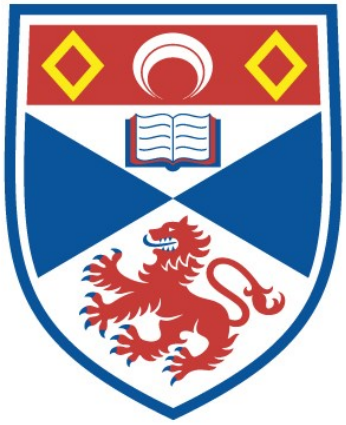
MOLECULAR SYSTEMATICS OF THE TROPICAL AQUATIC PLANT GENUS, 'CRYPTOCORYNE' FISCHER EX WYDLER (ARACEAE)

Ahmad Sofiman Othman

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



1997

Molecular Systematics of the Tropical Aquatic Plant Genus, *Cryptocoryne* Fischer ex Wydler (Araceae).

by

Ahmad Sofiman Othman

A thesis submitted to the
University of St. Andrews for
the degree of Doctor of Philosophy

School of Biological and Medical Sciences
University of St. Andrews
November 1997

This thesis is dedicated to my late grandparents, Yan, Tok Bedah and Tok Aji.

Abstract

Cryptocoryne is an aquatic plant genus occurring in tropical Asia. A number of species within the genus are widely used as foliage plants in tropical fish aquaria. The taxonomy and systematics of the genus have been based in the past on morphological characteristics, geographical distribution and chromosome number. The aim of this research was to construct a phylogeny for the genus based on molecular variation and to establish relationships among species in areas of different geographical origin and of different chromosome base number.

A phylogeny constructed from sequence variation of the internal transcribed spacer (ITS) region of nuclear ribosomal DNA, separated the genus into three distinct clades reflecting a strong geographical pattern. One clade consisted of species from mainland Asia, another was comprised of species from Sri Lanka while the third clade consisted of species from Malesia. However, these clades did not reflect rigidly the geographical origin of areas in which the genus is distributed. The ITS phylogeny also indicated that species with a base chromosome number of x=14 are derived from x=18 stock, and that x=11 has arisen twice in the course of the evolution of the genus. The origin of other chromosome base numbers, i.e. x=15 and x=17, was not clearly resolved. A phylogeny based on chloroplast (cp) DNA sequence variation was similar in topology to that of the ITS phylogeny although resolution of species relationships was not so good. Phenograms produced from RAPD variation gave a better resolution of affinities among closely related *Cryptocoryne* species from Sri Lanka, but differed in overall topology from the phylogenies obtained from ITS and cpDNA sequence variation.

Compelling evidence for the origin of the hybrid *C. x willisii* from *C. walkerii* and *C. parva* was obtained from the RAPD analysis, in that *C. x willisii* exhibited an additive RAPD profile which combined the profiles of its two putative parents. *C. x willisii* also contained an ITS sequence similar to that of *C. walkeri* in some accessions examined, and a cpDNA sequence similar to *C. walkeri* and its close relative *C. beckettii*.

COPYRIGHT

27 November, 1997. A. Sofiman Othman

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank my supervisor Dr. R.J. Abbott for his advice over the duration of my study and for patiently going through earlier drafts of this thesis. Special thanks are also due to Professor N. Jacobsen of RVAU, Copenhagen for his support in providing plant material and comments made on the research and Dr. M. Chase of Royal Botanic Gardens, Kew, for assissting me in DNA sequencing analysis. Grateful thanks are extended to Dr. M. Mansor, Dr. J. Bastmeijer, Mr. P. Boyce and Mr. C. Kidd for their enthusiastic collaborative support. In addition, I would like to thank Universiti Sains Malaysia for provision of my studentship. Finally, thanks are due to Mum and Dad for their encouragement and support and to all my friends and colleagues both in Plant Science Laboratory, University of St. Andrews and molecular systematics laboratory, Kew Gardens (from 1994 to 1997) who have taught and assisted me in molecular analysis and for providing an enjoyable working environment.

CONTENTS

Chapter 1

General introduction

General Introduction

Cryptocoryne is an aquatic plant genus within the aroid family, Araceae. It is native to south east Asia extending from Mainland India and Indo-China through Indonesia to Papua New Guinea. Species within the genus thrive in pristine rivers with sandy and rocky sediment (Plate 1.1 and 1.2) or in swampy areas under a thick canopy deep in the forest. Many of these areas which occur at elevations of 0 - 300 metres above sea level are usually not easily accessible by man. Nevertheless, such habitats are today under constant threat of modification and this may result in species extinctions in the near future. This is particularly apparent in Malaysia where primary forest has given way to agriculture and other development in the recent past. Some *Cryptocoryne* species are already reported as endangered. For example, Jacobsen (1987) has listed *C. elliptica* as one of the 20 most endangered plant species in Malaysia.

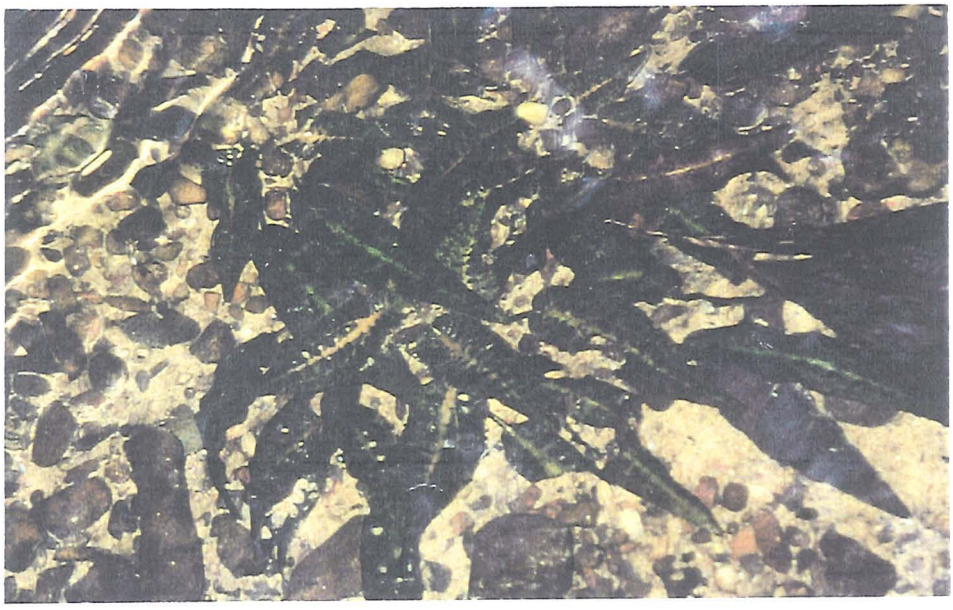
Until now, most work on the genus *Cryptocoryne* has been restricted to morphological description, investigation of chromosome numbers and the geographical distribution of species. Additional information on ecology, population structure and breeding system of species is lacking.

1.1 History, Evolution and Biogeography of the genus

The genus *Cryptocoryne* was established by Fischer in 1829 based on a species in India, *C. spiralis* (Jacobsen, 1979). Previously *C. spiralis* was placed into different genera and described as *Arum spiralis* by Rentzin (Jacobsen, 1979). More than 60 species of *Cryptocoryne* have since been

Plate 1.1.
A river in Kelantan, Malay Peninsula - an example of the type of habitat in which *Cryptocoryne* occurs.

Plate 1.2. *Cryptocoryne affinis* in its natural habitat - a closeup from Plate 1.1



described. Many of these are endemics, a situation which is unusual for aquatic plants, which normally have the potential to be widely distributed (Jacobsen, 1977). In the Malay Peninsula, ten *Cryptocoryne* species are recognised (Jacobsen, 1987; Mansor and Masnadi, 1994). These and almost all other species of *Cryptocoryne* in Malaysia are considered as endemics. They occur in restricted areas and are viewed as endangered species (Mansor and Masnadi, 1994).

The genus *Cryptocoryne* is distributed over an area that includes continental south Asia, from west India to south China, the Malay Peninsula as well as the island of Sri Lanka, Sumatra, Borneo, Java, Sulawesi (Celebes), the Philippines and Irian Jaya/Papua New Guinea. This area has been divided into four sectors based on differences in paleogeographical origin (Dunn and Dunn, 1977). The first sector is the Mainland sector which is made up of continental Asia including the island of Sri Lanka but excluding the Malay Peninsula (see Fig. 1.1). According to Dunn and Dunn (1977) this area has always been a continent and underwent only minor changes in its geographical structure during the Quaternary period (epoch Pleistocene). The second sector is the Sunda sector which comprises the Malay Peninsula and the islands of Sumatra, Borneo, Palawan, Bali, Java and other smaller islands surrounding these major islands. During the Pleistocene, this sector was influenced by extreme changes in sea level. During periods of low sea level, the sector consisted of one large land mass or Sundaland, whereas, at times of high sea level, it was fragmented into different islands as is the case today (Johns, 1995). A third sector, known as the Island sector contains the Philippines, Sulawesi and the Moluccas. These islands are thought to have

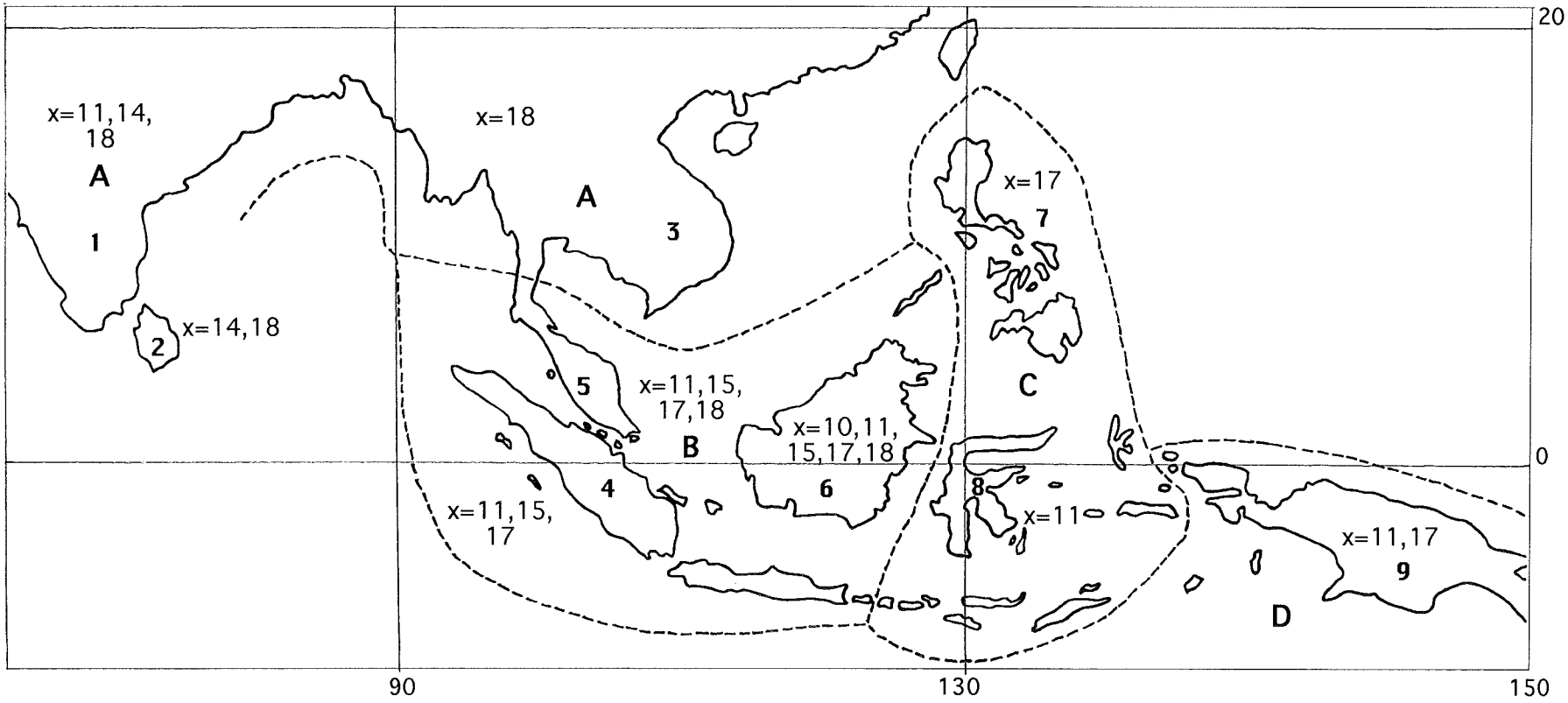


Figure 1.1. Map showing the distribution of different chromosome base numbers within *Cryptocoryne* and sectoral subdivision of south east Asia; A: Mainland sector; B: Sunda sector; C. Island sector and D: Sahul sector. 1: India; 2: Sri Lanka; 3: Indo-China; 4: Sumatra; 5: Malay Peninsula; 6: Borneo; 7: Philippine; 8: Sulawesi (Celebes) and 9: New Guinea.

remained separate islands during the Pleistocene, unconnected to each other or to the continental shelf. This belief stems from the fact that the depth of the straits between these islands is much deeper than those occurring between islands within the Sunda sector. The fourth and final sector of the region is the Sahul sector which comprises New Guinea, the Aru islands and Australia. This area was also influenced by changes in sea level during the Pleistocene. During periods of low sea level, land bridges connected the islands currently found in this area (Johns, 1995).

Cryptocoryne ciliata is the most widely distributed species in the genus. Its distribution extends from India through the Malay archipelago to New Guinea (Jacobsen, 1980b). It is usually found in the brackish water of tidal parts of rivers and coastal areas, and seems therefore, to have a tolerance to saline conditions. The species produces viviparous seeds which begin to develop into plants as soon as they have dropped into a suitable patch of mud below or away from the mother plant (Jacobsen, 1980b). The spread of the species is also aided by its ability to produce runners. These can become detached from the mother plant, after which they may be carried out to the open sea by water currents and wind and then transported to distant islands (Reumer, 1984).

The chromosome numbers of almost all recognised species within the genus are known and include the following: 2n = 20, 22, 28, 30, 34, 36, 42, 54, 60, 66, 68, 72, 85, 88, 90, 102, and 132 (Jacobsen, 1977, 1985; Arends *et al.*, 1982; Petersen, 1989). These chromosome numbers are considered to be related to six different base numbers viz. x = 10, 11, 14, 15, 17, and 18

(Jacobsen, 1977; Arends *et al.*, 1982). Triploid, tetraploid and hexaploid cytotypes have been recorded in several species, for example, in *C. spiralis*. The geographical distribution of the different chromosome base numbers in *Cryptocoryne* is illustrated in Figure 1.1.

Jacobsen (1977) has suggested that the predominant trend of chromosome evolution in *Cryptocoryne* is represented by the formation of an aneuploid series. Arends *et al.* (1982) have proposed that the base number of the genus is x=18 and this has been reduced by aneuploidy to x=10. The proposal that x=18 is the most primitive chromosome condition in *Cryptocoryne* is supported by the fact that all species of the genus *Lagenandra*, which is closely related to *Cryptocoryne* (*Lagenandra* and *Cryptocoryne* constitute the subtribe Cryptocoryninae) have chromosome numbers of 2n=36 (Reumer, 1984).

Reumer (1984) has argued, however, that the hypothesis of Arends *et al.* (1982) is not firmly supported by biogeographical data. Instead, he has proposed a phylogeny of the subtribe Cryptocoryninae based on both cytological and biogeographical data (Figure 1.2). The resulting phylogeny is thought to have a hypothetical primary base number of x_1=9, from which the following base chromosome numbers are derived:

a) x_1=11. This base number is proposed to have developed as a result of aneuploidal increase (9+2). The base number of x=10 may or may not have been an intermediate step, although the x=10 group from Borneo is considered by Reumer (1984) to have descended from x=11; b) x_2=18, derived from x=9 by means of euploid doubling. From x=18 it is thought that x=14, x=17 and x=15 have been derived by aneuploidal reduction.

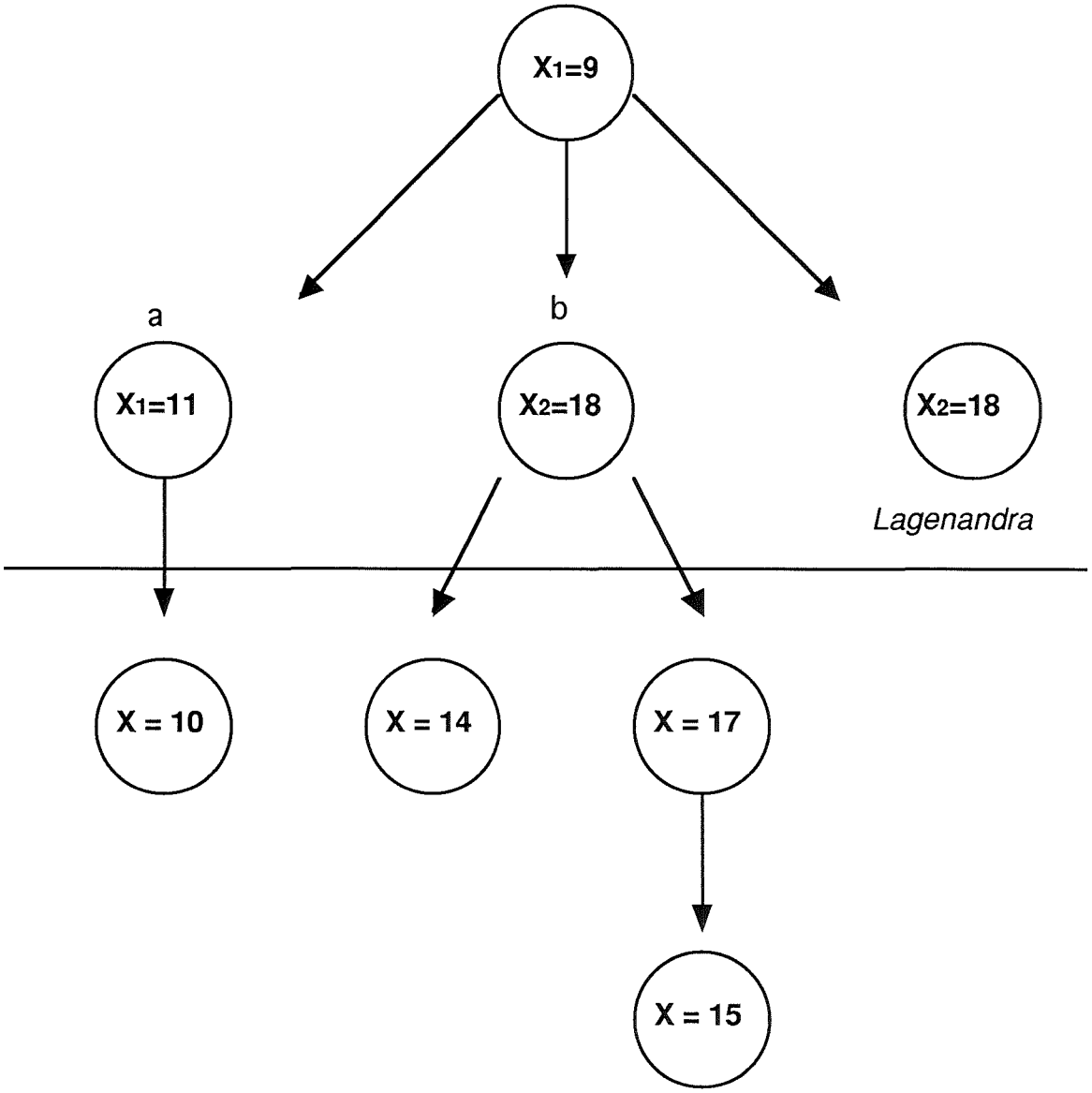


Figure 1.2. A phylogeny of subtribe Cryptocoryninae showing the relationships between the different chromosome base numbers. Above the line shows the evolutionary steps that have occured in the Indian subcontinent while below are the evolutionary steps that have taken place in other areas where the genus is found (after Reumer, 1984).

The presence of two separate pathways in the phylogeny (i.e. a and b, Fig. 1.2) would suggest that the genus *Cryptocoryne* is of polyphyletic origin.

1.2 Morphological Characters

The leaves of *Cryptocoryne* are differentiated into a petiole and a blade and occur in rosettes that develop at intervals along rhizomes. Leaf blades vary from oval to ribbon shaped, and often have markings or a blistered texture and undulating edges. Leaf shape, size and sometimes colour are very variable characters and their development depends to a large extent on local conditions (Rataj, 1975). Species growing in fast running water, have their leaves accomodated to the water stream (Rataj, 1975). Some species, for example, *C. minima* has green to purple ventral leaves, while the dorsal leaves are pale green with purple marking and purple veins (Jacobsen and Mansor, unpublish). The colouring of leaves, depends on the amount of light received; for example, plants growing in brightly lit areas tend to produce green leaves (Rataj, 1975).

The inflorescence of *Cryptocoryne* species is totally submerged except for the aerial opening of the spathe (Figure 1.3). The spathe is connate along its margin forming a watertight tube with an aerial opening at the top (the throat). A kettle at the base contains the male and female organs. The throat is sometimes bordered by a raised rim or collar. The limb or free margin is present at the upper most part of the spathe, which usually varies in shape and colour with a pointed tip or a long caudical in some species. In *C. ciliata,* for example, the edges of the limb are distinctly fringed. The colour and texture of the spathe, and presence and colour of the collar are important in species identification (Jacobsen, 1980a).

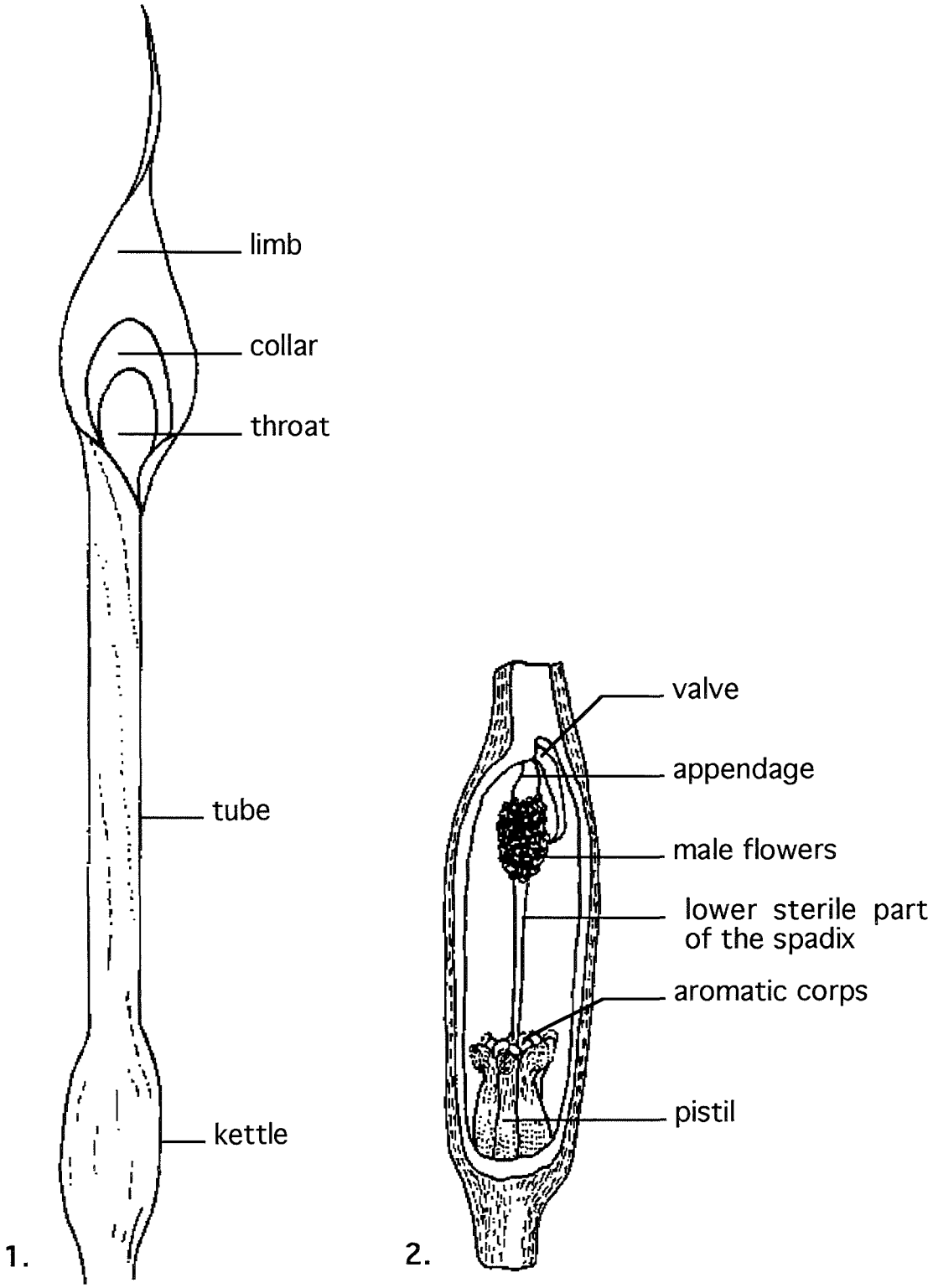


Figure 1.3. Floral morphology of *Cryptocoryne;* 1. The spathe, 2. The kettle with the inflorescence.

The spadix as is indicated by the name of the genus, *Cryptocoryne,* is totally hidden inside the kettle. A single whorl of female flowers, usually with 5 to 6 carpels, lies at the base of the spadix. In the middle of this female flowers, there is present a single whorl of abortive, modified female flowers known as the aromatic corps. Above this there is a long thin sterile zone which is topped by a cluster of male flowers. At the apex of the spadix there is a dome-shaped appendage present while a valve-like flap is attached to the kettle wall. The valve controls the opening of the passage into the kettle (Petch, 1928).

1.3 Propagation

Within *Cryptocoryne,* vegetative propagation occurs through the production of runners and dormant buds (Rataj, 1975). Sexual propagation occurs more frequently in species occupying terrestrial habitats or with an amphibious life form. Under conditions of total submergence or rapid water flow, amphibious species propagate vegetatively, but at low water levels or when the water current is slow, they flower and form fruits. Rataj (1975) stressed that *Cryptocoryne* is self-sterile. Fruits failed to be produced when two plants of the same vegetative origin were crossed; however, crosses between individuals from different vegetative origin formed fruits.

The genus *Cryptocoryne* is entomophilous, that is the flower is pollinated by insects. Saibeh and Mansor (1996) listed two species from the beetle family, Nitidulidae (*Haptoncus luteolus* and *H. minitus*), in addition to the fruit fly, *Drosophila melanogaster,* as the most probable pollinators of *C. ciliata.* This was based on the observation that these insect

species were frequently found in the plant's kettle. However, several other insects from the family Acrididae and Syrphidae were also found trapped occasionally inside the kettle, indicating that many different insects might effect pollination of the species (Saibeh and Mansor, 1996).

As with other aroids, both male and female organs of *Cryptocoryne* mature at slightly different times. The bionomy of flowering can be divided into three stages (Rataj, 1975). At the first stage the appendage of the spadix lies against the valve forcing the valve to touch the inner kettle wall. This blocks the passage from the tube into the kettle chamber. At this stage, the spathe is still closed and both the male and female organs remain immature. The second stage begins with the opening of the limb followed by submaturation of the female organs, and ovaries. The appendage then separates from the valve which opens an inwards passage allowing pollinators to enter the kettle. At this stage the pollinators which were inside the kettle cannot escape. The final stage begins when the stamen matures and secretes a sticky pollen liquid. This usually happens three or four days after the opening of the inflorescence. The pollinators, in their attempt to escape, might have sticky pollens attached onto parts of their bodies while climbing onto the male organs. If pollination is accomplished, the tension within the kettle wall and valve collapses. However, if pollination does not take place, the kettle will collapse naturally five to seven days after the opening of the spathe (Rataj, 1975), allowing pollinators to escape to pollinate other flowers.

1.4 Uses of *Cryptocoryne*

Cryptocoryne species are widely used as aquarium plants (Mansor and

Masnadi, 1994). They are collected and cultivated on a large scale as foliage plants for tropical aquaria. The main features that appeal to aquarists are the plant's small size and variation of leaf shape, texture and colour. In order to meet the demand of the aquarium industry, some species from the south of the Malay Peninsula and also from Sri Lanka have been collected in massive quantities and this has contributed to a considerable decline in the number of these species in their natural habitat (Mansor and Masnadi, 1994; de Graaf and Arends, 1986).

In Sri Lanka, many *Cryptocoryne* species are thought to have some medicinal properties. It has been reported that *Cryptocoryne* have been used by local physicians to treat patients with stomach ailments and helminthiasis (de Graaf and Arends, 1986).

1.5 Systematics of *Cryptocoryne*

After the establishment of the genus by Fisher in 1829, several more species of *Cryptocoryne* were described by Schott in the 1850s (Jacobsen, 1979) and by 1860 the number of species placed in the genus had increased to 12. In 1920, Engler described 28 species and divided the genus *Cryptocoryne* into the following three sections based on the morphology of the spathe (Jacobsen, 1979); Unitubulosae (spathe with a short tube), Bitubulosae (spathe with a long tube) and Ciliatae (spathe with cilia).

Based on morphological comparisons, the genus was later revised by Rataj (1975) who divided it into four subgenera and 16 sections (Table 1.1). However, Jacobsen (1977) noted that Rataj (1975) provided errornous description on several species and did not always take into account the

chromosome number and geographical distribution of certain species when placing them either into different sections, or into the same section, and this created more errors in his revision. For example, in subgenus *Submersina*, species with 2n=28 were placed by Rataj (1975) into two different sections, i.e. sect. Nevillae and sect. Walkeriae. This was because Rataj (1975) considered that species in these two sections had evolved independently. However, Jacobsen (1977) argued that this separation was not supported by the equivalent chromosome numbers of species in these different sections, nor by morphological comparison. Furthermore, according to Rataj (1975), section Thwaitesiae (Subgenus Submersina) contained a Sri Lankan species, *C. thwaitesii* with 2n=36, although other members placed into the same section with 2n=28, were from Borneo, the Malaysia Peninsula and Sumatra. Finally, according to Rataj (1975), section Auriculata (Subg. Submersina) contained a Sri Lankan species, *C. bogneri* with 2n=36 plus two species from the Philippines, i.e. *C. pygmaea* (2n=34) and *C. auriculata* (2n=34). Jacobsen (1977) argued against placing different endemic species in the same section if they have different chromosome numbers and widely different geographical distributions.

Table 1.1. Taxonomic division in *Cryptocoryne* species as recognised by Rataj (1975).

Taxa	Distribution
Subgenus Suturina	
Section 1. Spiralae	
- *C. spiralis*	India
- *C. burmensis*	Myanmar
Section 2. Gomeziae	
- *C. gomezi*	Bangladesh
- *C. ferruginea*	Sarawak, Borneo
Subgenus Myrioblastus	
- *C. ciliata* var. *ciliata*	Throughout Asia
- *C. ciliata* var. *latifolia*	Throughout Asia
Subgenus Terrestrina	

Section 3. Consobrinae	
- *C. cognata*	India
- *C. cruddasiana*	Myanmar
- *C. consobrina*	India
Section 4. Retrospiralae	
- *C. retrospiralis*	India
- *C. crispatula*	India
- *C. bertelihansenii*	Thailand
- *C. tonkinensis*	Vietnam
- *C. balansae*	Thailand
- *C. usteriana*	Philippines
Section 5. Costatae	
- *C. costata*	Thailand
- *C. korthausae*	Origin unknown
Subgenus Submersina	
Section 6. Thwaitesiae	
- *C. caudata*	Borneo
- *C. thwaitesii*	Sri Lanka
- *C. johorensis*	Malay Peninsula
- *C. pontederiifolia*	Sumatra
Section 7. Linguae	
- *C. lingua*	Borneo
- *C. versteegii*	New Guinea
Section 8. Auriculatae	
- *C. auriculata*	Philippines
- *C. pygmaea*	Philippines
Section 9. Nevillae	
- *C. nevillii*	Sri Lanka
- *C. parva*	Sri Lanka
- *C. lucens*	Sri Lanka
Section 10. Walkeriae	
- *C. axelrodii*	Sri Lanka
- *C. wendtii*	Sri Lanka
- *C. walkeri*	Sri Lanka
- *C. beckettii*	Sri Lanka
- *C. petchii*	Sri Lanka
Section 11. Ellipticae	
- *C. elliptica*	Malay Peninsula
- *C. gracilis*	Borneo
Section 12. Striolatae	
- *C. affinis*	Malay Peninsula
- *C. striolata*	Borneo
Section 13. Pallidinerviae	
- *C. pallidinervia*	Borneo
Section 14. Griffithiae	
- *C. zewaldiae*	Malay Peninsula
- *C. minima*	Malay Peninsula
- *C. griffithii*	Malay Peninsula
- *C. schulzei*	Malay Peninsula
- *C. nurii*	Malay Peninsula
- *C. bullosa*	Borneo

Section 15. Cordatae	
- *C. evae*	Thailand and Myanmar
- *C. siamensis*	Thailand
- *C. zonata*	Borneo
- *C. didericii*	Malay Peninsula
- *C. stonei*	Malay Peninsula
- *C. cordata*	Malay Peninsula
Section 16. Zukalae	
- *C. hejnyi*	Malay Peninsula
- *C. zukalii*	Malay Peninsula

The most recent treatment of the genus *Cryptocoryne* was reported by Jacobsen (1977) and Arends *et al.* (1982). In these treatments (Table 1.2), the genus was revised on the basis of geographical distributions, morphology and chromosome numbers. Jacobsen (1977) listed 34 *Cryptocoryne* species divided into eight groups. Arends *et al.* (1982) expanded the genus to include 48 species divided into 24 groups.

Table 1.2. List of *Cryptocoryne* groupings, mitotic numbers and geographical distribution (compiled from Jacobsen, 1977 and Arends *et al.*, 1982).

Taxa	Chromosome number	Distribution
Group 1. *C. striolata* group		
- *C. striolata*	20	Borneo
- *C. keei*	20	Borneo
Group 2. *C. ciliata*		
- *C. ciliata*	22, 33	Asia
Group 3. *C. spiralis*		
- *C. spiralis*	33, 66, 88, 132	India
Group 4. *C. becketti* group		
- *C. beckettii*	28, 42	Sri Lanka
- *C. walkeri*	28, 42	Sri Lanka
- *C. wendtii*	28, 42	Sri Lanka
- *C. undulata*	28, 42	Sri Lanka
- *C. parva*	28	Sri Lanka
- *C. nevillii*	28	Sri Lanka
- *C. X willisii*	28	Sri Lanka
Group 5. *C. pontederiifolia* group		
- *C. pontederiifolia*	30	Sumatra
- *C. moehlmannii*	30	Sumatra
Group 6. *C. villosa*		
- *C. villosa*	30	Sumatra
Group 7. *C. longicauda*		
- *C. longicauda*	30	Borneo, Malay Peninsula

Group 8. *C. scurrilis* group		
- *C. gasseri*	34	Sumatra
- *C. amicorum*	34	Sumatra
- *C. scurrilis*	34	Sumatra
Group 9. *C. griffithii* group		
- *C. minima*	34	Malay Peninsula
- *C. zewaldiae*	34	Malay Peninsula
- *C. griffithii*	34	Malay Peninsula
Group 10. *C. purpurea* group		
- *C. purpurea*	34	Malay Peninsula
- *C. zukalii*	34	Malay Peninsula
- *C. jacobsenii**	34	Malay Peninsula
- *C. didericii**	34	Malay Peninsula
Group 11. *C. cordata* group		
- *C. cordata*	34, 68, 85, 102	Malay Peninsula
- *C. zonata*	68	Borneo
- *C. grabowskii*	68	Borneo
- *C. edithiae*	68	Borneo
Group 12. *C. schulzei*		
- *C. schulzei*	34	Malay Peninsula
Group 13. *C. nurii*		
- *C. nurii*	34	Malay Peninsula
Group 14. *C. affinis*		
- *C. affinis*	34	Malay Peninsula
Group 15. *C. ferruginea* group		
- *C. ferruginea*	34	Borneo
- *C. tortilis*	34	Borneo
- *C. fusca*	34	Borneo
Group 16. *C. auriculata*		
- *C. auriculata*	34	Borneo
Group 17. *C. bullosa*		
- *C. bullosa*	34	Borneo
Group 18. *C. pallidinervia*		
- *C. pallidinervia*	34	Borneo
Group 19. *C. usteriana*		
- *C. usteriana*	34	Philippines
Group 20. *C. pygmaea*		
- *C. pygmaea*	34	Philippines
Group 21. *C. versteegii*		
- *C. versteegii*	34	New Guinea
Group 22. *C. lingua*		
- *C. lingua*	36	Borneo
Group 23. *C. thwaitesii* group		
- *C. thwaitesii*	36	Sri Lanka
- *C. alba*	36	Sri Lanka
- *C. bogneri*	36	Sri Lanka
Group 24. *C. albida* group		
- *C. albida*	36	Thailand
- *C. crispatula*	36, 54	IndoChina
- *C. retrospiralis*	36, 72	India

* These species may need to be transferred to another group as they are reported to occur in Sumatra and not the Malay Peninsula (Jacobsen and Mansor, unpublished).

The revision of the genus by Jacobsen (1977) and Arends *et al.* (1982) clearly indicated that there is present within *Cryptocoryne* a pattern in regard to morphological attributes, chromosome numbers, and geographical distribution. However, the relationships between species groups with different chromosome numbers, either from different or the same geographical area were unclear. For example, species present in the Malay Peninsula are separated into seven taxonomic groups, but the association between these different groups were not clearly defined. What is also unclear is the nature of association between the Malay Peninsula species group and other species groups from India and Mainland Asia as well as Sri Lanka which have different chromosome base numbers. It was decided, therefore, to conduct an analysis of molecular variation on representatives from these different groups to examine more critically their taxonomy groupings and interrelationships as proposed by Jacobsen (1977) and Arends *et al.* (1982).

1.6 Aims and objectives of the research

The main aim of the research reported in this thesis was to construct a molecular phylogeny of the genus so as to gain a better understanding of its systematics and evolution. Phylogenies based on nuclear and chloroplast DNA sequences were constructed and compared. In addition, a survey of variation of randomly amplified polymorphic DNA (RAPD) was conducted to aid the resolution of the phylogeny of closely related taxa. A final objective was to identify the diploid parents of certain hybrid taxa in the genus from a comparison of nuclear and chloroplast DNA sequences and also RAPD variation.

Chapter 2

Materials and methods

Material and Methods

2.1 Origin of plant material

Leaf material used for DNA extraction was collected either directly from the wild or was made available from cultivated sources. Some material from the Malay Peninsula was collected from the wild in 1994 and 1995, but the bulk of material was obtained from cultivated specimens maintained by Professor Niels Jacobsen of the Royal Veterinary and Agricultural University, Copenhagen. During the period of research, additional cultivated material was obtained from Jan Bastmeijer (Wageningen, The Netherlands) and Joseph Bogner (Munich Botanical Garden, Germany) and also from Kew Gardens. Material collected from the wild and obtained from Professor Jacobsen was cultivated in a green house at St. Andrews. In total, leaf material for analysis was made available from 25 species of *Cryptocoryne* and two species of *Lagenandra* (Table 2.1). The latter two species were used to root the phylogenetic tree with respect to *Cryptocoryne*.

Table 2.1. List of *Cryptocoryne* and *Lagenandra* species subjected to analysis.

Taxon	2n	Accession	Distribution area
C. ciliata (Roxb.) Schott	22	P1958/6013	*tropical Asia
C. spiralis (Retz.) Fischer ex. Wydler	66	NJ3129a2	India
C. beckettii Trimen	28	1671	Sri Lanka
C. walkeri Schott	28	NJ23-3	Sri Lanka
C. wendtii de Wit	28	P1961/5342	*Sri Lanka
C. x willisii Reitz	28	P1978/5045	Sri Lanka
C. x willisii Reitz	28	NJ23-25	Sri Lanka
C. x willisii Reitz	28	Kew3790	*Sri Lanka
C. undulata Wendt	28	NJ22-7	Sri Lanka
C. pontederiifolia Schott	30	USM9438	Sumatra
C. moehlmannii de Wit	30	P1989/5046	Sumatra
C. longicauda Engl.	30	USM9439	Malay Peninsula
C. minima Ridley	34	S1995/9201	Malay Peninsula

C. griffithii Schott	34	NJ85-30	Malay Peninsula
C. purpurea Ridley	34	Othman s.n.	Malay Peninsula
C. cordata Griffith	34	USM9139	Malay Peninsula
C. zonata de Wit	34	W534	Borneo
C. schulzei de Wit	34	USM8087	Malay Peninsula
C. affinis Hooker f.	34	USM8065	Malay Peninsula
C. pygmaea Merill	34	NJ3962	*Philippines
C. usteriana Engler	34	P1983/5448	Philippines
C. aponogetifolia Merill	34	P3401	Philippines
C. annamica Serebryani	34	M92/3205	Vietnam
C. alba de Wit	36	NJ3172-6	*Sri Lanka
C. albida Parker	36	P1958/5363	Thailand
C. crispatula var. *balansae* Engler	36	NJ3406	*Thailand
C. retrospiralis (Robx.) Kunth	36	P1977/5146	India
C. elliptica Hooker f.	34	USM8069	Malay Peninsula
Lagenandra meeboldi (Engl.) C.E.C. Fisch.	36	P1979/5019	India
L. ovata (L.) Thwaites	36	P1983/5653	*India

* the exact locations from where these taxa were found/collected were not recorded

These taxa were chosen on the basis of the availability of plants from cultivated and wild sources. Species from the four paeleogeographical areas described by Dunn and Dunn (1977) were represented and contained the range of chromosome found in the genus, except for x=10. The analysis of this group of species, therefore, allowed investigation of relationships between species from different geographical sectors and, also of different chromosome base number. A hybrid species, *C. x willisii,* and its putative parents, were also included in the analysis to assess the utility of molecular markers in identifying them from a phylogenetic reconstruction. Jacobsen (1981) showed that in hybridisation experiments, *C. x willisii* is of hybrid origin with *C. parva* as one parent and *C. walkeri* or *C. beckettii* as the other parent. Further it was also shown that both *C. walkeri* and *C. beckettii* can hybridise and backcross, and that these hybrid and backcross products can hybridise with *C. parva.* Variation in the colour of the limb and collar of the spathe, which is recorded within the hybrid, is due to

variation present in the parents.

2.2 Cultivation of plant material

Live plants were transplanted into a medium of peat moss/clay mixture (3:1) contained in pots. These pots were placed into a tank such that the water level reached half their height. Water temperature was maintained at 24°C, and the tank was covered with transparent plastic to maintain high humidity. A 12 hour photoperiod was provided using 400W mercury-vapour lamps (Osram). Water level was monitored weekly and two capsules of fertiliser (Annimix - a rhododendron organic fertiliser) were given to each plant once every three months. Under these conditions plants grew well and in species such as *C. minima, C. beckettii, C. walkeri, C. x willisii, C. usteriana, C. crispatula* var. *balansae* and *C. albida,* flowering was observed.

2.3 DNA extraction

Cryptocoryne leaves were dried with silica gel (Chase and Hills, 1991) and kept at -20°C prior to DNA extraction. Two extraction methods were employed, both of which were successful. The first used a 2% CTAB extraction buffer as described by Doyle and Doyle (1987), whilst the second method was that described by Whittemore and Schaal (1991), which is a modification of the former extraction procedure. Here, only the second extraction procedure will be described as it is the protocol that was used for all of the research described.

Leaf material that had been dried with silica gel was flash frozen with

liquid nitrogen and ground with a pestle and mortar to a fine powder. 10 ml of extraction buffer (2% Hexadecyltrimethylammonium bromide [CTAB]; 20 mM EDTA-Na_2; 1.4 M NaCl; 100 mM Tris-HCl pH8.0; 5% PVP-10; 0.2% mercaptoethanol) preheated to 65°C, were then added and mixed to produce a smooth green paste. The homogenate was transferred to a 50 ml oak ridge tube to which 5 ml of extraction buffer used to rinse the mortar and pestle were added. The mixture was incubated at 65°C for 30 mins and then allowed to cool for 5 mins before adding 10 ml of dichloromethane. The solution was mixed thoroughly to a single phase by inverting the tube. Balanced tubes were spun at 5,000 rpm in a Sorvall^R Superspeed centrifuge for 10 mins before transferring the uppermost aqueous layer to a clean oak ridge tube. Two thirds volume of ice-cold propan-2-ol was added to the aqueous layer and then mixed by inverting the tube to a single phase solution to precipitate DNA. Samples were left to stand for 1-2 hours at room temperature before the pellet was spun down at 5,000 rpm for 5 mins and resuspended in wash buffer (76% ethanol; 10 mM ammonium acetate) for 30 mins. The pellet was spun for a second time, the supernatant discarded, and the pellet was left to dry at room temperature for another half an hour before dissolving in 2 ml TE (10 mM Tris-HCl pH7.6; 1 mM EDTA-Na_2) at 4°C overnight.

2.4 DNA purification

Following extraction, DNA was purified using caesium chloride (CsCl) density gradients. This was done to obtain a pure extract of DNA which could be stored without deterioration for a long period. Purification of *Cryptocoryne* DNA using a CsCl density gradient was carried out as follows.

CsCl, Ethidium bromide (EtBr) and extracted DNA (2 ml obtained from the extraction procedure) were made up with TE solution to a total volume of 11 ml, with a final concentration of 0.75 g/ml of CsCl and 200 ug/ml EtBr. The mixture was put into a 12 ml Sorvall^{R} ultracentrifuge tube in a fixed angle Sorvall T865.1 rotor and centrifuged at 50,000 rpm in a Sorvall ultracentrifuge at 20^{o}C for 18-24 hours. The DNA band produced was removed from the CsCl gradient using a wide bore hypodermic needle and treated with TE-saturated butanol until all traces of EtBr had been removed. An equal volume of TE was added to reduce the CsCl concentration, and two volumes of propanol at room temperature were added to precipitate the DNA overnight again at room temperature. The sample was then spun at 8,000 rpm for 30 mins to pellet the DNA which was then air dried and resuspended in 1 ml TE and left to stand for 1 hour. Samples were divided into two 500 µl aliquots in 1.5 ml microfuge tubes and 1 ml of propan-2-ol was added to precipitate the DNA. Samples were spun down at 13,000 rpm in a microfuge to produce a DNA pellet which was then air dried. 100 µl of TE was added and the two aliquots of the DNA samples were combined to make a total volume of 200 µl. DNA concentration was determined by running samples through an 0.8% agarose gel together with a DNA standard of known concentration for comparison.

2.5 PCR amplification of a DNA fragment

Polymerase chain reaction (PCR) is a technique which allows the amplification of a specific DNA region that is flanked by known DNA sequences (Fig. 2.1). Amplification of a DNA fragment was achieved by using a thermostable polymerase and short, single-stranded

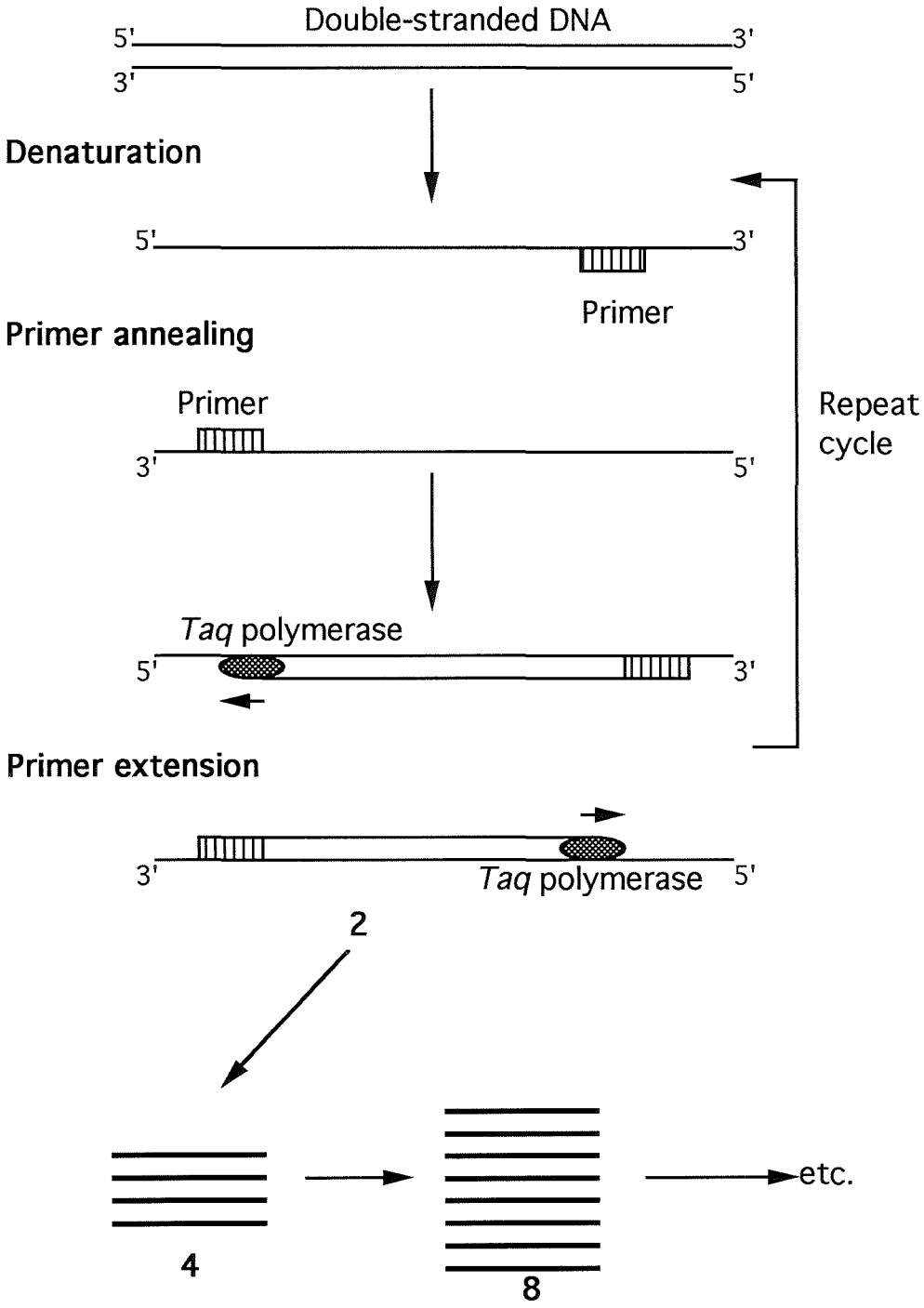


Figure 2.1. The polymerase chain reaction (PCR). PCR is a cycling process. Double-stranded DNA is denatured near the boiling temperature of water to give single strands. It is then cooled to allow primers to anneal specifically to the target region. DNA polymerase (thermostable) is used to synthesise complementary strands from the template strands by primer extension. The cycle is repeated many times. The numbers refer to the number of double-stranded DNA in the reaction; start off with 1, to produce 2 strands at the end of the first cycle, 4 at the end of the second , 8 at the end of the third cycle, and so on.

oligonucleotide primers which were complementary to the ends of a defined sequence of the DNA template. PCR involves three steps of a cycle; i) denaturation of DNA template at near boiling temperature of water; ii) annealing of oligonucleotide primers to the denatured template at a lower temperature; and iii) primer extension at 72°C using a thermostable DNA polymerase. These steps were repeated over several cycles (25-45 cycles) in order to amplify fragments.

An example of PCR used in this research was the amplification of the internal transcribed spacer region of the nucler ribosomal DNA (see chapter 3). This was carried out as follows. The PCR reaction with a total volume of 100 µl consisted of 10 µl of reaction buffer, 6 µl to 8 µl of 25 mM magnesium chloride, 10 µl of a 2 mM of dNTP (i.e. dATP, dCTP, dGTP and dTTP) solution mixture in equimolar ratio, 50 pmol of each primer, 25 to 100 ng of template DNA and 2.5 Units of *Taq* DNA polymerase form Promega. The reaction mixture was overlaid with two drops of mineral oil to prevent evaporation. Double-stranded PCR products were produced through 25-30 cycles of denaturation (97°C for 1 min), primer annealing (48°C for 1 min) and extension (72°C for 3 mins). An initial denaturation step (97°C for 1 min) preceded the cycles and a 7 mins final extension at 72°C followed the cycles. Primers used in this PCR reaction were ITS4 and ITS5 (see Fig. 3.1 in chapter 3).

In the case of the species considered to be of hybrid origin, *C. x willisii*, the ITS region from three accessions of this hybrid were amplified. From this pooled amplified ITS, different fragments were cloned (see section 2.7 in this chapter) in order to separate different repeat-types of the ITS region that may be present in the genome of the three three hybrid accessions (see

section 3.4.3) in order to separate different ITS repeat-types that may be present in the genome of the three hybrid accessions. From *C. x willisii* NJ23-25, cloned ITS from four different colonies were analysed, while for *C. x willisii* Kew3790 and *C. x willisii* P1978/5045 only one colony of cloned ITS each was analysed.

2.6 Cleaning the amplified product

Amplified DNA product has to be cleaned from excess unincorporated primers and nucleotides before being subjected to cycle sequencing. The PCR product was purified using the Wizard™ PCR Preps DNA Purification Systems (Promega). It comes with Direct Purification Buffer, Magic DNA Purification Resin and Minicolumns. Cleaning of amplified DNA product was carried out using the protocol described in Promega's Wizard™ PCR Preps DNA Purification System manual.

The product of each completed PCR reaction, excluding the mineral oil, was transferred to a clean 1.5 ml microcentrifuge tube. 100 µl of Direct Purification Buffer was added to the tube and the components were mixed by vortexing. 1 ml of Resin was then added and mixed by vortexing three times over a 1 min period.

A Wizard minicolumn was set up for each PCR product with a 2 ml syringe barrel attached on top and a lidless 1.5 ml microfuge tube placed at the bottom. The DNA/buffer/resin mixture was pipetted into the syringe barrel and was gently pushed through the minicolumn using a syringe plunger. Later the syringe was detached from the minicolumn and the plunger removed. The syringe barrel was then reattached to the column and 2 ml of 80% isopropanol were pipetted into the syringe and pushed

through the column. Once again the syringe barrel was removed and a new 1.5 ml microfuge tube was attached to the bottom of the minicolumn. The minicolumn/microfuge tube was then spun down at 13,000 rpm for 20 seconds to dry the column. The 1.5 ml microfuge tube was then discarded.

Another 1.5 ml microfuge tube was attached to the minicolumn and 50 µl of sterile distilled water was applied to the minicolumn to elute the DNA. The minicolumn was left to stand for at least 1 min before being centrifuged at 13,000 rpm for 20 seconds so as to elute the bound DNA fragment. 5 µl of purified and concentrated DNA were run through a 0.8% agarose gel to check its concentration, and the remaining DNA was stored at -20°C.

2.7 Cloning of a PCR product

DNA sequences of a gene family of nuclear origin might be present in multiple repeat types of the same length (Baldwin *et al.*, 1995). Using direct sequencing approach, this intragenomic repeat type variant can be tentatively diagnosed by two or more nucleotide states observed at a site, suggesting a superimposition of two or more sequence patterns (Baldwin *et al.*, 1995). In order to obtain a single repeat type sequence from pooled amplified fragments that might consist of multiple repeat types, cloning of fragments is necessary

The cloning of a gene or DNA fragment involves the use of a suitable carrier molecule or vector and a suitable living system or host cell in which the vector can be propagated. Once recombinant DNA molecules

have been constructed (i.e. the DNA fragment of interest has beenincorporated or ligated into a vector) and propagated in a host cell, they can be identified using a colour screening method achieved by using chromogenic substrates such as the X-Gal (5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside) screening system (Nicholl, 1994).

The vector used in this research was the commercially available pGEM^{R}-T vector system (Promega). This vector has an ampicillin resistant gene and multiple cloning sites within the coding region of the enzyme beta-galactosidase (*lac*Z). Successful insertion of a DNA fragment within these multiple cloning sites can be detected by colour screening using indicator agar plates with the X-Gal system. X-Gal is a colourless substrate for enzyme beta-galactosidase, usually synthesised by the bacterium *Escherichia coli* when lactose is available or when a lactose analogue such as IPTG (iso-propyl-thiogalactoside) is used (Nicholl, 1994). The enzyme beta-galactosidase will cleave X-Gal to give rise to galactose and an indoxyl derivative. This indoxyl derivative will oxidise in air to generate the dibromo-dichloro derivative which is blue. If a fragment has been successfully inserted into a vector within its functional beta gene, the resulting colony present on the screening plate will remain colourless as the beta reaction will not occur. However, a vector without recombinant DNA retains a functional beta gene and therefore gives rise to blue colonies. Once the desired colony was identified (i.e. a colourless or white colony), the recombinant sequence from this colony was directly amplified (i.e. to provide enough material for sequencing analysis).

2.7.1 Preparation of indicator agar plates and media broth

Agar plates for screening positive recombinant DNA colonies as

described in the previous section, were prepared as follows. Agar was prepared by adding 28 g of ready made nutrient agar (Oxoid) to 1l of distilled water and heating until dissolved. The agar was then autoclaved at 121°C, for 15 mins. After autoclaving, agar was cooled to 50°C in a water bath before adding the *lac*Z staining reagents and antibiotic. 1 ml ampicillin (100 mg/ml) was added together with 1 ml of 40 mg/ml dimethylformamide X-Gal and 10 ml of 100mM IPTG. 20 ml of agar media containing ampicillin and *lac*Z staining mixture was poured into 9 cm petri dishes. The lid of the dish was left open until the agar became hard. Plates were then left to dry for 15 mins at 50°C in an oven, upside down with the lid off before being stored at 4°C until required.

2.7.2 Making competent cells

The host cell that was used in the cloning procedure was *E. coli* DH5-alpha. Before the host cells were used for propagation of vectors, they needed to be made competent and this was achieved using a protocol described by Sambrook *et al.* (1989). Firstly, overnight culture was prepared, that is an *E. coli* was grown overnight in 10 ml nutrient broth at 37°C. 100 µl of overnight culture was then put into 10 ml fresh nutrient broth and incubated at 37°C, shaking at 300 rpm for 2.5 hours or until the density reading was 0.5-0.6 at 600 nm absorbance using a spectrophotometer (PYE UNICAM SP6-550). Cells at this stage were in the log phase of growth. The nutrient broth (10 ml) containing the cells were pipetted into several precooled 1.5 microfuge tubes and kept on ice for 15 minutes before centrifuging at 13,000 rpm for 30 seconds at room temperature. The supernatant was discarded and the pellet was resuspended in 500 µl 50 mM $CaCl_2$ and left to stand on ice for 30 mins. The solution was centrifuged

again for 30 seconds before discarding the supernatant and resuspending the pellet in 100 µl 50 mM $CaCl_2$. The cells were now competent for use and could be left on ice for up to 12 hours or mixed with 50% glycerol and stored at -70°C until required.

2.7.3 Ligation of a PCR fragment into a vector and transformation of competent cells

Ligation of a DNA fragment into a vector was performed using the protocol described in Promega's pGEM^R-T and pGEM^R-T Easy Vector Systems technical manual. A ligation reaction was conducted in an 0.5 ml microfuge tube by adding 1 µl T4 DNA ligase 10X buffer, 1 µl (50 ng) pGEM^R-T Easy Vector, 35-105 ng PCR product, 1 µl T4 DNA ligase (3 units/µl) and sterile distilled water to a final volume of 10 µl. These components were mixed by pipetting and the solution was incubated overnight at 4°C.

