((TAGATA)_2GA(TAAG)_2)$. *Encephalartos* and *Lepidozamia* shared three repeats; $TG(A)_n$, $(TAGATA)_2$gataagtaa$(AGAT)_3$ and $TA(T)_n$. Three tandem repeats $(TG(A)_n$, $(TAGATA)_2$gataagtaa$(AGAT)_3$ and $TG(A)_{12}G(A)_{16})$ were found in *Macrozamia*. In contrast, one tandem repeat as $TT(C)_8(TTT)_3TG(A)_{16}$, $(TAGATA)_2$ and $(TCC)_2(T)_{10}$ was represented by *Dioon*, *Bowenia* and *Stangeria*, respectively (Table 8).

1.1.4 *trn*S-*trnf*M spacer

The *trn*S-*trnf*M non-coding region ranging from 741(*M. communis*) to 894 bp (*M. riedlei*, *M. pauli-guilielmi*, *Z. vasquesii* and *B. spectabilis*) and the average length was 866 bp. It comprised of 62.3% of AT and 37.7% of GC with the average number of substitution per site of 0.0587. The alignment length was 961 bp presented in 19 indels. The number of AT repeated units in this sequence varied among genera. *Macrozamia* differed at the repeat $(AG(AT)_{9-14})$, *Bowenia* at $CT(AT)_6$, *Stangeria* at $TAC(AT)_5$ and *Microcycas* at $TAC(AT)_6$. *Zamia* and *Chigua* shared the units $GC(AT)_5$ and *Encephalartos* and *Lepidozamia* shared the units $CT(AT)_5$. This result implied that *Zamia* was closely related to *Chigua* as well as *Encephalartos* and *Lepidozamia*. *Ceratozamia* demonstrated a high variability of AT repeated patterns at $TAC(AT)_{9-13}$, $TAG(AT)_7$ and $TAC(AT)_2GT(AT)_4$. *Dioon* and *Cycas* showed similar patterns at $TGT(AT)_5$ and $TGT\text{-}T(AT)_4$, respectively. Genetic distances matrix based on the *trn*S-*trnf*M sequences was estimated using the formula Kimura-2 parameter (Table 7). Genetic distance for outgroup and ingroup was found ranging from 0.0724 (*Cycas-Dioon*) to 0.1301 (*Cycas-Stangeria*) and 0.0108 (*Chigua-Zamia*) to 0.1124 (*Chiqua-Stangeria*), respectively.

Copyright by Kasetsart University All rights reserved

Table 3 Characterization of non-coding region chloroplast DNA in cycad taxa.

	Combined	*atp*B-*rbc*L	*psb*M-*trn*D	*trn*L-*trn*F	*trn*S-*trnf*M
No. of taxa	45	45	45	45	45
No. of indels	-	13	26	17	19
Aligned sequences length	3897	689	1162	1128	961
No. of parsimony informative character	991	119	365	310	203
% of parsimony informative character	25.42	17.27	31.41	27.38	21.12
No. of MP trees	49,564	76,605	82,805	627,576	59,100
Length of MP trees	2799	330	1102	886	517
Consistency index (CI)	0.748	0.830	0.748	0.733	0.708
Retention index (RI)	0.829	0.859	0.823	0.806	0.885
Model	K81uf+G[1]	K81uf+G[1]	K81uf+G[1]	K81uf+G[1]	K81uf+G[1]
Gamma shape	0.8762	0.5816	1.8292	0.9146	0.4225
parameter	62.8	64.6	62.4	63.2	62.3
A+T%	37.2	35.4	37.6	36.8	37.7
G+C%	0.0737	0.0419	0.1182	0.0803	0.0587
The average number of substitution per site					

1 = Unequal-frequency Kimura 3-parameter plus Gamma

Copyright by Kasetsart University All rights reserved

Table 4 The number of base substitutions per site from averaging overall the *atp*B-*rbc*L sequence pairs is shown. All results are based on the pairwise analysis of 45 sequences**.**

Genus	*Cycas*	*Dioon*	*Zamia*	*Chigua*	*Microcycas*	*Macrozamia*	*Lepidozamia*	*Encephalartos*	*Bowenia*	*Ceratozamia*	*Stangeria*
Cycas											
Dioon	0.1130										
Zamia	0.1342	0.0787									
Chigua	0.1401	0.0858	0.0102								
Microcycas	0.1227	0.0703	0.0263	0.0325							
Macrozamia	0.1079	0.0467	0.0495	0.0541	0.0399						
Lepidozamia	0.1081	0.0467	0.0520	0.0571	0.0425	0.0093					
Encephalartos	0.1041	0.0462	0.0512	0.0560	0.0417	0.0093	0.0080				
Bowenia	0.1118	0.0531	0.0529	0.0611	0.0449	0.0206	0.0182	0.0173			
Ceratozamia	0.1406	0.0691	0.0695	0.0773	0.0680	0.0498	0.0462	0.0461	0.0514		
Stangeria	0.1357	0.0695	0.0715	0.0768	0.0666	0.0448	0.0394	0.0432	0.0472	0.0570	

Copyright by Kasetsart University All rights reserved

Table 5 The number of base substitutions per site from averaging overall the *psb*M-*trn*D sequence pairs is shown. All results are based on the pairwise analysis of 45 sequences**.**

Genus	*Cycas*	*Dioon*	*Zamia*	*Chigua*	*Microcycas*	*Macrozamia*	*Lepidozamia*	*Encephalartos*	*Bowenia*	*Ceratozamia*	*Stangeria*
Cycas											
Dioon	0.1360										
Zamia	0.2379	0.1128									
Chigua	0.2803	0.1373	0.0478								
Microcycas	0.3018	0.1684	0.0709	0.1317							
Macrozamia	0.1987	0.0865	0.1555	0.2252	0.2396						
Lepidozamia	0.2029	0.0905	0.1366	0.2084	0.2314	0.0660					
Encephalartos	0.2379	0.0688	0.1237	0.1882	0.2155	0.0418	0.0374				
Bowenia	0.2530	0.1750	0.1648	0.1780	0.2373	0.1538	0.1500	0.1300			
Ceratozamia	0.2809	0.1462	0.2148	0.2996	0.2948	0.1679	0.1588	0.1456	0.2595		
Stangeria	0.2312	0.1565	0.1984	0.2575	0.2842	0.1556	0.1565	0.1380	0.2157	0.2296	

Copyright by Kasetsart University All rights reserved

Table 6 The number of base substitutions per site from averaging overall the *trn*L-*trn*F sequence pairs is shown. All results are based on the pairwise analysis of 45 sequences**.**

Genus	*Cycas*	*Dioon*	*Zamia*	*Chigua*	*Microcycas*	*Macrozamia*	*Lepidozamia*	*Encephalartos*	*Bowenia*	*Ceratozamia*	*Stangeria*
Cycas											
Dioon	0.1698										
Zamia	0.1646	0.1216									
Chigua	0.1739	0.1260	0.0131								
Microcycas	0.2158	0.1638	0.0803	0.0892							
Macrozamia	0.1595	0.0879	0.0961	0.1052	0.1296						
Lepidozamia	0.1585	0.0828	0.0963	0.1080	0.1353	0.0534					
Encephalartos	0.1636	0.0970	0.1049	0.1152	0.1415	0.0616	0.0521				
Bowenia	0.1555	0.0765	0.0851	0.0928	0.1252	0.0541	0.0622	0.0703			
Ceratozamia	0.1549	0.0852	0.0767	0.0905	0.1152	0.0609	0.0653	0.0713	0.0503		
Stangeria	0.1676	0.1111	0.0921	0.1037	0.1294	0.0938	0.0866	0.1415	0.0882	0.0817	

Copyright by Kasetsart University All rights reserved

Table 7 The number of base substitutions per site from averaging overall the *trn*S-*trnf*M sequence pairs is shown. All results are based on the pairwise analysis of 45 sequences**.**

Genus	*Cycas*	*Dioon*	*Zamia*	*Chigua*	*Microcycas*	*Macrozamia*	*Lepidozamia*	*Encephalartos*	*Bowenia*	*Ceratozamia*	*Stangeria*
Cycas											
Dioon	0.0724										
Zamia	0.1128	0.0717									
Chigua	0.1119	0.0729	0.0108								
Microcycas	0.0948	0.0496	0.0347	0.0366							
Macrozamia	0.1056	0.0948	0.0948	0.0937	0.0672						
Lepidozamia	0.0888	0.0255	0.0593	0.0605	0.0447	0.0447					
Encephalartos	0.0891	0.0325	0.0655	0.0667	0.0471	0.0500	0.0153				
Bowenia	0.0918	0.0436	0.0760	0.0777	0.0518	0.0691	0.0323	0.0373			
Ceratozamia	0.1162	0.0664	0.0866	0.0877	0.0709	0.0813	0.0605	0.0660	0.0697		
Stangeria	0.1301	0.0975	0.1123	0.1124	0.0948	0.1084	0.0904	0.1009	0.0982	0.0933	

Copyright by Kasetsart University All rights reserved

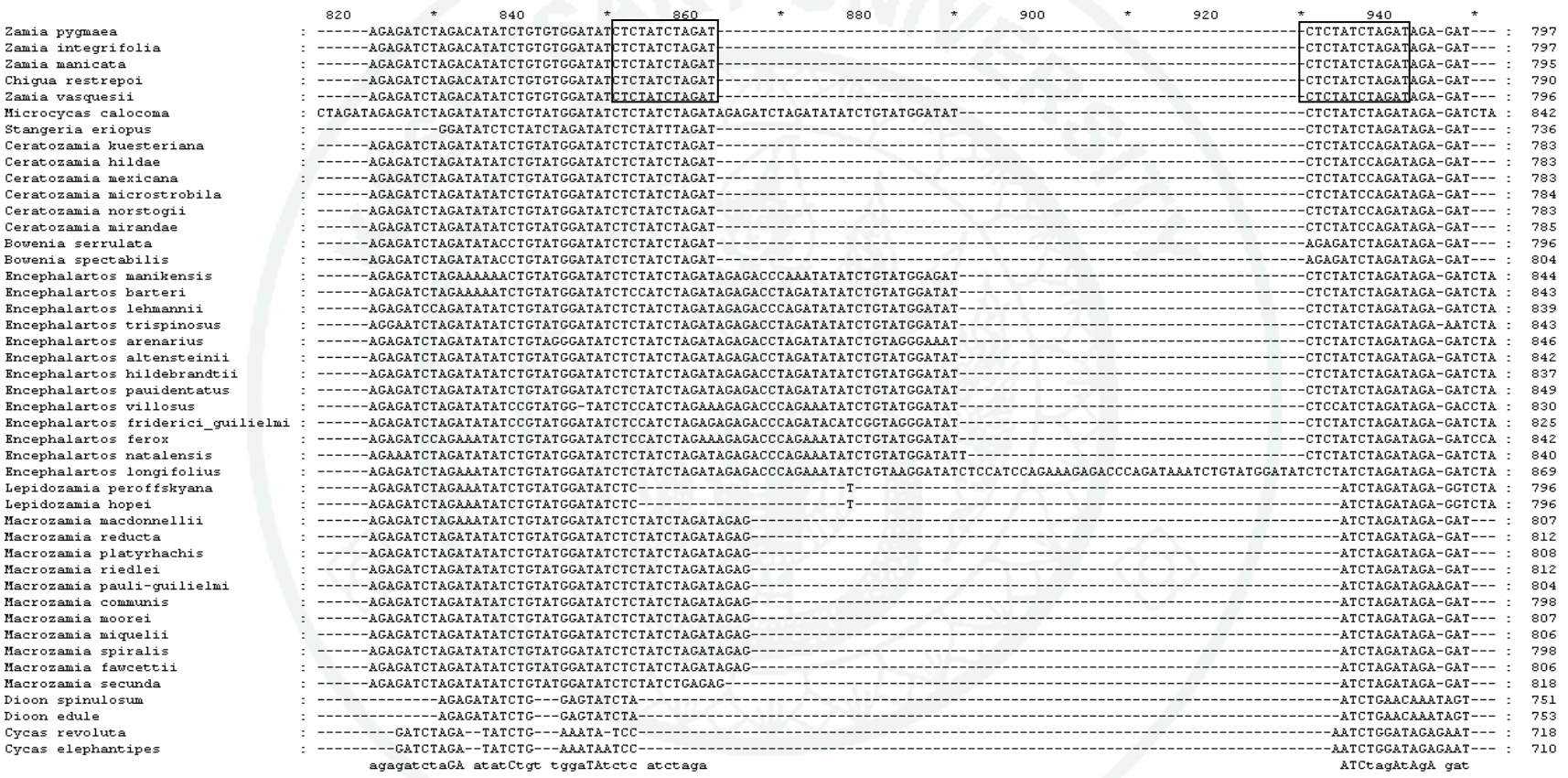


Figure 1 Partial alignment of chloroplast *trn*L-*trn*F region from 45 cycad species. The boxed region showed that *Chigua* and *Zamia* shared 12 bp (CTCTATCTAGAT) direct repeat sequences with flanking region.

Copyright by Kasetsart University All rights reserved

Table 8 Microsatellites containing the *trn*L-*trn*F region of cycad species are shown.

Specie	Tandem repeats
Z. pygmaea, *Z. integrifolia*	$TG(A)_{12}$
Z. vasquesii	$TG(A)_{11}$
Z. manicata	$TG(A)_{10}$
Mi. calocoma	$TG(A)_{22}$, $TCC(T)_{13}$
Ci. restrepoi	$TG(A)_{11}$
CZ. hildae CZ. kuesteriana *CZ. mexicana CZ. microstrobila* *CZ. mirandae CZ. norstogii*	$(TAGATA)_2GA(TAAG)_2$
E. paucidentatus	$AG(A)_{28}$, $(TAGATA)_2$gataagtaa$(AGAT)_3$, $TA(T)_{12}$
E. longifolius, *E. barteri*	$TG(A)_{25}$, $(TAGATA)_2$gataagtaa$(AGAT)_3$, $TA(T)_{15}$
E. arenarius	$AT(A)_{25}$, $(TAGATA)_2$gataagtaa$(AGAT)_3$, $TA(T)_{13}$
E. ferox	$AG(A)_{25}$, $TA(T)_{13}$
E. manikensis	$TG(A)_{23}$, $(TAGATA)_2$gataagtaa$(AGAT)_3$, $TA(T)_{15}$
E. trispinosus	$AG(A)_{23}$, $TA(T)_{12}$
E. lehmannii	$AG(A)_{23}$, $TA(T)_{11}$
E. altensteinii	$TG(A)_{22}$, $(TAGATA)_2$gataagtaa$(AGAT)_3$, $TA(T)_{10}$
E. hildebrandtii	$TG(A)_{22}$, $(TAGATA)_2$gataagtaa$(AGAT)_3$, $TA(T)_9$
E. friderici-guilielmi	$AG(A)_{22}$
E. natalensis	$AG(A)_{21}$, $(TAGATA)_2$gataagtaa$(AGAT)_3$, $TA(T)_{16}$
E. villosus	$TG(A)_{19}$, $TA(T)_{13}$
L. peroffskyana, *L. hopei*	$TG(A)_{25}$, $(TAGATA)_2$gataagtaa$(AGAT)_3$, $TA(T)_{12}$
M. reducta	$TG(A)_{26}$
M. secunda, *M. platyrhachis*, *M. macdonnellii*	$TG(A)_{25}$, $(TAGATA)_2$gataagtaa$(AGAT)_3$
M. moorei, *M. miquelii*	$TG(A)_{25}$
M. fawcettii	$TG(A)_{22}$, $(TAGATA)_2$gataagtaa$(AGAT)_3$
M. pauli-guilielmi	$TG(A)_{18}$, $(TAGATA)_2$gataagtaa$(AGAT)_3$
M. riedlei	$TG(A)_{12}G(A)_{16}$
M. spiralis, *M. communis*	$TG(A)_{15}$, $(TAGATA)_2$gataagtaa$(AGAT)_3$
D. edule	$TT(C)_8(TTT)_3TG(A)_{16}$
D. spinulosum	-
B. serrulata, *B. spectabilis*	$(TAGATA)_2$
S. eriopus	$(TCC)_2(T)_{10}$

Copyright by Kasetsart University All rights reserved

1.2 Phylogenetic analysis

1.2.1 *atp*B-*rbc*L intergenic spacer

NJ tree was constructed using Kimura 3-parameter plus Gamma (K81uf+G) model of nucleotide substitution and Gamma distribution shape parameter = 0.5816 (Figure 2A). *Zamia*, *Chigua*, *Microcycas*, *Stangeria* and *Ceratozamia* formed a polyphyletic group. *Zamia*, *Chigua* and *Microcycas* formed a group together with 100 % bootstrap support (BP), while *Macrozamia*, *Lepidozamia*, *Encephalartos* and *Bowenia* were grouped with low bootstrap support (65%). Furthermore, the placement of *Stangeria* as sister to *Ceratozamia* was strongly supported (BP=93%).

Maximum Parsimony analysis of the *atp*B-*rbc*L spacer (Figure 2B) yielded 76,605 trees of 330 length with a consistency index (CI) = 0.830 and retention index (RI) = 0.859. ML model was K81uf+G with base frequency of A = 0.2953, C=0.1770, G= 0.1954, T=0.3341**;** base substitution of A-C=1.000, A-G=1.8886, A-T=0.3915, C-G=0.3915, C-T=1.8886, G-T=1.0000 and Gamma distribution shape parameter = 0.5816 (Figure 2C). ML and MP analyses suggested that *Dioon* was a sister group (with >90% bootstrap support) to the remaining genera of Stangeriaceae and Zamiaceae. *Zamia*, *Chigua* and *Microcycas* grouped together with 100 % bootstrap support and *Stangeria* was closely related with *Ceratozamia* with 79% and 57% bootstrap support in MP and ML, respectively. In MP, *Macrozamia*, *Lepidozamia* and *Encephalartos* were placed within the same clade as a monophyletic group, while three genera in ML were unresolved.

Copyright by Kasetsart University All rights reserved

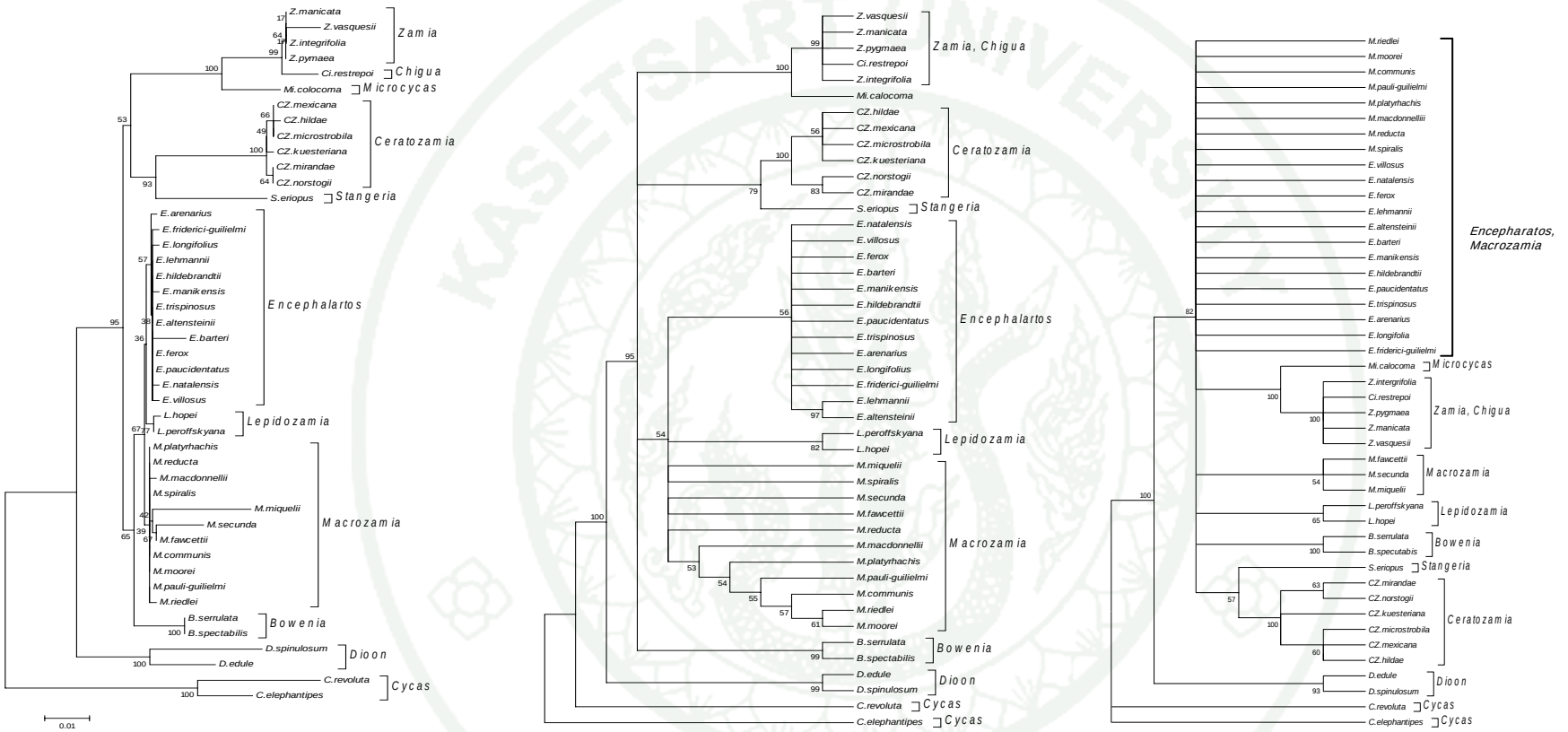


Figure 2 Phylogenetic trees construed by NJ (A), MP (B) and ML (C) from *atp*B-*rbc*L spacer sequence of 45 taxa of cycads

1.2.2 *psb*M-*trn*D intergenic spacer

NJ tree was constructed using Kimura 3-parameter plus Gamma (K81uf+G) model of nucleotide substitution and Gamma distribution shape parameter = 1.8292 (Figure 3A). Phylogenetic trees reconstructed by MP of the *psb*M-*trn*D non-coding region (Figure 3B) yielded 82,805 trees with the tree length of 1102 step, CI = 0.748 and RI = 0.823 and the number of informative sites was 365 (31.41%). ML model was K81uf+G with base frequency of A = 0.2981, C=0.2031, G= 0.1681, T=0.3308**;** base substitution of A-C=1.000, A-G=1.8995, A-T=0.5603, C-G=0.5603, C-T=1.8995, G-T=1.0000 and Gamma distribution shape parameter = 1.8292 (Figure 3C). All NJ, MP and ML produced similar tree topologies. Most genera were supported with > 50% bootstrap. *Dioon* was placed as a sister group to all other genera of Stangeriaceae and Zamiaceae. *Macrozamia*, *Bowenia*, *Lepidozamia* and *Encephalartos* were placed in the same clade supported with a >50% bootstrap value and *Bowenia* was a sister group to the other genera. Within this clade, low bootstrap support (68%) in ML and high bootstrap (92%) in MP and (81%) in NJ supported the tribe Encephalarteae as a monophyletic group. *Stangeria* was sister to the tribe Zamieae but weakly supported (BP<50%) in the ML and NJ trees, while in MP analyses this genus was sister to *Ceratozamia* (76%). Within tribe Zamieae, *Microcycas*, *Zamia* and *Chigua*, were grouped together with high bootstrap support (100%) in ML and MP and low bootstrap support (65%) in NJ as *Microcycas* was a sister group.

Copyright by Kasetsart University All rights reserved

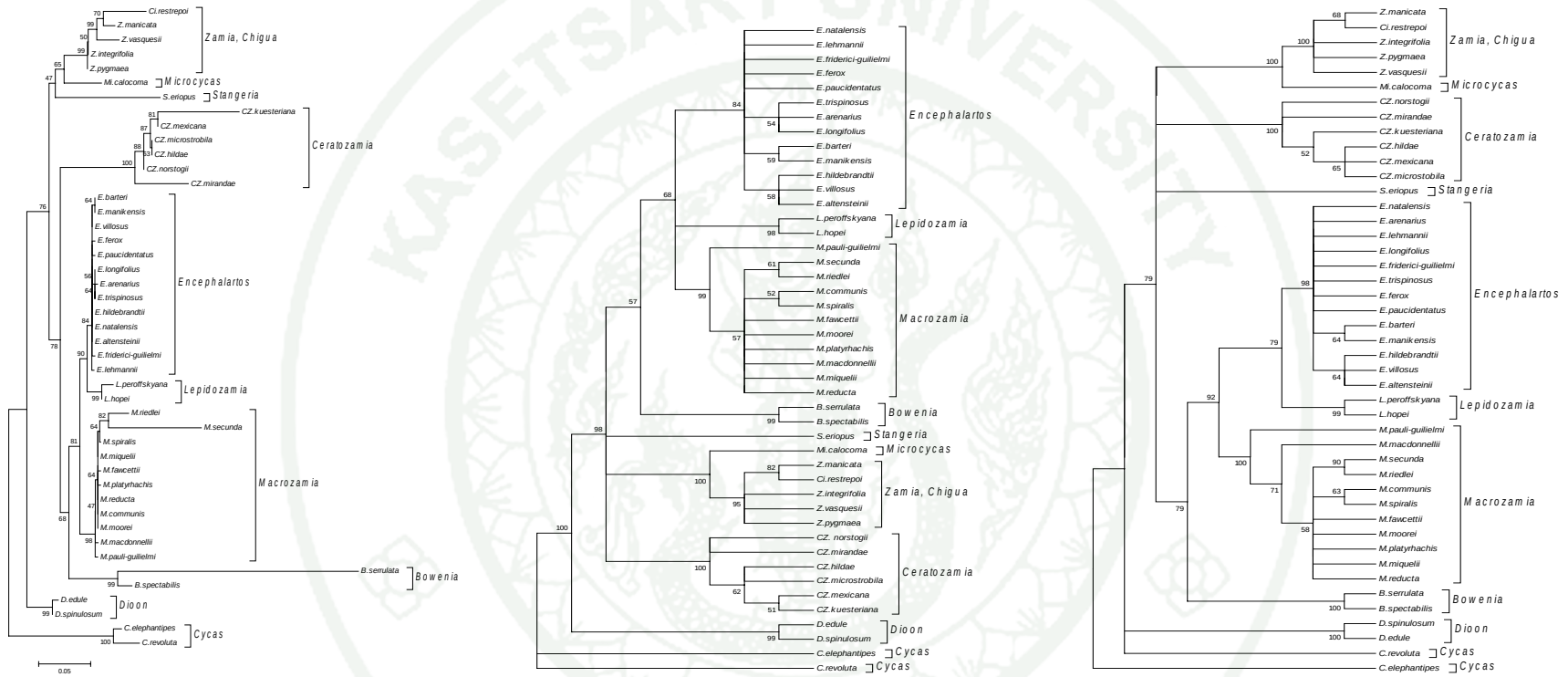


Figure 3 Phylogenetic trees construed by NJ (A), MP (B) and ML (C) from *psb*M-*trn*D spacer sequence of 45 taxa of cycads.

