CHAPTER 1

General introduction

1.1 Introduction to the taxonomy of Cyperaceae

Cyperaceae is the third largest monocotyledonous family. It is cosmopolitan, including 104 genera (Goetghebeur 1998) and c.5000 species (Bruhl 1995; Goetghebeur 1998), occurring in a range of different ecogeographic habitats and commonly dominant in swamps. This family shows the greatest diversity in the humid and semi-humid tropics, but members are also often dominant in temperate and cold temperate (Goetghebeur 1998) and even sometimes arctic regions of the world (e.g. *Carex* L. Aiken *et al.* 1999). Some members of this family are among the largest and most ecologically significant angiosperm genera (e.g. *Carex* L., c. 2000 spp.; *Cyperus* L., c. 600 spp.) (Starr *et al.* 2003)

Cyperaceae is currently regarded as one of the Commelinid monocot families in the order Poales (Chase *et al.* 2000). Some workers put this family in its own order Cyperales in the super order Commelinanae (Muasya *et al.* 2000a). According to Chase *et al.* (2000), the family is near Juncaceae and Thurniaceae and together they form the sedge clade.

Various botanists have produced superfamilial classifications of Cyperaceae. Cronquist (1998) placed this family within Cyperales with the family Poaceae, while Juncaceae as well as Thurniaceae were classified within Juncales. However, most studies (e.g. Takhtajan 1980; Dahlgren *et al.* 1985) have viewed Cyperaceae as closely linked to the Juncaceae, placing Cyperaceae, Juncaceae, and Thurniaceae in Cyperales, a taxonomy supported by cladistic analysis of morphological (Simpson 1995) and molecular (Chase *et al.* 1993; Plunkett *et al.* 1995) data.

The phylogeny of Cyperaceae inferred by Muasya *et al.* (1998) using DNA sequence data from 40 genera showed some incongruence with both Goetghebeur's (1986) and Bruhl's (1995) non-molecular classifications. Muasya *et al.* (2000a) showed that, in contrast to Plunkett *et al.* (1995), and similarly to the results of Muasya *et al.* (1998), Cyperaceae are monophyletic, and sister to *Oxychloe* (Juncaceae). Recent presentations at Monocots 3 (Drabkova *et al.* pers. comm.; Jones *et al.* pers. comm.—see www.monocots3.org/general.htm#) soundly confirm the sister relationship of Cyperaceae and the family Juncaceae.

1.1.1 Difficulties in Cyperaceae

Cyperaceae are known as a difficult family for taxonomists, particularly because the flowers are very small, the homology of the plant parts is mostly indistinct, and interpreting inflorescence morphology is very difficult (Bruhl 1995). Bruhl (1991) discussed difficulties of anatomy and morphology of flowers and inflorescences in the family.

Numerous authors have classified Cyperaceae into tribes and subfamilies. Jussieu (1789) was the first to provide a suprageneric grouping, and divided the family into unisexual-flowered versus hermaphrodite-flowered subfamilies. Later classifications have considered subfamilies as well as tribes in most cases (e.g. Bentham 1883; Clarke 1908; Koyama 1961; Schultze-Motel 1964; Koyama 1969a, 1971; Hooper 1973; Goetghebeur 1986; Bruhl 1990; Goetghebeur 1998).

The number of tribes has varied from six to seventeen. Table 1.1 (modified from Bruhl 1995) is a summary of the major classifications from 1883 to 1998 for the family.

In the classifications of Bentham (1883) and Schultze-Motel (1964, which followed Bentham's classification fairly closely) the family Cyperaceae are divided into six and eight tribes respectively, essentially based on flower structure. Clarke (1908) recognised six different tribes. His classification was based on flower structure as well. Koyama's (1961) classification was firstly limited largely to a consideration of floral and spikelet morphology but his later classifications also consider fruit anatomy and vegetative characters (Koyama 1969, 1971). Hooper (1973) did not use subfamilies and only recognised eight tribes based on floral characters. More recent classifications of the family have been based on a greater range of data, particularly vegetative morphology and anatomy, variation in photosynthetic pathways, embryo morphology, and vegetative ultrastructure (Bruhl 1995; Goetghebeur 1998). Using phylogenies based on morphological and embryological data, Goetghebeur divided the family into four subfamilies and 15 or 17 tribes (Goetghebeur 1986, 1998) while Bruhl (1995) classified the family into two subfamilies, Cyperoideae and Caricoideae, and twelve tribes (Table 1.1).

The theses of Goetghebeur (1986) and Bruhl (1990) and also their respective publications (e.g. Bruhl 1995; Goetghebeur 1998) are the latest comprehensive non-

Table 1.1. Comparison of the recent suprageneric classifications of Cyperaceae

Subfamilies, underlined, and tribes are listed following the order of Goetghebeur (1998). Codes to the left of names in columns 2–7 relate these names to the suggested classification in the first column (see Bruhl 1995).

Goetghebeur	Bruhl	Goetghebeur	Koyama	Schultze-Motel	Koyama	Bentham
1998	1995	1986	1969/71	1964	1961	1883
A. Cyperoideae	A: Cyperoideae	A: Cyperoideae	A: Scirpoideae	A: Cyperoideae	A: Scirpoideae	A: Monoclines
A1: Scirpeae	A1– 3, 5, 6: Scirpeae	A1: Scirpeae	A1– 6: Scirpeae	A1–5: Scirpeae#	A1– 6: Scirpeae	A1– 6: Scirpeae
A2: Fuireneae		A2: Fuireneae				
A3: Eleocharideae		A3: Eleocharideae				
A4: Abildgaardieae	A4: Abildgaardieae	A4: Abildgaardieae				
		A5: Ficinieae				
A5: Cypereae	A5: Cypereae	A5: Cypereae	A5: Cypereae	A5: Cypereae	A5: Cypereae	A5, D1, 2: Hypolytreae
A6: Dulichieae		A6: Dulichieae		A6: Dulichieae		
	A7: Arthrostylideae	A7: Arthrostylideae				
	B: Caricoideae				B: Rhynchosporoideae	
	A7: Rhynchosporeae	A7: Rhynchosporeae	A7: Rhynchosporeae	A7: Rhynchosporeae	A7: Rhynchosporeae	A7: Rhynchosporeae *
A7: Schoeneae	A7: Schoeneae	A7: Schoeneae				
				D1, 2: Hypolytreae		
B: Sclerioideae		B: Sclerioideae	B: Mapanioideae	B: Caricoideae		B: Diclines
B1: Cryptangieae	B1: Cryptangieae	B1: Cryptangieae	B1, 2: Lagenocarpeae	B1, 2: Lagenocarpeae	B1-4, C1: Sclerieae	B1, 2, 4: Cryptangieae
B2: Trilepideae	B2: Trilepideae	B2: Trilepideae				B2, 3, C1: Sclerieae
B3: Sclerieae	B3: Sclerieae	B3: Sclerieae	B3, 4: Sclerieae	B3, 4: Sclerieae		
B4: Bisboeckelereae	B4: Bisboeckelereae	B4: Bisboeckelereae				
C: Caricoideae		C: Caricoideae	C1: Caricoideae		C: Caricoideae	
C1: Cariceae	C1: Cariceae	C1: Cariceae		C1: Cariceae	C1: Cariceae	C1: Cariceae
D: Mapanioideae		D: Mapanioideae			D: Mapanioideae	
D1: Hypolytreae	D1, 2: Hypolytreae	D1: Hypolytreae	D1, 2: Mapanieae		D1, 2: Hypolytreae	
D2: Chrysitricheae		D2: Chrysitricheae				

molecular works on the relationships of t ıe genera of Cyperaceae. Both workers recognised a similar number of genera for the family, although they differed in explanations of homologies of some morphological characters, particularly of flowers and spikelet characters. For instance, the definition of flowers and spikelets has remained a matter of controversy, and mapanioids (including Hypolytreae and Mapanieae) have been differently assigned depending on whether their inflorescence units were interpreted as bisexual flowers with perianth or as unisexual spikelets (Bruhl 1991).

1.1.1.1 Problem of defining Scirpeae and related tribes

Bentham (1883) categorised Scirpeae in a broad way, as possessing a range of different spikelets (from solitary to clustered umbellate) with several or rarely 2 or 1 flowers and only 1 or 2 empty glumes at the base of the spikelets. However, these were all synapomorphies which later resulted in the introduction of a few other tribes out of Scirpeae such as Cypereae (Clarke 1908), Fimbristylideae (Metcalfe 1971; Raynal 1973), which later was classified as Abildgaardieae (Goetghebeur 1986, 1998; Bruhl 1995), Furineae (Goetghebeur 1986, 1998), Eleocharideae (Goetghebeur 1986, 1998), Ficinieae (Goetghebeur 1986), Dulichieae (Schultze-Motel 1964; Hooper 1973, Goetghebeur 1986, 1998) . Clarke (1908) narrowed Scirpeae and shifted many taxa to a new tribe Cypereae (which later included more taxa of Scirpeae; Koyama 1969a) mainly on the basis of flower structure. Later, the tribe Scirpeae was divided into Scirpeae and Dulichieae by Hooper (1973). Raynal (1973) extracted five different categories within *Scirpus* sensu lato including members of Scirpeae themselves. He stated that there are two tribes and two genera that must be separated from *Scirpus* sensu lato for distinct differences. According to his categorisation members of the tribe Fimbristyloideae (based on thickened style base in this group), tribe Cypereae (based on distichous glumes in this group), and genera *Fuirena* and *Eriophorum* which were classified within *Scirpus* sensu lato before must be considered as separate groups from Scirpeae in general and *Scirpus* in particular. Van der Veken (1965) had already classified Raynal's Fimbristyloideae as section *Nemum* (Desv. Ex Ham.) C.B. Clarke within the genus *Scirpus*.

Goetghebeur (1986) concluded that more genera should be excluded from Scirpeae. The analyses of Muasya *et al.* (2000a) based on rbcL and morphological data gave results that were largely different from Goetghebeur's (1986) morphological trees. So, excluding those genera from Scirpeae must be either delayed or be done tentatively until further molecular and non-molecular data are available.

Bruhl (1995) doubted the distinctness of Eleocharideae, Fuireneae, Ficinieae, and Dulichieae, and included them in the tribe Scirpeae. Traditionally *Eleocharis* has been placed within tribe Scirpeae, but Bruhl (1995) considered that the relationships among the diverse portions within the tribe, including *Bolboschoenus* (Ascherson) Palla, *Isolepis* R. Br., *Schoenoplectus* (Reichb.) Palla, *Scirpus* L. s.s., and *Trichophorum* Pers., and allied genera or tribes were in need of much more study.

1.1.2 Abildgaardieae

A tribe equivalent to Abildgaardieae was first recognised under the name Fimbristylideae (Reichenbach 1828). Lye (1973) created the word Abildgaardieae for the tribe. This tribe belongs to family Cyperaceae subfamily Cyperoideae (Bruhl 1995; Goetghebeur 1998) (Table 1.1).

Abildgaardieae is composed of six or seven genera and about 480 species mostly distributed in tropical or subtropical habitats with a few cosmopolitan species. The group ranges from tiny annuals a few centimetres tall, to tall rhizomatous perennials.

Fimbristylis (c. 300 spp.) is found in all pantropical to warm-temperate regions of the world with a heavy concentration in SE Asia, Malaysia, and NE Australia (Goetgheubeur 1998). *Bulbostylis* Kunth (150 spp.) is widely distributed in tropical or subtropical regions worldwide, especially in tropical Africa and South America (Lopez 1996). The species of *Bulbostylis* grow mainly in dry, sandy areas (Prata *et al.* 2001). *Abildgaardia* Vahl (17 spp.) is distributed widely in the tropics and subtropics, concentrated in Australia and Africa. *Nemum* Desv. ex Ham. (10 spp.) (Lye 1989) is found in tropical Africa. *Crosslandia* W.V.Fitzgerald (1–3 spp.) is endemic to northern Australia. The monotypic *Nelmesia* Van der Veken is endemic to the northern regions of Zaire (Goetghebeur 1998). *Tylocarya cylindrostachya* Nelmes, which has been considered a species of *Fimbristylis*, *F. nelmesii* Kern, by recent authors (e.g. Kern 1958; Simpson and Koyama 1998), is found in South East Asia.

Four of these genera, *Abildgaardia* (c. 9 spp.), *Bulbostylis* (c. 7 spp.), *Crosslandia* (1–3 spp.), and *Fimbristylis* (c. 85 spp.) are native in Australia.

1.1.2.1 Relationships of Abildgaardieae and its closest tribes

The circumscription and taxonomic position of Abildgaardieae and associated genera have varied greatly. As stated above, six or seven genera, including the large *Fimbristylis* Vahl, generally have been included in this tribe (e.g. Goetghebeur 1998), depending on whether *Tylocarya* is regarded as a separate genus or as belonging in *Fimbristylis*.

Bulbostylis and *Fimbristylis* have been, in some classifications, treated as close relatives of *Eleocharis* (Goetghebeur 1985; Kukkonen 1990). *Eleocharis* differs from *Fimbristylis* in the occurrence of hypogynous bristles and by the persistent style-base forming a button on the nut (Kern 1974). Goetghebeur (1986) separated Eleocharideae from Abildgaardieae by attributes such as reduced inflorescence to solitary spikelets, elaminate leaves, and *Eleocharis*-type embryo. The first two occur in at least some species of both of his neighbouring tribes (the Abildgaardieae and the Fuireneae) and as Bruhl (1995) explains 'the last attribute can be reduced to a matter of whether the first embryonic leaf primordium is exserted (*Eleocharis*-type) or not (*Schoenoplectus*-type)'. However, his observations and Goetghebeur's drawings (1986, p. 382) show this character to be variable for both *Eleocharis* and *Schoenoplectus* (Bruhl 1995).

The 'Eleocharideae' did not often group with Abildgaardieae in Bruhl's (1995) analyses; on the contrary it was often seemed to be close to Scirpeae. The 'Eleocharideae' and Abildgaardieae share zoned stigmatic papillae and enlargement of the style base (Bruhl 1995). This ambiguity was not resolved when C_4 species of *Eleocharis* (Bruhl *et al.* 1987; Ueno *et al.* 1988) were discovered, as these species are not similar to the Abildgaardieae C_4 species anatomically (Bruhl *et al.* 1987), ultrastructurally (Bruhl 1990; Bruhl and Perry 1990), or biochemically (Bruhl *et al.* 1987). Even though the sister group relationship between the Abildgaardieae and the 'Eleocharideae' postulated by Goetghebeur (1986) and implied by Raynal (1973) was not generally substantiated by Bruhl's (1995) analyses, Bruhl's four eleocharid groups (the C_3 and the C_4 species of *Eleocharis*, *Egleria* and *Websteria*) plus *Androtrichum* formed the sister group of the Abildgaardieae in a couple of analyses, and the 'Eleocharideae' constituted a clade within the Abildgaardieae in two other analyses of his (although Arthrostylideae were also included in these cases).