On the following day, the ligation mixture was centrifuged for 10 seconds at 13,000 rpm to collect contents at the bottom of the tube. 2 µl of the centrifuged ligation mixture was transferred to a 1.5 ml microfuge tube and placed on ice. 50 µl of competent cells were added to the tube and mixed by gently flicking the tube before placing it on ice again for 20 mins. The cells were heat shocked for 45-50 seconds in a thermal cycler at exactly 42°C and were immediately placed on ice for 2 mins.

950 µl of room temperature nutrient broth were added to the tube containing cells transformed with the ligation reaction and incubated for 1.5 hours at 37°C, at 150 rpm. 50 µl and 100 µl of each transformed culture

were spread onto two separate screening plates and were incubated for 16-24 hours at 37°C. White colonies observed after the incubation period (that might contain the recombinant DNA fragment of interest) were directly amplified using the white colony as template in a PCR reaction, so as to obtain enough DNA template for sequencing analysis.

2.8 DNA sequencing

DNA sequencing is now widely used in studies of plant systematics and evolution. The main principle behind the sequencing technique is that two single-stranded DNA molecules that differ in length by just a single nucleotide can be separated into distinct bands by electrophoresis in a polyacrylamide gel (Brown, 1994).

One method commonly used in DNA sequencing is the Sanger-Coulson chain termination procedure (Nicholl, 1994). It involves denaturation of double-stranded DNA and the attachment of a short oligonucleotide to each strand. This primer acts to initiate the synthesis of a new polynucleotide chain catalysed by a thermal stable DNA polymerase using four deoxyribonucleotide triphosphate (dNTPs - dATP, dCTP, dGTP and dTTP). However, in this reaction four modified nucleotides, dideoxynucleotides (ddNTPs) were also added. These modified nucleotides are incorporated into a growing polynucleotide chain in the same manner as normal nucleotides; however, instead of continuous extension of the chain, the intergration of a ddNTP blocks further chain elongation due to the fact that it lacks a hydroxyl group at the 3' position of its sugar component.

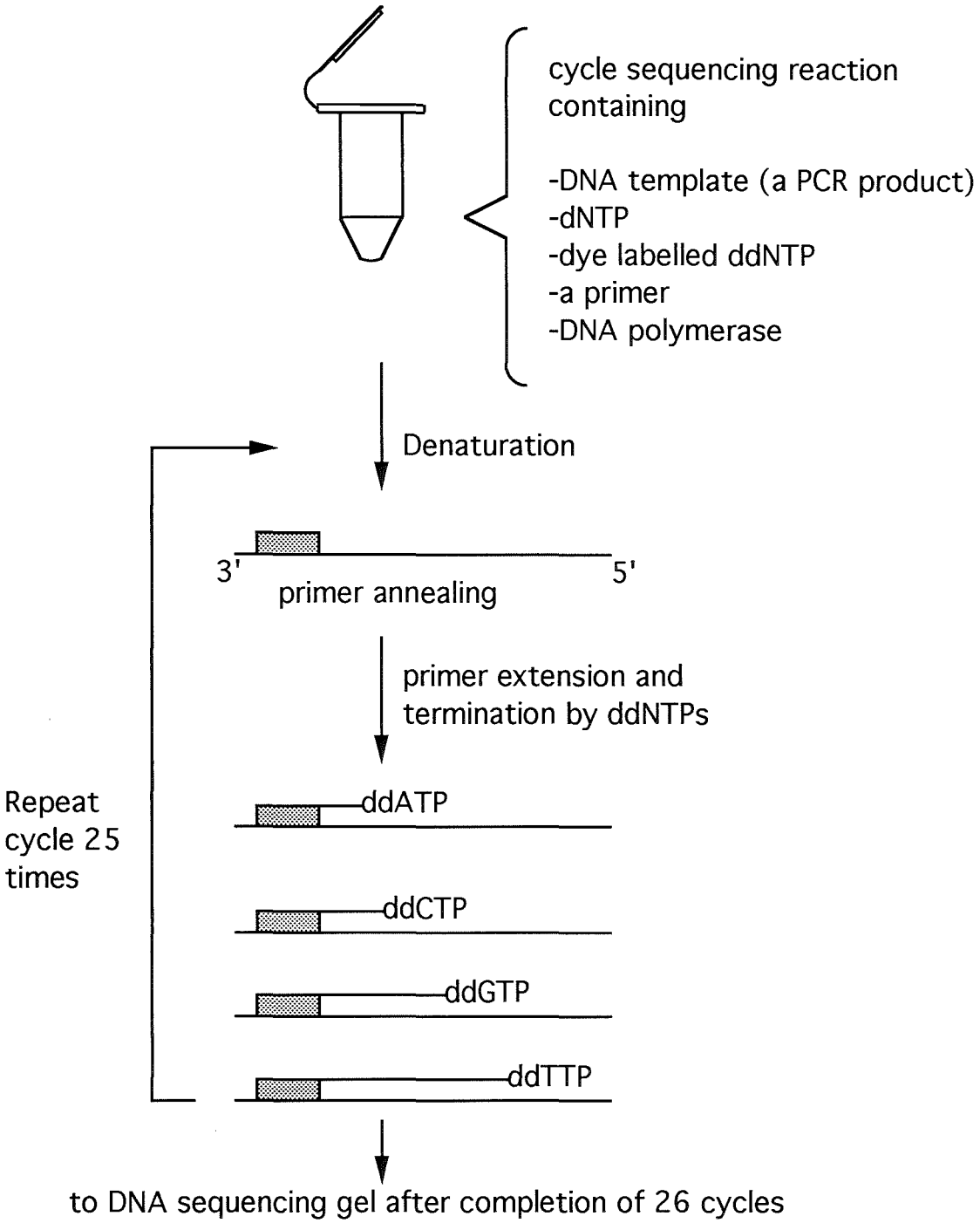


Figure 2.2. Cycle sequencing. Polymerase chain reactionproducts (DNA template) are mixed with a single primer, dNTPs, fluorescent dye-labelled ddNTPs and a thermostable DNA polymerase. Linear amplification is performed on a thermal cycler for 26 cycles with each cycle consisting of denaturation, annealing and extension as in normal PCR. Once these cycles are completed, the cycle sequencing products can be analysed by polyacrylamide gels and detected with an automated fluorescent sequencer (see text on explanation of cycle sequencing reaction conditions).

2.8.1 Cycle Sequencing

The sequencing technique used in the research reported in this thesis is known as cycle sequencing (Fig. 2.2). It uses the concept of the chain termination method described in the previous section. However, cycle sequencing allows sequencing of a double-stranded DNA molecule avoiding the need to obtain single-stranded DNA as a sequencing template. Moreover, it allows sequencing of a DNA fragment produced directly by the polymerase chain reaction (Brown, 1994).

The protocol for cycle sequencing used in this research was a modified protocol described in Perkin Elmer's ABI PRISM^{TM} Dye Terminator Cycle Sequencing Ready Reaction Kit manual. For each cycle sequencing reaction, the following reagents were added into a 0.2 ml microfuge tube: 30-90 ng of cleaned PCR product; 4 µl Terminator Ready Reaction Mix (Perkin Elmer); 5 ng of oligonucleotide primer; and sterile distilled water to a final volume of 10 µl . The Terminator Ready Reaction Mix is made up of dNTPs, ABI PRISM^{TM} Dye labelled ddNTPs terminators and AmpliTaq^{R} DNA polymerase,FS (fluorescent sequencing). Tubes were placed in a Gene Amp PCR System 9600 (Perkin Elmer) thermal cycler and subjected to a cycle sequencing thermal profile consisting of 26 cycles of 10 seconds denaturation at 94^{o}C, 5 seconds annealing at 50^{o}C and 4 mins extension at 60^{o}C.

The cycle sequencing product was cleaned by ethanol precipitation to remove any excess of dye terminator, and was then purified as described in Perkin Elmer's ABI PRISM^{TM} Dye Terminator Cycle Sequencing Ready Reaction Kit manual. To a 1.5 ml microfuge tube, 2.0 µl 3 M sodium acetate, pH4.5 and 2.5 µl absolute alcohol were added together with the

entire 10 µl of cycle sequencing product. The components were mixed thoroughly by vortexing, after which the mixture was left on a bench at room temperature for 5 mins before being placed on ice for 10 mins. The tube was then centrifuged at 13,000 rpm for 25 mins before drainning off the solution. 300 µl of 70% ethanol were added to the solution which was centrifuged again at 13,000 rpm for another 15 mins. The ethanol was drained off once more, and another step of cleaning with 300 µl 70% ethanol and centrifugation was repeated. The ethanol was drained off a final time and the tube was left to dry at 70°C in an oven.

9 µl of loading buffer (comprised of deionised formamide and 25 mM EDTA, pH 8 containing 50 mg/ml Blue dextran in a ratio of 5:1 formamide to EDTA/Blue dextran) were added to the dried template and mixed by vortexing before spinning down briefly at 13,000 rpm. The sample was then heated at 90°C for 2 mins to denature the cycle sequencing product and placed immediately on ice to prevent reannealing of the DNA before being loaded on to a 4% polyacrylamide gel in an ABI377 automated sequencer.

2.9 RAPD analysis

Part of the research involved surveying randomly amplified polymorphic DNAs (RAPDs) using the accessions of *Cryptocoryne* analysed. For RAPD analysis, the following reagents per sample were mixed together in a 0.5 ml 96 well-reaction plate: 1.25 µl of 0.2 mM of dNTP (consisting 0.2mM of each dATP, dCTP, dGTP and dTTP); 2.5 µl of 10x *Taq* Polymerase buffer; 5 pmol of primer; 0.5 units of *Taq* Polymerase (supplied by Promega); 10-15 ng of required DNA template and sterile

deionised distilled water to a final volume of 25 µl. The reaction mixture was overlaid with a drop of mineral oil to prevent evaporation.

The optimal PCR reaction cycle proceded as follows: i) 15 seconds at 94°C to denature the double-stranded DNA; ii) 45 seconds at 36°C to anneal primers; and (iii) 1 min 30 seconds at 72°C for primer extension. The first cycle was preceded by an initial denaturation step of 3 mins at 94°C. Also a 4 mins final extension period at 72°C was used to allow completion of extension and termination of the PCR reaction. 42 cycles involving steps (i), (ii) and (iii) were conducted, and included an increment of 0.4°C/second to step (ii) after the first cycle was completed. The PCR was carried out in a PTC-100, MJ Research thermal cycler.

Chapter 3

Variation within the internal transcribed spacer region of nuclear ribosomal DNA in Cryptocoryninae

3.1 INTRODUCTION

Ribosomal RNA (rRNA) is an essential component in protein synthesis. In plants, DNA sequences that code for rRNA are found in the mitochondrial, chloroplast and nuclear genomes (Hamby and Zimmer 1992).

Within the nuclear genome there are two ribosomal RNA gene families. The first encodes the ribosomal rRNA 18S-5.8S-26S subunits while the second encodes the 5S subunit. In the work reported in this chapter, attention was focused on the internal transcribed spacer (ITS) within the 18S-5.8S-26S gene family.

3.1.1 Structure and organisation of plant nuclear ribosomal DNA

The nuclear ribosomal DNA (nrDNA) of higher plants is arranged in tandem repeats at loci on one or more chromosomes. Each repeat unit contains a transcribed region which is separated from adjacent repeat units by a long, non-transcribed intergenic spacer (IGS) region (Hamby and Zimmer, 1992).

Within the transcribed region of nrDNA are three highly conserved ribosomal RNA regions and two non-coding spacer regions. Figure 3.1 shows that the three highly conserved regions that comprise the ribosomal RNA gene are arranged in a 5'-18S-5.8S-26S-3' order with the ITS spacers (designated as ITS1 and ITS2) flanking the 5.8S region. In addition, an external transcribed spacer region, the ETS, is situated at

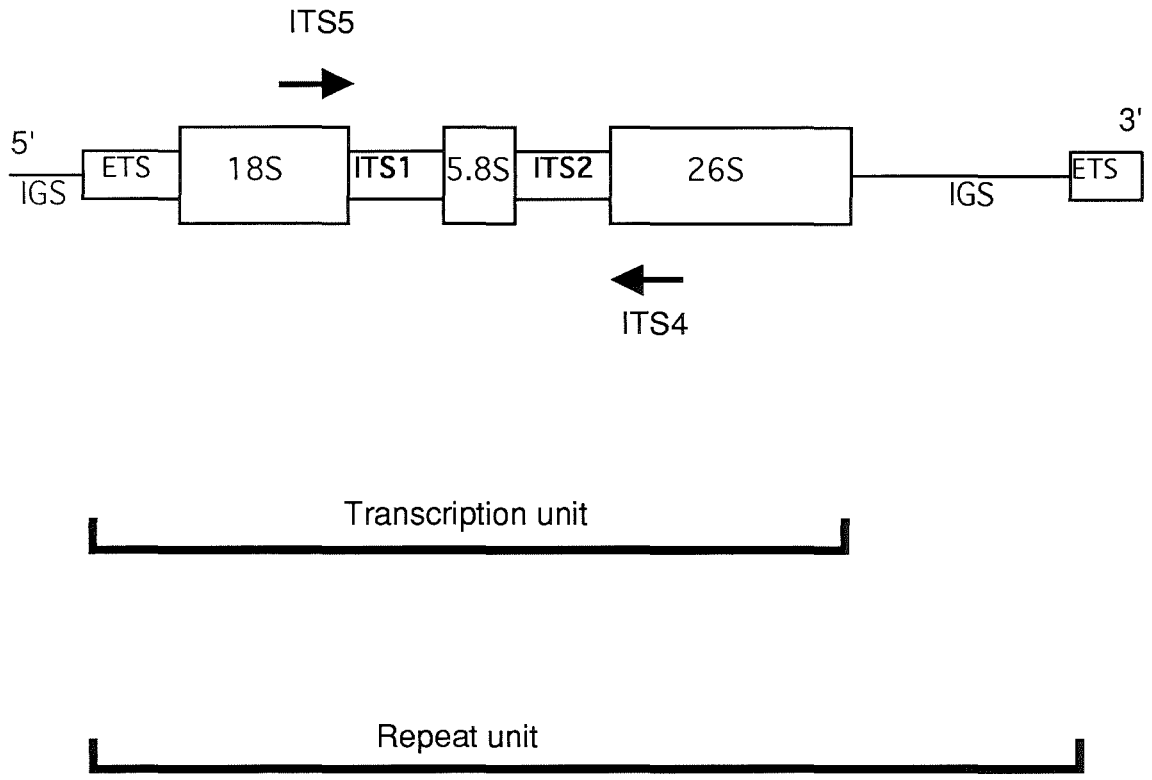


Figure 3.1. Organization of a repeat unit of nuclear ribosomal DNA showing the internal transcribed spacer (ITS) region, external transcribed spacer (ETS), intergenic spacer (IGS) and the 18S, 5.8S and 26S regions. Arrows indicate approximate positions of primers used to amplify the ITS region and for sequencing. Primers were from White *et al.* (1990). Primers sequences (5' to 3'): ITS4 = TCCTCCGCTTATTGATATGC; ITS5 = GGAAGTAAAAGTCGTAACAAGG.

the beginning of the 5' end of the transcribed unit. The whole region is transcribed as a single large precursor, and is subsequently processed into the 18S, 5.8S and 26S functional rRNA forms (Rogers and Bendich, 1987). The intergenic spacer (IGS) is, for convenience, divided into three subregions, i.e. a series of tandemly repeated subrepeats, which are flanked by a 3' end on one side and a 5' end at the other side (Appels and Dvorak, 1982). This repeated sequence varies interspecifically in length generally from 100 to 200 base pairs, while within species its length normally varies slightly (Jorgensen and Cluster 1988). The variation in length within the IGS is due to variation in the length of the subrepeats and the number of subrepeats present.

3.1.2 Use of rDNA ITS sequence variation in phylogenetic reconstruction.

Early studies of the ITS sequence revealed that it was more variable at the nucleotide level relative to other coding regions of nrDNA. Jorgensen and Cluster (1988) compared the nucleotide sequence of the coding region 5.8S between the pea (*Pisum sativum*) and the broad bean (*Vicia faba*) and also between the pea and the lupin (*Lupinus luteus*). They found that the number of base differences were one and two respectively. In contrast, sequence divergence for the ITS1 was much greater; for example, between the pea and lupin the percentage difference for one part of the ITS1 sequence was 16-18% while for the other part it rose to 55-58% (Jorgensen and Cluster, 1988).

Studies of restriction site variation of nrDNA between closely related plant species have revealed sequence variation within the ITS region, although this is not as great as is found within the IGS (Appels and

Dvorak, 1982; Kim and Mabry, 1991). In addition, studies based on restriction enzyme analyses, length variation has been reported in the ITS (Jorgensen and Cluster, 1988), although this is rare relative to length variation in the IGS which is very common (Rogers and Bendich, 1987). The occurrence of high nucleotide variability in combination with low length variation of the ITS suggested that this part of the rDNA gene could be used in reconstructing phylogenies within plant genera, in that sequences could be readily aligned across closely related species, and yet would contain sufficient sequence variation for the resolution of a phylogeny (Baldwin *et al.*, 1995).

Additional features that have made the ITS region of the 18S-26S ribosomal cistron suitable for use in phylogenetic reconstruction at both generic and species level are as follows. First, as a component of a nrDNA multigene family, the ITS region is highly repeated within the plant nuclear genome with copy number varying at the individual, population and species levels (Appels and Honeycutt, 1986; Roger and Bendich, 1987; reviewed by Hamby and Zimmer, 1992). Of interest in this respect is that a study of a large population of broad bean, showed copy number to vary from 500 to 44,000 per individual (Roger and Bendich, 1987). The high copy number per individual promotes detection, amplification, cloning and sequencing of the ITS (Baldwin *et al.*, 1995).

Secondly, although the ITS is quite small in size in the Angiosperms, (i.e. approximately 700 base pairs) it has been relatively easy to develop primers for its amplification, due to the fact that it is flanked by highly conserved regions (18S and 26S). Thus a set of primers (Fig. 3.1) designed for the ITS by White *et al.* (1990), has been used successfully for

amplification across a wide range of plant families.

Thirdly, and of considerable importance, is that the ITS, together with the rest of the nrDNA gene undergoes rapid concerted evolution, such that all members of the gene family exhibit the same sequence within an individual (Arnheim *et al.*, 1983). Two causes of concerted evolution are unequal crossing over and gene conversion. Unequal crossing over is a reciprocal recombination process where sequence duplication is created in one chromatid or chromosome and a complementary deletion occurs in the other. This process of unequal change might take place either between two sister chromatids of a chromosome during mitosis of a germ-line cell or between two homologous chromosomes during meiosis (Li and Graur, 1991). Figure 3.2 gives a hypothetical example of unequal crossing over within a multigene family which is composed of five different variants (a to e). After an initial unequal change, a duplication of c-type unit (a-b-c-c-d-e) occurs in one daughter chromosome while the other daughter chromosome loses the c-type unit (a-b-d-e; not shown). As this process is repeated a few times the daughter chromosome will become more homogenised and finally only one type of multigene family (in this example the c-type) will spread to fixation throughout the gene family.

In contrast to unequal crossing over, gene conversion is a non-reciprocal recombination process in which two sequences interact in such a way that one is converted by the other (see Lewin, 1994). Gene conversion is important in concerted evolution as it allows conversion between genes located at different chromosomal loci (non-allelic form). Figure 3.3 shows an example of a non-allelic gene conversion in which two wild type repeats are converted into a mutant type. As a result, the first daughter

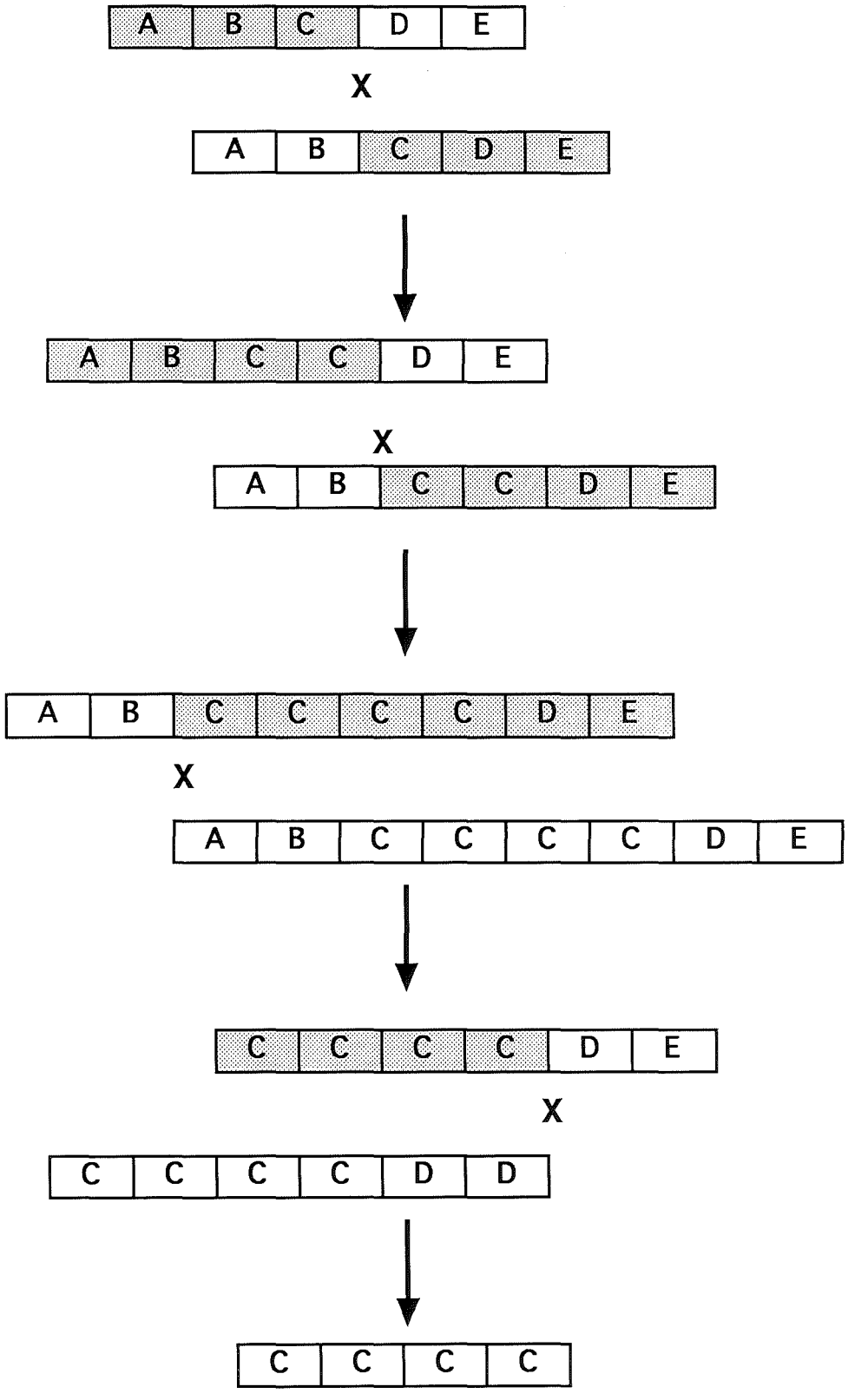


Figure 3.2. Model of concerted evolution by unequal crossing over (after Li and Graur, 1991). Shaded repeat units will form recombinant repeat units in daughter chromosomes. Repeated cycles of unequal crossing over events cause the duplicated genes on each chromosomes to become progressively more homogenised.

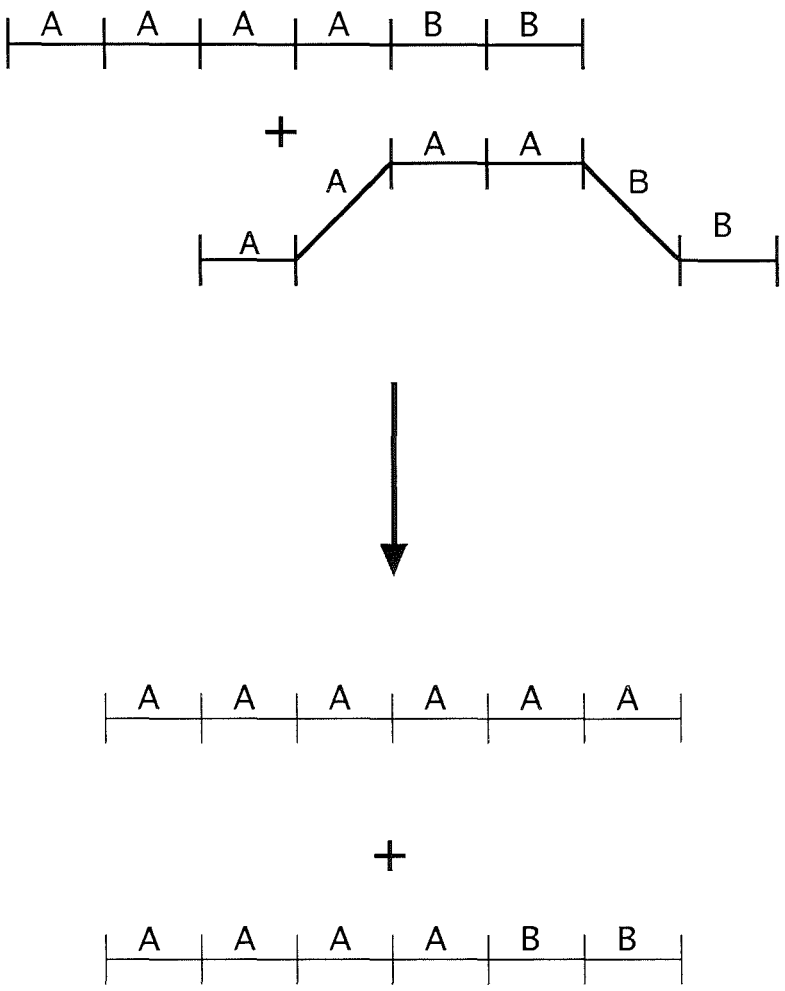


Figure 3.3. Model of concerted evolution by gene conversion (after Li and Graur, 1991). The B repeat type of the first chromatid was converted into the A repeat type. As a result one daughter chromatid has all the A repeat type while the other chromatid maintain the parental repeat type. Thus gene conversion changes the frequencies of the two types of repeats in only one of the daughter chromosomes but does not alter the total number of repeats in either chromosomes.

gene family becomes more homogeneous than the parental gene family while there is no change in the second daughter gene family. Gene conversion acts as a correction mechanism on dispersed repeats as well as tandem repeats (Li and Graur, 1991).

3.1.3 Use of ITS sequence variation in plant systematic and evolution.

Early investigation of the utility of the ITS sequence in studies of the reconstruction of angiosperm phylogenies was conducted by Baldwin(1992) who sequenced the ITS region of 12 representatives of the Compositae subtribe Madiinae and two outgroup species. He found that there was a high sequence alignability and minimal length variation among ITS1, 5.8S and ITS2 in the taxa surveyed which facilitated the determination of positional homology of nucleotide sites. In addition, phylogeny constructed among the Hawaiian silversword alliance species and California tarweed taxa based on the ITS sequences was highly congruent with a chloroplast DNA phylogeny constructed for the same group. Among other things, the ITS phylogeny suggested the monophyly of the Hawaiian silversword, originated from a California tarweed ancestor. It was also found that the Hawaiian silversword species are more closely related to two California tarweed genera, *Madia* and *Raillardiopsis* than to the other two California tarweed genera included in the analysis. Based on these findings, Baldwin concluded that the ITS sequence might provide a valuable source of nuclear phylogenetic markers in plants.

Further studies have confirmed that ITS sequences provide a most useful independent line of phylogenetic evidence in studies of plant systematics and evolution (Baldwin *et al.*, 1995). Thus, an improved resolution of relationships within lineages has been obtained using this

approach in the genera *Robinsonia* (Sang *et al.*, 1995), *Panax* (Wen and Zimmer, 1996) and *Gentiana* (Gielly *et al.*, 1996; Yuan *et al.*, 1996). From a survey of ITS sequence variation, Bayer *et al.* (1996) showed that the genus *Antennaria* is comprised of six monophyletic groups to confirm what had been proposed for the genus based on studies of morphological and chloroplast DNA restriction site variation. However, the ITS tree gave better support for the inclusion of *A. gayeri* within *Antennaria,* whereas previous morphological and cpDNA comparisons were inconclusive on this point. *A. gayeri* exhibits a tendency towards polygamodioecy (i.e. individuals produce either only male, female or hermaphrodite flowers) and also lacks basal leaves. In these traits, therefore, it resembles species in the genus *Anaphalis* rather than other *Antennaria* species. The ITS data, however, provided strong support for the inclusion of *A. gayeri* within *Antennaria.* ITS variation in *Antennaria* also suggested that the Argantae group of species was comprised of three taxa, *A. argantea, A. luzuloides* and *A. stenophylla* with the last two taxa considered to be sister taxa. In the ITS tree, this grouping was supported by six synapomorphies. In contrast, previous cladistic analysis using morphological characters placed *A. stenophylla* into the Dimorphae group along with *A. flagellaris* and *A. dimorpha.* The correct placement of *A. stenophylla* in the Argantae group, based on ITS sequence variation, draws support from the fact that both *A. stenophylla* and *A. luzuloides* are similar in having narrow linear leaves and small flowering heads.

Another study, by Sun *et al.* (1994), in the genus *Sorghum,* has also shown the value of ITS sequence variation in clarifying phylogenetic relationships within a genus. ITS sequence data showed that a member of

section *Parasorghum, Sorghum australiense* exhibits a close affinity to two members of section *Stiposorghum, S. matarankense* and *S. stipoideum.* Originally, all three of these species were grouped into one section, *Parasorghum,* based on their geographical distribution (restricted to Australia) and the fact that all three species had hairy nodes and simple penicle. However, these species were later separated into two different sections based on the shape of the callus and the length of awns, although defining the shape of callus as 'pointed' or 'obtuse' is difficult, and awn length may not be a good criterion for classification. The ITS phylogeny supported the original view that these three species should be placed in the same section and that it would be best to combine *Parasorghum* and *Stiposorghum* and conclude that *Parasorghum* is not monophyletic.

ITS phylogenies, in some instances, have provided new insights into relationships between taxa. For example, in the *Heuchera* group (Saxifragaceae), chloroplast DNA restriction site variation suggested that both northern and southern populations of *Tellima grandiflora* in the United States of America were not closely related. In contrast, the ITS tree for this group supported the notion that the southern and northern populations of *T. grandiflora* are conspecific as indicated by their morphology and allozyme data (Soltis and Kuzoff, 1995). It was concluded that introgressive hybridisation between *T. grandiflora* and a species of *Mitella* had led to chloroplast capture of *Mitella* chloroplast DNA by some populations of *T. grandiflora* (Soltis *et al.*, 1991) thus causing the high level of cpDNA divergence found in his species.

Additional studies have demonstrated that within the Angiosperms, ITS

sequences are useful for investigating phylogenetic relationships above the species level. Divergence between ITS sequences has been useful in resolving phylogenetic relationships between genera within the Agavaceae (Bogler and Simpson, 1996), Berberidaceae (Kim and Jensen, 1996), Rosaceae (Campbell *et al.*, 1995), Saxifragaceae (Soltis *et al.*, 1991; Soltis and Kuzoff, 1995), Asteraceae (Susanna *et al.*, 1995; Kim *et al.*, 1996a), Gentianaceae (Yuan and Kupfer, 1995), Caryophyllaceae (Oxelman and Liden, 1995), Poaceae (Hsiao *et al.*, 1995) and Ericaceae (Kron and King, 1996). In some of these studies, for example, within the Apiaceae, ITS sequences gave a better resolved phylogenetic tree than was obtained previously from an analysis of chloroplast DNA variation (Downie and Katzdownie, 1996).

ITS sequence variation has also proved useful in clarifying the origins of plant genera on oceanic islands. For example, Sang *et al.* (1994) used the ITS sequence to examine the holophyly of the genus *Dendroseris* on the Juan Fernandez Islands. Eleven species of *Dendroseris* that are found on the Juan Fernandez Islands were divided into three subgenera, and are considered endemic to the islands. The three subgenera are quite distinct morphologically and they are widely divergent from any known continental *Dendroseris*. This indicated that the genus might have originated from a single introduction to the islands. Chloroplast DNA restriction site mutation analysis provided weak support for holophyly of the island species with only three mutations occurring relative to the outgroup species used. In contrast, an analysis of ITS sequence variation provided much stronger support for holophyly with 25 mutations present compared to the outgroup. However, the ITS variation provided

insufficient resolution of relationships among the three subgenera. It was concluded that during a period of rapid morphological radiation among the three subgenera, an insufficient number of mutations accumulated in the ITS region to allow relationships between subgenera to be made evident. Other studies that have used ITS sequence variation to clarify the evolutionary history of plant genera endemic to oceanic island, have been conducted on the genus *Robinsonia* - also endemic to Juan Fernandez Islands (Sang *et al.*, 1995) and the genus *Schiedea* and *Alsinidendron* - endemic to the Hawaiian Islands (Soltis *et al.*, 1996).

Care must be taken in the use of ITS sequence variation in the analysis of reticulate evolution in plant genera. This is because concerted evolution may act across repeat units contributed by different parent species. Sequence homogenisation in the hybrid due to concerted evolution is likely to be reduced if: 1) the hybridisation event was recent, 2) the nrDNA repeat units are located at different chromosomal loci or on different chromosomes in the parental taxa, and gene conversion fails to operate in the hybrids, or 3) the hybrid is asexual (Baldwin *et al.*, 1995). When sequence homogenisation occurs, however, it can lead to some unexpected findings. For example, in a study of the origin of the allopolyploid, *Saxifraga osloensis* (Brochman *et al.*, 1996), direct sequencing of the ITS was carried out on hybrid individuals from four populations, together with diploid individuals from three maternal and two paternal populations of the parental species from Sweden. The allopolyploid which is endemic to Sweden, was found to possess an ITS sequence virtually the same as that of *S. adscendents*, (the maternal parent) and different from that of *S. tridactylis* (the paternal parent) due to

34 to 44 base substitutions. In contrast, an analysis of RAPD variation in the same material showed that *S. osloensis* normally exhibited the expected additive RAPD phenotype which combines the distinctive phenotypes of the two parent species.

However, sequence homogenisation is not always a problem and ITS sequencing can help in the detection of hybrids. For example, Campbell *et al.* (1993) found that in the genus *Amelanchier,* the ITS sequence of the putative parents were both present in a presumed hybrid, *A. x neglecta.* Similarly an allohexaploid species in the genus *Krigia,* exhibited the expected additive ITS sequences of the putative diploid and tetraploid parent species (Kim and Jansen, 1994).

In *Gossypium,* sequence homogenisation appears to have occured in some hybrid taxa in the direction of either parent, a phenomenon that could lead to much phylogenetic confusion should it be common. Wendel *et al.* (1995) found that whereas certain populations of an allopolyploid *Gossypium* species possessed the ITS sequence of one of the parental species, other populations of the same hybrid taxon possessed the ITS of the other parent.

3.1.4 Aims and objectives of sequencing the ITS of *Cryptocoryne* species.

The primary goal of sequencing the ITS region of *Cryptocoryne* species in this study was two-fold. First, to reconstruct a phylogeny of the genus *Cryptocoryne* and establish relationships between species from different geographical locations. Second to examine chromosomal evolution

within the genus. Previous studies have been inconclusive as to whether the aneuploid series within *Cryptocoryne* arose by ascending or descending euploidy or through both ascending and descending euploidy from an ancestor with a low chromosome base number (as suggested by Reumer, 1984). It was hoped that the phylogeny constructed from ITS data would shed some light on the evolution of chromosome numbers within *Cryptocoryne*.

3.2 Materials and Methods

3.2.1 Plant material

Twenty five species of *Cryptocoryne* representing all four geographical sections (see Reumer, 1984) were subjected to analysis. They represented all but one (x=10) of the different chromosome base numbers present within the genus, and 13 of the 24 taxonomic groups identified by Jacobsen (1977) and Arends *et al.* (1982). Two species of *Lagenandra,* the genus which forms the other half of the subtribe Cryptocoryninae were used as outgroup taxa. The list of species and accessions analysed is given in Table 2.1 (chapter 2).

3.2.2 DNA isolation, ITS amplification and sequencing

The extraction, purification and quantification of DNA from leaf samples was carried out as described in sections 2.3 and 2.4 (chapter 2). Further, the amplification of the ITS region as well as cleaning of the amplified products were carried out as outlined in section 2.5 and 2.6 (chapter 2) respectively. Sequencing of the ITS region was done using the same procedure as described in section 2.8.1 (chapter 2)

3.2.3 Sequence alignment and treatment of sequence insertions and deletions (indels)

DNA sequences from the ITS1, 5.8S and ITS2 regions were aligned manually by sequential pairwise comparisons. Alignment required interpretation of gaps which appeared in the sequences of different taxa throughout the three regions.

Currently there are two ways in which indels (insertions and deletions of nucleotides) can be incorporated in the phylogenetic analysis of a group of taxa (Wojciechowski *et al.* 1993). Each gap position can be treated as a missing data item, or alternatively as a new character, i.e. the fifth base. Treating gaps as missing data allows information to be retained on base substitutions occurring in those taxa within the indel region. However, it will exclude information regarding the evolutionary events or transformation involved in the insertion or deletion of bases. On the other hand, scoring indels as separate characters will increase the risk of overweighting indels in the analysis if adjacent gaps are non-independent due to erroneous decisions made during alignment (Baum *et al.*, 1994). In the present study indels were scored as missing data.

3.2.4 Phylogenetic inference

Several procedures are used in phylogenetic inference and these will be discussed below. Data obtained from DNA sequence can be categorised as either; 1) discrete characters and, ii) distance characters. Such characters are used most commonly to generate phylogenetic trees using parsimony or a distance approach.

3.2.4.1.1 Discrete Data

The use of sequence data for phylogenetic analysis is straightforward. For a set of DNA sequences, the character is represented by a corresponding position in the sequence, while the nucleotide observed (A, C, T and G) at such a position represent the character state. Character

distinction is dependent on two assumptions (Swofford *et al.* 1996). The first is positional homology or identity, that is the nucleotide observed at a given position in the taxa under study is said to have been derived from a single ancestral nucleotide for a particular position. Positional homology is established by aligning sequences, so as to minimise mismatches. Accurate alignment of indels (insertions or deletions) must be conducted (by inserting gaps) in order that positional homology is attained.

The second assumption is that characters are assumed to be independent variables, whose possible values are drawn from a collection of exclusive characters states. The assumption of independence among characters is common to most character-based methods of phylogenetic analysis.

3.2.4.1.2 Distance Data

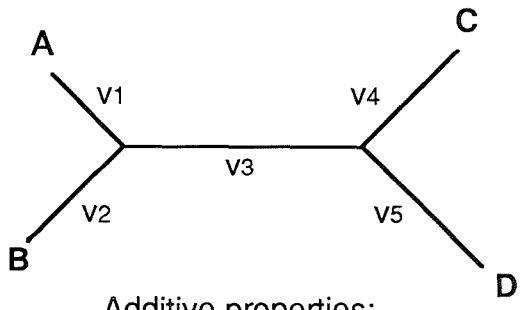
Distance data are data which reflect a measure of degree of dissimilarity (opposite to similarity) between two taxa or genes (Williams, 1992). Two taxa or genes are said to be identical to each other if they are separated by zero distance (zero dissimilarity or 100% similarity). In the case of DNA sequences, the level of dissimilarity between two sequences in a pairwise comparison is equal to the number of aligned sequence positions with non-identical bases divided by the total number of sequence positions compared (Swofford *et al.*, 1996).

For tree construction, distance data can be divided into two types, additive distance data and ultrametric distance data. A tree formed from

additive distance data (Fig. 3.4) is a tree where the evolutionary distance between any pair of taxa would be equal to the sum of the length of the branches connecting them (Swofford *et al.*, 1996). However, the true topology of an additive tree can only be obtained if no character changes its state more than once (Fitch 1981, 1984) or, in short, additivity can only be obtained when there is no homoplasy. If or when the data contains some homoplasy, new calculations will give a different set of values for branch lengths. Conflicting evidence will increase branch lengths and eventually may change branching order.

One method used to construct phylogenetic trees from additive distance data is the neighbour-joining method (Saitou and Nei, 1987), which does not assume the operation of a molecular clock. In this procedure, an initial distance matrix for all taxa is derived from the raw data. From this, a modified distance matrix is constructed where the separation of two taxa is based on their average divergence from all other taxa. A tree is constructed by linking the least distant pair of taxa within this modified distance matrix. After two taxa (or nodes) are linked, a third taxon is added to the tree which will replace the two terminal nodes of the tree with a single new node representing the common ancestor of the pair. This process of replacing two terminal nodes by one new node is repeated until only two nodes remain, separated by a single branch.

In contrast, a tree generated from ultrametric data is a tree in which the distance between any two taxa is equal to the sum of the branches joining them (additive distance) and this tree can be rooted so that all of the taxa are equidistant from the root (Fig. 3.5). Thus it is assumed that a



Additive properties:

$d_{AB} = V1+V2$

$d_{AC} = V1+V3+V4$

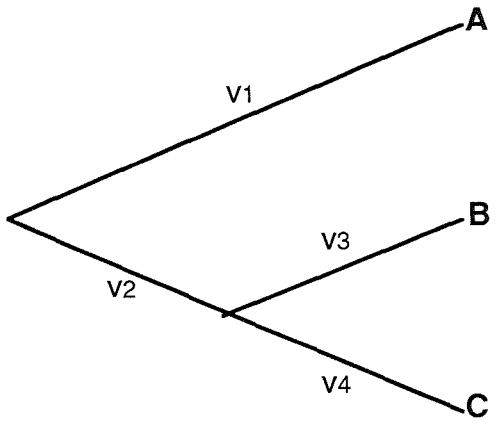
$d_{AD} = V1+V3+V5$

$d_{BC} = V2+V3+V4$

$d_{BD} = V2+V3+V5$

$d_{CD} = V4+V5$

Figure 3.4. An additive tree showing the relationships between four taxa; A,B,C and D. Both additive distances and trees do not make an assumption about rooting, and trees are unrooted. Note: dAB equals the distance between taxa A and B.



Additive properties

$d_{AB} = V1+V2+V3$

$d_{AC} = V1+V2+V4$

$d_{BC} = V3+V4$

Ultrametric properties:

$V3=V4$

$V1=V2+V3=V2+V4$

Figure 3.5. An ultrametric tree showing the relationships between three taxa; A, B and C. In addition to having additive distances properties where all taxon to taxon distances are the total of branch lengths joining them, every common ancestor is the same distance from all of its decendants. For example here, the most recent common ancestor of A and B is v_1 from A and v_2+v_3 from B, therefore $v_1=v_2+v_3$. And the most recent ancestor of B and C is v_3 from B and v_4 from C, therefore $v_3=v_4$.

molecular clock is operating at the same rate in all lineages. A tree is constructed by connecting the least distant pair of taxa then adding successively more distant taxa until all taxa have been joined into the tree. A common method used to construct trees from ultrametric data is the UPGMA (Unweighted Pair Group Mean Analysis) procedure.

3.2.4.2 Phenetic and Cladistic approaches to phylogeny reconstruction

There are two school of thought in regard to reconstructing phylogenies. One adopts a phenetic approach while the other uses cladistics. These two procedures will be briefly discussed in turn.

3.2.4.2.1 Phenetic analysis

Phenetic or numerical taxonomy was first developed in the 1960's and was based on the assumption that overall similarity was the key to an accurate classification (Siebert, 1992). The aim of phenetics was to overcome intuitive methods which considered that certain characters should be treated as more important in a classification and consequently given more weight than others. Phenetic classification aimed to be objective, explicit and repeatable both in evaluation of taxa and taxon recognition. This was to be achieved by producing consistent data matrices, examining large character sets and weighting all characters equally.

Classification was based on the construction of a phenogram (phenetic tree) produced from similarity or dissimilarity matrices generated from the original data and using a clustering algorithm.

Phenograms are composed of terminal taxa or operational taxonomic units (OTUs), which are clustered together at different hierarchical levels determined by overall percentage of similarity or dissimilarity. Evolutionary relationships between taxa, however, are not examined as no attempt is made to distinguish between homologous and homoplasious character. If a taxon has accumulated many autapomorphies (i.e. uniquely derived evolutionary characters), then it might be clustered further apart from other taxa, simply because of these autapomorphies. Moreover, similarities possibly due to parallel and convergent evolution are ignored (Scotland, 1992).

3.2.4.2.2 Cladistic analysis

Cladistic or phylogenetic systematics was first formulated by Willi Hennig in 1950 based on an idea that in sexually reproducing organisms, diversification in evolution is due to speciation, and speciation either happens or does not. Hennig also considered that most speciation were dichotomous and that sister species share unique characters (Panchen, 1992).

The pattern of speciation, and consequently the evolutionary history of taxa, is reflected by a cladogram or phylogenetic tree. This tree is generated using the derived or apomorphic characters that are shared by particular taxa. It, therefore, groups taxa based on their common ancestry. A succinct description of cladistic methods, outlining the key features of the approach, has been given by Funk (1995). Each terminal taxon or species in the cladogram is defined by having one or more unique

characters (autapomorphies), and two closely related taxa or sister groups are clustered together by sharing one or more characters that are uniquely derived for this pair (synapomorphies). Cladistic analysis differs from phenetic analysis in the way that homologous and synapomorphic characters are detected and treated. In phenetics analysis, no consideration is made of synapomorphies (shared derived character), symplesiomorphies (shared ancestral characters) and whether characters are truly homologous (similarity due to common ancestry). Consequently a detailed understanding of the evolutionary history of a group of taxa cannot be obtained from a phenetic analysis.

3.2.4.3 Phylogeny estimation using parsimony method

Parsimony is perhaps one of the most popular tree building methods used in cladistic analysis. The method is based on the hypothesis that simple character changes are more likely to have occurred than more complicated ones during the evolution of a group of taxa (Swofford *et al.*, 1996). Moreover, when there is no reason to think otherwise, two characters that appear to be the same should be treated as homologous. If, however, the character is clearly not homologous, i.e. it supports conflicting groups, the explanation that is the simplest should be chosen, i.e. the one that requires the smallest number of homoplasious characters and character loss.

The method of inferring phylogeny based on the concept of parsimony operates by selecting the tree(s) that has the shortest branch length for a particular data set. There are four different form of parsimony

used in phylogenetic reconstruction. These differ in their assumptions about character changes as outlined below:

1. Wagner parsimony. This approach was first used by Kluge and Farris (1969) and treated characters such that the change from one character state (A) to another (D) involved changes through intervening character state (B and C) in the transformation series. Character states were allowed to reverse freely.

2. Fitch parsimony. In this procedure, transformation from one character state to another can proceed in an unordered fashion, and therefore does not involve changes through intervening character states. Once again character states are allowed to reverse freely.

3. Dollo parsimony. In this approach, each shared derived character state (synapomorphy) is uniquely derived and appears only once in the tree. Character state reversal is allowed, but once a state is reversed it cannot reappear. Multiple origins of a derived state either by convergence or parallelism are also not permitted.

4. Camin-Sokal parsimony. This method makes the assumption that character state evolution is irreversible. This particular parsimony procedure is rarely used as the assumption of irreversibility of character states is very difficult to justify (Kitching, 1992).

A frequently used computer package for constructing phylogenetic

trees using parsimony is phylogenetic analysis using parsimony or PAUP (Swofford, 1993). This package allows the evaluation of the optimal tree topologies of constructed trees using the following three approaches:

1. Exhaustive search.

This approach is illustrated in Fig. 3.6. An initial tree for the first three taxa examined is constructed and the fourth taxon is added and evaluated in every topology (three different possible positions). Then each additional taxon is added and every single tree topology is evaluated as subsequent taxa are added. The difficulty with this search is that the number of trees increase rapidly with the addition of further taxa.

2. Branch-and-Bound search

The branch-and-bound approach to evaluating optimal tree topology employs a search procedure which has a provision for discarding trees without evaluating them in detail. It can provide an exact solution for a larger number of taxa than the exhaustive approach although in essence it closely resembles the exhaustive search method. In the first step of tree building (see Fig. 3.7), three taxa (A, B and C) are used to obtain the first tree. Then another taxon is added to this tree which will result in three possible tree arrangements (B1, B2 and B3). The fifth taxon may then be placed into five different places within each of the three trees yielding 15 trees altogether. However, if an upper bound (in regard to tree length) is incorporated into the search, then all trees which exceed this bound in length can be eliminated. In practice, when the upper bound is exceeded, the branch will be cut off and no evaluation will be done on trees with

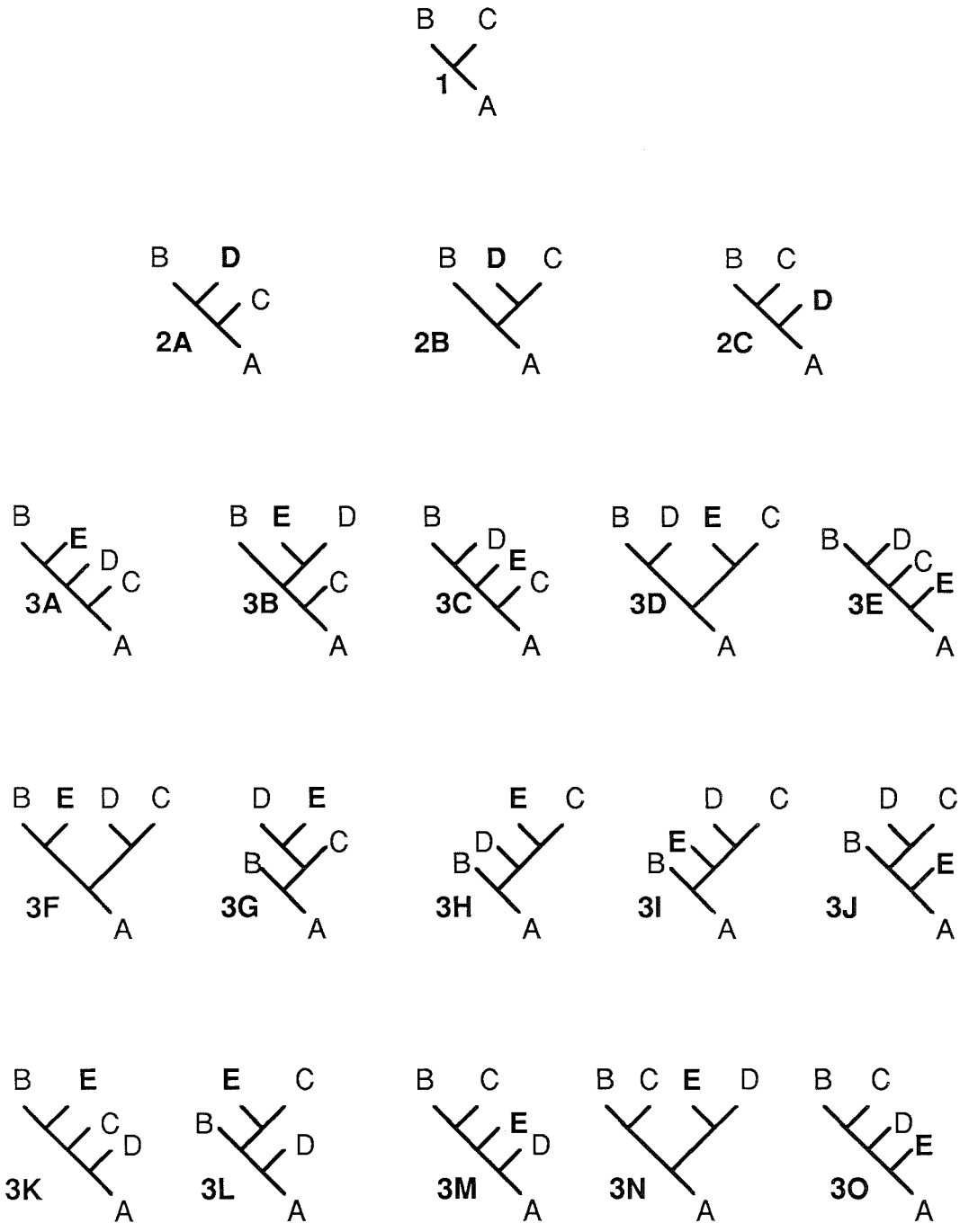


Figure. 3.6 Illustration of the exhaustive search strategy (after Swofford *et al.*, 1996).

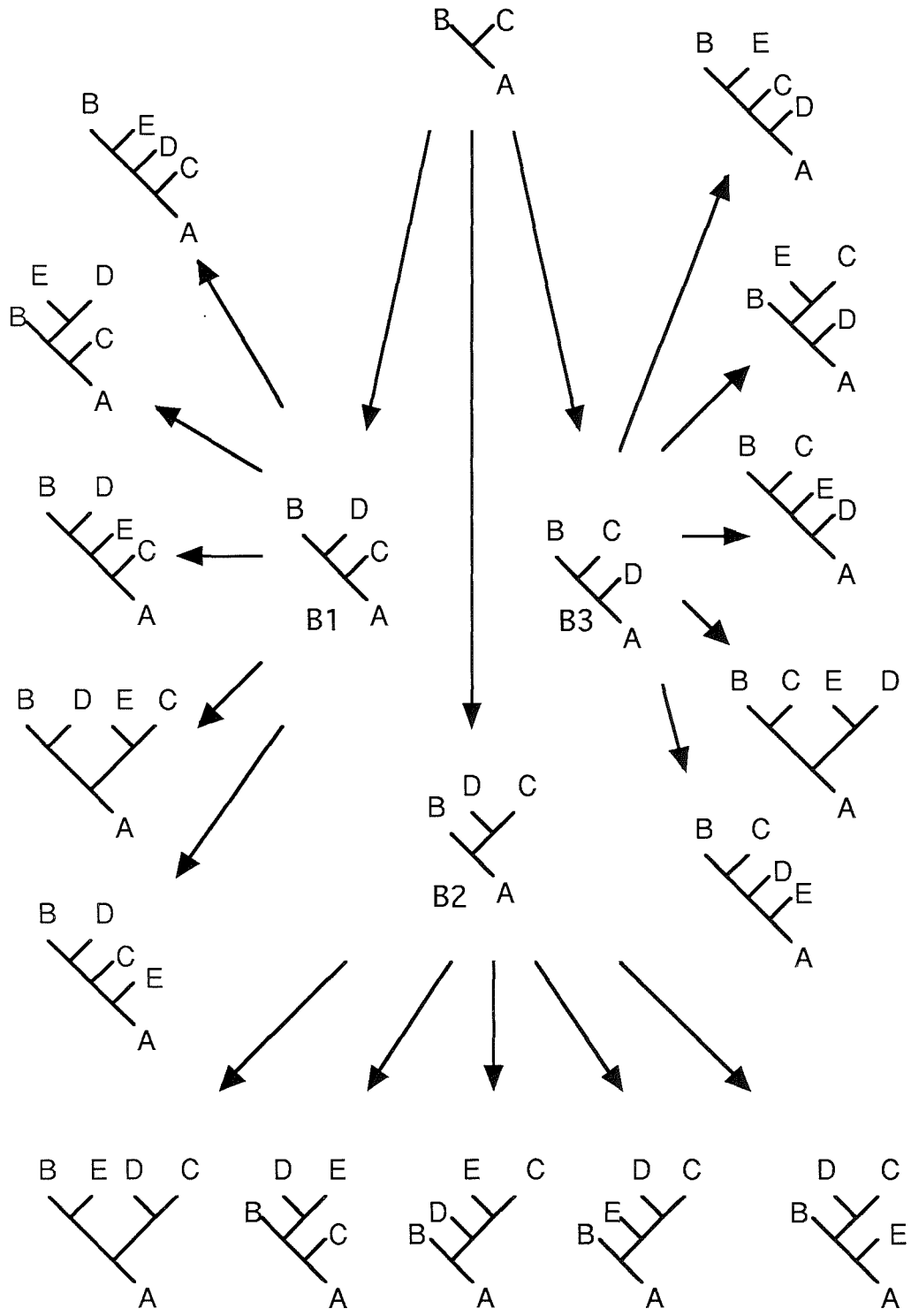


Figure 3.7. Illustration of search tree for branch-and-bound algorithm (after Swofford *et al.*, 1996).

additional branches connected to this branch. Thus it is possible to backtrack down the search tree and proceed up another branch and determine whether this produces a tree with a length below the upper bound.

3. Heuristic search.

A heuristic search is carried out if the data set is too large to allow the use of an exact method within a given period of computing time. The search begins by building a tree which may or may not be optimal in regard to its topology. The manner in which each taxon is added to a growing tree is constrained. This yields a tree that is optimal given the constraints to the way in which taxa were added, but need not necessarily be the best tree if all possible trees were evaluated. This method is rather analogous to proceeding along a reasonable path in search of a tree which might or might not contain the best tree without having the option of backtracking to explore other possibilities (Wiley *et al.*, 1991).

The heuristic search procedure may be improved in several ways so that the tree with the local optimum may approach a global optimum in regard to its topology. In PAUP, two procedures are used to achieve this, and they are (i) stepwise addition and (ii) branch swapping.

1. Stepwise addition

Stepwise addition is the common method for obtaining a starting point for further rearrangement of additional taxa to a growing tree. Firstly, three taxa are chosen for the initial tree. Then one of the unplaced taxa is selected for next addition. The trees which results from the addition of the

fourth taxon to the tree are evaluated and the one with the optimal score is saved for the next round. In the next round, the fifth taxon is placed along one of the five possible branches on the tree saved from the previous round. The evaluation procedure is repeated with the best tree saved for the next round. The process is terminated when all taxa have been joined to the growing tree. A problem with this approach is to decide which three taxa should be used for a starting tree, and which of the unplaced taxa should be attached to the tree next etc. One solution to this problem is to start the tree and add taxa to it in the same order as taxa are presented in the data matrix. An alternative approach is to check first all triplets of taxa and start the tree with the three taxa that produce the shortest tree. After this all unplaced taxa are considered for connecting to each branch of the tree and the tree with the smallest length for a given taxon-branch combination is chosen. This process of evaluating the length of each taxon-branch combination is repeated until all taxa are placed on the tree(s) with the smallest length (Swofford *et al.*, 1996).

2. Branch Swapping

Branch swapping is conducted by performing a set of predefined rearrangement of branches within trees. These arrangements are performed in the hope that a better tree will be found. If a better tree is found, a new rearrangement is initiated on this tree and the process is continued so as to come closer to a topology near to the global optimum. Three branch swapping algorithms are implemented in PAUP: Nearest Neighbour Interchanges (NNI, see Fig. 3. 8), subtree pruning and regrafting (SPR, see Fig. 3.9) and tree bisection and reconnection (TBR, see Fig. 3.10). The descriptions of these procedures in the figures are taken from Swofford *et al.* (1996).