Copyright by Kasetsart University All rights reserved

1.2.3 *trn*L-*trn*F intergenic spacer

NJ tree was constructed using Kimura 3-parameter plus Gamma (K81uf+G) model of nucleotide substitution and Gamma distribution shape parameter = 0.9146 (Figure 4A). The MP analysis (Figure 4B) yielded 627,576 trees with tree length of 886 steps (CI = 0.733, RI = 0.806) and 310 (27.38%) informative sites. ML model was K81uf+G with base frequency of A = 0.2901, C=0.1658, G= 0.1993, T=0.3447**;** base substitution of A-C=1.000, A-G=1.8995, A-T=0.5603, C-G=0.5603, C-T=1.8995, G-T=1.0000 and Gamma distribution shape parameter = 0. 9146 (Figure 4C). The ML, MP and NJ trees revealed similar topology. Two major clades were identified with >50 % bootstrap support and *Dioon* as sister to the other genera. *Macrozamia*, *Lepidozamia*, *Encephalartos* and *Bowenia* were placed in the same clade (60% BP in NJ and 50% BP in MP), while *Bowenia* was a sister group of the remaining genera with the exception of *Dioon* in ML tree. The high bootstrap support (100%) indicated that *Zamia*, *Microcycas*, *Chigua* were closely related genera. *Stangeria* was placed in the New World clade of genera (*Zamia*, *Microcycas*, *Chigua* and *Ceratozamia*) with high bootstrap support in the ML tree (83%), moderate bootstrap support in the MP tree (76% BP) and poor bootstrap support in the NJ tree (< 50%). These results indicated that *Stangeria* and *Bowenia* were polyphyletic group.

Copyright by Kasetsart University All rights reserved

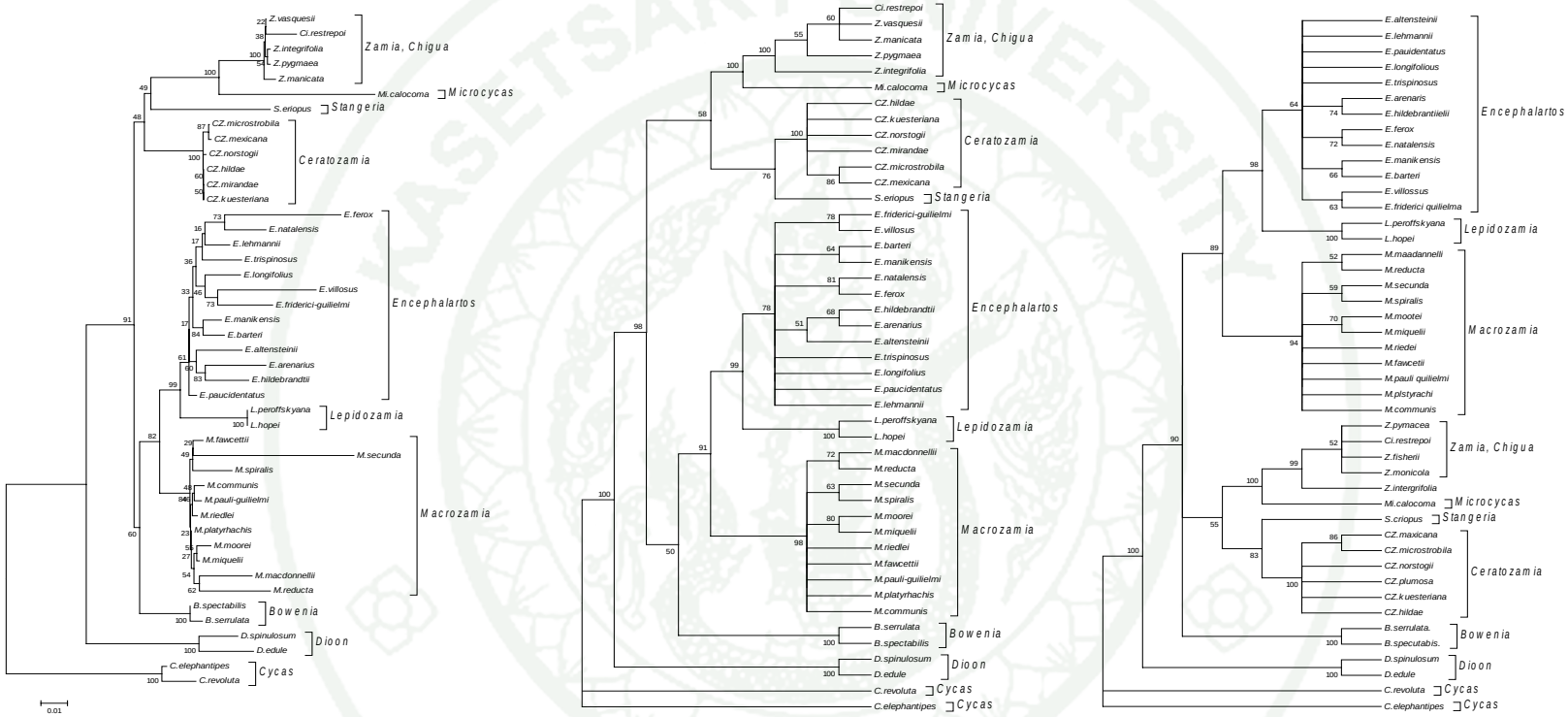


Figure 4 Phylogenetic trees construed by NJ (A), MP (B) and ML (C) from *trn*L-*trn*F spacer sequence of 45 taxa of cycads.

Copyright by Kasetsart University All rights reserved

1.2.4 *trn*S-*trnf*M region

NJ tree was constructed using Kimura 3-parameter plus Gamma (K81uf+G) model of nucleotide substitution and Gamma distribution shape parameter = 0.4225 (Figure 5A). The MP analysis (Figure 5B) yielded 59,100 trees with a tree length of 517 steps (CI = 0.708, RI = 0.855) where the number of informative sites was 203 (21.12%). ML model was K81uf+G with base frequency of A = 0.2901, C=0.1658, G= 0.1993, T=0.3447**;** base substitution of A-C=1.000, A-G=1.9781, A-T=0.4028, C-G=0.4028, C-T=1.9781, G-T=1.0000 and Gamma distribution shape parameter = 0.4225 (Figure 5C). ML analysis showed an identical tree topology compared to the NJ and MP analyses. Three analyses grouped *Dioon* as sister to the other genera (BP≥98%). The clade of *Macrozamia*, *Lepidozamia* and *Encephalartos* was monophyletic group. *Stangeria* and *Bowenia* were separated in different clade. *Bowenia* appeared as sister to sub-tribe Encephalartinae, whereas *Stangeria* was closely related to *Ceratozamia* with strong bootstrap support (89 % in MP, 87% in NJ 86% in ML). High bootstrap support (>90%) indicated that *Zamia* was inseparable from *Chigua*. In addition, *Zamia*, *Macrozamia* and *Chigua* were placed in the same group with high bootstrap support (99%).

Copyright by Kasetsart University All rights reserved

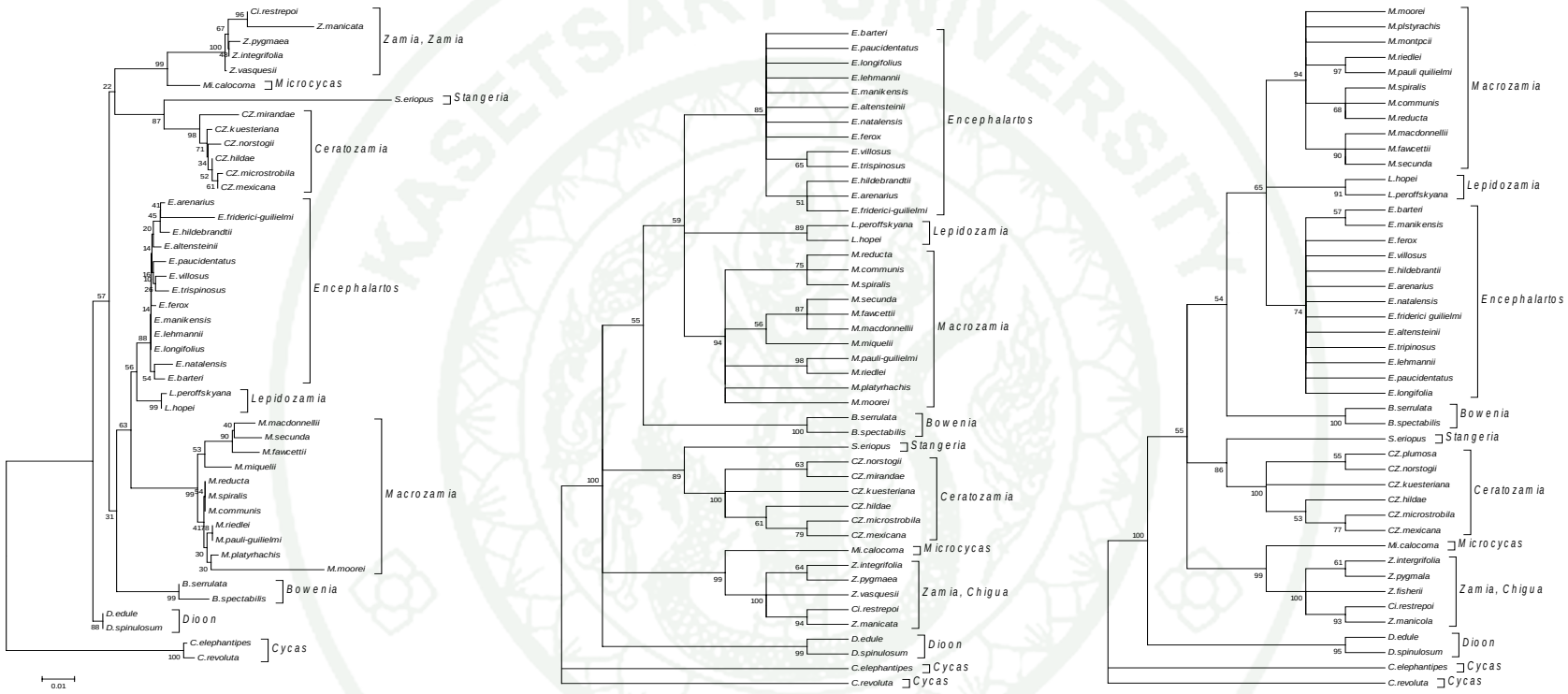


Figure 5 Phylogenetic trees construed by NJ (A), MP (B) and ML (C) from *trn*S-*trnf*M spacer sequence of 45 taxa of cycads.

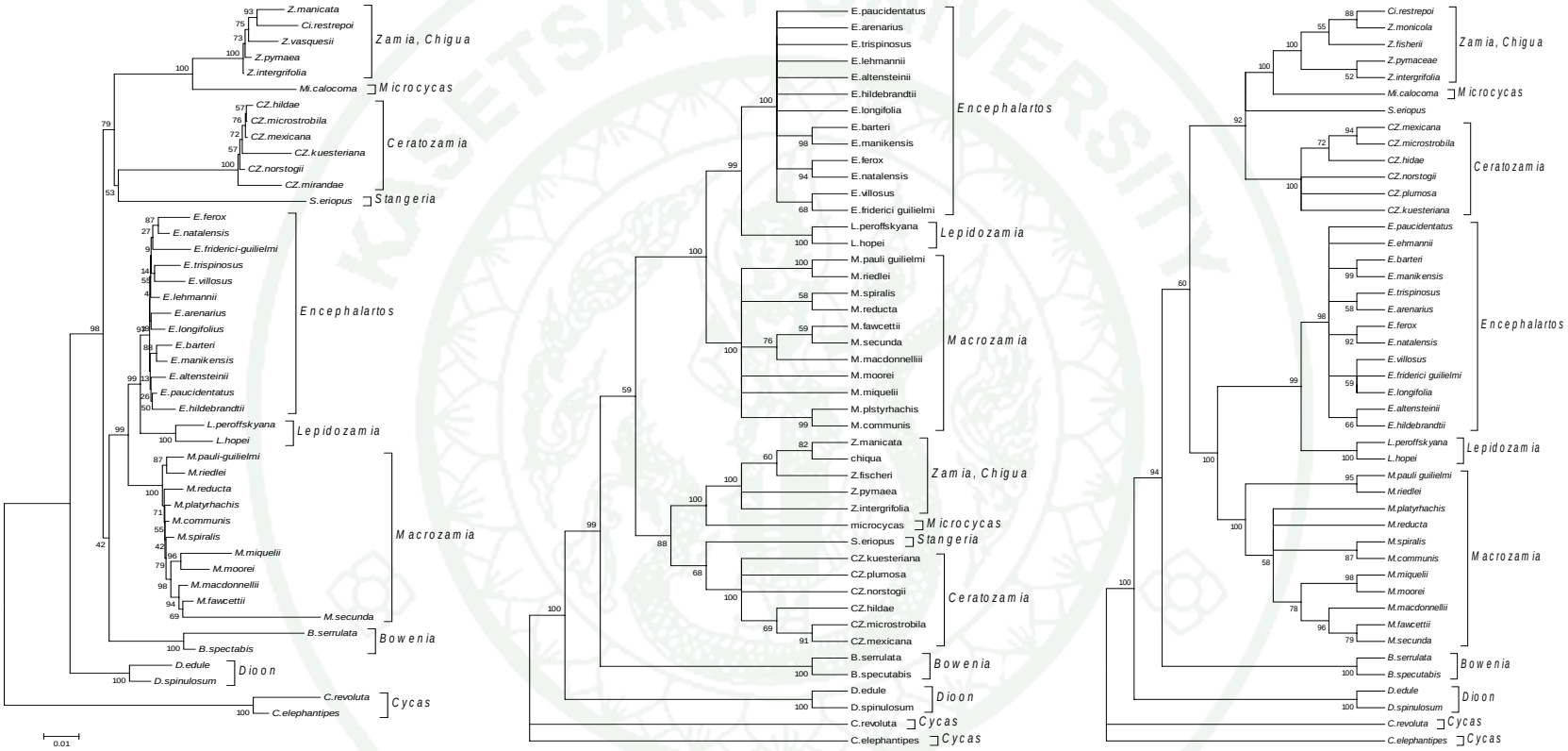
Copyright by Kasetsart University All rights reserved

1.2.5 Combined data of the *atp*B-*rbc*L, *psb*M-*trn*D, *trn*L-*trn*F and *trnS*-*trnf*M non-coding regions

An ILD test indicated that the data from the *atp*B-*rbc*L, *psb*M-*trn*D, *trn*L-*trn*F and *trnS*-*trnf*M non-coding regions were not significantly different (P= 0.7050). The average length of the combined data sequence was 3,279 bp and the alignment length was 3,697 bp. There were 900 informative sites (25.4%), 62.8 % (AT), 37.2% (GC) and the average number of substitutions per site was 0.0686.

NJ tree was constructed using Kimura 3-parameter plus Gamma (K81uf+G) model of nucleotide substitution and Gamma distribution shape parameter = 0.8762 (Figure 6A). The MP analysis generated 49,564 trees with 2,799 steps (CI = 0.748 and RI = 0.829). ML model (Figure 6B) was K81uf+G with base frequency of A = 0.2959, C=0.1803, G= 0.1893, T=0.3345**;** base substitution of A-C=1.000, A-G=1.9597, A-T=0.6029, C-G=0.6029, C-T=1.9597, G-T=1.0000 and Gamma distribution shape parameter = 0.8762 (Figure 6C). The ML, MP and NJ trees inferred from combining the four regions of non-coding chloroplast DNA provided higher resolution than the individual non-coding regions and revealed similar topologies in some clades of the *trn*L-*trn*F phylogenetic tree. Two clades were identified in three analyses; the first clade was subfamily Zamioideae and Stangeria with ≥ 79% bootstrap support. The second clade was tribe Encephalateae with high bootstrap support. All trees, the high bootstrap support (≥98%) indicated that *Bowenia* was placed as a sister group to all taxa with the exception of *Dioon*. *Bowenia* was a sister group of Zamiaceae family. The first clade, *Stangeria* was grouped with *Ceratozamia* in MP and NJ trees (bootstrap support >50%) while *Stangeria* was closely related with *Zamia*, *Chigua* and *Microcycas* in ML. In the second clade, *Lepidozamia* was closest to *Encephalartos* (99% bootstrap support) *Macrozamia* appeared as the sister group with BP≥99%. In all analyses the tribes Encephalateae (*Encephalartos*, *Macrozamia* and *Lepidozamia*) and Zamieae (*Zamia*, *Chigua* and *Microcycas*) formed monophyletic groups (BP ≥ 99%).

Copyright by Kasetsart University All rights reserved



(A) (B) (C)

Figure 6 Phylogenetic trees construed by NJ (A), MP (B) and ML (C) from combined data of 45 taxa of cycads.

Copyright by Kasetsart University All rights reserved

2. Phylogenetic relationships within the family Cycadaceae

Statistic analyses for NJ, MP and ML of each non-coding region and combined data sets are summarized in Table 9.

2.1 Nucleotide sequences and variation

2.1.1 *trn*S-*trn*G intergenic spacer

The length of the *trn*S-*trn*G region was variable among Cycadaceae family from 892-925 bp. This region consisted of 948 aligned nucleotides, of which 192 were variable and 144 were parsimony informative. The GC content was 36.8% and AT was 63.2%. The average number of substitution per site was 0.0295. Fourteen bp insertions (GTAAGAATCTATAA) were found at position 203-216 in section *Indosinenses*, whereas section *Cycas* (*C. edentata*, *C. seemannii, C. thouarsii, C. zeylanica* and *C. rumphii*) showed ATAAGAATCTATAA at the same position (Figure 7). Interestingly, the difference in the insertions between section *Indosinenses* and *Cycas* was found only one base transition, i.e., G↔A at position 203 except those of *C. bougainvilleana*, *C. media*, *C. pranburiensis*, *C. semota* whose sequence was exactly the same as those of *Indosinenses***.** In addition, the section *Indosinenses* and *Cycas* showed 2 minisatellite repeats (TCTATAAGTAAGAA). Section *Asiorientales* (*C. taitungensis* and *C. revoluta*), on the other hand, shared 2 microsatellite repeats (TTTAC) and they had 5 bp insertions at position 161-165. The *trn*S-*trn*G alignment of this section also provided 46 base substitution sites, two deletions and one insertion which indicated the high genetic diversity in section *Asiorientales*.

2.1.2 *psb*M-*trn*D intergenic spacer

The length of the *psb*M-*trn*D ranged from 978 bp in *C. rumphii* to 984 bp in *C. changjiangensis*. The sequence alignment was 1,004 nucleotide characters including 218 variation sites, of which 146 were parsimony informative

Copyright by Kasetsart University All rights reserved

sites. The CG content was 62.1 % and AT was 37.9% with the average number of substitution per site of 0.0308. This sequence did not show indel position. However, 5 base substitution sites were detected in section *Asiorientales*, 3 base substitution sites in section *Indosinenses* and *Cycas* and 1 base substitution site in section *Wadeanae*. All base substitution sites in section *Indosinenses*, *Cycas* and *Wadeanae* were transitions (i.e., A↔G and T↔C), while the section *Asiorientales* showed both transitions and transversions (i.e., T↔C, T↔G and A↔C).

2.1.3 *trn*L-*trn*F intergenic spacer

The length of the *trn*L-*trn*F was widely distributed ranging from 864 bp in *C. pranburiensis* to 911 bp in *C. bougainvilleana*. The sequence alignment was 993 bp. There were 186 variable sites, including 112 parsimony informative sites. The GC content was 62.7% and AT was 37.3%. The average number of substitution per site was 0.0265. This region represented 4 base substitution sites. One base substitution site was transition in section *Asiorientales* and one base substitution site in section *Indosinenses* and *Cycas* also showed transitions. Three species, *C. edentata*, *C. thouarsii* and *C. zeylanica*, in section *Cycas* had 2 base substitution sites which appeared transitions and transversions pattern.

2.1.4 *trn*S-*trnf*M region

The *trn*S-*trnf*M region contained three genes, trnG^{GCC}, ycf9 and psbZ. The length of this region was 929-937 bp. The sequence alignment was 1,049 bp. There were 110 variable sites; including 89 were parsimony informative sites. This sequence contained 62 % AT and 38 % GC with the low of the average number of substitution per site of 0.0158. Nevertheless, it showed 6 base substitution sites while 4 base substitution sites (transitions and transversions) were detected in section *Asiorientales* and 2 base substitution sites (transitions and transversions) were shown in *C. media* and *C. semota* of section *Cycas* (Table 10).

Copyright by Kasetsart University All rights reserved

Table 9 Characterization of non-coding region chloroplast DNA in *Cycas* taxa.

	trnS-trnG	*psb*M-*trn*D	*trn*L-*trn*F	*trn*S- *trnf*M	combined
No. of taxa	27	27	27	27	27
Aligned sequences length	948	1,004	993	1,049	2,944
No. of parsimony informative character	144	146	112	89	385
% of parsimony informative character	15.18	14.54	11.27	8.48	12.85
No. of MP trees	95720	15,413	89,863	76,298	95,870
Length of MP trees	234	243	199	124	702
Consistency index (CI)	0.9145	0.9465	0.0905	0.9113	0.8989
Retention index (RI)	0.8930	0.9363	0.8657	0.900	0.8704
Model	K81uf+G[1]	HKY+G[2]	K81uf+G[1]	HKY+G[2]	K81uf+G[1]
Gamma shape parameter	0.4393	1.0012	0.5231	0.2635	0.5470
A+T%	63.2	62.1	62.7	62	62.7
G+C%	36.8	37.9	37.3	38	37.3
Transitions/Transversions	1.063	1.232	1.21	0.917	0.95
The average number of substitution per site	0.0295	0.0308	0.0265	0.0158	0.0291

1 = Unequal-frequency Kimura 3-parameter plus Gamma

2 = Hasegawa-Kishino-Yano plus Gamma

Copyright by Kasetsart University All rights reserved

2.2 Phylogenetic analysis

2.2.1 *trn*S-*trn*G intergenic spacer

NJ tree was constructed using Kimura 3-parameter plus Gamma (K81uf+G) model of nucleotide substitution and Gamma distribution shape parameter = 0.4393. The tree topology indicated *Cycas* was monophyletic. NJ analysis showed that the 5 *Cycas* section formed two separate clades (Figure 8A). The first clade contained 5 *Cycas* section (*Asiorientales*, *Indosinenses*, *Cycas*, *Stangerioides* and *Wadeanae*) and represented 3 sub-clades. *C. praburiensis* and *C. bougainvilleana* were placed in the same sub-clade with low bootstrap support (66%) as well as *C. wadei* and *C. curranii*, *C. condaoensis* and *C. nongnoochiea* also clustered together with 65% and 59% bootstrap support, repectively. High bootstrap support (99%) indicated that *C. taitungensis* was closely related to *C. revoluta* and *C. seqmentilida* was a sister group. The second clade was 5 species of section *Cycas*, *C. edentata*, *C. rumphii*, *C. seemannii*, *C. thouarsii* and *C. zelanica*, which grouped as sister to the first clade.

MP analyses produced 95,720 trees with the length of 234 steps, a consistency index (CI) of 0.9145 and a retention index (RI) of 0.8930. The strict consensus tree derived from the MP trees is shown in Figure 8B. *C. wadei* and *C. curranii* were clustered together with 61%. In addition were clustered together *C. taitungensis* was closely related to *C. revoluta* with 100% bootstrap support and *C. seqmentilida* was a sister group.

ML model was K81uf+G with base frequency of A = 0.3036, C=0.1692, G= 0.2058, T=0.3214**;** base substitution of A-C=1.000, A-G=1.8384, A-T=0.5645, C-G= 0.5645, C-T=1.18364, G-T=1.0000 and Gamma distribution shape parameter = 0.4393. The ML topology was similar to NJ. ML tree could be divided into two clades (Figure 8C). The first clade comprised of five sections, *Asiorientales*, *Indosinenses*, *Cycas*, *Stangerioides* and *Wadeanae*. With the exception of five species in section *Cycas*, i.e., *C. edentata*, *C. rumphii*, *C. seemannii*, *C. thouarsii* and *C.*

Copyright by Kasetsart University All rights reserved

zelanica, formed the second clade. The first clade could be further divided into 3 sub-clades. *C. praburiensis* - *C. bougainvilleana* sub-clade was 55% bootstrap support. *C. wadei* and *C. curranii* were clustered together with 61%. *C. taitungensis* was closely related to *C. revoluta* and *C. seqmentilida* was a sister group.

2.2.2 *psb*M-*trn*D intergenic spacer

NJ tree of the *psb*M-*trn*D was constructed using K81uf+G model and Gamma distribution shape parameter = 1.0012. NJ tree was divided into two major clades (Figure 9A). The first clade included two sections (*Indosinenses* and *Cycas*). This clade could be further separated into two sub-clades. *C. clivicola*, *C. elephantipes*, *C. condaoensis* and *C. siamensis* grouped together with low bootstrap support (58%). *C. thou*arsii was grouped with *C. seemannii* with 70% bootstrap support and *C. bougainvilleana* was a sister group. The second clade consisted of sections *Asiorientales*, *Stangerioides* and *Wadeanae*. This clade was divided into three sub-clades. The high bootstrap support showed that *C. taitungensis* was closely related to *C. revoluta* as well as *C. curranii* and *C. wadei*. On the other hand, *C. hainanensis* clustered together with *C. changjiangensis* with low bootstrap support.