A few evolutionary classifications have been suggested for *Fimbristylis* and related genera founded on inflorescence morphology, embryology, anatomy, and physiology (e.g. Bruhl 1995); however. ideas differ about these schemes. Inquiries into the relationships within the tribe Abildgaardieae have been stifled by unclear homology of inflorescence and vegetative characters. The main problem of classification in the Abildgaardieae is the circumscription of the large genus *Fimbristylis* (more than 300 spp.) and the distinction of the related genera *Bulbostylis*, and *Abildgaardia*. These have been differently treated by different authors (Table 1.2) (e.g. Vahl 1805; Kunth 1837; Koyama 1961; Gordon-Gray 1971; Lye 1971, 1973, 1974a, 1974b, 1982, 1983; Kern 1974; Goetghebeur 1984a, 1984b, 1986; Goetghebeur and Coudijzer 1984, 1985).

Table 1.2. Comparison of the intratribal classifications of Abildgaardieae
The genera are listed following the order of Goetghebeur and Coudijzer (1985).

Goetghebeur & Coudijzer 1985	Lye 1973/74	Kern 1974	Koyama 1961	Kunth 1837	Vahl 1805
A: *Fimbristylis*	A: *Fimbristylis*	A,B: *Fimbristylis*	A,C: *Fimbristylis*	A: *Fimbristylis*	A,C: *Fimbristylis*
B: *Abildgaardia*	B,C: *Abildgaardia*		B: *Abildgaardia*	B: *Abildgaardia*	B: *Abildgaardia*
C: *Bulbostylis*		C: *Bulbostylis*	C: *Bulbostylis*		

Bruhl (1995) found that the phenetic similarity for Abildgaardieae was not constant, but linked the tribe to Cypereae and/or Scirpeae in most of his phenetic analyses. Bruhl (1995) in his cladistic analysis and Muasya *et al.* (1998) have mentioned Scirpeae as the closest tribe to Abildgaardieae. Goetghebeur (1986) distinguished Scirpeae s.l. as the closest tribe to Abildgaardieae, too, although he divided Scirpeae into six different tribes suggesting Eleocharideae and Fuireneae as the closest to Abildgaardieae. Kukkonen (1990) also considered *Eleocharis* as the closest genus to the members of Abildgaardieae although he believed that these members were in fact all part of Scirpeae. Bruhl noted the Scirpeae as the most poorly supported among the tribes recognised by his analyses, and explained it generally as the 'conservative and convenient means of dealing with these genera' (Bruhl 1995). The molecular study of Cyperales using *rbc*L (Muasya *et al.* 1998) suggested a part of *Schoenoplectus* to be a sister group of *Eleocharis*. Unfortunately, most clades in the resulted tree have insufficient bootstrap value to be accepted, including a clade placing *Schoenoplectus* together with *Eleocharis*.

Muasya *et al.* (2000a) presented the most recent hypothesis about the phylogeny of Abildgaardieae, which they considered as sister to Scirpeae. They found Arthrostylideae as the closest tribe to Abildgaardieae with a moderate support while both tribes were nested within Scirpeae with low support, leaving the issue of the position of the Abildgaardieae/Arthrostylideae group unresolved. The limits and relations of most of the genera in these tribes, particularly *Fimbristylis*, have remained unresolved in the studies uncertaken by Goetghebeur (1986, 1998), Bruhl (1995) and Muasya *et al.* 2000a.

1.1.3. Arthrostylideae

The Arthrostylideae (formally recognised and named by Goetghebeur 1986) consists of four genera: *Arthrostylis* (2 species), *Trachystylis* (monotypic), *Trichoschoenus* (monotypic), and *Actinoschoenus* (7 or 8 species including several Australian natives including the species reviewed by Raynal 1967). However, Goetghebeur changed his mind and subsequently (Goetghebeur 1998) considered all the genera of Arthrostylideae as closely related to Schoeneae, based on similarity of spikelet structure where flowers are wrapped by the wings of the next glume (Tables 1.1 and 1.3). Bruhl (1995), in contrast, found a close relationship between the C_3 species of *Abildgaardia* and *Fimbristylis* and some members of Arthrostylideae considering both tribes (Abildgaardieae and Arthrostylideae) as subgroups of Cyperoideae while treating Schoeneae under Caricoideae. Other studies suggest that the sister of Abildgaardieae is Arthrostylideae (one cladistic analysis by Bruhl 1995; Muasya *et al.* 2000a), Scirpeae (a few cladistic analyses by Bruhl 1995; Muasya *et al.* 2000b), or Eleocharideae (Goetghebeur 1986).

The sister group to the Arthrostylideae (i.e. usually excluding *Trichoschoenus*, which shows a strong affinity to the tribe Schoeneae- see discussion below) in most of the cladistic analyses by Bruhl (1995) was *Abildgaardia* (the C_3 species) or the Abildgaardieae generally, but Bruhl also found other sister group relationships in a few analyses. He elaborated on a hypothesis that *Abildgaardia* may be from different origins by indicating that C_3 species seemed to be more closely related to Arthrostylideae than Abildgaardieae (Tables 1.1 and 1.3).

Goetghebeur's (1986) treatment consistently favoured relationships between the Arthrostylideae and the tribes Rhynchosporeae and Schoeneae, however, it is not consistent enough because only two features support that cladogram in Bruhl's (1995) study. Goetghebeur (1986) placed all three tribes and Abildgaardieae in Cyperoideae while Bruhl (1995) believed that a thorough treatment of *Trichoschoenus* and the limits of the tribe Arthrostylideae are necessary to clarify the position of this genus. In most of Bruhl's analyses *Trichoschoenus* did not make a monophyletic clade with other genera of Arthrostylideae. Meanwhile these sister group relationships with Arthrostylideae are mostly based on the robust clade of *Actinoschoenus*, *Arthrostylis*, and *Trachystylis*, with *Trichoschoenus* as their sister group (Bruhl 1995) (although this result could be an artefact of inadequate material of *Trichoschoenus*). *Trichoschoenus* exhibits a number of characters, especially of the style and fruit (Raynal 1968; Bruhl 1995) and embryo (Bruhl 1995) that are distinct from other members of the tribe, but very close to some members of the tribe Schoeneae such as *Costularia* and *Oreobolus* (Table 1.3).

Other members of the tribe have been classified in various ways too. Kern (1974) included *Actinoschoenus* in *Fimbristylis*, in tribe Cypereae of subfamily Cyperoideae. Goetghebeur (1998) suggested that Abildgaardieae and Arthrostylideae might not be closely related and included the members of Arthrostylideae in Schoeneae. However, Bruhl (1995) demonstrated that some taxa from Abildgaardieae (e.g. *Abildgaardia* C_3 species) are closer to the tribe Arthrostylideae than the tribe Abildgaardieae.

Bruhl (1995) regarded the relationships of Rhynchosporeae and Arthrostylideae as equivocal. In his analyses the Arthrostylideae were included in Cyperoideae, but Rhynchosporeae were included in Caricoideae, whereas Goetghebeur (1986), like many other researchers before him (e.g. Kern 1974), included the Schoeneae, Rhynchosporeae and Arthrostylideae in Cyperoideae, as the first two tribes are sister groups and the latter one is the basal clade in his analysis. In fact, two nodes of Goetghbeur's (1986) cladogram were not supported by Bruhl's (1995) analyses and the relationship between the three tribes was judged by Bruhl to be an unresolved trichotomy. In general, Bruhl's (1995) recognition of the tribes Cypereae, Scirpeae, Abildgaardieae and Arthrostylideae as members of Cyperoideae is broadly compatible with Raynal's classification (1973) and is, as far as it goes, congruent with Goetghebeur's (1986) classification. Bruhl (1995) also interpreted some overlapping features apparent in his analyses between features of Caricoideae and Arthrostylideae as either support for putting Arthrostylideae in Caricoideae or just a homoplasy in cladistic analysis and an incidental compatibility in phenetic analysis.

The position of Arthrostylideae remained uncertain in Bruhl's (1995) work. He emphasised that nucleic acid sequencing would help to resolve the problems of Cyperaceae and its tribal limits (Bruhl 1995).

1.1.4. Genera of Abildgaardieae

Numerous synapomorphies within the genera of Abildgaardieae have been recognised from various datasets. All these genera share fimbristyloid embryos (Goetghebeur 1986). The style base (which is often thickened and deciduous or persistent on the nut) is sharply differentiated from the fruit apex in all the members of Abildgaardieae (Bruhl 1995). Combined molecular and non-molecular data (Muasya *et al.* 2000a) also supported the monophyly of the tribe Abildgaardieae with several synapomorphies among these genera. However, these analyses used a limited range of species so further sampling is needed to confirm these results.

Different analyses by Goetghebuer (1986) and Bruhl (1995) indicate that Abildgaardieae are monophyletic, paraphyletic, or polyphyletic.

Goetghebeur (1986), in general, showed Abildgaardieae as a monophyletic group using morphological and embryological characters. Bruhl (1995) found a monophyletic Abildgaardieae in some analyses using non-molecular data across various combinations of taxa:

1. All its genera but omitting C_3 species of *Abildgaardia.*
2. Omitting *Nelmesia* and *Nemum.*

Table 1.3. Suggested relationships and placement of the genera of tribe Arthrostylideae.

Classification	*Actinoschoenus*	*Arthrostylis*	*Trichoschoenus*	**C_3 spp. of** *Abildgaardia*	Sister group
Kern (1974); Koyama (1974)	in tribe Cyperaeae (as synonym of *Fimbristylis*)	–		–	–
Goetghebeur (1986)	✓	✓	✓	–	Rhynchosporeae or Schoeneae
Bruhl (1995)	✓	✓	✓ but not consistently. Also suggested to be close to some members of Schoeneae	close to Arthrostylideae	Rhynchosporeae or Schoeneae or C_3 species of *Abildgaardia* and *Fimbristylis* or Abildgaardieae as a whole
Goetghebeur (1998)	in tribe Schoeneae	in tribe Schoeneae	in tribe Schoeneae	–	–

– = not mentioned by that author; ✓ = that genus is recognised by that author in tribe Arthrostylideae

3. Omitting *Nelmesia* and *Nemum* and combining the C_3 and C_4 species of *Abildgaardia*.

Other phylogenetic analyses based on non-molecular data (Simpson 1995) and DNA sequencing data (Plunkett 1995; Muasya *et al.* 2000a) of Cyperaceae or Cyperales (sensu Dahlgren *et al.* 1985) also recover a monophyletic tribe Abildgaardieae. The main cause of Abildgaardieae appearing as para- or polyphyletic is the separation of C_3 species of *Abildgaardia* in Bruhl's (1995) study. The features of C_3 species of *Abildgaardia* were in conflict with the rest of Abildgaardieae when six characters (related to the leaf blade and photosynthetic characters) were scored in (Bruhl 1995).

Because generic boundaries within the tribe Abildgaardieae are founded largely on the form of the style base and embryological and micromorphological characters such as second embryonic leaf primordium and chromosome number, clarifying the generic boundaries of the genera that have been classified differently (either as an independent genus or as a section within another genus), is a big challenge. *Abildgaardia*, *Fimbristylis* and *Bulbostylis* have been variously treated as separated genera or united in various combinations (Table 1.2; Bruhl 1995). Sometimes they have all been combined under *Fimbristylis* (e.g. Bentham 1878; Koyama 1985). More often, *Abildgaardia* has been placed in *Fimbristylis*, while *Bulbostylis* has been considered a separate genus (e.g. Clarke 1902; Wilson 1983).

Vahl (1805), the founder of *Fimbristylis*, segregated from *Scirpus* just those species that have spirally arranged glumes, as well as a flat, bifid, ciliate, deciduous style with enlarged base. For the species with a similar flower structure but distichously arranged glumes, he created the genus *Abildgaardia*. The tristigmatic species were left in the genus *Scirpus* (Kern 1974).

All genera except *Nemum* have a distinct and thickened or slightly thickened style base. In *Nelmesia* and most species of *Bulbostylis*, the style base is persistent. Typically, *Fimbristylis* comprises species that have a deciduous style base and the orifice of the leaf sheath without long hairs (Goetghebeur 1998). *Bulbostylis*, *Abildgaardia* and *Crosslandia* have been segregated from *Fimbristylis* based on characters such as duration of style base upon the fruit, embryo type, fruit morphology in general and fruit wall ornamentation and cell pattern in particular, morphology of leaf sheaths, the type of spikelet compression, and less significantly (due to the plasticity of this character in some species of theses genera) the inflorescence (Kunth 1837; Van der Veken 1965; Gordon-Gray 1971; Lye 1971; Goetghebeur and Coudijzer 1984, 1985; Goetghebeur 1986; Bruhl 1995; Goetghebeur 1998).

Plants in *Crosslandia* have unisexual flowers, whereas plants in the other genera have bisexual flowers. *Nelmesia* is characterised among the members of Abildgaardieae as having subdistichous leaves and no primary bracts (or probably similar to glumes).

In the late 19th century Australian species of *Abildgaardia* were included in *Fimbristylis* (Bentham 1878) because of the lack of enough information to completely recognise the boundaries of these two genera and some overlapping of features in the two genera. Kern (1974) considered *Bulbostylis* to be morphologically clearly circumscribed but treated *Abildgaardia* as a section of *Fimbristylis* because circumscribing *Abildgaardia* and *Fimbristylis* is difficult. This difficulty is particularly due to the existence of species resembling in the arrangement of the glumes in *Abildgaardia* and (section *Fuscae* of) *Fimbristylis* and indistinctness of their boundaries as a result of the lack of this arrangement in some of their other species. Lye (1973) and Haines and Lye (1983) included all the *Bulbostylis* species (with Bulbostylis-type embryo and distichous glumes) in *Abildgaardia.*

Koyama (1961) did not distinguish between *Bulbostylis* and *Fimbristylis* and sank the former in the latter because of the existence of intermediates (*F. hispidula, B. pilosa*) that exhibited characteristics of both. He located *Fimbristylis* in tribe Scirpeae, but Goetghebeur (1986, 1998) and Bruhl (1995) considered *Bulbostylis* to be a member of Abildgaardieae and close to *Fimbristylis*. Bentham (1878) united *Bulbostylis* and *Fimbristylis*.

Van der Veken (1965), who studied the embryos of *Fimbristylis* and *Bulbostylis*, found that except for a few species of *Fimbristylis*, the embryos are of different types in the two genera, thus providing evidence to keep *Bulbostylis* separate from *Fimbristylis*. In a few species of *Fimbristylis* (e.g. *F. hispidula*), the embryo type is intermediate between the Bulbostylis- and Fimbristylis-types (Kern 1974).

The sister group relationships within Abildgaardieae were unresolved in Bruhl's (1995) and Goetghebeur's (1986) investigations. Bruhl (1995) found that *Crosslandia* and *Tylocarya* make a clade and their sister group is *Fimbristylis*; these three genera with *Abildgaardia* made a clade that with *Bulbostylis* formed a monophyletic group.

Fimbristylis is morphologically heterogeneous and numerous sections are recognised (Kern 1974). Latz (1990), when describing nine new Australian species in *Fimbristylis,*called for much more research on this poorly collected and diverse genus. It presents various taxonomic challenges, and no overall agreement about infrageneric classification of it has been reached to date.