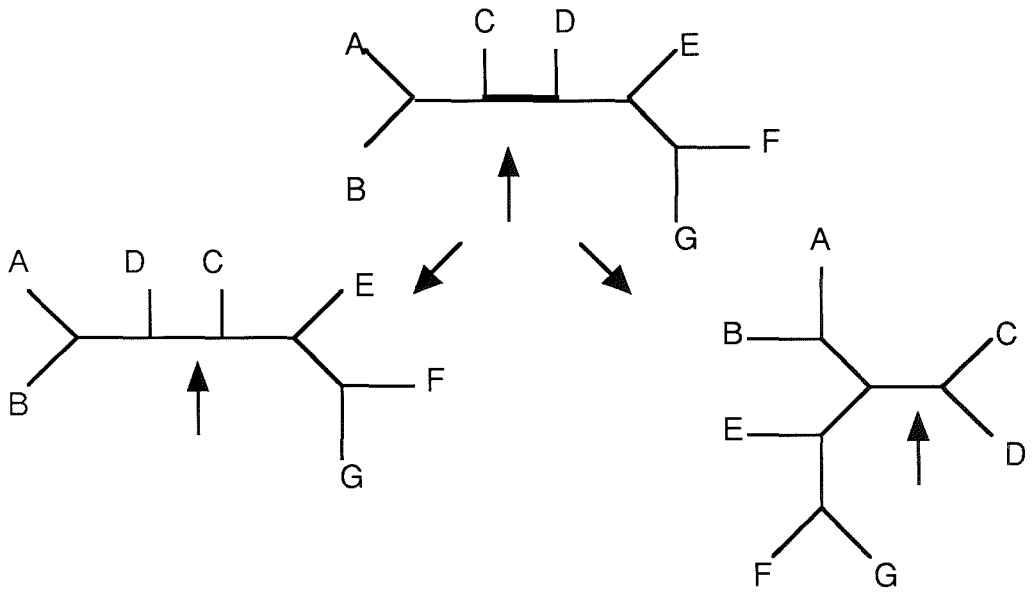


Figure 3.8. Branch swapping by nearest neighbour interchange (NNI). In this procedure each internal branch of the tree is a local region of four subtrees connected by internal branches and each interal branches can be considered to link two subtrees on one side to two on the other. In the example the highlighted branch connects the two left hand side subtrees of (A+B) and C to two right hand side subtrees D and (E+(F+G)). A NNI exchanges one subtree on the left hand side with one on the right. In this case either D or (E+(F+G)) can be exchanged with C or (A+B) to give two possible NNI rearrangements.

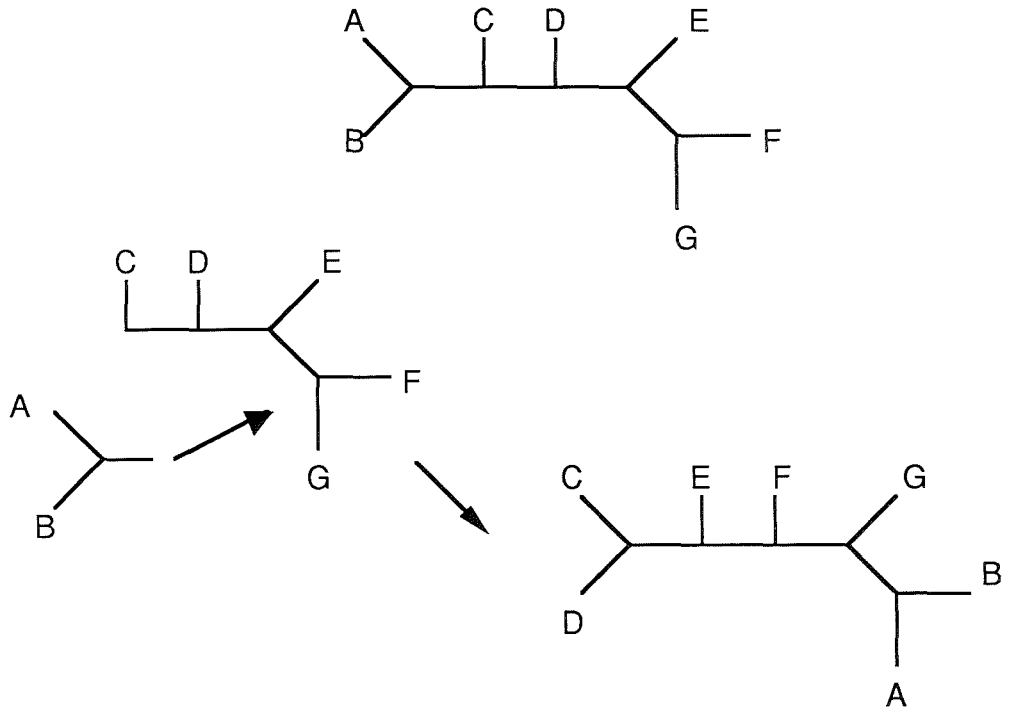


Figure 3.9. Branch swapping by subtree pruning and regrafting (SPR). In this method, the tree is divided into two, producing one subtree (A=B) with a free branch. This pruned subtree is then re-attached by joining its free branch to either a terminal or internal branch of the other subtree. All possible subtree removals and reattachment points are evaluated.

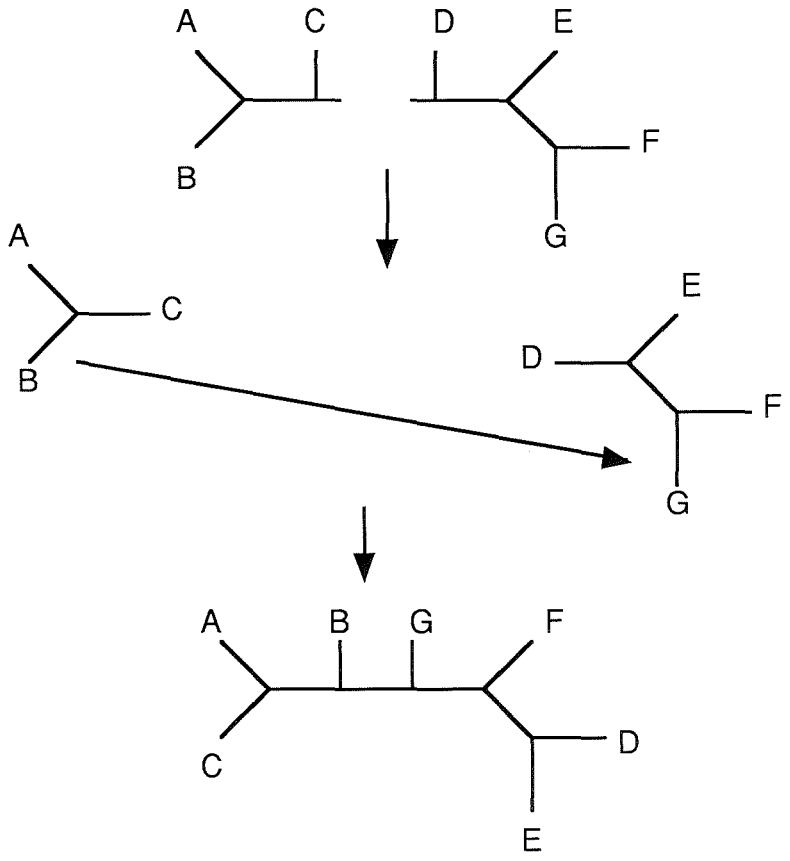


Figure 3.10. Branch swapping by tree bisection and reconnection (TBR). A tree is bisected along a branch resulting in two disjoint subtrees. The two subtrees are then reconnected by joining a pair of branches from each subtree. In the example above the new branch is created between the branches bearing taxa B and G. All possible bisections and reconnections are evaluated.

3.2.4.4 Consensus tree

Phylogenetic analysis conducted in the ways described above usually results in the production of a number of equally parsimonious trees which vary in topology. A consensus tree may then be constructed which summarises the information contained within the equally parsimonious trees. It is to be noted that a consensus tree does not necessarily give the best estimate of phylogenetic relationships among groups. It only summarises the phylogenies of the group under study and thus must be interpreted carefully.

Frequently a strict consensus tree is constructed. This is the most conservative approach. A strict consensus tree reflects only the groups that are found in all the equally parsimonious trees. In short it includes only those monophyletic groups that are totally unambiguous. This constraint is particularly restrictive and usually results in a tree with unresolved polytomies.

Another approach operates on the majority rule basis. This approach places less constraint on the production of a consensus tree in that a monophyletic group may be preserved in the consensus tree, even though some equally parsimonious trees may support a conflicting group. This is illustrated in Fig. 3.11, where in tree 1 it is assumed that A is more closely related to B than either is to C. In contrast, in trees 2 and 3 it is assumed that B is more closely related to C than either is to A. The majority consensus tree adopts the latter topology, whereas the strict consensus tree contains an unresolved polytomy for A, B and C. Other procedures for constructing consensus trees are described in Swofford *et al.* (1996) and Wiley *et al.* (1991).

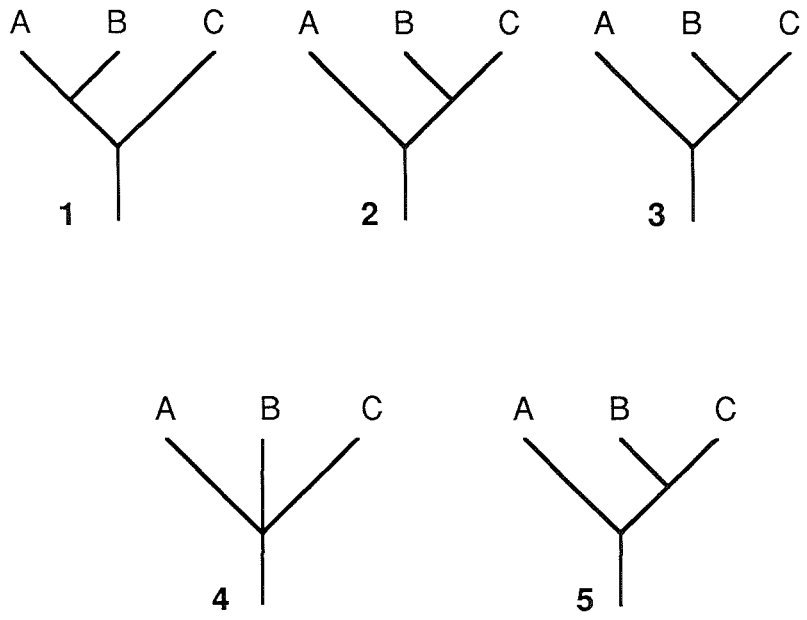


Figure 3.11. Three trees (1,2 and 3) and their strict consensus tree (4) and the majority consensus tree (5) for taxa A, B and C.

3.2.4.5 Inclusion of an outgroup

Another important criterion in phylogenetic analysis is the inclusion of an outgroup. An outgroup is any taxon used in phylogenetic analysis that is assumed to be phylogenetically outside the group of taxa under study. It is used for comparative purposes, usually in determining character polarity and assigning the direction of change of character state transformation, and also for determining the root of a phylogenetic tree (Swofford *et al.*, 1996). An outgroup is often chosen as a sister group in the sense that it is a taxon that is genealogically most closely related to the ingroup, but must not be the ancestor of the ingroup.

3.2.4.6 Measures of the robustness of a cladogram

A number of indices are used to measure the robustness of a cladogram. Kluge and Farris in 1969 introduced what is called the consistency index (CI) which remains the most widely used index today. This index is a measure of how the transformation series and the entire data matrix fits a tree's topology. In simple terms, a transformation series with little or no homoplasy will yield a high CI value (1 is the highest possible value) while those which exhibit high homoplasy have a lower value (0 is the lowest possible value). The consistency index is calculated using only synapomorphies and is the minimum number of changes or steps (m) necessary if all data agreed, divided by the actual number of steps (s) in the tree (i.e. CI = m/s).

Another index used is the retention index (RI) described by Farris in 1989. RI is used to express the amount of synapomorphy in a data set by

examining the actual amount of homoplasy as a function of the maximum possible homoplasy. RI can be thought of as the proportion of similarities in a tree due to synapomorphies. Further details on CI and RI and other indices of robustness of trees can be found in Wiley *et al.* (1991) and Siebert (1992).

3.2.4.7 Reliability of inferred trees

A question that is often asked by systematicists is how much 'confidence' can one place in the topology of a particular tree. According to Siebert (1992), this question is unanswerable as one needs to distinguish between psychological and statistical 'confidence', and in the case of a branching diagram, the topology either correctly reflects historical relationships or does not. However, there are two methods that have been used for attaching confidence limits to the branches of trees; these are the bootstrap method, which is widely used, the jack-knife method and decay indices. Only the bootstrap procedure will be mentioned here as it is the one more commonly used. The bootstrap procedure was introduced by Felsenstein (1985) and involves random sampling with replacement (compared to jack-kniffing which involves random sampling without replacement) either character rows or columns in a data set to build many bootstrap data sets of the same size as the original data set. Each bootstrap data set is then analysed using a heuristic or branch-and-bound search, to give a tree or a set of trees. This procedure of random sampling and tree generation is repeated at least 100 times and the percentage of occurrence of a particular group or component that appears among the bootstrap trees can be considered as an index of support for that group, although it is not a true confidence limit in a statistical sense.

3.2.5 Phylogenetic Analysis of the ITS Region Of *Cryptocoryne*

From the ITS sequences obtained for each accession of *Cryptocoryne* investigated, phylogenetic trees were reconstructed using Fitch parsimony with unweighted characters as implemented in PAUP 3.1.1 (Swofford, 1993). Due to the large number of taxa, heuristic searches were conducted. Under the heuristic search. RANDOM addition sequence procedure (i.e. the first three taxa for the starting tree were chosen at random) was employed using 100 replicates and TBR branch swapping was carried out. All minimal trees were saved and equally parsimonious trees obtained were then summarised using strict consensus.

Support for clades within a tree were determined by bootstrap analysis (Felsenstein, 1985) using 100 replicates of heuristic searches with random addition sequences. A maximum of 1000 trees were saved during each bootstrap replicate to reduce the analysis time.

3.3 Results

3.3.1 ITS structure, size and composition

Within the species of *Cryptocoryne* examined, the ITS1 sequence was found to be consistently shorter than the ITS2 sequence. The same was also true in both *Lagenandra* species investigated. Among *Cryptocoryne* DNAs, the ITS1 varied in length from 201 base pairs (bp) in *C. undulata* to 232 bp in *C. aponogetifolia,* while ITS 2 varied from 231 bp in *C. moehlmannii* to 259 bp in *C. elliptica* and *C. schulzei.* With respect to the 5.8S subunit, most species surveyed contained a sequence of 164 bp, which is consistent in length with most angiosperms; however, some variation was detected within *C. spiralis and C. moehlmanii* containing a 5.8S of 162 bp, while that of *C. pygmaea* contained 165 bp, *C. elliptica* with 166 bp and *C. zonata* with 168 bp. The aligned sequences of the entire ITS1, 5.8S subunit and ITS2 of all *Cryptocoryne* species studied are presented in Table 3.1.

Alignment of the ITS1 sequences of *Cryptocoryne* created one or more gaps at 67 positions or 27% of sites. Of these, 32 were autapomorphic. Similarly the aligned ITS2 sequences created one or more gaps at 81 positions or 27.6% of sites, which 22 were autapomorphic. Inclusion of the outgroup species among the aligned sequences created additional gaps; a single gap was needed for a single position within ITS 1 and 17 positions (one long indel) within the ITS2. In addition, their inclusion created gaps at 13 positions or 7.7% of sites in the 5.8S, of which four were autapomorphic.

Table 3.1. Aligned DNA sequences of the ITS region within the 18-26S nuclear ribosomal DNA of 25 representatives of *Cryptocoryne* and from two *Lagenandra* species. (Note: six races of ITS sequences obtained from three accessions of *C. x willisii* were included in the analysis, one ITS sequence for each *C. x willisii* Kew3790 and *C. x willisii* P1978/5045 accessions and four ITS sequences from *C. x willisii* NJ23-25.

Taxa[a]	Nucleotide sites[b]
	ITS 1 --->
	1 2 3 4 5 6 0 0 0 0 0 0 . . . . . .
1	TC-GTA-GTGA-CTGCG-AAGGATCATTG--TCGTTCC-GACC--AAACGACGGTACACC
2	----------GGACGGATCATTG--TCGTTCC-GACC--AAACGACGGCACACC
3	TCCGTAGGTGAACTGCGGAAGGATCATTG--TCGTTCCCGACC--AAACGACGGCACACC
4	TC-GTAGGTGAACTGCGGAAGGATCATTG--TCGTTCCCGACC--AAACGACGGCACACC
5	TC-GTAGGTGAACTGCGGAAGGATCATTG--TCGTTCCCGACC--AAACGACGGCACACC
6	TC-GTAGGTGAACTGCGGAAGGATCATTG--TCGTTCCCGACC--AAACGACGGCACACC
7	TC-GTAGGTGAACTGCGGAAGGATCATTG--TCGTTCCCGACC--AAACGACGGCACACC
8	TC-GTA-GTGA-CTGCGGAAGGATCATTG--TCGTTCC-GACC--AAACGACGGCACACC
9	TC-GTAGGTGAACTGCGGAAGGATCATTG--TCGTTCC-GACC--AAACGACGGCACACC
10	----------G--TCGTTCC-GACC--AAACGACGGAACACC
11	TC-GTAGGTGAACTGCGGAAGGATCATTG--TCGTTCCCGACC--AAACGACGGCACACC
12	TC-GTAGGTGAACTGCGGAAGGATCATTG--TCGTTCCCGACC--AAACGACGGCACACC
13	TC-GTAGGTGAACTGCGGAAGGATCATTG--TCGTTCC-GACC--AAACGACGGCACACC
14	T--GT--GTGAACTGCGGAAGGATCATTG--TCGTTCCCGACCGACAA-GACGGCACACC
15	TC-GTA-GTGA-CTGCGGAAGGATCAT-G--TCGT-CCCGAC---AAACGGCGGCACAC-
16	TC-GTAGGTGA-CTGCGGAAGGATCATTG--TCGTTCC-GACC--AAACGACGGCACACC
17	TC-GTAGGTGAACTGCGGAAGGATCATTG--TCGTTCC-GACC--AAACGACGGCACACC
18	TCCGTA-GTGAACTGCGGAAGGATCAT-G--TCGTTCC-GACC--AAACGAC-GCACACC
19	TC-GTAGGTGA-CTGCGGAAGGA-CATTG--TCGTTCC-GACC--AAACGACGGCACACC
20	-------GTGAACTGCGGAAGGATCATTG--TCGTTCC-GACC--AAACGACGGCACACC
21	TC-GTAGGTGA-CTGTT-AAGGA-CATTG--TCGTTGGAGAGG--AAACTACGGCACTCC
22	------GGTGAACTGCGGAAGGATCATTG--TCGTTCCCGACC--AAA-GACGGCACACC
23	--------TGA--TGAGGA-GG-TCAT-G--TCGT--CCGACC---AACGACGGCACAC-
24	TC-GTAGGTGAACTGCGGAAGGATCATTGGTTCGTTCCCGAGCGCAAACGACGGCACACC
25	TC-GTAGGTGAACTGCGGAAGGA-CATTG--TCGTTCC-GAGC--AAACGACGGCACACC
26	TC-GTAGGTGAACTGCGGAAGGATCATTG--TCGTTCCCGACC--AAACGACGGCACACC
27	TC-GTAGGTGCACTGCGGAAGGA-CATTG--TCGTTCC-GACC--AAACGACGGTACACC
28	-C-GTAGGTGAACTGCGGAGGGA-CATTG--TCGTTCC-GACC--AAACGACGGTACACC
29	TC-GTAGGTGAACTGCGGAAGGATCATTG--TCGTTCCCGACC--AAACGACGGTACACC
30	---GTA-GTGAACTGCGGAAGGATCATTG--TCGTTCC-GACC--AAACGACG-CACACC
31	--CGTAGGTGAACTGCGGAAGGATCATTG--TCGT-CCCGACC--AAAAGATGGTACACC
32	---GTAGGTGAACTGCGGAAGGATCATTG--TCGT-CCCGACC--AGAATATGGTACACC

Table 3.1. Continued

Taxa[a]	Nucleotide sites[b]
	70 . 80 . 90 . 100 . 110 . 120 .
1	GCG--ACTGTCCGCGCCCTCG---CGACGTCGGAGGGCAC-ACGA-CTCCGCG-CGTCCG
2	GCG-AACTGTCGCGCGCCCCCCGCGGGCGCCGGAGGAAAC---AACCTCCGCG-CGTCCG
3	GCG-AACTGTCCGCGCCGCCC-GCGGGCGCCGGAGGAAAC---GACCTCCGCG-CGTCCG
4	GCG-AACTGTCCGCGCCCCCC-GCGGGCGCCGGAGGAAAC---GACCTCCGCG-CGTCCG
5	GCG-AACTGTCCGCGCCCCCC-GCGGGCGCCGGAGGAAAC---GACCTCCGCG-CGTCCG
6	GCG-AACTGTCCGCGCCGCCC-GCGGGCGCCGGAGGAAAC---GACCTCCGCG-CGTCCG
7	GCG-AACTGTCCGCGCCCCCC-GCGGGCGCCGGAGGAAAC---GACCTCCGCG-CGTCCG
8	GCG-AA-TGTCCGCGCCGCCC-GCGGGCGCCGGAGGAAAC---GACCTCCGCG-CGTCCG
9	GCG-AACTGTCCGCGCCCCC--GCGGGCGCCGGAGGAAAC---GACCTCCGCG-CGTCCG
10	GCG-AACTGTCGCCGCGCCGCCGCGGGCGCCGGAGGAAAC---GACTCGCGCG---TCCG
11	GCG-AACTGTCCGCGCCCCCC-GCGGGCGCCGGAGGAAACA--GACCTCCGCG-CGTCCG
12	GCG-AACCGTCCGCGTCCCCG---CGACGCTGGAGGACAC---GACCTCCGCG-CGTCCG
13	GCG-AACCGTCCGCGTCCCCG---CGACGCTGGAGGACAC---GACCTCCGCG-CGTCCG
14	GCG-AACCGTCCGCGCCCCCC-GCCGACGCCGGAGGACGC---GACCTCCGCG-CGTCCG
15	GCG-AA-CGTCCGCGCCCCC--TCCGGCGCCGGAGGACACGACCTCCTCCGCG-CGTCCG
16	GCG-AA-CGTCCGCGCCCCC--GCCGGCGCCGGAGGACAC---GACCTCCGCG-CGTCCG
17	GCGGAACCGTCCGCGCCCCCC-GCCGGCGCCGGAGGACAC---GACCTCCGCG-CGTCCG
18	GCG-AACCGTCCGCGCCCCCC-GCCGGCGCCGGAGGACAC---GACCTCCGCG-CGTCCG
19	GCG-AACCGTCCGCGCCCCC--GCCGGCGCCGGAGGGCAC---GACCTCCGCG-CGTCCG
20	GCG-AACCGTCCGCGCCCCCC-GCCGGCGCCGGAGG-CAC---GACCTCCGCG-CGTCCG
21	GCG-AACCGTCCGCGCCCCCG-TCCGACGCCGGATGACGC---GACCTCCGCG-CGTCCG
22	GCG-AACCGTCCGCGCCCCCC-GCCGACGCCGGATGACGC---GACCTCCGCG-CGTCCG
23	GCG-AA-CGTC-GCGCCCC---GCCGGCGCGGAGGAACGC---GACCTCCGCG-CGTCCG
24	GCGGAACCGTCCGCGCCCCCCCGCCGGCGCCGGAGGACAC---GACCTCCGCG-CGCCCG
25	GCG-AACCGTCCGCGCCCCCCCGCCGGCGCCGGAGGACAC---AACCTCCGCG-CGTCCG
26	GCG-AACTGTCCGCGCCGCCC-GCCGGCGCCGGAGGACAC---GACCTCCGCG-CGTCCG
27	GCG-AACTGTCCGCGCCCC-TCGCCGACGACGTAGGGCAC-ACGACCCCCGCGGCGTTCG
28	GCG-AACTGTCCGCGCCCCCTCGCCGACGTCGGAGGGCAC-ACGACCCCCGCG-CGTCCG
29	GCG-AACTGTCCGCGCCCC-TCGCCGACGACGTAGGGCAC-ACGACCCCCGCG-CGTTCG
30	GCG-AACCGTCCGCGCCCC-ACGCCGACGCCGGAGGACAC---GACCCCCGCG-CGTCCG
31	GCG-AACTGTCCGCCCCCC---ACCGACGTCGGAGG-----ACGGCCTCCGTG-CTTCCG
32	GCG-AACTGTCCGCCCCC----ACCGACGTCGGAGG-----ACGGCCTCCGTG-CTTCCG

Table 3.1. Continued

Taxa[a]	Nucleotide sites[b]
	1 1 1 1 1 1 3 4 5 6 7 8 0 0 0 0 0 0 . . . . . .
1	AAC--TGTTAACAATCCATCCCC-GGCGCGGCATGCGCCAAGGAACACG-GACACAAAAA
2	GAC--CGCCGACGATCCACCCCCCGGCGCGGAACGCGCCAAGGAACACG-GACACAGAAA
3	GAC--CGCCAACGATCGACCCCCCGGCGCGGCACGCGCCAAGGAACACG-GACACAGAAA
4	GAC--CGCCGACGATCCACCCCC-GGCGCGGCACGCGCCAAGGAACACG-GACACAGAAA
5	GAC--CGCCGACGATCCACCCCC-GGCGCGGCACGCGCCAAGGAACACG-GACACAGAAA
6	GAC--CGCCAACGATCCACCCCCCGGCGCGGCACGCGCCAAGGAACACG-GACACAGAAA
7	GAC--CGCCGACGATCCACCCCC-GGCGCGGCACGCGCCAAGGAACACG-GACACAGAAA
8	GAC--CGCCAACGATCCACCCCCCGGCGCGGCACGCGCCAAGGAACACG-GACACAGAAA
9	GAC--CGCCGACGATCCACCCCCCGGCGCGGCACGCGCCAAGGAACACG-GACACAGAAA
10	GAC--CGCCGACGATCCACCCCCCGGCGCGGCACGCGCCAAGGAACACG-GACACAGAAA
11	GAC--CGCCGACGATCCACCCCC-GGCGCGGCACGCGCCAAGGAACACG-GACACAGAAA
12	AAC--TTTCGACCATCCACCCCC-GGCGCGGCACGCGCCAAGGAACACG-GACACAGAAA
13	AAC--TTTCGACCATCCACCCCC-GGCGCGGCACGCGCCAAGGAACACG-GACACAGAAA
14	GAC--TCCCGACGATCCACCCCC-GGCGCGGCACGCGCCAAGGAACACG-GACACAGAAA
15	AAC--TCTCGACGATCCACCCCC-GGCGCGGCACGCGCCAAGGAACACG-GACACAGAAA
16	AGC--TTTCGACGATCCACCCCC-GGCGCGGCACGCGCCAAGGAACACG-GACACAGAAA
17	AAC--TTTCGACGATCCACCCCC-GGCGCGGCACGCGCCAAGGAACACG-GACACAGGAA
18	GAC--TCTCGACGATCCACCCCC-GGCGCGGCACGCGCCAAGGAACACG-GACACAGAAA
19	GAC--TCTCGACGATCCACCCCC-GGCGCGGCACGCGCCAAGGAACACG-GACACAGAAA
20	GAC--TCTCGACGATCCACCCCC-GGCGCGGCACGCGCCAAGGAACACG-GACACGGAAA
21	AAG--TTTCGACAATCCACCCCC-GGCGCGGCCCGCGCCAGGGAACACG-GACACAGAAA
22	AAGCGTTTCGACAATCCACCCCC-GGCGCGGCACGCGCCAAGGAACACG-GACACAGAAA
23	AGC--TTTCGACGATCCACCC---GGCGCGGCACGCGCCAAGGAACACG-GACACAGAAA
24	GAGC-CTTCGACGATCCACACCC-GGCGCGGCACGCGCCAAGGAACACG-GACACAGAAA
25	GAG--TCTCGACGATCCACCCCC-GGCGCGGCACGCGCCAAGGAACACGCGACACAGAAA
26	GAC--CTTCGACGATCCACCCCC-GGCGCGGCACGCGCCAAGGAACACG-GACGCAGAAA
27	AGC--TGTTAACAATCCACCCCC-GGCGCGGCATGCGCCAAGGAACACG-GACACGAAGA
28	AGC--TATTAACAATCCACCCCC-GGCGCGGCATGCGCCAAGGAACACG-GACACGAAGA
29	AGC--TGTTAACAATCCACCCCC-GGCGCGGCATGCGCCAAGGAACACG-GACACGAAGA
30	GAC--TTTCGACAATCCACCCCC-GGCGCGGCACGCGCCAAGGAACACG-GACACAGAAA
31	GAC-----TAACAATTCATCCCC-GGCGCGGCATGCGCCAAGGAACACG-AACACAAA--
32	GAC-----TAACAATTCATCCCC-GGCGCGGCATGCGCCAAGGAACACG-AACACAAA--

Table 3.1. Continued

Taxa[a]	Nucleotide sites[b]
	1 2 2 2 2 2 9 0 1 2 3 4 0 0 0 0 0 0 . . . . . .
1	CA--CCCACGATCCCGAACCCGCTC-GGGGGACGGGGCGTGGAG--TGCCTCATGTGTGG
2	CGCGCCCACGATCCGGAACCCGATC-GGGGGACGGGGCGTGGGGCGCGCCTCACGTGCAG
3	CGCGCCCACGATCCGGAACCCGATC-GGGGGACGGGGCGTGGGGCGCGCCTCACGTGCAG
4	CGCGCCCACGATCCGGAACCCGATC-GGGGGACGGGGCGTGGGGCGCGCCTCACGTGCAG
5	CGCGCCCACGATCCGGAACCCGATC-GGGGGACGGGGCGTGGGGCGCGCCTCACGTGCAG
6	CGCGCCCACGATCCGGAACCCGATC-GGGGGACGGGGCGTGGGGCGCGCCTCACGTGCAG
7	CGCGCCCACGATCCGGAACCCGATC-GGGGGACGGGGCGTGGGGCGCGCCTCACGTGCAG
8	CGCGCCCACGATCCGGAACCCGATC-GGGGGACGGGGCGTGGGGCGCGCCTCACGTGCAG
9	CGCGCCCACGATCCGGAACCCGATC-GGGGGACGGGGCGTGGGGCGCGCCTCACGTGCAG
10	CGCGCCCACGATCCGGAACCCGATC-GGGGGACGGGGCGTGGGGCGCGCCTCACGTGCAG
11	CGCGCCCACGATCCGGAACCCGATC-GGGGGACGGGGCGTGGGGCGCGCCTCACGTGCAG
12	CG--CCCACGATCCGGAACCCGCAC-GGGGGACGGGGCGTGGCG--CGCCTCACGTGCAG
13	CG--CCCACGATCCGGAACCCGCAC-GGGGGACGGGGCGTGGCG--CGCCTCACGTGCAG
14	CG--CCCACGATCCGGAACCCGCAC-GGGGGACGGGGCGTGGCG--CGCCTCACGTGCAG
15	CG--CCCACGATCCGGAACCCGCGC-GGGCGACGGGGCGTGGCG--CGCCTCACGTGCAG
16	CG--CCCACGATCCGGAACCCGCAC-GGGGGACGGGGCGTGGCG--CGCCTCACGTGCAG
17	CG--CCCACGATCCGGAACCCGCAC-GGGGGACGGGGCGTGGCG--CGCCTCACGTGCAG
18	CG--CCCACGATCCGGAACCCGCAC-GGGGGACGGGTCGTGGCG--CGCCTCACGTGCGG
19	CG--CCCACGATCCGGAACCCGCAC-GGGGGACGGGTCGTGGCG--CGCCTCACTTGCGG
20	CG--CCCACGATCCGGAACCCGCAC-GGGGGACGGGTCGTGGCG--CGCCTCACGTGCGG
21	CG--CCCACAATCCGGAACCCGCAC-GGGGGACGGGGCGTGGCG-GCGCCTCGCGTGCAG
22	CGCGCGCACGATCCGGAACCCGCAC-GGGGGACGGGGCGTGGCG--CGCCTCGCGTGCAC
23	CG---CCACGATCCGGAACCCGCAC--GGGGAC-GGGCGTGGCG--CGCCTCACGTGCAG
24	CG--CCCACGATCCGGAACCCGCAC-GGGGGACGGGGCGTGGCG--CGCCTCGCGTGCAG
25	CG--CCCACGATCCGGAACCCGCAC-GGGGGACGGGGCGTGGCG--CGCCTCACGTGCAG
26	CG--ACCACGATCCGGAACCCGCTC-GGGGGACGGGGCGTGGTG--CGCCTCACGCGCAG
27	TG--CCCACGATCCGAAACCCGCTC-GGGGGACGGGGCGTGGCG--TGCCTCATGTGCGG
28	TG--CCCACGATCCCAAACCCGCTC-GGGGGACGGGGCGTGGCG--CGCCTCATGTGTGG
29	TG--CCCACGATCCGAAACCCGCTC-GGGGGACGGGGCGTGGCG--TGCCTCATGTGCGG
30	-GCGCCCACGATCCGGAACCCGCAC-GGGGGACGGGGCGTGGCG--CGCCTCACGTGCAG
31	--CGCCCACGATCCGAAACCCGCTCCGGGGGGACGGGGTGTGGCGCACCTCATATGTGTGG
32	--CGCCCACGATCCGAAACCCGCTCGGGGGGGACGGGGTGTGGCGCACCTCA--TGTGTGG

Table 3.1. Continued

Taxa[a]	Nucleotide sites[b]

```
                5.8 S
                --->
                2         2         2         2         2         3
                5         6         7         8         9         0
                0         0         0         0         0         0
                .         .         .         .         .         .

1       AGA---TCGAA-ACGACTCCCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAA---
2       AGC---TCGAA-ACGACTCCCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAA---
3       AGC---TCGAC-ACGACTCCCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAA---
4       GGC---TCGAA-ACGACTCCCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAA---
5       AGC---TCGAA-ACGACTCCCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAA---
6       AGC---TCGAA-ACGACTCCCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAA---
7       GGC---TCGAA-ACGACTCCCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAA---
8       AGC---TCGAA-ACGACTCCCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAA---
9       AGC---TCGAA-ACGACTCCCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAA---
10      AGC---TCGAA-ACGACTCCCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAA---
11      AGC---TCGAA-ACGACTCCCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAA---
12      AGC---TCGAA-ACGACTCCCGGCAACGGATATCTAGGCTCTCGAATCGATGAAGAA---
13      AGC---TCGAA-ACGACTCCCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAA---
14      AGC---TCGAAAACGACTCCCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAA---
15      AGC---TCGAA-ACGACTCCCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAA---
16      AGC---TCGAA-ACGACTCCCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAA---
17      AGC---TCGAA-ACGACTCCCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAA---
18      AGC---TCGAA-ACGACTCCCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAA---
19      AGC---TCGAA-ACGACTCCCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAA---
20      AGC---TCGAA-ACGACTCCCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAAGGA
21      AGTTTTTCGAACACGACTCCCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAA---
22      AGTTTTTCGAA-ACGACTCCCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAA---
23      AGC---TCGAA-ACGACTCCCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAA---
24      AGC---TCGAA-ACGACTCCCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAA---
25      AGC---TCGAA-ACGACTCCCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAA---
26      AGC---TCGAA-ACGACTCCCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAA---
27      AGA---TCGAA-ACGACTCCCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAA---
28      AGA---TCGAA-ACGACTCCCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAA---
29      AGA---TCGAA-ACGACTCCCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAA---
30      AGC---TCGAA-ACGACTCCCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAA---
31      AGC---TCGAA-ACGACTCCCGGCAACGGATATCTAGGCTCTTGCATCGATGAAGAA---
32      AGC---TCGAA-ACGACTCCCGGCAACGGATATCTAGGCTCTTGCATCGATGAAGAA---
```

Table 3.1. Continued

Taxa[a]	Nucleotide sites[b]
	3 3 3 3 3 3 1 2 3 4 5 6 0 0 0 0 0 0 . . . . . .
1	-CGTAGCGAAATGCGATAACGTGGTGTGAATTGCAGAATCCCGTGAACCATCGAATCTTT
2	-CGTAGCGAAATGCGATA-CGTGGTGTGAATTGCAGAATCCCGTGAACCATCGAATCTTT
3	-CGTAGCGAAATGCGATA-CGTGGTGTGAATTGCAGAATCCCGTGAACCATCGAATCTTT
4	-CGTAGCGAAATGCGATA-CGTGGTGTGAATTGCAGAATCCCGTGAACCATCGAATCTTT
5	-CGTAGCGAAATGCGATA-CGTGGTGTGAATTGCAGAATCCCGTGAACCATCGAATCTTT
6	-CGTAGCGAAATGCGATA-CGTGGTGTGAATTGCAGAATCCCGTGAACCATCGAATCTTT
7	-CGTAGCGAAATGCGATA-CGTGGTGTGAATTGCAGAATCCCGTGAACCATCGAATCTTT
8	-CGTAGCGAAATGCGATA-CGTGGTGTGAATTGCAGAATCCCGTGAACCATCGAATCTTT
9	-CGTAGCGAAATGCGATA-CGTGGTGTGAATTGCAGAATCCCGTGAACCATCGAATCTTT
10	-CGTAGCGAAATGCGATA-CGTGGTGTGAATTGCAGAATCCCGTGAACCATCGAATCTTT
11	-CGTAGCGAAATGCGATA-CGTGGTGTGAATTGCAGAATCCCGTGAACCATCGAATCTTT
12	-CGTAGCGAAATGCGATA-CGTGGTGTGAATTGCAGAATCCCGCGAACCATCGAATCTTT
13	-CGTAGCGAA-TGCGATA-CGTGGTGTGAATTGCAGA-TCCCGCGAACCATCGAATCTTT
14	-CGTAGCGAAATGCGATA-CGTGGTGTGAATTGCAGAATCCCGTGAACCATCGAATCTTT
15	-CGTAGCGAAATGCGATA-CGTGGTGTGAATTGCAGAATCCCGTGAACCATCGAATCTTT
16	-CGTAGCGACATGCGATA-CGTGGTGTGAATTGCAGAATCCCGTGAACCATCGAATCTTT
17	-CGTAGCGAAATGCGATA-CGTGGTGTGAATTGCAGAATCCCGTGAACCATCGAATCTTT
18	-CGTAGCGAAATGCGATA-CGTGGTGTGAATTGCAGAATCCCGTGAACCATCGAATCTTT
19	-CGTAGCGAAATGCGATA-CGTGGTGTGAATTGCAGAATCCCGTGAACCATCGAATCTTT
20	ACGTAGCGAAATGCGATA-CGTGGTGTGAATTGCAGAATCCCGTGAACCATCGAATCTTT
21	-CGTAGCGAAATGCGATA-CGTGGTGTGAATTGCAGAATCCCGCGAACCATCGAATCTTT
22	-CGTAGCGAAATGCGATA-CGTGGTGTGAATTGCAGAATCCCGCGAACCATCGAATCTTT
23	-CGTAGCGAAATGCGATA-CGTGGTGTGAATTGCAGAATCCCGTGAACCATCGAATCTT-
24	-CGTAGCGAAATGCGATA-CGTGGTGTGAATTGCAGAATCCCGCGAACCATCGAATCTTT
25	-CGTAGCGAAATGCGATA-CGTGGTGTGAATTGCAGAATCCCGTGAACCATCGAATCTTT
26	-CGTAGCGAAATGCGATA-CGTGGTGTGAATTGCAGAATCCCGTGAACCATCGAATCTTT
27	-CGTAGCGAAATGCGATA-CGTGGTGTGAATTGCAGAATCCCGTGAACCATCGAATCTTT
28	-CGTAGCGAAATGCGATA-CGTGGTGTGAATTGCAGAATCCCGTGAACCATCGAATCTTT
29	-CGTAGCGAAATGCGATA-CGTGGTGTGAATTGCAGAATCCCGTGAACCATCGAATCTTT
30	-CGTAGCGAAATGCGATA-CGTGGTGTGAATTGCAGAATCCCGTGAACCATCGAATCTTT
31	-CGTAGCGAAATGCGATA-CGTGGTGTGAATTGCAGAATCCCGTGAACCATCGAATCTTT
32	-CGTAGCGAAATGCGATA-CGTGGTGTGAATTGCAGAATCCCGTGAACCATCGAATCTTT

Table 3.1. Continued

Taxa[a]	Nucleotide sites[b]					
						ITS2 ->
	370	380	390	400	410	420
1	GAACGCAAGTTGCGCCCGAGGCCGTCGGGTCGAGG--CACGC-TGCCTGG-CGTCACACC					
2	GAACGCAAGTTGCGCCCGAGGCCGTCGGGCCGAGG-GCACGCCTGCCTGGGCGTCACGCC					
3	GAACGCAAGTTGCGCCCGAGGCCGTCGGGCCGAGG-GCACGCCTGCCTGGGCGTCACGCC					
4	GAACGCAAGTTGCGCCCGAGGCCGTCGGGCCGAGG-GCACGCCTGCCTGGGCGTCACGCC					
5	GAACGCAAGTTGCGCCCGAGGCCGTCGGGCCGAGG-GCACGCCTGCCTGGGCGTCACGCC					
6	GAACGCAAGTTGCGCCCGAGGCCGTCGGGCCGAGG-GCACGCCTGCCTGGGCGTCACGCC					
7	GAACGCAAGTTGCGCCCGAGGCCGTCGGGCCGAGG-GCACGCCTGCCTGGGCGTCACGCC					
8	GAACGCAAGTTGCGCCCGAGGCCGTCGGGCCGAGG-GCACGCCTGCCTGGGCGTCACGCC					
9	GAACGCAAGTTGCGCCCGAGGCCGTCGGGCCGAGG-GCACGCCTGCCTGGGCGTCACGCC					
10	GAACGCAAGTTGCGCCCGAGGCCGTCGGGCCGAGG-GCACGCCTGCCTGGGCGTCACGCC					
11	GAACGCAAGTTGCGCCCGAGGCCGTCGGGCCGAGG-GCACGCCTGCCTGGGCGTCACGCC					
12	GAACGCAAGTTGCGCCCGAGGCAGTCGGGCCGAGG-GCACGCCTGCCTGGGCGTCACGCA					
13	GAACGCAAGTTGCGCCCGAGGCCGTCGGGCCGAGG--CACGC-TGCCTGGGCGTCACGCC					
14	GAACGCAAGTTGCGCCCGAGGCCGTCGGGCCGAGG-GCACGCCTGCCTGGGCGTCACGCC					
15	GAACGCAAGTTGCGCCCGAGGCCGTCGGGCCGAGG-GCACGCCTGCCTGGGCGTCACGCC					
16	GAACGCAAGTTGCGCCCGAGGCCGTCGGGCCGAGG-GCACGCCTGCCTGGGCGTCACGCC					
17	GAACGCAAGTTGCGCCCGAGGCCGTCGGGCCGAGG-GCACGCCTGCCTGGGCGTCACGCC					
18	GAACGCAAGTTGCGCCCGAGGCCGTCGGGCCGAGG-GCACGCCTGCCTGGGCGTCACGCC					
19	GAACGCAAGTTGCGCCCGAGGCCGTCGGGCCGAGG-GCACGCCTGCCTGGGCGTCACGCC					
20	GAACGCAAGTTGCGCCCGAGGCCGTCGGGCCGAGG-GCACGCCTGCCTGGGCGTCACGCC					
21	GAACGCAAGTTGCGCCCGAGGCCGTCGGGCCGAGGAGCACGCCTGCCTGGGCGTCACGCC					
22	GAACGCAAGTTGCGCCCGAGGCCGTCGGGCCGAGG-GCACGCCTGCCTGGGCGTCACGCC					
23	GAACGCAAGTTGCGCCCGAGGCCGTCGGGCCGAGG-GCACGCCTGCCTGGGCGTCACGCC					
24	GAACGCAAGTTGCGCCCGAGGCCGTCGGGCCGAGG-GCACGCCTGCCTGGGCGTCACGCC					
25	GAACGCAAGTTGCGCCCGAGGCCGTCGGGCCGAGG-GCACGCCTGCCTGGGCGTCACGCC					
26	GAACGCAAGTTGCGCCCGAGGCCGTCGGGCCGAGG-GCACGCCTGCCTGGGCGTCACGCC					
27	GAACGCAAGTTGCGCCCGAGGCCGTCGGGTCGAGG-GCACGCCTGCCTGGGCGTCACACC					
28	GAACGCAAGTTGCGCCCGAGGCCGTCGGGTCGAGG-GCACGCCTGCCTGGGCGTCACACC					
29	GAACGCAAGTTGCGCCCGAGGCCGTCGGGTCGAGG-GCACGCCTGCCTGGGCGTCACACC					
30	GAACGCAAGTTGCGCCCGAGGCCGTCGGGCCGAGG-GCACGCCTGCCTGGGCGTCACGCC					
31	GAACGCAAGTTGCGCCCGAGGCCGTCGGGCCGAGG-GCACGCCTGCCTGGGCGTCACACC					
32	GAACGCAAGTTGCGCCCGAGGCCGTCGGGCCGAGG-GCACGCCTGCCTGGGCGTCACACC					

Table 3.1. Continued

Taxa[a]	Nucleotide sites[b]
	4 4 4 4 4 4 3 4 5 6 7 8 0 0 0 0 0 0 . . . . . .
1	TGTGTCGCTCCCC---GCACGC--CCGG----------CGTGTCGC-GGA-TG
2	TGTGTCGCTCCCCC--GCACGC-CTCGT----------CGTGCCGC-GGA-TG
3	TGTGTCGCTCCCCC--GCACGC-CTCGT----------CGTGCCGCGGGA-TG
4	TGTGTCGCTCCCCC--GCACGC-CTCGT----------CGTGCCGCGGGA-TG
5	TGTGTCGCTCCCCC--GCACGC-CTCGT----------CGTGCCGCGGGA-TG
6	TGTGTCGCTCCCCC--GCACGC-CTCGT----------CGTGCCGCGGGA-TG
7	TGTGTCGCTCCCCC--GCACGC-CTCGT----------CGTGCCGCGGGA-TG
8	TGTGTCGCTCCCCC--GCACGC-CTCGT----------CGTGCCGCGGGA-TG
9	TGTGTCGCTCCCCC--GCACGC-CTCGT----------CGTGCCGCGGGA-TG
10	TGTGTCGCTCCCCC--GCACGC-CTCGT----------CGTGCCGCGGGA-TG
11	TGTGTCGCTCCCCC--GCACGC-CTCGT----------CGTGCCGCGGGA-TG
12	TGTGTCGCTCCCCC--GCACGC-CTCGG----------CGTGTCGCGGGA-TG
13	TGTGTCGCTCCCC---GCACGC-CTCGG----------CGTGTCGCGGGA-TG
14	CGTGTCGCTCCCCC--GCACGC-CCCGG----------CGTGCCGCGGGA-AG
15	CGTGTCGCTCCCCCCCCGCACGC-CTCGG----------CGTGCCGCGGGGGAG
16	TGTGTCGCTCCCCC--GCACGC-CTCGG----------CGTGCCGCGGGA-AG
17	TGTGTCGCTCCCCC--GCACGC-CTCGG----------CGTGCCGCGGGA-AG
18	CGTGTCGCTCCCCC--GCACGC-CTCGG----------CGTGTCGCGGGA-AG
19	CGTGTCGCTCCCCC--GCACGC-CTCGG----------CGTGTCGCGGGA-AG
20	CGTGTCGCTCCCCC--GCACGC-CTCTG----------CGTGTCGCGGGA-AG
21	CGTGTCGCTCCCCC--GCACGCTCTCGG----------CGTGCCGCGGGA-AG
22	CGTGTCGCTCCCCC--GCACGC-CTCGG----------CGTGCCGCGGGA-AG
23	TGTGTCGCTCCCCC--GCACGC-CTCGG----------CGTGTCGCGGGA-AG
24	TGTGTCGCTCCCCC--GCACGC-CTCGG----------CGTGTCGCGGGA-TG
25	TGTGTCGCTCCCCC--GCACGC-CTCGG----------CGTGTCGCGGGA-TG
26	TGTGTCGCTCCCCC--GCACGC-CTCGG----------CGTGCCGCGGGA-TG
27	TGTGTCGCTCCCCC--GCACGC-CCCGG----------CGTGTGGCGGGA-CG
28	TGTGTCGCTGCCCC--GCACGC-CCCGG----------CGTGTGGCGGGA-TG
29	TGTGTCGCTCCCCC--GCACGC-CCCGG----------CGTGTGGCGGGA-CG
30	TGTGTCGCTCCCCC--GCACGC-CTCGG----------CGTGTCGGGGGA-CG
31	TATGTCGCTCCCCT--ACACGC-CTCGGTCGACCGTCGGCGCGGGCGTGTCGCGGGA-TG
32	TATGTCGCTCCCCT--ACACGC-CTCGGCCGACCGTCGGCGCGGGCGTGTCGCGGGA-TG

Table 3.1. Continued

Taxa[a]	Nucleotide sites[b]
	4 5 5 5 5 5 9 0 1 2 3 4 0 0 0 0 0 0 . . . . . .
1	CGGAGA-TCTGGCCTACCGTG-TC-CTCCGGCACGGCCGGCCGAAGCGCTCG---ACCC-
2	CGGAGA-TCTGGCC-ACCGTGCTC-CTCCGGCACGGCGGGCCGAAGCGCTCG---ACCC-
3	CGGAGA-TCTGGCCCACCGTGCTC-CTCCGGCACGGCGGGCCGAAGTGCTCG---ACCCC
4	CGGAGA-TCTGGCCCACCGTGCTC-CTCCGGCACGGCGGGCCGAAGCGCTCG---ACCCC
5	CGGAGA-TCTGGCCCACCGTGCTC-CTCCGGCACGGCGGGCCGAAGCGCTCG---ACCCC
6	CGGAGA-TCTGGCCCACCGTGCTC-CTCCGGCACGGCGGGCCGAAGTGCTCG---ACCCC
7	CGGAGA-TCTGGCCCACCGTGCTC-CTCCGGCACGGCGGGCCGAAGCGCTCG---ACCCC
8	CGGAGA-TCTGGCCCACCGTGCTC-CTCCGGCACGGCGGGCCGAAGCGCTCG---ACCC-
9	CGGAGA-TCTGGCCCACCGTGCTC-CTCCGGCACGGCGGGCCGAAGCGCTCG---ACCCC
10	CGGAGA-TCTGGCCCACCGTGCTC-CTCCGGCACGGCGGGCCGAAGTGCTCG---ACCCC
11	CGGAGA-TCTGGCCCACCGTGCTC-CTCCGGCACGGCGGGCCGAAGCGCTCG---ACCCC
12	CGGAGA-TCTGGCCCACCGTGTTC-CTCCGGCACGGCGGGCCGAAGCGCTCG---ACCCC
13	CGGAGA-TCTGGCC-ACCGTGT---CTCCGGCACGGC-GGCCGAAGCGCTCG---ACCC-
14	CGGAGA-TCTGGCCCACCGTGTTC-CTCCGGCACGGCGGGCCGAAGCGCTCG---ACCCC
15	CGGAGA-TCTGGCCCACCGTGTCG-CTCCGGCACGGCGGGCCGAAGCGCTGG---ACCC-
16	CGGAGA-TCTGGCCCACCGTGTTC-CTCCGGCACGGCGGGCCGAAGCGCTCG---ACCC-
17	CGGAGA-TCTGGCCCACCGTGTTC-CTCCGGCACGGCGGGCCGAAGCGCTCG---ACCCC
18	CGGAGA-TCTGGCCCACCGCGTTTCCTCCGGCACGGCGGGCCGAAGCGCTCG---ACCCC
19	CGGAGA-TCTGGCCCACCGCGTTC-CTCCGGCACGGCGGGCCGAAGCGCTCG---ACCCC
20	CGGAGA-TCTGGCCCACCGCGTTC-CTCCGGCACGGCGGGCCGAAGCGCTCG---ACCCC
21	CGGAGA-TCTGGCCCACCGTGTTCTCTCCGGCACGGCGGGCCGAAGCGCTCGACGACCCC
22	CGGAGA-TCTGGCCCACCGTGTTC-CTCCGGCACGGCGGGCCGAAGCGCTCGACGACCCC
23	CGGAGA-TCTGGCCCACCGTGTCC-CTCCGGCACGGCGGGCCGAAGCGCCCG---GCCCC
24	CGGAGA-TCTGGCCCACCGTGTTCCCTCCGGCACGGCGGGCCGAAGCGCTCG---ACCCC
25	CGGAGA-TCTGGCCCACCGTG-CCCCTCCGGCGCGGCGGGCCGAAGCGCTCG---ACCCC
26	CGGAGATTCTGGCCGACCGTGATC-CTCCGGCACGGCGGGCCGAAGCGCACG---ACCCC
27	CGGAGT-TGTGGCCTACCGTGTTC-CTCTGGGACGGTGGGCCGAAGCGCTCG---ACCCC
28	CGGAGT-TCTGGCCCACCGTGTTC-CTCCGGGACGGCGGGCCGAAGCGTTCG---ACCCC
29	CGGAGT-TGTGGCCTACCGTGTTC-CTCTGGGACGGTGGGCCGAAGCGCTCG---ACCCC
30	CGGAGA-TCTGGCCCACCGTGTTCCCTCCGGCACGGCGGGCCGAAGCGCGCG---ACCCC
31	CGGAGA-TCTGGCCTACCGTGTTC-CTCCGGCACGGCGGGCTGAAGCGCTCG---ACCC-
32	CGGAGA-TCTGGCCTACCGTGTTC-CTCCGGCACGGCGGGCTGAAGCGCTCG---ACCC-

Table 3.1. Continued

Taxa[a]	Nucleotide sites[b]
	550 560 570 580 590 600
1	TGCCGTAC-GTGTTGCG---GGCGCGACCGTGCGGCGAG-TGGTGGAC---GAAAGCTC-
2	TGCCGTGC-GTGTCGCGGG-CGCGC-ACGTTGCGGCGAG-TGGTGGAC---GAGAGCTC-
3	TGCCGTGC-GTGTCGCGGG-GGCGCGACGGTGCGGCGAG-TGGTGGAC---GAAAGCTC-
4	TGCCGTGC-GTGTCGCGGG-GGCGCGACGGTGCGGCGAG-TGGTGGAC---GAAAGCTC-
5	TGCCGTGC-GTGTCGCGGG-GGCGCGACGGTGCGGCGAG-TGGTGGAC---GAAAGCTC-
6	TGCCGTGC-GTGTCGCGGG-GGCGCGACGTTGCGGCGAG-TGGTGGAC---GAAAGCTC-
7	TGCCGTGC-GTGTCGCGGG-GGCGCGACGGTGCGGCGAG-TGGTGGAC---GAAAGCTC-
8	TGCCGTGC-GTGTCGCG---GGCGCGACGTTGCGGCGAG-TGGTGGAC----GAAGCTC-
9	TGCCGTGC-GTGTCGCGGG-GGCGCGACGTTGCGGCGAG-TGGTGGAC---GAGAGCTC-
10	TGCCGTGC-GTGTCGCGGG-GGCGCGACGTTGCGGCGAG-TGGTGGAC---GAAAGCTC-
11	TGCCGTGC-GTGTCGCGGG-GGCGCGACGGTGCGGCGAG-TGGTGGAC---GAAAGCTC-
12	TGCCGTAC-GTGTCGCGGG-GGCGCGACGGTGCGGCGAG-TGGTGGAC---CGAAGCTC-
13	TGCCGTAC-GTGTCGCG---GGCGCGACGGTGCGGCGAG-TGGTGGAC---CGAAGCTC-
14	TGCCGTAC-GTCTCGCGGG-GGCGCGACGGTGCGGCGAG-TGGTGGAC---CGAAGCTC-
15	TGTCGTAC-GTGTCGCGGG-GGCGCGACGGTGCGGCGAG-TGGTGGAC---CGAAGCTC-
16	TGCCGTAC-GTGTCGCGGG-GGCGCGACGGTGCGGCGAG-TGGTGGAC---CGAAGCTC-
17	TGCCGTAC-GTGTCGCGGG-GGCGCGACGGTGCGGCGAG-TGGTGGAC---CGAAGCTC-
18	TGCCGTAC-GTGTCGCGGG-GGCGCGACGGTGCGGCGAG-TGGTGGAC---CGAAGCTC-
19	TGCCGTAC-GTGTC-CGCG-GGCGCGACGGTGCGGCGAG-TGGTGGAC---CGAAGCTC-
20	TGCCGTAC-GTGTCGCGGG-GGCGCGACGGTGCGGCGAG-TGGTGGAC---CGAAGCTC-
21	TGCCGTACAGTGCCGCGGGGCGCGCGACGGTGCGGCGAG-TGGTGGAC---CGAAGCTC-
22	TGCCGTAC-GTGCCGCGGGGGCGACGACGGTGCGGCGAG-TGGTGGAC--ACGAAGCTC-
23	TGCCGTAC-GTGTCGCGGG-GGCGCGACGGTGCGGCGAG-TGGTGGAC---CGAAGCTC-
24	TGCCGTAACGTGTTCGCGG-GGCGCGACGGTGCGGCGAG-TGGTGGAC---CGAAGCTCT
25	TGCCGTAC-GTGTCGCGGG-GGCGCGACG-TGCGGCGAG-TG-TGGAC---CGAAGATC-
26	TGCCGTAC-GTGTCGCGGG-GGCGCGACGGTGCGGCGAG-TG-TGGTGGACGAAAGCTC-
27	TGCCGTAC-ATCTTGCGGC-GGCGCGACGGTGCGGCGAGGTGGTGGAC---GAAAGCTC-
28	TGCCGTAC-GTCTTGCGGC-GGCGCGATGGTGCGGCGAG-TGGTGGAC---GAAAGCTC-
29	TGCCGTAC-ATCTTGCGGC-GGCGCGACGGTGCGGCGAG-TGGTGGAC---GAAAGCTC-
30	TGCCGTAC-GTGTCGCG---GGCGCGACGGTGCGGCGAG-TGGTGGAC---CGAAGCTC-
31	TGCCGAAC-GTGTCGCGGT-GGCGCGACGGTGCGGCGAG-TGGTGGAC---GAAAGCTC-
32	TGCCGAAC-GTGTCGCGGT-GGCGCGACGGTGCGGCGAG-TGGTGGAC---GAAAGCCC-

Table 3.1. Continued

Taxa[a]	Nucleotide sites[b]
	610 620 630 640 650 660
1	-GAT-CACCTCGTCGCGTCGCGAAC-TTGCCGTAAGGCGGTGAGG---------TA-CAG
2	-GAT-CACCTCGTCGCGTCGCGAACCGCGCCGCAGGGCG--AGGAGGG---CCGTA-CAG
3	-GAT-CACCTCGTCGCGTCGCGAACCGCGCCGCAGGGCG--AGGAGGGGG-CCGTA-CAA
4	-GAT-CACCTCGTCGCGTCGCGAACCGCGCCGCAGGGCG--AGGAGGGGG-CCGTA-CGG
5	-GAT-CACCTCGTCGCGTCGCGAACCGCGCCGCAGGGCG--AGGAGGGGG-CCGTA-CGG
6	-GAT-CACCTCGTCGCGTCGCGAACCGCGCCGCAGGGCG--AGGAGGGGG-CCGTA-CAG
7	-GAT-CACCTCGTCGCGTCGCGAACCGCGCCGCAGGGCG--AGGAGGGGG-CCGTA-CGG
8	-GAT-CACCTCGTCGCGTCGCGAACCGCGCCGCAGGGCG--AGGAGGG---CCGTA-CAG
9	-GAT-CACCTCGTCGCGTCGCGAACCGCGCCGCAGGGCG--AGAAGGG---CCGTA-CAG
10	-GAT-CACCTCGTCGCGTCGCGAACCGCGCCGCAGGGCG--AGGAGGGGG-CCGTA-CAG
11	-GAT-CACCTCGTCGCGTCGCGAACCGCGCCGCAGGGCG--AGGAGGGGG-CCGTA-CGG
12	-GAT-CACCTCGTCGCGTCGCGAACCGCGCC------CGGCGA--GGGG-CCCGTA-CAA
13	-GAT-CA-CTCGTCGCGTCGCGAACCGCG-CGTAGG-CGGCGA--GGG---CCGTA-CAA
14	-GAT-CACCTCGTCGCGTCGCGAACCGCGCCGTAGGCCGGCGAGAGGGGG-CCGTA-CGA
15	-GAT-CACCTCGTCGCGTCGCGAACCGCGCCGTAGGGCGGCGA--GGG---CCGTA-CGA
16	-GAT-CACCTCGTCGCGTCGCGAACCGCGCCGTAGG-CGGCGAGAGGGG--CCGTA-CGA
17	-GAT-CACCTCGTCGCGTCGCGAACCGCGCCGTAGGGCGGAGAGAGGGG--CCGTA-CGA
18	-GAT-CACCTCGTCGCGTCGCGAACCGCGCCGTAGGGCGGCGAGAGGGGG-CCGTA-CAA
19	-GAT-CAC-TCGTCGCGTCGCGAACCGCGCCGTAGG-CGGCGAGAGGGG--CCGTA-CAA
20	-GAT-CACCTCGTCGCGTCGCGAACCGCGCCGTAGGGCGGCGAGAGGGGG-CCGTA-CAA
21	-GAT-CACCTCGTCGCGTCGCGAACCGCGCCGTAGGGCGGCGAGAGGGGG-CCGTATCAA
22	-GATACACCTCGTCGCGTCGCGAACCGCGCCGTAGGGCGGCGAGAGGGGGCCCGTA-CAA
23	-GAT-CACCTCGTCGCGTCGCGAACCGCGCCGTAGGGCGGCGAGAGGGGG-CCGTA-CAA
24	CGAT-CACCTCGTCGCGTCGCGAACCGCGCCGTAGG-CGGCGA--GGGG--CCGTA-CAA
25	-GAT-CACCTCGTCGCGTCGCGAACCGCGCCGTAGG-CGGCGACGGCCT--------CAC
26	-GAT-CACCTCGTCGCGTCGCGAACCGCGCCGTAGGGCGGTGATGGGG---CCGTA-CAA
27	-GGC-CAACTCGTCGCGTCGCGAACCGTGCCGTAAGGCGGTGC--GGG------TA-CAG
28	-GAT-CAACTCGTCGCGTCGCGAACCGTGCCGTAAGGCGGTGA--GGG------TA-CAG
29	-GG--CAACTCGTCGCGTCGCGAACCGTGCCGTAAGGCGGTG-GG---------TA-CAG
30	-GAT-CACCTCGTCGCGTCGCGAACCGCGCCGTAGGGCGGCGA--GGGG--CCG---CAC
31	-GAT-CATCTCGTCGCGTCGCGAACCGCGCCGT-GGGCGGTGAGAGGG------TA--CA
32	-GAT-CATCTCGTCGCGTCGCGAACCGTGCCGT-GGGCGGTGA--GGG------TA--CA

Table 3.1. Continued

Taxa[a]	Nucleotide sites[b]				
	6	6	6	7	7
	7	8	9	0	1
	0	0	0	0	0
	.	.	.	.	.
1	T-CAA--CCCACTCGCGGG--AGCGG-ATCAGACGATG-TCG-TCGCTGCCG				
2	AAGAA---CCGCTCCCGGG----CCGGATCCGACGACG-ACG-CCGCTCC-G				
3	GGGAA-ACCCGCTCACGGGGGGCGCGGATCCGACGACGGACGGCCGCTCCCG				
4	AAGAA-ACCCGCTCACGGGGGGCGCGGATCCGACGACGGACGGCCGCTCCCG				
5	AAGAA-ACCCGCTCACGGGGGGCGCGGATCCGACGACGGACGGCCGCTCCCG				
6	AAGAA-ACCCGCTCCCGGGGGCCGCGGATCCGACGACGGACGGCCGCTCCCG				
7	AAGAA-ACCCGCTCACGGGGGGCGCGGATCCGACGACGGACGGCCGCTCCCG				
8	AAGAA---CCGCTCCCGGG----CCGGATCCGACGACGGACG-CCGCTCCCG				
9	AAGAA---CCGCTCCCGGG----CCGGATCCGACGACG-ACG-CCGCTCC-G				
10	AAGAA-ACCCGCTCCCGGGGG--CCGGATCCGACGACGGACGGCCGCTCCCG				
11	AAGAA-ACCCGCTCACGGGG--CGCGGATCCGACGACGGACGGCCGCTCCCG				
12	GAAA---CCCACCCGCGGG---AGCGGATCCGAAGACGGACG-CCGCTCCCG				
13	GAA-----CCACCCGCGGG---AGCGGATCCGACGACGGACG-CCGCTCC-G				
14	GAGAGAACCCACCCGCGGG---AGCGGATCCGACGACGGACGGCCGCTCCCG				
15	GAA-----CCACC-GCGGG---AGCGGATCCGACGACGGACG-CCGCTGCCG				
16	GA------CCACCCGCGGG---AGCGGATC-GACGACGGACG-CCGCTCCCG				
17	GAA----CCCACCCGCGGG---AGCGGATCAGA-GA-GGACG-CCGCTCCCG				
18	GAA---ACCCACCCGCGGG---GGCGGATCCGACGACGGACGGCCGCTCCCG				
19	GAA-----CCACCCGCGGGG--GGCGGATCCGACGAC-GACGGCCGCTCCCG				
20	GAA---ACCCACCCGCGGGG--AGCGGATCCGACGACGGACGGCCGCTCCCG				
21	GAA---ACCCACCCGCGGGG--AGCGGATCCGACGACGGACGGACGCTCCCG				
22	GAA---ACCCACCCGCGGGG--AGCGGATCCGACGACGGACGGACGCTCCCG				
23	GAA---ACCCACCCGCGGGG--GGCGGATCCGACGAC-------CGCTCCCG				
24	GAA-GAACCCACCCGCGGG---AGCGGATCTGACGACGGACG-CCGCTCCCG				
25	GAGAAGAACCACCCGCGGG---CGGGGATC-GACGACGGACGGCCGCTCCCG				
26	GAA---ACCCGTCTCGCGGG--GGCGGATCCGACGACGGACGGCCGCTCCCG				
27	T-CAA--CCCACTCGCGGG---AGCGGATCCGACGACGGACGGTCGCTCCCG				
28	T-CAA--CCCACTCGCGGG---AGCGGATCCGACGATGGACG-TCGCTCCCG				
29	T-CAA--CCCACTCGCGGG---AGCGGATCCGACGACGGACG-TCG-TGCCG				
30	GAGAAGAACCACC-GCGGG---AGCGGATCCGACGACGGACG-CCGCTCCCG				
31	AGCA--ACCCACTCACGGG---AGCAGATCCGACGACGGACGGCCGCTCCCG				
32	AGCA--ACCCACTCGCGGG---AGCGGATCCGACGACGGACG-CCGCTCCCG				

Note. Vertical columns are nucleotide positions. Horizontal rows are individual DNA sequences. Positions are numbered consecutively from 1 to 712 (5' to 3') beginning at 18S/ITS 1 border and ending at ITS 2/26S subunit border. Arrows mark the borders of the ITS 1, 5.8S subunit and ITS 2. a; 1. *Cryptocoryne spiralis* NJ3129a2, 2. *C. beckettii* 1671, 3. *C. x willisii* P1978/5045, 4. *C. x willisii* NJ23-25/1, 5. *C. x willisii* NJ23-25/2, 6. *C. x willisii* NJ23-25/3, 7. *C. x willisii* NJ23-25/4, 8. *C. walkeri* NJ23-3, 9. *C. wendtii* P1961/5342, 10. *C. undulata* NJ22-7, 11. *C. x willisii* Kew3790, 12. *C. pontederiifolia* USM9638, 13. *C. moehlmannii* P19,89/5046, 14. *C. pygmaea* NJ3962, 15. *C. longicauda* USM9439, 16. *C. minima* S1995/9201, 17. *C. griffithii* NJ85-30, 18. *C. purpurea* Othman s.n., 19. *C. cordata* USM9139, 20. *C. zonata* W534, 21. *C. elliptica* USM8069, 22. *C. schulzei* USM8087, 23. *C. affinis* USM8065, 24. *C. aponogetifolia* P3401, 25. *C. annamica* M92/3205, 26. *C. alba* NJ3172-6, 27. *C. albida* P1958/5363, 28. *C. retrospiralis* P1977/5146, 29. *C. crispatula* var. *balansae* NJ3406, 30. *C. ciliata* P1958/6013, 31. *Lagenandra ovata* P1983/5653, and 32. *L. meeboldii* P1979/5019. b; sequence symbols: A, C, G, T = dATP, dCTP, dGTP, dTTP; hyphens = gaps or missing nucleotides.