MP analyses produced 15,413 trees with the length of 243 steps, a consistency index (CI) of 0.9465 and a retention index (RI) of 0.9363. The strict consensus tree derived from the MP trees is shown in Figure 9B. MP analysis was separated into two major clades. The first clade consisted of section *Indosinenses* and *Cycas* and it could be divided into two sub-clades. *C. noongnoochiae* and *C. bougainvilleana* were grouped together with low bootstrap support the same as *C. thouarsii* and *C. seemannii*. The second clade contained sections *Asiorientales*, *Stangerioides* and *Wadeanae*. Within the second clade, there were two well-support groups; one consisting of *C. revoluta* and *C. taitungensis* and another consisting of *C. wadei* and *C. curranii*, while *C. hainanensis* clustered together with *C. changjiangensis* with low bootstrap support.

Copyright by Kasetsart University All rights reserved

ML model was HKY+G with base frequency of A = 0.2806, C=0.2069, G= 0.1790, T=0.3335 and Gamma distribution shape parameter = 1.0012. ML was similar in topology to MP but the relationships of some species were resolved. ML tree was divided into two major clades (Figure 9C). The first consisted of sections *Indosinenses* and *Cycas*. The relationship within this clade was unresolved with the exception of *C. thouarsii* and *C. seemannii* which were closely related species with 63% bootstrap support. Sections *Asiorientales*, *Stangerioides* and *Wadeanae* formed a second clade. Within the second clade, the bootstrap was still high that it was 97% and 84% for *C. revoluta* - *C. taitungensis* sub-clade and *C. wadei* - *C. curranii* sub-clade, respectively. Nevertheless, *C. hainanensis* clustered together with *C. changjiangensis* with low bootstrap support (66%).

2.2.3 *trn*L-*trn*F intergenic spacer

NJ tree was constructed using K81uf+G model and Gamma distribution shape parameter = 0.5231. The NJ analysis showed that 5 *Cycas* sections could be divided into two clades (Figure 10A). The first clade included 2 sections, *Indosinenses* and *Cycas*. In this clade, *C. edentata* and *C. thouarsii* were grouped with 65% bootstrap support and they clustered with *C. zeylanica* having *C. seemannii* as a sister group. Furthermore, *C. semota* was closely related with *C. media* with 62% bootstrap support. The second clade consisted of the three remaining section, *Asiorientales*, *Stangerioides* and *Wadeanae*. Although the resolution in this clade was slightly lower (<50%), *C. revoluta* and *C. taitungensis* grouped together with 66% bootstrap support.

MP analysis produced 89,863 trees with the length of 218 steps, a consistency index (CI) of 0.269 and a retention index (RI) of 0.8806. The strict consensus tree derived from the MP trees is shown in Figure 10B. MP tree generated a similar topology with NJ tree. MP tree also was divided into two clades. The first main clade included sections *Indosinenses* and *Cycas*. Within this clade, *C. edentata*, *C. thouarsii* and *C. zeylanica* were grouped together with 84% bootstrap support and *C. semota* was closely related with C. *media* with 62% bootstrap support. The second

Copyright by Kasetsart University. All rights reserved

clade was sections *Asiorientales*, *Stangerioides* and *Wadeanae*. The second clade included two sub-clades. The first sub-clade consisted of *C. revoluta* and *C. taitungensis* (78% bootstrap support), while the second sub-clade consisted of *C. segmentifida* and *C. sexsiminifera*.

ML model was K81uf+G with base frequency of A = 0.3064, C=0.1783, G= 0.1896, T=0.3258**;** base substitution of A-C=1.000, A-G=2.5972, A-T=0.5056, C-G= 0.5056, C-T=2.5972, G-T=1.0000 and Gamma distribution shape parameter = 0.523. ML tree could be separated into two clades (Figure 10C). The first clade consisted of section *Indosinenses* and *Cycas*, while the second clade was composed of section *Asiorientales*, *Stangerioides* and *Wadeanae*. In the first clade, it was shown that the ML topology was similar to MP tree. However, the second clade had one sub-clade differed from MP that *C. changjiangensis* and *C.* hainanensis were placed to the same sub-clade with low bootstrap support.

2.2.4 *trnS- trnf*M region

NJ tree was constructed using HKY+G model and Gamma distribution shape parameter = 0.2635. The NJ analysis showed that all 5 *Cycas* sections formed a single clade (Figure 11A). Three species of section *Cycas* (*C. rumphii*, *C. media* and *C. semota*) were grouped with one species of section *Indosinenses* (*C. condaoensis*). *C. elephan*tipes was closely related to *C. clivicola* with 62% bootstrap support. The high bootstrap support indicated that *C. revoluta* was closely related to *C. taitungensis* as well as *C. hainanensis* and *C. changjiangensis*.

MP analysis produced 79,298 trees with the length of 124 steps, a consistency index (CI) of 0.9913 and a retention index (RI) of 0.9000. The strict consensus tree derived from the MP trees is shown in Figure 11B. MP tree differed from NJ tree. Sections *Indosinenses*, *Cycas*, *Stangerioides* and *Wadeanae* formed a group together, while section *Asiorientales* seemed to be a sister group. In addition,

Copyright by Kasetsart University All rights reserved

there was only one sub- clade (*C. hainanensis* and *C. changjiangensis*) with high bootstrap support.

ML model was HKY+G with with base frequency of A = 0.2875, C=0.1857, G= 0.1948, T=0.3320 and Gamma distribution shape parameter = 0.2635. ML tree generated a similar topology with NJ tree (Figure 11C). *C. rumphii* and *C. condaoensis* were placed in the same sub-clade and *C. media* was a sister group. *C. revoluta* and *C. taitungensis* formed sub-clade with 76% bootstrap support as well as *C. hainanensis* and *C. changjiangensis*. *C. elephantipes* was grouped with *C. clivicola* but they showed a low bootstrap support (50%).

2.2.5 Combined data

ILD test showed that four non-coding data was incongruent. Three data sets were selected (*trn*S-*trn*G, *psb*M-*trn*D and *trn*L-*trn*F) but they did not differ significantly in structure (ILD = 0.16). Therefore, they were combined into a single data set for phylogenetic analysis. The combined data consisted of 2,944 aligned nucleotides, of which 573 were variable sites and 385 were parsimony informative.

NJ analysis was constructed using K81uf+G and Gamma distribution shape parameter = 0.5470. The strict consensus of combined 3 regions supported the monophyly of genus *Cycas*. NJ tree could be divided into two main clades (Figure 12A). The first clade contained sections *Indosinenses* and *Cycas*. Within this clade, there were 2 sub-clades with bootstrap support > 50%. Five species of section *Cycas*; *C. rumphii*, *C. edentata*, *C. zeylanica*, *C. seemannii* and *C. thouarsii*, were grouped together with 59% bootstrap support. In addition, *C. elephantipes* was closely related to *C. siamensis*. The second clade contained 3 sections, *Asiorientales*, *Stangerioides* and *Wadeanae*. This clade had 3 strongly support sub-clades. The first sub-clade was *C. revoluta* and *C. taitungensis*. *C. curranii* and *C. wadei* formed a strongly support sub-clade and *C. sexsiminifera* appeared basal to this group. The third sub-clade was *C. changjiangensis* and *C. hainanensis* with less well support.

Copyright by Kasetsart University All rights reserved

MP analysis produced 89,863 trees with a length of 218 steps, a consistency index (CI) of 0.8989 and a retention index (RI) of 0.8704. *Cycas* was monophyletic (bootstrap support = 100%) and resolved into the similar two major clades to NJ tree with the exception of two species of section *Asiorientales* (*C. revoluta* and *C. taitungensis*) that they were a basal group (Figure 12B). The first clade was sections *Indosinenses* and *Cycas*. It showed only one sub-clade of *C. thouarsii* forming a group with *C. seemannii* while *C. rumphii*, *C. edentata* and *C. zeylanica* were a sister group. The second clade consisted of 2 sections of *Stangerioides* and *Wadeanae*. Within this clade, *C. wadei* was closely related to *C. curranii* with high bootstrap support as well as *C. changjiangensis* and *C. hainanensis*.

ML model was K18uf+G with base frequency of A = 0.2950, C=0.1874, G= 0.1894, T=0.3282**;** base substitution of A-C=1.000, A-G=2.3796, A-T=0.5660, C-G= 0.5660, C-T=2.3796, G-T=1.0000 and Gamma distribution shape parameter = 0.5978. The ML topology was also incongruent with MP and only two clades were recognized (Figure 12C). Sections *Indosinenses* and *Cycas* formed the first clade and the second clade included all the remaining sections. Although the relationship among species within the first clade was unresolved, five species of section *Cycas* (*C. rumphii*, *C. edentata*, *C. zeylanica*, *C. seemannii* and *C. thouarsii*) were grouped together with 65% bootstrap support. In contrast to the second clade, there were 3 sub-clades, *C.wadei* - *C. curranii* sub-clade, *C. changjiangensis*-*C. hainanensis* sub-clade, *C.revoluta*-*C. taitungensis* sub-clade, with high bootstrap support.

Copyright by Kasetsart University All rights reserved

Table 10 Base substitution sites in four non-coding regions of chloroplast DNA of 27 *Cycas* species.

species	*trn*S-*trn*G (position)				*psb*M-*trn*D (position)				*trn*L-*trn*F (position)			*trn*S- *trnf*M region (position)					
	169	212	291	462	524	654	687	965	286	291	392	529	614	713	796	951	987
C. chamaoensis	T	G	G	G	C	C	C	A	A	T	A	T	G	C	C	A	T
C. clivicola	T	G	G	G	T	C	C	A	A	T	A	T	G	C	C	A	T
C. condaoensis	T	G	G	G	T	C	C	A	A	T	A	T	G	C	C	A	T
C. elephantipes	T	G	G	G	T	C	C	A	A	T	A	T	G	C	C	A	T
C. lindstromii	T	G	G	G	T	C	C	A	A	T	A	T	G	C	C	A	T
C. nongnoochiae	T	G	G	G	T	C	C	A	A	T	A	T	G	C	C	A	T
C. siamensis	T	G	G	G	T	C	C	A	A	T	A	T	G	C	C	A	T
C. tansachana	T	G	G	G	T	C	C	A	A	T	A	T	G	C	C	A	T
C. bougainvilleana	T	G	G	G	T	C	C	A	A	T	A	T	G	C	C	A	T
C. media	T	G	G	G	T	C	C	A	A	T	A	T	G	C	C	C	C
C. semota	T	G	G	G	T	C	C	A	A	T	A	T	G	C	C	C	C
C. pranburiensis	T	G	G	G	T	C	C	A	A	T	A	T	G	C	C	A	T
C. edentata	T	A	A	A	T	C	C	A	C	G	A	T	G	C	C	A	T
C. zeylanica	T	A	A	A	T	C	C	A	C	G	A	T	G	C	C	A	T
C. thouarsii	T	A	A	A	T	C	C	A	C	G	A	T	G	C	C	A	T
C. seemannii	T	A	A	A	T	C	C	A	A	T	A	T	G	C	C	A	T
C. rumphii	T	A	A	A	T	C	C	A	A	T	G	T	G	C	C	A	T
C. balansae	G	-	G	G	C	T	T	G	A	T	G	T	G	C	C	A	T
C. changjiangensis	G	-	G	G	C	T	T	G	A	T	G	T	G	C	C	A	T
C. hainanensis	G	-	G	G	C	T	T	G	A	T	G	T	G	C	C	A	T
C. hoabinhensis	G	-	G	G	C	T	T	G	A	T	G	T	G	C	C	A	T
C. segmentifida	G	-	G	G	C	T	T	G	A	T	G	T	G	C	C	A	T
C. sexseminifera	G	-	G	G	C	T	T	G	A	T	G	T	G	C	C	A	T
C. revoluta	G	-	G	G	C	T	T	G	A	T	G	C	T	A	T	A	T
C. taitungensis	G	-	G	G	C	T	T	G	A	T	G	C	T	A	T	A	T
C. curranii	G	-	G	G	C	T	T	G	A	T	G	T	G	C	C	A	T
C. wadei	G	-	G	G	C	T	T	G	A	T	G	T	G	C	C	A	T

Copyright by Kasetsart University All rights reserved

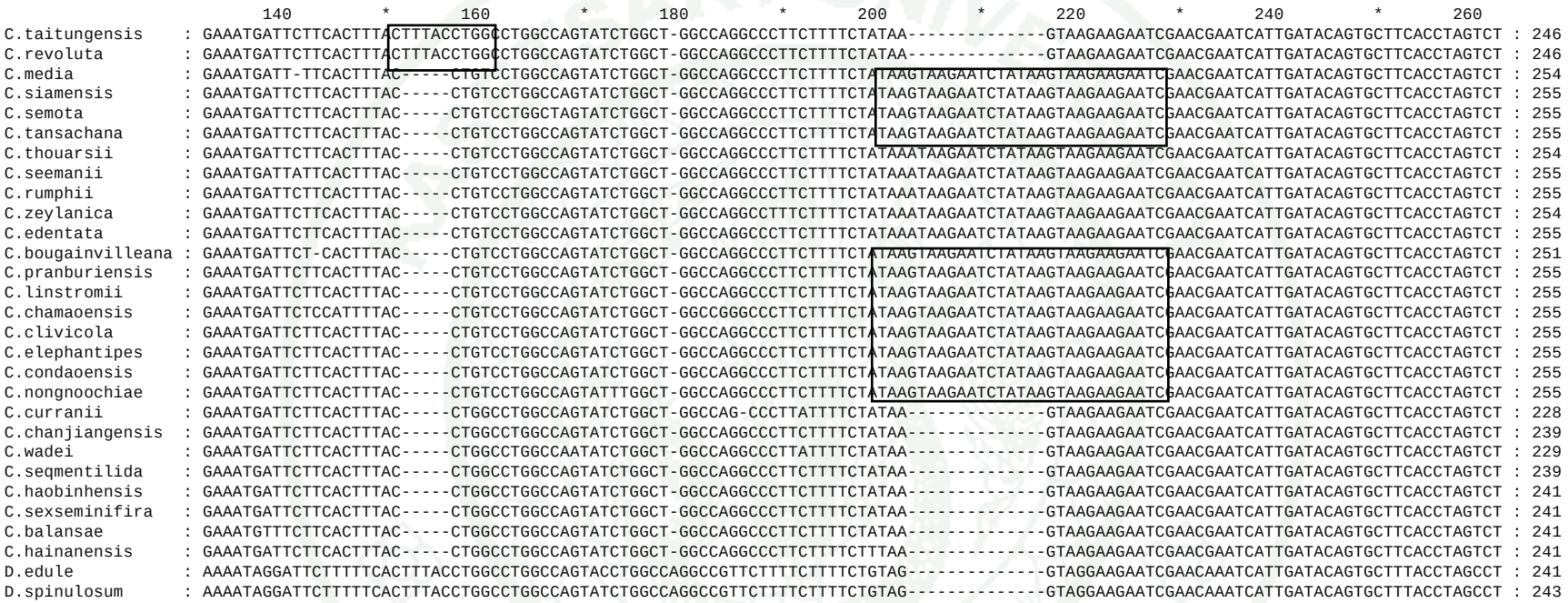


Figure 7 Partial alignment of chloroplast *trn*S-*trn*G region from 27 *Cycas* species. The boxed region show that minisatellite represent in section *Cycas* and *Indosinenses*, while microsatellite appear in section *Asiorientales*.

Copyright by Kasetsart University All rights reserved

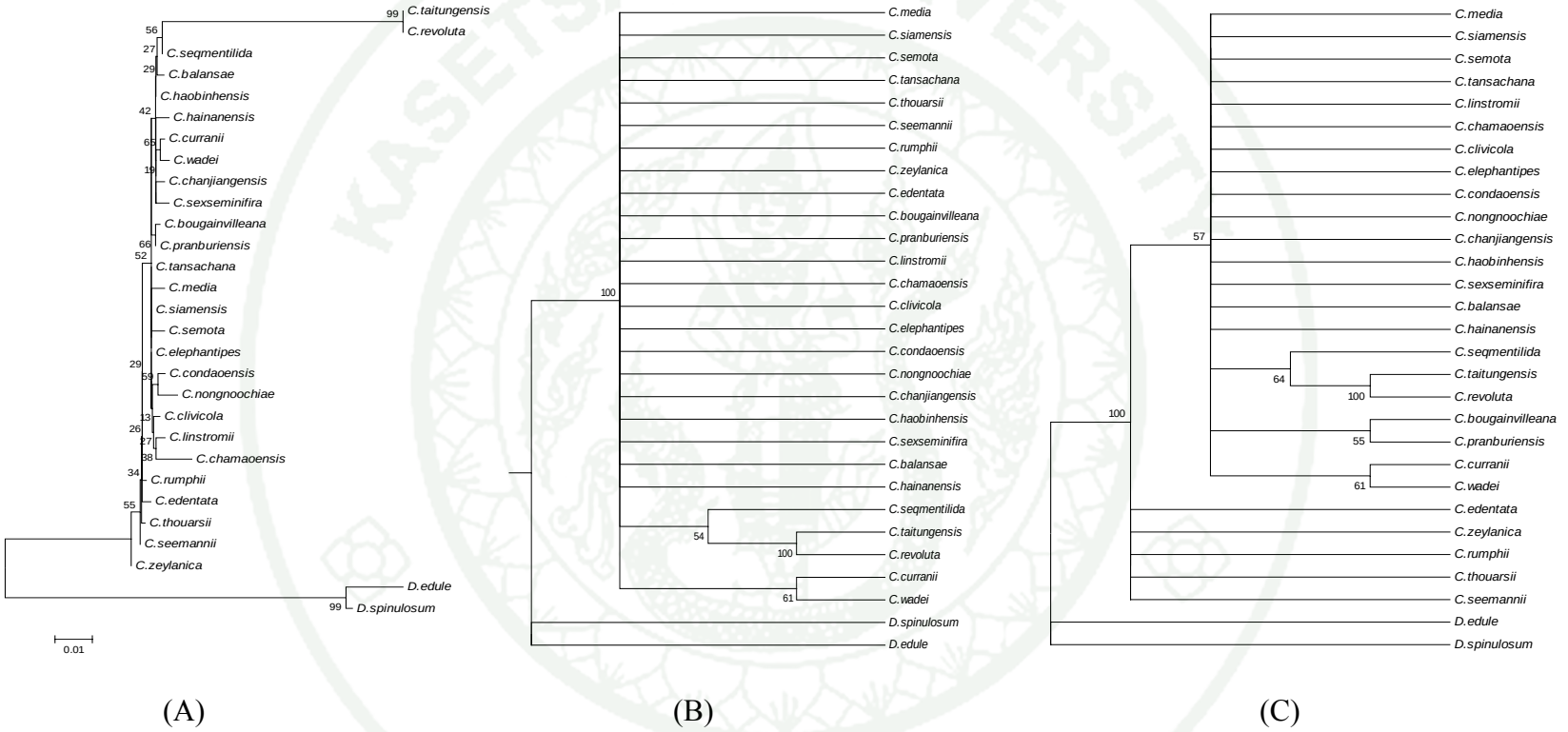


Figure 8 Phylogenetic trees construed by NJ (A), MP (B) and ML (C) from *trn*S-*trn*G spacer sequence of *Cycas* species.

Copyright by Kasetsart University All rights reserved

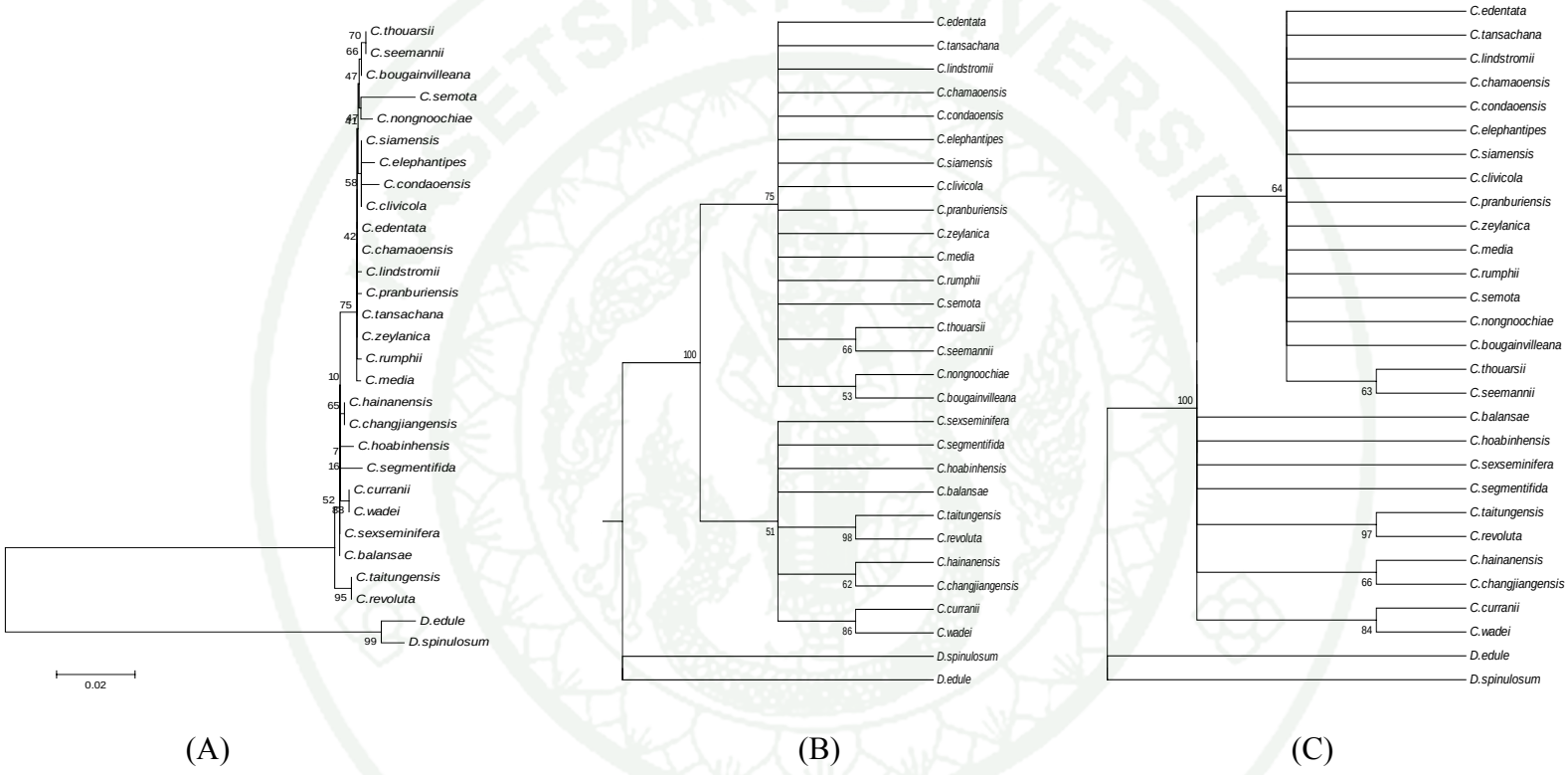
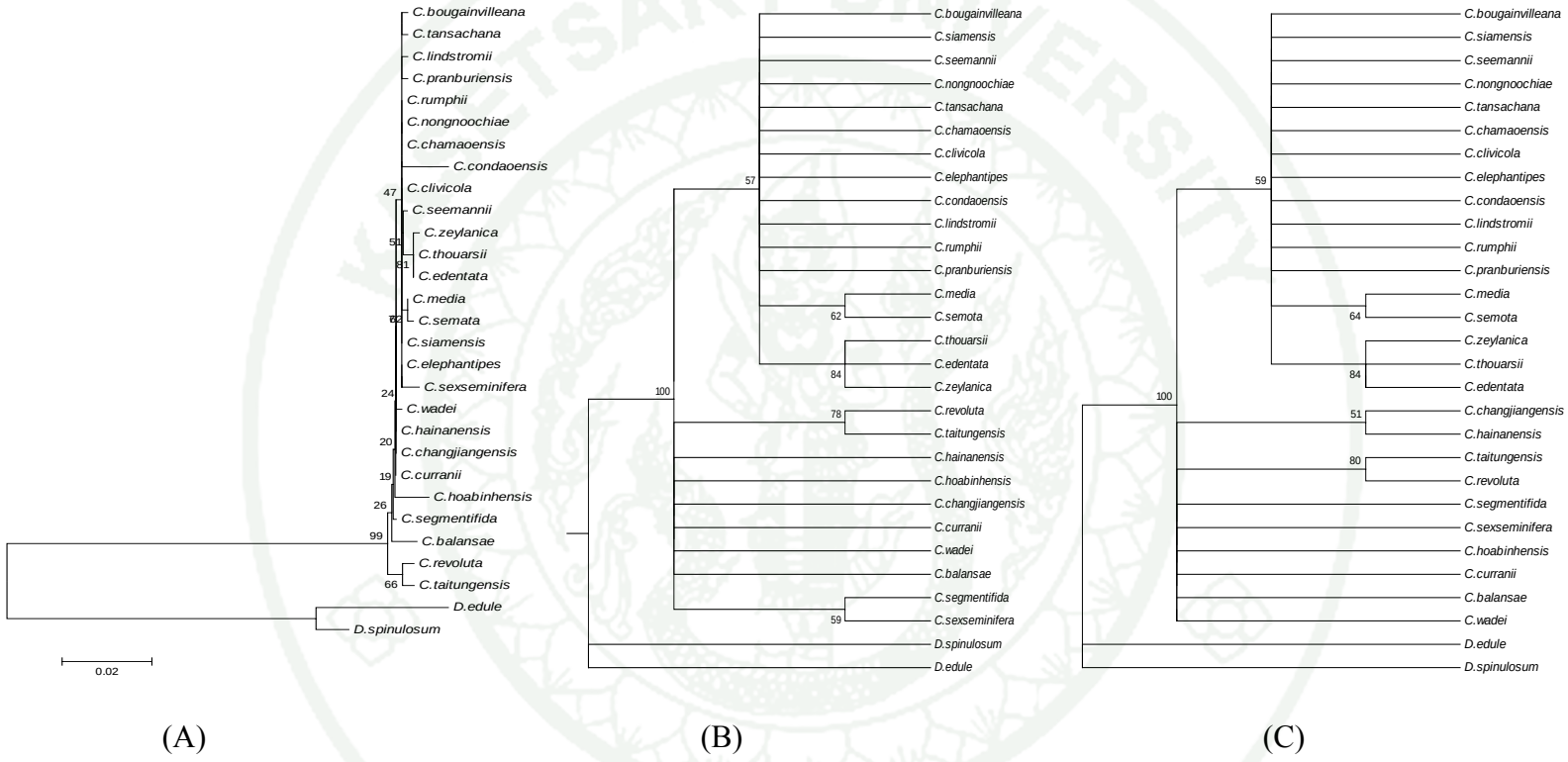


Figure 9 Phylogenetic trees construed by NJ (A), MP (B) and ML (C) from *psb*M-*trn*D spacer sequence of *Cycas* species.