Up until now, just a few evolutionary studies have been carried out on *Fimbristylis*; none of them provides an extensive evolutionary reconstruction for this genus (Goetghebeur 1986, 1998; Bruhl 1995; Muasya *et al.* 1998, 2000). Bentham (1878) divided *Fimbristylis* into five sections with one of them made up of four series. Current classifications follow a system suggested by Kern (1974), in principle. Kern classified *Fimbristylis* into eighteen sections founded on spikelet and nut characteristics but he recognised that this kind of classification may not be a reflection of evolutionary relationships (Kern 1974: p. 543). For example section *Leptocladae* Ohwi, typifically with long-ciliate glumes, may be an artificial section as indicated by Kern (1955), who expressed his concern that inserting *F. recta* in *Leptocladae* would be highly debatable due to the lack of thorough knowledge of all species of the genus. He had similar concerns about the section *Neodichelostylis* Camus. Sections *Signatae* Kern, *Mischospora* (Boeck.) Camus, and *Dipsaceae* Ohwi (1– few species each), are characterised by unusual ornamentation of the surface of their nuts, although the hypothesis that this is significant phylogenetically has not been tested cladistically.

1.2. Value of non-molecular data

Previous classifications were based on non-molecular characters, mainly morphology, but also anatomy, cytology, embryology, and physiology. Many characters are not known or difficult to assess across groups such as the genera of Cyperaceae; for example, in those taxa with capitate or spike-like inflorescences, the inflorescence prophylls might be indistinguishable from spikelet prophylls (Bruhl 1991, 1995). Other characters (e.g. style form: terete or flattened) may be somewhat difficult to interpret consistently because of the difference between fresh and herbarium material (Bruhl 1995). On the other hand some morphological and anatomical characters are fairly reliable, such as some features of the epidermis, types of silica bodies and anatomical characters related to photosynthetic pathways (Bruhl *et al.* 1987; Bruhl 1995; Bruhl and Perry ˙995). When morphology, anatomy, cytology, embryology, physiology, and chemistry of secondary products do not provide a good basis for distinguishing the limits of taxa, molecular data may provide a solution (Stevenson *et al.* 2000).

1.2.1 Suitability of pollen grains as a source of characters

There are advantages in using pollen grains as a valuable source of characters in systematic studies. Pollen grains are easy to collect and can be easily stored for at least several months (Ravikumar 1984). Srikanth (1981) in a detailed phylogenetic study of exine and aperture in the *Waltheria indica* complex, also correlated with the stigmatic papillae, was able to solve the taxonomic problem in the *W. indica* complex. Hence, it is clear that pollen grain characters are informative in biosystematics and phylogenetic studies, even at population and variety level (Walker and Doyle 1975; Nair and Ravi Kumar 1984; Harley and Zavada 2000).

1.2.1.1 Pollen grains in Cyperaceae

Several authors have categorised the pollen grains of Cyperaceae based on their shape, size and most of all the presence/absence and number of apertures (Van Wichelen *et al.* 1999). Up till now only fragmentary records on the general features of pollen grains of Cyperaceae have been published (see below for details). These studies have been often based on few taxa and are inadequate to determine whether pollen grain morphological characters might be valuable in understanding the taxonomy of Cyperaceae.

The ontogeny of Cyperaceae pollen grains has been studied by Shah (1962, 1967), Dunbar (1973), Strandhede (1973) and Makde (1982). With regard to morphogenetics, cyperaceous pollen grains should be described, as far as is known, as tetrads, with regard to morphology, however, they might better be described as pseudomonads (Selling 1947) or cryptotetrads (Erdtman 1952) because they result from a tetrad in which three of the four grains fail to develop. Moar (1993) named these types of tetrads as S-type tetrads, a tetrad in which only one member is fully developed. Another example for S-type tetrad is *Leucopogon fasciculatus* (Epacridaceae) (Moar 1993).

Cyperaceae and Juncaceae have these monad-like tetrads with no dividers between their four sections. One nucleus is usually developed, but the residual three are at the bottom of the pollen grains and degenerate, sometimes (for example in *Cladium mariscus*) appearing as a narrow sac-like projecting bulge (Erdtman 1969). Chanda and Ghosh (1978) believe the pollen of Cyperaceae arose from the Juncaceous model. Moreover, the pollen wall in Juncaceae, like that in Cyperaceae, shows simple stratification, which is faintly ornamented (Erdtman 1952; Nair 1970).

The works undertaken by Wodehouse (1935), Erdtman (1948, 1952), Kuprianova (1948), Cranwell (1952), and Ikuse (1956) contain data about pollen of many species of sedge. Sharma (1967), Padhye (1967), and Tiwari (1970) studied some Indian sedge s. Padhye (1966-67) described pollen grains in three Kyllinga species. Van Wichelen *et al.* (1999) studied some characters in Cyperaceae pollen grains using light microscopy (LM) and scanning electron microscopy (SEM) methods but they only used two species of Abildgaardieae: *Bulbostylis megastachys* and *Fimbristylis complanata*. Their study is a valuable reference in terms of comparing different preparation methods of the pollen grains in Cyperaceae both in SEM and LM. Generally their suggested pollen preparation methods are the principal methods used in this study, although they have been modified a little.

Koyama (1969b) found that the condition of the openings as well as exine ornamentation could provide a great deal of information for delimitation of infrafamilial taxa in the family. This is still considered to be correct way to describe Cyperaceae pollen as Padhye and Makde (1980) stated that the exine is slightly ornamented, being foveolate in most genera of Cyperaceae and occasionally, reticulate, e.g. in *Fimbristylis*, *Lipocarpha*, and *Eleocharis*, which may represent a more developed pattern in Cyperaceae Position and number of apertures seem to be very valuable too in the latter study suggesting that Koyama's selection criteria for pollen characters have been right. Harley and Zavada (2000) were silent about exine ornamentation at infrafamilial level but they did confirm that aperture characters such as aperture type, apertutre number, aperture position, aperture membrane, and aperture margin are significant characters in phylogenetic analysis of monocotyledons.

Koyama (1961) categorized three pollen grain types within Cyperaceae:

1. Prolate, with 1+3 if not 1+6 openings (the majority of Cyperaceae have such type);
2. Spheroidal, with lots of pores (genus *Machaerina*);
3. Spheroidal, inaperturate (representatives of the genus *Hypolytrum*).

Padhye and Makde (1980) recognised two types: the *Cyperus*-type with one distal colpus and the *Carex*-type as defined by Erdtman (1952). The *Cyperus*-type is found amongst members of *Bulbostylis*, *Courtoisina*, *Cyperus* p.p., *Eleocharis*, *Fimbristylis*, *Rhynchospora*, and *Scirpus* s.l.. The *Carex*-type was found in *Cyperus imbricatus* as well as members of *Pycreus* and *Lipocarpha*. The *Cyperus*-type seems to be arisen from the *Carex*-type by removal of side apertures, and hence is considered more evolved.

Dahlgren and Clifford (1982) stated that the standard pollen type in Cyperaceae is ulcerate and that some of the pollen grains also have three lateral pores or furrows.

Fernandez (1987) as quoted by van Wichelen *et al.* (1999) identified 6 forms of pollen grains among the Cyperaceae but her topology was based on 19 species only. The following forms were named:

4. *Cyperus longus*-type: pantoaperturate including 1 distal pore and 4-6 colpi;
5. *Cyperus michelianus*-type: pantoaperturate with several pores;
6. *Cladium mariscus*-type: inaperturate;
7. *Schoenus nigricans*-type: pantoapertureate with 1 distal pore together with 4-6 equatorial colpi but with dimensions different from the foremost type;
8. *Carex flacca*-type: pantoaperturate with 1 distal pore and 4-5 colpi;
9. *Carex hallerana*-type: pantoaperturate with 1 distal pore and 4-5 equatorial pores.

Bruhl (1995) recognised two major kinds of pollen grain types: one with few (<6) apertures (some 20 genera) and the other with several (>6) apertures (only in *Baumea*, *Machaerina*, and *Tricostularia*). Van Wichelen *et al.* (1999), who investigated the pollen of 30 Cyperaceae species using scanning electron and light microscopy, described pollen apertures in Cyperaceae as colpi, a distal ulcus, and pores. Only two species of Abildgaardieae were included in that study, namely

Fimbristylis complanata (with no apertures) and *Bulbostylis megastachys* (1 distal ulcus + 4–6 lateral colpi).

Erdtman (1952) described the pollen of Cyperaceae as having 1–4 apertures, elongate or ± spheroidal. Pollen in *Lepironia mucronata* (Fig. 1.1) has an ulceroid opening at its thick side and three lateral poroid or elongate apertures (or aperturoids, occasionally one of these is as big as the terminal ulceroid aperture). This is a common aperture type in the family. *Calyptrocarya glomerulata* has spheroidal grains with four faint apertures. Erdtman (1952) noticed that *Mapania* and its relatives have morphologically different pollen grains from other members of the family having a distinct ulcerate aperture. The grains of *Hypolytrum schraderianum* and *Mapania amphivaginata* are also spheroidal though with only one such marked aperture. In *Mapania humilis* flat triangular pollen with poroid apertures at the two ends of the triangle have been occasionally found (Erdtman 1952).

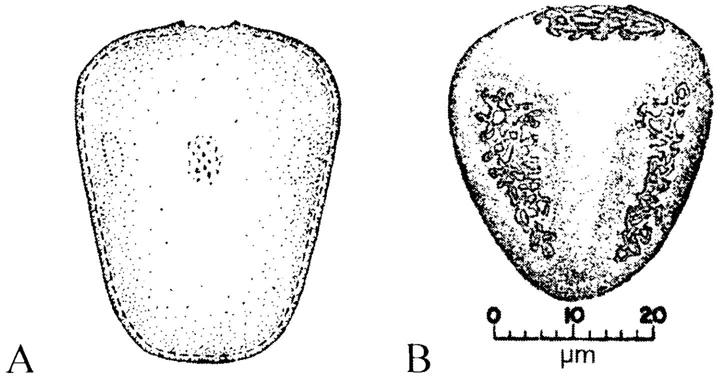

Fig. 1.1. Pollen grains of *Lepironia mucronata* (A) and *Carex pringlei* (B). These are the commonest pollen shapes in Cyperaceae (modified from Erdtman 1952 (A) and Davis 2000 (B)).

1.2.2. Embryomorphology of Abildgaardieae

Embryological studies (Van der Veken 1965; Goetghebeur 1986) indicated that *Fimbristylis*, *Crosslandia* (both with Fimbristylis-type embryo; with basal coleoptile and lateral coleorhiza), *Bulbostylis* (Bulbostylis-type; basal coleoptile and coleorhiza), and *Abildgaardia* (Abildgaardia-type; similar to Bulbostylis-type with very minor differences including very well-developed second leaf primordium and well developed third leaf primordium), are grouped in the same category (fimbristyloid-type). *Arthrostylis* and *Actinoschoenus* and their allies have a very different kind of embryo (Carex-type). Embryo-types in this group are being studied further by Kerri Clarke (PhD student, UNE; pers. comm.).

1.3. Molecular systematic studies of Abildgaardieae

The difficulties in circumscribing the tribe Abildgaardieae make it necessary to look for other data sources (e.g. molecular information) to find further information concerning relationships. There have been only a few DNA studies in the tribe Abildgaardieae. Muasya *et al.* (1998) in a study of suprageneric relationships suggested that a number of genera of the tribe Abildgaardieae might be rooted in a different way from the rest of the tribe (Fig. 1.2).

1.3.1 Relationships of Abildgaardieae

Muasya *et al.* (1998) suggested that tribes Abildgaardieae, Cariceae, Rhynchosporeae and Scirpeae together with *Hellmuthia* make a clade with low bootstrap and jackknife values (50 and 56 respectively). A recent molecular phylogenetic analysis of Cyperaceae using the data from *rbc*L (the chloroplast gene encoding the large subunit of ribulose-1,5-biphosphate carboxylase) sequences (Muasya *et al.* 2000a) has greatly improved the knowledge of limits and relationships within Abildgaardieae. This study did not confirm the previous hypothesis gained from morphological data that Abildgaardieae and Arthrostylideae do not constitute a monophyletic group. However, some uncertainties still continue.

Muasya *et al.* (2000a) sampled only 4 species from 4 of the 7 genera of Abildgaardieae using *rbc*L whereas the total number of species in this tribe is about 400 species. These sedges were analysed for separate and combined DNA and morphological data of the genera of Cyperaceae. Muasya *et al.* (2000a) suggested that the homologies of subfamilies should be reviewed again in comparison to phylogeny using molecular data.

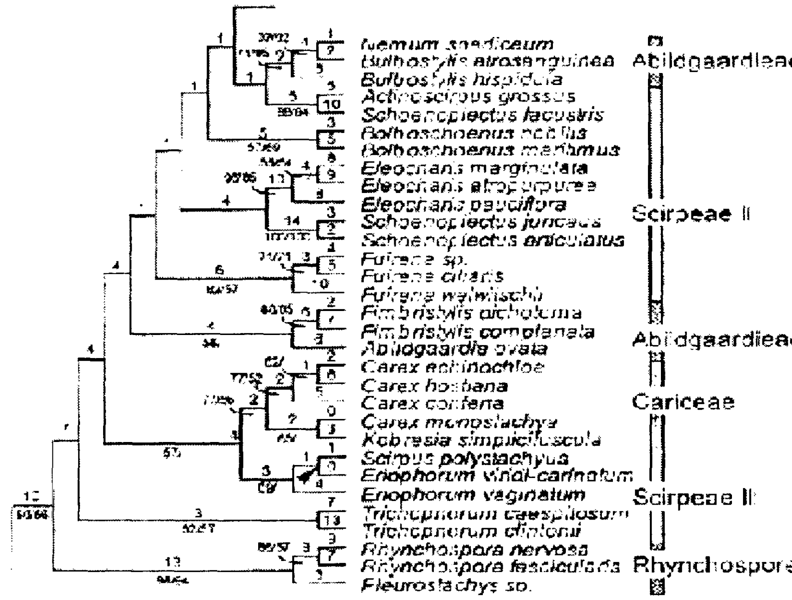

Fig. 1.2. Six diverging groups of the family Cyperaceae starting with Rhynchosporeae headed for Abildgaardieae I inferred from *rbc*L gene sequence data. Branch lengths (above the branches) together with bootstrap/jackknife values (below the branches) shown (Muasya *et al.* 1998).

1.3.1.1. Relationships of Abildgaardieae inferred from separate molecular and morphological data

The results of *rbc*L data (Muasya *et al.* 1998) and morphology (Bruhl 1995) individually indicated that *Abildgaardia* and *Fimbristylis* were sister groups while *Bulbostylis* is either further away from (Muasya *et al.* 1998) or less closely related to (Bruhl 1995) these two genera. These genera in turn are sister groups in morphological studies. Morphological analysis suggested that *Eleocharis* was basal and closely related to Abildgaardieae. *rbc*L data, however, introduced the genus *Fuirena* as the basal genus for the Abildgaardieae sister groups. The place of *Eleocharis* was not the same in the two analyses. *Arthrostylis* has made a clade with *Eleocharis*, which was not as close as other genera to the genera of the tribe Abildgaardieae in *rbc*L analysis (Muasya *et al.* 2000a). On the other hand, other studies did not show strong support for *Bulbostylis* (Abildgaardieae) as the sister lineage to *Eleocharis*.