3.3.2 ITS nucleotide site variation

Following alignment of the ITS sequences, a total of 712 characters (positions) were available for comparison. However, it was necessary to exclude 23 positions from ITS2 (positions 643-665) prior to phylogenetic analysis because of alignment ambiguity. Of the remaining unambiguous aligned positions, 202 or 29.3% were variable, i.e. varied for more than one DNA nucleotide type. Approximately 49.5% of the variable sites were contained within ITS1, 46% in ITS2 and only 4.5% in the 5.8S subunit.

Of these variable characters, 118 or 58.4% were potentially informative phylogenetically i.e. possessed nucleotides states shared by at least two DNAs. It was found that 53.4% of the sites were contained in ITS1, 43.2% in ITS2 and 3.4% in 5.8S subunit.

3.3.3 Phylogenetic analysis

2112 most parsimonious trees were generated from Fitch parsimony analyses of potentially most informative sites from *Cryptocoryne* species and outgroup (*Lagenandra*) ITS sequences. One of these equally parsimonious trees is shown in Fig. 3.12. These 2112 most parsimonious trees require 312 steps with a consistency index of 0.776. The phylogenetic tree in Figure 3.12 showed that the genus *Cryptocoryne* is monophyletic relative to the sister genus *Lagenandra*. Within *Cryptocoryne*, three clades could be identified. The first clade contained species from mainland Asia, i.e. *C. spiralis, C. albida, C. retrospiralis* and *C. crispatula* var. *balansae*.

The second clade consisted of species found only on the island of Sri

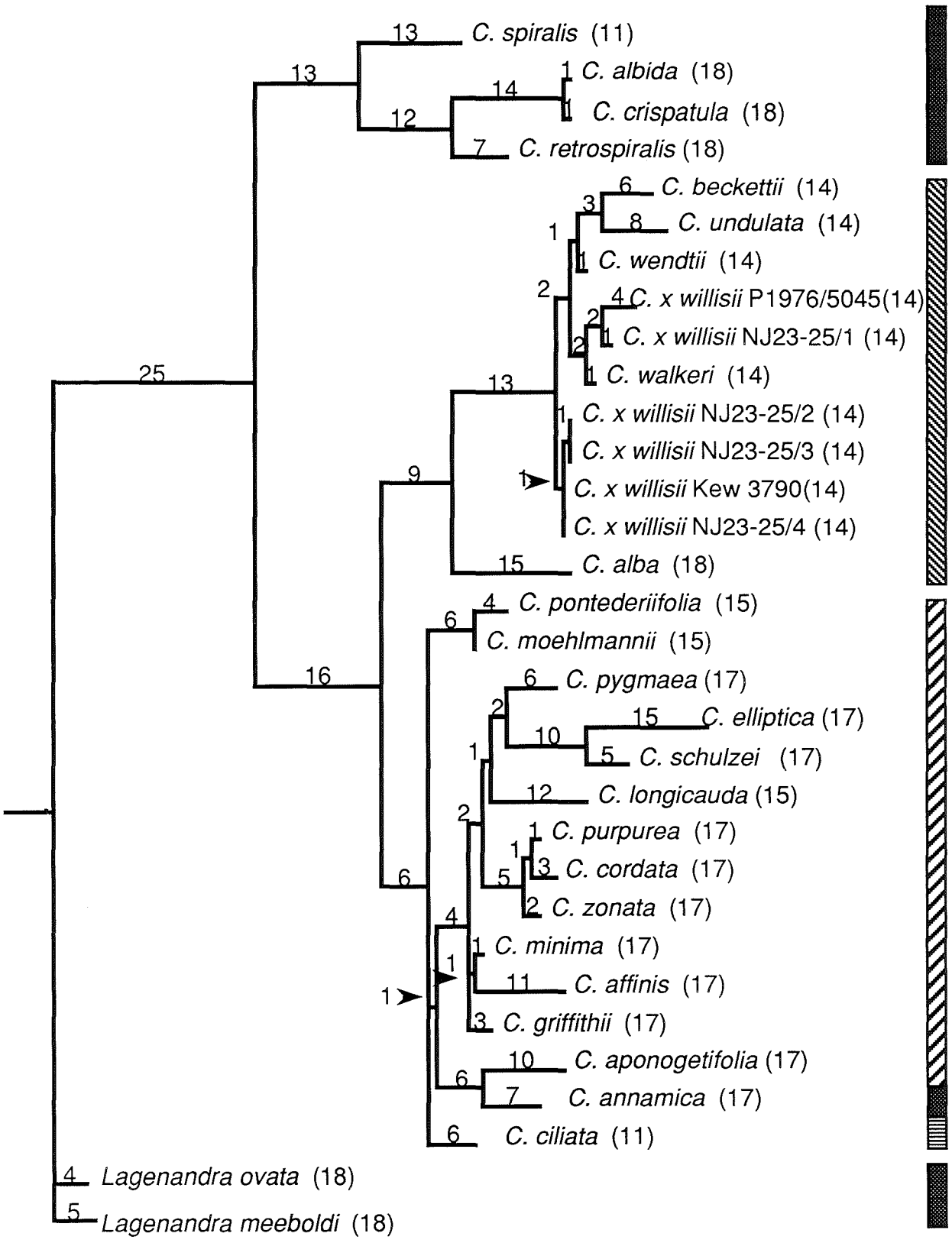


Figure 3.12. One of the 2112 equally parsimonious trees from analysis of the ITS region of 25 species of *Cryptocoryne* and two species of *Lagenandra*. Numbers above each branch are the branch lengths. Numbers in brackets are chromosome base numbers.

Lanka. Within this clade *C. alba* was found to be the sister taxon to the rest of the *Cryptocoryne* species of Sri Lanka included in the analysis.

The third clade consisted of species mainly from the Malesia floristic region with the exception of *C. annamica* which is found in Vietnam (mainland Asia) and *C. ciliata* which is found throughout Asia.

The equally parsimonious trees were summarised using a strict consensus tree (Fig. 3.13) including the bootstrap values supporting each branch, and a majority rule consensus tree (Fig. 3.14) . Both the strict consensus and majority consensus trees showed the monophyly of *Cryptocoryne* and divided the genus into three clades. However, the resolution within the *C. beckettii* group (the sister group of *C. alba*) which constitute the Sri Lanka clade, was not well supported by bootstrap values in the strict consensus tree.

The strict consensus and the majority rule consensus trees differ in their topology for *Cryptocoryne* species within the Malesia clade. The majority rule consensus tree provides an insight to species relationships which were not made evident in the strict consensus tree due to the latter conservative nature in dealing with conflicting groups. Within the majority rule consensus tree, the clade containing species from Malesia region form a tetrachotomy with one particular subclade consisted of all the Malay Peninsula species (*C. elliptica, C. schulzei, C. longicauda, C. purpurea, C. cordata, C. minima, C. affinis* and *C. griffithii*) including a species from the Philippines (*C. pygmaea*) and a Borneo species (*C. zonata*). Within this subclade, *C. affinis* appears to have a close relationship with *C.*

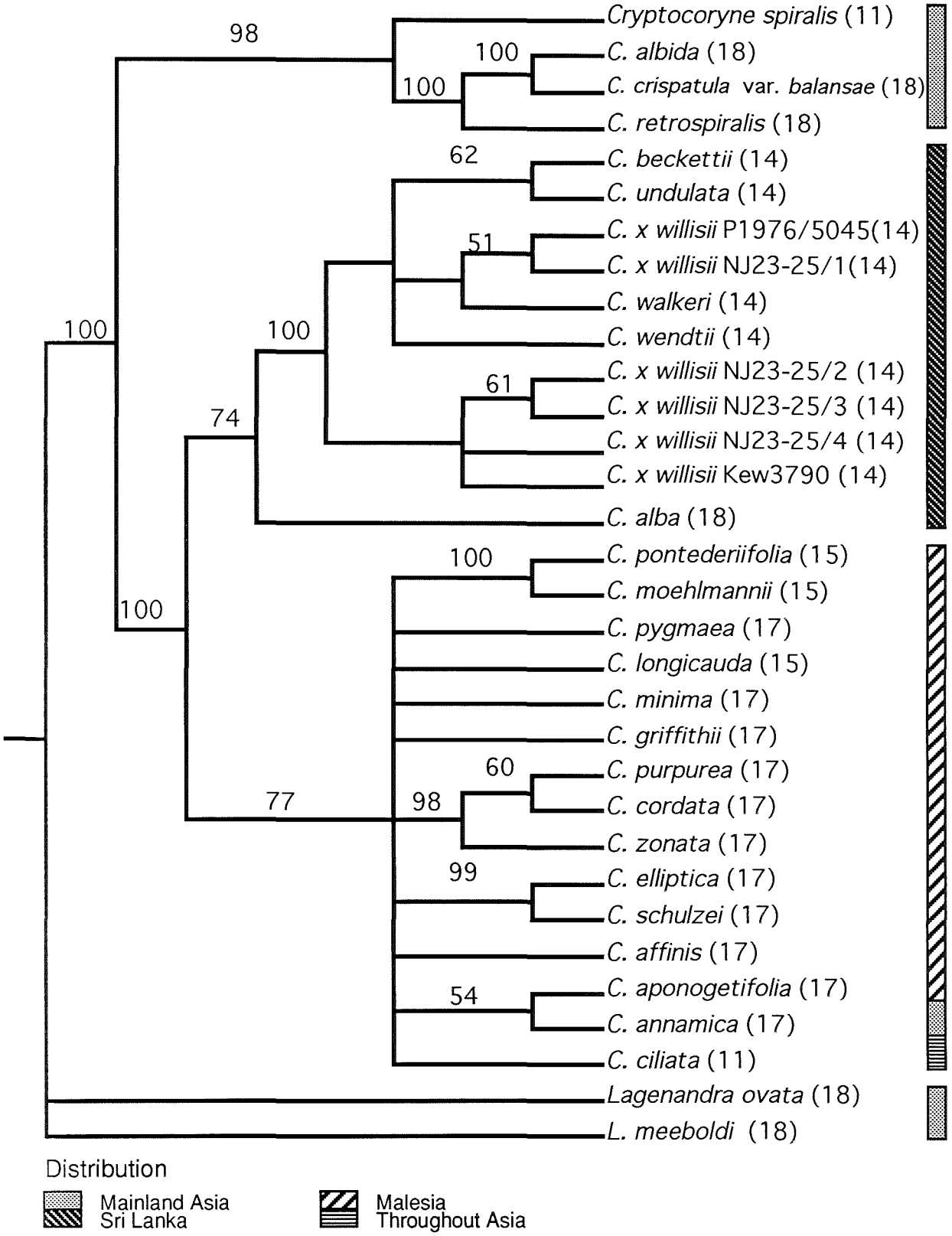


Figure 3.13. Strict consensus of 2112 equally parsimonious trees of 25 species of *Cryptocoryne* and two species of an outgroup genus, *Lagenandra*. Values given above the branches are bootstrap values (100 replicates). Number in brackets are chromosome base numbers.

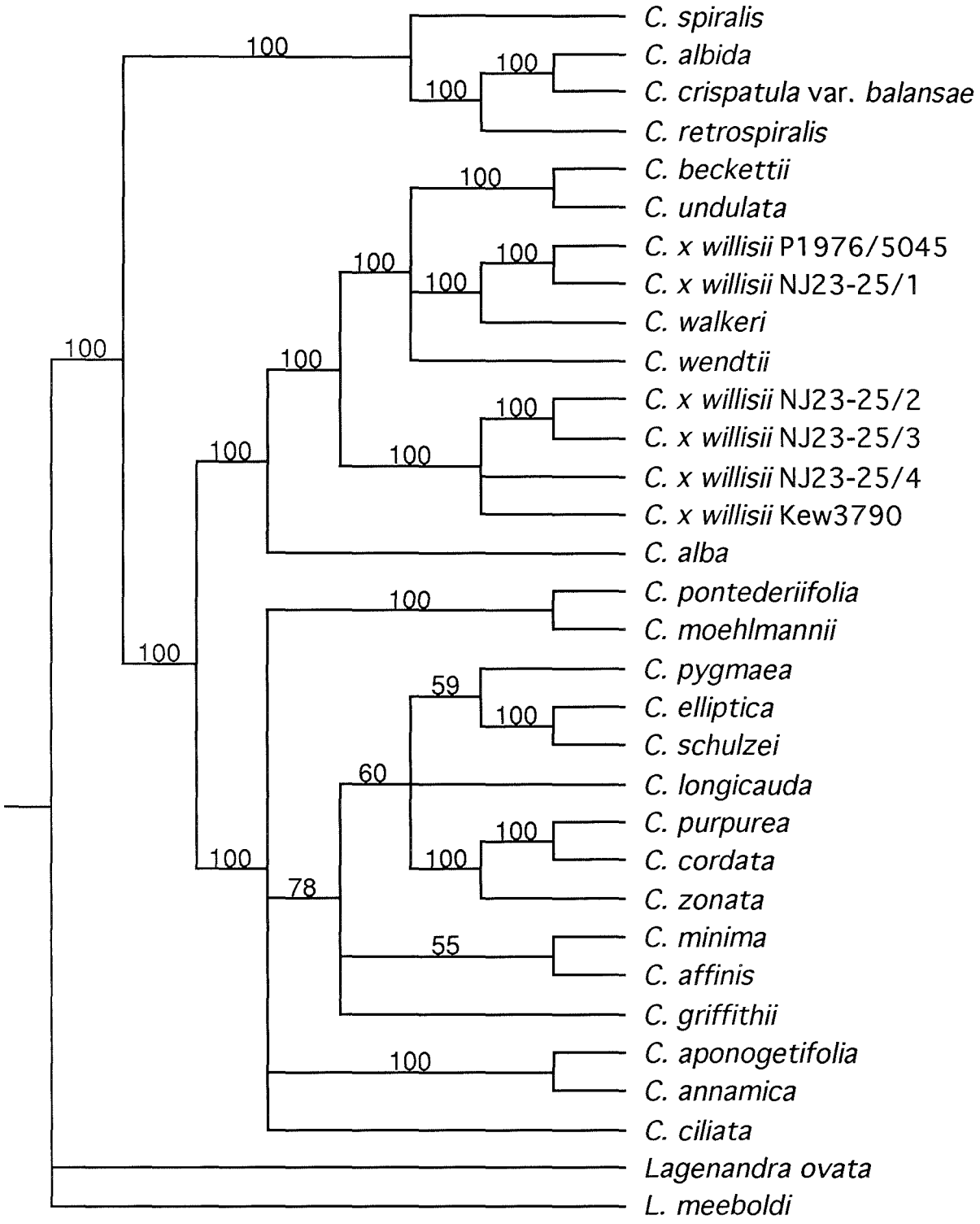


Figure 3.14. A majority rule consensus of 2112 equally parsimonious trees of 25 species of *Cryptocoryne* species and two species of the outgroup genus, *Lagenandra*. Values given above the branches are the percentage or number of time of the particular branch occur in the 2112 most parsimonious trees.

minima, and *C. pygmaea* is the basal sister taxon to *C. elliptica* and *C. schulzei.*

3.4 Discussion

3.4.1 ITS phylogenetic resolution: Biogeography of *Cryptocoryne*

Several conclusions can be drawn about the biogeographic history of the genus *Cryptocoryne* based on the ITS phylogeny (Fig. 3.12). The genus can be separated into three clades which are comprised of species from different geographical regions. *Cryptocoryne* taxa which appear to be basal to the tree are comprised of species from the Indian subcontinent and elsewhere on mainland Asia. Within this mainland Asia clade, *C. spiralis* occupies the basal position which suggests that it is the most primitive member of the genus *Cryptocoryne*. This agrees with previous hypotheses for the genus based on morphological examination. *Cryptocoryne spiralis,* in common with *Lagenandra* taxa retains a spathe with a spirally twisted limb and lacks a tube. In contrast, all other *Cryptocoryne* species produce a spathe which includes a tube.

The placement together of the other three species, *C. albida, C. retrospiralis* and *C. crispatula* var. *balansae,* which make up the first clade in the ITS phylogeny is also congruent with previous systematic treatments. These three species were placed into one taxonomic group by Arends *et al.* (1982) as they differ from the rest of *Cryptocoryne* species by having a spathe with a long tube and a spirally twisted limb which lacks a collar (Jacobsen, 1980a). Thus, a spirally twisted limb has been retained by all taxa of this particular lineage. *C. retrospiralis* occurs mostly in the western part of India and northern Bangladesh, while *C. crispatula,* in general, is found in west Bangladesh throughout Thailand to Southern China. In contrast, *C. albida* is only present at low elevations in Myanmar

and south Thailand. The ITS phylogeny would suggest that the ancestor of *C. crispatula* var. *balansae* and *C. albida* had differentiated in western India and then migrated and radiated throughout Bangladesh, Myanmar, Thailand and South China. This would fit in with the proposal of Jacobsen (1977) that evolution within *Cryptocoryne* can be considered as an example of 'island speciation', where 'islands' are represented by different river systems in South Asia which are isolated from each other, thus allowing allopatric speciation to take place.

A second conclusion that can be drawn from the ITS phylogeny is in regard to the clade which consists of species exclusively found in the island of Sri Lanka. Morphologically and cytologically, *Cryptocoryne* of Sri Lanka can be separated into two taxonomic groups. The first taxonomic group consists of species with x=18, i.e. *C alba, C thwaitesii* and *C. bogneri* (the latter two species were not available for analysis), while the second group consists of species with x=14, i.e. *C. beckettii, C. wendtii, C. walkeri, C. undulata, C. x willisii, C. parva* and *C. nevillii* (*C. nevillii* was not available for analysis, whilst analysis on material of *C. parva* failed to give a reliable ITS sequence). These two taxonomic groupings were also evident in the ITS phylogeny. The group which is basal to this clade is *C. alba.* This species is characterised by having a spathe with distinctly long, smooth and caudate white limb, which lacks a collar, and is often red spotted on its inner surface. de Graaf and Arends (1986) considered *C. alba* together with *C. thwaitesii* and *C. bogneri* as local endemics as they occur in a very restricted area within Sri Lanka. Species within the other group (known as the *C. beckettii* group by Arends et al., 1982), are categorised by having a spathe with a long tube (except in *C. parva*) and a limb with a prominent

collar (except in *C. walkeri*). Most species of this group are concentrated in the central part of the island, apart from *C. nevillii* which is found in the eastern part, and *C. wendtii* which occupies the north western part of Sri Lanka. Species of the *C. beckettii* group are mainly distributed north of the 7th degree of latitude while *C. alba* is found only to the south of this latitude (de Graaf and Arends, 1986). Thus it is not surprising, that the ITS sequence of *C. alba* differs from that of the *C. beckettii* group and reflects the differences that are also evident between the two groups in morphology and chromosome number (see section 3.4.2 in this chapter). The clade containing the *C. beckettii* group (see Fig. 3.12 and Fig 3.13) was not very well resolved based on the ITS sequence variation. This suggests that after the ancestor of this group split from *C. alba*, it became isolated to radiate rapidly in different river systems of Sri Lanka and insufficient time has passed for the accumulation of many mutations within the ITS region of the rDNA gene.

The third clade in the ITS phylogeny is, with the exception of *C. annamica* and *C. ciliata*, comprised of species from an area known as the Malesia floristic region. The Malesia region is a centre of tropical plant diversity which extends from the Malay Peninsula through the archipelago in South East Asia and includes New Guinea. According to Michaux (1991) the geological origin of this area is almost exclusively Gondwanic (a great supercontinent during late Paleozoic and Mesozoic around 400 million years ago which later separated into South America, Africa, Antarctica, Australia and India) apart from north Borneo, Southern Palawan and the shelf of Borneo.

The resolution of species relationships within the Malesia clade is not

satisfactory, and a polytomy for all species within this clade is present in the strict consensus tree (Fig. 3.13). Nevertheless, several interesting findings are evident. One intriguing finding involves the placement of *C. annamica* and *C. aponogetifolia* within the same monophyletic group (although with a poor support - 54% bootstrap). *C. annamica* is only found in Vietnam while *C. aponogetifolia* occurs in the Philippines. These two locations are separated by a deep ocean, but the distance between them is not great and it is, therefore, conceivable that their ancestor was dispersed to both areas.

A close association was also established between *C. pontederiifolia* and *C. moehlmannii* in clade 3. These two tidal zone species are found only in Sumatra. Similarly, *C. elliptica* and *C. schulzei* were shown to be closely related. The latter species is found only in the southern part of the peninsula, while *C. elliptica* occurs nowadays only in the north, athough it was previously recorded in the south. Their recent common ancestry is, therefore, not surprising.

Another interesting finding involves the monophyletic group within clade 3 which contains *C. purpurea, C. cordata* and *C. zonata*. Jacobsen (1977) pointed out that pollen of *C. purpurea* is completely sterile and one explanation for this is that the species is of hybrid origin with *C. cordata* and *C. griffithii* as its putative parents (Jacobsen and Mansor, unpublish). The spathe of *C. purpurea* has a broad collar zone, which is a characteristic of the spathe of *C. cordata*, and a rough purple limb which might be inherited from *C. griffithii*. Moreover, *C. purpurea* has been found in the Malay Peninsula at Mengkibol and Kota Tinggi in the state of Johore where the distributions of *C. cordata* and *C. griffithii* overlap. The ITS tree

suggests that *C. purpurea* resemble more closely to *C. cordata* than *C. griffithii*. This could stem from concerted evolution occurring within *C. purpurea* such that its ITS is being homogenised towards that of *C. cordata*. Further analysis, involving a larger representative samples of *C. purpurea*, *C. cordata* and *C. griffithii* is required to investigate this possibility in greater detail.

The majority rule consensus tree (Fig. 3.14) does, of course, provide a better resolution of species relationships within the Malesian clade. An interesting feature of this tree is the placement of *C. pygmaea*, a Philippines species, as sister taxon to *C. elliptica* and *C. schulzei*. The association of *C. pygmaea* with the two Malay Peninsula species rather than to other Philippine species is of interest as morphologically it does not resemble other Philippine species.

Another interesting feature of the majority rule consensus tree concerns the close relationship evident between *C. minima* and *C. affinis*. This association is not expected as *C. minima* was thought to be more closely related to *C. griffithii* than to *C. affinis* (Arends *et al.*, 1982).

3.4.2 Chromosomal evolution

Petersen (1989) in her study on chromosome numbers within Araceae has suggested that *Cryptocoryne* together with *Lagenandra* originated from an ancestral genus with a basal chromosome number of x=9. A likely candidate would be a genus of the Madagascan tribe Arophytae, which has the basic number x=9. However, studies of chloroplast RFLP within Araceae (French *et al.*, 1995) failed to support this assumption, placing

subtribe Schismatoglottidinae as sister to Cryptocoryninae with Arophytae only distantly related. In addition, there appear to be no support for this proposal either based on morphological comparisons (Grayum, 1990). Initial ITS amplification on material of *Arophyton* (Arophytae) resulted in a PCR product which was longer (around 50 bp longer) than all of the PCR products obtained from materials of *Cryptocoryne* and *Lagenandra* used in this study. Unfortunately, the sequencing of the pooled PCR product of *Arophyton* was not successful.

Reumer (1984), however, supported the view that x=18 might have originated from x=9 by the simple neans of euploid doubling and concluded that *Cryptocoryne* is of polypheletic origin with one lineage having x=18 as the ancestral state, and another with x=11 as the other ancestral state. He believed that further chromosome number changes within the genus occurred as losses or gains of whole chromosomes due to meiotic irregularities. The ITS phylogeny suggests, however, that *Cryptocoryne* is monophyletic and that the genus originated from an ancestor with x=18. The ITS phylogeny also suggests that the basal chromosome number of x=18 genome is maintained in at least two lineages (Mainland Asia and Sri Lanka) and perhaps also within the Malesia lineage which is likely to contain *C. lingua* (x=18) from Borneo, a species that was not available for ITS analysis. It is feasible that historically there was an easterly dispersal of taxa with x=18 from the Indian subcontinent towards the Malesia region. Within the Sri Lanka lineage, the basal species, *C. alba* has x=18 and it is possible, therefore, that x=14 arose in this lineage from x=18 through aneuploidy, as suggested by Reumer (1984).

The ITS phylogeny did not give a clear resolution of the origin of species within the Malesian region, with x=15 and x=17. Three species included in the present analysis had a base chromosome number of x=15. These were *C. pontederiifolia*, *C. moehlmannii* and *C. longicauda*. The first two of these species occur only in Sumatra, while *C. longicauda* has a wider distribution area which includes Sumatra, the Malay Peninsula and Borneo. In contrast, *Cryptocoryne* species with x=17 are more widely dispersed, being found in Sumatra, the Malay Peninsula, Borneo, the Philippines, Indo-China as well as New Guinea. Bearing in mind this difference, and also the much greater morphological variation that exists between with x=17, Reumer (1984) suggested that species with x=15 might have originated from x=17 and afterwards spread over the Malay Peninsula, Sumatra and Borneo at the time when these areas formed one single land mass (Sundaland). The ITS phylogeny, however, suggests that taxa with x=15 might have arisen twice, once in Sumatra and secondly in the Malay Peninsula (see Fig. 3.12 and Fig. 3.14), and it is not clear whether x=15 has originated from x=17 stock or vice versa. It should also be mentioned here that perhaps the sampling was not comprehensive enough to enable the resolution on the origin of both x=15 and x=17. Reumer further suggested that x=17 originated from x=18. Unfortunately, this could not be tested by the phylogeny constructed as the only Malesian species known to have x=18, *C. lingua*, was not available for analysis.

One particular difficulty that requires explanation concerns the presence of *C. spiralis* with x=11 in the mainland Asia clade. If we accept that x=18 is the ancestral chromosome state, and that this base number has been clearly maintained in the mainland Asia lineage, how might we explain the large shift from the ancestral type to x=11, and then the further shift

back to x=18, as indicated by the ITS phylogeny? How this may have been achieved remains a mystery. The only other species in the genus with x=11 is placed in a completely different lineage (clade 3) and only helps us to conclude that during the course of evolution of the genus, x=11 genome arose twice in an independent fashion.

3.4.3 Parentage of hybrid species

As mentioned in the method section of this chapter, the ITS region of three individuals from three different accessions of *C. x willisii*, a naturally occurring hybrid in Sri Lanka, were analysed. The ITS region from these accessions was cloned prior to sequencing, so as to separate the different ITS types that might be present within the hybrid genome. For *C. x willisii* accession NJ23-25 four clones of the ITS region from four different colonies were sequenced, while only one each was sequenced from accession P1976/5045 and accession Kew3790.

In the ITS phylogeny (Fig. 3.12), the ITS sequences of *C. x willisii* fall into two subclades of clade 2. The ITS sequence from *C. x willisii* accession P1976/5045 and one of the four ITS sequences from *C. x willisii* accession NJ23-25 were placed together with *C. walkeri*. This suggests that one of the parents of the hybrid is *C. walkeri*. None of the ITS sequences from *C. x willisii* accessions grouped with *C. beckettii* which would indicate that this species was not its other parent. Previous results from hybridisation and morphological investigations indicated that *C. x willisii* has *C. parva* as one of its parent and either *C. beckettii* or *C. walkeri* as the other parent (Jacobsen, 1981). In the current study a reliable ITS sequence could not be obtained for *C. parva*. It might be speculated, however, that had one been

obtained, it would have fell into the same group as the ITS sequences from *C. x willisii* accession Kew3790 and the three remaining ITS sequences from *C. x willisii* accession NJ23-25. If this were so, then the parentage of the hybrid would be firmly demonstrated to be *C. walkeri* and *C. parva*.

The occurrence of different races of ITS sequence within *C. x willisii* accession NJ23-25, which are placed into two separate subclades in clade 2, would suggest that ITS sequence homogenisation through concerted evolution has not taken place within this hybrid. However, proof of this requires the sequence of *C. parva* to be known, assuming that *C. parva* is one of the parents of *C. x willisii* accession NJ23-25. An alternative to concerted evolution is backcrossing which may have played a role in the maintenance of only one parental ITS type within the hybrid genome. An initial interspecific hybridisation, following recombination, may have resulted in a hybrid with both parental ITS repeat-types. However, after backcrossing and perhaps through subsequent vegetative reproduction, one of the parental ITS repeat-type was lost while the other parental ITS type was maintained within a hybrid and perhaps within the entire hybrid population. The parental ITS repeat-type which is maintained in a hybrid population may be determined to a large extent by which of the parental species the hybrid form backcrosses with.

Chapter 4

Chloroplast DNA variation and phylogenetic relationships within the Genus *Cryptocoryne*

4.1 Introduction

4.1.1 Structure of chloroplast (cp) DNA

The chloroplast DNA (cpDNA) molecule forms a closed circle comprised of three different regions; a large single copy (LSC) region, a small single copy (SSC) region and two regions in reverse orientation, the inverted repeat (IR) (Fig. 4.1). The inverted repeat separates the SSC from the LSC. The chloroplast genome varies little in size, structure and gene content among angiosperms (Olmstead and Palmer, 1994). In size, the cpDNA genome ranges from 120 to 217 kilo bases (kb) (Downie and Palmer,1992). The lower extreme of this variation is found in a group of legumes (Palmer and Thompson, 1982) which lack one copy of the inverted repeat, and also in the conifers (Raubeson and Jansen, 1992), and the parasitic plant, *Epifagus* (Orobanchaceae) which contain a large deletion in their cpDNA genome (de Pamphilis and Palmer, 1990). The largest cpDNA genomes have been reported in *Spirodella oligorrhiza* (180kb) and *Pelagonium x hortorum* (217kb; Palmer, 1985). The increased size of the cpDNA genome of *Pelargonium x hortorum* is due to an increase in the length of the inverted repeat. Between these extremes in genome size, the majority of angiosperms have a cpDNA genome that falls within the range of 135 to 160kb (Palmer, 1985; Olmstead and Palmer, 1994). In regard to structure, a notable variant form of cpDNA has been found in the legumes, all of which contain a 50kb inversion within their LSC (Doyle *et al.*, 1992).In addition, some tribes of Papilionoideae have lost a large inverted repeat.

4.1.2 Favourable features of chloroplast DNA for phylogenetic studies

Surveys of cpDNA variation offer several advantages in phylogenetic

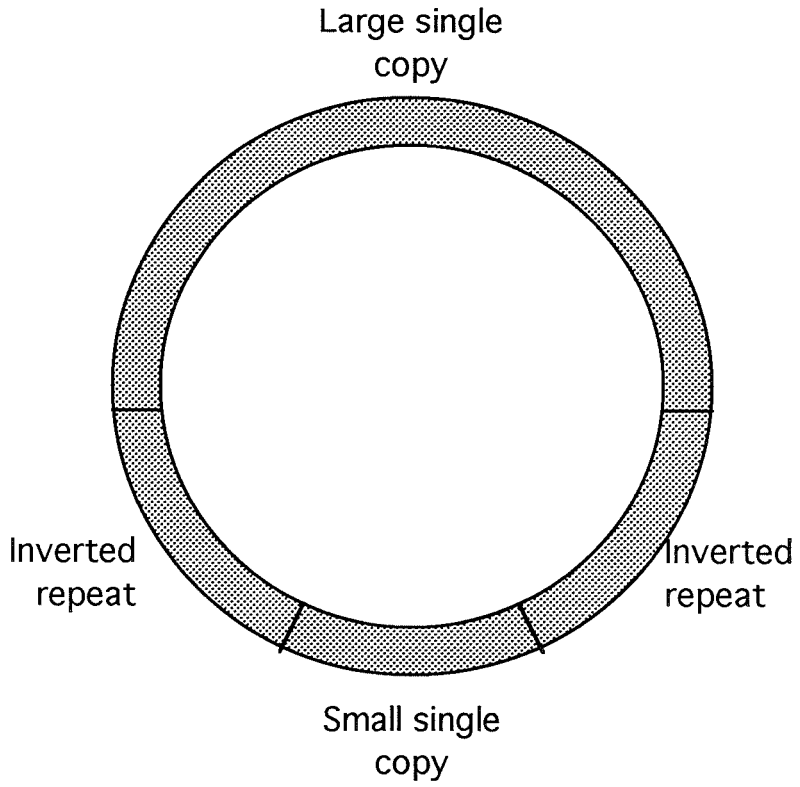


Figure 4.1. A schematic diagram of the circular chloroplast DNA genome showing the two inverted repeats (IR) which separates the large single copy (LSC) from the small single copy (SSC).

studies. The molecule is highly conserved in terms of nucleotide substitution. This slow rate of nucleotide substitution permits the molecule to be used in resolving phylogenetic relationships especially at deep levels of evolution (Clegg and Zurawski, 1992). As cpDNA is highly conserved in size, structure and gene content, any change in its structure, arrangement or content of the genome might have significant phylogenetic implication (Downie and Palmer, 1992). Genes which are encoded within the chloroplast genome evolve at different rates and surveys of these genes with different evolutionary rates can, therefore, provide phylogenetic information at various taxonomic levels.

One important feature of cpDNA in phylogenetic studies is its mode of inheritance. The cpDNA is clonally inherited (Palmer *et al.*, 1988) predominantly through the maternal parent in most angiosperms. It is also transmitted paternally in some angiosperms and this mode of transmission is frequent in the gymnosperms (Harris and Ingram, 1991). In plants in which biparental transmission occurs, the chloroplast genome of the offspring are not known to recombine, but simply segregate somatically. It is important to assess the mode of plastid inheritance in a genus as an incorrect assumption can lead to errors in studies of the parentage of polyploid taxa, the origin of hybrids and gene flow (Harris and Ingram, 1991).

The chloroplast genome is comparatively small in size and this makes it easy to carry out restriction site mapping (Palmer, 1987). However, the size is large enough to allow many restriction sites to be sampled per restriction enzyme (Olmstead and Palmer, 1994), and it is possible to visualise all digested fragments in a single agarose gel (Crawford, 1990).

Much work has been carried out in using cpDNA for systematic studies. With the determination of complete cpDNA sequences of *Nicotiana tabacum* (Shinozaki *et al.*, 1986), *Marchantia polymorpha* (Ohyama *et al.*, 1986), *Oryza sativa* (Hiratsuka *et al.*, 1989) and *Epifagus virginiana* (Wolfe *et al.*, 1992), information on the structure and gene content of the chloroplast genome from different lineages can be compared with relative ease (Olmstead and Palmer, 1994). For example, the comparison of the gene content and order between a monocotyledon species, *Oryza sativa* and a dicotyledon species, *Nicotiana tabacum*, showed that they differ three inversions, the absence of three protein genes, and gene duplication and rearrangements associated with the movement of the ends of the inverted repeat, which represent a derived changes within the monocotyledon lineage.

The chloroplast molecule is present in many copies per chloroplast. There are around 20 to 200 copies of cpDNA present in a mature chloroplast (Palmer, 1987) and this high copy number per cell facilitates DNA extraction and analysis.

4.1.3 Chloroplast DNA in phylogenetic studies: restriction site mapping of chloroplast DNA

The conventional method used for analysing chloroplast DNA for phylogenetic analysis involves the analysis of restriction fragment variation of the total chloroplast genome. The method relies on the Southern blotting procedure which is carried out as follows. First the intact chloroplast genome is digested with restriction enzymes and the

resulting fragments are separated in an agarose gel. The DNA is transferred on to a nylon membrane and fragments are visualised by hybridisation to radio- or fluorescent labelled probes. Probes are fragments of the chloroplast genome.

The analysis of restriction fragment variation of cpDNA has been widely used in elucidating taxon relationships in phylogenetic studies both at higher and lower taxonomic levels. At the higher level, Kim and Jansen (1995) employed this approach to determine relationships between genera within the family Berberidacae. The results they obtained were congruent with those of a previous *rbc*L gene sequence analysis in recognising four major chromosomal groups (x=10, 8, 7 and 5). However, their findings provided no support for fragmenting the family into a set of smaller families as suggested by previous treatments. Overall, their cpDNA phylogeny supported the morphological classification of the family outlined by Loconte and Estes (1989).

Restriction site analysis of cpDNA variation was also employed by French *et al.* (1995) in a study of phylogenetic relationships between 86 genera of Araceae and the related genera *Lemna* and *Acorus*. Their results supported the monophyly of many tribes of Araceae, and contrary to previous belief, the monoecious aroids appeared to form a monophyletic lineage rather than being composed of polyphyletic lineages. Support was also obtained for including the Lemnaceae within the tribe Aroidae (Araceae).

At the lower taxonomic level, many studies have examined cpDNA

restriction site variation to elucidate the phylogeny of species within particular plant genera, e.g. in *Leucaena* (Harris *et al.*, 1994), *Anemone* (Hoot, 1995), *Erythrina* (Bruneau, 1996), *Cornus* (Xiang *et al.*, 1996), *Fragaria* (Harrison *et al.*, 1997) and *Andira* (Pennington, 1996). Gillies and Abbott (1996) used this approach to examine phylogenetic relationships within the papilionoid tropical forage legume genus *Stylosanthes*. The phylogenies they obtained with Wagner parsimony divided *Stylosanthes* into four separate clades, and overall species relationships were congruent with those previously established from other character comparisons.

Analysis of restriction site variation has also been of value in resolving species phylogenies in *Magnolia* (Qiu *et al.*, 1995) and *Symplocarpus* (Wen *et al.*, 1996). In both of these cases, the cpDNA phylogeny helped to improve an understanding of the evolution and time of divergence of taxa that are disjunctly distributed in eastern Asia and eastern North America.

Similarly, cpDNA restriction site variation has been informative in clarifying the colonisation and evolution of plant species on oceanic islands. For example, Francesco-Ortega *et al.* (1996) showed that *Argyranthemum* (Asteraceae), which is the largest endemic plant genus in Macaronesia, is a monophyletic group that has speciated only recently. One of two major cpDNA lineages recognised is restricted to the northern part of the Macaronesian archipelago (Madeira, Desertas and Selvagens) while the other comprises those taxa endemic to the southern portion of the archipelago, i.e. the Canary islands. Two major radiations were recognised within the Canary island taxa; one of these was restricted to

ecological zones influenced by the northeastern trade winds, while the other occurred at sites unaffected by these winds.

Other studies that have involved surveys of cpDNA restriction site variation to examine the colonisation and evolution of plant groups on oceanic islands include those by Baldwin *et al.* (1990) on the Hawaiian Silversword alliance, Mes and Hart (1996) on the Macaronesia genus, *Aeonium* and Ito and Pak (1996) on the genus *Crepidiastrum.*

CpDNA restriction site variation has also been of value in determining the origin of polyploid species (see Soltis and Soltis, 1995). For example, from an analysis of such variation, Harris and Ingram (1992) showed that the maternal parent of the newly evolved allohexaploid species, *Senecio cambrensis* (2n=6x=60) was the tetraploid *S. vulgaris* (2n=4x=40) rather than the diploid *S. squalidus* (2n=2x=20). Moreover, *S. cambrensis* was polymorphic for cpDNA haplotype which confirmed that the species had arisen on at least two separate occasions following hybridisation between *S. vulgaris* and *S. squalidus.*

Although most studies of cpDNA restriction site variation have utilised Southern hybridisation and probing to resolve mutations, more recently an alternative method has been used for the same purpose which takes advantage of PCR. This approach involves the amplification and subsequent digestion with restriction enzymes of a segment of cpDNA genome usually shorter than 4 kb. The restriction enzymes used to digest the amplified product are normally four base cutters and the fragments produced are separated in an agarose gel and stained with ethidium

bromide. This technique has advantages where amounts of DNA for analysis are limited; moreover it is fast and no blotting was required. It also avoids the use of radioisotopes, as probing is not required. Several studies have examined phylogenetic relationships between plant taxa using this method. These include studies on *Astragalus* - Fabaceae (Liston, 1992), Datiscaceae (Rieseberg *et al.*, 1992), *Astragalus* - Fabaceae (Liston and Wheeler, 1994), *Prunus* (Badeness and Parfitt, 1995), Astragalinae - Fabaceae (Ding *et al.*, 1995), Papaveraceae (Schwarzbach and Kadereit, 1995), Papaveraceae subfamily Papaveroideae (Jork and Kadereit, 1995), families of conifers (Tsumura *et al.*, 1995), Dipterocarpaceae (Tsumura *et al.*, 1996) and Scrophulariaceae (Wolfe *et al.*, 1997).

4.1.4 Chloroplast DNA in phylogenetic studies: DNA sequencing

While restriction site analysis represents an indirect comparison of DNA sequence variation, DNA sequencing, provides a direct comparison. Olmstead and Palmer (1994) stated two criteria that should be met in the choice of a sequence of cpDNA for phylogenetic purposes. First, the sequence should be of sufficient length to provide enough informative nucleotide positions for phylogenetic reconstruction, and secondly the selected DNA sequence should have a substitution rate which is appropriate to the phylogenetic problem being addressed.

Currently there are 20 gene sequences within the cpDNA genome which have been used for phylogenetic analysis (Olmstead and Palmer, 1994). These gene sequences vary widely in evolutionary rate, and may be used, therefore for phylogenetic purposes at various taxonomic level. However, within the angiosperms, phylogenetic analysis based on sequences from cpDNA has so far been restricted to only a few gene sequences, and most

notably the gene which encodes the large subunit of ribulose-1,5-biphosphate carboxylase (*rbc*L). The great interest in using the *rbc*L gene sequence for phylogenetic reconstruction stems from the abundance of information available on the enzyme and its encoding gene, and its importance in photosynthesis. A concerted effort of obtaining and analysing *rbc*L sequences data to examine the phylogeny of seed plants has been reported by Chase *et al.* (1993). In certain lineages, the *rbc*L gene sequence has been found to evolve at a high rate, e.g. within the Geranium family (Price and Palmer, 1993) and in *Pelargonium* in particular (Price and Palmer, unpublished data). In *Geranium*, Pax *et al.*(1997) found that Hawaiian species of *Geranium* have a very similar *rbc*L gene sequence to one another, showing them to comprise a monophyletic group that is closely related to American *Geranium* rather than *Geranium* from Asia or the Pacific.

Several other cpDNA gene sequences currently used in phylogenetic analysis are the *rps*4 (SouzaChies *et al.*, 1997), *rpo*A (Petersen and Seberg, 1997), *ndh*F (Terry *et al.*, 1997), *rpL*16 (Jordan *et al.*, 1996), *mat*K (Johnson and Soltis, 1994; Steel and Vigalys, 1994) and *atp*B (Jensen *et al.*, 1995; Hoot and Crane, 1995) sequences. Some of these genes have higher substitution rates and, therefore, might be useful in addressing phylogenetic questions at lower taxonomical levels. For example, the *mat*K gene sequence exhibit twice the rate of nucleotide substitution relative to the *rbc*L gene sequence in the Polemoniaceae (Steele and Vigalys, 1994). Similarly, the *ndh*F gene (which putatively encodes a chlororespiratory peptide) comprises a sequence which is 50% longer and twice as variable as the *rbc*L sequence (Olmstead *et al.*, 1993; Olmstead and Palmer, 1994; Olmstead and Sweere, 1994).

Sequences of non-coding regions of the cpDNA genome have also been used in phylogenetic reconstruction e.g. the intergenic spacer region between the *trn*L(UAA)3' exon and *trn*F(GAA). Gielly and Taberlet (1996) reported that this intergenic spacer region gave a reasonable resolution of taxon relationships within *Gentiana*, although the resolution was lower for than a nuclear-based phylogeny (Gielly *et al.*, 1996). They concluded that the intergenic sequence was more appropriate for assessing intergeneric rather than intrageneric relationships. The same sequence has also been examined to assess phylogenetic relationships within the genus *Aconitum* (Kita *et al.*, 1995), within the genus *Sedum* (Kim *et al.*, 1996b) and among genera of Macaronesian family Sempervivoideae (Mes *et al.*, 1996). Similarly, the spacer region between the *atp*B and *rbc*L gene has been used in phylogenetic studies of plants (Sovalainen *et al.*, 1994; Manen and Natali, 1995).

4.1.5 Aims and objectives of cpDNA analysis

The primary goals of the survey of cpDNA variation in *Cryptocoryne* that is reported in this chapter were two-fold. First, to obtain a phylogeny for the genus based on an organelle genome for comparison with that constructed from the nuclear gene (rDNA ITS) sequence variation reported in chapter 3. Second, to compare and contrast the levels of cpDNA variation resolved by (i) restriction site analysis of PCR amplified cpDNA regions, (ii) restriction analysis of the total cpDNA genome using conventional Southern hybridisation and probing procedures, and (iii) sequencing particular segments of the cpDNA genome.

4.2 Materials and Methods

4.2.1 Plant Material

Most of the plant material subjected to an analysis of cpDNA variation was the same as that examined for rDNA ITS sequence variation (see chapter 3). Details of the accessions assayed are given in the results section of this chapter.

4.2.2 Methods

Three different techniques were used to resolve cpDNA variation within *Cryptocoryne*. These techniques included the analysis of (i) restriction site variation within a PCR-amplified cpDNA region, (ii) restriction site variation of total cpDNA genome, and (iii) sequence variation within particular segments of the chloroplast genome.

4.2.2.1 Restriction site variation within a PCR amplified cpDNA regions

An aliquot (1 µl; 50-100 ng) of purified DNA extracted (as described in chapter 2) from plant material was used as DNA template for PCR amplification. The PCR reaction mixture contained 1.5 - 3 mM $MgCl_2$, 100 µM of each of the four dNTPs, 50 pmol of each primer, 10 µl 10x Taq polymerase buffer, 1 unit of *Taq* polymerase and sterile distilled water made up to a final volume of 100 µl. The PCR reaction was overlaid with two drops of mineral oil. The amplification was carried out in a Techne PHC-3 thermal cycler using 1 cycle of 4 mins at 94°C, 30 cycles of 45 seconds at 94°C, 1 min at 48 - 60°C, 2-4 mins at 72°C and one cycle of 10 mins at

72°C. The PCR product was visualised by UV illumination after electrophoresis on 0.8% gel and staining with ethidium bromide. The eleven universal primers used to amplify eleven non-coding regions of cpDNA (Taberlet *et al.*, 1991; Demesure *et al.*, 1995) are listed in table 4.1.

After amplification, 5 - 10 µl of PCR product were subjected to restriction digestion reaction after adding 2.5 µl of restriction enzyme buffer, 0.25 units of restriction enzyme and sufficient sterile distilled water to make up a 25 µl reaction solution. The mixture was left for three hours or overnight at a temperature required for complete digestion. The resulting fragments, together with a standard DNA size marker (1 kb DNA ladder), were separated in 1.4 - 4% agarose gel, stained with ethidium bromide and visualised under UV light. The presence/absence of a restriction site was coded as the character state for each mutation for phylogenetic analysis. The computer program PAUP (Swofford, 1993) was used to reconstruct the phylogeny using Fitch parsimony (unweighted characters). Heuristic searches were conducted with the RANDOM addition sequence strategy using 100 replicates and TBR branch swapping.

4.2.2.2 Restriction site variation of total cpDNA genome

Total DNA of each *Cryptocoryne* species was extracted using the CTAB method, and was purified by passing extracts down caesium chloride gradients as described in chapter 2. 500 ng of DNA were then digested with 6 units of restriction enzyme according to the manufacturer's instructions. The DNA fragments produced, were separated in 1% (for 6 base cutter enzyme) or 1.4% (for 4 base cutter enzyme) agarose gels. The separated

Table 4.1. Description of 11 pairs of primers for amplication of non coding regions of plant cpDNA.

No	Primer 1	Primer 2	Source
1	*trn*K[tRNA-His(GUG)] 5'-ACGGGAATTGAACCCGCGA-3'	*trn*K[tRNA-Lys(UUU)exon1] 5'-CCGACTAGTTCCGGGTTCGA-3'	Demesure *et al.*, 1995
2	*trn*K[tRNA-Lys(UUU)exon1] 5'-GGGTTGCCCGGGACTCGAAC -3'	*trn*K[tRNA-Lys(UUU)exon2] 5'-CAACGGTAGAGTACTCGGCTTT TA-3'	Demesure *et al.*, 1995
3	*trn*C[tRNA-Cys(GCA)] 5'-CCAGTTCAAATCTGGGTGTC -3'	*trn*D[tRNA-Asp(GUC)] 5'-GGGATTGTAGTTCAATTGGT-3'	Demesure *et al.*, 1995
4	*trn*D[tRNA-Asp(GUC)] 5'-ACCACTTGAACTACAATCCC -3'	*trn*T[tRNA-Thr(GGU)] 5'-CTACCACTGAGGTTAAAAGGG -3'	Demesure *et al.*, 1995
5	*psb*C[psII 44kd protein] 5'-GGTCGTGACCAAGAAACCAC -3'	*trn*S[tRNA-Ser(UGA)] 5'-GGTTCGAATCCCTCTCTCTC-3'	Demesure *et al.*, 1995
6	*trn*S[tRNA-Ser(UGA)] 5'-GAGAGAGAGGGATTCGAACC -3'	*trn*fM[tRNA-fMet(CAU)] 5'-CATAACCTTGAGGTCACGGG-3'	Demesure *et al.*, 1995
7	*psb*A[PSI (P700 apoprotein A1)] 5'-ACTTCTGGTTCCGGCGAACGAA- 3'	*trn*S[tRNA-Ser(GGA)] 5'-AACCACTCGGCCATCTCTCCTA -3'	Demesure *et al.*, 1995
8	*trn*S[tRNA-Ser(GGA)] 5'-CGAGGGTTCGAATCCCTCTC -3'	*trn*T[tRNA-Thr(UGU)] 5'-AGAGCATCGCATTTGTAATG-3'	Demesure *et al.*, 1995
9	*trn*M[tRNA-Met(CAU)] 5'-TGCTTTCATACGGCGGGAGT -3'	*rbc*L[RuBisCO large subunit] 5'-GCTTTAGTCTCTGTTTGTGG-3'	Demesure *et al.*, 1995
10	*trn*T(UGU) 5'-CATTACAAATGCGATGCTCT-3'	*trn*L(UAA)3' exon 5'-GGGGATAGAGGGACTTGAAC -3'	Taberlet *et al.*, 1991
11	*trn*L(UAA)5'exon 5'-CGAAATCGGTAGACGCTACG -3'	*trn*F(GAA) 5'-ATTTGAACTGGTGACACGAG -3'	Taberlet *et al.*, 1991

fragments were transferred from agarose gels on to nylon membranes by Southern blotting (Southern, 1975). The gels were trimmed, photographed and immersed in denaturation buffer (1.5 M NaCl; 0.5 M NaOH) for 30 mins to denature the DNA fragments in the gel. After 30 mins, gels were rinsed with distilled water and immersed in neutralisation buffer (1.5 M NaCl; 0.5 Tris-HCl pH7.2; 1 mM EDTA-Na_2) for another 30 mins. Gels were then placed on a blotting apparatus (as illustrated in Nicholl, 1994 pp. 118), and were left overnight to allow DNA transfer from the gel onto the nylon membrane using 20x SSC (3 M NaCl; 0.3 M trisodium citrate) as the transfer buffer. The next day, the nylon filters were rinsed in 2x SSC and left to air dry. DNA fragments were bound to the nylon membrane by exposure to ultra-violet radiation for 30 seconds. The DNA fragments of interest on the nylon membrane were visualised by hybridisation with labelled probes. During the course of this research two methods of labelling probes were used. Initially DNA probes were radio-labelled with ^{32}P, but later, a chemilluminescent labelling procedure was used. Both of these methods are described below.