Copyright by Kasetsart University All rights reserved



(A) (B) (C)

Figure 10 Phylogenetic trees construed by NJ (A), MP (B) and ML (C) from *trn*L-*trn*F spacer sequence of *Cycas* species.

Copyright by Kasetsart University All rights reserved

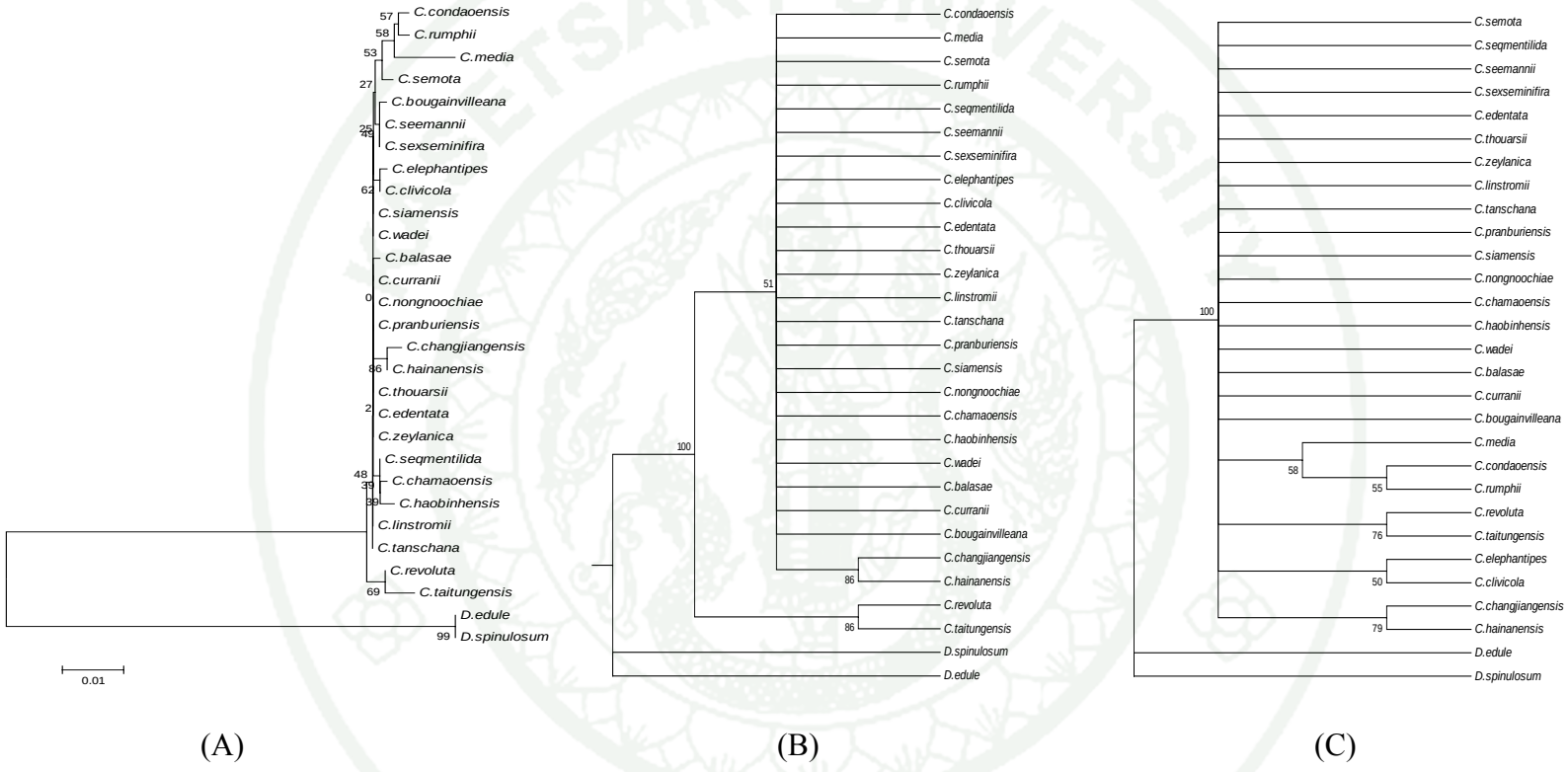


Figure 11 Phylogenetic trees construed by NJ (A), MP (B) and ML (C) from *trn*S-*trnf*M spacer sequence of *Cycas* species.

Copyright by Kasetsart University All rights reserved

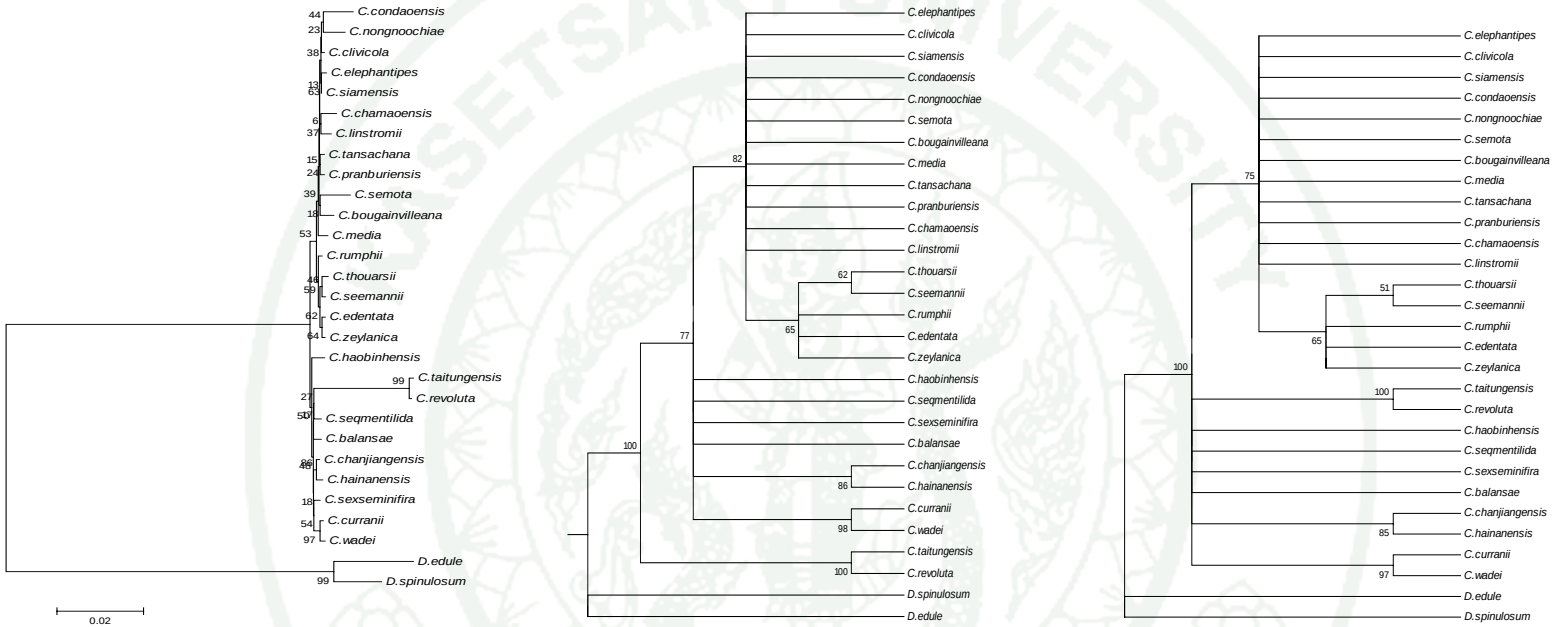


Figure 12 Phylogenetic trees construed by NJ (A), MP (B) and ML (C) from combined data set of *trn*S- *trn*G, *psb*M-*trn*D and *trn*L-*trn*F regions.

Copyright by Kasetsart University All rights reserved

3. Sex Identification in *Cycas*

3.1 Screening and isolation of male and female specific RAPD fragment

A total of 80 RAPD primers were used to screen for the marker associated with gender in *Cycas*. The only one primer OPB-18 was shown to produce male (M) and female (F) specific RAPD fragments in *C. clivicola* (F, M), *C. chamaoensis* (F, M), *C. elephantipes* (F, M), *C. siamensis* (M), *C. pectinata* (M) and C. *nongnoochiae* (F). The male specific RAPD fragment was approximately 500 bp, while the female specific RAPD fragment was approximately 400 bp (Figure 13). In addition, UBC 485 RAPD primer produced male specific DNA fragment in *C. clivicola*.

3.1.1 Converted of the RAPD OPB 18 marker into SCAR marker

The male and female sex specific RAPD fragments were cloned and sequenced. The sequences are shown in Figures 14 and 15. The sequence of male specific RAPD fragment was 487 bp with GC content of 40.6% and AT content of 59.4%, while the sequence of female specific RAPD fragment was 427 bp with GC content of 38.6% and AT content of 61.4%. These results indicated that the sequence of sex specific RAPD fragment contained AT rich region. The first ten nucleotide of the male and female specific DNA sequence completely matched with the RAPD primer OPB-18 (CCACAGCAGT). Interestingly, the female specific sequence contained the simple sequence repeats (SSR) motifs $(TA)_2$ 6 times, while $(TG)_2$ 6 times appeared in the male specific sequence. Blast search in EMBL plant library of female specific sequence did not show significant homology to any known sequences. On the other hand, the male specific sequence was found at 70% homology between 300 bp on 5′ end with Expressed Sequence Tag (ETS) of *C. rumphii* ovule. Further analysis was made by designing specific primer from male and female sequences to detect sex of *Cycas*. Two primer pairs (Table 11), M1 (forward) and M2 (reverse) were designed from male specific fragment (Figure 14A), while F1 (forward) and F2 (reverse) were designed from female specific fragment (Figure 14B).

Copyright by Kasetsart University All rights reserved

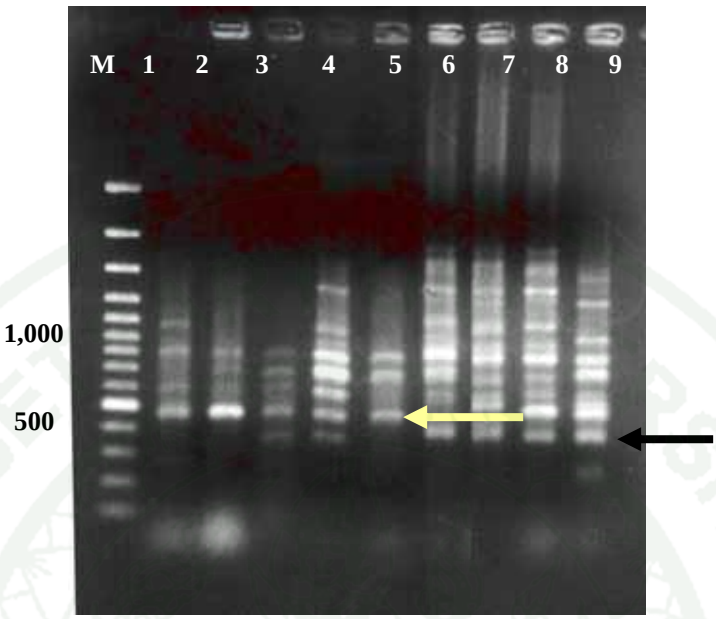
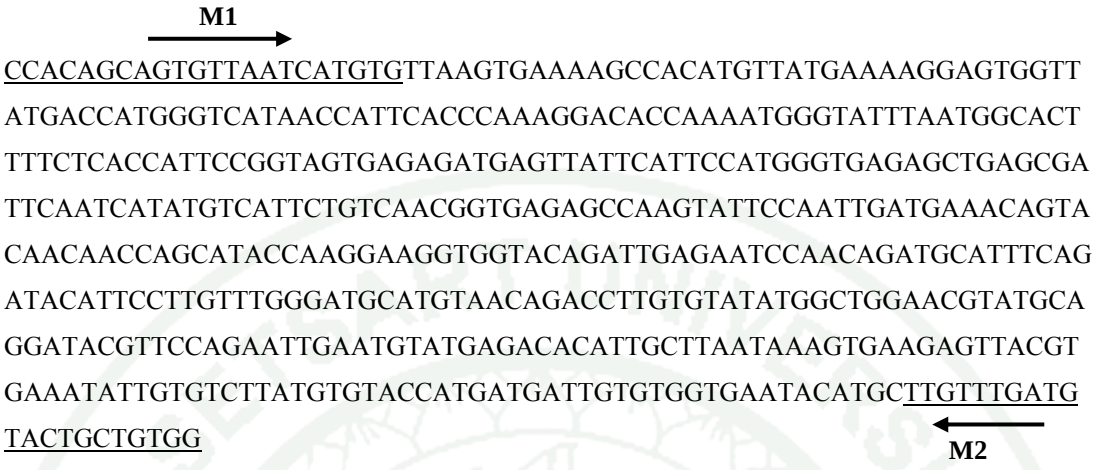


Figure 13 Amplification profile of RAPD marker OPB-18. Two arrows indicated male (white arrow) and female (black arrow) specific RAPD fragments: M = 100 bp molecular marker, lane 1 = *C. clivicola* (male), lane 2 = *C. chamaoensis* (male), lane 3 = *C. siamensis* (male), lane 4 = *C. pectinata* (male), lane 5 = *C. elephantipes* (male), lane 6 = *C. nongnoochiae* (female), lane 7 = *C. clivicola* (female), lane 8 = *C. elephantipes* (female) and lane 9 = *C. chamaoensis* (female).

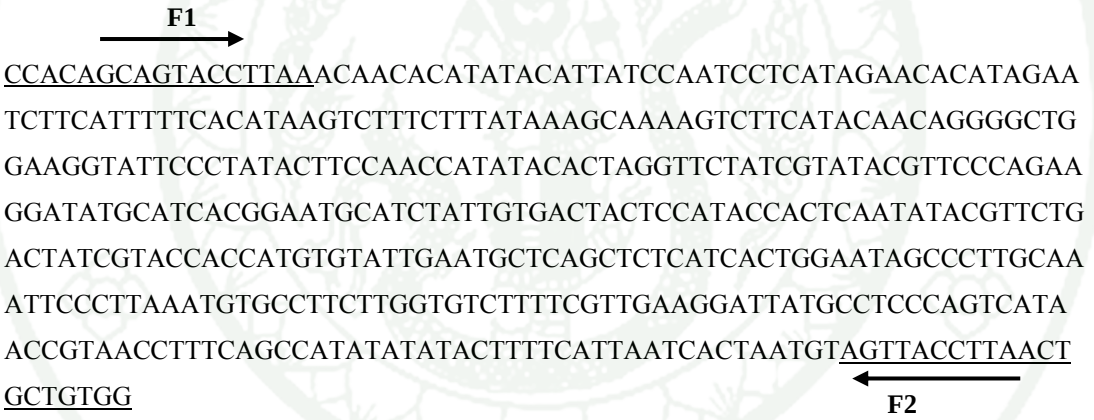
Table 11 Sequences of male and female sex specific primers.

Name	Sequence (5'-3')	Anealing Temperature (°C)
M 1 (Male)	CCACAGCAGTTTAATCATGTG	53
M 2 (Male)	CCACAGCAGTACATC AAACAA	
F 1 (Female)	CCACAGCAGTACCTTAAACAA	53
F 2 (Female)	CCACAGCAGTTAAGGTAACT	

Copyright by Kasetsart University All rights reserved



(A)



(B)

Figure 14 Male specific sequence of *Cycas* (A). Two designed primers, M1 and M2 (underlined) in accordance with male specific RAPD cloned fragment. Female specific sequence of *Cycas* (B). Two designed primers, F1 and F2 (underlined) in accordance with female specific RAPD cloned fragment.

Copyright by Kasetsart University All rights reserved

3.1.2 PCR amplification of male and female specific SCAR marker

Two sets of primer pair M1-M2 (male) and F1-F2 (female) were used to amplify genomic DNA of *C. rumphii* (male and female), *C. edentata* (male and female), *C. pectinata* (male) and *C. chamaoensis* (female). The results showed that two sets of primer could be used to amplify all male and female *Cycas* (Figure 15). Therefore, the six PCR fragments from M1-M2 primers and the six PCR fragments from F1-F2 primers were cloned and sequenced. The alignment of M1-M2 sequences is shown in Figure 16A, while that of F1-F2 sequences is shown in Figure 16B. The results of alignment obtained from the sequences amplification of M1-M2 primers showed no difference between male and female, but those from F1-F2 primers showed one insertion (ACA/GG/TAATGCATCT) in male. Subsequently, three clones of each species (male and female *C. edntata*, *C. rumphii* and *C. chamaoensis* and male *C. pectinata*) from F1-F2 PCR amplification fragments were sequenced to confirm this insertion site. However, the alignment of 21 clones indicated that this insertion site was not restricted presence only in female but shown to be absence or presence on both gender (Figure 17).

Copyright by Kasetsart University All rights reserved

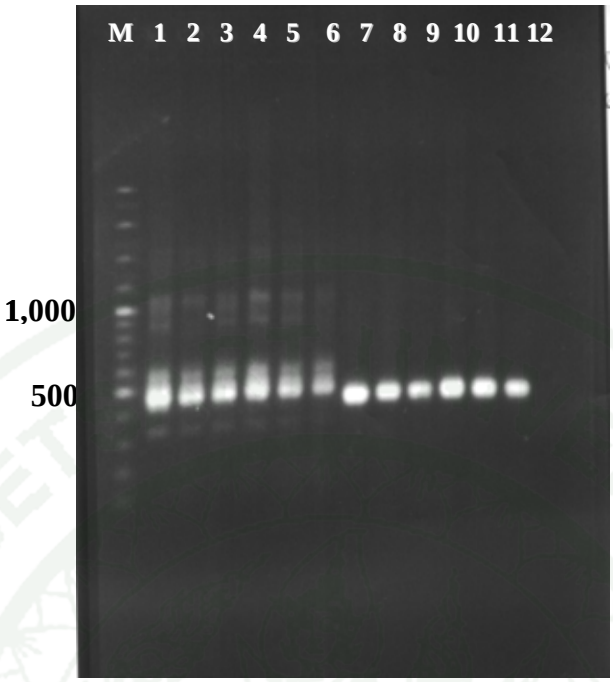


Figure 15 Gel electrophoresis of the amplification products obtained from sex specific primers; M1-M2 primers (lanes 1-6) and F1-F2 primers (lanes 7-12). M = 100 bp molecular marker, lanes 1, 7 = *C. rumphii* (male), lanes 2, 8 = *C. edentata* (male), lanes 3, 9 = *C. pectinata* (male), lanes 4, 10 = *C. rumphii* (female), lanes 5, 11 = *C. edentata* (female), lanes 6, 12 = *C. chamaoensis* (female).

Copyright by Kasetsart University All rights reserved

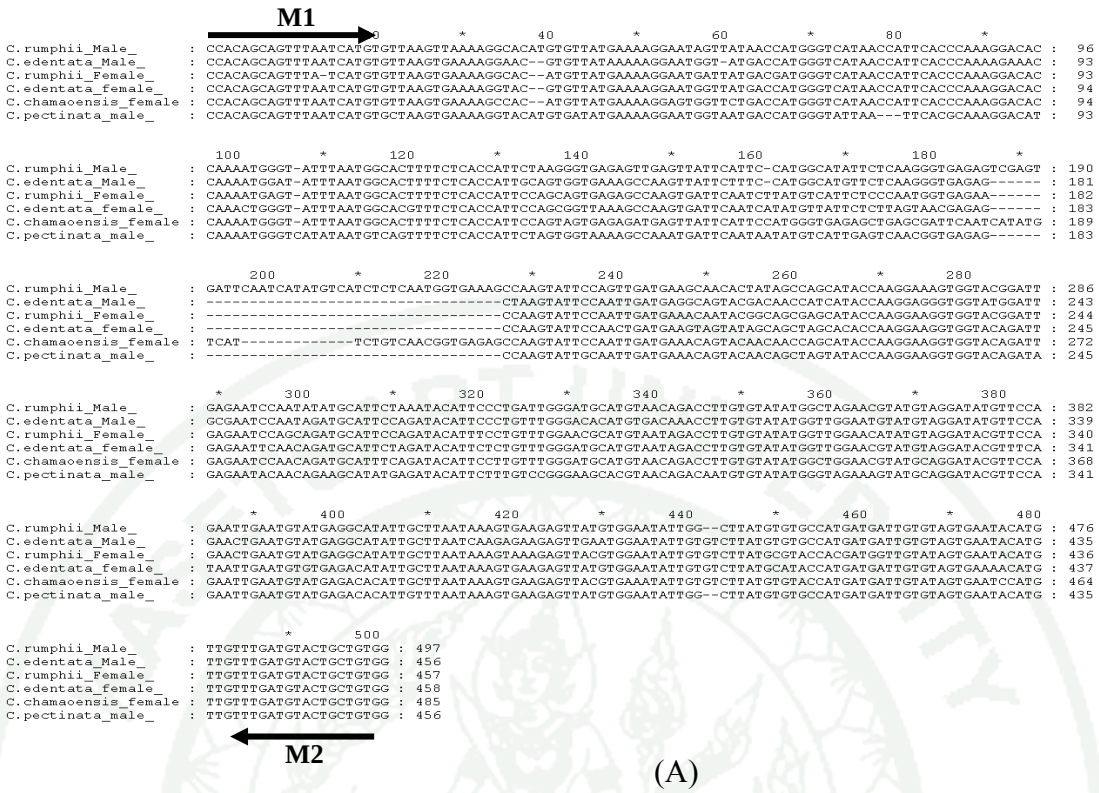


Figure 16 The alignment of six PCR sequences from M1 and M2 primers (A) and the alignment of six PCR sequences from F1 and F2 primers (B).

Copyright by Kasetsart University All rights reserved

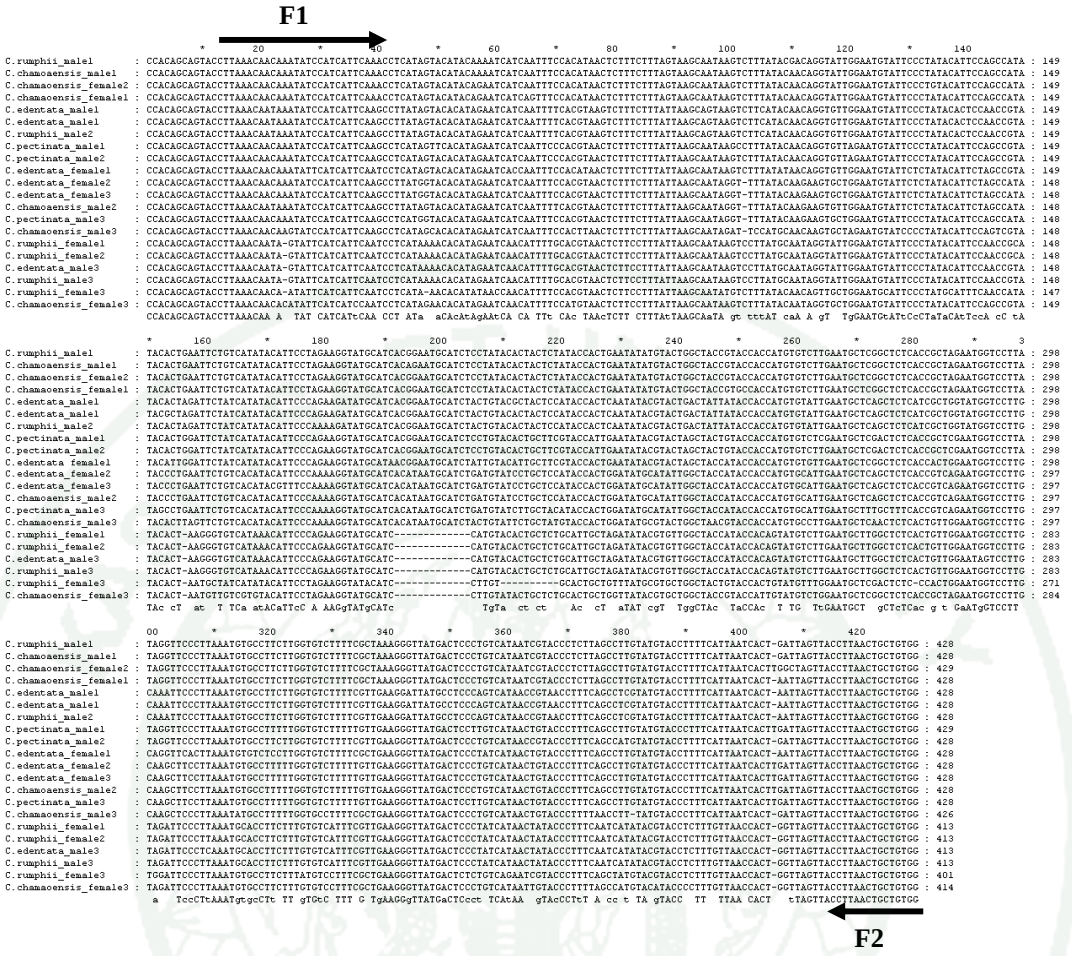


Figure 17 The alignment of 21 cloned sequences from F1 and F2 primers.