The results obtained from ITS data by Roalson and Friar (2000) using the species of *Eleocharis*, *Bulbostylis*, *Scirpus* and *Fuirena* were partly congruent with previous studies provided by Goetghebeur (1986, 1998), Bruhl (1995), and Muasya *et al.* (2000a) on relationships between Scirpeae and Abildgaardieae. Roalson and Friar (2000) in this study of intrageneric relationships on *Eleocharis* suggested *Bulbostylis* (as one of the genera of Abildgaardieae) to be the most closely related lineage to *Eleocharis* of the lineages included in their study (particularly the species of *Scirpus*), but this is weakly supported with a 52% bootstrap. They only sampled two species of one genus of five genera in Abildgaardieae, thus intra- and intertribal relationships for Abildgaardieae could not be explained in this study.

1.3.1.2. Relationships of Abildgaardieae inferred from combined data

The main findings of the analysis of combined morphological and molecular data were (Muasya *et al.* 2000a) (Fig. 1.3):

1. There is no doubt about the close phylogenetic relationships among the genera of Abildgaardieae and Arthrostylideae.
2. The genera of the tribes Abildgaardieae and Arthrostylideae were sister to a complex of genera including those in Cypereae, two different groups of Scirpeae and some genera of Hypolytreae.
3. A very close relationship was found between Scirpeae and the Abildgaardieae-Arthrostylideae clade. In other words, this clade has evolved from within the tribe Scirpeae. This supported Bruhl (1995), who suggested a close relationship between Abildgaardieae and Arthrostylideae rather than Abildgaardieae and Scirpeae.
4. There was moderate to uncertain support for the situation of the Abildgaardieae-Arthrostylideae clade (76%) compared to most of the other clades of Cyperaceae. Therefore, this clade merits further attention.
5. The *rbc*L sequence confirmed earlier views (e.g. Bruhl 1990, 1995) about the close relationship between *Arthrostylis* (one of the genera of Arthrostylideae) and Abildgaardieae. Within the Arthrostylideae-Abildgaardieae lineage, Arthrostylideae is sister to Abildgaardieae (17 nucleotide substitutions, reliability 76%); Abildgaardieae, in turn, is also a well-supported clade (13 nucleotide substitutions, 91% bootstrap value). Using computer simulations and a laboratory-generated phylogeny, Hillis and Bull (1993) demonstrated that, for a range of states, clades with a

bootstrap value of more than 70% have a high probability of being real monophyletic groups. Although the evidence for relation with Abildgaardieae is informative, nevertheless assessment of the limits of Arthrostylideae are debatable because only one species of one genus out of four genera in this tribe has been sampled and this species has placed in very different situations using *rbc*L, morphology, and combined data.

6. The phylogenetic relationship between *Nemum* and *Bulbostylis* is strong with 100% bootstrap value and thirteen synapomorphies. The number of characters that show synapomorphy between the *Nemum-Bulbostylis* clade and *Fimbristylis* was seven. More data are needed to confirm the relationship among these three genera. However, the bootstrap value in another study (Muasya *et al.* 2000b) for a clade including *Abildgaardia* and *Fimbristylis* based on *rbc*L and *trnL*–*trn*F (*trnL*–F) data was 100%.
7. *Arthrostylis* is basal and related to *Abildgaardia*, *Fimbristylis* species diverged after *Abildgaardia*; and *Bulbostylis-Nemum* clade was terminal. Therefore, four genera of Abildgaardieae (*Nemum*, *Bulbostylis*, *Fimbristylis*, and *Abildgaardia*) form a monophyletic group, and these form the sister group of Arthrostylideae. Therefore the results strongly support the monophyly of Abildgaardieae but only moderately to weakly support the monophyly of Abildgaardieae plus *Arthrostylis*. In summary these taxa show the following relationships: (*Arthrostylis* (*Abildgaardia* (*Fimbristylis* (*Bulbostylis*, *Nemum*)))).

In another recent study (Muasya *et al.* 2000b) with emphasis on subfamily Cyperoideae (*sensu* Bruhl 1995) and based on the parsimony algorithm of PAUP* (Swofford 2000), *Fimbristylis* and *Abildgaardia* make a clade with 100% bootstrap value but only one species of each genus has been sampled. Meanwhile, other genera of the tribe have not been studied in this work.

The combined data in the study of Cyperaceae by Muasya *et al.* (2000a) have shown a medium level of homoplasy between morphological and DNA data and the topology of the tree resulted from combined data has been similar to that related to DNA data. However, relationships among the genera of Abildgaardieae and between Abildgaardieae and Arthrostylideae remained unresolved by *rbc*L sequence data due to the low level of sequence changes in this gene.

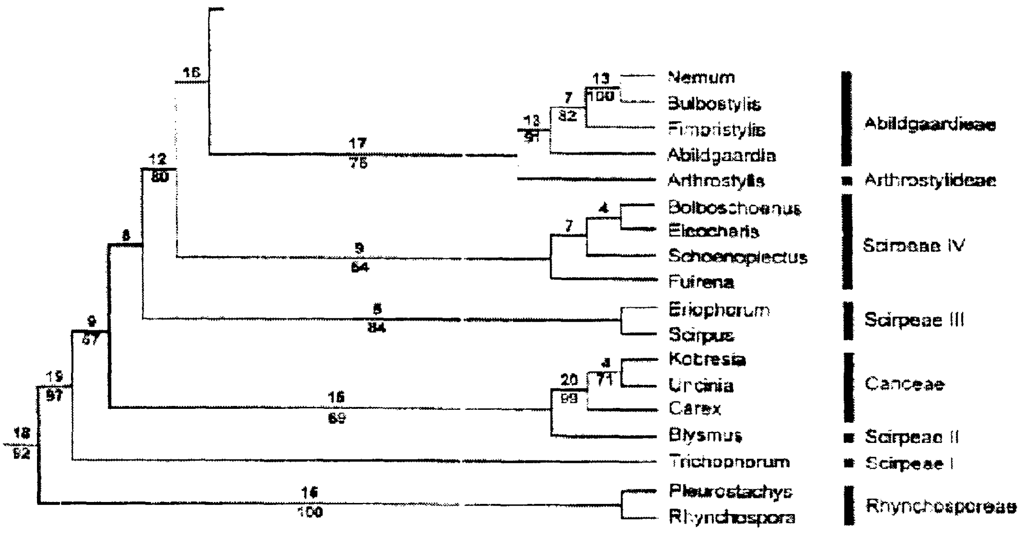

Fig. 1.3. Part of a tree derived from combined *rbc*L and morphological data using successively weighted characters showing tribes Rhynchosporeae to Abildgaardieae. Numbers above the branches indicate branch lengths. Bootstrap percentages are shown below the branches (Muasya *et al.* 2000a).

1.4. Aims of this study

It is clear that additional data are needed to resolve the relationships within those clades identified in Scirpeae, Abildgaardieae, Arthrostylideae, Schoeneae, and Cypereae by *rbc*L. This present study aims to explore phylogenetic relationships and taxonomic limits of Abildgaardieae and close tribes, and the phylogenetic positions of the taxa that have been grouped in Abildgaardieae.

Specific aims are:

1. To resolve (using pollen morphological characters and DNA sequences) weakly supported relationships that have been recovered in non-molecular analyses by previous researchers.
2. To test the taxonomic value of pollen characters in the tribes Abildgaardieae, Arthrostylideae and their allies; and identification of the level at which these characters are informative.
3. To evaluate conflicts between previous molecular and non-molecular studies by studying molecular (DNA) sequence data of *trn*L–F IGS [the intergenic spacer between tRNA Leucine (UAA) and tRNA Phenylalanine genes] and *trn*L intron (the intron that the tRNA Leucine contains) of cpDNA, and ITS (internal transcribed spacer) of nrDNA. These two regions seemed suitable for

phylogenetic assessment at the tribe level in Cyperaceae (see discussion in Chapter 2).

4. To test the outcomes of previous molecular research based on *rbc*L gene sequences claiming the monophyly of Abildgaardieae and sister relationships between this tribe and Arthrostylideae.

Tribal monophyly and limits of the Abildgaardieae will be assessed using a wide range of specimens. Members of the close tribes are used as outgroups in cladistic analysis. As there are different classifications of the genera of Arthrostylideae, I also investigate whether tribe Arthrostylideae can be considered monophyletic. This work will compare directly chloroplast *trnL*–F and ITS ribosomal DNA data at the tribal level in Cyperaceae. This study will include representatives of one genus that has not been previously studied in molecular analyses, i.e. *Crosslandia.*

In this study, I will focus on the generic level relationships, studying material of almost all genera and a range of species covering most sections in the main genus *Fimbristylis*. I will expand the range of collections from the contemporaneous morphological and anatomical study of K. Clarke (Ph.D. project, UNE) using molecular and pollen data to assess the phylogenetic relationships between Abildgaardieae and closely related tribes and evaluate previous works on the morphological similarity of this group. The project has the advantage of abundance of the members of Abildgaardieae in Australia and therefore can help Australian biodiversity projects as well as having global taxonomic significance.

In summary, these are the questions that must be answered in this project:

1. Is the tribe Abildgaardieae monophyletic?
2. Are the taxa within Abildgaardieae monophyletic?
3. What is the evolutionary order of occurrence of the genera of Abildgaardieae?
4. Are the new molecular and micromorphological characters efficient in delimitation of monophyletic groups and for classification at different taxonomic levels?
5. Are the previous infrageneric classifications of the genus *Fimbristylis* unfailing?
6. Which character set gives more resolution to the relationships between Abildgaardieae and its closest tribes and within Abildgaardieae?

7. Which character set provides more support for the results inferred from previous studies?
8. What is the effect of combining two molecular data sets and molecular and micromorphological data sets on the whole resolution of the relationships at inter- and infra-tribal level in this study?

CHAPTER 2

Molecular Systematics

2.1 Introduction

2.1.1 Problems with morphological data

The morphological characters that form the basis for conventional taxonomic schemes have made problems such as parallelism or convergence for systematists, especially at higher levels of classifications. These uncertainties make the majority of morphological data ambiguous and several authorities recommend using a range of sources of data in evolutionary studies (Clegg 1993).

Different interpretations of homology are causes of inaccuracy when assessing morphological characters (Kitching *et al.* 1998). High levels of homoplasy in addition to a common lack of distinct morphological attributes indicate the high dependence on molecular characters in finding the clades, as well as confidence about their accuracy. Homoplasy is perceived widely either as noise in a phylogenetic analysis or as harmful in phylogenetic reconstruction; however, with sufficient data, homoplasy can be reasonably interpreted in a parsimony analysis by different optimisation methods such as Wagner (Farris 1970), Fitch (1971), Camin-Sokal (Camin and Sokal 1965), and generalised optimisations (Swofford and Olsen 1990).

2.1.2 Link between phenotype and genotype

For decades, a goal of many systematists has been to break the barrier of the phenotype and find characters that have a more basic nature free of the effects of the environment (Hollingsworth *et al.* 1999). Molecular techniques are worthwhile in analysing taxonomically complicated groups due to their potential of offering systematic data independent from morphology, enabling researchers to organise a large range of comparative analyses necessary for reconstructing a phylogenetic record of those groups (Yen and Olmstead 2000).

Genetic variation appears to be responsible for a lot of the natural, structural, physiological and even behavioural differences we observe among genera and species of living beings. Much of the research in modern systematics relies on the idea that data, for example nucleotide chains, are representative of the organism's original genome. The source of every genetic difference is DNA so some researchers believe that molecular data are the aptest data to address questions related to evolution (Avise 1994). By studying the differences in nucleic acids among species one learns in which way and at what time they developed (Norman and Christidis 2000), although this is not possible without fossil records (but see Hillis *et al.* 1996). Recombinant DNA technology has permitted systematists to test genotypic variation directly (Doyle 1993).

2.1.3 Phytochemistry as a data source

In the 1960s and 1970s, some secondary chemical metabolites such as phenolics, waxes, flavonoids, terpenoids, alkaloids, and related compounds, were the focus of investigations due to their closer relationship with the genome compared with morphological characters. In other families (e.g. Towers *et al.* 1966; Harborne 1969; Wofford 1974; von Rudloff 1975; McMi lan *et al.* 1975; Seaman and Mabry 1979), and later in Cyperaceae (e.g. Harborne *et al.* 1985; Manhart 1990), various kinds of these products have been used for systematic purposes.

Such compounds, however, are produced in more than one way and also they are the products of a chain of biosynthetic reactions whose qualities are not as independent from environmental factors as was previously thought (Doyle 1993). Moreover, there is not any reason for bel eving in data obtained from secondary chemical metabolites to have any bonus compared with other characters, such as the morphological, cytological, or anatomical characters. For example, the existence or lack of a particular compound does not a pear to be, a priori, more important taxonomic evidence than the presence or absence of petals. This is due to the controlling role of several to large numbers of genes on these micromolecular features, i.e. alleles at various loci affect the outcome. Thus, the strict meaning of both morphological and micromolecular differences is often unclear from a genetic standpoint (Woodland 1997).

2.1.4 Molecular sequence data for phylogenetic reconstruction

The molecular genetic revolution has generated an explosion in phylogenetic reconstruction. It has been evident that molecular sequences contain useful information about evolutionary history (Zuckerkandl and Pauling 1965; Fitch and Margoliash 1967).

The explosion in systematic informatio 1 derived from research of DNA has reformed the tradition of systematics (Olmstead and Palmer 1994), resolving problems that were unresolved by the use of morphological data (Hilu and Liang 1997). Recent technical progress makes the study of nucleic acids a quite routine procedure. This has led to a considerable increase in using DNA approaches for comparative evolutionary, ecological and behavioural studies (Norman and Christidis 2000).

Evolution proceeds by the increase of DNA variation, therefore, taxa that are strongly related have homologous DNA arrangements and taxa that are distant will show quite dissimilar DNA arrangements. This simple relationship between the divergence of sequences and the period s nce two species shared a common ancestor enables us to reconstruct a model of phyl ogenetic relationships between organisms.

Phylogenetic studies use two major types of nucleic acid data: differences in gene content or order (generally in cpDNA in plants; Downie and Palmer 1992) and nucleotide substitutions. Chloroplast DNA sequence study is a useful tool to estimate evolutionary relationships among plants at higher levels (Palmer 1987; Palmer *et al.* 1988; Clegg *et al.* 1991a; Clegg *et al.* 1991b).

2.1.5 Proteins

Proteins are more closely tied to the genotype than micromolecules or secondary chemical metabolites are, and electropho etically assayed variation of isozymes is used efficiently to assess the relationships among individuals, populations, and other closely related taxa (Doyle 1993). Neither of these approaches had been as efficient on phylogenetic reconstruction as nucleic acid analysis approaches because the latter directly deals with the genetic material, despite some massive efforts in protein sequencing (Martin and Dawd 1991).

2.2 Techniques

2.2.1 PCR technique

Much effort has been made in recent years to analyse DNA data to get the best estimate of phylogeny. The past ten years have witnessed a considerable change from very primitive, long, slow, and less informative methods to shorter, faster, and more informative methods of DNA analyses in biological studies. This change has been brought about as a result of the development of the polymerase chain reaction (PCR) technique (Mullis and Faloona 1987). PCR is a technique for biochemically amplifying (making millions of new copies of) a specific gene sequence. Karry Mullis, an American, developed PCR in 1987, for easing genetic studies. The technique was a remarkable breakthrough, allowing molecular studies that were previously very time-consuming, laborious, and expensive, to be accomplished with relative ease and at lower cost.