4.2.2.2.1 Use of radiolabelled cpDNA probes

Probes were radio-labelled using the random oligomer method of Feinberg and Vogelstein (1983). Thus, for a single membrane, 60 ng of probe DNA was made up with distilled water to a final volume of 20 µl and was boiled for 5 mins in a water bath to denature the probe. Denatured probes were immediately placed on ice to prevent the separated DNA strands from reannealing. The following reagents were later added: 5 µl 1 M HEPES pH6.6; 5 µl DTM (100 µM of each dATP, dGTP and dTTP, which

is made up in TM - 250 mM Tris-HCL pH8.0; 25 mM $MgCl_2$; 50 mM b-mercaptoethanol); 1.4 µl OL (1 mM Tris-HCL pH7.5; 1 mM EDTA; 90 OD units/ml hexaoligodeoxyribonucleotide); 1 µl 10 mg/ml BSA; 2.5 units of Klenow (DNA polymerase); 10 µCi ^{32}P-dCTP. The reaction mixture was left to stand at room temperature for five hours to allow radiolabelling to take place.

Before hybridisation, the nylon membrane was subjected to a prehybridisation treatment. This involved placing it in a hybridisation tube with 50 ml buffer III (0.6 M NaCl; 10 mM PIPES pH6.8; 1 mM EDTA; 10x modified Denhardt's solution, made up with 0.2% (w/v) gelatine; 0.2% (w/v) ficoll-400; 0.2% (w/v) PVP-360; 1% SDS; 0.05% tetrasodium pyrophosphate) and 10 µg/ml denatured salmon testes DNA. Hybridisation tubes were left to rotate in a hybridisation oven (Techne hybridiser HB1) at 65°C for six hours.

After DNA probes were radio-labelled and the prehybridisation treatment was done, the probes were boiled for 5 mins to ensure that they were single stranded before being injected into the hybridisation tubes containing the nylon membranes with a 25-gauge hypodermic needle and syringe. Membranes were left overnight at 65°C for hybridisation to take place. They were then removed from the hybridisation tubes and washed twice in 2x SSC and 0.5% SDS at room temperature for 30 mins followed by one wash of 2x SSC and 0.5% SDS at 65°C for 30 mins.

The membranes were prepared for autoradiography by covering them

with plastic sheeting to prevent them from drying out. They were then placed in autoradiography cassettes with X-ray film placed on top of them and left at -70°C for the length of time required. Autoradiographs were developed using a developing machine (RGII Fuji X-ray film processor).

Once autoradiographs were developed, nylon membranes were stripped of radioactive probe by immersing them in boiling 0.1% SDS. They were then left to cool at room temperature before being checked for radiation. Membranes with some traces of radiation on them were subjected to another round of the same stripping procedure until all traces of radioactivity had been removed. They were then sealed in plastic bags and stored at -20°C until another probing was required.

4.2.2.2.2 Use of chemiluminescent labelled cpDNA probe

Chemiluminescent labelled cpDNA probes were produced following the protocols described by Hoisington *et al.* (1994) and by the manufacturer Boehringer. Probes were labelled using the random hexamer labelling procedure, with DIG-dUTP, and used in conjunction with CSPD, a chemiluminescent substrate, which emits light after an enzymatic reaction has taken place. The labelling procedure was as follows. Approximately 120 ng (for the first membrane with additional 60 ng for each additonal membranes used) of probe DNA was made up with water to a final volume of 13.5 µl and was boiled for 10 mins. DNA probe was immediately quenched on ice to prevent it from reannealing before adding 2 µl of Klenow buffer (10x 100 mM Tris-HCl pH8.5, 100 mM

$MgCl_2$), 2 µl OL (1 mM Tris-HCl pH7.5; 1 mM EDTA; 90 OD units/ml hexaoligodeoxyribonucleotide) 2 µl DIG labelling mix (1 mM dATP, 1 mM dGTP, 1 mM dCTP, 0.65 mM dTTP, 0.35 mM DIG-dUTP pH7.5), and 2.5 units Klenow (DNA polymerase).

For prehybridisation treatment, up to 5 nylon membranes were placed in a hybridisation tube with a mesh in between each membranes with 28 µl (20 µl for the first membrane and additional 2 ml for each additional membranes) of prehybridisation buffer (5x SSC; 0.1% Na-Sarkosyl; 0.02% SDS; 0.5% Blocking stock which is made up of 10% blocking reagent (Boehringer Mannheim) in buffer 1[0.1 M Maleic acid; 0.15 M NaCl]). Prehybridisation was carried out for two hours in a hybridisation oven at 65°C.

Labelled DNA probe was purified before use by first precipitating it with 2.5 µl of 4 M LiCl and 75 µl of cold 96% ethanol left at -70°C for 30 mins before spinning down briefly and discarding the supernatant. It was then washed with 50 µl of cold 70% ethanol, before spinning down again and discarding the supernatant. Finally the probe was dissolved in 50 µl TE and left for 30 mins at 37°C.

Purified probe was boiled for 5 mins and immediately placed on ice before adding 14 ml (10 ml for the first membrane and an additional 1 ml for each additional membranes) of prewarmed (65°C) hybridisation buffer (made up with the same ingredients as the prehybridisation buffer). The prehybridisation buffer was poured out of the hybridising tube before the mixture of hybridisation buffer and labelled probe was poured in. The membranes were left to hybridise overnight at 65°C.

After hybridisation, membranes were removed from the tube and washed first in low stringency washing solution (2x SSC and 0.1% SDS) for five min twice. A second wash involved a medium stringency solution (1x SSC and 0.1% SDS) carried out twice for 15 min at 65°C each wash. Membranes were rinsed in a mixture of buffer 1 and 0.3% Tween 20 before adding 25 ml/membrane buffer 2 (a 1:10 dilution of blocking stock in buffer 1) and left on a shaker (100 rpm) at room temperature for 30 mins. For each membrane 1.7 µl anti-DIG in 25 ml buffer was added and left for 30 mins at room temperature.

Membranes were washed twice in a mixture of buffer 1 and 0.3% Tween 20 (250 ml/membrane) for 15 mins each wash. Another wash was carried out in 20 ml/membrane buffer 3 (0.1 M Tris-HCl pH9.5; 0.1 M NaCl; 50 mM $MgCl_2$) for 5 mins. 25 ml of CSPD solution was added to the membranes, which were left for 5 mins.

The membranes were prepared for autoradiography by wrapping each of them in plastic sheeting to prevent them from drying out and incubating them at 37°C for up to 15 mins to enhance the luminescence. Membranes were placed into autoradiograph cassettes with X-ray films placed on top and exposed from 15 mins up to 16 hours at room temperature before the films were developed.

After autoradiographs had been developed, the labelled probe was stripped off the membranes by washing twice for 15 mins at 37°C in a mixture of 0.2 M NaOH and 0.1% SDS. Membranes were rinsed in 2x SSC before storing in a sealed plastic bag at -20°C until required for another probing.

4.2.3 Direct sequencing of a segment of the cpDNA genome

The cpDNA segment that was sequenced for analysis was the 5' *trn*K intron and the first 141 bp of the *mat*K gene (Fig. 4.2). This segment was amplified using the primers *trn*K3914F (developed by Jerry Learn - see Johnson and Soltis, 1995) and *mat*K122SR (developed by the author). The PCR reaction mixture contained 10 µl of 10x *Taq* buffer, 3.5 mM $MgCl_2$, 10 µl of 2 mM dNTP in equimolar ratio, 100 ng of each primer, 25-100 ng DNA template, 2.5 units of *Taq* polymerase and was made up to 100 µl total volume with sterile distilled water. Amplification was carried out using a Techne PHC-3 thermal cycler with 26 cycles of 1 min at 94°C, 1 min at 50°C, 3 mins at 72°C and 7 mins at 72°C. The amplification product was cleaned before cycle sequencing was carried out as described in chapter 3.

DNA sequences were aligned by eye and required interpretation of gaps scattered within the 5' *trnK* intron and part of the sequenced *mat*K gene. Gaps were scored as additional presence/absence characters and included in the sequence data matrix for phylogenetic analysis (Oxelman and Liden, 1995).

Phylogenetic trees were generated from the sequences using the computer program PAUP (Swofford, 1993) with characters weighted equally (Fitch parsimony). Heuristic searches were conducted with the RANDOM addition sequence strategy using 100 replicates and TBR branch swapping. Saving all minimal trees (MULPARS) and accelerated transformation (ACCTRAN) were used in conjunction with the addition sequence regime. Branch zero length were collapsed to reduce the number of most parsimonious tree. Descriptive statistics that reflect the amount of

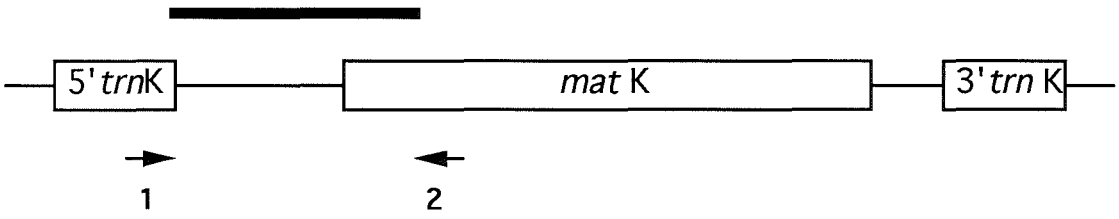


Figure 4.2. Relative location of amplification and sequencing primers used for sequencing the 5' end of the *trn*K intron and a part of *the mat*K gene sequence. Arrow labelled 1 represents the primer *trn*K-3914F (monocot) (5'-ATCTGGGTTGCTAACTCAATGG-3') designed by Jerry Learn, while arrow labelled 2 represents the primer *mat*K-122SR (5'-TGAAAGAGAAGCGGGTA-3') designed by the author. The shaded region above the figure indicates the region sequenced for species included in this study.

phylogenetic signal in the parsimony analysis were computed as the consistency index (CI: Kluge and Farris, 1969) and retention index (RI: Swofford, 1993). Sets of equally parsimonious trees were summarised as a 50% majority rule consensus tree and a strict consensus tree. Bootstrap analysis (Felsenstein, 1985) was conducted using 100 replicates of heuristic searches with random addition sequences. A maximum of 1000 trees were saved during each bootstrap replicate to reduce the analysis time.

4.3 Results

4.3.1 Restriction site variation within a PCR amplified cpDNA region

Initial amplification was successful for three of the eleven sets of primers used. A product of 3100 bp was amplified using primers *trn*M[tRNA-Me(CAU)] and *rbc*L[RuBisCO large subunit], one of 1550 bp was produced using the primer pair *psb*C[psII 44kd protein] and *trn*S[tRNA-Ser(UGA)] while another of 2500 bp was amplified with primers *trn*K[tRNA-Lys(UUU)exon1] and *trn*K[tRNA-Lys(UUU)exon2]. Most of the other primer sets either failed or produced only a very faint PCR product or multiple PCR products (primer pair *trn*L(UAA)5'exon and *trn*F(GAA) differed in this respect by producing multiple PCR products - Plate 4.1).

A total of 26 restriction enzymes (Table 4.2) were used to digest each of the first two amplification products (i.e. the 3100 bp and the 1550 bp PCR products). However, for the 1550 bp products, only five of the enzymes produced restriction site variation among the *Cryptocoryne* species and the two *Lagenandra* species examined. From this a total of seven character changes were recorded, all of which were unique to particular taxa. Restriction site variation was detected with seven of the 26 restriction enzymes used to digest the 3100 bp amplified product (e.g. in Plate 4.2), yielding 10 character changes of which four were autapomorphic. Eight restriction enzymes (Table 4.2) were used to digest the 2500 bp PCR product amplified by the primer pair *trn*K[tRNA-Lys(UUU)exon1] and *trn*K[tRNA-Lys(UUU)exon2], and six of these resolved variation (e.g. in Plate 4.3). A total of 12 character changes were recorded, four of which were unique to particular taxa. Owing to the low number of synapomorphic characters obtained, further analysis using additional restriction enzymes was not

Plate 4.1. Multiple amplification products obtained when using primers *trn*L(UAA)5'exon and *trn*F(GAA). Two species were used in this preliminary amplification *C. ciliata* P1958/6013 (lane 1-3, 7-9, 13-15) and *C. minima* S1995/9201 (lane 4-6, 10-12, 16-18). Several combinations of different amounts DNA and $MgCl_2$ concentrations were tried in the amplification reactions. Three different DNA concentrations were used, 25ng (lane 1, 4, 7, 10, 13, 16), 50ng (lane 2, 5, 8, 11, 14, 17) and 75ng (lane 3, 6, 9, 12, 15, 18). Three different $MgCl_2$ concentrations were used; 2mM (lane 1-6), 2.5mM (lane 7-12) and 3 mM (lane 13-18).

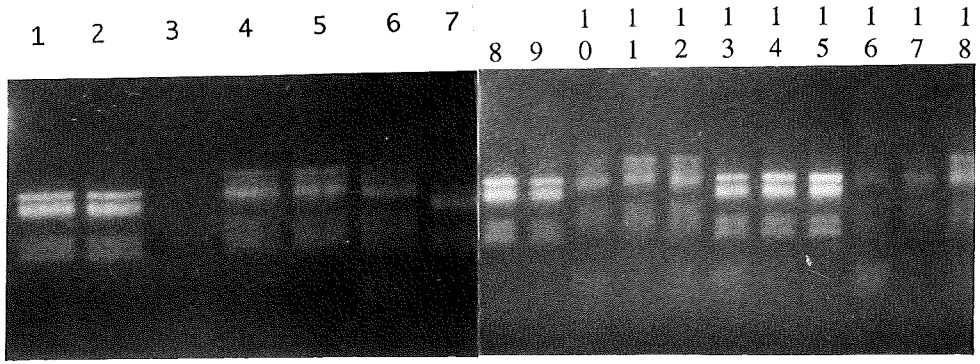
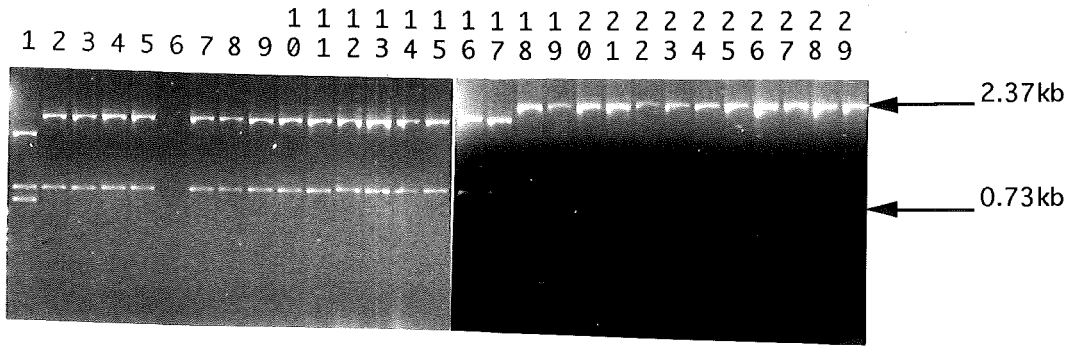


Table 4.2.Details of restriction enzymes used in the PCR-RFLP analyses of chloroplast DNA.

Restriction enzyme	Recognition site (5'-3')	Incubation temperature
Alu I^	AG*CT TC*GA	37°C
Bam HI^	G*GATCC CCTAG*G	37°C
Bcl I^	T*GATCA ACTAG*T	50°C
Bgl II	A*GATCT TCTAG*A	37°C
Bst EII	G*GTNACC CCANTG*G	60°C
Bst OI	CC*(A/T)GG GG(T/A)*CC	60°C
Cfo I	GCG*C C*GCG	37°C
Cla I^	AT*CGAT TAGC*TA	37°C
Dra I	TTT*AAA AAA*TTT	37°C
Eco RI^	G*AATTC CTTAA*G	37°C
Eco RV	GAT*ATC CTA*TAG	37°C
Hae III	GG*CC CC*GG	37°C
Hind III^	A*AGCTT TTCGA*A	37°C
Hinf I	G*ANTC CTNA*G	37°C
Hpa II	CC*GG GG*CC	37°C
Kpn I	GGTAC*C C*CATGG	37°C
Mlu I	A*CGCGT TGCGC*A	37°C
Nru I	TCG*CGA AG*CGCT	37°C
Nsi I	ATGCA*T T*ACGTA	37°C
Pst I	CTGCA*G G*ACGTC	37°C
Pvu II	CAG*CTG GTC*GAC	37°C
Rsa I^	GT*AC CA*TG	37°C
Sma I	CCC*GGG GGG*CCC	25°C
Taq I	T*CGA AGC*T	65°C
Xba I	T*CTAGA AGATC*T	37°C
Xho I^	C*TCGAG GAGCT*C	37°C

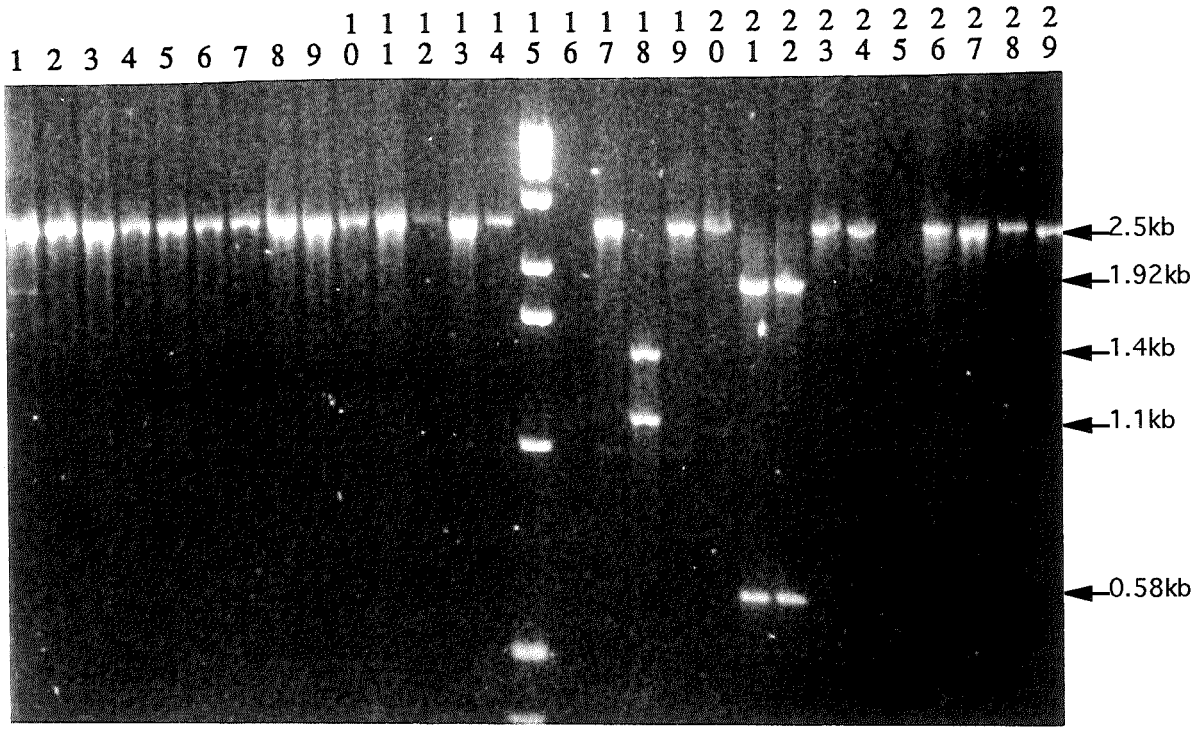
Note: ^ denotes the eight restriction enzymes used to digest the 2500 bp PCR product. The cutting sites are indicated by the asterick (*).

Plate4.2. Restriction site variation within the chloroplast DNA fragment amplified using primer pair *trn*M[tRNA-Met(CAU)] and *rbc*L[RuBisCo large subunit] digested with BstO1.



from left, 1. *C. ciliata* P1958/6013, 2. *C. spiralis* NJ3129a2, 3. *C. beckettii* 1671, 4. *C. walkeri* NJ23-3, 5. *C. wendtii* P1961/5342, 6. *C. undulata* NJ22-7, 7. C. *x willisii* P1978/5045, 8. *C. x willisii* Kew 3790, 9. *C. cognata* 28Yadav, 10. *C. pontederiifolia* USM9638, 11. *C. moehlmannii* P1989/5046, 12. *C. longicauda* USM9439, 13. *C. minima* S1995/9201, 14. *C. griffithii* NJ85-30, 15. *C. purpurea* Othman s.n., 16. *C. cordata* USM9139, 17. *C. zonata* W534, 18. *C. elliptica* USM8069, 19. *C. schulzei* USM8087, 20. *C. affinis* USM8065, 21. *C. aponogetifolia* P3401, 22. *C. usteriana* P1983/5448, 23. *C. annamica* M92/3205, 24. *C. versteegii* P1952/5376, 25. *C. alba* NJ3172-6, 26. *C. albida* P1958/5363, 27. *C. retrospiralis* P1977/5146, 28. *C. crispatula* var. *balansae* NJ3406, 29. *Lagenandra meeboldi* P1979/5019.

Plate 4.3. Restriction site variation within chloroplast DNA fragment amplified using primer pair *trn*K[tRNA-Lys(UUU)exon1] and *trn*K[tRNA-Lys(UUU)exon2] digested with *Xho*1.



from left, 1. *C. ciliata* P1958/6013, 2. *C. spiralis* NJ3129a2, 3. *C. beckettii* 1671, 4. *C. walkeri* NJ23-3, 5. *C. wendtii* P1961/5342, 6. *C. undulata* NJ22-7, 7. *C. x willisii* P1978/5045, 8. *C. x willisii* Kew 3790, 9. *C. cognata* 28Yadav, 10. *C. pontederiifolia* USM9638, 11. *C. moehlmannii* P1989/5046, 12. *C. longicauda* USM9439, 13. *C. minima* S1995/9201, *14. C. griffithii* NJ85-30, 15. 1kb standard, 16. *C. cordata* USM9139, 17. *C. zonata* W534, 18. *C. elliptica* USM8069, 19. *C. schulzei* USM8087, 20. *C. affinis* USM8065, 21. *C. aponogetifolia* P3401, 22. *C. usteriana* P1983/5448, 23. *C. annamica* M92/3205, 24. *C. versteegii* P1952/5376, 25. *C. alba* NJ3172-6, 26. *C. albida* P1958/5363, 27. *C. crispatula* var. *balansae* NJ3406, 28. *Lagenandra ovata* P1983/5653, 29. *L. meeboldi* P1979/5019.

conducted.

The presence/absence of restriction sites for the 27 *Cryptocoryne* species (including two accessions of *C. x willisii*) and two species of *Lagenandra* are listed in Table 4.3, with details of restriction sites given in Table 4.4. 174 most parsimonious phylogenetic trees with 33 steps were generated from the data set and one of these parsimonious trees is illustrated in Fig. 4.3. It is evident that the tree, is largely unresolved and uninformative due to the low level of synapomorphic characters. Consequently it contains little phylogenetic information of value.

4.3.2 Restriction site variation across the entire cpDNA genome

Cryptocoryne DNA was completely digested by all restriction enzymes used in this study. However, hybridisation of *Lactuca sativa* cpDNA probes to *Cryptocoryne* DNA fragments on the nylon membranes was not very successful, and even when success was achieved, the interpretation of the resulting bands was very difficult. For example, hybridisation of cpDNA probes pLsc 7 and pLsc 8 to *Cryptocoryne* DNA fragments, resulted in fragments very much larger than the size of the pLsc 7 (7.7kb) and pLsc 8 (7.0) probes used respectively (see Plate 4.4, Plate 4.5 and Plate 4.6). There was little difference between the quality of autoradiograph produced using chemiluminescent or radiolabelled probes, although a much shorter X-ray film exposure time was required for chemiluminescent probes.

A major disadvantage of the technique was the substantial amount of DNA required for each restriction digest. Initially 500 ng of total DNA was

Table 4.3 Restriction site data of the amplified cpDNA regions used for the construction of *Cryptocoryne* phylogenetic trees. Numbers at the top of each column refer to characters listed in table 4.3. The binary data represent present (1) and absent (0) of site mutations. Missing data are denoted by '?'. Numbers in parentheses after taxon names are the accession numbers of each taxon.

Taxa	Character 1111111112222222222 12345678901234567890123456789
C. ciliata (P1958/6013)	1010000001101010000110100 1001
C. spiralis (NJ3129a2)	1000000001011010000111100 1001
C. beckettii (1671)	1000000101011010000100100 1001
C. walkeri (NJ23-3)	1000000101011011000110100 1001
C. wendtii (P1961/5342)	1000000101011011000110100 1001
C. undulata (NJ22-7)	1000000101011011000110100 1001
C. x willisii (P1978/5045)	1000000101011011000110100 1001
C. x willisii (Kew 3790)	1000000101011010000110100 0001
C. cognata (28 Yadav)	1000000001011010000110100 1001
C. pontederiifolia (USM9638)	1000000101011010000110100 0101
C. moehlmannii (P1989/5046)	1000000101011010000110100 0101
C. longicauda (USM9439)	1000000101011010000110100 1000
C. minima (S1995/9201)	1000000101011010000110000 1001
C. griffithii (NJ85-30)	1000000101011010000110000 1001
C. purpurea (Othman s.n.)	0100000001011010?????????????
C. cordata (USM9139)	1000000101011010?????????????
C. zonata (W534)	1000100100001010100110100 0011
C. elliptica (USM8069)	1000000100001010000110101 1001
C. schulzei (USM8087)	1000000100001010000110100 0011
C. affinis (USM8065)	1000000100001010000110100 1000
C. aponogetifolia (P3401)	1000000100001010000100110 1001
C. usteriana (P1983/5448	1000000100001010000100110 1001
C. annamica (M92/3205)	1001010110001010000110100 1001
C. versteegii (P1952/5376)	1000000100001000011110100 1001
C. alba (NJ3172-6)	1000000100000110000110100 1001
C. albida (P1958/5363)	1000000100001000000110100 1001
C. retrospiralis (P1977/5146)	100000010000111??????????????
C. crispatula var. *balansae* (NJ3406)	1000001100001010000110100 1001
Lagenandra ovata (P1983/5653)	1000000??????????000101001001
L. meeboldi (P1979/5019)	1000000100001010000010100 1001

Table 4.4. Details of the restriction site characters included in Table 4.3. These sites have been scored as either 1 (present) or 0 (absent) in Table 4.3. Numbers in parentheses under the heading primer pair denote the primer number in Table 4.1. A question mark under the heading mutations indicated that a particular fragment(s) was not detected.

Character	Primer pair	Enzyme	Mutation (bp)
1	*psb*C-*trn*S (5)	*Cfo* I	1550 -> 1190 + 360
2	*psb*C-*trn*S (5)	*Cfo* I	1550-> 850 + 700
3	*psb*C-*trn*S (5)	*Hae* III	1215 -> 1100 + ?
4	*psb*C-*trn*S (5)	*Hinf* I	215 -> 170 + ?
5	*psb*C-*trn*S (5)	*Rsa* I	1200 -> 755 + 445
6	*psb*C-*trn*S (5)	*Taq* I	1360 -> 1100 + 260
7	*psb*C-*trn*S (5)	*Taq* I	1360 -> 735 + 625
8	*trn*M-*rbc*L (9)	*Bcl* I	3100 -> 1635 + 1465
9	*trn*M-*rbc*L (9)	*Eco* RI	3100 -> 2570 + 530
10	*trn*M-*rbc*L (9)	*Bst* EIII	3100 -> 2375 + 725
11	*trn*M-*rbc*L (9)	*Bst* EIII	2370 -> 1720 + 610
12	*trn*M-*rbc*L (9)	*Bst* OI	3100 -> 2370 + 730
13	*trn*M-*rbc*L (9)	*Bst* OI	2370 -> 1730 + 640
14	*trn*M-*rbc*L (9)	*Dra* I	3100 -> 2780 + 320
15	*trn*M-*rbc*L (9)	*Hpa* I	1440 -> 915 + 525
16	*trn*M-*rbc*L (9)	*Rsa* I	270 -> 195 + ?
17	*trn*K-*trn*K (2)	*Rsa* I	475 -> 420 + ?
18	*trn*K-*trn*K (2)	*Alu* I	1650-> 1190+ 460
19	*trn*K-*trn*K (2)	*Alu* I	580 -> 552 + ?
20	*trn*K-*trn*K (2)	*Bgl* II	1900 -> 1700 + 200
21	*trn*K-*trn*K (2)	*Bgl* II	800 -> 600 + 200
22	*trn*K-*trn*K (2)	*Hind* III	1500 -> 850 +650
23	*trn*K-*trn*K (2)	*Hind* III	920 -> 740 + 180
24	*trn*K-*trn*K (2)	*Xho* I	2500 -> 1920 + 580
25	*trn*K-*trn*K (2)	*Xho* I	2500 -> 1440 + 1100
26	*trn*K-*trn*K (2)	*Eco* RI	1820 -> 1580 + 240
27	*trn*K-*trn*K (2)	*Eco* RI	1820 -> 1210 + 610
28	*trn*K-*trn*K (2)	*Eco* RI	344 -> ?
29	*trn*K-*trn*K (2)	*Bcl* I	980 -> 800 + 180

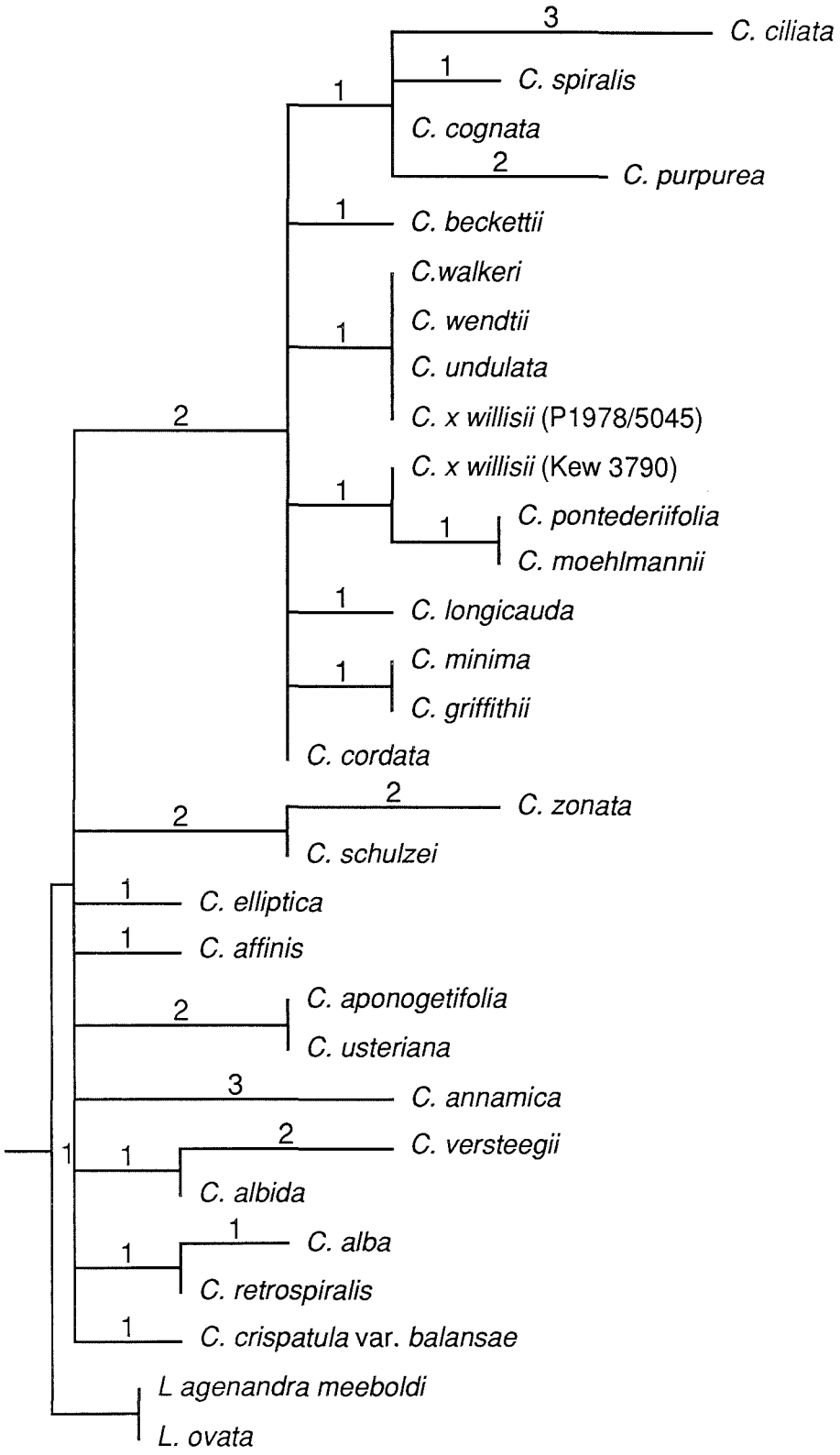
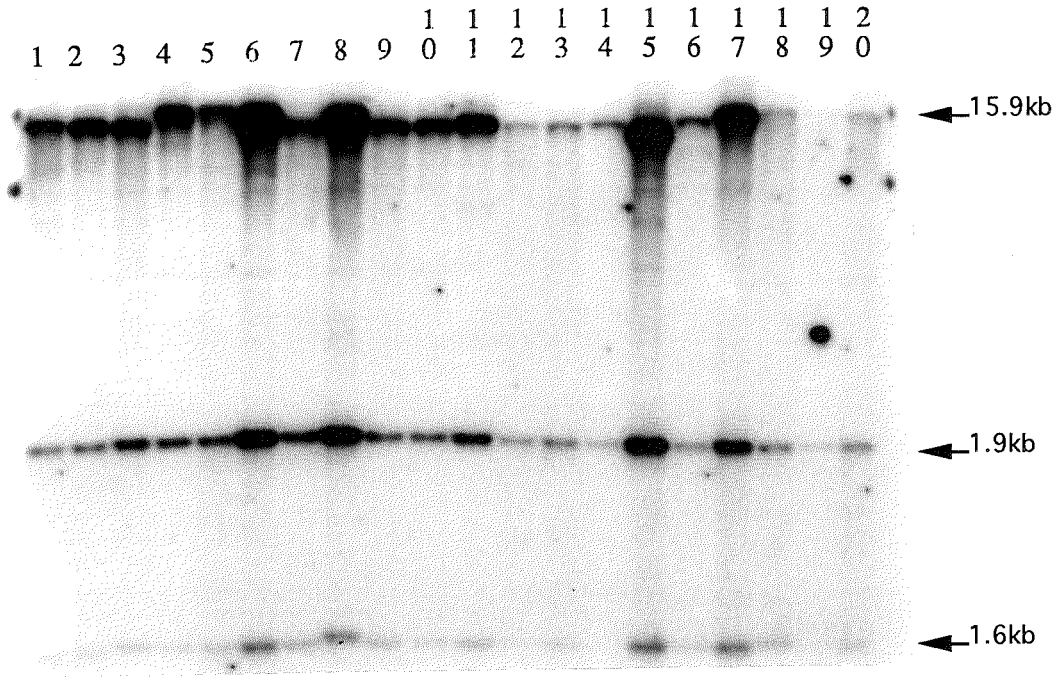


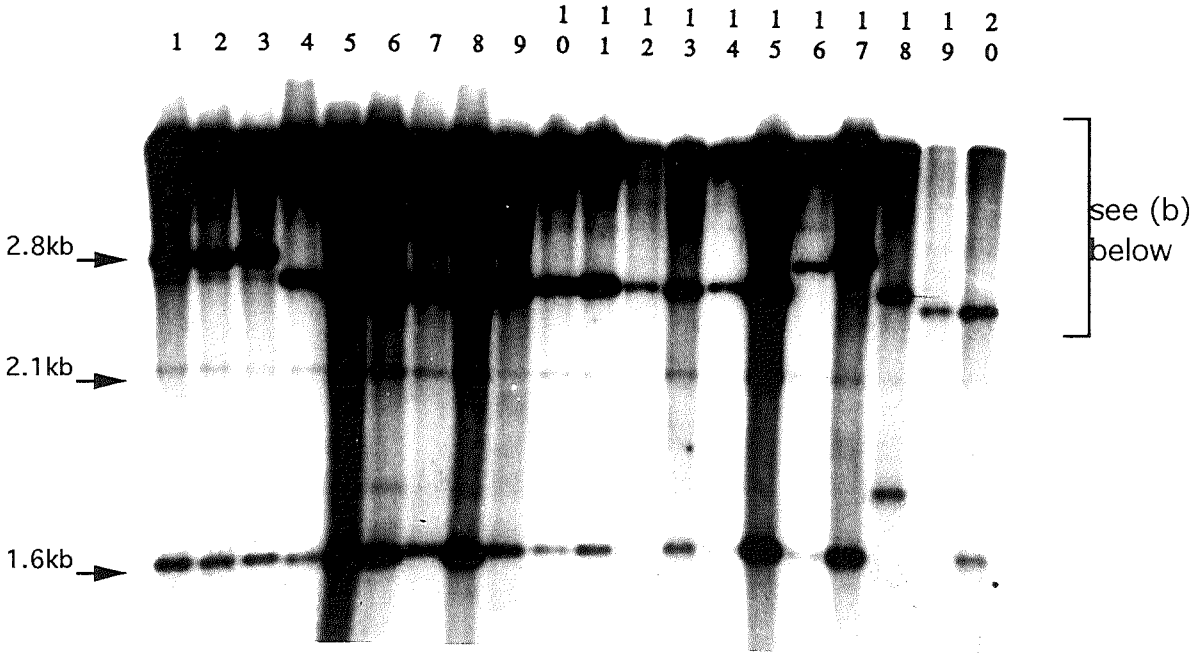
Figure 4.3. One of 174 equally parsimonious trees with 33 steps (CI=0.879; RI=0.911) generated from combined restriction site data from the amplified cpDNA regions. The trees were unresolved due to lack of synapomorphic characters.

Plate 4.4 *Cryptocoryne* chloroplast DNA digested with *Eco*RI and probed with pLsc7.

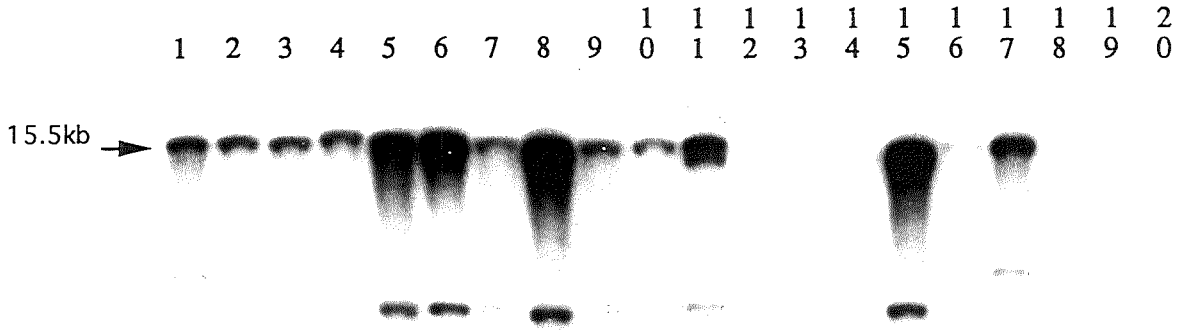


from left 1. *C. spiralis* NJ3125, 2. *C. spiralis* NJ3129a2, 3. *C. spiralis* NJ327-73c, 4. *C. beckettii* NJ2898, 5. *C. beckettii* 1671, 6. *C. beckettii* P1976/5044, 7. *C. walkeri* NJ23-3, 8. *C. walkeri* ADG315, 9. *C. wendtii* P1961/5342, 10. *C. wendtii* NJ2781, 11. *C. wendtii* P1974/5024, 12. *C. x willisii* NJ24-3, 13. *C. x willisii* NJ23-25, 14. *C. x willisii* P1978/5045, 15. *C. undulata* P1976/5045, 16. *C. cognata* 28Yadav, 17. *C. moehlmannii* P1989/5046, 18. *C. longicauda* USM9439, 19. *C. cordata* USM9139, 20. *C. cordata* USM9140.

Plate 4.5 a & b. *Cryptocoryne* chloroplast DNA digested with *Eco*RI and probed with pLsc8.



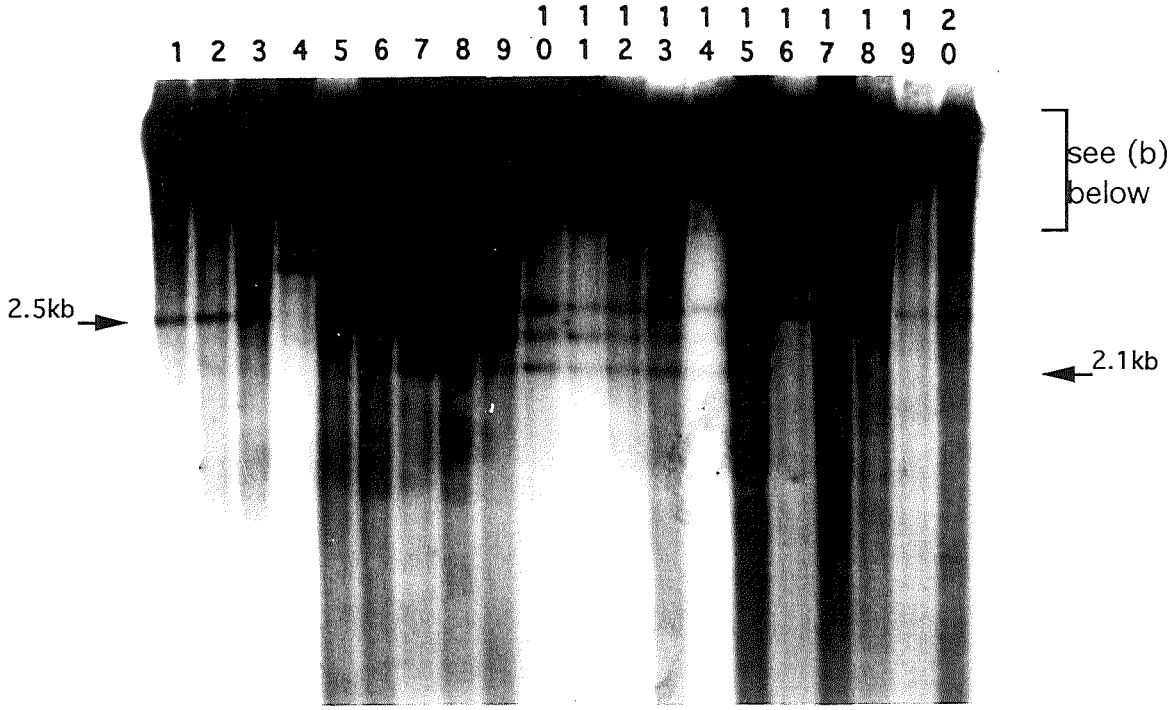
a



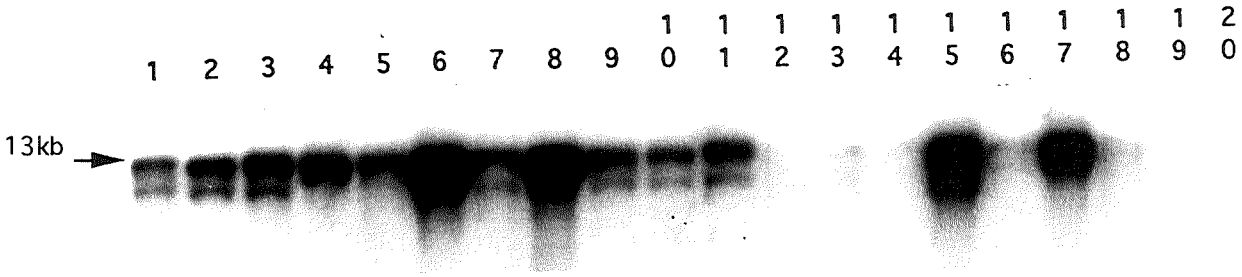
b

from left 1. *C. spiralis* NJ3125, 2. *C. spiralis* NJ3129a2, 3. *C. spiralis* NJ327-73c, 4. *C. beckettii* NJ2898, 5. *C. beckettii* 1671, 6. *C. beckettii* P1976/5044, 7. *C. walkeri* NJ23-3, 8. *C. walkeri* ADG315, 9. *C. wendtii* P1961/5342, 10. *C. wendtii* NJ2781, 11. *C. wendtii* P1974/5024, 12. *C. x willisii* NJ24-3, 13. *C. x willisii* NJ23-25, 14. *C. x willisii* P1978/5045, 15. *C. undulata* P1976/5045, 16. *C. cognata* 28Yadav, 17. *C. moehlmannii* P1989/5046, 18. *C. longicauda* USM9439, 19. *C. cordata* USM9139, 20. *C. cordata* USM9140.

Plate 4.6 a & b. *Cryptocoryne* chloroplast DNA digested with *Bam*HI and probed with pLsc8.



a



b

from left 1. *C. spiralis* NJ3125, 2. *C. spiralis* NJ3129a2, 3. *C. spiralis* NJ327-73c, 4. *C. beckettii* NJ2898, 5. *C. beckettii* 1671, 6. *C. beckettii* P1976/5044, 7. *C. walkeri* NJ23-3, 8. *C. walkeri* ADG315, 9. *C. wendtii* P1961/5342, 10. *C. wendtii* NJ2781, 11. *C. wendtii* P1974/5024, 12. *C. x willisii* NJ24-3, 13. *C. x willisii* NJ23-25, 14. *C. x willisii* P1978/5045, 15. *C. undulata* P1976/5045, 16. *C. cognata* 28Yadav, 17. *C. moehlmannii* P1989/5046, 18. *C. longicauda* USM9439, 19. *C. cordata* USM9139, 20. *C. cordata* USM9140.

used per digest (i.e. an equivalent amount to that used in a study on *Stylosanthes* by Gillies, 1994). Moreover, despite using the same amount of total DNA for all taxa, the intensity of the resulting hybridising DNA fragments visualised sometime varied between taxa. This variation could stem from an error in estimating accurately DNA concentrations. It is feasible that in taxa where fragment intensity was reduced, the proportion of cpDNA to nuclear DNA might be lower than normal. It is known that the proportion of cpDNA to nuclear DNA in total DNA extraction varies from taxon to taxon and tissue to tissue (Olmstead and Palmer, 1994).

Due to the mixed results obtained, the difficulty experienced in interpreting the bands observed, and the large amounts of DNA required for analysis, it was decided not to continue with this form of methodology after the initial work had been completed.

4.3.3 Direct sequence analysis of part of the cpDNA genome

The nucleotide composition of the entire 5' *trn*K intron together with the first 141bp of the *mat*K gene sequence for *Cryptocoryne* species and the outgroup species of *Lagenandra* are listed in Table 4.5. The length of the 5' *trn*K intron in *Cryptocoryne* species ranged from 680bp (*C. affinis)* to 708 bp (*C. annamica*), while in *Lagenandra* it was 683bp. Ambiguous nucleotides were encountered at two positions, one at position 459 in *C. cognata* (coded as R, i.e. either A or G) and the other at position 862 in *C. retrospiralis* (coded as Y, i.e. C or T). These ambiguous nucleotide states for each taxon were coded as missing data for phylogenetic analysis.

A data matrix of 743 characters was necessary to align the sequence of the

Table 4.5. Aligned DNA sequences of the 5' *trn*K intron together with a part of *mat*K genic region of chloroplast DNA from 27 representatives of *Cryptocoryne* species and from one outgroup species of *Lagenandra*.

Taxa[a]	Nucleotide sites[b]
	1 2 3 4 5 6 0 0 0 0 0 0 . . . . . .
1	GCGTAT-ATCTTT-ACACATTGGGATGAAGCAACAAATTCGTCCAGACTTTTGGTAGAGT
2	GGCTATAATCTTTTACACATTTGGATGAAGCAACAAATTCGTCCAGACTTTTGGTAGAGT
3	CGCTATAATCTTTTACACATTTGGATGAAGCAACAAATTCGTCCAGACTTTTGGTAGAGT
4	GGCTATAATCTTT-ACACATTTGGATGAAGCAACAAATTCGTCCAGACTTTTGGTAGAGT
5	GGCTAT-ATCTTT-ACACATTTGGATGAAGCAACAAATTCGTCCAGACTTTTGGTAGAGT
6	G--TATAATCTTTTACACATTTGGATGAAGCAACAAATTCGTCCAGACTTTTGGTAGAGT
7	CGCTATAATCTTTTACACATTGGGATGAAGCAACAAATTCGTCCAGACTTTTGGTAGAGT
8	GG-TATAATCTTT-ACACATTGGGATGAAGCAACAAATTCGTCCAGACTTTTGGTAGAGT
9	CGCTTTAATCTTTTACACGTTGGGATGACCCAACAAATTCGTCCAGACTTTTGGTAGAGT
10	GGCTATAATCTTTTACACGTTGGGATGAAGCAACAAATTCGTCCAGACTTTTGGTAGAGT
11	GGCTATAATCTTTTACACATTGGGATGAAGCAACAAATTCGTCCAGACTTTTGGTAGAGT
12	--CT-TAAT-----ACACATTGGGATGAAGCAACAAATTCGTCCAGACTTTTGGTAGAGT
13	GGCTATAATCTTTTACACATTGGGATGAAGCAACAAATTCGTCCAGACTTTTGGTAGAGT
14	GCCTATAATCTTTTACACATTGGGATGAAGCAACAAATTCGTCCAGACTTTTGGTAGAGT
15	CGCTATAATCTTTTACACATTGGGATGAAGCAACAAATTCGTCCAGACTTTTGGTAGAGT
16	GCGTTTAATCTT--ACACATTGGGATGAAGCAACAAATTCGTCCAGACTTTTGGTAGAGT
17	GGCTATAATCTTT-ACACATTGGGATGAAGCAACAAATTCGTCCAGACTTTTGGTAGAGT
18	GGCTATTATCTTT-ACACCTTGGGATGAAGCAACAAATTCGTCCAGACTTTTGGTAGAGT
19	GGTTATAATCTTTTACACATTGGGATGAAGCAACAAATTCGTCCAGACTTTTGGTAGAGT
20	GGCTATAATCTTTTACACATTGGGATGAAGCAACAAATTCGTCCAGACTTTTGGTAGAGT
21	GGCTATAATCTTTTACACATTGGGATGAAGCAACAAATTCGTCCAGACTTTTGGTAGAGT
22	GGCTATAATCTTTTACACATTGGGATGAAGCAACAAATTCGTCCAGACTTTTGGTAGAGT
23	GGCTATAATCTTTTACACATTGGGATGAAGCAACAAATTCGTCCAGACTTTTGGTAGAGT
24	GGCTATTATCTTTTACACATTGGGATGAAGCAACAAATTCGTCCAGACTTTTGGTAGAGT
25	GGCTATTATCTGTTACACATTGGGATGAAGCAACAAATTCGTCCAGACTTTTGGTAGAGT
26	GCGTAT-ATCTTT-ACACATTGGGATGAAGCAACAAATTCGTCCAGACTTTTGGTAGAGT
27	GCGTATAATCTTT-ACACATTGGGATGAAGCAACCAATTCGTCCAGACTTTTGGTAGAGT
28	CGCTTTAATCTTTTACACATTGGGATGAAGCAACAAATTCGTCCAGACTTTTGGTAGAGT

Table 4.5. Continued

Taxa[a]	Nucleotide sites[b]
	70 . . . 80 . . . 90 . . . 100 . . . 110 . . . 120
1	CTATAAGACCACGACTGATCCTGAAGGGTAATGAATGGAAAAAATAGCATGTCGTATTA-
2	CTATAAGACCACGACTGATCCTGAAGGGTAATGAATGGAAAAAATAGCATGTCGTATTAT
3	CTATAAGACCACGACTGATCCTGAAGGGTAATGAATGGAAAAAATAGCATGTCGTATTAT
4	CTATAAGACCACGACTGATCCTGAAGGGTAATGAATGGAAAAAATAGCATGTCGTATTAT
5	CTATAAGACCACGACTGATCCTGAAGGGTAATGAATGGAAAAAATAGCATGTCGTATTAT
6	CTATAAGACCACGACTGATCCTGAAGGGTAATGAATGGAAAAAATAGCATGTCGTATTAT
7	CTATAAGACCACGACTGATCCTGAAGGGTAATGAATGGAAAAAATAGCATGTCGTATTA-
8	CTATAAGACCACGACTGATCCTGAAGGGTAATGAATGGAAAAAATAGCATGTCGTATTA-
9	CTATAAGACCACGACTGATCCTGAAGGGTAATGAATGGAAAAAATAGCATGTCGTATTA-
10	CTATAAGACCACGACTGATCCTGAAGGGTAATGAATGGAAAAAATAGCATGTCGTATTA-
11	CTATAAGACCACGACTGATCCTGAAGGGTAATGAATGGAAAAAATAGCATGTCGTATTA-
12	CTATAAGACCACGACTGATCCGGAAGGGTAATTAATGGAAAAAATAGCATGTCGTATTA-
13	CTATAAGACCACGACTGATCCGGAAGGGTAATGAATGGAAAAAATAGCATGTCGTATTA-
14	CTATAAGACCACGACTGATCCTGAAGGGTAATGAATGGAAAAAATAGCATGTCGTATTA-
15	CTATAAGACCACGACTGATCCTGAAGGGTAATGAATGGAAAAAATAGCATGTCGTATTA-
16	CTATAAGACCACGACTGATCCGGAAGGGTAATGAATGGAAAAAATAGCATGTCGTATTA-
17	CTATAAGACCACGACTGATCCTGAAGGGTAATGAATGGAAAAAATAGCATGTCGTATTA-
18	CTATAAGACCACGACTGATCCTGAAGGGTAATGAATGGAAAAAATAGCATGTCGTATTA-
19	CTATAAGACCACGACTGATCCTGAAGGGTAATGAATGGAAAAAATAGCATGTCGTATTA-
20	CTATAAGACCACGACTGATCCTGAAGGGTAATGAATGGAAAAAATAGCATGTCGTATTA-
21	CTATAAGACCACGACTGATCCTGAAGGATAATGAATGGAAAAAATAGCATGTCGTATTA-
22	CTATAAGACCACGACTGATCCTGAAGGATAATGAATGGAAAAAATAGCATGTCGTATTA-
23	CTATAAGACCACGACTGATCCTGAAGGGTAATGAATGGAAAAAATAGCATGTCGTATTA-
24	CTATAAGACCACGACTGATCCTGAAGGGTAATGAATGGAAAAAATAGCATGTCGTATTA-
25	CTATAAGGCCACGACTGATCCTGAAGGGTAATGAATGGAAAAAATAGCATGTCGTATTA-
26	CTATAAGACCACGACTGATCCTGAAGGGTAATGAATGGAAAAAATAGCATGTCGTATTA-
27	CTATAAGACCACGACTGATCCTGAAGGGTAATGAATGGAAAAAATAGCATGTCGTATTA-
28	CTATAAGACCACGACTGATCCTGAAGGGTAATGAATGGAAAAAATAGCATGTCGTATTA-

Table 4.5. Continued

Taxa[a]	Nucleotide sites[b]
	1301 1401 1501 1601 1701 1801
1	----------TATTAATTTAATTAGTAA--------TGTAGAACTCTAAAAATAGATCA
2	TATTATATTAATATTAATTTAATTAGTAA--------TGTAGAACTCTAAAAATAGATCA
3	TATTATATTAATATTAATTTAATTAGTAA--------TGTAGAACTCTAAAAATAGATCA
4	TATTATATTAATATTAATTTAATTAGTAA--------TGTAGAACTCTAAAAATAGATCA
5	TATTATATTAATATTAATTTAATTAGTAA--------TGTAGAACTCTAAAAATAGATCA
6	TATTATATTAATATTAATTTAATTAGTAA--------TGTAGAACTCTAAAAATAGATCA
7	-----TATTAATATTAATTGAATTAGTAA--------TGTAGAACTCTAAAAATAGATCA
8	----------TATTAATTTAATTAGTAA--------TGTAGAACTCTAAAAATGGATCA
9	----------TATTAATTTAATTAG---TTTAGTAATGTAGAACTCTAAAAATGGATCA
10	----------TATTAATTTAATTAG---TTTAGTAATGTAGAACTCTAAAAATGGATCA
11	----------TATTAATTTAATTAGTAATTTAGTAATGTAGAACTCTAAAAATAGATCA
12	----------TATTAATTTAATTAGTAA--------TGTAGAACTCTAAAAATAGATCA
13	----------TATTAATTTAATTAGTAA--------TGTAGAACTCTAAAAATGGATCA
14	----------TATTAATTTAATTAGTAA--------TGTAGAACTCTAAAAAAAGATCA
15	----------TATTAATTTAATTAGTAA--------TGTAGAACTCTAAAAAAAGATCA
16	----------TATTAATTTAATTAGTAA--------TGTAGAACTCTAAAAATAGATCA
17	----------TATTAATTTAATTAGTAA--------TGTAGAACTCTAAAAATAGATCA
18	----------TATTAATTTAATTAGTAA--------TGTAGAACTCTAAAAATAGATCA
19	----------TATTAATTTAATTAGTAA--------TGTAGAACTCTAAAAATAGATCA
20	-----TATTAA-ATTAATTTAATTACTAA--------TGTAGAACTCTAAAAATAGATCA
21	-----TATTAATATTAATTTAATTAGTAA--------TGTAGAACTCTAAAAATAGATCA
22	-----TATTAATATTAATTTAATTAGTAA--------TGTAGAACTCTAAAAATAGATCA
23	-----TATTAATATTAATTTAATTAGTAA--------TGTAGAACTCTAAAAATAGATCA
24	----------TATTAATTTAATTAGTAA--------TGTAGAACTCTAAAAATAGATCA
25	----------TATTAATTTAATTAGTAA--------TGTAGAACTCTAAAAATAGATTA
26	----------TATTAATTTAATTAGTAA--------TGTAGAACTCTAAAAATAGATTA
27	----------TATTAATTTAATTAGTAA--------TGTAGAACTCTAAAAATAGATCA
28	----------TATTAATTTAATTAGTAA--------TGTAGAACTCTAAAAATAGATCA