Copyright by Kasetsart University All rights reserved

3.1.3 RAPD UBC 485 sex specific marker of *C. clivicola*

Eighty random primers were used for RAPD fingerprinting to search for sex specific *C. clivicola* fragment. RAPD amplification with primer UBC 485 produced a fragment of about 450 bp, which was specific for male *C. clivicola* (Figure 18A). Male specific UBC 485 fragment was, thereafter, excised from agarose gel and cloned. Then, the male specific UBC 485 cloned was sequenced. The sequence length was 427 bp with high AT content (64.6%). Furthermore, five microsatellites, i.e., $(CATT)_2$, $(TCA)_2$, $(AAG)_2$, $(TAA)_2$ and $(TATC)_2$ were found in this sequence.

This sequence was aligned with the EMBL database by BLAST and the result showed that it produced 59 % homology between 300 nucleotides on 5′ end with *Nicotiana langsdorffii* x *Nicotiana sanderae* and 57 % homology between 300 nucleotides on 5′ end with EST of *Artemisia annua*.

Scar marker was further designed from RAPD UBC 485 of *C. clivicola* (Figure 19). Clivicola 485_F and Clivicola 485_R produced 427 bp in both female and male *C. clivicola* (Figure 20).

.

Copyright by Kasetsart University All rights reserved

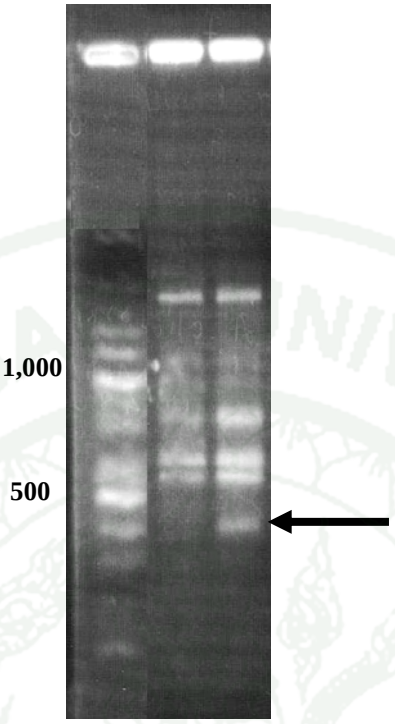


Figure 18 PCR amplification with UBC 485 primer in *C. clivicola*. Male specific band was marked with arrow: M =100 bp molecular marker, male = male *C. clivicola* and female = female *C. clivicola*.

Clivicola485_F

AGAATAGGGCTATATCCTTATGGGGGAATAGGTGCTTCTTATTAAATTTGTATAGGTTGGA
GTTTGACAACTTACTCAAATGAACAAATTATGGATGTAAATGAATCCATTGTCCTAATTAG
AGGCCTTTGAGATTATGGTTAGGGCCTTAGATCTCATAGTACCAGTTGTGCCTTATGTCGT
GGGCTGGGCTTCTTAATCTTCTTTAAGCCTGCTTAGATTTTGATGAATTCTAAGGTGAGGC
ATTTTATAATTATAGAGCGACATATACTTGCACTCATTTTGAGCACTTACACATAGACCTT
TAGATCAAGGACATTCATAAAGATATCCATCTTTTGATTTCTAAGTAACACATCCTTGAAT
GAATGTAATGGATCATGGAAATGATAGATAAATCTCTTATGAGCTTATCTTGGCCCTATTC
T

Clivicola485_R

Figure 19 The sequence of male specific RAPD UBC 485. Clivicola 485_F and Clivicola 485_R primers were underlined.

Copyright by Kasetsart University All rights reserved

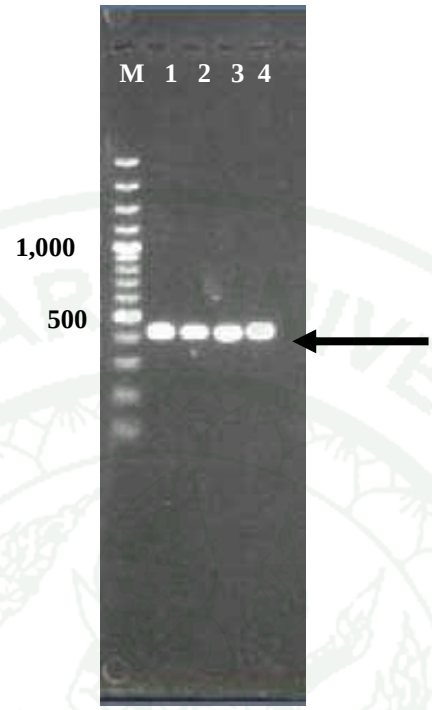


Figure 20 PCR amplification with Clivicola485_F and Clivicola485_R. Male specific bands were marked with arrow: M= 100 bp molecular marker, lanes 1, 2 = female *C. clivicola* and lanes 3, 4 = male *C. clivicola*.

Copyright by Kasetsart University All rights reserved

3.2 ISSR PCR produced sex specific DNA fragment

3.2.1 ISSR PCR produced sex specific DNA fragment of *C. clivicola*

Twenty-nine ISSR primers were screened on 2 bulks (male and female *C. clivicola*) and two ISSR primers, 001 and 864, producing 850 bp (ISSR 001) and 800 bp (ISSR 864) shown in male (Figures 21A and 21B). Male specific ISSR fragments of both ISSR primers were excised from agarose gel and cloned. Two cloned fragments were sequenced and the sequences are shown in Figures 22 and 23. The male specific ISSR 001 sequence was 876 bp and AT rich (62.5%). It contained microsatellites which were $(AGA)_2$, $(GATGA)_2$, $(GATAA)_2$, $(GAA)_2$, $(CAA)_2$, $(AAT)_2$, $(ATA)_2$, $(TCA)_2$, $(TTA)_2$, $(TGG)_2$, $(TA)_3$, $(TTG)_2$, $(TGGA)_2$, $(GAG)_2$, $(GTT)_2$. The male specific ISSR 864 sequence was 802 bp and contained 60.4% of AT. A total of 22 microsatellite sites were found in this region. These were $(TT)_3$, $(CA)_3$, $(GAA)_2$, $(ACA)_3$, $(CAA)_3$, $(AAC)_2$, $(ACA)_2$, $(ACC)_2$, $(ATA)_2$, $(TAA)_2$, $(CAT)_2$ 2 sites, $(TGA)_2$, $(TCA)_2$, $(AAG)_2$, $(TAC)_2$, $(GAT)_2$, $(TGC)_2$, $(GATT)_2$, $(GCAA)_2$, $(ATAG)_2$, $(ATTCA)_2$.

Blast search in EMBL plants library showed that between 300 bp on 3′ end of the male specific ISSR 001sequence showed 60% homology with ESTs from young male strobilus of *Zamia fischeri*. The male specific ISSR 864 sequence contained about 73% homology between160-450 nucleotide on 5′ end with ETSs from sporophyll of *C. rumphii*.

Based on two completed sequencing data, two pairs of SCAR primer were designed (Table 12) from male specific sequences ISSR 001 and 864. Clivicola001_F and Clivicola001_R were designed from male specific ISSR 001 sequence while Clivicola 864_F1, Clivicola 864_F2, Clivicola 864_R1 and Clivicola 864_R2 were designed from male specific ISSR 864 sequence. These primers were used to amplify the three male and three female *C. clivicola*. The SCAR primers Clivicola 001_F and Clivicola 001_R amplified a male specific band at an annealing temperature of 58°C in the male and female genomic DNA (Figure 24A). Clivicola864_F1 and Clivicola864_R1 amplified a band of 800 bp at high annealing

Copyright by Kasetsart University All rights reserved

temperature of 60°C in all male and one female genomic DNA (Figure 24B), while Clivicola864_F2 and Clivicola864_R2 amplified a band of 300 bp at annealing temperature of 58°C in all male and female genomic DNA (Figure 24C).

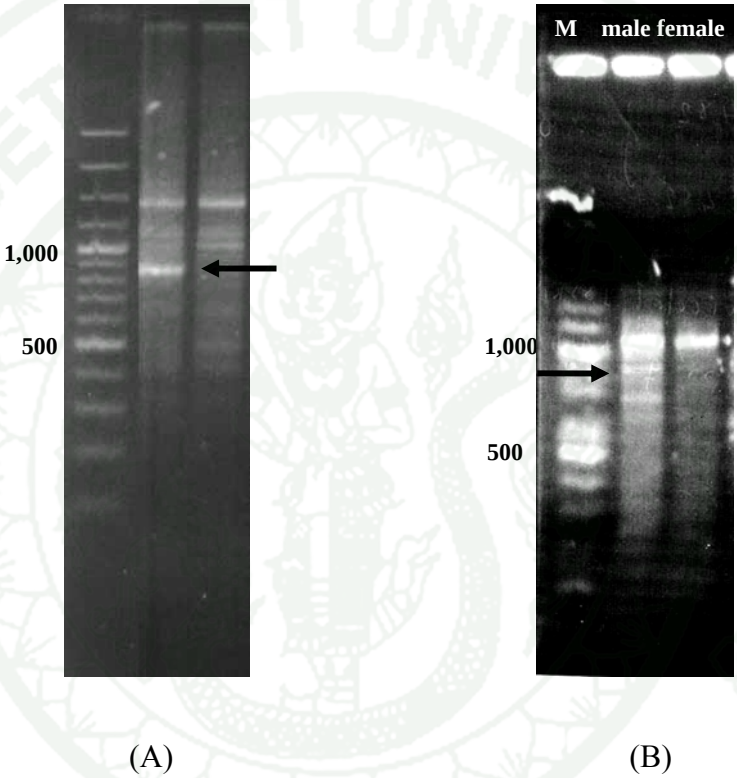


Figure 21 PCR amplification with ISSR 001 primer (A) and ISSR 864 primer (B) in *C. clivicola*. Male specific band was marked with arrow: M = 100 bp molecular marker, male = male *C. clivicola* and female = female *C. clivicola*.

Copyright by Kasetsart University All rights reserved

Clivicola 001-F →

CACACACACACACACAGGTTTGAGATGCGTACACGCCAGAAGAAGCCATGCGTTGGGATG
AGATGAGAGGAACACGTAACAAAACATGGCAAGTTCATAGAGTCGCCCAACGAAAGGAA
TAGAGTAGGTTTATGCTACAGTTCAGAGTCACCTAGTCAGTGTGATAAGATAACATTGCAC
ACTGAAGAAAATGATCTAATACTAAAAACCACAGTCCTTCATTGGACAGAAAATAGTCTA
ACATAGTAAGAGATGCCACAGTCCCATAGGGGAAAAGCACAATCACACACTTAATGTGTA
GTTTGTTAAGACATGTGATGTTAGCCATGTAAAATAATGTTTAATGTTCTCTTGCTCAAAC
TTTGATTAAGGTGGCATATATACACATGTTAATGTAAACTCCCATTCGGAGGTGAGAAGG
CTTTGAAATATGATTGGTTCATCATGGGGTGGATCCCACTTGAGAAACTTGAGAGTGGAA
GGAGTAACTTTATTATGCGATTAACAGAGTAAAGTTAGACTCCAACCTTCTCTTGTGGTTT
TTCCCTAGTTTGGGTTTTCCACATTTATGTCTAGTGTCTCATGGTGGTTATGCATGGTCTTA
TATTTATCTTGATTTGCATAATGTTTGTGTATATACCTTATGATTTGTTTTTATGCTTCAACT
ATTATGTTATTCTTTGACATGGTATATTAATCTATTGTTGCATGCTAGACTTGTTGTGGATT

← Clivicola001_R

GGATTGTTTCTAACATTTGGTATCAAAGGTTTCTTACATGTGGTATCAGAGTCAATGGTTTT
GAAGGTTTAGATGAGGAGAACCTATTTTGATAGGTTGTTAGTCATGTTGATTTAGAAGCTT
GTGTGTGTGTGTGTGTGTGCCTGT

Figure 22 The sequence of male specific ISSR 001. Clivicola 001_F and Clivicola 001_R primers were underlined.

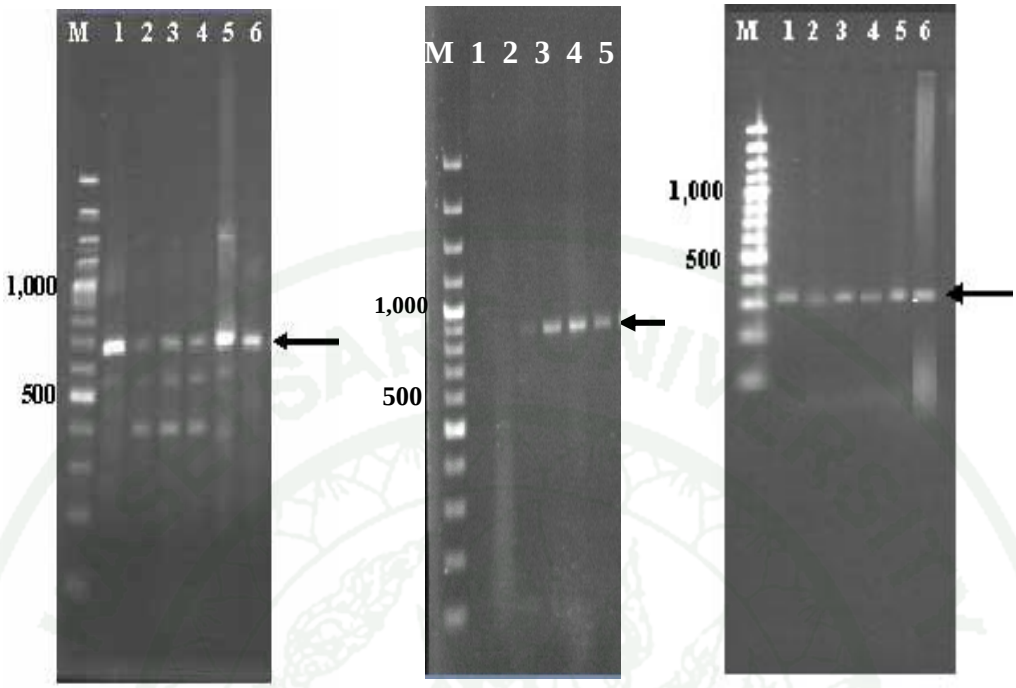
Copyright by Kasetsart University All rights reserved

Clivicola 864_F1

ATGATGATGATGATGCAAAAAGGATGCACTCACGAGATGATTAATGTTCTCCACCCCA
TCAAACATGCCATTGATGGGGACTTTAAGACCTAGTGTATCACTTGGGCTTTTCTGATT
Clivicola 864_F2
CAATTCAGTCAGCATGTCCCAAATCACAGAAGCTTTCACTAAAGACCCTTGACTTTAA
AGCTTTAGCTGAGAAGTTTATTGAACATGATAAGACCTTTGTGACCAACTTCCTAATAT
TAGATTCACTTCCAAAATGTAGCTTTGAAGTAGAAGTGGACGAGGAAGTTTTTTTCACT
AAACAACAAAGTTACCCCTCACACAGAGGAAGAAAAGCTTGCAAAAGAGGCAGGGGC
AAGCAATAGGGACAACAACAAAGGCAACAACAAGAGTAGCAATAACAAAACACCTA
ATATAGATAGTCTCCCCAATAAGGACAACATCAGAATAACAATTAAAGAACCACCAAT
Clivicola 864_R2
CAGAGATAATATGGGCAAATTAATAACAACACTCCCATGTTCCAAAGGTGTAATAGGA
GAGGACACTTAGCCAAAGATTGCTGGATTGATTGCAATAGATTGCCACATCATTAGTC
ACAGCATTAGGCTCATAATGTTGATGAGGGAAATAACGATATGTTTGCAGGAGCGAAT
TTCATTGAGGCACAAGTTCATCAAAGAGGGGATTCATCCACATCATCAACTGCTATAT
GACCAAAGAAGTATGTTGAGGCTCATTGTGGTACTACCATTTGTCAGCTCGATTTGCCA
CTACCTTAATCACTTGGTTAGGCCAATGCTGCCATCATCATCATCAT
Clivicola 864_R1

Figure 23 The sequence of male specific ISSR 864. Clivicola 864_F1, Clivicola 864_R1, Clivicola 864_F2 and Clivicola 864_R2 primers were underlined.

Copyright by Kasetsart University All rights reserved



(A) (B) (C)

Figure 24 PCR amplification with Clivicola001_F and Clivicola001_R primers (A), Clivicola864_F1 and Clivicola864_R1 primers (B) and Clivicola864_F2 and Clivicola864_R2 primers (C). Male specific band were marked with arrow: M= 100 bp molecular marker, lanes 1, 2, 3 = female *C. clivicola*, lanes 4, 5, 6 = male *C. clivicola*.

Copyright by Kasetsart University All rights reserved

3.2.2 ISSR PCR produced sex specific DNA fragment of *C. edentata*

Twenty-nine ISSR primers were screened on 2 bulks (male and female *C. edentata*). One ISSR primer (ISSR 836) produced a male specific fragment of 800 bp (Figure 25). This fragment was excised from agarose gel, cloned and sequenced. The male specific ISSR 836 sequence was 747 bp and contained 60.6% of AT. Microsatellites were detected in this sequence. There were 11 microsatellite sites, $(AA)_3$ 2 sites, $(GAA)_2$, $(AGG)_2$ 2 sites, $(AGA)_2$, $(AAT)_2$, $(ACC)_2$, $(GAA)_2$ 2 sites, $(ACCTAA)_2$ (Figure 26).

Blast search in EMBL plants library showed that the male specific ISSR 836 sequence produced 79% homology between 200 nucleotides on 3′ end with ESTs from *C. rumphii* ovules, 62% homology between 424-730 nucleotides on 5′ end with ESTs from *Zamia fischeri* megasporophylls and ovules, 60% homology between 417-730 nucleotides on 5′ end with female gametophytes of *Zamia vazquezii* and 58% homology between 461-730 nucleotides on 5′ end with ESTs from young male strobilus of *Zamia fischeri*.

SCAR primers were designed from the male specific ISSR 836 sequence. Edentata836_F1 and Edentata836_R1 were designed to amplify a 747 bp fragment, while Edentata836_F2 and Edentata836_R2 were designed to amplify a 633 bp fragment. The results showed that Edentata836_F1 and Edentata836_R1 SCAR primers could amplify the expected band in male and female *C. edentata* at an annealing temperature of 58°C as well as Edentata836_F2 and Edentata836_R2 SCAR primers (Figures 27A and 27B).

Copyright by Kasetsart University. All rights reserved

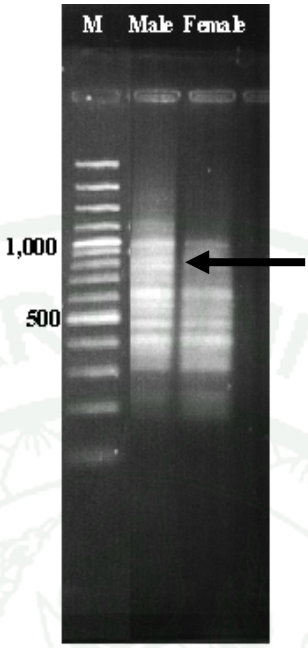


Figure 25 PCR amplification using ISSR 836 primer in *C. edentate*. Male specific band was marked with arrow: M = 100 bp molecular marker, male = male *C. edentata* and female = female *C. edentata*.

Copyright by Kasetsart University All rights reserved

Edentata836_F1

AGAGAGAGAGAGAGAGTAGGTGGAAGAAGAGGAGGAAAATGACGTTCCTTTGGGA
TTTATACCTAAACCTAACTTGAATGGATGGTTAGGGTTAGGCTAGGATTCAAGGGGTGC
TTAAGCCACATGGCAATTTCGTGAAAAAAATGAAAATTTCCACAGTAGATTAAAAACA

Edentata836_F2

ATATTTTATGACAATCGCCCCATGACAAACTTTGCGATACTGTGAAGTTTACTGCAGTA
CTGCAACATTTGCTATGGTGTAGTGACAAAGGCAAAATCAACATAGAAATAGGTGCCT
GGGAGCCTTGAGGGAGCGTACCCCAAGCCTAGGAGGGGCTAGCCCTCTATCCAAGTGG
CTACGGATAATCATATAGAGAGGATGACCAAAACCGGATGGGGTACCGAAGTGGTAAG
GTTCCCAAGAAATTGATAACTGGAATAGGGAGCTATCCTCATGTAGAATTTTCATCCAA
TACCTTTGAAAGTCCAAAATTGAAAGACCATGAATGTTGATTTTTAACCTATAAAAAAG
AAATGAATGTTATTCTCCATGTAGTTGATAAGTGATATCAATATTTACTCAGATCCAAA
TTCACTATTTTAACTTACCACCATGGTCTCAATCACCTCCATGATCAACCTATCCTTTAT

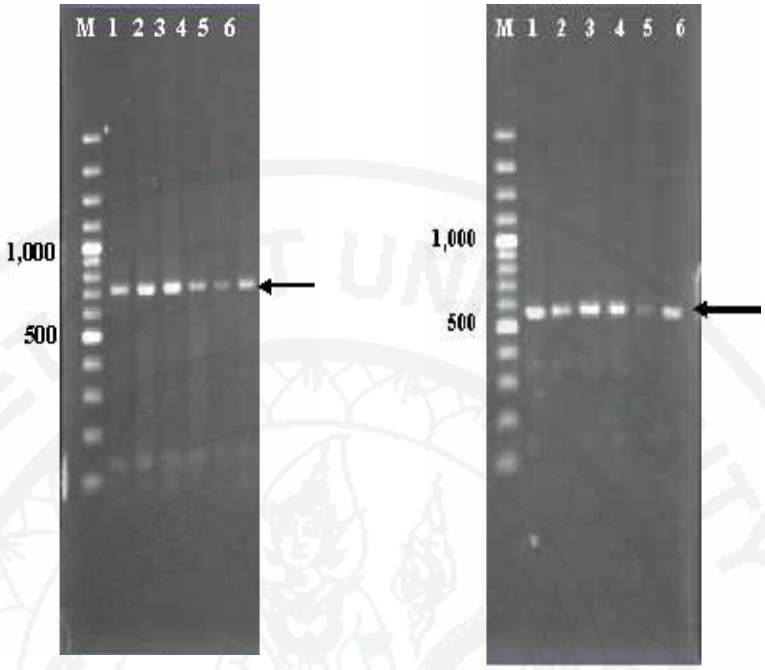
Edentata836_R2

GAAGAACAAGCAAAGTGGTTCAATCACCTCTAAGTGTTCAACTTCACCATATGTTACAA
AAAGGGCAAGGACAATGTTATAGTTGATGCTCTCTCTCTCTCTCTT

Edentata836_R1

Figure 26 The sequence of male specific ISSR 836 from *C. edentata*. Edentata836_F1, Edentata 836_R1, Edentata836_F2 and Edentata 836_R2 primers were underlined.

Copyright by Kasetsart University All rights reserved



(A) (B)

Figure 27 PCR amplification with Edentata836_F1 and Edentata836_R1 primers (A) and Edentata836_F2 and Edentata836_R2 primers (B). Male specific band were marked with arrow. M= 100 bp molecular marker, lanes 1, 2, 3 = female *C. edentata*, lanes 4, 5, 6 = male *C. edentata.*

Copyright by Kasetsart University All rights reserved

3.2.3 ISSR sex specific marker of *C. chamaoensis*

Twenty-nine ISSR primers were screened on 2 bulks (male and female *C.chamaoensis*). One ISSR primer (ISSR 001) produced a male specific fragment of 900 bp (Figure 28). This fragment was again excised from agarose gel and cloned. The clone was sequenced (Figure 29). The length of the male specific ISSR 001 sequence was 871 bp with AT content of 68.4%. Thirteen microsatellite sites were found in this sequence; $(AA)_2$, $(TT)_2$, $(TAA)_2$, $(GTT)_2$, $(GAA)_2$, $(AGA)_2$, $(TTG)_2$, $(ATT)_2$, $(TA)_2$, $(ATAG)_2$, $(GAGT)_2$, $(ATAA)_2$ and $(TGAT)_2$.

Blast search in EMBL plants library showed that the male specific ISSR 001 sequence produced 55% homology between 750 nucleotides on 5′ end with ESTs from *Beta vulgaris* and *Codonopsis lanceolata*, while it produced 56% homology between 550 nucleotides on 5′ end with ESTs from *Arabidopsis thaliana*.

SCAR markers were designed based on the male specific ISSR 001 sequence. Chamaoensis001_F (forward) and Chamaoensis001_R (reverse) were designed to amplify 820 bp sex specific fragment. The result showed that the sex specific fragments were obtained in male and female (Figure 30).

Copyright by Kasetsart University All rights reserved

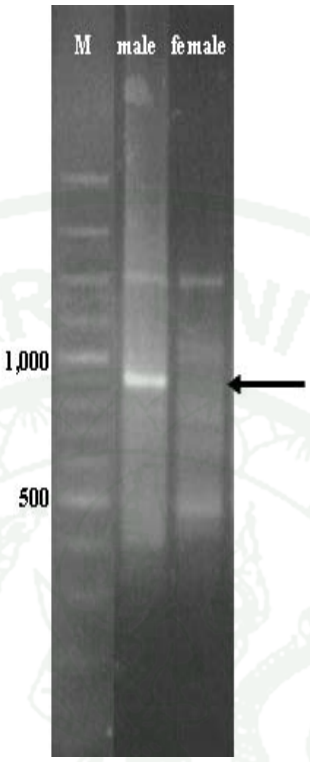


Figure 28 PCR amplification using ISSR 001 primer in *C. chamaoensis* (A)
. Male specific band was marked with arrow: M = 100 bp molecular marker, male = male *C. chamaoensis* and female = female *C. chamaoensis*.