Since then, there has been a real explosion in molecular studies on plant systematics at different levels, from algae to flowering plants (Doyle 1993). PCR is a technically easy procedure that synthetically generates nearly one million copies of a specified short DNA sequence from the genomes of the organisms being analysed. This amplification procedure is crucial as single cells include minute amounts of genes that cannot be detected using standard analytic methods (Norman and Christidis 2000).

The most important characteristic of the polymerase chain reaction is its ability to amplify very specific fragments of nucleic acids (Mullis and Faloona 1987). Many articles on the subject are nowadays being published, although these are just the earliest steps (Doyle 1993). Regions targeted by PCR for DNA sequencing are usually 300–1000 bases in length, and thus constitute a very small proportion of the organism's total DNA.

2.2.2 DNA sequencing

2.2.2.1 Transition, transversion and codons

A nucleotide substitution from one purine to another purine (e.g. A → G), or from one pyrimidine to another pyrimidine (e.g. T → C) is called transition. Transversion, on the other hand, is defined as a nucleotide substitution from a purine to a pyrimidine (e.g. A → C), or vice versa (e.g. T → G). In protein coding regions of DNA/RNA each triplet of nucleotides plays a role in the composition of a protein molecule that will be constructed based on the sequences of these 'codons'. Each amino acid (structural unit of proteins) is coded or recognised by one or more three-nucleotide codons because there are only 20 amino acids, which are coded by 61 out of 64 possible combinations of three-nucleotides or codons. The other three combinations do not code any amino acid.

2.2.2.2 Character weighting

Methods to make corrections for consequences of variation in base composition among taxa on phylogenetic analysis also have been developed (e.g. Sidow and Wilson 1991; Steel *et al.* 1993; Lockhart *et al.* 1994; Lake 1994). Knowledge about different types of relations among sequence positions (Wheeler and Honeycutt 1988; Korber *et al.* 1993) and differences in probabilities of change across sites has led to objective criteria for differential character weighting. This is because of the findings of Zurawski *et al.* (1984) and Zurawski and Clegg (1984) who showed that most nucleotide substitutions occur as silent changes in the third position of codons. Therefore, weighting by character (e.g. codon position) as well as character state conversion (like transitions and transversions) have been most common in DNA sequence studies (e.g. Lake 1987; Albert *et al.* 1993).

2.2.2.3 Sequence variation

Sequence variation in the coding and spacer regions of a number of chloroplast, mitochondrial, and nuclear DNA sequences has been widely used to find the best regions on the genome that can provide more reliable information about phylogenetic relationships (Hilu and Liang 1997). More rapidly evolving cpDNA genes, introns, and intergenic spacers, and the non-coding portions of nuclear ribosomal RNA have been useful for phylogeny reconstruction of taxa at lower levels (Olmstead and Palmer 1994).

Studies of DNA that deal with phylogenetic uncertainty across an evolutionary period among organisms must include more conservative DNA sequences (Hillis and Davis 1986) because sequences with rapid divergence are not informative on deep phylogeny. The latter sequences are useful tools for studying evolutionary matters on the rank of infrageneric taxa and closely related groups.

Rates of nucleotide substitution do not appear to be fast and hence provide a suitable window to find the phylogenetic relationships of the plants at deep (i.e. old) levels of evolution (Clegg *et al.* 1994). Furthermore, structural rearrangements are common, with many inversions or deletions distinguished in flowering plants (Olmstead and Palmer 1994).

Inversions or large deletions are rare events that might be unlikely to occur independently in the chloroplast genomes of unrelated taxa (Doyle 1993). Inversions in cpDNA provided evidence applicable to long-standing taxonomic questions such as the relationships of g asses to other monocots (Doyle *et al.* 1992). Structural mutations of the cpDNA can, however, provide misleading information for phylogenetic reconstruction (Doyle 1993). For example, the inverted repeat is absent from the chlor plast genomes of conifers (Strauss *et al.* 1988; Raubeson and Jansen 1992) as well as some members of the angiosperm family Fabaceae. Conifers and Fabaceae are very widely unrelated phylogenetically that there is little question that the inverted repeat loss is independent in the two groups, and in any case detailed studies have shown that it is not the same copy of the inverted repeat that is lost (Doyle 1993).

The rare transfer of a gene from the organellar genomes to the nucleus may be a good evolutionary character, although the lack of a gene might be unreliable in phylogenetic studies. Certainly, not all plant groups possess cpDNA variation between their subgroups. Even in those that do, the existence of an uncommon mutation does not show the relationships within the group having the mutation, nor among the group and other groups that do not have the mutation (Doyle 1993). A number of other genes are useful for comparisons at the lower phylogenetic levels.

The likelihood of transition is more than transversion because the latter needs more structural changes and therefore more biochemical reactions than the former. For instance in mtDNA, a bias toward transitions can be extreme. Protein-coding genes as well as the control segment of mitochondrial DNA may accumulate transitions 10 or more times more quickly than transversions in some species (Brown *et al.* 1982; Irwin *et al.* 1991; Kocher and Wilson 1991). To complicate concerns further, the proportional probabilities of changeovers between a specific pair of nucleotides (for example, A ↔ G transitions) would be asymmetric, resulting in biased base composition in the sequences compared. Correction for such inequalities by weighted estimators of sequence diversity increases the limitation of consistency in assessments of evolutionary relationships from nucleic acid sequence characters (Felsenstein 1988; Huelsenbeck 1995).

The rate of transversion substitutions are likely to be affected by various factors, intrinsic genetic or external environmental. If it is true, then transversions, and subsequently nr/nv (transition/transversion) values, might not represent particularly conserved characters for phylogenetic work, but are rather a product of an intrinsic nucleotide composition pattern that characterises a lineage (Hilu and Liang 1997). This was found for most of the lineages of the grass family (Liang and Hilu 1996).

2.2.2.4. Advantages

Perhaps the greatest advantage of DNA alignment studies lies in the range of phylogenetic depth to which sequence data can be applied. The potentially great number of characters presented in DNA sequencing data is known as the most valuable advantage in phylogenetic analyses (Miyamoto and Cracraft 1991; Donoghue and Sanderson 1992). One last and significant advantage for sequencing studies is the possibility of addition of other taxa to the existing records simply through entering an aligned order of bases (e.g. Chase *et al.* 1993; Chase *et al.* 2000).

2.2.3 Analysing DNA

Comparison of DNA sequences is the area of molecular systematics in which huge progress is being seen. DNA sequencing is believed to be the most sensitive method for assessing genetic variation. It has long been clear that DNA sequences include useful data about evolutionary history. These sequences provide information on the order of nucleotides (G, C, T, and A) in a nucleic acid fragment and permit the exact nature or location of every mutation to be determined. In theory, any degree of divergence appears to be amenable for comparative sequencing studies (Olmstead and Palmer 1994).

Inversions or significant deletions have been considered reliable markers of relationship when shared by two or more genera. Such characters are able to provide key insights into the evolution of living beings. The case of a 22-kilobase inversion whose distribution has suggested the identity of the basal genera in the largest flowering plant family, the Asteraceae (Jansen and Palmer 1987a) is a good example of the use of these characters. Similarly, cpDNA inversions provide data of significance to old taxonomic challenges such as the links between grasses and other plants (Doyle *et al.* 1992) or the identity of the earliest-diverging lineage in vascular plant evolution (Raubeson and Jansen 1992).

The utility of these mutations is supported by the good connection between their distribution and traditional taxonomic groupings. For example, a nearly 20-kilobase deletion has removed one entire copy of the inverted repeat from a large number of genera in several tribes of the legume family; the distribution of the deletion agrees well with phylogenetic concepts based on such conventional characters as morphology, biogeography, chemistry, and chromosome number (Lavin *et al.* 1990).

Additionally, singular structural rearrangements (e.g. inversions and intron losses) in the chloroplast genome have been investigated to identify monophyletic groups. However, the occurrence of these rearrangements is too rare to be considered as a useful character for reconstructing the phylogeny of the lower levels such as inter- and intra-species levels (Jansen and Palmer 1987; Bruneu *et al.* 1990; Lavin *et al.* 1990; Olmstead *et al.* 1990; Downie *et al.* 1991: Raubeson and Jansen 1992).

2.2.3.1 Genes and gene trees

A gene is composed of many nucleotides; each of which carries data linking the gene to the species of origin, not just to the species in which it is found now. It is rarely stated explicitly that a terminal taxon on trees obtained from molecular data shows a gene-, molecule-, or organelle-genome- (not a taxon) phylogeny. However, the results are used to describe the phylogenetic relationships among the taxa not their molecules or genes (Doyle 1992, 1993; Avise 1994). Assuming that the genes assessed are genuinely homologous, gene trees and organismal phylogenies can differ because of keeping the ancestral diversity, or interrelationships among populations and/or species. This is a serious concern for chloroplast and mitochondrial DNA because the consequences of these complex gene flows or hybridisations are possibly retained by later generations (Doyle 1992; Degnan 1993).

Phylogeny is a representation of the diverging history of the paths of heredity of living beings. The paths of heredity represent the route of genes through different generations, and the pattern of branching is a gene tree. But different genes may have different histories, or different paths of heredity. When we track down the history of various genes in different species, we find the paths of heredity for the species, and, in fact, we draw a phylogenetic tree for the species, which is called the species tree, representing the evolutionary relationships among the species (Nei and Kumar 2000).

Hybridisation among species that have been separated for long time ago can result in the introduction of a maternally inherited gene and, therefore, discrepancies of gene and species phylogenies (Avise 1994). However, if phylogeny is assessed in several ways, these inconsistencies can give insights into the phylogenetic history of a species and the effect of hybridisation in this history (e.g. Arnold *et al.* 1991; Whittemore ar d Schall 1991; Dowling and De Mrais 1993).

In a species tree, the bifurcations correspond to the moment of the occurrence of a new species, i.e., the moment when two species separated from each other. The gene tree differs from the species tree in two main points. First, the difference of two genes from two species may have occurred before the two species split (Fig. 2.1). The second problem with gene tre:s is the difference between the topology of the gene tree may be different from the branching pattern of the species tree. The reason of this, is genetic polymorphism in the ancestral species (Nei and Kumar 2000).

Differences in gene trees of different populations and closely correlated species would happen purely due to lineage sorting effects (e.g., Hey and Kilman 1993; Slade *et al.* 1994). These concerns can be overcome by combining data across a large number of loci (Pamilo and Nei 1)88; Atchley and Fitch 1991; Slade *et al.* 1994), however there might be potential disadvantages, such as bias in favour of particular loci, in combined analysis of data from distinct gene trees (Bull *et al.* 1993; de Queiroz 1993).

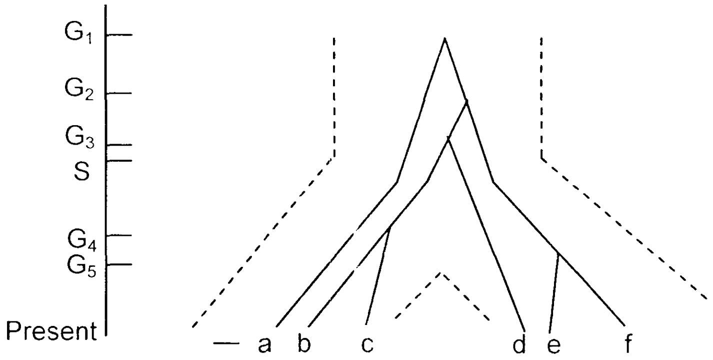

Fig. 2.1. Diagram showing gene splitting (G1– G5) may happen before or after population splitting (S) if a population is genetically polymorphic. The evolution of separation of genes resulting in six alleles (a – f) is shown in solid lines whereas broken lines show speciation. (Figure 5.6 of Nei and Kumar 2000)

2.2.4 Markers and population variation

In plants, chloroplast or mitochondrial genomes have not been variable enough at lower levels (Palmer 1987; Crawford 1989) and nuclear genome markers like

RAPDs (Random Amplification of Polymorphic DNA) that amplify DNA segments by randomly selected primers are typically used to study populations (e.g. Waycott 1998; Bussel 1999).

2.2.4.1 DNA fingerprinting

Another DNA-based method that has been widely employed in biological analyses is DNA fingerprinting or DNA profiling. A disadvantage of this technique is its need for large quantities of undamaged DNA, which might require destructive experimental methods (Norman and Christidis 2000).

2.2.4.1.1. RAPDs

RAPD markers, based on amplification of DNA by one small primer using PCR technique, result in multiple products from the whole genome (Williams *et al.* 1990). The primer amplifies a DNA fragment flanked between two primer-binding regions in reversed orientation.

Where sensitive methods such as nucleic acid sequencing or targeted polymerase chain reaction amplificat on have been preferred, RAPD PCR appears to be less appropriate than these methods for comparative studies of differences between taxa and that is why it has had limited use in biological analyses of taxa at higher levels (Norman and Christidis 2000). RAPD is not a popular marker because it is difficult to reproduce it. However, it is still an advantageous starting method for newly established laboratories due to the negligible inputs needed for preparation, the ability to use a general set of primers for all species, simplicity of the process, and the need to just a relatively small amount of DNA.

2.2.4.1.2. Microsatellites

Microsatellite technique is another ncreasingly recommended technique for studies at intraspecific levels.Simple sequence repeat (SSR) markers or microsatellites are tandem repeats of simple sequences of DNA that turn up abundantly and randomly in most of the eukaryotes (Akkaya *et al.* 1992). In particular, dinucleotide microsatellites, such as (AG)n, (AC)n, and (AT)n, are reportedly abundant and highly polymorphic in eucaryotic genomes. Howeve, this method (as RAPD) does not satisfy phylogenetic approaches due to doubts about homology of alleles detected between species (FitzSimmons *et al.* 1995; Smith *et al.* 1995). For microsatellite analysis, sometimes alleles that produce either no protein product or a non-functional protein product (under the conditions analysed) do not amplify due to large size of the resulted polynucleotides or changes in the neighbouring primer sites. Therefore, there is no universal primer even often for two closely related species. This means a lot of energy and time for designing new primers unless a primer has already been designed for the species under study. Advantages of the microsatellite technique are its speed and precision.

2.2.5. Restriction site mapping

Two approaches have provided most data from the chloroplast DNA for phylogeny reconstruction. One of these approaches is restriction site mapping of the whole chloroplast genome (Palmer *et al.* 1988; Jansen *et al.* 1991; Olmstead and Palmer 1992), and another one is sequencing of the gene *rbc*L (Doebley *et al.* 1990; Soltis *et al.* 1990; Giannasi *et al.* 1992; Olmstead *et al.* 1992).