Table 4.5. Continued

Taxa[a]	Nucleotide sites[b]					
	1	2	2	2	2	2
	9	0	1	2	3	4
	0	0	0	0	0	0
	.	.	.	.	.	.
1	TCCTTACTAGATCGGTACAAAAAACCTTTGATTTTAGGAAGGGGATAAAATGAATTCACA					
2	TCCTTACTAGATCGGTACAAAAAACCTTTGATTTTAGGAAGGGGATAAAATGAATTCACA					
3	TCCTTACTAGATCGGTACAAAAAACCTTTGATTTTAGGAAGGGGATAAAATGAATTCACA					
4	TCCTTACTAGATCGGTACAAAAAACCTTTGATTTTAGGAAGGGGATAAAATGAATTCACA					
5	TCCTTACTAGATCGGTACAAAAAACCTTTGATTTTAGGAAGGGGATAAAATGAATTCACA					
6	TCCTTACTAGATCGGTACAAAAAACCTTTGATTTTAGGAAGGGGATAAAATGAATTCACA					
7	TCCTTACTAGATCGGTACAAAAAACCTTTTATTTTAGGAAGGGGATAAAATGAATTCACA					
8	TCCTTACTAGATCGGTACAAAAAACCTTTGATTTTAGGAAGGGGATAAAATGAATTCACA					
9	TCCTTACTAGATCGGTACAAAAAACCTTTGATTTTAGGAAGGGGATAAAATGAATTCACA					
10	TCCTTACTAGATCGGTACAAAAAACCTTTGATTTTAGGAAGGGGATAAAATGAATTCACA					
11	TCCTTACTAGATCGGTACAAAAAACCTTTGATTTTAGGAAGGGGATAAAATGAATTCACA					
12	TCCTTACTAGATCGGTACAAAAAACCTTTGATTTTAGGAAGGGGGTAAAATGAATTCACA					
13	TCCTTACTAGATCGGTACAAAAAACCTTTGATTTTAGGAAGGGGGTAAAATGAATTCACA					
14	TCCTTACTAGATCGGTACAAAAAACCTTTGATTTTAGGAAGGGGATAAAATGAATTCACA					
15	TCCTTACTAGATCGGTACAAAAAACCTTTGATTTTAGGAAGGGGATAAAATGAATTCACA					
16	TCCTTACTAGATCGGTACAAAAAACCTTTGATTTTAGGAAGGGGGTAAAATGAATTCACA					
17	TCCTTACTAGATCGGTACAAAAAACCTTTGATTTTAGGAAGGGGATAAAATGAATTCACA					
18	TCCTTACTAGATCGGTACAAAAAACCTTTGATTTTAGGAAGGGGATAAAATGAATTCACA					
19	TCCTTACTAGATCGGTACAAAAAACCTTTGATTTTAGGAAGGGGATAAAATGAATTCACA					
20	TCCTTACTAGATCGGTACAAAAAACCTTTGATTTTAGGAAGGGGATAAAATGAATTCACA					
21	TCCTTACTAGATCGGTACAAAAAACCTTTGATTTTAGGAAGGGGATAAAATGAATTCACA					
22	TCCTTACTAGATCGGTACAAAAAACCTTTGATTTTAGGAAGGGGATAAAATGAATTCACA					
23	TCCTTACTAGATCGGTACAAAAAACCTTTGATTTTAGGAAGGGGATAAAATGAATTCACA					
24	TCCTTACTAGATCGGTACAAAAAACCTTTGATTTTAGGAAGGGGATAAAATGAATTCACA					
25	TCCTTACTAGATCGGTACAAAAAACCTTTGATTTTAGGAAGGGGATAAAATGAATTCACA					
26	TCCTTACTAGATCGGTACAAAAAACCTTTGATTTTAGGAAGGGGATAAAATGAATTCACA					
27	TCCTTACTAGATCGGTACAAAAAACCTTTGATTTTAGGAAGGGGATAAAATGAATTCACA					
28	TCCTTACTAGATCGGTACAAAAAACCTTTGATTTTAGGAAGGGGATAAAATGAATTCACA					

Table 4.5. Continued

Taxa[a]	Nucleotide sites[b]
	250 260 270 280 290 300
1	TTTACTGTTTGGTCGAGTGAATAAATGGATAGAGTCTTATGGCTCCAATTCTCGGTAAAG
2	TTTACTGTTTGGTCGAGTGAATAAATGGATAGAGTCTTATGGCTCCAATTCTCGGTAAAG
3	TTTACTGTTTGGTCGAGTGAATAAATGGATAGAGTCTTATGGCTCCAATTCTCGGTAAAG
4	TTTACTGTTTGGTCGAGTGAATAAATGGATAGAGTCTTATGGCTCCAATTCTCGGTAAAG
5	TTTACTGTTTGGTCGAGTGAATAAATGGATAGAGTCTTATGGCTCCAATTCTCGGTAAAG
6	TTTACTGTTTGGTCGAGTGAATAAATGGATAGAGTCTTATGGCTCCCATTCTCGGTAAAG
7	TTTACTGTTTGGTCGAGTGAATAAATGGATAGAGTCTTATGGCTCCAATTCTCGGTAAAG
8	TTTACTGTTTGGTCGAGTGAATAAATGGATAGAGTCTTATGGCTCCAATTCTCGGTAAAG
9	TTTACTGTTTGGTCGAGTGAATAAATGGATAGAGTCTTATGGCTCCAATTCTCGGTAAAG
10	TTTACTGTTTGGTCGAGTGAATAAATGGATAGAGTCTTATGGCTCCAATTCTCGGTAAAG
11	TTTACTGTTTGGTCGAGTGAATAAATGGATAGAGTCTTATGGCTCCAATTCTCGGTAAAG
12	TTTACTGTTTGGTCGAGTGAATAAATGGATAGAGTCTTATGGCTCCAATTCTCGGTAAAG
13	TTTACTGTTTGGTCGAGTGAATAAATGGATAGAGTCTTATGGCTCCAATTCTCGGTAAAG
14	TTTACTGTTTGGTCGAGTGAATAAATGGATAGAGTCTTATGGCTCCAATTCTCGGTAAAG
15	TTTACTGTTTGGTCGAGTGAATAAATGGATAGAGTCTTATGGCTCCAATTCTCGGTAAAG
16	TTTACTGTTTGGTCGAGTGAATAAATGGATAGAGTCTTATGGCTCCAATTCTCGGTAAAG
17	TTTACTGTTTGGTCGAGTGAATAAATGGATAGAGTCTTATGGCTCCAATTCTCGGTAAAG
18	TTTACTGTTTGGTCGAGTGAATAAATGGATAGAGTCTTATGGCTCCAATTCTCGGTAAAG
19	TTTACTGTTTGGTCGAGTGAATAAATGGATAGAGTCTTATGGCTCCAATTCTCGGTAAAG
20	TTTACTGTTTGGTCGAGTGAATAAATGGATAGAGTCTTATGGCTCCAATTCTCGGTAAAG
21	TTTACTGTTTGGTCGAGTGAATAAATGGATAGAGTCTTATGGCTCCAATTCTCGGTAAAG
22	TTTACTGTTTGGTCGAGTGAATAAATGGATAGAGTCTTATGGCTCCAATTCTCGGTAAAG
23	TTTACTGTTTGGTCGAGTGAATAAATGGATAGAGTCTTATGGCTCCAATTCTCGGTAAAG
24	TTTACTGTTTGGTCGAGTGAATAAATGGATAGAGTCTTATGGCTCCAATTCTCGGTAAAG
25	TTTACTGTTTGGTCGAGTGAATAAATGGATAGAGTCTTATGGCTCCAATTCTCGGTAAAG
26	TTTACTGTTTGGTCGAGTGAATAAATGGATAGAGTCTTATGGCTCCAATTCTCGGTAAAG
27	TTTACTGTTTGGTCGAGTGAATAAATGGATAGAGTCTTATGGCTCCAATTCTCGGTAAAG
28	TTTACTGTTTGGTCGAGTGAATAAATGGATAGAGTCTTATGGCTCCAATTCTCGGTAAAG

Table 4.5. Continued

Taxa[a]	Nucleotide sites[b]
	3 3 3 3 3 3 1 2 3 4 5 6 0 0 0 0 0 0 . . . . . .
1	AAAAAGCGACGAGCTTATGTTCTTAATTTGAATGGTTACCCGATCTAATCTAATTAGACG
2	AAAAAGCGACGAGCTTATGTTCTTAATTTGAATGGTTACCTGATCTAATCTAATTAGACG
3	AAAAAGCGACGAGCTTATGTTCTTAATTTGAATGGTTACCTGATCTAATCTAATTAGACG
4	AAAAAGCGACGAGCTTATGTTCTTAATTTGAATGGTTACCTGATCTAATCTAATTAGACG
5	AAAAAGCGACGAGCTTATGTTCTTAATTTGAATGGTTACCTGATCTAATCTAATTAGACG
6	AAAAAGCGACGAGCTTATGTTCTTAATTTGAATGGTTACCTGATCTAATCTAATTAGACG
7	AAAAAGCGACGAGCTTATGTTCTTAATTTGAATGGTTACCTGATCTAATCTAATTAGACG
8	AAAAAGCGACGAGCTTATGTTCTTAATTTGAATGGTTACCCGATCTAATCTAATTAGACG
9	AAAAAGCGACAAGCTTATGTTCTTAATTTGAATGGTTACCTGATCTAATCTAATTAGACG
10	AAAAAGCGACGAGCTTATGTTCTTAATTTGAATGGTTACCTGATCTAATCTAATTAGACG
11	AAAAAGCGACGAGCTTATGTTCTTAATTTGAATGGTTACCTGATCTAATCTAATTAGACG
12	AAAAAGCGACGAGCTTATGTTCTTAATTTGAATGGTTACCTGATCTAATCTAATTAGACG
13	AAAAAGCGACGAGCTTATGTTCTTAATTTGAATGGTTACCTGATCTAATCTAATTAGACG
14	AAAAAGCGACGAGCTTATGTTCTTAATTTGAATGGTTACCTGATCTAATCTAATTAGACG
15	AAAAAGCGACGAGCTTATGTTCTTAATTTGAATGGTTACCTGATCTAATCTAATTAGACG
16	AAAAAGCGACGAGCTTATGTTCTTAATTTGAATGGTTACCTGATCTAATCTAATTAGACG
17	AAAAAGCGACGAGCTTATGTTCTTAATTTGAATGGTTACCTGATCTAATCTAATTAGACG
18	AAAAAGCGACGAGCTTATGTTCTTAATTTGAATGGTTACCTGATCTAATCTAATTAGACG
19	AAAAAGCGACGAGCTTATGTTCTTAATTTGAATGGTTACCTGATCTAATCTAATTAGACG
20	AAAAAGCGACGAGCTTATGTTCTTAATTTGAATGGTTACCTGATCTAATCTAATTAGACG
21	AAAAAGCGACGAGCTTATGTTCTTAATTTGAATGGTTACCCGATCTAATCTAATTAGACG
22	AAAAAGCGACGAGCTTATGTTCTTAATTTGAATGGTTACCCGATCTAATCTAATTAGACG
23	AAAAAGCGACGAGCTTATGTTCTTAATTTGAATGGTTACCCGATCTAATCTAATTAGACG
24	AAAAAGCGACGAGCTTATGTTCTTAATTTGAATGGTTACCCGATCTAATCTAATTAGACG
25	AAAAAGCGACGAGCTTATGTTCTTAATTTGAATGGTTACCTGATCTAATCTAATTAGACG
26	AAAAAGCGACGAGCTTATGTTCTTAATTTGAATGGTTACCTGATCTAATCTAATTAGACG
27	AAAAAGCGACGAGCTTATGTTCTTAATTTGAATGGTTACCTGATCTAATCTAATTAGACG
28	AAAAAGCGACGAGCTTATGTTCTTAATTTGAATGGTTACCCGATCTAAT-----TAGACG

Table 4.5. Continued

Taxa[a]	Nucleotide sites[b]

	3	3	3	4	4	4
	7	8	9	0	1	2
	0	0	0	0	0	0
	.	.	.	.	.	.
1	TTAAAAATGGATTAGTACCTGATACGGGAAAGGGTTCTCCTGTGAGTGGATCATCATTTA					
2	TTAAAAATGGATTAGTACCTGATACGGGAAAAGGTTCTCCTGTGAGTGGATCATCATTTA					
3	TTAAAAATGGATTAGTACCTGATACGGGAAAAGGTTCTCCTGTGAGTGGATCATCATTTA					
4	TTAAAAATGGATTAGTACCTGATACGGGAAAAGGTTCTCCTGTGAGTGGATCATCATTTA					
5	TTAAAAATGGATTAGTACCTGATACGGGAAAAGGTTCTCCTGTGAGTGGATCATCATTTA					
6	TTAAAAATGGATTAGTACCTGATACGGGAAAAGGTTCTCCTGTGAGTGGATCATCATTTA					
7	TTAAAAATGGATTAGTACCTGATACGGGAAAAGGTTCTCCTGTGAGTGGATCATCATTTA					
8	TTAAAAATGGATTAGTACCTGATACGGGAAAGGGTTCTCCTGTGAGTGGATCATCATTTA					
9	TTAAAAATGGATTAGTACCTGATACGGGAAAGGGTTCTCCTGTAAGTAGATCATCATTTA					
10	TTAAAAATGGATTAGTACCTGATACGGGAAAGGGTTCTCCTGTAAGTAGATCATCATTTA					
11	TTAAAAATAGATTAGTACCTGATACGGGAAAGGGTTCTCCTGTGAGTGGATCATCATTTA					
12	TTAAAAATGGATTAGTACCTGATACGGGAAAGGGTTCTCCTGTGAGTGGATCATCATTTA					
13	TTAAAAATGGATTAGTACCTGATACGGGAAAGGGTTCTCCTGTGAGTGGATCATCATTTA					
14	TTAAAAATAGATTAGTACCTTATACGGGAAAGGGTTCTCCTGCGAGTGGATCATCATTTA					
15	TTAAAAATAGATTAGTACCTTATACGGGAAAGGGTTCTCCTGTGAGTGGATCATCATTTA					
16	TTAAAAATGGATTAGTACCTGATACGGGAAAGGGTTCTCCTGTGAGTGGATCATCATTTA					
17	TTAAAAATAGATTAGTACCTGATACGGGAAAGGGTTCTCCTGTGAGTGGATCATCATTTA					
18	TTAAAAATAGATTAGTACCTGATACGGGAAAGGGTTCTCCTGTGAGTGGATCATCATTTA					
19	TTAAAAATGGATTAGTACCTGATACGGGAAAGGGTTCTCCTGTGAGTGGATCATCATTTA					
20	TTAAAAATGGATTAGTACCTGATACGGGAAAGGGTTCTCCTGTGAGTAGATCATTATTTA					
21	TTAAAAATGGATTAGTACCTGATACGGGAAAGGGTTCTCCTGTGAGTGGATCATCATTTA					
22	TTAAAAATGGATTAGTACCTGATACGGGAAAGGGTTCTCCTGTGAGTGGATCATCATTTA					
23	TTAAAAATGGATTAGTACCTGATACGGGAAAGGGTTCTCCTGTGAGTGGATCATCATTTA					
24	TTAAAAATGGATTAGTACCTGATACGAGAAAGGGTTCTCCTGTGAGTGGATCATCATTTA					
25	TTAAAAATGGATTAGTACCTGATACGGTAAAGGGTTCTCCTGTGAGTGGATCATCATTTA					
26	TTAAAAATGGATTAGTACCTGATACGGGAAAGGGTTCTCCTGTGAGTGGATCATCATTTA					
27	TTAAAAATGGATTAGTACCTGATACGGGAAAGGGTTCTCCTGTGAGTGGATCATCATTTA					
28	TTAAAAATGGATTAGTACCTGATACGGGAAAGGGTTCTCCTGTGAGTGGATCATCATTTA					

Table 4.5. Continued

Taxa[a]	Nucleotide sites[b]
	4 4 4 4 4 4 3 4 5 6 7 8 0 0 0 0 0 0 . . . . . .
1	CTTTTTGAATGAATCCTAACTATTCTCCATTA-----TAGATAGGGCAGAGTAGAGATAG
2	CTTTTTGAATGAATCCTAACTATTCTCCATTA-----TGGATAGGGCAGAGTAGAGATGG
3	CTTTTTGAATGAATCCTAACTATTCTCCATTA-----TGGATAGGGCAGAGTAGAGATGG
4	CTTTTTGAATGAATCCTAACTATTCTCCATTA-----TGGATAGGGCAGAGTAGAGATGG
5	CTTTTTGAATGAATCCTAACTATTCTCCATTA-----TGGATAGGGCAGAGTAGAGATGG
6	CTTTTTGAATGAATCCTAACTATTCTCCATTA-----TGGATAGGGCAGAGTAGAGATGG
7	CTTTTTGAATGAATCCTAACTATTCTCCATTA-----TGGATAGGGCAGAGTAGAGATGG
8	CTTTTTGAATGAATCCTAACTATTCTCCATTA-----TRGATAGGGCAGAGTAGAGATAG
9	CTTTTTGAATGAATCCTAACTATTCTCCATTA-----TGGATAGGGCAGAGTAGAGATGG
10	CTTTTTGAATGAATCCTAACTATTCTCCATTA-----TGGATAGGGCAGAGTAGAGATGG
11	CTTTTTGAATGAATCCTAACTATTCTCCATCATATTATGGATAGGGCAGAGTAGAGATGG
12	CTTTTTGAATGAATCCTAACTATTCTCCATTA-----TGGATAGGGCAGAGTAGAGATGG
13	TTTTTTGAATGAATCCTAACTATTCTCCATTA-----TGAATAGGGCAGAGTAGAGATGG
14	CTTTTTGAATGAATCCTAACTATTCTCCATTATATTATGGATAGGGCAGAGTAGAGATGG
15	CTTTTTGAATGAATCCTAACTATTCTCCATTATATTATGGATAGGGCAGAGTAGAGATGG
16	CTTTTTGAATGAATCCTAACTATTCTCCATTA-----TGGATAGGGCAGAGTAGAGATGG
17	CTTTTTGAATGAATCCTAACTATTCTCCATTATATTATGGATAGGGCAGAGTAGAGATGG
18	CTTTTTGAATGAATCCTAACTATTCTCCATTATATTATGGATAGGGCAGAGTAGAGATGG
19	CTTTTTGAATGAATCCTAATTATTTTCCATTA-----TGGATAGGGCAGAGTAGAGATGG
20	CTTTTTGAATGAATCCTAACTATTCTCCATTA-----TGGATAGGGCAGAGTAGAGATGG
21	CTTTTTGAATGAATCCTAACTATTCTCCATTA-----TGGATAGGGCAGAGTAGAGATGG
22	CTTTTTGAATGAATCCTAACTATTCTCCATTA-----TAGATAGGGCAGAGTAGAGATAG
23	CTTTTTGAATGAATCCTAACTATTCTCCATTA-----TAGATAGGGCAGAGTAGAGATAG
24	CTTTTTGAATGAATCCTAACTATTCTCCATTA-----TAGATAGGGCAGAGTAGAGATAG
25	CTTTTTGAATGAATCCTTACTATTCTCCATTA-----TGGATAGGGCAGAGTAGAGATGG
26	CTTTTTGAATGAATCCTAACTATTCTCCATTA-----TGGATAGGGCAAAGTAGAGATGG
27	CTTTTTGAATGAATCCTAATTATTCTCCATTA-----TGGATAGGGCAGAGTAGAGATGG
28	CTTTTTGAATGAATCCTAACTATTCTCCATTA-----TCGATAGGGCAGAGTGGAGATGG

Table 4.5. Continued

Taxa[a]	Nucleotide sites[b]
	490 500 510 520 530 540
1	ATGTGTAAAAGAAAGAGCATACTGATAAAGTGATTCCAAAATCAAAAGAGCGATTGGGTT
2	ATGTGTAAAAGAAAGAGCATACTGATAAAGTGATTCCAAAACCAAAAGAGCGATTGGGTT
3	ATGTGTAAAAGAAAGAGCATACTGATAAAGTGATTCCAAAACCAAAAGAGCGATTGGGTT
4	ATGTGTAAAAGAAAGAGCATACTGATAAAGTGATTCCAAAACCAAAAGAGCGATTGGGTT
5	ATGTGTAAAAGAAAGAGCATACTGATAAAGTGATTCCAAAACCAAAAGAGCGATTGGGTT
6	ATGTGTAAAAGAAAGAGCATACTGATAAAGTGATTCCAAAACCAAAAGAGCGATTGGGTT
7	ATGTGTAAAAGAAAGAGCATACTGATAAAGTGATTCCAAAACCAAAAGAGCGATTGGGTT
8	ATGTGTAAAAGAAAGAGCATACTGATAAAGTGATTCCAAAATCAAAAGAGCGATTGGGTT
9	ATGTGTAAAAGAAAGAGCATACTGATAAAGTGATTCCAAAATCAAAAGAGCGATTGGGTT
10	ATGTGTAAAAGAAAGAGCATACTGATAAAGTGATTCCAAAATCAAAAGAGCGATTGGGTT
11	ATGTGTAAAAGAAAGAGCATACTGATAAAGTGATTCCAAAATCAAAAGAGCGATTGGGTT
12	ATGTGTAAAAGAAAGAGCATACTGATAAAGTGATTCCAAAATCAAAAGAGCGATTGGGTT
13	ATGTGTAAAAGAAAGAGCATACTGATAAAGTGATTCCAAAATCAAAAGAGCGATTGGGTT
14	ATGTGTAAAAGAAAGAGCATACTGATAAAGTGATTCCAAAATCAAAAGAGCGATTGGGTT
15	ATGTGTAAAAGAAAGAGCATACTGATAAAGTGATTCCAAAATCAAAAGAGCGATTGGGTT
16	ATGTGTAAAAGAAAGAGCATACTGATAAAGTGATTCCAAAATCAAAAGAGTGATTGGGTT
17	ATGTGTAAAAGAAAGAGCATACTGATAAAGTGATTCCAAAATCAAAAGAGCGATTGGGTT
18	ATGTGTAAAAGAAAGAGCATACTGATAAAGTGATTCCAAAATCAAAAGAGCGATTGGGTT
19	ATGTGTAAAAGAAAGAGCATACTGATAAAGTGATTCCAAAATCAAAAGAGCGATTGGGTT
20	ATGTGTAAAAGAAAGAGCATACTGATAAAGTGATTCCAAAATCAAAAGAGCGATTGGGTT
21	ATGTGTAAAAGAAAGAGCATACTGATAAAGTGATTCCAAAATCAAAAGAGCGATTGGGTT
22	ATGTGTAAAAGAAAGAGCATACTGATAAAGTGATTCCAAAATCAAAAGAGCGATTGGGTT
23	ATGTGTAAAAGAAAGAGCATACTGATAAAGTGATTCCAAAATCAAAAGAGCGATTGGGTT
24	ATGTGTAAAAGAAAGAGCATACTGATAAAGTGATTCCAAAATCAAAAGAGCGATTGGGTT
25	ATGTGTAAAAGAAAGAGCATACTGATAAAGTGATTCCAAAATCAAAAGAGCGATTGGGTT
26	ATGTGTAAAAGAAAGAGCATACTGATAAAGTGATTCCAAAATCAAAAGAGCGATTGGGTT
27	ATGTGTAAAAGAAAGAGCATACTGATAAAGTGATTCCAAAATCAAAAGAGCGATTGGGTT
28	ATGTGTAAAAGAAAGAGCATACTGATAAAGTGATTCCAAAATCAAAAGAGCGATTGGGTT

Table 4.5. Continued

Taxa[a]	Nucleotide sites[b]
	550 560 570 580 590 600
1	GCAAAAATAAA------GGATTTTTAACC-TTTTCTTTTTCTTGTTATGTTAACGGACAT
2	GCAAAAATAAA------GGATTTTTAACC-TTTTCTTTTTCTTGTTATGTTAACGGACAT
3	GCAAAAATAAA------GGATTTTTAACC-TTTTCTTTTTCTTGTTATGTTAACGGACAT
4	GCAAAAATAAA------GGATTTTTAACC-TTTTCTTTTTCTTGTTATGTTAACGGACAT
5	GCAAAAATAAA------GGATTTTTAACC-TTTTCTTTTTCTTGTTATGTTAACGGACAT
6	GCAAAAATAAA------GGATTTTTAACCCTTTTCTTTTTCTTGTTATGTTAACGGACAT
7	GCAAAAATAAA------GGATTTTTAACC-TTTTCTTTTTCTTGTTATGTTAACGGACAT
8	GCAAAAATAAA------GGATTTTTAACC-TTTTCTTTTTCTTGTTATGTTAACGGACAT
9	GCAAAAATAAAAATAAAGGATTTTTAACC--TTTTTTTTTCTTGTTATGTTAACGGACAT
10	GCAAAAATAAAAATAAAGGATTTTTAACC--TTTTTTTTTCTTGTTATGTTAACGGACAT
11	GCAAAAATAAA------GGATTTTTAACC-TTTTCTTTTTCTTGTTATGTTAACGGACAT
12	GCAAAAATAAA------GGATTTTTAACC-TTTTCTTTTTCTTGTTATGTTAACGGACAT
13	GCAAAAATAAA------GGATTTT-GACCCTTTTCTTTTTCTTGTTATGTTAACGGACAT
14	GCAAAAATAAA------GGATTTTTAACC-TTTTCTTTTTCTTGTTATGTTAACGGACAT
15	GCAAAAATAAA------GGATTTTTAACC-TTTTCTTTTTCTTGTTATGTTAACGGACAT
16	GCAAAAATAAA------GGATTTTTAACC-TTTTTTTTTTCTTGTTATGTTAACGGACAT
17	GCAAAAATAAA------GGATTTTTAACC-TTTTCTTTTTCTTGTTATGTTAATGGACAT
18	GCAAAAATAAA------GGATTTTTAACC-TTTTCTTTTTCTTGTTATGTTAATGGACAT
19	GCAAAAATAAA------GGATTTTTAACC-TTTTCTTTTTCTTGTTATGTTAACGGACAT
20	GCAAAAATAAA------GGATTTTTAACC-TTTTCTTTTTCTTGTTATGTTAACGGACAT
21	GCAAAAATAAA------GGATTTTTAACC-TTTTCTTTTTCTTGTTATGTTAACGGACAT
22	GCAAAAATAAA------GGATTTTTAACC-TTTTCTTTTTCTTGTTATGTTAACGGACAT
23	GCAAAAATAAA------GGATTTTTAACC-TTTTCTTTTTCTTGTTATGTTAATGGACAT
24	GCAAAAATAAA------GGATTTTTAACC-TTTTCTTTTTCTTGTTATGTTAACGGACAT
25	GCAAAAATAAA------GGATTTTTAACC-TTTTCTTTTTCTTGTTATGTTAACGGACAT
26	GCAAAAATAAA------GGATTTTTAACT-TTTTATTTTTCTTGTTATGTTAACGGACAT
27	GCAAAAATAAA------GGATTTTTAACC-TTTTCTTTTTCTTGTTATGTTAACGGACAT
28	GCAAAAATAAA------GGATTTTTAACCCTTTTCTTTTTCTTGTTATGTTAACGGACAT

Table 4.5. Continued

Taxa[a]	Nucleotide sites[b]
	6 6 6 6 6 6 1 2 3 4 5 6 0 0 0 0 0 0 . . . . . .
1	AACCCA-TTTGATCTGGAAAGATAGGAAGAAGATCTAGAGATTAGACTTTCTGTTTTCGA
2	AAACCAATTTGATCTGGAAAGATAGGAAGAAGATCTAGAGATTAGACTTTCTGTTTTCGA
3	AAACCAATTTGATCTGGAAAGATAGGAAGAAGATCTAGAGATTAGACTTTCTGTTTTCGA
4	AAACCAATTTGATCTGGAAAGATAGGAAGAAGATCTAGAGATTAGACTTTCTGTTTTCGA
5	AAACCAATTTGATCTGGAAAGATAGGAAGAAGATCTAGAGATTAGACTTTCTGTTTTCGA
6	AAACCAATTTGATCTGGAAAGATAGGAAGAAGATCTAGAGATTAGACTTTCTGTTTTCGA
7	AAACCAATTTGATCTGGAAAGATAGGAAGAAGATCTAGAGATTAGACTTTCTGTTTTCGA
8	AAACCAATTTGATCTGGAAAGATAGGAAGAAGATCTAGAGATTAGACTTTCTGTTTTCGA
9	AAACCAATTTGATCTGGAAAGATAGGAAGAAGATCTAGAGATTAGACTTTCTGTTTTCGA
10	AAACCAATTTGATCTGGAAAGATAGGAAGAAGATCTAGAGATTAGACTTTCTGTTTTCGA
11	AAACCAATTTGATCTGGAAAGATAGGAAGAAGATCTAGAGATTAGACTTTCTGTTTTCGA
12	AAACCAATTTGATCTGGAAAGATAGGAAGAAGATCTAGAGATTAGACTTTCTGTTTTCGA
13	AAACCAATTTGATCTGGAAAGATAGGAAGAAGATCTAGAGATTAGACTTTCTGTTTTCGA
14	AAACCAATTTGATCTGGAAAGATAGGAAGAAGATCTAGAGATTAGACTTTCTGTTTTCGA
15	AAACCAATTTGATCTGGAAAGATAGGAAGAAGATCTAGAGATTAGACTTTCTGTTTTCGA
16	AAACCAATTTGATCTGGAAAGATAGGAAGAAGATCTAGAGATTAGACTTTCTGTTTTCGA
17	AAATCAATTTGATCTGGAAAGATAGGAAGAAGATCTAGAGATTAGACTTTCTGTTTTCGA
18	AAATCAATTTGATCTGGAAAGATAGGAAGAAGATCTAGAGATTAGACTTTCTGTTTTCGA
19	AAACCAATTTGATCTGGAAAGATAGGAAGAAAATCTCGAGATTAGACTTTCTGTTTTCGA
20	AAACCAATTTTATCTGGAAAGATAGGAAGAAGATCTAGAGATTAGACTTTCTGTTTTCGA
21	AAACCAATTTGATCTGGAAAGATAGGAAGAAGATCTAGAGATTAGACTTTCTGTTTTCGA
22	AAACCAATTTGATCTGGAAAGATAGGAAGAAGATCTAGAGATTAGACTTTCTGTTTTCGA
23	AAACCAATTTGATCTGGAAAGATAGGAAGAAGATCCAGAGATTAGACTTTCTGTTTTCGA
24	AAACCAATTTGATCTGGAAAGATAGGAAGAAGATCTAGAGATTAGACTTTCTGTTTTCGA
25	AAACCAATTTGATCTGGAAAGATAGGAAGAAGATCTAGAGATTAGACTTTCTGTTTTCGA
26	AAACCAATTTGATCTGGAAAGATAGGAAGAAGATCTAGAGATTAGACTTTCTGTTTTCGA
27	AAACCAATTTGATCTGGAAAGATAGGAAGAAAATCTCGAGATTAGACTTTCTGTTTTCGA
28	AAACCAATTTGATCTGGAAAGATAGGAAGAAGATCTAGAGATTAGACTTTCTGTTTTCGA

Table 4.5. Continued

Taxa[a]	Nucleotide sites[b]
	6 6 6 7 7 7 7 8 9 0 1 2 0 0 0 0 0 0 . . . . . .
1	GGTCACTTTTATCA-CTGTATATATA----------CCTTATATACATA------
2	GGTCACTTTTATCA-CTGTATATATA----------CCTTATATACATA------
3	GGTCACTTTTATCA-CTGTATATATA----------CCTTATATACATA------
4	GGTCACTTTTATCA-CTGTATATATA----------CCTTATATACATA------
5	GGTCACTTTTATCA-CTGTATATATA----------CCTTATATACATA------
6	GGTCACTTTTATCACCTGTATATATA----------CCTTATATACATA------
7	GGTAACTTTTATCA-CTGTATATATA----------CCTTATATACATA------
8	GGTCACTTTTATCA-CTGTATATATA----------CCTTATATACATA------
9	GGTCACTTTTATCA-CTGTATATATA----------CCTTATATACATA------
10	GGTCACTTTTATCA-CTGTATATATA----------CCTTATATACATA------
11	GGTCACTTTTATCA-------ATATA----------CCTTATATACATA------
12	GGTCACTTTTATCA-CTGTATATATA----------CCTTATATACATA------
13	GGTCACTTTTATCA-CTGTATATATA----------CCTTATATACATA------
14	GGTCACTTTTATCA-------ATATA----------CCTTATATACATA------
15	GGTCACTTTTATCA-------ATATA----------CCTTATATACATA------
16	GGTCACTTTTATCA-CTGTATATATA----------CCTTATATACATA------
17	GGTCACTTTTATCA-------ATATA----------CCTTATATACATA------
18	GGTCACTTTTATCA-------ATATA----------CCTTATATACATA------
19	GGTCACTTTTATCA-CTGTATATATA----------TACCTTATATACATA------
20	GGTCACTTTTATCA-CTGTATATATACCTTATATACATATACCTTATATACATA------
21	GGTCACTTTTATCA-CTGTATATATA----------CCTTATATACATA------
22	GGTCACTTTTATCA-CTGTATATATA----------CCTTATATACATA------
23	GGTCACTTTTATCA-CTGTATATATA----------CCTTATATACATA------
24	GGTCACTTTTATCA-CTGTATATATA----------CCTTATATACATA------
25	GGTCACTTTTATCA-CTGTATATATA----------TCTTATATACATACTTGAT
26	GGTCACTTTTATCA-CTGTATATATA----------CCTTATATACATA------
27	GGTCACTTTTATCA-CTGTATATATA----------TACCTTATATACATA------
28	GGTCACTTTTATCA-CTGTATATATA----------CCTTATATACATA------

Table 4.5. Continued

Taxa[a]	Nucleotide sites[b]

matK .--->

	7 3 0 .	7 4 0 .	7 5 0 .	7 6 0 .	7 7 0 .	7 8 0 .
1	-CTTGTTCTGACCATATCGCACTATGTATCATTTGCTGACCCAAGAAATGTTTCT-----					
2	-CTTGTTCTGACCATATCGCACTATGTATCATTTGCTGACCCAAGAAATGTTTCT-----					
3	-CTTGTTCTGACCATATCGCACTATGTATCATTTGCTGACCCAAGAAATGTTTCT-----					
4	-CTTGTTCTGACCATATCGCACTATGTATCATTTGCTGACCCAAGAAATGTTTCT-----					
5	-CTTGTTCTGACCATATCGCACTATGTATCATTTGCTGACCCAAGAAATGTTTCT-----					
6	-CTTGTTCTGACCATATCGCACTATGTATCATTTGCTGACCCAAGAAATGTTTCT-----					
7	-CTTGTTCTGACCATATCGCACTATGTATCATTTGCTGACCCAAGAAATGTTTCT-----					
8	-CTTGTTCTGACCATATCGCACTATGTATCATTTGCTGACCCAAGAAATGTTTCT-----					
9	-CTTGTTCTGACCATATCGCACTATGTATTATTTGCTGACCCAAGAAATGTTTCT-----					
10	-CTTGTTCTGACCATATCGCACTATGTATTATTTGCTGACCCAAGAAATGTTTCT-----					
11	-CTTGTTCTGACCATATTGCACTATGTATCATTTGCTGACCCAAGAAATGTTTCT-----					
12	-CTTGTTCTGACCATATCGCACTATGTATCATTTGCTGACCCAAGAAATGTTTCTTTTTC					
13	-CTTGTTCTGACCATATCGCACTATGTATCATTTGCTGACCCAAGAAATGTTTCT-----					
14	-CTTGTTCTGACCATATCGCACTATGTATCATTTGCTGACCCAAGAAATGTTTCT-----					
15	-CTTGTTCTGACCATATCGCACTATGTATCATTTGCTGACCCAAGAAATGTTTCT-----					
16	-CTTGTTCTGACCATATTGCACTATGTATTATTTGCTGACCCAAGAAATGTTTCT-----					
17	-CTTGTTCTGACCATATCGCACTATGTATCATTTGCTGACCCAAGAAATGTTTCT-----					
18	-CTTGTTCTGACCATATCGCACTATGTATCATTTGCTGACCCAAGAAATGTTTCT-----					
19	-CTTGTTCTGACCATATCGCACTATGTATTATTTGCTGACCCAAGAAATTTTTCT-----					
20	-CTTGTTCTGACCATATCGCACTATGTATCATTTGCTGACCCAAGAAATGTTTCT-----					
21	-ATTGTTCTGACCATATCGCACTATGTATCATTTGCTGACCCAAGAAATGTTTCT-----					
22	-CTTGTTCTGACCATATCGCACTATGTATCATTTGCTGACCCAAGAAATGTTTCT-----					
23	-CTTGTTCTGACCATATCGCACTATGTATCATTTGCTGACCCAAGAAATGTTTCT-----					
24	-CTTGTTCTGACCATATCGCACTATGTATCATTTGCTGACCCAAGAAATGTTTCT-----					
25	ACTTGTTCTGACCATATCGCACTATGTATCATTTGCTGACCCAAGAAATATTTCT-----					
26	-CTTGTTCTGACCATATCGCACTATGTATCATTTGCTGACCCAAGAAATGCTTCT-----					
27	-CTTGTTCTGACCATATCGCACTATGTATTATTTGCTGACCCAAGAAATTTTTCT-----					
28	-CTTGTTCTGACCATATCGCACTATGTATCATTTGCTGACCCAAGAAATGTTTCT-----					

Table 4.5. Continued

Taxa[a]	Nucleotide sites[b]
	790 800 810 820 830 840 . . . . . .
1	-GGTTCAAGTCTAATTAATAATTAAAAATGGAAGAATTAAAAGGATATTTAGAAAAAACT
2	-GGTTCAAGTCTAATTAATAATTAAAAATGGAAGAATTAAAAGGATATTTAGAAAAAACT
3	-GGTTCAAGTCTAATTAATAATTAAAAATGGAAGAATTAAAAGGATATTTAGAAAAAACT
4	-GGTTCAAGTCTAATTAATAATTAAAAATGGAAGAATTAAAAGGATATTTAGAAAAAACT
5	-GGTTCAAGTCTAATTAATAATTAAAAATGGAAGAATTAAAAGGATATTGAGAAAAAACT
6	-GGTTCAAGTCTAATTAATAATTAAAAATGGAAGAATTAAAAGGATATTTAGAAAAAACT
7	-GGTTCAAGTCTAATTAATAATTTAAAATGGAAGAATTAAAAGGATATTTAGAAAAAACT
8	-GGTTCAAGTCTAATTAATAATTAAAAATGGAAGAATTAAAAGGATATTTAGAAAAAACT
9	-GGTTCAAGTCTAATTAATAATTAAAAATGGAAGAATTAAAAGGATATTTAGAAAAAACT
10	-GGTTCAAGTCTAATTAATAATTAAAAATGGAAGAATTAAAAGGATATTTAGAAAAAACT
11	-GGTTCAAGTCTAATTAATAATTAAAAATGGAAGAATTAAAAGGATATTTAGAAAAAACT
12	TGGTTCAAGTCTAATTAATAATTAAAAATGGAAGAATTAAAAGGATATTTAGAAAAAACT
13	-GGTTCAAGTCTAATTAATAATTTAAAATGGAAGAATTAAAAGGATATTTAGAAAAAACT
14	-GGTTCAAGTTTAATTAATAATTAAAAATGGAAGAATTAAAAGGATATTTAGAAAAAACT
15	-GGTTCAAGTCTAATTAATAATTAAAAATGGAAGAATTAAAAGGATATTTAGAAAAAACT
16	-GGTTCAAGTCTAATTAATAATGAAAAATGGAAGAATTAAAAGGATATTTAGAAAAAACT
17	-GGTTCAAGTCTAATTAATAATTAAAAATGGAAGAATTAAAAGGATATTTTGAAAAAACT
18	-GGTTCAAGTCTAATTAATAATTAAAAATGGAAGAATTAAAAGGATATTTTGAAAAAACT
19	-GGTTCAAGTCTAATTAATAATTAAAAATGGAAGAATTAAAAGGATATTTAGAAAAAACT
20	-GGTTCAAGTCTAATTAATAATTTAAAATGGAAGAATTAAAAGGATATTTAGAAAAAACT
21	-GGTTCAAGTCTAATTAATAATTTAAAATGGAAGAATTAAAAGGATATTTAGAAAAAACT
22	-GGTTCAAGTTTAATTAATAATTAAAAATGGAAGAATTAAAAGGATATTTAGAAAAAACT
23	-GGTTCAAGTCTAATTAATAATTAAAAATGGAAGAATTAAAAGGATATTTAGAAAAAACT
24	-GGTTCAAGTTTAATTAATAATTAAAAATGGAAGAATTAAAAGGATATTTAGAAAAAACT
25	-GGTTCAAGTCTAATTAATAATTAAAAATGGAAGAATTAAAAGGATATTTAGAAAAAACT
26	-GGTTCAAGTCTAATTAATAATTTAAAATGGAAGAATTAAAAGGATATTTAGAAAAAACT
27	-GGTTCAAGTCTAATTAATAATTAAAAATGGAAGAATTAAAAGGATATTTAGAAAAAACT
28	-GGTTCAAGTCTAATTA------AAAAATGGAAGAATTAAAAGGATATTTCGAAAAAACT

Table 4.5. Continued

Taxa[a]	Nucleotide sites[b]
	8 8 8 8 8 5 6 7 8 9 0 0 0 0 0 . . . . .
1	AGATCTATACAACAACACTTCCTATACCCGCTTCTCTTTCAAGAGTCTAT
2	AGATCTAAACAACAACACTTCCTATACCCGCTTCTCTTTCAAGAGTCTAT
3	AGATCTAAACAACAACACTTCCTATACCCGCTTCTCTTTCAAGAGTCTAT
4	AGATCTAAACAACAACACTTCCTATACCCG-TTCTCTTTCAAGAGTCTAT
5	AGATCTAAACAACAACACTTCCTATACCCGCTTCTCTTTCAAGAGTCTAT
6	AGATCTAAACAACAACACTTCCTATACCCGCTTCTCTTTCAAGAGTCTAT
7	AGATCTAAACAACAACACTTCTTATACCCGCTTCTCTTTCAAGAGTCTAT
8	AGATCTAAACAACAACACTTCTTATACCCGCTTCTCTTTCAAATATC-AT
9	AGATCTAAACAACAACACTTCTTATACCCGCTTCTCTTTCAAGAGTCTAT
10	AGATCTAAACAACAACACTTCTTATACCCGCTTCTCTTTCAAGAGTCTAT
11	AGATCTAAACAACAACACTTCTTATACCCGCTTCTCTTTCAAGAGTCTAT
12	AGATCTAAACAACAACACTTCTTATACCCGCTTCTCTTTCAAGAGTCTAT
13	AGATCTAAACAACAACACTTCTTATACCCGCTTCTCTTCCAAGAGTCTAT
14	AGATCTAAACAACAACACTTCTTATACCCGCTTCTCTTTCAAGAGTCTAT
15	AGATCTAAACAACAACACTTCTTATACCCGCTTCTCTTTCAAGAGTCTAT
16	AGATCTAAACAACAACACTTCTTATACCCGCTTCTCTTTCAAGAGTCTAT
17	AGATCTAAACAACAACACTTCTTATACCCGCTTCTCTTTCAAGAATCTAT
18	AGATCTAAACAACAACACTTCTTATACCCGCTTCTCTTTCAAAATATCAT
19	AGATCTAAACAACAACACTTCTTATACCCGCTTCTCTTTCAAGAGTCTAT
20	AGATCTAAACAACAACACTTCTTATACCCGCTTCTCTTTCAAGAGTCTAT
21	AGATCTAAACAACAACACTTCCTATACCCGCTTCTCCTTCAAGAGTCTAT
22	AGATCTAAACAACAACACTTCCTATACCCGCTTCTCTTTCAAGAGTCTAT
23	AGATCTAAACAACAACACTTCYTATACCCGCTTCTCTTTCAAGAGTCTAT
24	AGATCTAAACAACAACACTTCCTATACCCGCTTCTCTTTCAAGAGTCTAT
25	GGATCTAAACAACAACACTTCTTATACCCGCTTCTCTTTCAAGAGTCTAT
26	GGATCTAAACAACAACACTTCTTATACCCGCTTCTCTTTCAAGAGTCTAT
27	AGATCTAAACAACAACACTTCTTATACCCGCTTCTCTTTCAAGAGTCTAT
28	AGATCGAAACAACAACACTTCCTATACCCGCTTCTCTTTCAAGAGTCTAT

Note. Vertical columns are nucleotide positions. Horizontal rows are individual DNA sequences. Positions are numbered consecutively beginning with 5, end of the 5' *trn*K intron ending at 141 bp of the *mat*K gene sequence. Arrow mark the borders of the 3' end of the 5' *trn*K intron and the start codon of the *mat*K gene.

a;

1. *Cryptocoryne spiralis* NJ3129a2,
2. *C. beckettii* 1671,
3. *C. x willisii* P1978/5045,
4. *C. walkeri* NJ23-3,
5. *C. wendtii* P1961/5342,
6. *C. undulata* NJ22-7,
7. *C. parva* NJ3405,
8. *C. cognata* 28Yadav,
9. *C. pontederiifolia* USM9638,
10. *C. moehlmannii* P1989/5046,
11. *C. pygmaea* NJ3962,
12. *C. affinis* USM8065,
13. *C. minima* S1995/9201,
14. *C. griffithii* NJ85-30,
15. *C. purpurea* Othman s.n.,
16. *C. cordata* USM9139,
17. *C. elliptica* USM8069
18. *C. schulzei* USM8087,
19. *C. aponogetifolia* P3401,
20. *C. annamica* M92/3205,
21. *C. alba* NJ3172-6,
22. *C. albida* P1958/5363,
23. *C. retrospiralis* P1977/5146,
24. *C. crispatula* var. *balansae* NJ3406,
25. *C. ciliata* P1958/6013,
26. *C. ferruginea* NJ78-44,
27. *C. usteriana* P1985/5448,
28. *Lagenandra meeboldi* P1979/5019,

b; sequence symbols: A=dATP; C=dCTP; G=dGTP, T=dTTP; R=A/G; Y=C/T; hyphens = insertion or deletions

5' *trn*K intron over the *Cryptocoryne* and *Lagenandra* taxa examined. Within *Cryptocoryne,* one or more gaps were required at 78 positions or 10.5% of sites and of these gaps, seven were autapomorphic. With the inclusion of *Lagenandra,* a further five gaps (all autapomorphic) were needed. Following alignment, it was evident that 57 or 7.7% of positions were variable and, of these, 26 were potentially informative phylogenetically.

The sequence of the first 141bp of the *mat*K gene was also used in the analysis. Alignment of these sequences required gaps to be placed at six positions in all *Cryptocoryne* species except *C. affinis.* This latter species, therefore, contained an insertion of 6 bp within this part of the sequence. A deletion of 6 bp was detected in *Lagenandra,* but was absent from all *Cryptocoryne* species examined. The alignment of the partial *mat*K gene sequence resulted in a matrix of 147 characters (from position 744 to 890 in Table 4.5), and it was evident that 20 or 14.2% of these positions were variable and 11 were potentially informative. In regard to the gaps that were scored, 12 were potentially informative phylogenetically (Table 4.5).

Phylogenetic analysis of the 5' *trn*K intron and the partial *mat*K gene sequence resulted in the production of 12 equally parsimonious trees that were 148 steps in length with a consistency index of 0.764 and a retention index of 0.715. One of the 12 equally parsimonious trees is presented in Figure 4.4. The phylogeny, rooted with respect to *Lagenandra meeboldi* showed there present two 'clades' within *Cryptocoryne* . One of the two 'clades' contains *Cryptocoryne* species from India and mainland Asia, while the other is comprised of species mostly from Sri Lanka and Malesia.

Table 4.6. List of gaps present in data matrix in table 4.5 coded as additional characters used in phylogenetic analysis of the 5' *trn*K intron and the *mat*K gene sequence. Taxa 1-28 are in the same order as the one listed in table 4.5., 0=absence of gap, 1= present of gap.

Character	Position of gap/gaps from 5' *trn*K intron and partial *mat*K gene data matrix	Taxa 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8
1	1-2	0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
2	2-3	0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
3	3	0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
4	5	0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
5	7	1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
6	10-14	0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
7	13-14	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
8	14	1 0 0 1 1 0 0 1 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 1 1 0
9	120-125	0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0
10	120-131	1 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1
11	132	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0
12	147-149	0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
13	150-157	1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
14	350-354	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
15	453-457	1 1 1 1 1 1 1 1 1 1 0 1 1 0 0 1 0 0 1 1 1 1 1 1 1 1 1 1
16	552-557	1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
17	565	0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
18	570	1 1 1 1 1 0 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0
19	570-571	0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
20	607	1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
21	675	1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
22	675-681	0 0 0 0 0 0 0 0 0 0 1 0 0 1 1 0 1 1 0 0 0 0 0 0 0 0 0 0
23	687-699	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0
24	687-701	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 0 1
25	715-721	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1
26*	776-781	1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
27*	798-863	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1

Note. * Denote characters/gaps present within *mat*K gene

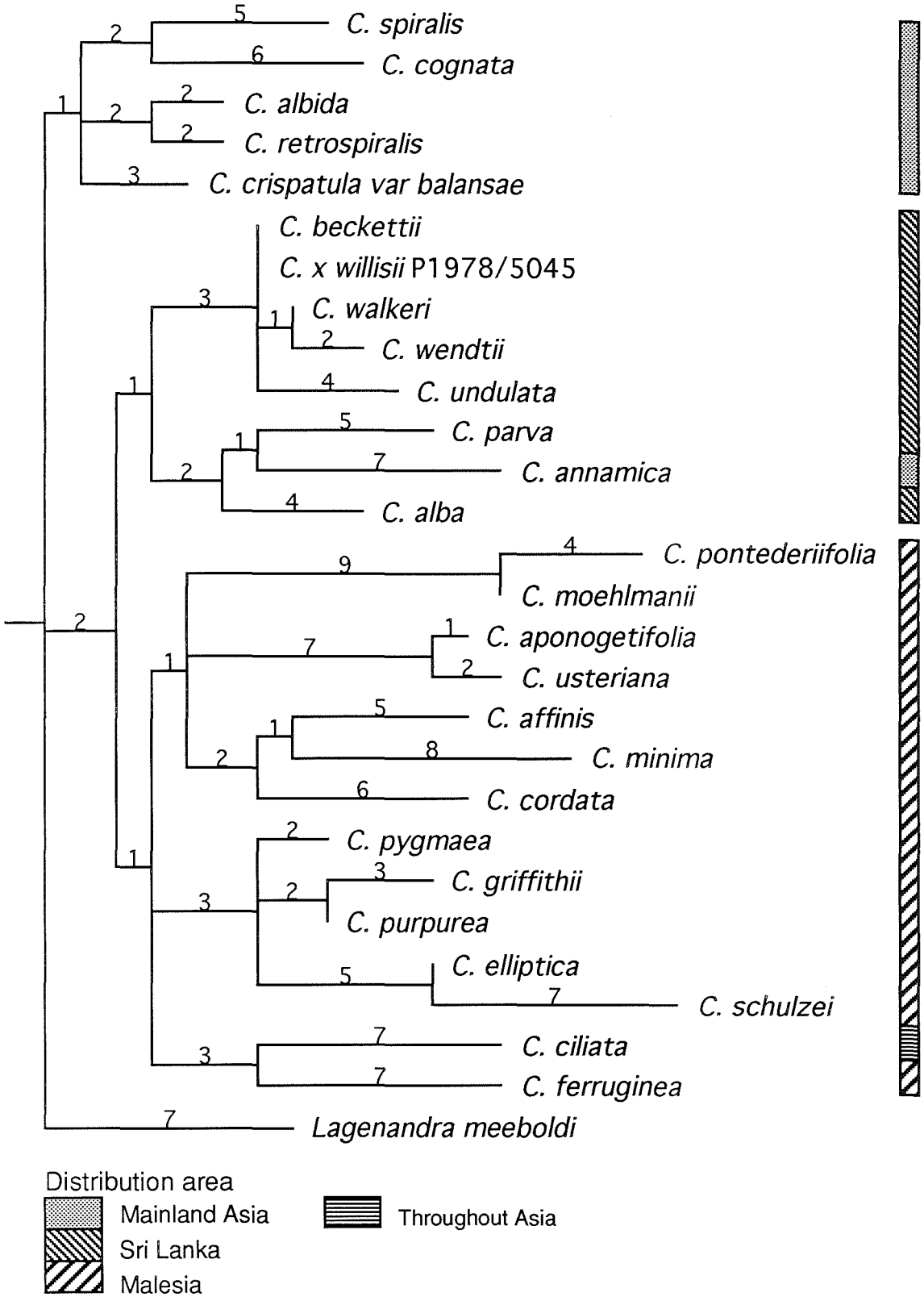


Figure 4.4. One of the 12 most parsimonious trees generated from the 5' *trn*K intron and partial *mat*K gene sequence from 27 *Cryptocoryne* species and one species of the outgroup genus *Lagenandra*. The numbers above the branch are the number of steps. The tree is 148 in length with CI=0.764 and RI=0.715.

The first clade contains *C. spiralis* and *C. cognata* (which appear to be closely related), *C. albida,* and *C. retrospiralis* (also closely related), and *C. crispatula* var. *balansae.* Within the second clade two subclades are recognised. The first subclade is comprised of species mainly from Sri Lanka, while the second subclade contains species mainly from Malesia. The first subclade can be subdivided further into two species groups, with one group containing *C. beckettii, C. x willisii, C. walkeri, C. wendtii* and *C. undulata* all of which occur in Sri Lanka, while the other species group is comprised of *C. alba, C. parva* and *C. annamica.* The positioning of the three species in this latter group is somewhat surprising as *C. annamica* is native to Vietnam (Mainland Asia), while *C. alba* and *C. parva* occur only in Sri Lanka.

Within the second subclade, which is comprised only of *Cryptocoryne* species from South East Asia, *C. pontederiifolia* is closely allied to *C. moehlmannii* (both from Sumatra), *C aponogetifolia* is closely related to *C. usteriana,* while *C. ciliata* groups with *C. ferruginea.*

From the 12 equally parsimonious trees, a 50% majority rule consensus tree (Fig. 4.5) and a strict consensus tree (Fig 4.6) including the bootstrap values were produced. Both types of consensus trees were similar in topology in regard to the clade containing the Malesian species, *C. minima, C. cordata* and *C. affinis.* Whereas The 50% majority consensus tree placed these three species into a subclade, the strict consensus collapsed this subclade into a polytomy. The same topology was also observed in one of the 12 most parsimonious trees given in Fig. 4.4 and it differs from the strict consensus tree in regard to the three Malesian species mentioned above i.e. they were placed into one group. In addition,

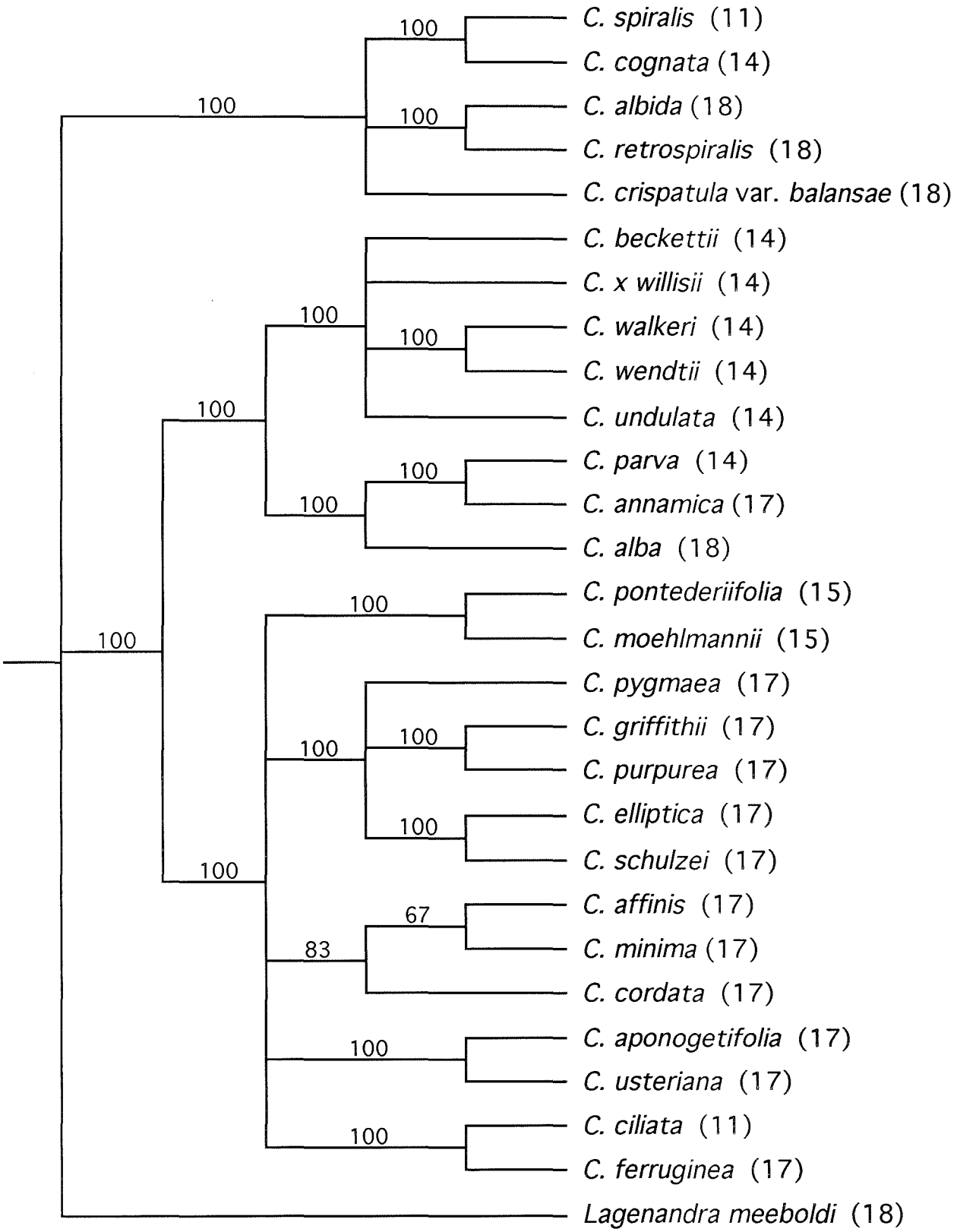


Figure 4.5. 50% majority-rule consensus tree of the 12 equally parsimonious trees. Values on each branches reflect the percentage of each branch present in the 12 trees. Numbers in parenthesis after each taxa are the chromosome base numbers.

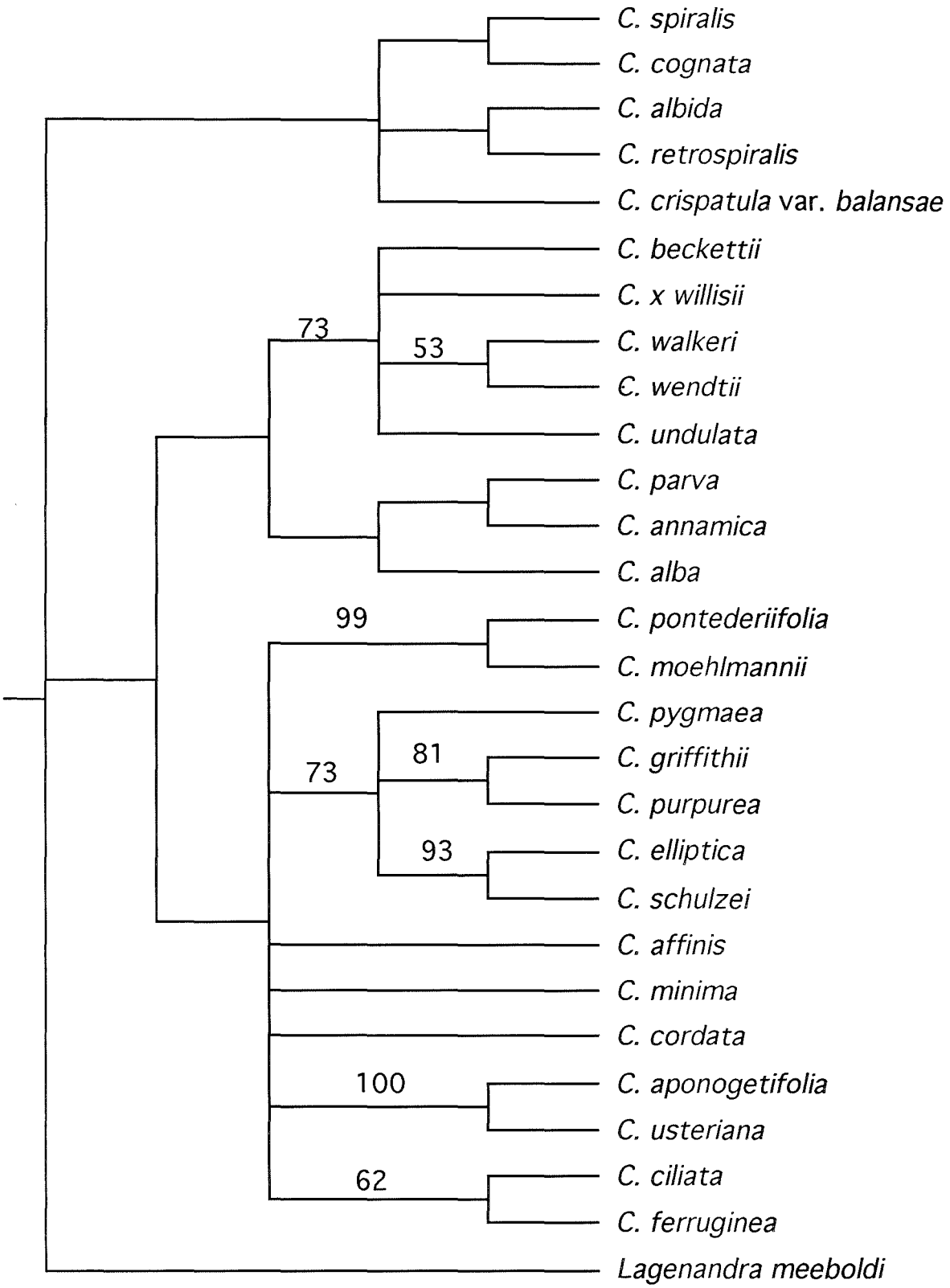


Figure 4.6. A strict consensus of the 12 most parsimonious tree generated from the cpDNA data. Values given above the branches are bootstrap values (100 replicates).

in Fig. 4.4, the species group containing *C. minima, C. cordata* and *C. affinis* was placed within a subclade containing three species group, including *C. aponogetifolia* and *C. usteriana* which as one species group, and *C. pontederiifolia* and *C. moehlmannii* as another species group. In the strict consensus (Fig. 4.6) this subclade was collapsed and only the latter two groups could be recognised as separate from one another.