Copyright by Kasetsart University All rights reserved

Chamaoensis 001_F →

CACACACACACACACAGGTATATATATATATATTGGATTTCATTACTTCTCAGCTTTGT
ATATCCAAACATAGAGCAAAATATCTATGCAAAAGCTATATGGTTGTGACAAAAGTTG
TATATTGGTAACCTCAAGGGTTGTATAACTACATATACAAAACCAAATTTAGAATGAA
AGAGCAAGTCATACAACTGCATGTTGGAGTGAGTATATAGATCATACGGCTAGGCCTA
ATAAGGGTATGTGGGCTATAGCATTCTTTTGGTTTTGATGCCAAATAAGAGTGCATAAA
TAACCTAATTTTACTTTCCAATATAAACCCTACCTTTAGTACCAATTCAAAATGAAACT
AATGGTTGTTTTTAGTTTAATTTCTACATTTTATTCAATGTGAATTTCAAAGAACATACA
AACATAATGAAGGTTGCTTTTCAATTTAATCCTACGCATGTTTTTGAAGACTATTTATCT
TATGTTTACTGTAGAACATTGGACATTTATATGAAGAAATAATTTTTTGTTTATGTACAT
GAAATATTGTTTGGTATCACGTTTACATTATGAAATGTAAAATGATTGTGTTAGAAGAG
ATTGTTGTCCACTTTATCAAGCAGTTGGAAGACTTGTACAAACCATGTTAGGCTTGCCA
TAGGCATCCTCAAAATCCCTTGAACATGTAGGTTCATAACTGGTAAGGAAAGGCAGAT
TATTATGGAAAAATAAAGGAAACAAAATCCATGATTGATTACCTATATATAATTGTGC
ATGTATGACTAAGTCGTTGCCCTTAATAGATAGTATGAAAATGATCGTGCTCTTGTGGA
TAAGACCTAATGTAAAAAAGTACCAAGCTCTTGTGTGTGTGTGTGTGTG

← Chamaoensis 001_R

Figure 29 The sequence of male specific ISSR 001 from *C. chamaoensis*. Chamaoensis 001_F and Chamaoensis 001_R were underlined.

Copyright by Kasetsart University All rights reserved

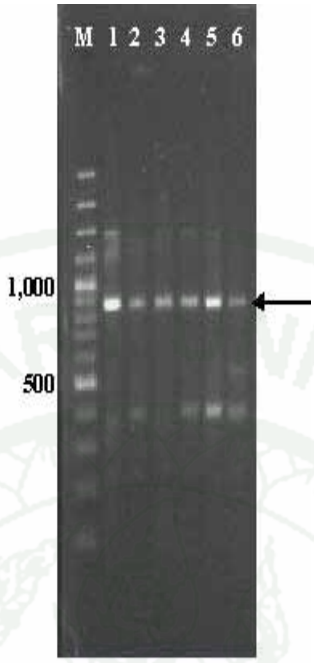


Figure 30 PCR amplification with Chamaoensis001_F and Chamaoensis001 _R primers. Male specific band were marked with arrow: M= 100 bp molecular marker, lanes 1, 2, 3 = female *C. chamaoensis*, lanes 4, 5, 6 = male *C. chamaoensis.*

Copyright by Kasetsart University All rights reserved

Table 12 Sequence of oligonucleotide primers for each SCAR locus derived from RAPD and ISSR primers.

Name	Sequence	Annealing temperature (°C)
Clivicola485_F	AGAATAGGGCTATAT CCTTATGG	60
Clivicola485_R	AGAATAGGGCCAAGATAAGCTC	
Clivicola001_F	CACAGGTTTGAGATGCGTACA	60
Clivicola001_R	CCAATCCACAACAAGTCTAGCA	
Clivicola864_F1	ATGATGATGATGATGCAAAAAGGATG	60
Clivicola864_R1	ATGATGATGATGATGTGGCAGCAT	
Clivicola864_F2	CTTTCACTAAAGACCCTTGACT	60
Clivicola864_R2	TCTGATTGGTGGTTCTTTAATTG	
Edentata836_F1	AGAGAGAGAGTAGGTGGAAGAAGA	58
Edentata836_R1	AAGAGAGAGAGAGAGCATCAACT	
Edentata836_F2	AATCGCCCCATGACAAACT	58
Edentata836_R2	CATGGAGGTGATTGAGACCAT	
Chamaoensis001_F	CACACACACACACACACAGGT	58
Chamaoensis001_R	CACACACAAGAGCTTGGTACTTTT	

Copyright by Kasetsart University All rights reserved

DISCUSSION

1. Phylogenetic relationships within the family Zamiaceae and Stangeriaceae

1.1 Sequence variability and tandem repeats among genera

All non-coding sequences in this study were AT-rich. According to Li (1997) most non-coding spacer and pseudogenes are highly A+T because of low functional constraints. Although *atp*B-*rbc*L spacer sequence was highly conserved region, the rate of nucleotide substitution of *atp*B-*rbc*L spacer of Cycad was higher than Rubiaceae (Manen and Natali, 1995), *Hylocomiaceae* and *Pleurozium* (Chiang, 2000) and mosses (Chiang and Schaal, 2000). It was noteworthy that the evolutionary rate of *atp*B-*rbc*L spacer in Cycad was higher than the other plants. The chloroplast genome contains simple sequence repeats similar to those found in nuclear genome and can provide polymorphism as a DNA marker (Powell *et al*., 1995). The *atp*B-*rbc*L spacer sequence of *Dioon* showed tetranucleotide 2 repeat units that one unit was insertion (indel) position. Indel or inversion might be a major driving force in sequence evolution (Britten *et al*., 2003). In contrast to previous studies, microsatellites in *atp*B-*rbc*L spacer of Coffea and *Fraxinus* were mononucleotide repeats (Harbourne *et al*., 2005; Tesfaye *et al*., 2007). The long length of microsatellite indicated that *Dioon* had higher mutation rate than other genera.

Encephalartos and *Macrozamia* revealed variable patterns of tandem repeats of *trn*L-*trn*F non-coding region supporting the highly genetic diversity in these genera (Byrne and James, 1991; Prakash and Staden, 2008; Prakash *et al*., 2008). Furthermore, *Encephalartos* and *Lepidozamia* showed a similar pattern of tandem repeat which differed only in the repeat number. This result indicated that the two genera were closely related.

The neotropical genus, *Dioon* is composed of 13 species (Hill, 2004). *Dioon* occurs in restricted to one Mexican species (*D. mejiae*) from Honduras. Previous molecular investigated of *Dioon* species, located along the Pacific seaboard

Copyright by Kasetsart University All rights reserved

of Mexico, based on ITS region of the ribosomal DNA and the *trnL–F* region of the chloroplast DNA. Wagner parsimony analysis on a 187 character matrix yielded two equally parsimonious trees and the consensus tree had two well-defined major clades. The first was composed of *D. mejae, D. rzedowskii*, and *D. spinulosum*; the second was composed of *D. califanoi, D. caputoi, D. edule* var. *angustifolium, D. edule* var. *edule, D. holmgrenii, D. merolae, D. purpusii, D. tomaselli* var. *sonorense*, and *D. tomaselli* var. *tomaselli* (González *et al*., 2008) and chloroplast DNA restriction fragment length polymorphism analysis (Moretti *et al*., 1993) indicated that this genus comprises two major clades (one including *D. spinulosum*, the other *D. edule*). These results showed that most clades obtained during analyses corresponded with previously recognized species within *Dioon*. In this study, tandem repeat data also indicated that *D. spinulosum* and *D. edule* could be separated into two groups. Repeat structures in chloroplast DNA are mostly found in AT-rich and often show long length of mononucleotide repeats (Kelchner, 2000). The tandem repeats in this study were AT-rich and similar to the repeat units found in the chloroplast $tRNA^{LEU}$intron of some plants (e.g. alfalfa (*Medicago sativa* L) (Skinner 2000), *Orchis palustris* (Cafasso *et al*., 2001), *Anacamptis palustris* (Cozzolino *et al*., 2003; Cozzolino *et al*., 2004) and Taxaceae and Cephalotaxaceae (Hao *et al*., 2009)). Although the $(A/T)_n$ mononucleotide repeats are rare in the plastid genome, they can be useful for identification and evolution studies at low taxonomic levels (Deguilloux *et al*., 2004; Jakobsson *et al*., 2007; Guicking *et al*., 2008).

The non-coding sequences of *Chigua* and *Zamia* shared some insertions, deletions and base substitutions. There were large indel in *psb*M-*trn*D and *trn*L-*trn*F non-coding sequences. The indels in cpDNA are small ranging from 1 bp to 10 bp and in large indels up to 1,200 bp (Palmer, 1991). These indels represent sequence similarity with the flanking region of the indel site and could provide useful information to resolve intraspecific level (Ingvarsson *et al*., 2003). Furthermore, the large deletions in the *trn*L-*trn*F non-coding sequence of *Chigua*, *Microcycas* and *Zamia* showed nucleotide direct repeat sequences within the flanking region and may be related with the formation of hairpins or stem-loop structure (Mes *et al*., 2000). The *trnL-trnF* region is located in the large single-copy region of the chloroplast

Copyright by Kasetsart University All rights reserved

genome. It consists of the *trnL* gene, a group I intron, and the *trnL-F* intergenic spacer. The polymorphic pattern in the *trnL-trnF* region in Cycad revealed a higher level of mutation points. Similarity in the presence of flanking repeated sequences of large deletions in *Pedicularis* were considered useful for phylogenetic reconstruction (Yang and Wang, 2007). McDade and Moody (1999) used sequence data from *trn*L-*trn*F non-coding sequence to study phylogenetic relationships among Acanthaceae. Their results indicated that this region is useful in addressing phylogenetic among genera.

The sequences of *trn*S-*trn*fM region showed a high variability of AT repeated patterns in this study. In addition, the number of AT repeats in this region was used to predict of the curcumin content in the rhizome of *Curcuma longa* and four *Curcuma* plant species, *C. aromatica*, *C. longa*, *C. xanthorrhiza* and *C. zedoaria*, were identified using this region (Minami *et al*., 2009). Genetic variation within and among host-specific subspecies of *Viscumalbum were* studied by ITS, *trn*H-*trn*K, *trn*S-*trnf*M and *trn*L-*trn*F sequences. The results showed that *trn*S-*trn*fM region provide more nucleotide substitution than other regions (Zuber and Widmer, 2000).

1.2 Phylogenetic relationships within the family Zamiaceae and Stangeriaceae

ML, MP and NJ analyses of the *atp*B-*rbc*L, *psb*M-*trn*D, *trn*L-*trn*F and *trnS-trnf*M sequence data sets and the combined data set indicated that the families Zamiaceae and Stangeriaceae were not monophyletic. These results strongly supported *Dioon* as the basal group to the remaining taxa and were consistent with previous phylogenetic studies using molecular data (Rai *et al*. 2003; Bogler and Francisco-Ortega, 2004; Chaw *et al*., 2005; Zgurski *et al*., 2008). Based on the morphological classification of Stevenson (1992), *Dioon* was placed in subfamily Encephalartoideae which also included *Encephalartos*, *Macrozamia* and *Lepidozamia* (Stevenson, 1992). It is the only genus in the Encephalartoideae with an apical cone dome like the genera in the Stangerioideae, Bowenioideae, Zamioideae and Cycadaceae. The sporophylls are loosely imbricate and end in an upturned shield which clearly separates this genus from other genera in the subfamily

Copyright by Kasetsart University All rights reserved

Encephalartoideae. *Dioon* could be perhaps interpretated as somewhat intermediate between *Cycas* and the other genera of cycads (Bogler and Francisco-Ortega, 2004).

All of our analyses indicated that the tribe Encephalarteae of Stevenson (1992), comprising *Encephalartos*, *Macrozamia* and *Lepidozamia*, was monophyletic and that *Encephalartos* was closely related to *Lepidozamia*. Three genera have a number of shared features, large cones, sporophylls with diamond-shaped tips (shields) and ovules that fully developed even in the absence of pollination (Stevenson, 1992). *Macrozamia* and *Lepidozamia* are restricted to Australia, while *Encephalartos* occurs in Africa, hence it is possible that *Lepidozamia* was separated from a common ancestor with *Encephalartos* before Africa and Australia were separated in the disintegration of Gondwanaland (Bogler and Francisco-Ortega, 2004). Furthermore, the monosaccharide distribution patterns of the mucilages were characteristic at the genetic level as *Lepidozamia* and *Encephalartos* showing similar pattern (De Luca *et al*., 1982). Our results revealed that *Encephalartos* and *Lepidozamia* shared nucleotides insertions in the *psb*M-*trn*D and *trn*L-*trn*F non-coding regions. In addition, the average of base substitution rate per site indicated that these two genera were closely related. While molecular dating of these lineage divergences remained controversial (Hermsen *et al*., 2006), It is unlikely that the current distribution of *Encephalartos* and *Macrozamia* was due to the long distance dispersal rather than continental drift (Treutlein and Wink, 2002).

The African *Stangeria* and Australian *Bowenia* were previously classified in the family Stangeriaceae (Stevenson, 1992). *Bowenia* is unique among cycad in having bipinnately compound leave. *Stangeria* has an underground stem, fernlike leaves with permanent midribs on the leafets and medium-size cones with tightly overlapping sporophyll tip. *Bowenia* and *Stangeria* share two morphology characters, they absence of cataphylls on vegetative shoots and the circular pattern of vascular bundles in the petioles. In the current study, *Stangeria* was not found to be closely related to *Bowenia*, similar to the results reported by other groups (Treutlein and Wink, 2002; Hill *et al*., 2003; Rai *et al*., 2003; Bogler and Francisco-Ortega, 2004; Chaw *et al*., 2005; Zgurski *et al*., 2008). These previous studies were, however,

Copyright by Kasetsart University All rights reserved

inconsistent in the position of *Stangeria* and *Bowenia* to the other genera. Our data indicated that *Stangeria* should be placed within or adjacent to the subfamily Zamioideae, while *Bowenia* was sister to the tribe Encephalarteae or to the remaining taxa with the exception of *Dioon*. *Stangeria* was sister to *Ceratozamia* which shared nucleotide insertion on the *atp*B-*rbc*L spacer and nucleotide insertion and deletion on the *trn*S-*trn*fM non-coding region. The sequence data indicated that *Bowenia* was isolated and not closely related to *Stangeria*.

Further evidence for these proposed relationships could be found in chromosome morphology. The chromosome number of *Stangeria* and *Ceratozamia* is 2n=16, whereas that of *Bowenia* is 2n=18 (Kokubugata *et al*., 2000, 2004). *Stangeria* and *Ceratozamia* have similar karyotypes composed of 12 metacentric (m), two submetacentric (sm) and two telocentric (t) chromosomes based on 5S ribosomal probes (Kokubugata *et al*., 2004). Fluorescent *in situ* hybridization (FISH) study of rRNA sites on somatic chromosomes of *Stangeria* revealed that seven of the sixteen rDNA sites on *Stangeria* corresponded with the seven rDNA sites on *Ceratozamia*. This result implied that not only was *Stangeria* karyomorphologically but also molecular-cytologically closely related to the genus *Ceratozamia* (Kokubugata *et al*., 2002). Fossils attributed to *Stangeria* have been found in Argentina (Artabe and Stevenson, 1999) and Europe (Kvaćek and Manchester, 1999). *Ceratozamia* (now restricted to the New World) has had fossils from Europe, Mexico and Central America attributed to it (Kvaćek, 2002).

The ML, MP and NJ of combined data analyses suggested that *Zamia*, *Chigua* and *Microcycas* formed a strongly supported monophyletic group. The species of *Chigua* (from Columbia) was embedded within the genus *Zamia* and appeared closely related to the Columbian *Zamia, Z. manicata*. This result corresponded with the results of Caputo *et al*. (2004) who investigated the molecular phylogeny of *Zamia* based on the sequences of the internal transcribed spacer 2 of the nuclear ribosomal DNA combined with a morphological data set and found that *Chigua* formed a clade with *Z. manicata*. *Chigua* is a genus of two species, *C. restrepoi* and *C. bernalii*, both from northwestern Columbia. *Zamia* has

Copyright by Kasetsart University All rights reserved

approximately 60 known species (Hill *et al*., 2004) widely distributed across the central Americas. *Chigua* and *Zamia* both have swollen petiole bases and articulate leaflets. In addition, they showed the same composition and distribution of monosaccharide in their mucilage (Stevenson and Siniscalco Giliano, 1989). The chromosome number of *Chigua* is 2n=18 (Caputo *et al*., 1996). Among Cycadales, *Zamia* is the only genus which shows interspecific and intraspecific aneuploidy, ranging from 2n=16 to 28 (Caputo *et al*., 1996). In ML analysis from combined data, *Z. intergrifolia* and *Z. pygmaea* were grouped together which supported the chromosome number of the two species (2n=16) and the same relative location and number of 45S rDNA signal (Tagashira and Kondo, 2001). The study of chromosome structure in *Zamia* using FISH of 45S rDNA and 5S rDNA showed that the aneuploid *Zamia* resulted from centric fission or fusion of chromosomes with structure differentiation of heterochromatin 45S rDNA regions (Tagashira and Kondo, 2001). Therefore, these results provided conclusive evidence that *Chigua* as a genus was invalid, and that the single species, *Z. restrepoi* was a distinct species within the genus *Zamia*, as already reported by Lindstrom (2009).

1.3 Phylogenetic utility of non-coding chloroplast DNA region for resolving intergeneric relationships of cycads

In the present study, we selected the four non-coding chloroplast loci; *atp*B-*rbc*L, *psb*M-*trn*D, *trn*L-*trn*F and *trn*S- *trnf*M, to reconstruct phylogenetic trees of the living cycads based on a subset of 43 species. The *psb*M-*trn*D region presented the greatest percentage of parsimony informative (PI) characters (total number of parsimony informative characters divided by the aligned length), while the *trn*L-*trn*F, *trnS*-*trnf*M and the *atp*B-*rbc*L regions provided the second, third and fourth, respectively. However, these regions did not produce clear resolution in some clades based upon bootstrap values. Although the percentage of PI characters for the *trn*L-*trn*F region was less than that for the *psb*M-*trn*D region, the former region provided higher bootstrap support for resolving several clades. This degree of resolution based upon the different loci contrasts to a range of studies where the *trnS*-*trnf*M region provided the greatest number of Potentially Informative Character (PIC), followed by

Copyright by Kasetsart University All rights reserved

the *psb*M-*trn*D region, and with the *trn*L-*trn*F region being the least useful (Shaw *et al*., 2005). Furthermore, Small *et al*. (1998) reported that PIC of *atp*B-*rbc*L intergenic spacer was not found in *Gossypium*.

The *atp*B-*rbc*L intergenic spacer region has been proven to be highly conservative for cycads as evidenced by the low percentage of PI characters and the number of nucleotide substitutions between in-groups and out-groups. The indels showed intergeneric variation but within genera there was low bootstrap support. However, this region showed 58 positions of base substitution. The rate of indels in more distantly related species is significantly lower than that of nucleotide substitutions (Golenberg *et al*., 1993). The indel in *psb*M-*trn*D region were reputed to be large (>100 bp) as shown by Shaw *et al*. (2005) and Johnson *et al*. (2008). This region presented difficulties in sequencing for some cycad species. In the previous analysis, *psb*M-*trn*D showed high interspecific variability in *Atropa belladonna* and *Nicotiana tabacum* (Solanaceae) (Kress *et al*., 2005). Although this represented a fast rate of evolution, the resolution of some clades was ambiguous especially within the subfamily Zamioideae and *Stangeria*.

The *trn*L-*trn*F region presented an even lower percentage of PI characters than the *psb*M-*trn*D region. Nevertheless, it was quite surprising that this region did provide several clades with reasonably bootstrap values. These indels and tandem repeats in this region often provide extremely potential useful for phylogenetic analysis within species (Aoki *et al*., 2003; Jung *et al*., 2004; Horning and Cronn, 2006; Yang and Wang, 2007). The *trnS*-*trnf*M region was generally considered to comprise a rapidly evolving non-coding region (Shaw *et al*., 2005). Mort *et al*. (2007) studied phylogenetic of closely related taxa using nine plastid loci. The result showed that *trn*S-*trn*fM region may provide the most parsimony informative characters but was not useful in supporting nodes in general. However, our work on cycads found a low percentage of PI characters for this region based on coding for the three genes, $trnG^{GCC}$, ycf9 and psbZ. The demonstrated conservatism of this region in cycads means that the region was not suitable for determining low-level phylogenetic relationships in the group.

Copyright by Kasetsart University All rights reserved

1.4 A revised suprageneric classification for the living cycads

The four cpDNA non-coding regions that could be amplified in cycad genera revealed different topologies for individual phylogenetic trees and provided phylogenetic resolution based on variation of topology and the associated bootstrap values. The *trn*L-*trn*F region provided the greater bootstrap support and useful data to resolve phylogenetic relationships.

All four regions indicated that *Dioon* was separated from the remaining genera. If the hypothesis that most cycad genera (with the exception of *Cycas*) were relatively recently diverged, even though the lineage dates to the Permian, then the existence of Paleocene fossils attributed to *Dioon* (Hermsen *et al.*, 2006) tends to support this antiquity and early divergence of the genus. Stevenson (1992) included *Dioon* as the sole genus in the tribe *Diooeae* in the subfamily *Encephalartoideae* of the Zamiaceae; however, it is more appropriate to place *Dioon* within its own subfamily of the Zamiaceae due to its basal position.

Family: *Zamiaceae* (Reichenbach,1837)
Subfamily: *Dioonioideae* (P.Sangin *et al. stat.nov)*
Tribe: *Diooeae* (Schuster, 1932).
Genus: Dioon (Lindley, 1843)
(Type: *Dioon edule* Lindley,1843)

Microcycas and *Zamia* were closely related and *Chigua* was not monophyletic, being embedded in *Zamia. Chigua* has now been relegated to the synonymy of *Zamia* (Lindstrom, 2009). *Chigua, Microcycas* and *Zamia* were all included in the subtribe *Microcycadinae* by Stevenson (1992) and this group appeared to be well-supported by this work, albeit as comprising only the latter two genera.

The continued recognition of the Stangeriaceae comprising *Bowenia* and *Stangeria* was not supported. The African *Stangeria* was most closely related to the genera of the subfamily Zamioideae and *Bowenia* formed a clade (in which it is basal)

Copyright by Kasetsart University All rights reserved

with the family Zamiaceae based on the combined data set. We suggest the following revised suprageneric classification for the extant cycads this classification streamlines (Table 13) and simplifies that of Stevenson (1992) and takes into account the clade strength from this study and others (Hill *et al*., 2003; Rai *et al*., 2003; Bogler and Francisco-Ortega, 2004).

.

Copyright by Kasetsart University All rights reserved

Table 13 A revised suprageneric classification for the living cycads.

Family	Subfamily	Tribe	Genus
Cycadaceae Persoon (contains only the genus *Cycas*)	-	-	*Cycas* L.
Zamiaceae	Encephalartoideae D.Stevenson	Encephalarteae	*Encephalartos* Lehmann *Lepidozamia* Rigel *Macrozamia* Miquel
	Zamioideae	Ceratozamieae D.Stevenson	*Ceratozamia* Brongniart *Stangeria* T.Moore
	Zamioideae	Zamieae	*Microcycas* (Miquel) A.DC. *Zamia* L.
	Dioonioideae* (contains only the genus *Dioon*)	-	*Dioon* Lindley
	Bowenioideae Pilger (contains only the genus *Bowenia*)		*Bowenia* Hook. ex Hook.f.

* *stat.nov*. (P.Sangin, A.J.Lindstrom, G.Kokubugata, P. I. Forster & M.Mingmuang)

Copyright by Kasetsart University All rights reserved

2. Phylogenetic relationships within the family Cycadaceae

2.1 Sequence variability in the family Cycadaceae

In the present study, the results 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 *psb*M-*trn*D, *trn*L-*trn*F and *trn*S-*trnf*M 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 numbers of 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 *trn*S-*trn*G region showed the high divergence and variability to resolve relationships among the closely related taxa (Small *et al*., 2005; Ickert-Bound and Wen, 2006). Although *trn*S- *trnf*M region showed the greater number of PICs reporting by Shaw *et al*. (2005), these was lower percent of parsimony informative character in this study.

The *trn*S-*trn*G region of section *Indosinenses* and section *Cycas* contained minisatellite repeats which revealed high levels of polymorphism in the two sections. This result strongly indicated that two sections were closely related and implied that they may have evolved from a common ancestor. Minisatellites are generally found in nuclear genome but are rarely found in the non-coding chloroplast DNA (Sun and Ma, 2009). Therefore, it should be noted here that this is the first report to detect a minisatellite of the *trn*S-*trn*G region on *Cycas*. The minisatellite repeats in this study were AT-rich that slipped-strand misparing was a major mechanism of the high repetitive sequences (Levinson and Gutman, 1987) and similar to the repeat units observed in some plants e.g. alfalfa (*Medicago sativa* L) (Skinner, 2000)**,** *Orchis palustris* (Cafasso *et al*., 2001), *Anacamptis palustris* (Cozzolino *et al*., 2003) and Taxaceae and Cephalotaxaceae (Hao *et al*., 2009). Furthermore, Isoda *et al*. (2000) have classified *Abies* (Pinaceae) based on non-coding region of chloroplast DNA suggesting that the tandem repeat in this genus was not only variable in copy

Copyright by Kasetsart University All rights reserved

number but in the number of unit types. Although minisatellite sequences were found in the nuclear genomes and seemed to be rare in chloroplast DNA (Berg *et al*., 1995), several studies indicated that minisatellite was present in non-coding regions of chloroplast DNA. The *Cycas* minisatellite identified in this study also appeared in a non-coding sequence.

Furthermore, the results showed that the *trn*S-*trn*G region contained 2 microsatellite repeats in *C. taitungensis* and *C. revoluta* and they were placed in the same clade which is consistent with morphological classification. The fluorescent *in situ* hybridization (FISH) analysis using 5S ribosomal DNA (5S rDNA) probe indicated that *C. taitungensis* and *C. revoluta* showed a 5S rDNA site on the same position, the interstitial region near the terminal in two t chromosomes (Kokubugata, 2003). In addition, phylogenetic study on ribosomal ITS region of mitochondria DNA (mtDNA) and the intergenic spacer between atpB and rbcL genes of chloroplast (cpDNA) were used to trace phylogeographical relationships in *C. taitungensis* and *C. revoluta* and showed the paraphyly of both loci in the two species (Chaiang *et al*., 2009). It implied that these two species shared common ancestors corresponding with the present study.