Analysis of restriction fragment variation has been the primary method for analysing cpDNA for evolutionary purposes (Olmstead and Palmer 1994). Chloroplast DNA restriction mapping has been most popular in studies at species level, where several studies have been conducted in taxa such as ferns (e.g. Gastony *et al.* 1992) and flowering plants (Soltis *et al.* 1992a). However, the flexibility of the method has made it useful from the intraspecific level (Soltis *et al.* 1992b) to intergeneric (Jansen *et al.* 1990; Doyle and Doyle 1993) or even interfamilial levels (Downie and Palmer 1992b). Restriction endonucleases vary considerably in the number of recognition sites in chloroplast DNA; thus, both the amount and selection of specific enzymes are important considerations for systematic studies (Doyle 1993).

2.2.5.1. Limitations

According to Olmstead and Palmer (1994) several factors constrain the usefulness of restriction site mapping in phylogenetic studies:

1. The conservative nature of cpDNA evolution places a practical lower limit to phylogenetic analysis. Closely related species in many studies have been found to be identical using this method (Schilling and Jansen 1989; Wallace and Jansen 1990; Olmstead *et al.* 1990; Riesbereg *et al.* 1991).

2. At greater molecular distance an upper limit might be also reached, where restriction site homology can no longer be determined confidently. Molecular divergence may be too great to permit comparative mapping of the whole genome in families that are either old or that have accelerated rates of cpDNA evolution.
3. A substantial amount of DNA (10–100 microgram) might be needed for the many restriction enzyme digests for genomic DNA required for whole-genome mapping except in Southern Blots.
4. Scoring and handling of restriction site mapping data are largely manual and cumbersome.
5. All types of restriction site studies need relatively high molecular weight DNA (Olmstead and Palmer 1994), whereas badly degraded DNA will work as a template for PCR amplification of small fragments of DNA, because of the enormous copies that are made by the technique, thereby enabling the use of herbarium specimens and even, very rarely, fossils as the sources of DNA (Golenberg *et al.* 1990).

Insertions and deletions that are too small for detection by restriction site analysis can be identified by PCR and sequencing and used as characters in a phylogenetic analysis.

2.2.5.2. Restriction site mapping vs. sequence data

Results from Solanaceae suggested that at the species level of cpDNA divergence, neither *rbc*L nor *ndh*F sequences are more advantageous than restriction site mapping in terms of the data quantity, but *rbc*L and *ndh*F sequences are less homoplasious than restriction site mapping data (Olmstead and Sweere 1994). However, in another study on the members of Asteraceae, homology was less for the *rbc*L gene sequences than the restriction sites (Kim *et al.* 1992). Converting the raw restriction site data to characters for a cladistic analysis seems to be much more time consuming than for sequence data, but the random distribution of characters throughout the chloroplast genome may be an offsetting advantage for restriction site mapping. The risk of contamination during PCR might be extremely severe for sequencing studies, too. To reduce this risk, all supplies used in DNA extraction should be disposable, or should be treated to destroy any trace DNA that might remain from a prior extraction (an acid or bleach bath, or exposure to short wavelength UV radiation for 3-5 minutes should degrade any residual DNA) (Olmstead and Palmer 1994). However, a study in which cpDNA restriction site and sequence data have been collected shows the relative value of both (Olmstead and Palmer 1994).

2.2.5.3. Structural changes in restriction site mapping

Restriction site mapping studies also identify structural rearrangements, mostly in the form of insertions and deletions (Palmer *et al.* 1988). These structural variations have been rarely included in a phylogenetic study (e.g. Sytsma and Gottlieb 1986; Soltis *et al.* 1990) but there are exceptions (e.g. Doebley *et al.* 1987). The stated reason for this exclusion is usually that homology of length variants appears to be difficult to determine. Homology of restriction site variants cannot solve the uncertainty of consistency indices (in many analyses below 0.5), suggesting that the presence or absence of a restriction site is not always a good indication of homology.

2.2.5.4. Restriction fragment length polymorphism (RFLP)

The principle of the restriction fragment length polymorphism markers is a specific enzymatic fragmentation of the whole genomic DNA followed by the separation of these fragments by gel electrophoresis based on their sizes (Helentjares *et al.* 1986). Following a Southern blotting, the immobilisation of the processed DNA on a layer and hybridisation with a radioactive probe sequence are achieved. The layer is subsequently placed next to an x-ray film. After the development of the film, bands are generated from each spot that probe base arrangement matched with the DNA sequence. RFLP markers are repeatable in laboratories as well as among parental lines. Their major weaknesses are their need to large quantities of DNA, and being expensive and time consuming.

2.2.5.5. AFLP as a mixture of restriction site and PCR methods

Amplified fragment length polymorphisms (AFLPs) bring the restriction site aspect of RFLP and the exponential amplification aspects of PCR together (Vos *et al.* 1995). Amplification is fulfilled in two steps: pre-selective and selective amplifications. For the first step, adapter oligonucleotides are matched to each end of the restriction sites. These adapter oligonucleotides subsequently work as common binding points for the annealing stage of PCR. For the second step, different primer sequences are used. These comprise an adapter oligonucleotide, the restriction site, and a selective sequence of bases. This results in discriminatory amplification of the fragments with the primer extensions matching the nucleotides that flank the restriction sites. The number of segments that may be produced varies between 50 and 100 on a denatu ing polyacrylamide gel. Furthermore, using fluorescent primers, it is possible to use three pairs of primers (with different fluorescent labels) to detect even more polymorphism. This method is used in inter- and intra-specific levels rather th in higher levels and has the disadvantage of dominant nature that makes it unable to distinguish homozygotes from heterozygotes, preventing estimation of inbreeding from the measurement of heterozygosity. It is also a very expensive method.

2.3 Character evolution

Well-defined molecular phylogenies pr ovide a framework for morphological evolution studies. Such a framework perr iits biologists to ask if specific characters or character states have appeared several times and in some instances it can lead to the discovery of the genetic determinants of morphological characters (Clegg 1993).

2.4 Genomes

Each plant cell has three genomes: one in the nucleus, another genome in the chloroplasts, and one in the mitochondria. The two primary sources of DNA variation selected for evolutionary studies are cpDNA (Palmer 1987; Palmer *et al.* 1988; Olmstead *et al.* 1990; Clegg *et al.* 1991; Clegg and Zurawski 1992; Downie and Palmer 1992b) and the nuclear ribosomal repeat region (Knaak *et al.* 1990; Baldwin 1992; Hamby and Zimmer 1992). The chloroplast genome has been the focus of most molecular phylogenetic wo ks in plants because in general, total-DNA analysis would not be used above the family level, or across such diverse families as the Fabaceae and Onagraceae. This restriction seems to be because of too much DNA variation, both in alignment and structure (Dowling *et al.* 1996).

2.4.1 Genome inheritance

The nuclear genome is inherited biparer tally. In contrast, the phylogeny based upon chloroplast DNA (cpDNA) sequence data shows the maternal evolution because cpDNA is inherited maternally in most angiosperms (Koch and Al-Shehbaz 2000). cpDNA is evolutionarily a conservative DNA molecule (Curtis and Clegg 1984).

2.4.2 Nature of cpDNA

The chloroplast genome is a prokaryotic DNA, circular-shaped double-stranded nucleic acid molecule usually 120,000 to 217,000 base pairs long (Palmer 1987). It exists inside chloroplast, in high copy number, and has a super-coiled, tertiary structure. This genome is an important component of total DNA, known to encode roughly 100 genetic functions.

With the exception of a pair of identical large inverted repeats discovered in nearly all the plants, chloroplast DNA is made up predominantly from single-copy sequences, another property that is useful for molecular systematists. Thus, cpDNA sequences provide an uncomplicated evolutionary model in which past events are much easier to reconstruct than in that of a typical nuclear gene (Doyle 1993).

2.4.2.1 cpDNA for phylogeny

cpDNA is the most widely studied plant genome regarding both molecular organisation and evolution (Clegg *et al.* 1994). The functional categories in cpDNA are protein-coding genes, introns, and DNA regions that do not code for tRNA, ribosomal RNA, or protein. Non-coding DNA in chloroplast is a small proportion of the whole molecule relative to the proportion of non-coding sequence in the nuclear genome. For example, only 32% of the rice cpDNA molecule is non-coding (Fig. 2.2). A number of recent studies have revealed very complicated patterns of mutation in non-coding regions (Clegg *et al.* 1994). Different parts of the chloroplast genome might be under different selection systems and have properties that make them more or less suitable for phylogenetic studies (Oxelman *et al.* 1999).

2.4.2.2 *rbc*L

This single copy gene is approximately 1431 bp in length and has a fairly conservative rate of evolution imposed by the function of ribulose 1, 5-biphosphate carboxylase/oxygenase (RuBisCo) (Johnson and Soltis 1994). This enzyme is part of a larger enzyme that catalyzes the combination of carbon dioxide and ribulose 1, 5-biphosphate into two molecules of 3-phosphoglycerate (Palumbi 1996). *rbc*L codes for the large subunit of this enzyme (Backlund *et al.* 2000). Coen *et al.* (1977) and McIntosh *et al.* (1980) did cloning and determination of sequences of *rbc*L gene for the first time by working on maize (*Zea mays*). This slowly evolving determinative gene seems to be particularly useful for broad analyses of higher-level relationships (Doyle 1993).

2.4.2.2.1 *rbc*L for deep plant phylogeny

The chloroplast *rbc*L gene has been used for phylogenetic studies in many plant families (e.g. Ritland and Clegg 1987; Kim *et al.* 1992; Olmstead *et al.* 1992; Chase *et al.* 1993; Conti *et al.* 1993; Kron *et al.* 1993; Morgan and Soltis 1993; Olmstead *et al.* 1993; Price and Palmer 1993; Bremer *et al.* 1994; Olmstead and Reeves 1995; Duval and Morton 1996; Gustafsson *et al.* 1996; Muasya *et al.* 1998; Soltis *et al.* 1998; Oxelman *et al.* 1999; Backlund *et al.* 2000; Muasya *et al.* 2000a,b; Savolainen *et al.* 2000). Thus, the tempo and mode for *rbc*L evolution is well known compared with other plant chloroplast genes (Kim and Jansen 1995).

Using *rbc*L arrangement data, the Cycads-Angiosperms and Monocotyledons-Dicotyledons divergence times were estimated (Wolfe *et al.* 1989a; Bousquet *et al.* 1992). *rbc*L is less useful when answering relationships amongst closely linked genera, for instance *Hordeum*, *Triticum* and *Aegilops* (Gielly and Taberlet 1994b). Regions of DNA that do not vary are constrained by functional requirements, while other regions that more frequently change might be less constrained and in some conditions, these regions may provide responses to adaptive changes (Hilu and Liang 1997). This leads us to the fact that the former regions are more useful for phylogenetic studies at higher levels where changes are clear only when the significant evolutionary differences are visible while the latter regions are useful for studying phylogeny at lower levels where only very small changes makes it possible to identify the species or infraspecific taxa.

The spacers separating the gene repeats vary enough for using at intergeneric (Baldwin 1992) and even infraspecific levels (Schaal and Learn 1988). The most widely studied non-coding region of the chloroplast genome, in the grass family (Poaceae), is a sequence near the *rbc*L gene, known as *atp*B-*rbc*L non-coding spacer. This region is nearly as long as the *rbc*L gene in rice therefore it is a long non-coding region in the genome.

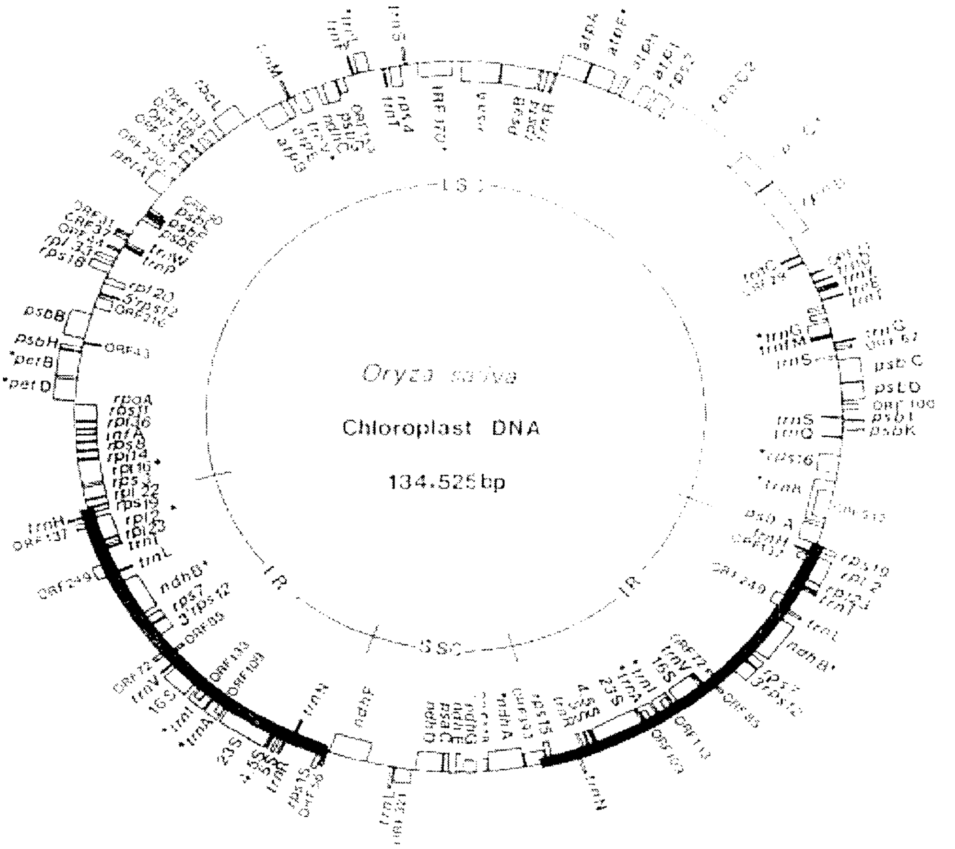

Fig. 2.2. Genetic circle map of the *Oryza sativa* chloroplast genome drawn to scale. Genes shown on the outside of the circle are encoded on the A strand and transcribed counter-clockwise. Genes on the inside are encoded on the B strand and transcribed clockwise. Asterisks denote split genes. LSC = large single copy region, IR = inverted repeat, SSC = small single copy region (figure 1 of Hiratsuka *et al.* 1989).

The *rbc*L gene has a reasonable length (nearly 1500 bp) to give sufficient information for phylogenetic reconstruction. An issue of a prestigious scientific journal includes several papers referring to the evolutionary significance of *rbc*L; for example, Chase *et al.* (1993) revealed the ease of use of more than 500 sequences of this gene in seed plants. Among chloroplast sequences, *rbc*L clearly leads in terms of the number of species examined. Like other conservative protein-coding genes, major parts of the variation of *rbc*L gene occur as silent (synonymous) substitutions; this feature together with the low level of insertion/deletion (except at the 3′ end of the gene) make the alignment of this gene unambiguous (Doyle 1993). Unambiguous alignment is one of several special advantages for the *rbc*L gene for the study of seed plants evolutionary history. A second important advantage is the conservative rate of nucleotide substitution. An important consequence of the conservative rate of substitution is that man-made oligonucleotide primers can be designed and used universally (Clegg 1993).