4.4 Discussion

Despite the ease with which restriction site analysis of amplified cpDNA regions may be carried out, in the present study it did not provide enough variation for an accurate phylogenetic resolution of the taxa examined. A contributing factor which might have hampered the detection of variation was the use of agarose gel. This might be inappropriate for visualising variation between small fragments produced with a 4 base cutter restriction enzyme. A polyacrylamide gel is likely to enable better resolution of fragments of small size or that differ in size by a single nucleotide; however, it was envisaged that the level of variation detected when using such gel would still be insufficient for a satisfactory phylogenetic analysis. Restriction analysis of the total cpDNA genome also did not generate satisfactory results, mainly due to problems in the hybridisation of probe DNA to the digested *Cryptocoryne* DNA. Use of more suitable probes might overcome these problems in the future. Of the three approaches used in the research reported in this chapter, sequencing of a particular segment of the cpDNA genome provided sufficient variation to allow analysis of phylogenetic relationships within *Cryptocoryne* to be conducted.

4.4.1 Phylogenetic relationships based on cpDNA sequence variation

Twelve most parsimonious phylogenetic trees generated for *Cryptocoryne* species from the sequence of the 5' *trn*K intron and the first 141 bp of the *mat*K gene sequence were only partially resolved due to a low number of synapomorphic characters recorded. This was the case even

though indels were included as characters. One of the 12 most parsimonious trees obtained is shown in Figure 4.4. This together with the 50% majority rule consensus tree (Fig. 4.5) and the strict consensus tree (Fig 4.6) showed that after rooted with respect to *Lagenandra*, there present two clades within *Cryptocoryne*. Clade 1 contained species from mainland Asia, i.e. *C. spiralis*, *C. cognata*, *C. albida*, *C. retrospiralis* and *C. crispatula* var. *balansae*, while clade 2 was comprised of the remaining species examined. Generally the topologies of the two consensus trees were similar except in regard to the grouping of three Malay Peninsula species, i.e. *C. cordata*, *C. affinis* and *C. minima*. In the 50% majority rule consensus tree, *C. cordata* was placed as sister taxon to the other two species, while in the strict consensus tree this association was collapsed into a polytomy.

C. albida, *C. crispatula* var. *balansae* and *C. retrospiralis*, which are all members of the mainland Asia clade, were previously considered to be very closely related by Jacobsen (1980a) based on morphological similarity, particularly in regard to the structure of the spathe. The placement of these three species together based on cpDNA sequence information, adds support to their close relationships deduced from morphological affinities. The two other taxa present within this clade were *C. spiralis* and *C. cognata*, which are both native to India. However, the relationship of these two species to the three other member of the clade was not resolved. The placement of *C. cognata* within this clade was somewhat surprising given that its chromosome number of 2n=28 is the same as that of several species native to Sri Lanka. Moreover, in regard to leaf and spathe structure, *C. cognata* appears to resemble more closely Sri Lankan species rather than those from mainland Asia, particularly *C. spiralis*, which is the

only species in the genus which possesses a spathe that lacks a tube.

Cryptocoryne species from Sri Lanka have traditionally been divided into two taxonomic groups, one of which is called the *C. beckettii* group, while the other is composed of *C. alba, C. thwaitesii* and *C. bogneri* (Jacobsen, 1977). Unfortunately, neither of the latter two species were available for analysis. It was of interest, nonetheless, that the analysis of cpDNA sequence variation also divided the Sri Lankan species into two groups. In one of these groups was placed *C. beckettii, C. walkeri, C . wendtii, C. undulata* and *C. x willisii.* Certain species within this group are known to hybridise with each other in areas of overlapping natural habitats, for example a fairly large population of the hybrid *C. x willisii* has been found in a natural site where both putative parental taxa, *C. walkeri* and *C. parva* occurred (de Graaf and Arends, 1986). Hybridisation is most likely to have increased in recent years as these species have invaded areas disturbed by man. A clear resolution of the phylogenetic relationships within this group of taxa was not obtained from the cpDNA sequence variation and it would seem that improved resolution will only be achieved if a more rapidly evolving cpDNA sequence becomes available for use.

The second species group within the Sri Lankan sub-clade consisted of *C. alba* and *C. parva* which are both native to the island, and *C. annamica* which is native to Vietnam on mainland Asia. This grouping is surprising in that there is no similarity on the basis of morphological characters, geographical distribution or chromosome numbers to provide support for a close association between *C. annamica* and *C. alba* or *C. parva.* Moreover, the phylogeny constructed from sequencing the internal

transcribed spacer of nuclear ribosomal DN (chapter 3) placed *C. annamica* in a Malesian clade in close association with the Philippines' species *C. aponogetifolia*. The occurrence of this latter association finds support also from the similarity in overall morphological structure (leaf and spathe) between *C. annamica* and *C. aponogetifolia* (Jacobsen per. comm.). It is difficult at the moment to account for the particular discordance between the two phylogenies constructed from ITS and cpDNA sequences in regard to the relationships between these species.

The placement of *C. parva* within this particular group also requires explanation. *C. parva* with 2n=28, was thought to be more closely related to species which comprise the first Sri Lankan sub-clade, i.e. the sub-clade that includes *C. beckettii, C. walkeri, C wendtii, C. undulata* and *C. x willisii*. Species within this group normally have a spathe with a long tube (except for *C. parva*) and a limb with a prominent collar (except for *C. walkeri*). In contrast, *C. alba* has a spathe with a long smooth caudate limb and lacks a collar. Moreover, whereas the "*C. beckettii*" group (including *C. parva*) occurs in spring and river habitats, *C. alba* is found in a dark forest habitat i.e. within deep creeks which are heavily shaded by trees and shrubs (de Graaf and Arends, 1986). Finally, *C. alba* (2n=36) has never been known to hybridise with members of *Cryptocoryne* that possess 2n=28 chromosomes.

The second sub-clade within clade 2 of the cpDNA phylogeny consisted of species mainly from the Malesia floristic region. Although clear relationships between species within this clade remained unresolved, several points can be made. A close relationships was evident between the two Sumatran species *C. pontederiifolia* and *C. moehlmannii,* as might

have been expected from their similar morphologies, chromosome numbers and ITS sequence (chapter 3). Similarly, two Philippine species, *C. aponogetifolia* and *C. usteriana* were grouped together, although another Philippine species, *C. pygmaea*, was embedded in a grouping comprised of species from the Malay Peninsula (*C. griffithii, C. purpurea, C. elliptica* and *C. schulzei*). Arends *et al.* (1982) placed *C. pygmaea* and *C. usteriana* into separate taxonomic group based on morphological comparisons and, Jacobsen (pers. comm.), suggested that *C. aponogetifolia, C. usteriana* and *C. annamica* are more closely related to one another than *C. pygmaea* is to either of these species.

Another pair of species which form a sister group within this sub-clade includes *C. ciliata* and *C. ferruginea*. This particular pairing in the phylogeny is surprising in view of the fact that the two species differ in chromosome number and morphology. *C. ferruginea* which is only present in Borneo, has a chromosome number of 2n=34 and possesses a spathe with a caudate limb. In contrast, *C. ciliata* has 2n=22 chromosomes and is the most widely distributed *Cryptocoryne* species known, being found from India to Indo-China, Malaysia, Indonesia, Borneo and New Guinea. However, in a locality in Borneo i.e. in a stream at Batu Kitang, a population of *C. ferruginea* has been reported to grow alongside *C. lingua* in fresh and brackish water tidal zone with isolated populations of *C. ciliata* further downstream (Jacobsen, 1980b). It is plausible that these species once grew in contact with each other and that hybridisation and an exchange of cpDNA genomes occurred between them. This may have resulted in the capture by one of the species of the other species cpDNA genome which would explain the unexpected similarity between these two species in respect to their cpDNA. Event of chloroplast capture

through hybridisation or introgression is not uncommon and it can occur at a variety of taxonomical level, and in some instances can occur in plant group which are not noted for hybridisation (Rieseberg and Soltis, 1991). However, it is difficult to ascertain if this is the case with *C. ciliata* and *C. ferruginea*, especially as the origin of *C. ciliata* is unknown and the sampling was extremely limited.

A larger grouping within the Malesian subclade was comprised of *C. pygmaea*, *C. griffithii*, *C. purpurea*, *C. elliptica* and *C. schulzei*. The close association between *C. griffithii* and *C. purpurea* is of interest as *C. purpurea* has been assumed to be of hybrid origin exhibiting low pollen fertility and a spathe which combines a broad collar zone typical of *C. cordata*, and a purple limb typical of *C. griffithii* (Jacobsen, 1977). Moreover, *C. purpurea* occurs in the Malay Peninsula in areas where its putative parents, *C. cordata* and *C. griffithii* once overlapped. Whereas a comparison of the ITS sequences (Chapter 3) revealed a close association between *C. purpurea* to *C. cordata*, the cpDNA result links *C. purpurea* more closely to *C. griffithii*. It seems feasible, therefore, that *C. purpurea* is indeed of hybrid origin with *C. griffithii* having acted as the maternal parent while *C. cordata* was the male parent.

Finally, it is of interest to note that the cpDNA based phylogeny placed *C. ciliata* and *C. spiralis* into two quite separate clades. This was also the case in the phylogeny reconstructed from ITS sequence variation (Chapter 3). As both of these species have a chromosome number of x=11, it might be concluded that in the course of evolution of the genus *Cryptocoryne*, the chromosome base number x=11 arose on at least two independent occasions.

4.4.2 Phylogeography

There is some correspondence between the geographic pattern of cpDNA variation recorded in *Cryptocoryne* and the cpDNA phylogeny resolved for the genus. The three major clades of the genus which contain species native to three different geographic areas based on ITS sequence variation,were also partly recognised in the cpDNA phylogeny. One particular anomaly, however, concerned the positioning of *C. annamica* (a mainland Asian species assumed to be closely related to *C. aponogetifolia* - species from Philippines) within the morphologically established Sri Lankan species sub-clade.

In contrast with the nrDNA ITS tree, the cpDNA tree was more poorly resolved and consequently caution must be exercised in regard to detailed phylogenetic interpretation.

Chapter 5

RAPD variation among *Cryptocoryne* species

5.1 Introduction

The procedure for resolving randomly amplified polymorphic DNA (RAPD) was introduced by Williams *et al.* (1990) as a method for detecting DNA polymorphisms and genetic markers in plants. This PCR based technique involves the amplification of random DNA fragments using a single primer of arbitrary nucleotide sequence of 10 base in length. The primer sites are thought to be randomly distributed and polymorphism between RAPD profiles can be attributed to a range of processes, including nucleotide substitution which create or destroy primer sites, and insertion, deletion or inversion of either priming sites or segments between priming sites (Williams *et al.*, 1993). In addition to polymorphism in product size distribution within a RAPD profile, polymorphism with respect to product intensity can also be expected (Caetano-Anolles *et al.*, 1991; Williams *et al.*, 1990; Williams *et al.*, 1993). Polymorphism in product intensity might be the result of low copy number of products, competition between molecular species, heterozygosity or partial mismatching of primer sites (Adams and Demeke, 1993; Venugopal e*t al.*, 1993; Williams *et al.*, 1993).

The resolution of RAPDs offers several advantages over other molecular techniques among detecting DNA variations in plants. First, as with other PCR based techniques, the method requires a small quantity of DNA for a reaction, and so only a small amount of leaf material is needed for DNA extraction. This is an attractive option especially when working with limited material such as herbarium specimens or plants with relatively few leaves. Further, the method is quick in that it can be completed within hours, and is fully automated such that a large number of samples can be handled at a particular time. Additionally, no prior

knowledge of the DNA sequence of the plant genome is required for the employment of RAPD primers and these primers can be used on a relatively wide array of plant taxa.

Due to its technical simplicity, the procedure that resolves RAPDs has been applied widely in various studies of plant diversity; for example, in studies of plant population genetics, fingerprinting plant genome for species or cultivar identification, as well as the determination of hybrids. The method has also been employed in the construction of plant phylogenies. A brief review of each of these applications is given below.

5.1.1 Population genetics

RAPD has contributed considerably in studies of plant population genetics. For example, Yeh *et al.* (1995) used RAPDs to assess the level of variation within and among natural population of trembling aspen, *Populus tremuloides*. The results showed that there was high RAPD variation within but less among populations of this species. These results were congruent with those obtained from isozyme analysis (Cheliak and Dancik, 1982; Jelinski and Cheliak, 1992) and support the conclusion reached by Hamrick and Godt (1989) that the long lived, wind pollinated, outcrossing trembling aspen with a wide and continuous range will retain considerable amounts of genetic diversity, though exhibit little genetic differentiation among populations.

Similarly, Nesbitt *et al.* (1995) used RAPDs to assess the genetic structure of populations of the forest tree species, *Eucalyptus globulus*, from Australia. This species has been divided into four subspecies. In

many localities where it occurs, it is considered to be intermediate morphologically between subspecies *globulus* and the other three subspecies. RAPD analysis, however, did not reveal an intermediate phenotype for speciments from these localities. Again, within *E. globulus*, the majority of RAPD variation was greater within rather than between populations. In contrast, in *Hordeum spontaneum*, a species that shows a high level of selfing, RAPD variation was low within populations (Dawson *et al.*, 1993).

5.1.2 Fingerprinting plant genomes

One of the main applications of RAPDs is fingerprinting plant genomes for species or cultivar identification. One example of this use is in the elit poplar (*Populus* spp.), which is widely used for wood production in temperate regions. This fast growing and short rotation tree species is in high demand for a wide array of wood products ranging from plywood to paper. The high success of interspecific hybridisation and the ease of producing asexual propagated clones has made poplar a good species for commercial plantations, and clones have been exchanged among countries around the world. Castiglione *et al.* (1993) used RAPDs to fingerprint 32 clones of different species of poplar which covered almost 50% of the world cultivated poplars.

Similarly Virk *et al.* (1995) used RAPDs to fingerprint different rice cultivars. Maintenance and exploitation of rice germplasm have been difficult due to the large number of accessions available. To increase the efficiency of handling large numbers of accessions, duplicates were removed and a 'core' collection was established. Virk *et al.* (1995) found

that RAPD provided an efficient method to classify and identify previously uncharacterised accessions of *Oryza japonica* and *O. indica*. RAPDs have also proved satisfactory in fingerprinting cultivars of *Musa* (Howell *et al.*, 1994), snap bean (*Phaseolus vulgaris*; Skroch and Nienhuis, 1995) and species of *Brassica* (Demeke *et al.*, 1992).

5.1.3 Investigation on plant of putative hybrid origin

RAPD markers can be used successfully in studies aimed at determining the parentage of a hybrid species. For example, Brochmann *et al.* (1996) showed that the putative allopolyploid *Saxifraga osloensis* exhibited an additive RAPD profile possessing 11 markers of *S. adscendens* (one of its putative parents) and eight markers of *S. tridactylis* (the other putative parent). Similarly Wang *et al.* (1994) reported that almost all RAPD bands resolved in *Paulownia taiwaniana* were shared with either *P. fortunei* or *P. kawakamii* or both. Moreover, the number of polymorphic bands between *P. taiwaniana* and *P. fortunei* was roughly equivalent to that number shared between *P. taiwaniana* and *P. kawakamii*. These findings together with the results of a survey of cpDNA-RFLP variation provided a strong evidence that *P. taiwaniana* was a hybrid product of *P. fortunei* and *P. kawakamii*.

Crawford *et al.* (1993) used RAPD markers to test the hypothesis of the origin of the intergeneric hybrid *x Margyracaena skottsbergii*, a genus endemic to the island Masatierra in the Juan Fernandez archipelago. *x Margyracaena skottsbergii* has been interpreted as a hybrid between *M. digynus* (endemic to Masatierra) and *Acaena argentea* (an introduced species). Results from RAPD analysis showed the presence of 18 species-

specific bands from *A. argentea* and 27 from *M. digynus* in *x Margyracaena skottsbergii* thus providing strong support for its hybrid origin. In contrast, RAPD markers did not provide any evidence that *A. ovalifolia* (a species native to Masatierra) as one of the parents of the hybrid genus. Other studies of similar nature include that of Marsolais *et al.* (1993) and Arnold (1993).

5.1.4 Phylogenetic analysis

RAPD variation has also been used for reconstructing plant phylogenies. An example of a study of this kind is that by Graham and Nichol (1995) who constructed a dendrogram for 13 species of raspberry (*Rubus*) using RAPD markers. The 13 species used represented three subgenera (*Idaeobats, Eubats* and *Anoplobats*) of the twelve recognised in *Rubus*. A total of 372 polymorphic markers were obtained from ten primers. A dendrogram separated the species into the three subgenera with the exception of *R. macrae,* a rare tropical species which traditionally was placed within the subgenus *Idaeobats*. RAPD data showed that *R. macrea* had only 26% similarity to other species of *Idaeobats,* which was equivalent to its level of similarity percentage to species of *Eubats*.

Similarly, RAPDs have been used in assessing the relationship between autumn buttercup (*Ranunculus acriformis* var. *aestivalis*) and other closely related *Ranunculus* species (Van Buren *et al.,* 1994). Over 350 markers were generated from 23 random primers, and the RAPD data were subjected to both phenetic and cladistic analysis. Results from the cladistic analysis were congruent with those of the phenetic analysis, in

that five distinct clades synonymous with the five taxa included were recognised. As a consequence, the autumn buttercup was elevated to species rank, i.e. *R. aestivalis.*

Lifante and Aguinagalde (1996) used RAPD markers to ratify the specific status of three species of *Asphodelus* sect. *Verinea,* i.e. *Asphodelus fistulosus, A. ayardii* and *A. tenuifolius.* Five populations of each species were examined in this study and eight RAPD primers were used to generate RAPD bands. UPGMA analysis of the similarity matrix resulted in a dendrogram with three distinct clusters, one for each species. Interpopulation diversity was highest in *A. tenuifolius* (inbreeder) followed by *A. ayardii* (outbreeder) and lowest in *A. fistulosus* (inbreeder). It was suggested that the discrepancy in interpopulation diversity between the two inbreeding species was because *A. fistulosus* is of recent origin. Further, *A. fistulosus* was observed to lack species-specific RAPD bands and produced fragment pattern that indicated an amphidiploid origin from *A. tenuifolius* and *A. ayardii.*

Stammers *et al.* (1995) used RAPD to distinguish species-specific markers and phylogenetic relationships between closely related species of *Lolium* and *Festuca.* The phylogenetic tree constructed using the least squares method of Fitch and Margoliash (1967) was generally in agreement with previous taxonomic treatment. For example, *Festuca pratensis* was placed closer to members of *Lolium* than to species of *Festuca* which was also indicated by seed protein variation. The results of a survey of cpDNA variation also showed *F. pratensis* to be in close affinity with *L. multiflorum,* and divergent from other *Festuca.* Somewhat unexpectedly,

the RAPD analysis further showed that two taxonomically similar *Festuca* species, *F. drymeja* and *F. lasto,* were widely separated from each other in the RAPD generated phylogram.

Despite the relative ease and wide applications of the RAPD methodology, caution must be taken when using RAPD generated fragments for phylogenetic purposes. There are intellectual as well as practical problems in using RAPD for phylogenetic purposes. The main concern involves the assumption that fragments of similar mobility observed in an agarose gel are homologous DNA sequences. However, fragments of the same size may not be homologous due to size convergence or limitation in gel resolution (Smith *et al.*, 1994) and deletion, insertion and primer loss/gain events (Williams *et al.*, 1993). In addition primer site distribution may result in products that display partial homology, thus are not independent (Smith *et al.*, 1994). The presence of non-homologous co-migrating fragments might provide information on convergence rather than close relationships between taxa (Stammers *et al.*, 1995).

In theory, RAPD technique randomly samples the entire plant genome, i.e. nuclear, mitochondrial and chloroplast genomes, which might include repetitive or single copy DNA sequence (Williams *et al.*, 1993). However, not much is known in regard to the distribution of such genomes or sequences, both within and between RAPD profiles (Smith *et al.*, 1994). McCauley (1995) suggested that the distribution of organelle (mitochondria and chloroplast) and nuclear markers within a RAPD profile might have important consequences for data interpretation since the population genetics of the two genomes are different.

Characters used in phylogenetic analysis are assumed to be independent. RAPD products however, might not be independent especially when the amplified fragments are associated with repetitive sequences in the genome, when allelic relationships are unknown or if there is heteroduplex formation (Ayliffe *et al.*, 1994; Ellsworth *et al.*, 1993; Novy and Vorsa, 1996). In addition, band amplification might be influenced by competition between primer binding sites making the assumption of independence untenable (Williams *et al.*, 1993).

On the technical aspect, reproducibility of RAPD bands is a well acknowledged problem. Reproducible results within a laboratory can be achieved using a strict standard protocol (eg. Williams *et al.*, 1993) with emphasis placed upon factors that can influence RAPD reproducibility which include the primer concentration, the source of *Taq* DNA polymerase, magnesium concentration, template concentration, the type of thermalcycler and temperature profile. However, problems may arise when transferring markers between laboratories even if a strict regime is observed.

Primers used in RAPD analysis usually have random nucleotide contents. However, the majority of RAPD primers have a relatively high GC contents, i.e. 60-70% (Bowditch *et al.*, 1993). In plant genomes, GC content is not evenly distributed (Li and Graur, 1991), thus it is feasible that RAPD procedure may indeed screen GC-rich regions. High GC content of RAPD primers is necessary for successful annealing at lower temperature to allow imperfect priming (Welsh and McClelland, 1994), but the imperfect priming might result in different primers that bind at the same priming sites. Thus, products obtained from different primers

might therefore be either identical or interdependent.

Another factor that might influence a RAPD profile is competition among priming sites. Competition might occur between sites within a genome, between genomes or between genotypes (Heun and Helentjaris, 1993). Competition among priming sites has been shown to be a major source of error in genotyping *Brassica* which leads to an underestimate of genetic relatedness (Hallden *et al.*, 1996)

Because of the problems associated with the RAPD technique, it is unwise to use RAPDs strictly for phylogenetic analysis. Extreme care must be taken in order to minimise errors that might influence the final result.

5.1.5 Aims and Objectives

The main objective of using RAPD analysis in the present study was to examine genetic relationships between selected *Cryptocoryne* species especially those species whose relationships were not well resolved in the phylogenies constructed from rDNA ITS and chloroplast DNA sequence variation. The utility of RAPD markers in resolving relationships among closely related *Cryptocoryne* species will also be assessed.

5.2 Materials and Methods

5.2.1 Plant material

Most of the plant material used in the survey of RAPD variation was the same as that used in the examination of ITS sequence variation (chapter 3). However, for groups of species whose relationships were well resolved in both the ITS and cpDNA trees (chapter 3 and 4), only one species was represented in the RAPD analysis. For example, the relationship between *C. moehlmannii* and *C. pontederiifolia* was well resolved in the phylogenies based on cpDNA and ITS variation, and consequently only one of these two species, *C. moehlmannii* was included in the RAPD analysis. For the same reason, other species excluded from the RAPD analysis were *C. schulzei* (only *C. elliptica* is included), *C. retrospiralis*, *C. albida* and *C. crispatula* var. *balansae* (only *C. spiralis* is included as a representative of the mainland Asia clade). Further details of the plant material used in the RAPD analysis are given in the results section of this chapter.

5.2.2 Experimental details

DNA template for RAPD analysis was extracted using the method described in chapter 2. Reagents for each RAPD reaction as well as the details of the RAPD reaction conditions, are described in section 2.9 of chapter 2. The six random primers used in the survey - OPA2, OPA11, OPB7, OPH4, OPH9 and OPH15 - were obtained from Operon Technologies in Kit A, B and H respectively. They were chosen for use on the basis of the regularity with which they amplified their target sequence.

5.2.3 Data analysis

DNA fragments were assumed to represent individual alleles at particular loci and were scored as present/absent in a binary matrix. From these data, genetic similarities between each pair of taxa were calculated using the asymetric similarity measure Jaccard's coefficient (Jaccard, 1908) as implemented in the software program RAPDistance (Armstrong*et al.*, 1995) using the equation;

$$F = \frac{Mxy}{(M - Nxy)} ;$$

where Mxy represents the number of amplified fragments shared by taxa x and y, M represents the total number of scorable bands (character), and Nxy represents the number of absent bands shared by the two taxa (Jaccard, 1908). The Jaccard's coefficient which measures the proportion of shared products presence, is one of the most widely used similarity measures for RAPD data. Jaccard's coefficient is used for similarity measures as its statistical properties are well understood (Real and Vargas, 1996).

The measure of genetic distance was obtained as 1 - F. A phenetic analysis was carried out on the genetic distance matrix using both Neighbour Joining (NJ) and Unweighted Pair Group Mean Analysis (UPGMA) methods, as implemented in the software package PHYLIP (Felsenstein, 1993).

5.3 Results

The six random primers used in the survey generated a total of 75 polymorphic DNA fragments. These DNA bands were identified on the basis of being bright and reproducible. Details of the presence/absence of RAPD bands are presented in Table 5.1.

Measures of similarity between each pair of species, based on the proportion of shared bands calculated using Jaccard's index (Jaccard, 1908) ranged from 0.25 to 1.0 (Table 5.2).

Phenograms for the 22 *Cryptocoryne* species included and a single species of *Lagenandra* based on genetic distance, were constructed using the NJ and UPGMA procedures, and are illustrated in Figure 5.1 and Figure 5.2 respectively.

5.3.1 NJ phenogram

The 22 species of *Cryptocoryne* in the NJ tree (Fig. 5.1) can be devided into four groups (A, B, C and D). Group A was comprised of two subgroups, with the first subgroup comprised of *C. walkeri*, *C. x willisii* and *C. parva* and the second subgroup consisting of *C. beckettii*, *C. undulata* and *C. wendtii*. All six species in group A are native to Sri Lanka. *C. cognata* (India) was found to be in close association with species in group A. Group B was also comprised of two subgroups, one of which contained *C. minima* (from the Malay Peninsula), *C. moehlmannii* (Sumatra), *C. griffithii* (Malay Peninsula) and *C. alba* (Sri Lanka), while the other subgroup comprised *C. zonata* (Borneo), *C. purpurea* and *C. cordata* (both species from the Malay Peninsula). The third group, C, was again

Table 5.1. Binary data matrix showing the presence (1) and absence (0) of 75 polymorphic RAPD fragments. The primers used to generate these characters were: OPA2: character 1-11; OPA11: 12-26; OPB15: 27-38; OPH4: 39-47; OPH9: 48-58; OPH15: 59-75. Note: * denotes RAPD fragments uniquely shared between *C. x willisii* and *C. walkeri* while ** denotes RAPD fragment uniquely shared between *C. x willisii* and *C. parva*.

Taxa\Character	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15 *	16 *	17	18	19	20	21	22	23	24	25	26	27
1. *C. spiralis* NJ3129	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0
2. *C. beckettii* 1671	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0
3. *C. walkeri* NJ23-3	0	0	0	0	0	0	1	1	0	1	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
4. *C. wendtii* P1961/5342	0	0	1	0	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5. *C. undulata* NJ22/7	0	0	1	0	0	0	1	1	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0
6. *C. parva* NJ3405	0	0	0	0	0	0	1	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
7. *C. x willisii* P1978/5045	0	0	0	0	0	0	1	1	0	1	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
8. *C. cognata* 28Yadav	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
9. *C. moehlmannii* P1989/5046	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10. *C. minima* S1995/9201	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11. *C. griffithii* NJ85-30	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12. *C. purpurea* Othman s.n.	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0
13. *C. cordata* USM9139	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
14. *C. zonata* W534	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
15. *C. elliptica* USM8069	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
16. *C. affinis* USM8065	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0
17. *C. ferruginea* NJ78-44	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
18. *C. aponogetifolia* P3401	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0
19. *C. annamica* M92/3205	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	0	0	0	0	1
20. *C. pygmaea* NJ3962	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
21. *C. alba* NJ3172-6	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22. *C. ciliata* P1958/6013	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23. *Lagenandra meeboldi*P1979/5019	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0

Table 5.1 continue

Taxa/character	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58
1.	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
2.	0	1	0	0	1	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	1	1	0	0	0	1	0	0
3.	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	1	0	1	0	1	1	0	0	0	0	0	0
4.	0	0	1	0	0	0	0	1	0	0	0	0	1	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0
5.	0	1	0	0	1	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0
6.	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
7.	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0
8.	0	0	1	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
9.	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
10.	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
11.	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
12.	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13.	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0
14.	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15.	1	0	0	0	0	1	0	0	1	0	1	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	1	1	0	1	0
16.	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	1	0	1	0
17.	0	0	0	1	0	0	1	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0
18.	1	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	1	0	0	1
19.	0	0	0	1	0	0	0	0	0	1	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0
20.	0	0	0	1	0	0	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
21.	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
22.	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
23.	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0

Table 5.1 continue

Taxa/character	59	* 60	61	62	63	64	65	66	67	68	69	** 70	71	72	73	74	75
1.	0	0	0	0	1	0	0	0	0	0	0	0	0	1	1	0	1
2.	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
3.	0	1	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0
4.	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0
5.	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
6.	0	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0	0
7.	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1
8.	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
9.	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11.	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0
12.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13.	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
14.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15.	0	0	1	0	0	0	1	0	0	1	0	0	0	0	1	0	0
16.	0	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0
17.	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0
18.	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
19.	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0
20.	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
21.	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	1	0
22.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23.	0	0	1	1	0	0	0	0	1	0	0	0	0	1	0	0	0

Table 5.2. Pairwise genetic distance for 22 *Cryptocoryne* taxa listed in Table 5.1.

1	0.85	0.89	0.88	0.89	0.91	0.89	0.89	1.00	0.94	1.00	1.00	1.00	1.00	0.93	0.96	0.96	0.92	0.96	0.90	0.95	0.87	0.86
2	0.65	0.56	0.30	0.79	0.69	0.86	1.00	0.95	0.66	1.00	0.86	0.95	0.97	0.97	1.00	1.00	0.97	0.96	0.92	1.00	1.00	
3	0.69	0.65	0.57	0.25	0.96	0.95	0.91	0.86	0.95	0.86	0.90	0.94	0.90	0.96	0.98	0.97	0.96	0.83	1.00	0.96		
4	0.43	0.72	0.68	0.91	0.95	0.90	0.90	0.95	0.85	0.89	0.94	0.93	0.96	1.00	0.97	0.96	0.82	0.95	0.96			
5	0.74	0.69	0.86	0.95	0.90	0.91	0.95	0.81	0.90	0.94	0.93	0.96	1.00	0.97	0.96	0.88	1.00	0.96				
6	0.47	0.82	0.94	0.80	0.88	0.94	0.81	0.86	0.93	0.76	0.90	1.00	0.96	0.95	0.78	1.00	0.90					
7	0.91	0.95	0.90	0.91	0.95	0.86	0.89	0.94	0.85	0.96	1.00	0.97	0.96	0.87	1.00	0.96						
8	0.92	0.73	0.93	0.92	0.85	0.91	0.96	0.90	0.88	0.95	1.00	0.94	0.80	1.00	0.94							
9	0.62	0.67	0.75	0.80	0.71	0.85	0.88	0.86	0.94	1.00	1.00	0.75	1.00	0.77								
10	0.70	0.78	0.82	0.75	0.86	0.82	0.78	0.89	1.00	0.93	0.67	1.00	0.87									
11	0.80	0.73	0.62	0.75	0.89	0.80	0.89	0.90	0.94	0.58	1.00	0.87										
12	0.80	0.50	0.85	0.88	0.86	0.94	0.89	1.00	0.85	0.87	0.86											
13	0.62	0.81	0.76	0.87	0.89	0.90	1.00	0.78	1.00	0.94												
14	0.90	0.87	0.85	0.94	0.95	1.00	0.73	1.00	0.93													
15	0.67	0.79	0.77	0.83	0.96	0.88	0.95	0.79														
16	0.75	0.83	0.96	1.00	0.80	1.00	0.86															
17	0.91	0.77	0.76	0.83	0.93	0.90																
18	0.89	0.86	0.96	0.87	0.91																	
19	0.70	0.92	0.89	0.96																		
20	0.89	0.92	1.00																			
21	1.00	0.76																				
22	0.93																					
23																						

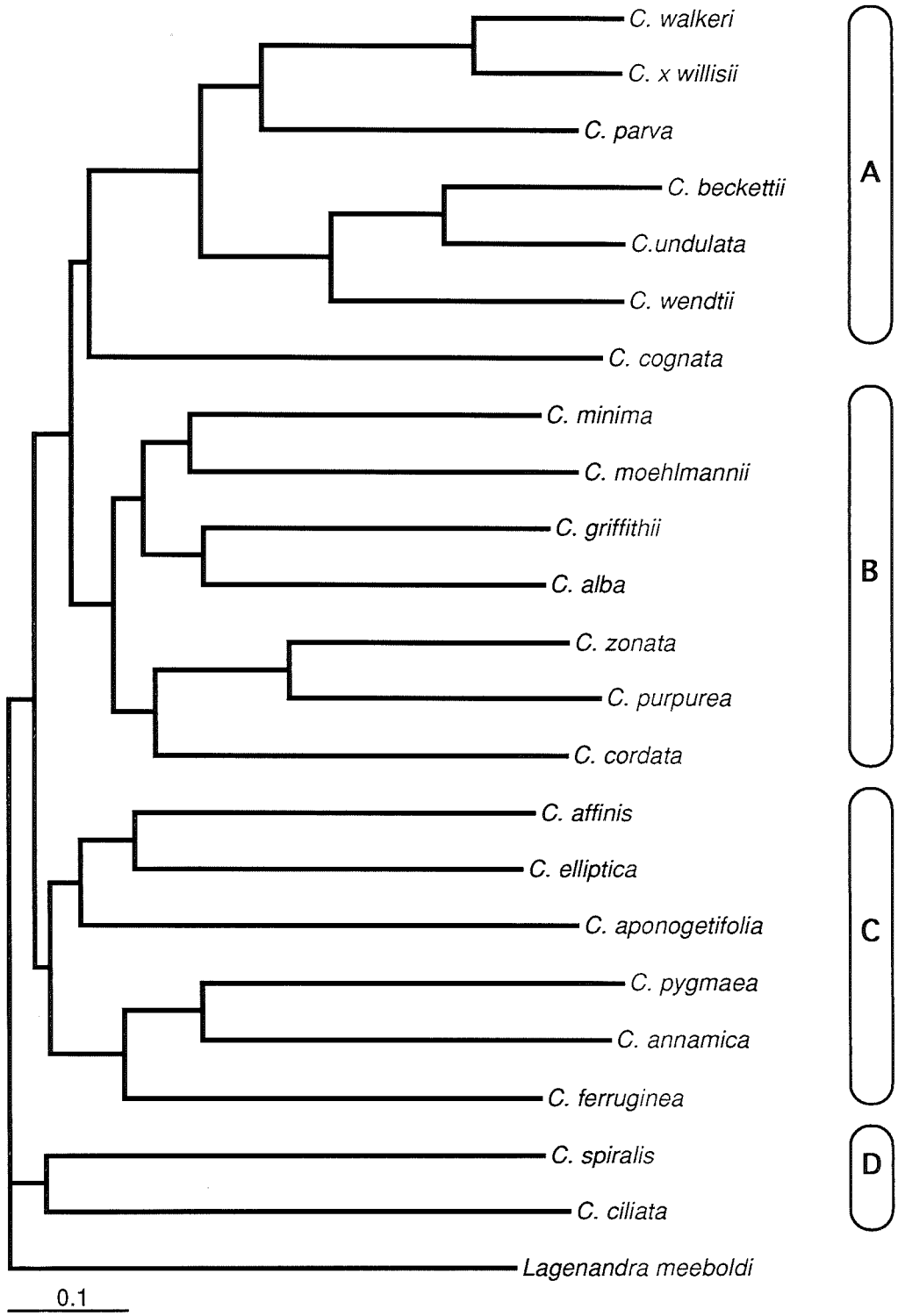


Figure 5.1. A Neighbour-Joining tree for 22 *Cryptocoryne* species and one species of the genus *Lagenandra* constructed from RAPDistance. The genus can be divided into four groups, A, B, C and D.

made up of two subgroups. The first subgroup contained *C. affinis, C elliptica* (both from the Malay Peninsula) and *C. aponogetifolia* (Philippine), while the second subgroup included *C. pygmaea* (Philippine), *C. annamica* (Vietnam) and *C. ferruginea* (Borneo). The final group, D, was made up of two species, *C. spiralis* (India) and *C. ciliata* (a widely distributed species).

5.3.2 UPGMA analysis

The 22 *Cryptocoryne* species in the UPGMA phenogram (Fig. 5.2) can be devided into five groups (A, B, C D and E). Group A of the UPGMA tree, was similar in composition to group A in the NJ phenogram, being comprised of two subgroups with one subgroup made up of *C. walkeri, C. x willisii* and *C. parva* and the other subgroup consisting of *C. beckettii, C. undulata* and *C. wendtii.* Group B consisted of two species, *C. spiralis* and *C. ciliata* which also formed a discrete group (D) in the NJ tree. Group C consisted of *C. pygmaea, C. annamica* and *C. ferruginea* which were also found to be closely associated in the NJ tree, while the fourth group, D, was comprised of the same species as occurred in group B in the NJ tree. Finally, the fifth group E, was made up of *C. affinis, C. elliptica* and *C. aponogetifolia* which were associated into a subgroup of group C in the NJ tree. A most notable difference between the UPGMA and NJ trees was that in the UPGMA tree, *C. cognata* was no longer placed in close association with members of group A (both trees) but instead was placed in close affinity to the group containing *C. minima, C. moehlmannii, C. griffithii, C. alba, C. zonata, C. purpurea* and *C. cordata.*

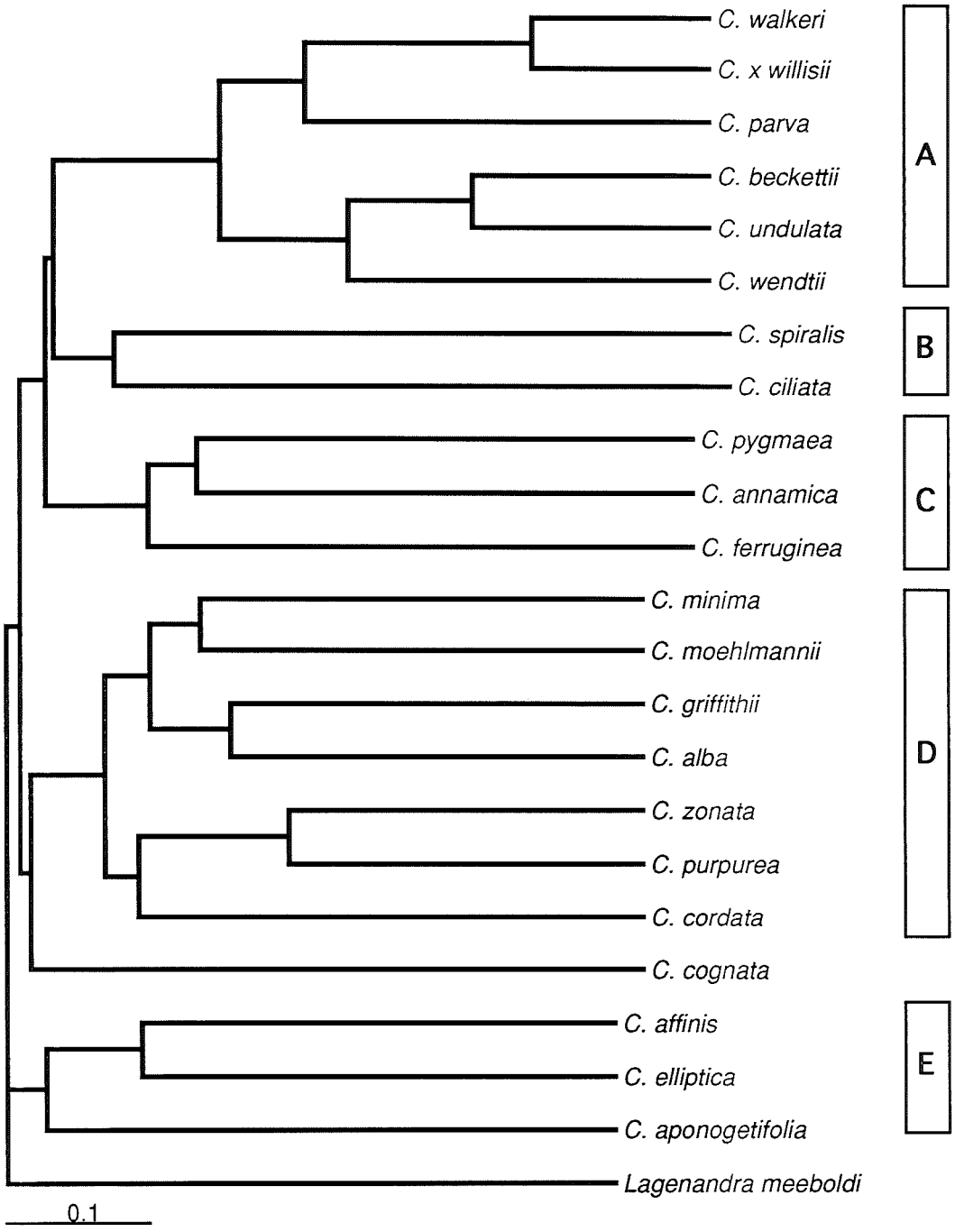


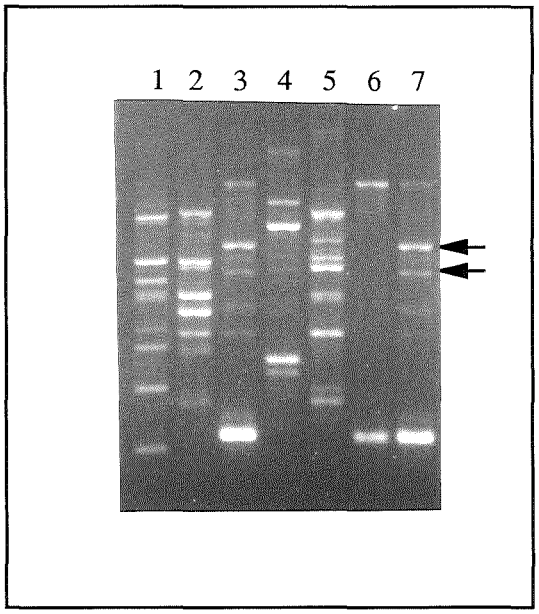
Figure 5.2. UPGMA phenogram generated from RAPDistance for species of *Cryptocoryne* and one species of the genus *Lagenandra.* The genus *Cryptocoryne* can be devided into five groups, A, B, C, D and E.

5.3.3 RAPD profile among closely related *Cryptocoryne* species of Sri Lanka

Inspection of the data for the closely related species from Sri Lanka (*C. beckettii, C. walkeri, C. wendtii, C. undulata, C. parva* and *C. x willisii* - Table 5.1) revealed that this group was polymorphic for 30 bands. Within this group, *C. beckettii* and *C. x willisii* were each characterised by unique bands, 2 and 75 respectively while two unique bands were present in *C. walkeri* (63 and 66) and *C. parva* (42 and 64), and three were present in *C. wendtii* (30, 44 and 74). Further inspection showed that *C. x willisii* (a species of hybrid origin) shared three bands uniquely with *C. walkeri* (e.g. Plate 5.1) and one band uniquely with *C. parva*. It did not, however, share any bands uniquely with its three other closely related species. These results support the hypothesis of a hybrid origin of *C. x willisii* with *C. walkeri* and *C. parva* as its putative parents.

A comparison of all RAPD bands produced by *C. x willisii, C. parva* and *C. walkeri,* showed 14 polymorphic markers to be present within the RAPD data matrix. Of these, seven were shared between *C. x willisii* and *C. walkeri* while only one was shared between *C. x willisii* and *C. parva*. Three bands were unique to *C. walkeri,* two unique to *C. parva,* and one unique to *C. x willisii.*

Plate 5.1. RAPD profile of *C. spiralis* and *Cryptocoryne* species from Sri Lanka obtained using primer OPA11. Arrows show the RAPD bands shared uniquely between *C. walkeri* and *C. x willisii* (character 15 and 16 in Table 5.1).



from left, 1. *C. spiralis* NJ3129a2, 2. *C. beckettii* 1671, 3. *C. walkerii* NJ23-3, 4. *C. wendtii* P1961/5342, 5. *C. undulata* NJ22-7, 6. *C. parva* NJ3405, 7. *C. x willisii P1978/5045*.

5.4 Discussion

The phenograms generated for 22 species of *Cryptocoryne* species based on NJ and UPGMA analysis are the same in topology with two exceptions. The first concerns the placement of *C. cognata* (an Indian species) within the two trees. In the NJ tree, *C. cognata* was placed in close association with group A, i.e. with *C. walkeri* and its closely related species, while within the UPGMA tree, *C. cognata* was placed in close association with *C. cordata* and its close relatives. The placement of *C. cognata* in different places within the two trees indicated that it is unresolved, probably due to low number of RAPD bands shared between *C. cognata* and the two species groups. Another difference between the two phenograms involves the position of the subgroup comprising *C. pygmaea, C. annamica* and *C. ferruginea*. This was placed together with another subgroup - *C. affinis, C. elliptica* and *C. aponogetifolia* - in the NJ phenogram, while in the UPGMA tree, it formed a separate group, C. In the UPGMA tree, the final group, E, comprised of *C. affinis, C. elliptica* and *C. aponogetifolia*. In each of these cases, the branch that incorporated a species or species subgroup within a group (i.e. *C. cognata* together with group A in the NJ tree or the inclusion of *C. pygmaea, C. annamica, C. ferruginea, C. affinis, C. elliptica* and *C. aponogetifolia* into one group in the NJ tree) was short compared to the branch length that separates groups or subgroups. This suggests that the affinity between groups or subgroups must be treated with caution.

An interesting feature of the RAPD analysis concerns the group of Sri Lankan species comprised of *C. walkeri, C. parva, C. x willisii, C. wendtii, C. beckettii* and *C. undulata*. It has been proposed that *C. x willisii* is of

hybrid origin with *C. parva* acting as one parent, and *C. walkeri* or *C. beckettii* as the other parent. Jacobsen (1981) managed to produce *C. x willisii* artificially when crossing *C. parva* with either *C. walkeri* or *C. beckettii*. In addition, *C. beckettii* can hybridise successfully with *C. walkeri* and backcrosses can hybridise with *C. parva* to produce *C. x willisii* like progeny. Morphologically, *C. x willisii* is characterised by having a roughened purple limb of the spathe with a purple collar (Plate 5.2). Sometimes individual hybrids possess a greenish limb usually with a purple collar, or a purple limb with a yellowish collar. The distinctly roughened, purple limb of the spathe is also the limb characteristic of *C. parva*. In addition, *C. parva* has a distinct purple to dark purple collar. In this study *C. x willisii* was placed together with *C. parva* and *C. walkeri* into a subgroup, and showed a particular close relationship to *C. walkeri*. This seems to suggest that the accession of *C. x willisii* included in this study, is a hybrid derivative of *C. parva* and *C. walkeri*. Almost complete additivity of the RAPD markers of these two putative parents was observed in *C. x willisii*, apart, that is, from one unique band present in the hybrid. It would seem, therefore, that since the initial hybridisation event took place, there has been insufficient time for the hybrid to accumulate new mutations at RAPD loci. However, it should be noted that this results is preliminary as the sample size used is extremely limited and test for fragments homology was not carried out. A similar situation where RAPD has been succesfully used to test the hyphotesis of plant of hybrid origin was reported, for example, in the polyploid *Saxifraga osloensis* (Brochmann *et al.*, 1996). Although RAPD markers represent anonymous genomic sites, they are distributed throughout the genome and appear to be biparentally inherited (Yu and Pauls, 1993; Huff and Bara, 1993; Brochmann *et al.*, 1996). The present study has indicated that RAPD

Plate 5.2. The spathe of *C. x willisii* with purple limb.

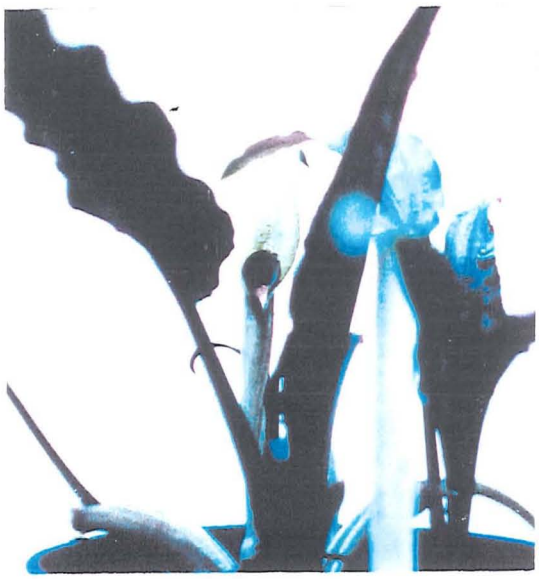


Plate 5.3. *C. beckettii* - a spathe with olive yellow limb and a distinct dark brown collar margin

Plate 5.4. The spathe of *C. walkeri*. Notice the smooth collar zone as compared to distinct collar margin present in *C. beckettii*. Both limb and collar are bright green to yellow.

analysis may be of value in identifying the progenitors of a hybrid species.

Previous morphological studies have concluded that the Sri Lankan species, *C. beckettii, C. walkeri, C. wendtii* and *C. undulata* are very closely related (Jacobsen, 1977, Arends *et al.*, 1982, Nicolson, 1988). These species have a rather uniform structure of the spathe but differ slightly in its colour. *Cryptocoryne beckettii,* for example, has a spathe with an olive yellow limb and a collar that is reddish purple or brown (Plate 5.3). In contrast, *C. walkeri* has a bright green to yellow limb, and a collar zone (not a distinct collar margin as present in *C. beckettii*) of the same colour as the limb (Plate 5.4). *Cryptocoryne wendtii* is characterised by having a yellow to purplish brown limb and brown to black purple collar. Moreover, it differs from its close relatives by having bullate to elliptic, and sometimes narrow ovate leaves, and not ovate to narrow ovate leaves. Finally, *C. undulata* has a cream yellowish to brownish limb with a collar of the same colour.

Phylogenetic trees constructed from cpDNA sequence (chapter 4) and nuclear DNA sequence (chapter 3) variation, did not resolve clearly the relationships between members of this closely related group of Sri Lankan species. In contrast, the NJ and UPGMA phenograms generated from RAPD data placed *C. walkeri, C. x willisii* and *C. parva* into a subgroup separate from *C. beckettii, C. undulata* and *C. wendtii* which were placed into another subgroup. In the latter subgroup, *C. beckettii* appeared to be more similar to *C. undulata* than to *C. wendtii.* Thus, the RAPD data appear to have clarified relationships within this closely related group of *Cryptocoryne* species which remained unresolved based on cpDNA and

nuclear DNA sequence variation. This indicates that RAPD data are particularly useful in providing additional information on the affinity between closely related species.

Although of value in this respect, some of the affinities suggested by the RAPD trees are difficult to explain. For example, the RAPD tree indicated that *C. ciliata* and *C. spiralis* (each having a basal chromosome number of x=11) were closely related, whereas they were placed into separate clades in the cpDNA and nuclear DNA trees. Moreover, their different morphologies would not suggest a close relationship to exist between these two taxa. One explanation, however, is perhaps as with *C. cognata*, the number of RAPD bands observed in *C. ciliata* was rather low and the majority of the bands appear to be shared with *C. spiralis*. Rather than showing a clear association between *C. ciliata* and *C. spiralis*, result from RAPD might rather indicate that there is not enough resolution on the affinities between species, thus some of the species grouping might not be significant.

Another inconsistency concerns the postulated hybrid origin of *C. purpurea*. Based on the low pollen fertility (Jacobsen, 1977) and morphological characters, it has been suggested that *C. purpurea* is a hybrid derivative of *C. cordata* and *C. griffithii* (Jacobsen and Mansor, unpublished). If this is correct, it would be expected that *C. purpurea*, *C. cordata* and *C. griffithii* should be placed into the same group or subgroup within the RAPD phenograms. However, in both the NJ and UPGMA trees, *C. purpurea* was associated with *C. zonata* in the same subgroup, while *C. griffithii* was placed in another subgroup. Of equal surprise was the finding that *C. griffithii* appeared to be more closely related to *C. alba*, a

species from Sri Lanka, than to other Malay Peninsula species. Further incongruencies between the RAPD trees and those generated from cpDNA and ITS sequence variation are evident with respect to species in group C and E in the UPGMA tree (which comprise group C in the NJ tree). Group E in the UPGMA tree contains two Malay Peninsula *Cryptocoryne* species, *C. affinis* and *C. elliptica* and also *C. aponogetifolia* which is native to the Philippines. In the cpDNA and ITS trees, however, *C. affinis* was more closely related to *C. minima,* while *C. elliptica* was closely associated with *C. pygmaea* (a Philippine species). In general, the RAPD trees did not show the geographical pattern of separation of *Cryptocoryne* species that was evident in both cpDNA and ITS trees.

The incongruencies between the RAPD, and the cpDNA and ITS trees mentioned above, raises concern on the usefulness of RAPD generated patterns for inferring relationships, especially between species which are distantly related. Several problems have been recognised in using RAPD data in this type of study. One of the more important problems is the extent to which it is possible to infer homology between bands showing the same migration distance in a gel. RAPD bands which migrate to the same distance in a gel are assumed to have originated from amplification of the same DNA sequence. However, it is not clear how often this assumption holds. Several authors (Gillies, 1994; Stemmers *et al.*, 1995; Rieseberg, 1996) have tested the homology of co-migrating bands using one of two methods: (i)restriction digestion of the bands; and (ii)hybridisation of co-migrating bands using the Southern blotting procedure. Rieseberg (1996) found that 91% of 220 co-migrating fragments of three wild sunflower species displayed some level of homology (i.e. congruent in restriction digest profile) while Gillies (1994) using the same

test found only 58% of co-migrating bands to be homologous in a more limited study of species of the genus *Stylosanthes*. Stodhard (1997) however, points out that the use of Southern hybridisation and restriction enzyme might not prove homology after all since the products that share a sequence identity may be paralogous (homologous due to duplication event) rather than orthologous (homologous due to speciation event). In the end, each co-migrating RAPD products has to be sequenced in order to distinguish orthology from paralogy. The possibility of a relatively high level of non-homology of co-migrating bands would contribute to the construction of inaccurate RAPD trees if homology was not tested. Unfortunately in the present study the homology of co-migrating RAPD bands was not tested due to a time constraint.

Another practical problem encountered in this work is the use of a very diverse plant material, i.e. species from both closely and distantly related. This causes difficulties which might eventually led to error in scoring, especially to the extent to which it is possible to infer homology. Rieseberg (1996) suggested that RAPD may be useful for investigations within species or between closely related species. In addition, the similarity coefficient used to calculate similarity values are affected by two types of error associated with RAPD methodology (Lamboy, 1994a; 1994b). The two errors are: i) false positive (a product that is present but should be absent), and ii) false negative (a product that is absent but should be present). False positives are scored in instances of mistaken product homology, while false negatives are scored when products are present below the limit of gel resolution (Smith *et al.*, 1994). These two types of error present within RAPD data may influence the resulting similarity matrices dramatically and consequently the clustering procedure (Harris, in press).

Another potential problem in using the presence/absence of RAPD fragments as a basis for examining species relationships arises due to the occurrence of competition among RAPD primer sites (Smith *et al.*, 1994; Williams *et al.*, 1993). Sometimes the absence of a band may simply be due to the fact that there is competition between a number of the same priming sites for a given primer. Consequently, amplification does not occur even though the appropriate sequence for primer annealment is present.

In the final analysis, RAPD data could be of value in testing the hyphothesis of hybrid origin of *C. x willisii.* A comprehensive sampling and testing of fragments similarity will hopefully add proof of its hybrid origin. On the species affinity within the genus, it is thought that the lack of RAPD bands in certain species might have contributed to some of the anomalies accountered in the phenetic trees. In addition, a survey which involves such diverse species within the genus could have introduced errors in the assumption of fragments identity within the RAPD data which in turn might have a significant effects on the final results.

Chapter 6

General discussion

The results presented in this thesis provide an insight into the value of molecular generated data in elucidating species relationships within *Cryptocoryne*. Both the ITS and cpDNA sequence analyses presented in chapter 3 and 4 respectively, resulted in two independent data sets which complemented each other in establishing evolutionary relationships within the genus. A third data set obtained from an analysis of RAPD variation in the genus (chapter 5), produced some contrasting results but also provided valuable information in regard to the genomic affinity of closely related species in Sri Lanka, and the parentage of a hybrid species native to Sri Lanka.

6.1 Evolutionary relationships within the genus *Cryptocoryne*

The ITS generated phylogeny (Fig. 3.12) revealed a geographical pattern of species relationships within the genus and subdivided it into three clades. The first clade consisted of species from the Indian subcontinent and Mainland Asia, i.e. *C. spiralis, C. albida, C. retrospiralis* and *C. crispatula*. However, *C. annamica*, a species found in Vietnam, was excluded from this clade. Relationships between species within the clade were highly resolved as reflected by the high bootstrap values supporting relevant branches. A second clade was comprised of species found exclusively in Sri Lanka. Within this clade, two groups were recognised, one consisted of *C. alba*, and the other of *C. beckettii, C. wendtii, C. walkeri, C. undulata and C. x willisii*. This species grouping was congruent with

previous morphological and cytological examination of *Cryptocoryne* species present in Sri Lanka (Jacobsen, 1977; Arends *et al.*, 1982). However, the relationship between *C. beckettii* and *C. undulata* and the other species within the group was not very well resolved.

The third clade within the ITS tree contained species from the Malesia floristic region, a region of high tropical plant diversity, extending from the Malay Peninsula including Sumatra, Java, Borneo and the Philippines through the archipelago to New Guinea. Although the resolution of species relationships within the Malesia clade was not satisfactory, several conclusions could be drawn. For example, the close association between the two Sumatran species included in the analysis, *C. pontederiifolia* and *C. moehlmannii,* and the close association between two Malay Peninsula species, *C. elliptica* and *C. schulzei,* were evident in that both formed monophyletic groups. One intriguing finding was that a species from the Philippines, *C. pygmaea,* appeared to be closely related to species from the Malay Peninsula namely *C. schulzei* and *C. elliptica* (Fig. 3.12 and 3.14) rather than to another Philippines' species, *C. aponogetifolia.* The close relationship between *C. aponogetifolia* and the Vietnamese species, *C. annamica* based on the ITS phylogeny was previously recognised by Jacobsen (pers. comm.) from morphological comparisons. He linked these two species with *C. usteriana* which was not analysed in the present study.