Most of the microsatellite polymorphism found in this region were mononucleotide repeats $(T)_{n,}$ the same as those of cpDNA which consist of only homonucleotide strand of A_n or T_n (Provan *et al*., 1999). The levels of polymorphism in microsatellites were quite variable across loci and across species. The molecular mechanisms for the development of microsatellite variation are not completely understood. The most common mutational mechanism affecting microsatellite is replication slippage, which is a process involving a gain or contraction of one or more repeat units (Levinson and Gutman, 1987). Other microsatellite mutations are caused by unequal crossing over, nucleotide substitutions, or duplication events (Bachtrog *et al*., 1999, Barrier *et al*., 2000). In addition, Tesfaye *et al*. (2007) suggested that the length variation of microsatellite mostly due to the microstructural mutations. Previous studies in other plants also showed that chloroplast genome contained tandem repeats sequences which provided polymorphism DNA markers (Powell *et*

Copyright by Kasetsart University All rights reserved

al., 1995; King and Ferris, 2002; Ohta *et al*., 2005; Wanke *et al*., 2007) and therefore, chloroplast microsatellite have been preferably used for studies in plant ecology and evolution (Provan *et al*., 2001; Matsuoka *et al*., 2002).

It could be concluded that the minisatellite and microsatellite repeats may provide a valuable marker to infer pattern of molecular relationships in taxa that contain these repeats and have been shown to be useful for resolving phylogenetic relationships among closely related taxa.

2.2 Phylogeny relationships within the family Cycadaceae

The NJ, MP and ML from combined data set of 28 *Cycas* species produced similar tree topologies. These results confirmed that *Cycas* was monotypic. Three analyses clearly showed that section *Cycas* and *Indosinenses* were placed in the same clade, while the other clade 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 they belong to the same subsection *Rumphiae* of section *Cycas*. This was in agreement with the study by Keppel *et al*. (2008) who investigated the relationships within subsection *Rumphiae* of the section *Cycas* using morphological and allozyme data. Although subsection *Rumphiae* are widely distributed from Malaysia to East Africa and Western Pacific oceans (Hill, 1994), this molecular study suggested that within subsection *Rumphiae* showed low genetic variation. These results supported a long-distance seed dispersal mechanism especially that usually occur in *Cycas*. It is known that some species of *Cycas* process seeds with special adaptation, a spongy layer in the seed, which give positive buoyancy in seawater (Dehgan and Yuen, 1983) and hence, their wide distribution while maintaining the genetic composition.

Copyright by Kasetsart University. All rights reserved

In NJ analysis, *C. elephantipes* and *C. siamensis* were grouped together which is in agreement with geographic distribution in Thailand (Hill and Yang, 1999; Lindstrom and Hill, 2002) having two species belong to the section *Indosinenses*. Furthermore, in three analyses, it was found that the two species of section *Wadeanae* (*C. wadei* and *C. curranii*) formed a clade together, 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, the Ryukyu Island in Japan and Fukien province of China (Norstog and Nichols, 1997), while *C. taitungensis* is endemic species in Taiwan (Shen *et al*., 1994). However, the high bootstrap support implied that *C. revoluta* and *C. taitungensis* were closely related and they shared identical base substitutions and indel positions.

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

2.3 Monophyly of Cycadaceae

The results from individual data and combined data suggested that Cycadaceae was a monophylytic group. This finding was congruent with morphological data classification. The female cones of *Cycas* are indeterminate, ovules are born on loosely arranged whorls of megasporophylls, but other genera form a determinate female cone and Cycas has platyspermic seeds (Stevenson, 1992).

Copyright by Kasetsart University All rights reserved

In addition, recent molecular analysis indicated that *Cycas* was identified in the monophyletic group and it is a basal group of remaning genera (Hill *et al*., 2003; Rai *et al*., 2003; Bogler and Francisco-Ortega, 2004).

3. Sex identification markers of *Cycas*

In the present study, only two primers from 80 RAPD primers showed polymorphic bands between male and female *Cycas* which revealed that the chromosome region controlling sex determination in *Cycas* was small and contained very few genes which supported the finding of Alstrom-Rapaport *et al*. (1998) that identified RAPD marker in basket willow. Of the 1,080 RAPD bands examined from 380 primers, only a single 560 bp band was shown to be linked to a sex determination locus. Vinod *et a*l. (2007) identified sex specific marker in dioecious *Pandanus fascicularis* used 89 random amplification polymorphic DNA primers and only one primer amplified a band in the males that was absent in the females. In addition, our result was in accordant with *Cycas* sex chromosome studies that the XY-type sex chromosome system of *Cycas* was investigated by *in situ* hybridization (ISH) and FISH using probes of the telomere sequences and the ribosome DNA (18S rRNA). The short telocentric Y chromosome was clarified to lack of the proximal region and the heterochromatin (Hizume *et al*., 1998). In the female of *C. revoluta*, chromosome numbers 21 and 22 was shown to be long and J-shaped but in the male, only chromosome number 21 was long and J-shaped while chromosome number 22 was short and J-shaped. This suggested that chromosome number 21 and 22 may be the sex chromosome (Segawa *et al*., 1971). The Y chromosome of *Cycas* was smaller than X chromosome and it was possible that the Y chromosome contained sex determination.

In this study, only three ISSR primers amplified a specific fragment for males *Cycas* and the sequence of these primers were di-nucleotide units. This result supported previous studies that there were two ISSR primers could be used to amplify male specific fragments in hop (*Humulus lupulus* L.) using di-nucleotide repeat primers (Danilova *et al*., 2006). Sharma *et al*. (2008) identified the sex of

Copyright by Kasetsart University All rights reserved

Simmondsia chinensis using 42 ISSR primers and only one primer showed sex specific in bulk analysis. This study and other studies showed that these ISSR primers were di-nucleotide repeat units. It indicated that almost target di-nucleotide units were more abundant than tri-nucleotide in *Cycas*. In addition, most of the sex-linked markers showed male associated because male are heterogametic sex (Ainsworth, 2000). The sequence derived from male specific ISSR markers of this result had a GC content of 32-42.5% agreement with male-specific sequences reported in other plants, i.e., 39.9-40.4% for *Cannabis sativa* (Mandolino *et al*., 1998), 33-45% for *Aucuba japonica* (Maki, 2009), 41% for *Pandanus fascicularis* L. (Vinod *et al*., 2007).

Recently, only two researches have studied sex identification in Cycad. The sex discrimination in *Encephalartos natalensis* using RAPD marker was a first report in Cycad. Of the 140 primers were used to amplify the male and female but only one primer (OPD-20) generated a specific band in female (Prakash and Staden, 2006). Gangaopadhyay *et al*. (2007) detected sexual dimorphism of *C. cirsinalis* by RAPD and 2 RAPD primers (OPB-01 and OPB-05) amplified a polymorphism bands between male and female. However, this work was the first report using Sequence Characterized Amplified Region (SCAR) marker was derived from RAPD and ISSR specific band of *Cycas*. These specific sequences also presented mononucleotide, di-nucleotide and tri-nucleotide repeats which indicated that they were repeat-rich regions. Like the finding of X and Y sex chromosomes in *Silene latifolia* had repetitive DNA and shared a significant repetitive DNA content with the autosomes (Scutt *et al*., 1997). In addition, Y chromosome of *Silene latifolia* accumulated repetitive sequence than X chromosome (Marais *et al*., 2008) in accordance with the Y chromosomes in dioecious liverwort *Marchantia polymorpha* which are rich in repeats and transposable elements (Yamato *et al*., 2006). The dioecious plant *Rumex acetosa* has 2 group of sex determination: XX in female and XY_1Y_2 in male. Both Y chromosomes contain a tandem repetitive DNA sequence and dispersed repetitive DNA sequence (Shibata *et al*., 1999). Moreover, Parasnis *et al*. (1999) investigated the presence of microsatellite $(GATA)_n$ to identify sex-specific differences in papaya using southern blotting method and $(GATA)_4$ revealed male specific hybridization in

Copyright by Kasetsart University All rights reserved

papaya. This work and many studies indicated that the sex specific sequences represented repetitive DNA sequences and theses sequence may be associated with the sex determining regions or sex chromosome.

The 9 SCAR primers were designed from one female and six males RAPD and ISSR specific sequences. Most of the SCAR primers could amplify in both males and females *Cycas* with the exception of Clivicola 864 primer amplified in all males and one female. It was concluded that RAPD and ISSR primers in this study could be used to separate sex in *Cycas* but SCAR primers converted from RAPD and ISSR sex-specific fragments did not retain the original sex-specificity. It was possible that these regions did not represent DNA in a heterogametic sex chromosome. ISSR primers were used to discover sex-specific markers that a female-specific band was found in *Pseudocalliergon trifarium*. However, the males and females shared the sequenced 182 bp from female-specific primer. It was not known that male and female specific sequences of *P. trifarium* was either in autosomes or represented partly homologous sex-chromosomal sequences (Korpelainen *et al*., 2008). McLetchie and Tuskan (1994) investigated sex markers in *Populus* using 1,219 RAPD primers. No band was found with 100% within a single sex. Amplified Fragment Length Polymorphism (AFLP) was employed for identification of sex-linked marker in *Pangasianodon gigas* and *P. hypophthalmus*. No primers amplified sex-specific fragment for *P. gigas* while 31 of the 102 primers combination showed sex-specific fragments for *P. hypophthalmus*. However, none of 45 SCAR primers showed sex specificity between males and females *P. hypophthalmus*. The failure to identify sex-associated markers suggested that sex chromosomes or sex determination mechanisms may be absent in *P. hypophthalmus* and *P. gigas* (Sriphairoj *et al*., 2007). Yamato *et al*. (2007) studied Y chromosome genes of *Marchantia polymorpha*. They identified 64 genes, of which 14 were detected in the male genome and expressed in the male reproductive organs. At least 6 males of these genes expressed both in thalli and female sexual organs suggesting that the X and Y chromosome of *M. polymorpha* shared the same ancestral autosome. Several studies have not succeeded in identifying sex marker and these finding supported the results in this study. It is possible that sex

Copyright by Kasetsart University All rights reserved

determination in *Cycas* was controlled by genes located on both autosomes and sex chromosomes.

In addition, the molecular mechanism controlling development of sexual characters in *Cycas edentata* were investigated and found Fortune-1 (*Ft-1*), a novel gene which expressed in male organs of *C. edentata*. *Ft-1* expression was enhanced in male cone, the cone axis, microsporophyll and microsporangia but was reduced in ovule and undetectable in megasporophyll. *Ft-1* was weakly expression in males and female leave. The *Ft-1* gene was significantly overexpression in male reproductive organs of *C. edentata* (Zhang *et al*., 2002). Therefore, the level of gene expression may be used to distinguish the sex genotype of *Cycas*. Moreover, the sequences of male specific ISSR primers in this study were found partial homology with ETSs from the ovule of *C. rumphii* and ESTs from young male strobilus of *Zamia fischeri*. SCAR primers were amplified both male and female *Cycas* genomes because these regions may present in X and Y chromosomes.

.

Copyright by Kasetsart University All rights reserved

CONCLUSION

1. Phylogenetic relationships within the family *Zamiaceae* and *Stangeriaceae*

The results obtained from the molecular phylogenetic analyses using 4 regions of non-coding chloroplast DNA (*atp*B-*rbc*L, *psb*M-*trn*D, *trn*L-*trn*F and *trn*S-*trnf*M) did not agree with morphological data. Zamiaceae and Stangeriaceae were not monophyletic group of which *Bowenia* and *Stangeria* were separated within the different clade. *Dioon* was showed as the most basal genus. Therefore, *Dioon* was placed within its own subfamily of the Zamiaceae due to its basal position. The subfamily Dioonioideae was newly recognized. *Encephalartos* and *Lepidozamia* were closer to each other than *Macrozamia* with these three genera forming a monophyletic group. *Microcycas* and *Zamia* were closely related and *Chigua* was not monophyletic, being embedded in *Zamia. Chigua* has now been relegated to the synonymy of *Zamia*. The *trn*L-*trn*F region provided the greater bootstrap support and useful data to resolve phylogenetic relationships.

2. Phylogenetic relationships within the family Cycadaceae

This was the first phylogenetic study of 27 *Cycas* species based on *trn*S-*trn*G, *psb*M-*trn*D, *trn*L-*trn*F and *trn*S- *trnf*M regions of non-coding chloroplast DNA. Three region *trn*S-*trn*G, *psb*M-*trn*D, *trn*L-*trn*F could be combined into a single data set for phylogenetic analysis and the results showed that *Cycas* could be separated into two clades. The first clade consisted of two sections, *Cycas* and *Indosinenses*, while the other clade contained *Asiorientales*, *Wadeanae* and *Stangerioides*. The newly discovered 5 species of subsection *Rumphiae* were separated into two groups (*C. rumphii* and *C. edentata*) based on base substitution pattern. The *trn*S-*trn*G region contained two microsatellite repeats in *C. taitungensis* and *C. revoluta* and they were placed in the same clade which is consistent with the morphological classification. It should be noted here that this is the first report to detect a minisatellite of the *trn*S-*trn*G region on *Cycas*.

Copyright by Kasetsart University. All rights reserved

3. Sex identification markers of *Cycas*

This study provided some new information on sex identification using 29 ISSR and 80 RAPD primers. One RAPD primer (OPB-18) showed polymorphic bands between males and females *Cycas* while UBC 485 produced a specific fragment in male *C. clivicola*. In addition two ISSR primers (001 and 864) showed male specific fragments of *C. clivicola* and ISSR 001 also amplified male specific fragment of *C. chamaoensis*. The male specific fragment of *C. edentata* was amplified by ISSR 836. These specific fragment sequences were converted into SCAR primer but these markers could not be used to identify the sexes of *Cycas*.

Copyright by Kasetsart University All rights reserved

LITERATURE CITED

Ainsworth, C. 2000. Boy and girl come out to play: the molecular biology of dioecious plants. **Ann. Bot.** 86: 211-221.

Alstrom-Rapaport, C., M. Lascour, Y.C. Wang, G. Roberts and G.A. Tuskan. 1998. Identification of a RAPD marker linked to sex determination in the basket willow (*Salix viminalis* L.). **J. Hered.** 89: 44-49.

Aoki, K., T. Suzuki and N. Murakami. 2003. Intraspecific sequence variation of chloroplast DNA among the component species of evergreen broad-leaved forests in Japan. **J. Plant Res.** 116: 337-344.

Arnedo-Andres, M.S., R. Gil-Ortega, M. Luis-Arteaga and J.I. Hormaza. 2002. Development of RAPD and SCAR markers linked to the Pvr4 locus for resistance to PVY in pepper (*Capsicum annuum* L.). **Theor. Appl. Genet.** 105: 1067-1074.

Arroyo-Garca, R., F. Lefort, M.T. de Andrs, J. Ibáñez, J. Borrego, N. Jouve, F. Cabello and J.M. Martínez-Zapater. 2002. Chloroplast microsatellite polymorphisms in *Vitis* species. **Genome** 45: 1142–1149.

Artabe, A.E., A.B. Zamuner and D.W. Stevenson. 2005. A new genus of Late Cretaceous cycad stem from Argentina, with reappraisal of known forms. **Alcheringa** 29: 87-100.

Bachtrog D., S. Weiss, B. Zangerl, G. Brem and C. Schlötterer. 1999. Distribution of dinucleotide microsatellites in the *Drosophila melanogaster* genome. **Mol. Biol. Evol.** 16: 602–610.

Copyright by Kasetsart University All rights reserved

Barrier, M.E. Friar, R. Robichaux and M. Purugganan. 2000. Interspecific evolution in plant microsatellite structure. **Gene** 241: 101–105.

Bausher, M. G., N. D. Singh, S. B. Lee, R. K. Jansen and H. Daniell. 2006. The complete chloroplast genome sequence of *Citrus sinensis* (L.) Osbeck var 'Ridge Pineapple': organization and phylogenetirelationships to other angiosperms. **BMC Plant Biol.** 6:21.

Bellstedt, D.U., H.P. Linder and E.H. Harley. 2001. Phylogenetic relationships in Disa based on non-coding trnL-trnF chloroplast sequences: Evidence of numerous repeat regions. **Am. J. Bot.** 88: 2088–2100.

Berg, T., T. Moum and S. Johansen. 1995. Variable numbers of simple tandem repeats make birds of the order Ciconiiformes heteroplasmic in their mitochondrial genomes**. Curr. Genet.** 27: 257-262.

Bogler, D.J. and J. Francisco-Ortega. 2004. Molecular systematic studies in Cycads: Evidence from trnL intron and ITS2 rDNA sequences. **Bot. Rev.** 70: 260-273.

Bornet, B. and M. Branchard. 2001. Nonanchored Inter Simple Sequence Repeat (ISSR) marker: reproducible and specific tools for genome fingerprinting. **Plant Mol. Biol. Rep**. 19: 209-215.

Britten, R.J., L. Rowen, J. Williams and R.A. Cameron. 2003. Majority of divergence between closely related DNA samples is due to indels. **Proc. Natl. Acad. Sci. USA.** 100: 4661–4665.

Bryan, G.J., J.M. McNicoll, G. Ramsay, R.C. Meyer and W.S. De Jong. 1999. Polymorphic simple sequence repeat markers in chloroplast genomes of Solanaceous plants. **Theor. Appl. Genet.** 99: 859-867**.**

Copyright by Kasetsart University All rights reserved

Byrne, M.and S.H. James. 1991. Genetic diversity in the cycad *Macrozamia riedlei*. **Heredity** 67: 35-39.

Cafasso, D., G. Pellegrino, A. Musacchio, A. Widmer and S. Cozzolino. 2001. Characterization of a minisatellite repeat locus in the chloroplast genome of *Orchis palustris* (Orchidaceae). **Curr. Genet.** 39: 394-398.

Caputo, P., S. Cozzolino, L. Gaudio, A. Moretti and D.W. Stevenson. 1996. Karyology and phylogeny of some Mesoamerican species of *Zamia* (Zamiaceae). **Am. J. Bot.** 83: 1513-1520.

_____ , _____ , De Luca, A. Moretti and D.W. Stevenson. 2004. Molecular phylogeny of *Zamia*. pp. 149-159. *In* Walters T and R. Osborne, eds. **Cycad Classification Concepts and Recommendations**. CABI Pub, Wallingford, U.K.

Charlesworth, D. 2002. Plant sex determination and sex chromosomes. **Heridity** 88: 94-101.

Chaves-Bedoya, G. and V. Nunez. 2007. A SCAR marker for the sex types determination in Colombian genotype of *Carica papaya*. **Euphytica** 153: 215-220.

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

Chiang, T.Y., B.A. Schaal and C.I. Peng. 1998. Universal primers for amplification and sequencing a noncoding spacer between the *atp*B and *rbc*L genes of chloroplast DNA. **Bot. Bull. Acad. Sin.** 39: 245-250.

Copyright by Kasetsart University All rights reserved

_____, _____ and _____. 2000. Molecular evolution of the *atpB-rbcL* noncoding spacer of chloroplast DNA in the moss family Hylocomiaceae. **Bot. Bull. Acad. Sin.** 41: 85-92.

_____, K-H. Hung, S-J. Moore, X-J. Ge, S. Huang, T-W. Hsu, B.A. Schaal and T.Y. Chiang. 2009. Paraphyly of organelle DNAs in *Cycas* Sect. *Asiorientale*s due to ancient ancestral polymorphisms. **BMC Evol. Biol.** 9: 161.

Cozzolino, S., D. Cafasso, G. Pellegrino, A. Musacchio and A. Widmer. 2003. Molecular evolution of a plastid tandem repeat locus in an Orchid lineage. **J. Mol. Evol.** 57: S41-S49.

_____, _____, _____, _____ and _____. 2004. Hypervariable plastid locus variation and intron evolution in the *Anacamptis palustris* lineage. **Genome** 47: 999-1003.

Curtis, E.S. and M.T. Clegg. 1984. Molecular evolution of chloroplast DNA sequences**. Mol. Biol. Evol.** 1: 291-301.

Dane, F., P. Lang and R. Bakhtiyarova. 2004. Comparative analysis of chloroplast DNA variability in wild and cultivated *Citrullus* species. **Theor. Appl. Genet.** 108: 958–966.

Danilova, T.V. and G.I. Karlov. 2006. Application of inter simple sequence repeat (ISSR) polymorphism for detection of sex-specific molecular marker hop (*Humulus lupulus* L.). **Euphytica** 151: 15-21.

Deguilloux, M.F., M.H. Pemonge and R.J. Petit. 2004. Use of chloroplast microsatellites to differentiate oak populations. **Ann. For. Sci.** 61: 825-830.

Copyright by Kasetsart University All rights reserved

Dehgan, B. and C.K.K. Yuen. 1983. Seed morphology in relation to dispersal, evolution, and propagation of *Cycas* L. **Bot. Gaz.** 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.

De Luca, P., A. Moretti, S. Sabato and G. Siniscalco Gigliano. 1982. A comparative study of cycad mucilages. **Phytochemistry** 21: 1609-1611.

Des Marais, D.L., A.R. Smith, D.M. Britton and K.M. Pryer. 2003. Phylogenetic relationships and evolution of extant horsetails, *Equisetum*, based on chloroplast DNA sequence data (*rbc*L and *trn*L-F). **Int. J. Plant Sci.** 164: 737–751.

Didion, B.A., A.A. Paszek, G.A. Rohrer, H.S. Sun and C.K. Tuggle. 2000. Rapid communication: genetic linkage and physical mapping of a porcine sequence characterized amplified region (SCAR) to chromosome 6. **J. Anim. Sci.** 78: 3195-3196.

Dioh, W., D. Tharreau and M-H. Lebrun. 1997. RAPD-based screening of genomic libraries for positional cloning. **Nucleic Acids Res**. 25: 5130-5131.

Di Stilio, V., R.V. Keseli and D.L. Mulcahy. 1998. A pseudoautosomal random amplified polymorphic DNA marker for the sex chromosomes of *Silence dioica*. **Genetic** 149: 2057-2062.

Downie, S.R. and J.D Palmer. 1991. Use of chloroplast DNA rearrangements in reconstructing plant phylogeny, pp 14–35. *In* P.S Soltis, D.E. Soltis and J.J. Doyle,eds. **Molecular Systematics of Plants.** Chapman and Hall, New York.

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

Copyright by Kasetsart University All rights reserved

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

Gangopadhyay, G., S.K. Roy, K. Ghose, R. Poddar, T. Bandyopadhyay, D. Basu and K.K Mukherjee. 2007. Sex detection of *Carica papaya* and *Cycas circinalis* in pre-flowering stage by ISSR and RAPD. **Curr. Sci.** 92: 524-526.

González, D., A.P. Vovides and C. Barcenas. 2008. Phylogenetic relationships of the Neotropical Genus *Dioon* (Cycadales, Zamiaceae) based on nuclear and chloroplast DNA sequence data. **Syst. Bot.** 33: 229-236.

Golenberg, E.M., M.T. Clegg, M.L. Durbin, J. Doebley, and D.P. Ma. 1993. Evolution of a noncoding region of the chloroplast genome. **Mol. Phylogent. Evol.** 2**:** 52-64.

Griffiths, R. and B. Tiwari. 1993. The isolation of molecular genetic markers for the identification of sex. **Proc. Natl. Acad. Sci. USA.** 90: 8324-8326.

Guicking, D., T. Kpöger-Kilian, K. Weising and F. R. Blattner. 2008. Single nucleotide sequence analysis: a cost-and time-effective protocol for the analysis of micro-satellite and indel-rich chloroplast DNA regions. **Mol. Ecol. Res.** 8: 62-65.

Hallick, R.B., L. Hong, R.G. Drager, M.R. Favreau, A. Monfort, B. Orsat, A. Spielman and E. Stutz. 1993. Complete sequence of *Euglena gracilis* chloroplast DNA. **Nucleic Acid Res.** 21: 3537–3544.

Hao, D.C., B.L. Huang, S.L. Chen and J. Mu. 2009. Evolution of the chloroplast *trn*L-*trn*F region in the Gymnosperm Lineages Taxaceae and Cephalotaxaceae. **Biochem. Genet.** 47: 351-369.

Copyright by Kasetsart University All rights reserved

Harbourne, M. E., G. C. Douglas, S. Waldren and T. R. Hodkinson. 2005. Characterization and primer development for amplification of chloroplast microsatellite regions of *Fraxinus excelsior*. **J. Plant Res.** 118: 339-341.

Harvery, C.F., G.P. Gill, L.G. Fraser and M.A. McNeilage. 1997. Sex determination in *Actinidia*. 1. Sex-linked markers and progeny sex ratio in diploid *A. chinensis*. **Sex Plant Reprod.** 10: 149-154.

Hermsen, E.J., T.N. Taylor, E.L. Taylor and D.W. Stevenson. 2006. Cataphylls of the middle Triassic cycad *Antarcticycas schopfii* and new insights into cycad evolution. **Am. J. Bot.** 93: 724-738.

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.

_____ , M.W. Chase, D.W. Stevenson, H.G. Hills and B. Schutzman. 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.

_____ . 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.

_____ , 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.

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

Copyright by Kasetsart University All rights reserved

Hizume, M., N. Kurose, F. Shibata and K. Kondo. 1998. Molecular cytogenetic studies on sex chromosomes and proximal heterochromatin containing telomere-like sequence in *Cycas revolute*. **Chromosome Science** 2: 63-72.

Hormaza, J.I., L. Dollo and V.S. Polito. 1994. Identification of a RAPD marker linked to sex determination in Pistacia vera using bulked segregant analysis. **Theor. Appl. Genet**. 89: 9-13.

Horning, M.E. and R.C. Cronn. 2006. Length polymorphism scanning is an efficient approach for revealing chloroplast DNA variation. **Genome** 49: 134-142.

Huang, S., Y. C. Chiang, B. A. Schaal, C. H. Chou, and T. Y. Chiang. 2001. Organelle DNA phylogeography of *Cycas taitungensis*, a relict species in Taiwan. **Mol. Ecol.** 10: 2669–2681.