As Olmstead and Reeves (1995) note the phylogenetic utility of this gene is sometimes restricted to above family levels (but see contrary views in Conti *et al.* 1993; Gadek and Quinn 1993; Kron *et al.* 1993; Price and Palmer 1993; Soltis *et al.* 1993). This is due to its slow evolutionary rate (Kim and Jansen 1995). Direct estimates of rates of base substitution in the gene confirmed that the rate of accumulation of nucleotide change was remarkably slower than in plant and animal nuclear genes (Zurawski *et al.* 1984; Wolfe *et al.* 1987; Zurawski and Clegg 1987; Wolfe *et al.* 1989b). At the intrafamilial level, *rbc*L sometimes does not provide adequate information to show evolutionary relations (Clegg 1993; Olmstead and Reeves 1995), such as within the Asteraceae (Kim *et al.* 1992), the Cornaceae (Xiang *et al.* 1993), and the Saxifragaceae (Soltis *et al.* 1993). Furthermore, in Triticeae of the Poaceae, relationships among the closely related genera *Hordeum*, *Triticum*, and *Aegilops* could not be resolved (Doebley *et al.* 1990; Gaut *et al.* 1992). These lower levels need a gene or region of DNA with faster evolution because the comparison will be among much closer taxa.

Gaut *et al.* (1992) found rate homogeneity for *rbc*L sequences in well-defined families, but substantial rate heterogeneity in interfamilial contrasts. Even above the family level, the data provided by this gene might fail to give fully resolved phylogenetic trees in some plant groups (Kim and Jansen 1995). Sequencing studies focusing on individual families and sufficient sampling to examine infrafamilial relationships (Conti *et al.* 1993; Doebley *et al.* 1990; Kim *et al.* 1992; Olmstead and Sweere 1994; Soltis *et al.* 1993) have shown the necessity of using a longer gene to gain resolution in phylogenetic reconstruction. Interest thus exists in finding another useful region (or regions) on the genome that evolves faster than *rbc*L does to facilitate lower-level phylogenetic reconstructions (Johnson and Soltis 1994).

Stevenson *et al.* (2000) found that *rbc*L has one of the lowest retention indices (RI) and data decisiveness among genes (nuclear, mitochondrial, and chloroplast). RI is the relative amount of similarity within a character, which is known as synapomorphy on a tree. Therefore, this gene is one of the best genes for parsimony analyses.

2.4.3 mtDNA

Plant chloroplast DNA is like animal mitochondrial DNA, which has provided a great deal of information for evolutionary investigations (Doyle 1993). Plants, too, have mitochondrial genomes, but mitochondrial DNA has not been found to be of comparable value by plant systematists or evolutionists. Unlike the small, compact, structurally stable, but rapidly evolving animal mitochondrial DNA, that of angiosperm plants varies widely in size, structure, and gene order (Newton 1988; Palmer 1990, 1992) even within the same family, and evolves more slowly than the chloroplast genome, making whole-genome restriction site studies difficult. Nevertheless, some mitochondrial genes might provide useful data in evolutionary systematic studies involving distant relationships (Palmer 1992).

According to Olmstead and Palmer (1994), in comparison with mitochondrial DNA, three features of the chloroplast genome offer distinct advantages in phylogenetic investigations at the species level and above. First, the approximately tenfold larger size of the cpDNA and six times greater amount of protein genes provide a much larger database for restriction site studies and greater choice for sequence comparisons. Second, the greater than tenfold lower silent replacement rate for cpDNA compared with animal mitochondrial DNA (Wolfe *et al.* 1987) makes the exact comparison of nucleotide sequences for higher-level phylogenetic studies more feasible for cpDNA than for mtDNA. Third, structural rearrangements, although infrequent in both cpDNA and animal mtDNA, are somewhat more common in cpDNA, with many inversions and nucleotide deletions distinguished in angiosperms (Jansen and Palmer 1987a; Downie and Palmer 1992b).

2.5 Molecular clock

Early indications of the strong relationship between measurements of genome divergence and divergence time (Zuckerkandl and Pauling 1962) raised the exciting possibility that molecular comparisons could provide indications of the time of divergence for taxa where no fossils exist. Although most scientists now acknowledge an apparent correlation linking amount of DNA divergence and time, some evidence (reviewed in chapter 12; Gillespie 1991; Avise 1994) indicates sufficient rate heterogeneity that one should not assume that rates are equal on an *a priori* basis. For instance, Gillespie (1987) calculated the proportion of the variance of the number of the substitutions to the mean number of substitutions that occur along a lineage as ranging from 1 to 35 for amino acid substitutions, indicating considerable fluctuation in evolutionary rate (see also Lynch and Jarrell 1993). This has considerable implications for modern systematics.

Since DNA mutations accumulate at a quite constant rate, the quantity of variations observed between the DNA sequences of two species can be used to estimate the relative time that those two species have been diverging (Norman and Christidis 2000). Constancy of rates has been a principle of the unbiased theory of molecular evolution, is an assumption of a few methods for estimating phylogeny (chapter 11; Hillis *et al.* 1996), and is widely assumed in estimating time since divergence (chapter 12; Hillis *et al.* 1996).

The conceptual questions about the process of evolution over time as well as improvement of molecular and statistical techniques and softwares has forced the macromolecules to be in the centre of attention for answering these time-dependent biological processes. Morphological data do not seem to be able to answer these questions as logically as molecular data are.

2.6 Separate vs. combined analyses

It is advantageous comparing data from molecular and non-molecular sources individually and in combination and sometimes 'conditional data combination' (Huelsenbeck *et al.* 1996) is recommended. The latter is recommended when either simultaneous or partitioned analysis is preferred to the other method. The 'conditional' is based on heterogeneity between the data partitions. As Bull *et al.* (1993) stated, when different datasets result in different trees and the difference is too much to be justified, the analyses should be done separately rather than simultaneously. In many instances genomic phylogenies support accepted morphological analyses but add more resolution to that group's evolutionary description (Norman and Christidis 2000).

Combined molecular data sets have been used before and presented more resolution and internal support for relationships than individual data sets (e.g. Soltis *et al.* 1993, 1996; Johnson and Soltis 1994, 1995; Plunkett 1994; Olmstead and Sweere 1994, 1995; Xiang *et al.* 1998; Bäcklund *et al.* 2000; Olmstead *et al.* 2000; Persson 2000). However, there are some exceptions to combining data obtained from different genes from cpDNA and nrDNA (Mason-Gamer and Kellogg 2000). Mason-Gamer and Kellogg (2000) found that the morphological data have higher level of homoplasy compared with the molecular data.

2.7 Advantages and disadvantages of various methods and gene regions for molecular systematics

2.7.1 Chloroplast gene regions other than *rbc*L

2.7.1.1 *ndh*F

Initial applications of another chloroplast gene (*ndh*F) sequences for systematic works centred on intrafamilial relationships (e.g. Olmstead and Sweere 1994; Kim and Jansen 1995; Olmstead and Reeves 1995; Scotland *et al.* 1995; Neyland and Urbatsch 1996: Catalan *et al.* 1997; Terry *et al.* 1997a,b). These works resulted in obtaining more parsimony-informative characters than for *rbc*L. An extensive sequence comparison of this gene (*ndh*F) obtained from the main clades of the biggest plant family (Asteraceae) has shown that the *ndh*F gene is about three times more informative than *rbc*L. This is partly because it is substantially longer. For example, the *ndh*F gene is 2133 base pairs long in tobacco (Shinonzaki *et al.* 1986; Olmstead *et al.* 1993), which is 1.5 times longer than *rbc*L. Amino acid sequence divergence of the *ndh*F gene is four times greater than *rbc*L in comparisons between tobacco (a dicot) and rice (a monocot). Furthermore, *ndh*F evolves twice as rapidly as *rbc*L (Kim and Jansen 1995). The *ndh*F gene (on the basis of observed homologies) codes for subunit 6 of NADH-dehydrogenase and is suggested as an alternative for *rbc*L (Sugiura 1989). The average number of base substitutions per 30 base pairs is about 15 in *ndh*F whereas it was about 9 in *rbc*L gene, making *ndh*F suitable for analyses at the generic or higher levels (Olmstead and Sweere 1994; Clark *et al.* 1995; Bohs and Olmstead 1997; Terry *et al.* 1997c). Furthermore, the number of informative sequences has been reported as three times more in *ndh*F than in *rbc*L (e.g. Kim and Jansen 1995; Oxelman *et al.* 1999). For these reasons *ndh*F sequences should provide greater resolution of intrafamilial relationships than *rbc*L (Kim and Jansen 1995; Olmstead and Reeves 1995).

In many families of angiosperms *ndh*F phylogenies (even with restricted taxon sampling) have showed more homology and phylogenetic information than *rbc*L as indicated by a higher consistency index (e.g. Olmstead and Sweere 1994; Kim and Jansen 1995; Scotland *et al.* 1995; Oxelman *et al.* 1999). The *ndh*F data also is efficacious with greater support for the individual branches (Oxelman *et al.* 1999). Comparison of sequences of 18 chloroplast genes in angiosperms showed that only six genes including *ndh*F, *rpo*C1, and *rpo*C2 are much longer and evolve faster than *rbc*L (Wolfe 1991; Olmstead and Palmer 1994; Kim and Jansen 1995). The location of *rpo*C1 and *rpo*C2 genes (F g. 2.1) is in the large single-copy region of the plastid genome and these genes along with *rpo*B encode three subunits of the chloroplast RNA polymerase (Downie *et al.* 1996). Chloroplast genes that evolve faster than *rbc*L can be extremely useful for clarifying relationships at lower levels than that of family (Xiang *et al.* 1998). *ndh*F is used in most of the studies because it is bigger and has evolved slightly faster than *rpo*C1. Another reason for this choice is that rpoC1 is sometimes absent chloroplast DNA of angiosperms (Kim and Jansen 1995). Olmstead *et al.* (2000) showed that the results obtained from the analysis of the *ndh*F gene are largely consistent with their earlier works on the *rbc*L gene, but provide more resolution and bootstrap support for the tree.

2.7.1.1.1 Combined data from *rbc*L and *ndh*F

The combined sequence data from *rbc*L and *ndh*F has been analysed in some families using parsimony methods (e.g. Backlund *et al.* 2000). Results from these studies based on both separate and combined analyses of nucleotide sequence data of these genes were considerably congruent (e.g. Backlund *et al.* 2000). The other approach (combining cpDNA sequences) has provided useful and informative data in several plant systematic studies (e.g. Olmstead and Sweere 1994; Olmstead and Reeves 1995; Scotland *et al.* 1995; Chase and Cox 1998; Soltis *et al.* 1998; Oxelman *et al.* 1999). In a study on the family Buddlejaceae, the results obtained from combined analysis are strongly supported by *ndh*F data alone (Oxelman *et al.* 1999). In such combined analyses, parsimony and jackknife analyses should be performed for each of the two genes independently (Backlund *et al.* 2000). Parsimony is the general ideal criterion for picking among different hypotheses that explains the processes in the simplest way and with the least complications. Jackknife (and bootstrap) is (are) statistical procedure(s) for achieving a better estimate of the parametric variance of a distribution from small samples than the observed sample variance by averaging pseudoreplicate variances (Kitching *et al.* 1998).

2.7.1.1.2 Variation within *ndh*F

The 5′ region of the *ndh*F gene is very different from the 3′ end in both rate and pattern of sequence changes. These differences probably reflect different operational constraints in these two regions. The 5′ region of *ndh*F represents evolutionary models similar to those from *rbc*L whereas about 60% of informative sites occur in the 3′ region (Kim and Jansen 1995). The two patterns of base substitutions in *ndh*F are specifically advantageous in phylogenetic reconstruction due to the utility of the stable and changeable regions of the gene to distinguish older and newer groups, respectively. Thus, the longer the gene the more information we get, and the higher sequence divergence the more reliable data for discovering phylogenetic relationships. In general the 5′ end is more conservative than the 3′ end (Kim and Jansen 1995).

2.7.1.2 *mat*K

Recent studies (e.g. Johnson and Soltis 1994,1995; Plunkett 1994; Steele and Vilgalys 1994) have shown the benefits of using *mat*K (previously ORFK; Neuhaus and Link 1987; Wolfe *et al.* 1992) sequence data to successfully resolve generic and even species-level relationships (Steele and Vigalys 1994; Kron 1997; Brochman *et al.* 1998). The *mat*K gene gives reliable information at a wide range of phylogenetic levels form tribe to ordinal levels as well (Hilu and Liang 1997). This chloroplast gene codes for a maturase enzyme. Moreover, the *mat*K gene is part of the intron of the transfer RNA gene for lysine (*trn*K) (Neuhaus and Link 1987; Johnson and Soltis 1994, 1995). The presence of the gene in the parasitic *Epifagus* (Orobanchaceae), a taxon with 65% lost chloroplast genes indicates the importance of the *mat*K gene (Table 2.1) in terms of the basic and stable role that it has in plants. In fact, both exons of the *trn*K gene that flank the *mat*K were lost, leaving the gene intact (Wolfe *et al.* 1992), except for a large deletion in the coding region. Olmstead and Palmer (1994) reported that among 20 genes utilised in molecular systematics, the *mat*K gene has the highest substitution rate.

2.7.1.2.1 Variation within *mat*K

The presence of the relatively conserved region (3′ end) and a less conserved 5′ part provides two sets of data that can be used at different levels of taxonomy (Hilu and Liang 1997). Sequence studies in the Saxifragaceae support the usefulness of the 5′ end of *mat*K at the intrafamilial level (Johnson and Soltis 1995).

Table 2.1. Twenty taxa, their families, and the sequence length (base pairs) of the *mat*K gene used in Hilu and Liang's (1997) study (modified from Hilu and Liang 1997).

Species	Family	Length (bp)
Marchantia polymorpha	Marchantiaceae	1113
Pinus contorta	Pinaceae	1548
Pinus thunbergii	Pinaceae	1548
Saxifraga integrifolia	Saxifragaceae	1518
Sullivanita sullivantii	Saxifragaceae	1521
Sinapis alba	Brassicaceae	1575
Solanum tuberosum	Solanaceae	1530
Nicotiana tabacum	Solanaceae	1530
Epifagus virginiana	Orobanchaceae	1320
Hordeum vulgare	Poaceae	1515
Oryza sativa	Poaceae	1629
Avena sativa	Poaceae	306
Dactyloctenium aegypticum	Poaceae	306
Eleusine indica	Poaceae	306
Paspalum almum	Poaceae	306
Arundo donax	Poaceae	306
Phyllostachys aurea	Poaceae	306
Cyperus strigosus	Cyperaceae	306
Smilax rotundifolia	Smilacaceae	306
Joinvellea plicata	Joinvilleaceae	306

In the Saxifragaceae, the *mat*K gene was found to evolve approximately two to three times faster compared with *rbc*L (Johnson and Soltis 1994, 1995). The sequences of *mat*K in Polemoniaceae have a total rate twice as fast as in *rbc*L (Steele and Vilgalys 1994). Sequence data for *mat*K has been shown to have more variation than *rbc*L in some studies (Johnson and Soltis 1995; Gadek *et al.* 1996). The total number of nucleotide replacements per site in *mat*K has been shown to be 2.1 times that of *rbc*L in Cornales (Xiang *et al.* 1998), 2.3 times in Apiales (Plunkett 1994), and 3 times in Saxifragaceae (Johnson and Soltis 1994,1995). Systematic studies using the sequences of this gene at inter- and intrafamilial levels found that indels can occur in *mat*K and some of these indels are informative (e.g. Johnson and Soltis 1994, 1995; Steele and Vilgalys 1994; Plunkett *et al.* 1996, 1997; Kron 1997; Xiang *et al.* 1998). For instance indels and nucleotide substitutions at the 3′ end of *mat*K provide valuable phylogenetic information in the family Poaceae and might yield data useful for phylogenetic investigations of other plant families. Moreover, Xiang *et al.* (1998) observed five indels corresponding to the five alignment gaps in *mat*K, but no indels in *rbc*L. However, Plunkett *et al.* (1997) noted that indels usually are not abundant in the sequences of this gene. The relatively high rate of transposition and reasonable length underscore the usefulness of the *mat*K gene in systematic studies.