The phylogenetic trees generated from cpDNA sequence variation were not as well resolved as those constructed from ITS sequence variation. The lower level of resolution was due to the presence of fewer synapomorphic characters despite scoring indels as additional characters.

Nonetheless, the cpDNA phylogeny, to an extent, complemented the ITS phylogeny especially in separating the genus into three clades, which reflected the geographical distribution of species in the genus. The position of several species within these clades was, however, rather intriguing. For example, the placement of *C. cognata* with a chromosome base number of x=14 within the mainland Asia clade, rather than with other *Cryptocoryne* species with x=14 occupying the Sri Lanka clade, was rather unexpected. Although placement of this species fits nicely with the geographical distribution of other member species of its clade, it would suggest that *Cryptocoryne* species with x=14 may have arisen on two independent occasions during the evolution of the genus.

Within the cpDNA phylogenetic tree, *C. alba, C. parva* and *C. annamica* were placed in the same subgroup. This again was somewhat surprising as whereas the first two species occur in Sri Lanka, *C. annamica* is found only in Vietnam and does not bear any morphological resemblance to species of *Cryptocoryne* from Sri Lanka. In view of the unexpectedness of this result, it is advised that additional material of *C. annamica* is examined in the future to confirm this finding. The close association of *C. parva* with *C. alba* was also incongruent with the results of morphological comparisons which indicate that *C. parva* with 2n=28 (x=14), is more closely related to other species with x=14 which comprise the subclade of Sri Lankan species that includes *C. beckettii, C. walkeri, C wendtii, C. undulata* and *C. x willisii* (Jacobsen, 1977; Arends *et al.*, 1982). *C. alba* and *C. parva* also occupy very different habitats (de Graaf and Arends, 1986), and *C. alba* with 2n=36 has never been recorded to hybridise with species with 2n=28.

Of the two types of sequence variation used in the current study to reconstruct a phylogeny for *Cryptocoryne*, ITS sequence variation provided a better resolution of species relationships than did cpDNA variation. However, although the ITS phylogeny separated species according to their geographical distribution, the pattern of distribution did not exactly match the paleogeographical origin of geographical areas as described by Dunn and Dunn (1977). For example, the mainland Asia sector as described by the two authors was comprised of continental Asia and Sri Lanka. The ITS phylogeny, on the other hand partitioned *Cryptocoryne* species from continental Asia (India Burma and Thailand) and species from Sri Lanka into two distinct clades.

In regard to the postulated aneuploid series present within the genus (Reumer, 1984), the ITS phylogeny supports the idea that the ancestral chromosome number of *Cryptocoryne* is x=18, and that species in Sri Lanka with x=14 originated from the x=18 stock. However, the pathway by which other chromosome base numbers arose is less certain. Both the ITS and cpDNA phylogenies indicated that some chromosome base numbers, for example x=11 had multiple origins, i.e. arose twice independently in the course of the evolution of the genus. However, the origin of x=15 and x=17 is not made clear by either phylogeny.

In a further analysis, both types of sequence variation were combined and subjected to phylogenetic analysis using PAUP to investigate if the resolution of species relationships might be improved. A total of 8012 most parsimonious trees was obtained and are summarised using both the

strict consensus tree (Fig. 6.1) and 50% majority rule consensus tree (Fig. 6.2). The strict consensus tree was not well resolved and much information was lost with only two clades recognised; one clade was comprised of species from mainland Asia while the other was comprised of species from Sri Lanka and Malesia. In contrast, the 50% majority rule consensus tree showed a much better resolution with the three clades resolved in the ITS and cpDNA phylogenies clearly present. One feature of particular interest within the majority rule consensus tree concerned the placement of *C. annamica* (from Vietnam) as sister taxon to two Philippine species, *C. aponogetifolia* and *C. usteriana* within the Malesia clade. In the cpDNA tree, *C. annamica* was placed together with *C. alba* and *C. parva* while *C. aponogetifolia* and *C. usteriana* formed a monophyletic group. In the ITS tree, *C. usteriana* was not included and *C. aponogetifolia* and *C. annamica* formed a monophyletic group. The placement of *C. annamica* as sister taxon to *C. aponogetifolia* and *C. usteriana* in the phylogeny generated from the combined data set is in agreement with previous views that these three species are closely related based on morphology (Jacobsen, pers. comm.).

Judged overall, the species groupings within both the ITS and cpDNA phylogenies was to some extent congruent with the taxonomic groupings identified in *Cryptocoryne* by Jacobsen (1977) and Arends *et al.* (1982). However, whereas the classifications by Jacobsen (1977) and Arends *et al.* (1982) divided the genus into a large number of different species groups, the molecular phylogenies show that three major and clear lineages are present in the genus.

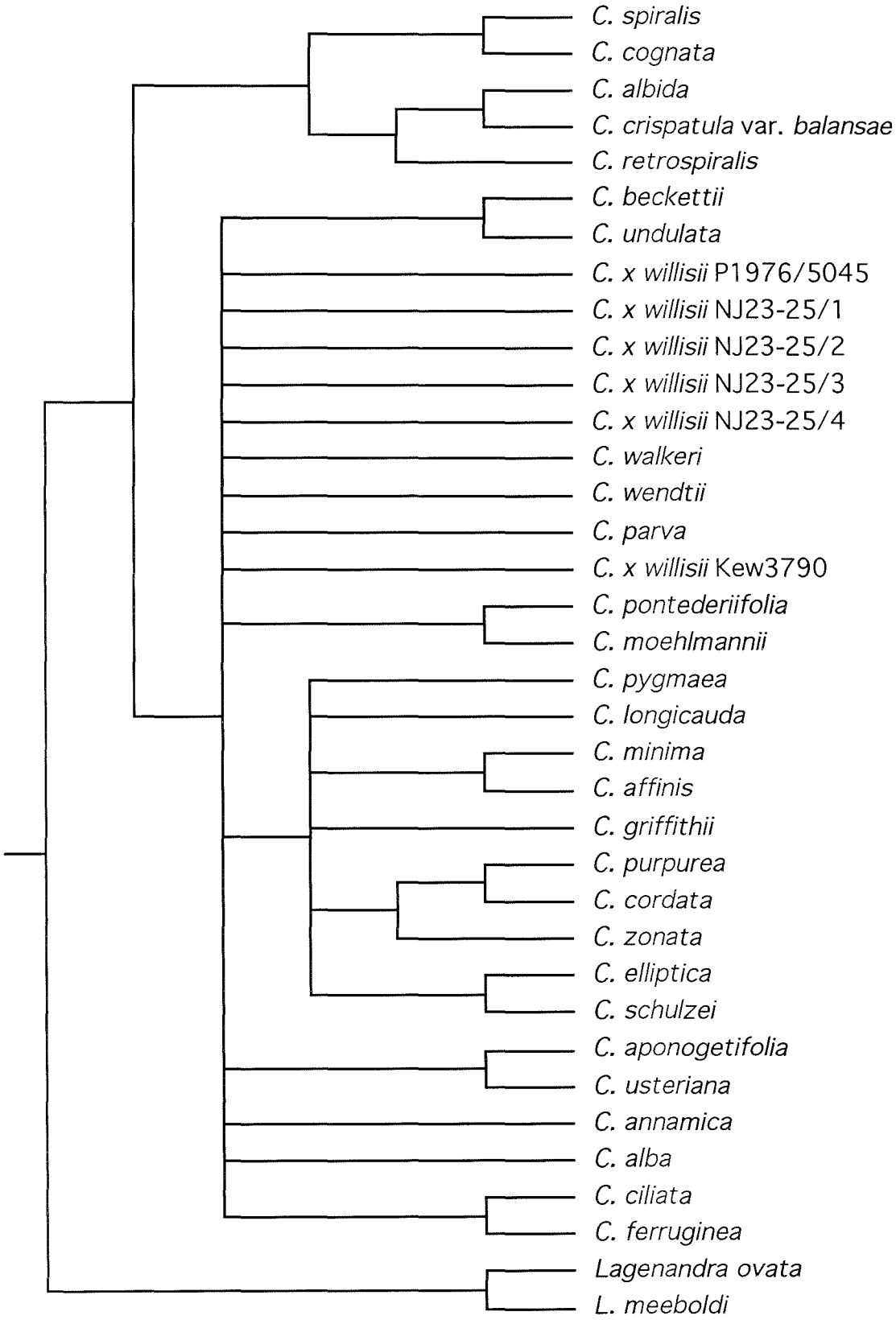


Figure 6.1. A strict consensus tree of the 8012 most parsimonious trees obtained from heuristic search as implemented in PAUP from combined ITS nrDNA and cpDNA sequence variation reported in chapter 3 and 4 for *Cryptocoryne* species and two species of *Lagenandra*.

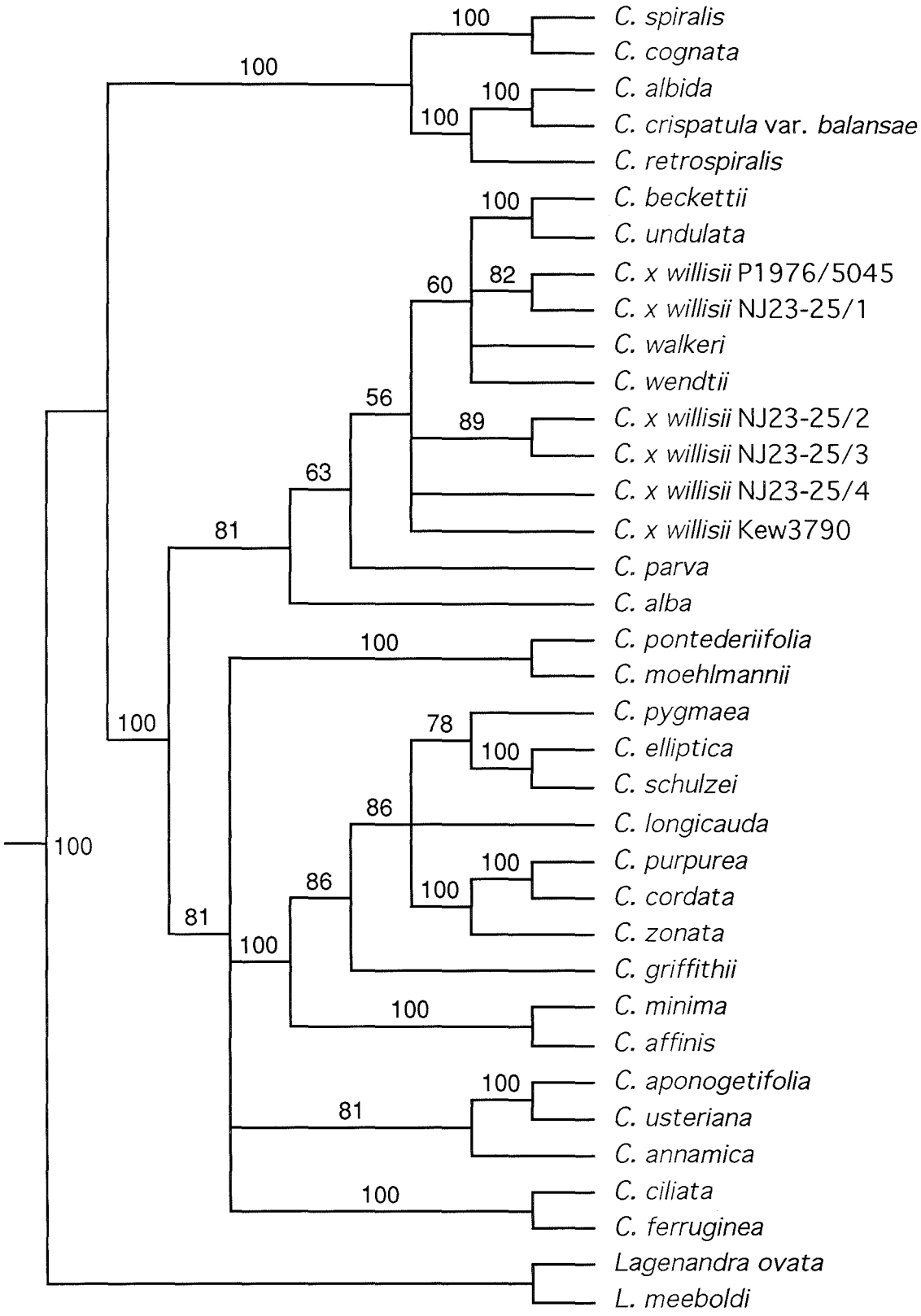


Figure 6.2. A 50% majority rule consensus tree of 8012 most parsimonious trees generated from combined ITS rDNA and cpDNA sequence variation obtained from chapter 3 and 4 respectively for *Cryptocoryne* species and two species of *Lagenandra*. Numbers above each branch represents the percentage in which a particular branch occurs in the 8012 most parsimonious trees.

6.2 Parentage of hybrid species

A subsidiary aim of the research conducted was to assess the utility of molecular markers in the determination of species of hybrid origin and the discovery of the parents of such species. *C. x willisii* is postulated to be a naturally occurring hybrid present in Sri Lanka based on morphological comparisons and crossing experiments. Jacobsen (1981) artificially produced *C. x willisii* by crossing *C. parva* with either *C. beckettii* or *C. walkeri*. In addition, *C. beckettii* was shown to hybridise successfully with *C. walkeri* and backcrosses hybridised with *C. parva* to produce C. x willisii-like progeny.

In the current study, cloned ITS sequences of a number of accessions of *C. x willisii* fell into one or other of two subclades of the ITS generated phylogeny. One of these subclades contained the ITS sequences of two accessions of *C. x willisii* (accession P1976/5045, and one of four cloned ITS sequences from accession NJ23-25) in addition to *C. walkeri* while the other subclade was comprised of the ITS sequences of the other accessions of *C. x willisii* investigated (i.e. accession Kew3790, and three cloned ITS sequence from *C. x willisii* accession NJ23-25). These results suggest that *C. walkeri*, and not *C. beckettii* acted as one of the parents of *C. x willisii*. Unfortunately, the ITS sequence of *C. parva* was not available for comparison due to difficulties with its amplification. It is hoped that the sequence of this species will be made available in the near future.

In contrast to ITS sequence information, cpDNA sequence variation did not provide good evidence for *C. walkeri* (or *C. parva*) being the parents of *C. x willisii*. On the contrary, the cpDNA phylogeny placed *C. x*

willisii more closely to *C. beckettii* than to *C. walkeri* although all three species were positioned within the same subclade. Close inspection of the cpDNA sequences of *C. x willisii* and *C. beckettii* revealed that they both share one character (T- position 14 in Table 4.4) which was absent in *C. walkeri*. This might indicate that *C. beckettii* was the maternal parent of *C. x willisii* assuming that cpDNA is maternally inherited in *Cryptocoryne* as is the case in most Angiosperms. It should be noted here that at present there has not been any study conducted concerning the mode of inheritance of cpDNA within Araceae

However, much more compelling evidence for the correct parentage of *C. x willisii* emerged from the results of the RAPD analysis. *C. x willisii* exhibited an additive RAPD profiles sharing three bands uniquely with *C. walkeri* and one band uniquely with *C. parva*. It did not share any other bands uniquely with other *Cryptocoryne* species from Sri Lanka. Furthermore, the RAPD generated phenogram, placed *C. x willisii* together with *C. parva* and *C. walkeri* into a separate subgroup with *C. x willisii* being particularly closely related to *C. walkeri*. Thus from these results it may be concluded that *C. x willisii* is of a hybrid origin between *C. parva* and *C. walkeri* and that *C. beckettii* was not directly involved in its parentage. It should be noted here that the results presented was preliminary and by no mean final as the sampling of materials was very minimal.

C. purpurea is another species within the genus that has been postulated to be a naturally occurring hybrid (Jacobsen, 1977). Inspections of morphological characters and geographical distributions have indicated

that *C. cordata* and *C. griffithii* are its putative parents (Jacobsen and Mansor, unpublished). In the present study, the ITS sequence of *C. purpurea* was found to closely resemble the ITS of *C. cordata* (both species were placed together within a subclade of the ITS phylogeny), but differed substantially from the ITS sequence of *C. griffithii*. In contrast, the cpDNA of *C. purpurea* appeared to be more similar to that of *C. griffithii* than to *C. cordata*. Thus the ITS and cpDNA sequence variation data, when combined, support the postulated hybrid origin of *C. purpurea*, indicating that *C. griffithii* acted as the maternal parent and that the ITS sequence of *C. purpurea* has been homogenised to resemble that of *C. cordata*. However, the results from RAPD analysis were less conclusive on the parentage of *C. purpurea*. RAPD data placed *C. purpurea* and its putative parents within the same species group in the RAPD phenogram; however, *C. purpurea* was placed in a subgroup which included *C. zonata* and *C. cordata*, with *C. purpurea* in close association with *C. zonata*. Further, *C. griffithii* was placed in another subgroup in close association with *C. alba*. It would be of value in the future to examine in more detail material of *C. purpurea* and its close relatives with a range of molecular markers to obtain a better assessment of its postulated hybrid origin.

6.3 Concluding remarks and future research

The results presented in this thesis provide only a partial understanding of the phylogeny of the genus *Cryptocoryne*. Clearly much more work is left to be done if we are to gain a more comprehensive understanding of phylogenetic relationships between all *Cryptocoryne* species. It shall be necessary to examine species which were not available

for inclusion in the present study, and it would also be of value to examine levels of intraspecific variation especially in those species considered to be endangered. The latter is especially important in Malaysia where most *Cryptocoryne* species are found in restricted areas, and are subjected to disturbance. There has been a decline in the number of populations of several *Cryptocoryne* species in recent years (Mansor and Masnadi, 1994) and conservation measures are required if extinction is to be prevented. Molecular techniques are useful and effective tools that may be used to investigate the pattern of variation within and between species in the genus and so provide essential background for implementing a satisfactory conservation strategy for *Cryptocoryne* species.

Bibliography

References

Adams, R.F and T. Demeke 1993. Systematic relationships in *Juniperus* based on random amplified polymorphic DNAs (RAPDs). *Taxon* 42:553-571.

Appels, R. and J. Dvorak 1982. The wheat ribosomal DNA spacer region: Its structure and variation in populations and among species. *Theoretical and Applied Genetics* 63:337-348.

Appels, R. and R.L. Honeycutt 1986. rDNA: Evolution over a billion years. In: *DNA Systematics* (ed. S. Dutta), 81-135. CRC, Boca Raton, Florida.

Arends, J.C., J.D. Bastmeijer and N. Jacobsen 1982. Chromosomes numbers and taxonomy in *Cryptocoryne* (Araceae) II. *Nordic Journal of Botany* 2:453-463.

Armstrong, J., A. Gibbs, R. Peakall and G. Weiller 1995. *RAPDistance Program version 1.03 for the Analysis of Patterns of RAPD Fragments.* Australian National University, Canberra.

Arnheim, N. 1983. Concerted evolution in multigene families. In: *Evolution of Genes and Proteins* (eds. M. Nei and R. Koehn), 38-61. Sinauer Associates, Sunderland, MA.

Arnold, M.L. 1993. *Iris nelsonii* (Iridaceae): origin and genetic composition of a homoploid hybrid species. *American Journal of Botany* 80:577-583.

Ayliffe, M.A., G.J. Lawrence, J.G. Ellis and A.J. Pryor 1994. Heteroduplex molecules formed between allelic sequences cause nonparental RAPD bands. *Nucleic Acids Research* 22:1632-1636.

Badeness, M.L. and D.E. Parfitt 1995. Phylogenetic relationships of cultivated *Prunus* species from an analysis of chloroplast DNA variation. *Theoretical and Applied Genetics* 90:1035-1041.

Baldwin, B.G., D. W. Kyhos and J. Dvorak 1990. Chloroplast DNA evolution and adaptive radiation in the Hawaiian silversword alliance (Asteraceae-Madiinae). *Annals of the Missouri Botanical Garden* 77:96-109.

Baldwin, B.G. 1992. Phylogenetic utility of the internal transcribed spacer of nuclear ribosomal DNA in plants: An example from the Compositae. *Molecular Phylogenetics and Evolution* 1:3-16.

Baldwin, B.G., M.J. Sanderson, J.M. Porter, M.F. Wojciechowski, C.S. Campbell and M.J. Donoghue 1995. The ITS region of nuclear ribosomal DNA - A valuable source of evidence on Angiosperm phylogeny. *Annals of the Missouri Botanical Garden* 82:247-277.

Baum, D.A., K.J. Sytsma and P.C. Hoch 1994. A phylogenetic analysis of *Epilobium* (Onagraceae) based on nuclear ribosomal DNA sequences. *Systematic Botany* 19:363-368.

Bayer, R.J., D.E. Soltis and P.S. Soltis 1996. Phylogenetic inferences in *Antennaria* (Asteraceae: Gnaphalieae: Cassiniinae) based on the sequences from nuclear ribosomal DNA internal transcribed spacers (ITS). *American Journal of Botany* 83:516-527.

Bogler, D.J. and B.B. Simpson 1996. Phylogeny of Agavaceae based on ITS rDNA sequence variation. *American Journal of Botany* 83:1225-1235.

Bowditch, B.M., D.G. Albright, J.G.K. Williams and M.J. Braun 1993. Use of amplified polymorphic DNA markers in comparative genome studies. *Methods in Enzymology* 224:294-309.

Brochmann, C., T. Nilsson and T.M. Gabrielsen 1996. A classic example of postglacial allopolyploid speciation confirmed by nucleotide sequences and RAPD markers: *Saxifraga osloensis*. Acta University Uppsala. *Symb. Ups.* 31:75-89.

Brown, T. A. 1994. *DNA Sequencing: The Basics*. IRL Press at Oxford University Press.

Bruneau, A. 1996. Phylogenetic and biogeographical patterns in *Erythrina* (Leguminosae:Phaseoleae) as inferred from morphological and chloroplast DNA characters. *Systematic Botany* 21: 587-605.

Caetano-Anolles, G. B. Bassam and P.M Gresshoff 1991. DNA amplification fingerprinting using very short arbitrary primers. *Biotechnology* 9:553-557.

Campbell, C.S., B.G. Baldwin, M.J. Donohue and M.F. Wojciechowski 1993. Toward a phylogeny of *Amelanchier* (Rosaceae: Maloideae): Evidence from sequences of the internal transcribed specers (ITS) of nuclear ribosomal DNA (nrDNA). *American Journal of Botany* 80 (supplement, abstract):135-136.

Campbell, C.S., M.J. Donoghue, B.G Baldwin. and M.F. Wojciechowski 1995. Phylogenetic relationships in Maloideae (Rosaceae) - Evidence from sequences of the internal transcribed spacers of nuclear ribosomal DNA and its congruence with morphology. *American Journal of Botany* 82:903-918.

Castiglione, S., G. Wang, G. Damiani, C. Bandi, S. Bisoffi and F. Sala 1993. RAPD fingerprints for identification and for taxonomic studies of elite poplar (*Populus* spp.) clones. *Theoretical and Applied Genetics* 87:54-59.

Chase, M.W. and H.G. Hills 1991. Silica gel: An ideal material for field preservation of leaf samples for DNA studies. *Taxon* 40: 215-220.

Chase, M.W., D.E. Soltis, R.G. Olmstead, D. Morgan, D.H. Les, B.D. Mishler, M.R. Duvalll, R. Price, H.G. Hills, Y.-L. Qui, K.A. Kron, J.H. Rettig, E. Conti, J.D. Palmer, J.R. Manhart, J. Systma, H.J. Michaels, W.J. Kress, K.G. Karol, W.D. Clark, M. Hedren, B.S. Gaut, R.K. Jansen, K.-J. Kim, C.F. Wimpee, J.F. Smith, G.R. Furnier, S.H. Strauss, Q.-Y. Xiang, G.M. Plunkett, P.S. Soltis, S.E. Williams, P.A. Gadek, C.J. Quinn, L.E. Eguiarte, E. Golenberg, G.H. Learn, S.

Graham, S.C.H. Barret, S. Dayanandan and V. Albert 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbc*L. *Annals of the Missouri Botanical Garden* 80:528-580.

Cheliak, W.M. and B.P Dancik, 1982. Genic diversity of natural populations of a clone-forming tree *Populus tremuloides*. *Canadian Journal of Genetics and Cytology* 24:611-616.

Clegg, M.T. and G. Zurawski 1992. Chloroplast DNA and the study of plant phylogeny: Present status and future prospect. In: *Molecular Systematics of Plants* (eds. P.S. Soltis, D.E. Soltis and J.J. Doyle), 1-13. Chapman and Hall, London.

Crawford, D.J. 1990. *Plant Molecular Systematics: Molecular Approaches.* Wiley Interscience Publication.

Crawford, D.J., S. Brauner, M.B. Cosner and T.F. Stuessey 1993. Use of RAPD markersto document the origin of the intergeneric hybrid *xMargyracaena skottsbergii* (Rosaceae) on the Juan Fernandez Islands. *American Journal of Botany* 80:89-92.

Dawson, I.K., K.J. Chalmers, R. Waugh and W. Powell 1993. Detection and analysis of genetic variation in *Hordeum spontaneum* populations from Israel using RAPD markers. *Molecular Ecology* 2:151-159.

de Graaf, A. and J.C. Arends 1986. The occurance of *Cryptocoryne* and *Lagenandra* (Araceae) on Sri Lanka. *Nordic Journal of Botany* 6:757-764.

Demeke, T, R.P. Adams and R. Chibbar 1992. Potential taxonomic use of random amplified polymorphic DNA (RAPD): A case study in *Brassica*. *Theoretical and Applied Genetics* 84:990-994.

Demesure, B., N. Sodzi and R.J. Petit 1995. A set of universal primers for amplification of polymorphic non-coding regions of mitochondria and chloroplast DNA in plants. *Molecular Ecology* 4:129-131.

De Pamphilis, C.W. and J.D. Palmer 1990. Loss of photosynthetic and chlororespitatory genes from the plastid genome of a parasitic flowering plant. *Nature* 348:337-339.

Ding, S.L., H.Y. Gu, L.J. Qu and Z.L. Chen 1995. A preliminary study on the use of the PCR amplified products in the systematic investigation of the subtribe Astragalinae (Fabaceae). *Acta Botanica Sinica* 37:97-102.

Downie, S.R. and J.D. Palmer 1992. Use of chloroplast DNA rearrangements in reconstructing plant phylogeny. In: *Molecular Systematics of Plants* (eds. P.S. Soltis, D.E. Soltis and J.J. Doyle), 14-35. Chapman and Hall, London.

Downie, S.R. and D.S. Katzdownie 1996. A molecular phylogeny of Apiaceae subfamily Apoideae - Evidence from nuclear ribosomal DNA internal transcribed spacer sequences. *American Journal of Botany* 83:234-251.

Doyle, J.J., M. Lavin and A. Bruneau 1992. Contributions of molecular data to Papilionoid legume systematics. In: *Molecular Systematics of Plants* (eds. P.S. Soltis, D.E. Soltis and J.J. Doyle), 223-251. Chapman and Hall, London.

Doyle, J.J. and J.L. Doyle 1987. A rapid isolation procedure for small amounts of leaf tissue. *Phytochemical Bulletin* 19:11-15.

Dunn, F.L. and D.F. Dunn 1977. Maritime adaptations and exploitation of marine resources in Sundaic Southeast Asian prehistory. *Modern Quaternary Research of South East Asia* 3: 1-28.

Ellsworth, D.L., K.D. Rittenhouse and R.L. Honeycutt 1993. Artifactual variation in randomly amplified polymorphic DNA banding patterns. *Biotechniques* 14:214-217.

Farris, J.S. 1989. The retention index and rescaled consistency index. *Cladistics* 5:417-419.

Feinberg, A.P. and B. Vogelstein 1983. A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Analytical Biochemistry* 132:6-13.

Felsenstein, J. 1985. Confidence limit on phylogenies : An approach using the bootstrap. *Evolution* 39:783-791.

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

Fitch, W.M. 1981. A non-sequential method for constructing trees and hierarchical classifications. *Journal of Molecular Evolution* 18:30-37.

Fitch, W.M. 1984. Cladistic and other methods: Problems, pitfalls, and potentials. In: *Cladistic Perspectives on the Reconstruction of Evolutionary History* (eds. T. Duncan and T.G. Stuessey), 221-252. Columbia University Press, New York.

Fitch, W.M. and E. Margoliash 1967. Construction of phylogenetic trees. *Science* 155:279-284.

Francesco-Ortega, J., R.K. Jansen and A. Santosguerra 1996. Chloroplast DNA evidence of colonisation, adaptive radiation and hybridisation in the evolution of the macaronesian flora. *Proceedings of the National Academy of Sciences, United States of America* 93:4085-4090.

French, J.C., M.G. Chung and Y.K. Hur 1995. Chloroplast phylogeny of the Ariflorae. In: *Monocotyledons :Systematics and Evolution* (eds. P.J. Rudall, P.J. Cribb, D.F. Cutler and C.J. Humphries), 255-276. Royal Botanic Gardens, Kew.

Funk, V.A. 1995. Cladistic methods. In: *Hawaiian Biogeography* (eds. W.L. Wagner and V.A. Funk), 30-38. Smithsonian Institute Press, Washington.

Gielly, L. and P. Taberlet 1996. A phylogeny of the European Gentians inferred from chloroplast *trn*L(UAA) intron sequences. *Botanical Journal of The Linnean Society* 120:57-75.

Gielly, L., Y.M. Yuan, P. Kupfer and P. Taberlet 1996. Phylogenetic use of noncoding regions in the genus *Gentiana* L - Chloroplast *trn*L(UAA) intron versus nuclear ribosomal internal transcribed spacer sequences. *Molecular Phylogenetics and Evolution* 5:460-466.

Gillies, A.C.M. 1994. *Molecular Analysis of Genetic Diversity and Evolutionary Relationships in the Tropical Legum Genus Stylosanthes* (Aubl.) Sw. Ph.D. Thesis, University of St. Andrews.

Gillies, A.C.M. and R.J. Abbott 1996. Phylogenetic relationships in the genus *Stylosanthes* (Leguminosae) based upon chloroplast DNA variation. *Plant Systematics and Evolution* 200:193-211.

Graham, J. and R.J. Nichol 1995. An examination of the ability of RAPD markers to determine the relationships within and between *Rubus* species. *Theoretical and Applied Genetics* 90:1128-1132.

Hallden, C., M. Hansen, N.O. Nilsson, A. Hjerdin and T. Sall 1996. Competition as a source of errors in RAPD analysis. *Theoretical and Applied Genetics* 93:1185-1192.

Hamby, R.K. and E.A. Zimmer 1992. Ribosomal RNA as a phylogenetic tool in plant systematics. In: *Molecular Systematics of Plants* (eds. P.S. Soltis, D.E. Soltis and J.J. Doyle), 50-91. Chapman and Hall, London.

Hamrick, J.L. and M.J.W. Godt 1989. Allozyme diversity in plant species. In: *Plant Population Genetics, Breeding and Genetic Resources* (eds. A.D.H. Brown, M.T. Clegg, A.L. Kahler and B.S. Weir), 43-63. Sinauer Associates, Sunderland, Massachusetts.

Harris, S.A. and R. Ingram 1991. Chloroplast DNA and biosystematics: The effects of intraspecific diversity and plastid transmission. *Taxon* 40:393-4412.

Harris, S.A. and R. Ingram 1992. Molecular systematics of the genus *Senecio* L. II. Hybridisation in a British polyploid complex. *Heredity* 69:1-10.

Harris, S.A., C.E. Huges, R. Ingram and R.J. Abbott 1994. A phylogenetic analysis of *Leucaena* (Leguminosae:Mimoisoideae). *Plant Systematics and Evolution* 19:1-26.

Harris, S. A. in press. RAPDs and systematics - a useful methodology? In: *Advances in Plant Molecular Systematics - Systematic Association, Special Volume* (eds. P. Hollingworth, R.M. Bateman and R. J. Gornall). Chapman and Hall, London.

Harrison, R.E., J.J. Luby and G.R. Furnier 1997. Chloroplast DNA restriction fragment variation among strawberry (*Fragaria* spp.) taxa. *Journal of the American Society for Horticultural Science* 122:63-68.

Heun, M. and T. Helentjaris 1993. Inheritance of RAPDs in F1 hybrids of corn. *Theoretical and Applied Genetics* 85:961-968.

Hiratsuka, J., H. Shimada, R. Whittier, T. Ishibashi, M. Sakamoto, M. Mori, C. Kondo, Y. Honji, C.-R. Sun, B.-Y. Meng, Y.-Q. Li, A. Kanno, Y. Nishizawa, A. Hirai, K. Shinozaki and M. Suguira 1989. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: Intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Molecular and General Genetics* 217:185-194.

Hoisington, D., M. Khairallah and D. Gonzalez-de-Leon 1994. *Laboratory Protocols: CIMMYT Applied Molecular Genetics Laboratory*. Second edition. Mexico, D. F.: CIMMYT.

Hoot, S.B. 1995. Phylogenetic relationships in *Anemone* (Ranunculaceae) based on DNA restriction site variation and morphology. *Plant Systematics and Evolution* S9:295-300.

Hoot, S.B. and P.R. Crane 1995. Interfamilial relationships in the Ranunculidae based on molecular systematics. *Plant Systematics and Evolution* S9:119-131

Howell, E.C., H.J. Newbury, R.L. Swennen, L.A. Withers and B.V. Ford-Lloyd 1994. The use of RAPD for identifying and classifying *Musa* germplasm. *Genome* 37:328-332.

Hsiao, C., N.J. Chatterton, K.H. Asay and K.B. Jensen 1995. Molecular phylogeny of the Pooideae (Poaceae) based on nuclear rDNA (ITS) sequences. *Theoretical and Applied Genetics* 90:389-398.

Huff, D.R. and J.M. Bara 1993. Determining genetic origins of abberant progeny from facultative apomictic Kentucky bluegrass using a combination of flow cytometry and silver-stained RAPD markers. *Theoretical and Applied Genetics* 87:201-208.

Ito, M. and J.H. Pak 1996. Phylogenetic relationships of *Crepidiastrum* (Asteraceae) of the Bonin (Ogasawara) islands based chloroplast DNA restriction site variation. *Journal of Plant Research* 109:277-280.

Jaccard, P. 1908. Nouvelles recherches sur la distribution florale. *Bulletin de la Societe Vaudense des Sciences Naturelles* 44:223-270.

Jacobsen, N. 1977. Chromosomes numbers and taxonomy in *Cryptocoryne* (Araceae). *Botanical Notiser* 130:71-87.

Jacobsen, N. 1979. *Cryptocoryne.* Clausen Boger, Aschehoug. Copenhagen.

Jacobsen, N. 1980a. The *Cryptocoryne albida* group of mainland Asia (Araceae). *Meded. Landbouwhogeschool Wageningen* 19:183-204.

Jacobsen, N. 1980b. Does *Cryptocoryne ferruginea* flower at full moon? *Aroideana* 3:111-116.

Jacobsen, N. 1981. *Cryptocoryne undulata.* Wendt und Bemerkungen zu anderen Arten. *Aqua Planta* 81:31-7.

Jacobsen, N. 1985. The *Cryptocoryne* (Araceae) of Borneo. *Nordic Journal of Botany* 5:31-50.

Jacobsen, N. 1987. Portraits of threatened plant, *Cryptocoryne elliptica. Malayan Naturalis* 40:34-35.

Jelinski, D.E. and W.M. Cheliak 1992. Genetic diversity and spatial subdivision of *Populus tremuloides* (Salicaceae) in a heterogenous landscape. *American Journal of Botany* 79:728-736.

Jensen, U., S.B. Hoot, J.T. Johansson and K. Kosuge 1995. Systematics and phylogeny of the Ranunculaceae - A revised family concept on the basis of molecular data. *Plant Systematics and Evolution* S9: 273-280.

Johns, R.J. 1995. Malesia - An introduction. *Curtis' Botanical Magazine* 12:52-62.

Johnson, L.A. and D.E. Soltis 1994. *mat*K DNA sequences and phylogenetic reconstruction in Saxifragaceae s. str. *Systematic Botany* 19:143-156.

Johnson, L.A. and D.E. Soltis 1995. Phylogenetic inference in Saxifragaceae sensu stricto and *Gilia* (Polemoniaceae) using *mat*K sequences. *Annals of Missouri Botanical Garden* 82:149-175.

Jordan, W.C., M.W. Courtney and J.E. Neigel 1996. Low level of intraspecific genetic variation at a rapidly evolving chloroplast DNA locus in North American duckweeds (Lemnaceae). *American Journal of Botany* 83:430-439.

Jorgensen, R.A. and P.D. Cluster 1988. Modes and tempos in the evolution of nuclear ribosomal DNA: New characters for evolutionary studies and new markers for genetic and population studies. *Annals of the Missouri Botanical Garden* 75:1238-1247.

Jork, K.B and J.W. Kadereit 1995. Molecular phylogeny of the Old World representatives of Papaveraceae subfamily Papaveroideae with special emphasis on the genus *Meconopsis. Plant Systematics and Evolution* [supplement] 9:171-180.

Kim, K.J. and T.J. Mabry 1991. Phylogenetic and evolutionary implications of nuclear ribosomal DNA variation in dwarf dandelions (*Krigia*, Lactuceae, Asteraceae). *Plant Systematics and Evolution* 177:53-69.

Kim, K.J. and R.K. Jensen 1994. Comparisons of phylogenetic hypotheses among different data sets in dwarf dandelions (*Krigia*, Asteraceae): Additional information from internal transcribed spacer sequences of nuclear ribosomal DNA. *Plant Systematics and Evolution* 190:157-185.

Kim, K.J. and R.K. Jensen 1995. Phylogenetic implications of chloroplast DNA variation in the Berberidaceae. *Plant Systematics and Evolution* S9:341-349.

Kim, Y.D. and R.K. Jensen 1996. Phylogenetic implications of *rbc*L and ITS sequence variation in the Berberidaceae. *Systematic Botany* 21:381-396.

Kim, Y.D., D.J. Crawford and R.K. Jensen 1996a. Phylogenetic relationships among the genera of the subtribe Sonchinae (Asteraceae): Evidence from ITS sequences. *Systematic Botany* 21:417-432.

Kim, J.H., H. T. Hart and T.H.M. Mes 1996b. The phylogenetic position of east Asian *Sedum* species (Crassulaceae) based on cpDNA *trn*L (UAA) - *trn*F (GAA) intergenic spacer sequence variation. *Acta Botanica Neerlandia* 45:309-321.

Kita, Y., K. Ueda and Y. Kadota 1995. Molecular phylogeny and Evolution of the Asian *Aconitum* subgenus *Aconitum* (Ranunculaceae). *Journal of Plant Research* 108:429-442.

Kitching, I.J. 1992. Tree-building techniques. In: *Cladistics: A practical Course in Systematics* (eds. P.L. Forey, C.J. Humphries, I.L. Kitching, R.W. Scotland, D.J. Siebert and D.M. Williams), 44- 71. Oxford Science Publications, London.

Kluge, A.G. and J.S. Farris 1969. Quantitative phyletics and the evolution of anurans. *Systematic Zoology* 18:1-32.

Kron, K.A. and J.M. King 1996. Cladistic relationships of *Kalmia, Leiophyllum* and *Loiseluria* (Phyllodoceae, Ericaceae) based on *rbc*L and nuclear ribosomal ITS data. *Systematic Botany* 21:17-29.

Lamboy, W. 1994a. Computing genetic similarity coefficients from RAPD data: correcting for the effects of PCR artefacts caused by variation in experimental conditions. *PCR Methods and Applications* 4:38-43.

Lamboy, W. 1994b. Computing genetic similarity coefficients from RAPD data: the effects of PCR artefacts. *PCR Methods and Applications* 4:38-43.

Lewin, B. 1994. *Genes V*. Oxford University Press.

Li, W.H. and D. Graur 1991. *Fundamentals of Molecular Evolution*. Sinauer Associates, Sunderland, MA.

Lifante, Z.D. and I. Aguinagalde 1996. The use of random amplified polymorphic DNA (RAPD) markers for the study of taxonomical relationships among species of *Asphodelus* sect. *Verinea* (Asphodelaceae). *American Journal of Botany* 83:949-953.

Liston, A. 1992. Variation in the chloroplast genes *rpo*C1 and *rpo*C2 of the genus *Astragalus* (Fabaceae): Evidence from restriction site mapping of a PCR-amplified fragment. *American Journal of Botany* 79:953-961.

Liston, A. and J.A. Wheeler 1994. The phylogenetic position of the genus *Astragalus* (Fabaceae): Evidence from the chloroplast genes *rpo*C1 and *rpo*C2. *Biochemical Systematics and Ecology* 22:377-388

Loconte, H. and J.R. Estes 1989. Phylogenetic systematics of Berberidaceae and Ranunculales (Magnoliidae). *Systematic Botany* 14:565-579.

Manen, J.F. and A. Natali, 1995. Comparison of the evolution of ribulose-1,5-biphosphate carboxylase(*rbc*L) and *atp*B-*rbc*L noncoding spacer sequences in a recent plant group, the Tribe Rubieae (Rubiaceae). *Journal of Molecular Evolution* 41:920-927.

Mansor, M. and M. Masnadi 1994. *Cryptocoryne elliptica,* an endangered amphibious plant in Pondok Tanjung Forest Reserve, Peninsular Malaysia. *Aquatic Botany* 47:91-96.

Marsolais, J.V., J.S. Pringle and B.N. White 1993. Assessment of random amplified polymorphic DNA (RAPD) as genetic markers for determining the origin of interspecific lilac hybrids. *Taxon* 42:531-537.

McCauley, D.E. 1995. The use of chloroplast DNA polymorphism in studies of gene flow in plants. *Trends in Ecology and Evolution* 10:198-202.

Mes, T.H.M. and H.T. Hart 1996. The evolution of growth forms in the macaronesian genus *Aeonium* (Crassulaceae) inferred from chloroplast DNA RFLPs and morphology. *Molecular Ecology* 5:351-363.

Mes, T.H.M., J. vanBrederode and H.T. Hart 1996. Origin of the woody macaronesian Sempervivoideae and the phylogenetic position of the east African species of *Aeonium. Botanica Acta* 109:477-491.

Michaux, B. 1991. Distributional patterns and tectonic development in Indonesia: Wallace reinterpreted. *Australian Systematic Botany* 4:25-36.

Nesbitt, K.A., B.M. Potts, R.E. Vaillancourt, A.K. West and J.B. Reid 1995. Partitioning and Distribution of RAPD variation in a forest tree species, *Eucaliptus globulus* (Myrtaceae). *Heredity* 74:628-637.

Nicholl, D.S.T. 1994. *An Introduction to Genetic Engineering.* Cambridge University Press, Cambridge.

Nicolson, D.H. 1988. Araceae. In: *A Revised Handbook of the Flora of Ceylon* vol. VI (eds. M.D. Dasanayake and F.R. Fosberg), 17-101. Model Press Pvt. Ltd, New Delhi.

Novy, R.G. and N. Vorsa 1996. Evidence for RAPD heteroduplex formation in cranberry: implications for pedigree and genetic-relatedness studies and a source of co-dominant RAPD markers. *Theoretical and Applied Genetics* 85:989-993.

Ohyama, K., H. Fukuzawa, T. Kohchi, H. Shirai, T. Sano, S. Sano, K. Umesono, Y. Shiki, M. Takeuchi, Z. Chang, S. Aota, H. Inokuchi and H. Ozeki 1986. Chloroplast gene organisation deduces from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature* 322:572-574.

Olmstead, R.G., B. Bremer, K.M Scott and J.D. Palmer 1993. A molecular systematic analysis of the Asteridae *sensu lato* based on *rbc*L sequences. *Annals of the Missouri Botanical Garden* 80:700-722.

Olmstead, R.G and J.D. Palmer 1994. Chloroplast DNA systematics: A review of methods and data analysis. *American Journal of Botany* 81:1205-1224

Olmstead, R.G. and J.A. Sweere 1994. Combining data in phylogenetic systematics: An empirical approach using three molecular data sets in the Solanaceae. *Systematic Biology* 43:467-481.

Oxelman, B. and M. Liden 1995. Genetic boundaries in the Tribe Sileneae (Caryophyllaceae) as inferred from nuclear rDNA sequences. *Taxon* 44:525-542.

Palmer, J.D and W.F. Thompson 1982. Chloroplast rearrangements are more frequent when a large inverted repeated sequence is lost. *Cell* 29:537-550.

Palmer, J.D. 1985. Evolution of chloroplast and mitochondrial DNA in plants and algae. In: *Monographs in Evolutionary Biologgy: Molecular Evolutionary Genetics* (ed. R.J. MacIntyre), 131-240. Plenum Press, New York.

Palmer, J.D. 1987. Chloroplast DNA evolution and biosystematic uses of chloroplast DNA variation. *American Naturalis* 130:456-529.

Palmer, J.D., R.K. Jansen, H.K. Michaels, M.W. Chase and J.R. Manhart 1988. Chloroplast DNA variation in plant phylogeny. *Annals of the Missouri Botanical Garden* 75:1180-1206.

Panchen, A.L. 1992. *Classification, Evolution and the Nature of Biology.* Cambridge University Press, Cambridge.

Pax, D.L., R.A. Price and H.J. Michaels 1997. Phylogenetic position of the Hawaiian Geramiums based on *rbc*L sequences. *American Journal of Botany* 84:72-78.

Pennington, R.T. 1996. Molecular and morphological data provide phylogenetic resolution of different hierarchical levels in *Andira*. *Systematic Biology* 45:496-515.

Petch, T. 1928. Notes on *Cryptocoryne*. *Annals of the Royal Botanical Garden, Peradeniya* 11:11-26.

Petersen, G. 1989. Cytology, systematics and chromosomes numbers of Araceae. *Nordic Journal of Botany* 9:119-166.

Petersen, G. and O. Seberg 1997. Phylogenetic analysis of the Triticeae (Poaceae) based on *rpo*A sequence data. *Molecular Phylogenetic and Evolution* 7:217-230.

Price, R.A. and J.D. Palmer 1993. Phylogenetic relationships of the Geraniaceae and Geraniales from *rbc*L sequence comparisons. *Annals of the Missouri Botanical Garden* 80:661-671.

Qiu, Y.L., M.W. Chase and C.R. Park 1995. A chloroplast DNA phylogenetic study of the eastern Asia eastern North-America disjunct section Rytidospermum of *Magnolia* (Magnoliaceae). *American Journal of Botany* 82:1582-1588.

Rataj, K. 1975. Revision of the genus *Cryptocoryne* Fischer. *Studie CSAV*, c. 3:1-74.

Raubeson, L.A. and Jansen, J.D. 1992. A rare chloroplast-DNA mutation is shared by all conifers. *Biochemical Systematics and Ecology* 20:17-24.

Real, R. and J.M. Vargas 196. The probabilistic basis of Jaccard's index of similarity. *Systematic Biology* 45:380-385.

Reiseberg, L.H. and D.E. Soltis 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants* 5:65-84.

Reiseberg, L.H., M.A. Hanson and C.T. Philbrick 1992. Androdioecy is derived from dioecy in Datiscaceae: Evidence from restriction site mapping of PCR-amplified chloroplast DNA fragments. *Systematic Botany* 17:324-336.

Rieseberg, L.H. 1996. Homology among RAPD fragments in interspecific comparisons. *Molecular Ecology* 5:99-105.

Reumer, J.W.F. 1984. Cytotaxonomy and evolution in *Cryptocoryne* (Araceae). *Genetica* 65:149-158.

Rogers, S.O. and A.J. Bendich 1987. Ribosomal RNA genes in plants: Variability in copy number and in the intergenic spacer. *Plant Molecular Biology*. 9:509-520.

Saibeh, K. and Mansor, M. 1996. Differential patterns of insect populations visiting *Cryptocoryne ciliata* flowers. *Malaysian Applied Biology* 25:1-6.

Saitou, N. and M. Nei, 1987. The neighbour-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4:406-425.

Sambrook, J., E. F. Fritsch and T. Maniatis 1989. *Molecular Cloning - Laboratory Manuals*. Second edition. Cold Spring Harbor Laboratory Press.

Sang, T., D.J. Crawford, S.-C. Kimand and T.F. Stuessy 1994. Radiation of the endemic genus *Denderosis* (Asteraceae) on The Juan Fernandez Islands: Evidence from sequence of the ITS regions of nuclear ribosomal DNA. *American Journal of Botany* 81:1494-1501.

Sang, T., D.J. Crawford, T.F. Stuessy and M. Silva O. 1995. ITS sequences and the phylogeny of the genus *Robinsonia* (Asteraceae). *Systematic Botany* 20:55-64.

Schwarzbach, A.E. and J.W. Kadareit 1995. Rapid radiation of Northern American desert genera of the Papaveraceae: Evidence from restriction site mapping of PCR-amplified chloroplast DNA fragments. *Plant Systematics and Evolution* [supplement] 9:159-170.

Scotland, R. W. 1992. Cladistic theory. In: *Cladistics: A Practical Course in Systematics* (eds. P.L. Forey, C.J. Humphries, I.L. Kitching, R.W. Scotland, D.J. Siebert and D.M. Williams), 3-13. Oxford Science Publications, London.

Shinozaki, K., M. Ohme, M. Tanaka, T. Wakasugi, N. Hayashida, T. Matsubayashi, N. Zaita, J. Chunwongse, J. Obokata, K. Yamaguchi-Shinozaki, C. Ohto, K. Yorazawa, B.Y. Meng, M. Sugita, H. Deno, T. Kamagoshira, K. Yamada, J. Kusuda, F. Takaiwa, A. Kato, N. Tohdo, H. Shimada and M. Sugiura 1986. The complete nucleotide sequence of the tobacco chloroplast genome: Its organisation and expression. *The EMBO Journal* 5:2043-2049.

Siebert, D. J. 1992. Tree statistics; trees and confidence; consensus trees; alternatives to parsimony; character weighting; character conflict and its resolution. In: *Cladistics: A Practical Course in Systematics* (eds P.L. Forey, C.J. Humphries, I.L. Kitching, R.W. Scotland, D.J. Siebert and D.M. Williams), 72-88. Oxford Science Publications, London.

Skroch, P.W. and J. Nienhuis 1995. Qualitative and quantitative characterization of RAPD variation among snap bean (*Phaseolus vulgaris*) genotypes. *Theoretical and Applied Genetics* 91:1078-1085

Smith, J.J., J.S. Scott-Craig, J.R. Leadbetter, G. Bush, D.L. Roberts and D.W. Fulbright 1994. Characterization of random amplified polymorphic DNA (RAPD) products from *Xanthomonas camprestris* and some comments on the used of RAPD products in phylogenetic analysis. *Molecular Phylogenetics and Evolution* 3:135-145.

Soltis, D.E., T.E. Collier and M.L. Edgerton 1991. Chloroplast DNA variation within and among genera of the *Heuchera* group (Saxifragaceae): Evidence from chloroplast transfer and paraphyly. *American Journal of Botany* 78:1091-1112.

Soltis, D.E. and R.K. Kuzoff 1995. Discordance between nuclear and chloroplast phylogenies in the *Heuchera* group (Saxifragaceae). *Evolution* 49:727-742.

Soltis, P.S. and D.E. Soltis 1995. Plant molecular systematics - Inferrence of phylogeny and evolutionary processes. *Evolutionary Biology* 28:139-194.

Soltis, P.S., D.E. Soltis, S.G. Weller, A.K. Sakai and W.L. Wagner 1996. Molecular phylogenetic analysis of the Hawaiian endemics *Schiedea* and *Alsinidendron* (Caryophyllaceae). *Systematic Botany* 21:365-379.

Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* 98:503-517.

SouzaChies, T.T., G. Bittar, S. Nadot, L. Carter, E. Besin and B. Lejeune 1997. Phylogenetic analysis of of Iridaceae with parsimony and distance methods using plastid gene *rps*4. *Plant Systematics and Evolution* 204:109-123.

Sovalainen, V., J.F. Manen, E. Douzery and R. Spichiger 1994. Molecular phylogeny of families related to Celastrales based on *rbc*L 5' flanking sequences. *Molecular Phylogenetics and Evolution* 3:27-37.

Stammers, M., J. Harris, G.M. Evans, M.D. Hayward and J.W. Forster 1995. Use of random PCR(RAPD) technology to analyse phylogenetic relationships in *Lolium/Festuca* complex. *Heredity* 74:19-27.

Steele, K.P. and R. Vilgalys 1994. Phylogenetics analyses of Polemoniaceae using nucleotide sequences of the plastid gene *mat*K. *Systematic Botany* 19:126-142.

Stothard, J.R. 1997. Phylogenetic inference with RAPDs: some observations involving computer simulation with viral genomes. *Journal of Heredity* 88:222-228.

Sun, Y., D.Z. Skinner, G.H. Liang and S.H. Hulbert 1994. Phylogenetic analysis of *Sorghum* and related taxa using internal transcribed spacers of nuclear ribosomal DNA. *Theoretical and Applied Genetics* 89:26-32.

Susanna, A., N.G. Jacas, D.E. Soltis, and P.S. Soltis, 1995. Phylogenetic relationships in tribe Cardueae (Asteraceae) based on ITS sequences. *American Journal of Botany* 82:1056-1068.

Swofford, D.L. 1993. *PAUP: Phylogenetic Analysis Using Parsimony,* version 3.1. Illinois Natural History Survey, Champaign, IL.

Swofford, D.L., G.J. Olsen, P.J. Waddell and D.M. Hillis 1996. Phylogenetic inference. In: *Molecular Systematics* 2nd edition (eds. D.M. Hillis, C. Moritz and B.K. Marble), 407-514. Sinauer Associates, Sunderland, MA.

Taberlet, P., L. Gielly, G. Pautou and J. Bouvet 1991. Universal primers for amplification of 3 non-coding regions of chloroplast DNA. *Plant Molecular Biology* 17:1105-1109.

Terry, R.G., G.K. Brown and R.G. Olmstead, 1997. Examination of subfamilial phylogeny in Bromeliaceae using comparative sequencing of the plastid locus *ndh*F. *American Journal of Botany* 84:664-670.

Tsumura, Y., K. Yoshimura, N. Tomaru and K. Ohba 1995. Molecular phylogeny of conifers using RFLP analysis of PCR-amplified specific chloroplast genes. *Theoretical and Applied Genetics* 91:1222-1236.

Tsumura Y., T. Kawahara, R. Wickneswari and K. Yoshimura 1996. Molecular phylogeny of Dipterocarpaceae in South East Asia using RFLP of PCR-amplified chloroplast genes. *Theoretical and Applied Genetics* 93:22-29.

Van Buren, R., K.T. Harper, W.R. Andersen, D.J. Stanton, S. Seyoum and J.L. England 1994. Evaluating the relationships of autumn Buttercup (*Ranunculus acriformis* var. *aestivalis*) to some close congeners using random amplified polymorphic DNA. *American Journal of Botany* 81:514-519.

Venugopal, G.,S. Mohapatra, D. Salo and S. Mohapatra 1993. Multiple mismatch annealing: basis for random amplified polymorphic DNA fingerprinting. *Biochemical and Biophysical Research Communications* 197:1382-1387.

Virk, P.S., B.V. Ford-Lloyd, M.T. Jackson and H.J. Newbury 1995. Use of RAPD for the study of diversity within plant germplasm collections. *Heredity* 74:170-179.

Wang, W.Y, R.C. Pai, C.C. Lai and T.P. Lin 1994. Molecular evidence for the hybrid origin of *Paulownia Taiwania* based on RAPD markers and RFLP of chloroplast DNA. *Theoretical and Applied Genetics* 89:271-275.

Wendel , J.F., A. Schnabel and T. Seelanan 1995. An unusual ribosomal DNA sequence from *Gossypium gossypioides* reveals ancient, cryptic, intergenomic introgression. *Molecular Phylogenetic and Evolution* 4:298-313.

Welsh, J. and M. McCleland 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acid Research* 18:7213-7218.

Wen, J. and E.A. Zimmer 1996. Phylogeny and biogeography of *Panax* L (The Ginseng Genus, Araliaceae)- Inferences from ITS sequences of nuclear ribosomal DNA. *Molecular Phylogenetics and Evolution* 6:167-177.

Wen, J., J. Jansen and K. Kilgore 1996. Evolution of the eastern Asia and eastern North American disjunct genus *Symplocarpus* (Araceae): Insights from chloroplast DNA restriction site data. *Biochemical Systematics and Ecology* 24:735-747.

White, T.J., T. Bruns, S. Lee and J. Taylor 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: *PCR Protocols:A Guide to Methods and Applications* (eds. M. Innis, D. Gelfand, J. Snisky and T. White), 315-322. Academic Press, San Diego, California.

Whittemore, A.T and B.A. Schaal 1991. Interspecific gene flow in sympatric oaks. *Proceedings of the National Academy of Sciences, United States of America* 88:2540-2544.

Wiley, E.O., D. Siegal-Causey, D.R. Brooks and V.A. Funk 1991. *The Compleat Cladist: A Primer of Phylogenetic Procedure.* The University of Kansas, Lawrence, Kansas.

Williams, D.M. 1992. DNA analysis: methods. In: *Cladistics: A Practical Course in Systematics* (eds. P.L. Forey, C.J. Humphries, I.L. Kitching, R.W. Scotland, D.J. Siebert and D.M. Williams), 102-123 Oxford Science Publications, London.

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 Research* 18:6531-6535.

Williams, J.G.K., M.K. Hanafey, J.A. Rafalski and S.V. Tingey 1993. Genetic analysis using random amplified polymorphic DNA markers. *Methods in Enzymology* 218:704-740.

Wojciechowski, M.F., M.J. Sanderson, B.G. Baldwin and M.J. Donoghue 1993. Monophyly of aneuploid *Astragalus* (Fabaceae): Evidence from nuclear ribosomal DNA internal transcribed spacer sequence. *American Journal of Botany* 80:711-722.

Wolfe K.H., C.W. Morden and J.D. Palmer 1992. Function and evolution of a minimal plastid genome from a nonphptosynthetic parasitic plant. *Proceedings of the National Academy of Sciences, United States of America* 89:10648-10652.

Wolfe, A.D., W.J. Elisens, L.E. Watson and W. DePamphilis 1997. Using restriction site variation of PCR amplified cpDNA genes for phylogenetic analysis of Tribe Cheloneae (Scrophulariaceae). *American Journal of Botany* 84:555-564.

Xiang, Q.Y., S.J. Brunsfeld, D.E. Soltis and P.S. Soltis 1996. Phylogenetic relationships in *Cornus* based on chloroplast DNA restriction sites: Implications for biogeography and character evolution. *Systematic Botany* 21:515-534.

Yeh, F.C., D.K.X. Chong and R.C. Yang 1995. RAPD variation within and among natural populations of trembling Aspen (*Populus tremuloides* Michx) from Alberta. *Journal of Heredity* 86:454-460.

Yu, K.F. and K.P. Pauls 1993. Segregation of random amplified polymorphic DNA markers and strategies for molecular mapping in tetraploid alfalfa. *Genome* 36:844-851.

Yuan, Y.M and P. Kupfer 1995. Molecular phylogenetics of the subtribe Gentianinae (Gentianaceae) inferred from the sequences of the internal transcribed spacers (ITS) of nuclear ribosomal DNA. *Plant Systematics and evolution*. 196:207-226.

Yuan, Y.M., P. Kupfer and J.J. Doyle 1996. Infrageneric phylogeny of the genus *Gentiana* (Gentianaceae) inferred from nucleotide sequences of the internal transcribed spacers (ITS) of nuclear ribosomal DNA. *American Journal of Botany*. 83:641-652.