_____ , W.K. Wang, C.I. Peng and T.Y. Chiang. 2005. Phylogeography and conservation genetics of *Hygrophila pogonocalyx* (Acanthaceae) based on *atp*B-*rbc*L noncoding spacer cpDNA. **J. Plant Res.** 118: 1-11.

Huaracha, E., M. Xu and S.S. Korban. 2004. Narrowing down the region of the *Vf* locus for scab resistance in apple using AFLP-derived SCARs. **Theor. Appl. Genet.** 108: 274-279.

Ickert-Bond, S.M. and J. Wen. 2006. Phylogenetic and biogeography of Altingiaceae: evidence from combined analysis of five non-coding chloroplast regions. **Mol. Phylogenet. Evol.** 39: 512-528.

Ingvarsson, P.K., S. Ribstein and D.R. Taylor. 2003. Molecular evolution of insertions and deletion in the chloroplast genome of *Silene*. **Mol. Biol. Evol.** 20: 1737-1740.

Irish, E.E. and T. Nelson. 1986. Sex Determination in Monoecious and Dioecious Plants. **Plant Cell** 1: 737-744.

Copyright by Kasetsart University All rights reserved

Isoda, K., S. Shiraishi and H. Kisanuki. 2000. Classifying *Abies* species (Pinaceae) based on the sequence variation of a tandemly repeated array found in the chloroplast DNA *tm*L and *tm*F intergenic spacer. **Silvae Genetica** 49: 161–165.

Jakobsson, M., T. Säll, C. Lind- Halldén and C. Halldén. 2007. Evolution of chloroplast mononucleotide microsatellites in *Arabidopsis thaliana*. **Theor. Appl. Genet.** 114: 223-235.

Jiang, C. and K. C. Sink. 1997. RAPD and SCAR markers linked to the sex expression locus *M* in asparagus. **Euphytica**. 94: 329-33.

Johnson, L.A.S. 1959. The families of cycads and the Zamiaceae of Australia. **Proc. Linn. Soc. New South Wales.** 84: 64-117.

Johnson, L.A., L.M. Chan, T.L. Weese, L.D. Busby and S. McMurry. 2008. Nuclear and cpDNA sequences combined provide strong inference of higher phylogenetic relationships in the phlox family (Polemoniaceae). **Mol. Phylogenet. Evol**. 48: 997-1012.

Juarez, C. and J.A. Banks. 1998. Sex determination in plants. **Curr. Opin. Plant Biol.** 1: 68-72.

Jung, Y.H., E.Y. Song, S.J. Chun, K.C. Jang, M. Kim, S.H. Kang and S.C. Kim. 2004. Phylogenetic analysis of plastid trnL-trnF sequences from *Arisaema* species (Araceae) in Korea. **Euphytica** 138: 81-88.

Kelchner**,** S. A. 2000**.** The evolution of non-coding chloroplast DNA and its application in plant systematics. **Ann. Missouri Bot. Gard.** 87: 482-498.

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.

Copyright by Kasetsart University All rights reserved

Khadka, D.K., A. Nejidat, M. Tal and A Golan-Goldhirsh. 2002. DNA markers for sex: molecular evidence for gender dimorphism in dioecious *Mercurialis annua* L. **Mol. Breeding** 9: 251-257.

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.

King, R.A. and C. Ferris. 2002. A variable minisatellite sequence in the chloroplast genome of *Sorbus* L. (Rosaceae: Maloideae). **Genome** 45: 570-576.

Kokubugata, G., K. Kondo, G.W. Wilson, L.M. Randall, A. van der Schans and D.K. Morris. 2000. Comparison of karyotype and rDNA-distribution in somatic chromosomes of *Bowenia* species (Stangeriaceae, Cycadales). **Aust. Syst. Bot.** 13: 15-20.

_____ , K.D. Hill and K. Kondo. 2002. Ribosomal DNA distribution in somatic chromosomes of *Stangeria eriopus* (Stangeriaceae, Cycadales) and molecular–cytotaxonomical relationships to the other cycad genera. **Brittonia** 54: 1–5.

_____ , A.P. Vovides and K. Kondo. 2004. Mapping 5S ribosomal DNA on somatic chromosome of four species of *Ceratozamia* and *Stangeria eriopus* (Cycadales). **Bot. J. Linn. Soc.** 145: 499-504.

Korpelainen, H., I. Bisang, L. Hedenas and J. Kolehmainen . 2008. The First Sex-Specific Molecular Marker Discovered in the Moss ***Pseudocalliergon trifarium*.** J. Hered. 99:581-587.

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

Copyright by Kasetsart University All rights reserved

Kvaćek, Z. and S.R. Manchester. 1999. *Eostangeria Barthel* (extinct Cycadales) from the paleogene of western North America and Europe. **Int. J. Plant Sci.** 160: 621-629.

_____ . 2002. A new tertiary *Ceratozamia* (Zamiaceae, Cycadopsida) from the European Oligocene. **Flora** 197: 303-316.

Lee, C. and J. Wen. 2004. Phylogeny of *Panax* using chloroplast *trn*C – *trn*D intergenic region and the utility of *trn*C – *trn*D in interspecific studies of plants. **Mol. Phylogenet. Evol.** 31: 894-903.

Levinson, G. and G.A. Gutman. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. **Mol. Biol. Evol.** 4: 203-221**.**

Li, W.H. 1997. **Molecular Evolution**. Sinauer Associates, Sunderland, MA, USA

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

_____ , _____ and L.C. Stanberg. 2008. The genus *Cycas* (Cycadaceae) in the Philippines. **Telopea** 12: 119-145.

Maki, M. 2009. Development of SCAR markers for sex determination in the dioecious shrub *Aucuba japonica* (Cornaceae). **Genome** 52: 231–237.

Mandolino, G., A. Carboni, S. Forapani, V. Faeti and P. Ranalli 1999. Identification of DNA markers linked to the male sex in dioecious hemp (Cannabis sativa L.). **Theor. Appl. Genet.** 98: 86–92.

Copyright by Kasetsart University All rights reserved

Manen, J-F.and A. Natali, 1995. Comparison of the evolution of ribulose-1, 5-biphosphate carboxylase (*rbc*L) and atpB-*rbc*L noncoding spacer sequence in a recent plant group, the tribe*Rubieae* (*Rubiaceae*). **J. Molec. Evol.** 41: 920–927.

Marais, G. A., M. Nicolas, R. Bergero, P. Chambrier and E. Kejnovsky. 2008. Evidence for degeneration of the Y chromosome in the dioecious plant Silene latifolia. **Curr. Biol.** 18: 545-549.

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.

McDade, L.A. and M.L. Moody. 1999. Phylogenetic relationships among Acanthaceae: evidence from noncoding trnL-trnF chloroplast DNA sequences **Am. J. Bot.** 86: 70–80.

McLetchie, D.N., G.A. Tuskan. 1994. Gender determination in *Populus.* **Norw. J. Agric. Sci.** 18:57–66.

Mes, T.H.M., P. Kuperus, J. Krischer, J. Stepanek, P. Oosteveld, H. Storchova and J.C. den Nijs. 2000. Hairpins involving both inverted and direct repeats are associated with homoplasious indels in non-coding chloroplast DNA of *Taraxacun* (Lactuceae: Asteraceae). **Genome** 43: 634-641.

Michelmore, R.W., I. Paran and R.V. Kesseli. 1991. Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic region by using segregating populations. **Proc. Natl. Acad. Sci. USA.** 88: 9828-9832.

Copyright by Kasetsart University All rights reserved

Minami, M., K. Nishio, Y. Ajioka, H. Kyushima, K. Shigeki, K. Kinjo, K. Yamada, M. Nagai, K. Satoh and Y. Sakurai. 2009. Identification of Curcuma plants and curcumin content level by DNA polymorphisms in the trnS-trnfM intergenic spacer in chloroplast DNA. **J. Nat. Med.** 63: 75–79.

Ming, R., J. Wang, P.H. Moore and A.H. Paterson. 2007. Sex chromosome in flowering plants. **Am. J. Bot.** 94: 141-150.

Moore, R.C. 2004. Autosomes behaving badly. **Heredity** 93: 126-127.

Moretti, A., P. Caputo, S. Cozzolino, P.D. Luca, L. Gaudio, G.S. Gigliano and D.W. Stevenson. 1993. A phylogenetic analysis of *Dioon* (Zamiaceae). **Am. J. Bot.** 80: 204-214.

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.

Mulcahy, D.L., N.F. Weeden, R. Kesseli and S.B. Carroll. 1992. DNA probes for the Y-chromosome of *Silene latifolia*, a dioecious angiosperm. **Sex Plant Reprod.** 5: 86-88.

Nakao, S., S. Matsunaga, A. Sakai, T. Kuroiwa and S. Kawano. 2002. RAPD isolation of a Y chromosome specific ORF in a dioecious plant, *Silene latifolia*. **Genome** 45: 413-420.

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**. Ithaca. New York: Cornell University Press.

Copyright by Kasetsart University All rights reserved

Ohta, S., C. Nishitani and T. Yamamoto. 2005. Chloroplast microsatellites in *Prunus*, Rosaceae. **Mol. Ecol. Notes** 5: 837–840.

Pakkad, G., S. Ueno and H. Yoshimaru. 2008. Genetic diversity and differentiation of *Quercus semiserrata* Roxb. in northern Thailand revealed by nuclear and chloroplast microsatellite markers. **Forest Ecol. Manag.** 255:1067–1077.

Palmer, J.D. 1991. Plastid chromosomes: structure and evolution, pp 5–53. *In* L.Bogorad and I.K. Vasil (eds.). **The Molecular Biology of Plastids**. Cell Culture and Somatic Cell Genetics of Plants, Vol. 7A. Academic Press, San Diego.

Paran, I. and R.W. Michelmore. 1993. Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. **Theor. Appl. Genet.** 85: 985-993.

Parasnis, A.S., W. Ramakrishna, K.V. Chowdari, V.S. Gupta and P.K. Ranjekar. 1999. Microsatellite (GATA)n reveals sex specific differences in Papaya. **Theor. Appl. Genet.** 99: 1047–1052.

_____ , V.S. Gupta, S.A. Tamhankar and P.K. Ranjekar. 2000. A high reliable sex diagnostic PCR assay for mass screening of papaya seedling. **Mol. Breeding** 6: 337-344.

Pombert, J.F., C. Lemieux and M. Turmel. 2006. The complete chloroplast DNA sequence of the green alga *Oltmannsiellopsis viridis* reveals a distinctive quadripartite architecture in the chloroplast genome of early diverging ulvophytes. **BMC Biol**. 4: 3

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

Copyright by Kasetsart University All rights reserved

Powell, W., M. Morgante, R. McDevitt, G.G. Vendramin and J.A. Rafalski. 1995. Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. **Proc. Natl. Acad. Sci. USA.** 92: 7759-7763.

Prakash, S. and J.V. Staden. 2006. Sex identification in *Encephalartos natalensis* (Dyer and Verdoorn) using RAPD markers. **Euphytica** 152: 197–200.

_____ and _____ . 2008. Genetic variability and species identification within *Encephalartos* using Random Amplified Polymorphic DNA (RAPD) marker. **S. African J. Bot.** 74: 735-739.

Prakash, S., N. Grobbelaar and J.V. Staden. 2008. Diversity in *Encephalartos woodii* collections based on Random Amplified DNA markers (RAPD's) and Inter-Specific Sequence Repeats (ISSR's). **S. African J. Bot.** 74: 341-344.

Provan, J., N. Soranzo, N.J. Wilson, D.B. Goldstein and W. Powell. 1999. A low mutation rate for chloroplast microsatellites. **Genetics** 153: 943–947.

_____, 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.

Copyright by Kasetsart University All rights reserved

Ruas, C.F., D.J. Fairbanks, R.P. Evans, H.C. Stutz, W.R. Andersen and P.M. Ruas. 1998. Male-specific DNA in the dioecious species *Atriplex garrettii* (Chenopodiaceae). **Am. J. Bot.** 82: 162-167.

Sakamoto, K., T. Abe, T. Matsuyama, S. Yoshida, N. Ohmido, K. Fukui and S. Satoh. 2005. RAPD markers encoding retrotransposable element are linked to the male sex in Cannabis sativa L. **Genome** 48: 931-936.

Sambrook, J., E.F. Fritsoh and T. Maniatis. 1989. **Molecular Cloning: A laboratory Manual, 2nd Ed.** Cold Spring Harbor Laboratory Press, Plainview, New York.

Sanger, F., S. Nicklen and A.R. Coulson. 1997. DNA sequencing with chain-terminating inhibitors. **Proc. Natl. Acad. Sci. USA.** 74: 5463-5467.

Schwendemann, A.B., T.N. Taylor and E.L. Taylor. 2009. Pollen of the Triassic cycad *Delemaya spinulosa* and implications on cycad evolution. **Rev. Palaeobot. Palynol.** 156: 98-103.

Scutt, C.P., Y. Li, S.E. Robertson, M.E. Willis and P.M. Cilmartin. 1997. Sex determination in dioecious *Silene latifolia*. Effects of the Ychromosome and the parasitic smut fungus *(Ustilago violacea)* on gene expression during flower development. **Plant Physiol.** 114: 969-979.

Segawa, M., S. Kishi and S. Tatsuno. 1971. Sex chromosomes of Cycas revoluta. **Jpn. J. Genet.** 46: 33-39.

Selvaraj, D., R.K. Sarma and R. Sathishkumar. 2008. Phylogenetic analysis of chloroplast *matK* gene from Zingiberaceae for plant DNA barcoding. **Bioinfomation** 3: 24-27.

Copyright by Kasetsart University All rights reserved

Sharma, K., V. Agrawal, S. Gupta, R. Kumar and M. Prasad. 2008. ISSR marker-assisted selection of male and female plants in a promising dioecious crop: jojoba (*Simmondsia chinensis*). **Plant Biotechnol. Rep.** 2: 239–243.

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

Shibata, F., M. Hizume and Y. Kuroki. 1999. Chromosome painting of Y-chromosomes and isolation of Y-chromosome-specific repetitive sequence in the dioecious plant *Rumex acetosa*. **Chromosoma** 108: 266–270**.**

Skinner, D.Z. 2000. Non-random chloroplast DNA hypervariability in *Medicago sativa*. **Theor. Appl. Genet.** 101: 1242-1249.

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.

Sriphairoj, K., W. Kamonrat, U. Na-Nakorn. 2007. Genetic aspect in broodstock management of the critically endangered Mekong giant catfish, *Pangasianodon gigas* in Thailand. **Aquaculture** 264: 36–46.

Copyright by Kasetsart University All rights reserved

Stehlik, I. and F.R. Blattner. 2004. Sex-specific SCAR markers in the dioecious plant *Rumex nivalis* (Polygonaceae) and implication for the evolution of sex chromosomes. **Theor. Appl. Genet.** 108: 238-242.

Stevenson, D.W. 1990. Morphology and systematics of the Cycadales. **Mem. NY Bot. Gard.** 57: 8–55.

_____ . 1992. A formal classification of the extant cycads. **Brittonia** 44: 220-223.

_____ and G. Siniscalco Giliano. 1989. The systematic value of the monosaccharide composition and distribution pattern of cycad mucilages. **Biochem. Syst. Ecol.** 17: 185–190.

Stoneberg-Holt, S.D., L. Horová and P. Bureŝ. 2004. Indel patterns of the plastid DNA *trnL-trnF* region within the genus *Poa* (Poaceae). **J. Plant Res.** 117: 393–407.

Su, Y. J., T. Wang, B. Zheng, Y. Jiang and G. P. Chen. 2005. Genetic differentiation of relictual populations of Alsophila spinulosa in southern China inferred from cpDNA *trn*L–F noncoding sequences. **Mol. Phylogenet. Evol.** 34: 323–333.

Sun, G. and X. Ma. 2009. Nucleotide diversity and minisatellite in chloroplast Asp(GUC)-Thr(GGU) region in *Elymus trachycaulus* complex, *Elymus alaskanus* and *Elymus caninus*. **Biochem. Syst. Ecol.** 37: 67-75.

Suyama, Y., H. Yoshimura and Y. Tsumura. 2000. Molecular phylogenetic position of Japanese *Abies* (Pinaceae) based on chloroplast DNA sequences. **Mol. Phylogenet. Evol.** 16: 271–277.

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

Copyright by Kasetsart University All rights reserved

Tagashira, N. and K. Kondo. 2001. Chromosome phylogeny of *Zamia* and *Ceratozamia* by means of Robertsonian changes detected by fluorescence *in situ* hybridization (FISH) technique of rDNA. **Pl. Syst. Evol.** 227: 145-155.

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.

Tesfaye, K., T. Borsch, K. Govers and E. Bekele. 2007. Characterization of *Coffea* chloroplast microsatellites and evidence for the recent divergence of *C. arabica* and *C. eugeniodes* chloroplast genomes. **Genome** 50: 1112–1129.

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.

Tian, S., L-C. Luo, S. Ge and Z-Y. Zhang. 2008. Clear Genetic Structure of *Pinus kwangtungensis* (Pinaceae) Revealed by a Plastid DNA Fragment with a Novel Minisatellite. **Ann. Bot.** 102: 69-78.

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

Turmel M, C. Otis and C. Lemieux. 1999. The complete chloroplast DNA sequence of the green alga *Nephroselmis olivacea*: insights into the architecture of ancestral chloroplast genomes. **Proc. Natl. Acad. Sci. USA.** 96: 10248–10253.

Copyright by Kasetsart University All rights reserved

_____ , _____ and _____ . 2005. The complete chloroplast DNA sequences of the charophycean green algae *Staurastrum* and *Zygnema* reveal that the chloroplast genome underwent extensive changes during the evolution of the Zygnematales. **BMC Biol.** 3: 22.

Urasaki, N., M. Tokumoto, K. Tarora, Y. Ban, T. Kayano, H. Tanaka, H. Oku, I. Chinen and R. Terauchi. 2002. A male and hermaphrodite specific RAPD marker for papaya (*Carica papaya* L.). **Theor. Appl. Genet.** 104: 281-285.

Vinod, M.S., P.S. Raghavan, S. George and A. Parida. 2007. Identification of a sex-specific SCAR maker in dioecious Pandanus fascicularis L. (Pandanaceae). **Genome** 50: 834-839.

Wanke, S., M.A. Jaramillo, T. Borsch, M.S. Samain, D. Quandt and C. Neinhuis. 2007. Evolution of Piperales-*mat*K gene and *trn*K intron sequence data reveal lineage specific resolution contrast. **Mol. Phylogenet. Evol.** 42: 477–497.

Williams, J.G.K., A.R. Kubelik, K.J. Livak, J.A. Rafalski and S.V. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. **Nucleic Acid Res.** 18: 6531-6535.

Wu**,** C.S., Y.N. Wang, S.M. Liu and S.M. Chaw. 2007. Chloroplast genome (cpDNA) of Cycas taitungensis and 56 cp protein-coding genes of Gn**et**um parvifolium: insights into cpDNA evolution and phylogeny of extant seed plants. **Mol. Biol. Evol.** 24: 1366–1379.

Xiang, J. Q-Y., M. L. Moody, D. E. Soltis, C. Z. F. Fan and P. S. Soltis. 2002. Relationships within Cornales and circumscription of Cornaceae-*matK* and *rbcL* sequence data and effects of outgroups and long branches. **Mol. Phylogenet. Evol.** 24: 35-57.

Copyright by Kasetsart University All rights reserved

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.

Yamato, K.T., K. Ishizaki, M. Fujisawa, S. Okada, S. Nakayama, M. Fujishita, H. Bando, K. Yodoya, K. Hayashi, T. Bando, A. Hasumi, T. Nishio, R. Sakata, M. Yamamoto, A. Yamaki, M. Kajikawa, T. Yamano, T. Nishide, S-H. Choi, Y. Shimizu-Ueda, T. Hanajiri, M. Sakaida, K. Kono, M. Takenaka, S. Yamaoka, C. Kuriyama, Y. Kohzu, H. Nishida, A. Brennicke, T. Shin-i, Y. Kohara, T. Kohchi, H. Fukuzawa and K. Ohyama. 2007. Gene organization of the liverwort Y chromosome reveals distinct sex chromosome evolution in a haploid system. **Proc. Natl. Acad. Sci. USA.** 104: 6472–6477.

Yang, F.S. and X.Q. Wang. 2007. Extensive length variation in the cpDNA trnT-trnF region of hemiparasitic *Pedicularis* and its phylogenetic implication. **Pl. Syst. Evol.** 264: 251-264.

Yoo, K-O., P.P. Lowry II and W. Jun. 2002. Discordance of chloroplast and nuclear ribosomal DNA data in *Osmorhiza* (Apiaceae). **Am. J. Bot.** 89: 966-971.

Yuan, Y.W. and R.G. Olmstead. 2008. A species level phylogenetic study of the *Verbena* complex (Verbenaceae) indicates two independent intergeneric chloroplast transfers. **Mol. Phylogent. Evol.** 48: 23–33.

Yu, C.Y, Hu, S.W. H.X. Zhao, A.G. Gua and G.L. Sun. 2005. Genetic distances revealed by morphological characters, isozyme, protein and RAPD markers and their relationships with hybrid performance in oilseed rape (*Bassica napus* L.). **Theor. Appl. Genet.** 110: 511-518.

Copyright by Kasetsart University All rights reserved

Yue, J.P., H. Sun, D.A. Baum, I.L. Al-Shehbaz and R. Ree. 2009. Molecular phylogeny of *Solms-laubachia* (Brassicaceae) s.l., based on multiple nuclear and plastid DNA sequences, and its biogeographic implications. **J. Syst. Evol.** 47: 402-415.

Zgurski, J.M., H.S. Rai, Q.M. Fai, Bogler, D.J., J. Francisco-Ortega, S.W. Graham. 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.

Zhifeng, G. and B.A. Thomas. 1989**.** A review of fossil cycad megasporophylls, with new evidence of *Crossozamia pommel* and its associated leaves from the lower Permian of Taiyuan, China. **Rev. Palaeobot. Palynol.** 60: 205–223.

Zhang, P., K-H. Pwee, H.T.W. Tan and P.P. Kumar. 2002. Cloning and characterization of *Fortune-1*, a novel gene with enhanced expression in male reproductive of *Cycas edentata*. **Mech. Dev.** 114: 149-152.

Zietkiewicz, E., A. Rafalski and D. Labuda. 1994. Genome fingerprint by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. **Genomic** 20: 176-183.

Zuber, D. and A. Widmer. 2000. Genetic evidence for host specificity in the hemi-parasitic *Viscum album* L. (Viscaceae). **Mol. Ecol.** 9: 1069-1073.

Copyright by Kasetsart University. All rights reserved

APPENDIX

Copyright by Kasetsart University All rights reserved

Appendix Table A1 Non-coding chloroplast DNA regions amplified with primers.

Region	Primers	References
*trn*L^{UAA}-*trn*F^{GAA}	*trn*L^{UAA}: CGAAATCGTAGACGCTACG	Taberlet *et al*. (1991)
	*trn*F^{GAA}: ATTTGAACTGGTGACACGAG	Taberlet *et al*. (1991)
trnS-*trnf*M	*trnS*: GAGAGAG AGGG ATTGCAACC	Demesure *et*
	*trnf*M: CATAACCTTGAGGTCAGCGG	*al*.(1995)
		Demesure *et*
*trn*S^{GUC}-*trn*G ^{UCC}	*trn*S^{GUC}: GCCGCTTTAGTCCACTCAGC	*al*.(1995)
	*trn*G ^{UCC}: GAACGAATCACACTTTTACCAC	
		Hamilton (1999)
*psb*M-*trn*D	*psb*M: AGCAATAAATGCRAGAATATT	Hamilton (1999)
	TACTTCCAT	
	*trn*D: GGGATTGTAGYTCAATTGGT	Shaw *et al*. (2005)
*atp*B-*rbc*L	*atp*B : ACATCKARTACKGGACCAATAA	Shaw *et al*. (2005)
	*rbc*L: AACACCAGCTTTRAATCCAA	
		Chiang *et al*. (1998)
		Chiang *et al*. (1998)

Copyright by Kasetsart University All rights reserved

Appendix Table A2 Twenty-nine ISSR primers used in this study.

Name	Sequence	Annealing Temperature (°C)
ISSR 000	$(CA)_8AT$	50
ISSR 001	$(CA)_8RG$	52
ISSR 002	$(GTG)_5AC$	55
ISSR 003	$TCGC(CT)_6C$	55
ISSR 807	$(AG)_8T$	50
ISSR 808	$(AG)_8C$	52
ISSR 810	$(GA)_8AT$	50
ISSR 811	$(GA)_8AC$	52
ISSR 815	$(TC)_8G$	52
ISSR 817	$(CA)_8A$	50
ISSR 819	$(GT)_8A$	50
ISSR 822	$(TC)_8A$	50
ISSR 835	$(AG)_8YC$	53
ISSR 836	$(AG)_8YA$	50
ISSR 840	$(GA)_8AYT$	50
ISSR 842	$(GA)_8YG$	52
ISSR 843	$(CT)_8RA$	50
ISSR 847	$(CA)_8RC$	52
ISSR 851	$(GT)_8YG$	52
ISSR 853	$(TC)_8RT$	50
ISSR 857	$(AC)_8YG$	52
ISSR 861	$(ACC)_6$	58
ISSR 864	$(ATG)_5$	40
ISSR 876	$(GATA)_2\ (GACA)_2$	45
ISSR 880	$(GGAGA)_3$	50
ISSR 882	$VBV(AT)_7$	35
ISSR 884	$HBHA(GA)_6G$	50
ISSR 888	$BDB(CA)_7$	50
ISSR 891	$HVHT(GT)_6G$	50

Copyright by Kasetsart University All rights reserved

CURRICULUM VITAE

NAME : Miss Pattamon Sangin

BIRTH DATE : December 17, 1976

BIRTH PLACE : Nakhonratchasima, Thailand

EDUCATION :

YEAR	INSTITUTE	DEGREE/DIPLOMA
1997	Khonkaen Univ.	B.Sc. (Biochemistry)
2001	Kasetsart Univ.	M.Sc. (Cell and molecular biology)

SCORLARSHIP/AWARDS : Thai Government Scholarship 2005-2008

Copyright by Kasetsart University All rights reserved