The use of the entire *mat*K gene is not always as instructive or important as using sections of the gene. In fact, some sectors of *mat*K can provide more noise than other parts. It has been stated that despite *mat*K's rapid evolution and a high potential of showing homoplasy caused by multiple base replacements at higher taxonomic levels, it allows recovery of relationships at the ordinal level. Moreover, detailed investigation of different sections of *mat*K showed the need to evaluate all the regions of a gene for evolutionary and systematic studies. For instance the position of some taxa in Saxifragaceae was quite different using the different regions of the *mat*K gene (Hilu and Liang 1997).

2.7.1.2.2 Comparison of *rbc*L and *mat*K

The rate and pattern of variation in this gene sequence suggests that *mat*K is less practically constrained than *rbc*L (Xiang *et al.* 1998; Hilu and Alice 1999). In functionally constrained genes, such as *rbc*L, substitutions of the nucleotides translate into lower amino acid variation (Hilu and Liang 1997). Neuhaus and Link (1987) report a 66% amino acid similarity between the peptides translated from *mat*K in *Nicotiana* and *Sinapis*, whereas *rbc*L shows 93% amino acid similarity in these two members of two different families. As Liang and Hilu (1996) reported, *mat*K sequences are 14.9% informative in grasses, which is higher than the 11.5% obtained from the *rbc*L sequences. In addition, according to Xiang *et al.* (1998), *mat*K has a much higher A–T content than *rbc*L, which makes DNA amplification easier in *mat*K. Transversions (changes between classes) are considered the most informative kind of nucleotide substitutions in phylogenetic studies (Quicke 1993). Transition/transversion ratios have been observed to be 2.0 for recently evolved *mat*K sequences and exceed 0.4 for highly diverged sequences (Holmqu st 1983). A lower transition/transversion ratio for *mat*K compared to *rbc*L means that *mat*K is a more informative gene (Xiang *et al.* 1998).

The information content of variable sites of *mat*K is 1.5 times higher than *rbc*L. The length of the gene is, however, less than the *rbc*L gene (Liang and Hilu 1996), which is a disadvantage in DNA sequencing, particularly in comparison to *ndh*F gene (1500 base pairs compared to 2133). Information from *mat*K sequences in phylogenetic studies is in complete congruence with those from nuclear and chloroplast genes (Liang and Hilu 1996). The *mat*K gene with its described features and high rates of base replacement represents a molecule that gives insight into the evolutionary and systematic problems at different levels.

2.7.1.2.3 Combined data from *rbc*L and *mat*K

For instance, more resolution was achieved among the lineages of Cornales from analyses of a combined data set compared with analysis of either *rbc*L or *mat*K data separately. The combined *rbc*L-*mat*K analysis revealed the same main subclades inside *Cornus* found in the *mat*K analysis alone (Xiang *et al.* 1998).

2.7.1.3 Non-coding regions of DNA

2.7.1.3.1 Introduction

There is increasing interest in pract cal analysis at low taxonomic levels of the sequences of those regions of chloroplast that do not code for a protein. Recognition of some limitations in coding DNA to resolve relations at low levels prompted the search for chloroplast introns and intergenic spacers of phylogenetic utility. Underlying this challenge was the assumption that these regions are subjected to limited or otherwise no particular pressure.

Kelchner (2000) has described non-coding regions in general as follows:

1. Non-coding regions appear to be highly organised and their elements evolve non-randomly and non-independently.
2. The resulting gaps in the aligned sequences may be phylogenetically valuable information to be applied in the phylogenetic studies.

2.7.1.3.2 Variation in non-coding regions

Non-coding regions of cpDNA are likely to evolve more rapidly than coding regions (Curtis and Clegg 1984; Wolfe *et al.* 1987; Zurawski and Clegg 1987; Clegg and Zurawski 1991; Taberlet *et al.* 1991; Bohle *et al.* 1994; Gielly and Taberlet 1994a; Sang *et al.* 1997), and therefore they are likely to be more useful below the family level. Among closely related species, non-coding regions exhibit very high level of variation of sequence in cpDNA compared to the coding region, making the former more useful at lower taxonomic levels. The analyses of these regions that display more mutations than other regions extend the level of taxonomic resolution of the cpDNA molecule and so estimate divergence among closely related species, even at the intraspecific level (Taberlet *et al.* 1991). Insertions and deletions are common in non-coding regions and might be caused by intramolecular recombination (Ogihara *et al.* 1988; Palmer 1991), slipped-strand mispairing (Takaiwa and Sugiura 1982), and stem-loop formation (Sears *et al.* 1996). There was an idea that these regions might experience random and individual mutations, both in mode and distribution.

2.7.1.3.2.1 Basis of variation of non-coding regions

Comparative studies on non-coding chloroplast DNA sequences mainly over the last decade (e.g. Palmer 1985; Blasko *et al.* 1988; vom Stein and Hatchel 1988; Wolfson *et al.* 1991; Golenberg *et al.* 1993; Gielly and Taberlet 1994a; Morton 1995; Downie *et al.* 1996a; Kelchner and Wendel 1996; Kelchner and Clark 1997; Sang *et al.* 1997) have agreed with the inference of particular underlying mutational mechanisms responsible for generating sequence diversity within non-coding fragments of the chloroplast DNA.

2.7.1.3.2.1.1 Slipped-Strand Mispairing (SSM)

SSM is a key, even primary, factor in length mutations within non-coding regions of the chloroplast, mitochondria, and nuclear genomes (e.g. Levinson and Gutman 1987; Hancock 1995; Wolfson *et al.* 1991; Kelchner and Clark 1997; Sang *et al.* 1997).

Strings of single nucleotide repeats, mainly of adenine or thymine, appear regularly in non-coding cpDNA, and slipped-strand mispairing may generate length mutations within these strings. Representation of lengthy repeated sequences in these regions of cpDNA would then be the "snapshot" of chains experiencing frequent insertions or deletions at this locality (Kelchner 2000).

2.7.1.3.2.1.2 Stem-loop secondary structure

Commonly prominent in spacers as well as introns in the cpDNA is the rate and size of expected secondary structures referred to as "stem-loops". Stem-loops are assumed to appear during single-stranding events when inverted repeats meet to form a region of pairing (the stem) surmounted by their interceding sequence (the loop). Such structures have been broadly discussed in ribosomal DNA, where ITS together with 18S coding regions have been of specific interest to plant phylogeny researchers (Baldwin *et al.* 1995; Soltis *et al.* 1997; Soltis and Soltis 1998 discuss these structures within ITS and 18S rDNA and the phylogenetic consequences of these structures). Gielly and Taberlet (1994a) reported quite a few probable stem-loops within the *trn*L–*trn*F (*trn*L–F) region including nine stem-loops in the *trn*L intron only.

2.7.1.3.2.1.3 Nucleotide substitutions

Nucleotide substitutions have been generally described as happening more in non-coding than in coding regions (Wolfe *et al.* 1987; Zurawski and Clegg 1987; Olmstead and Palmer 1994; Hoot and Douglas 1998; however, see Sang *et al.* 1997, for an exception). Percent AT quantity is extremely variable in non-coding cpDNA region, though it will be higher than the average value for the chloroplast genome (Shimada and Sugiura 1991; Downie *et al.* 1996a; Small *et al.* 1998). Kajita *et al.* (1998) announced an AT content of 67% in the *trn*L–F intergenic spacer together with the *trn*L region, Kelchner and Clark (1997) showed 70.5% AT content in the intron of the *rpl*16 chloroplast genome and Small *et al.* (1998) reported an incredible 71.5% AT content in the intergenic spacer *trn*T–L in *Gossypium*. Undoubtedly, this wide variation in AT value in non-coding chloroplast genome has quite a few as still undetermined implications for evolutionary analysis of non-coding DNA data. At least, it shows a strong base composition bias into the analysis.

Attributes of the molecular development of a non-coding region influence the manner of mutation and the allocation of base substitution occurrences in a spacer region or intron (Kelchner 2000). There is a linkage between transition/transversion ratios and neighboring base composition in non-coding regions (Morton 1995a, b; Morton *et al.* 1997; Savolainen *et al.* 1997), suggesting that A and/or T will demonstrate a significant tendency toward transversion mutations. Kelchner (2000) has confirmed the bias in non-coding regions including transition/transversion replacement ratios caused by the effect of nearby bases. Such a bias limits potential nucleotide substitutions at such sites, raising the possibility of parallelism and reversals, particularly if the site is subjected to multiple hits. Besides, one would expect transversion substitutions to be more common in high AT content data sets.

2.7.1.3.2.1.4 Intramolecular recombination

Intramolecular recombinations on a genomic scale have been suggested between nearby or very close repetitions in chloroplast DNA (Howe 1985; Palmer *et al.* 1985; Palmer *et al.* 1987; Blasko *et al.* 1988; Ogihara *et al.* 1988; Milligan *et al.* 1989; Kanno and Hirai 1992; Kanno *et al.* 1993; Morton and Clegg 1993; Hoot and Palmer 1994). In a non-coding region comparison, such a major recombination in the specific region of study would result in indels of surprising size, which contain sequence content not readily identifiable in origin (Kelchner 2000).

Recombination events might operate on a larger scale in an isolated non-coding fragment. Occasionally one finds extensive removed sequence within an alignment with no apparent mechanistic explanation, presence of a small or moderately sized inversion, or a large insertion showing little congruence with surrounding sequence pattern. Such mutations indicate intramolecular recombination, and they frequently occur in the loop regions of probable secondary structures. Sequences involved in stem-loops might be particularly subject to these events because of the conserved inverted repeats and mutationally flexible loop. Therefore, such structures could experience interactive recombination with other stem-loops, particularly with those existing in complementary sequence position.

2.7.1.3.3 Taxonomic value of non-coding DNA

Many studies have shown the phylogenetic usefulness of individual non-coding fragments in cpDNA:

- *trn*L–F spacer (e.g. Gielly and Taberlet 1994a; Mes and t'Hart 1994; van Ham *et al.* 1994; Sang *et al.* 1997; Cros *et al.* 1998; Bayer and Starr 1998);
- *trn*T–L spacer (Bohle *et al.* 1994, 1997; Small *et al.* 1998);
- *rpo*A-*pst*D and *rps*11-*rpo*A spacers (Peterson and Seberg 1997);
- *atp*B-*rbc*L spacer (Golenberg *et al.* 1993; Hodges and Arnold 1994; Natali *et al.* 1995; Samuel *et al.* 1997; Savolainen *et al.* 1997; Setoguchi *et al.* 1997; Hoot and Douglas 1998);
- *rbc*L-*psa*I spacer (Morton and Clegg 1993);
- *psb*A-*trn*H spacer (Aldrich *et al.* 1988; Sang *et al.* 1997);
- *acc*D-*psa*I spacer (Small *et al.* 1998);
- *rpl*16-*rpl*14 and *rps*8-*rpl*14 spacers (Wolfson *et al.* 1991);
- intron surronding *mat*K (Johnson and Soltis 1994);
- *rpo*C1 intron (Downie *et al.* 1996a, 1996b; Asmussen and Liston 1998; Downie *et al.* 1998);
- *rpl*16 intron (Jordan *et al.* 1996; Kelchner 1996; Kelchner and Clark 1997; Schnabel and Wendel 1998; Baum *et al.* 1998; Small *et al.* 1998);
- *trn*L intron (Sang *et al.* 1997; Bayer and Starr 1998; Kajita *et al.* 1998; Bayer *et al.* 2000);
- *rps*16 intron (Liden *et al.* 1997; Oxelman *et al.* 1997); and
- *ndh*A intron (Small *et al.* 1998).

Aligning these regions needs to be undertaken cautiously. How best to align the sequence matrix, determine homology and undertake the phylogenetic analysis of such regions is a critical issue. The nature of evolution that occurs in these regions may have considerable repercussions for the precision of distance, maximum likelihood, and parsimony methods.

2.7.1.3.4 Alignment of non-coding DNA

Golenberg *et al.* (1993) were the first to describe a standard for arranging gaps in a line for non-coding cpDNA matrices. Hoot and Douglas (1998) revised the gap alignment method suggested by Golenberg *et al.* (1993), starting the nomenclatural process for illustrating gap styles. Although a categorization system is not a key factor for misma ch treatment in phylogenetic studies, it might be worthwhile in collecting information about inferred mutational mechanisms if universally employed in non-coding DNA analyses.

2.7.1.3.4.1 Homology

A common form of indel in these regions is the direct repeat of a neighbouring sequence, as demons rated in figure 2.3 by an inserted repeat unit **ataaa** (Golenberg *et al.* 1993; Hoot and Douglas 1998). Homology can be extremely uncertain in these reproduced nucleotides. Therefore, such parts are removed from study as would-be phylogenetic characters (a conservative approach) or featured as coded gap features corresponding to length of the repeat string (often becoming highlv homoplasious in the context of a resulting topology). It is probably most sensible to delete such regions from consideration in a phylogenetic analysis (Kelchner 2000). Homology may be shown by the size of indels in the gap, nevertheless such a theory is not risk free (Kelchner 2000).

```
1. ATAAAACAAA-----GAGCG
2. ATAAAATAAAataaaGAGCG
3. ATAAAATAAA-----GAGCG
4. ATAAAATAAA-----GAGCG
```

Fig. 2.3. Type 1b gap as described by Golenberg *et al.* (1993) and Hoot and Douglas (1998). An inserted replicate of such nature might be large and difficult to mark as an duplicate unit during alignment.

2.7.1.3.5 Analysis of non-coding DNA

As a support for alignment, insertions or deletions caused by Slipped-Strand Mispairing would find the number of gaps (Kelchner 2000). Homology assessment at stem-loops can be difficult or impossible, and the conservative approach of deleting the non-homologous areas from the data matrix could be, in some cases, adopted (Kelchner 2000).

2.7.1.3.6 *trn*L intron and *trn*L–*trn*F intergenic spacer

Taberlet *et al.* (1991) used the cpDNA section bounded by *trn*L 5′ exon and *trn*F for the first time. Since then, this *trn*L intron region and the region between *trn*L and *trn*F have been used in phylogenetic studies of a broad range of plant families frequently (Gielly and Taberlet 1994a; van Ham *et al.* 1994; Gielly *et al.* 1996; Briggs *et al.* 2000; Kusumi *et al.* 2000; Muasya *et al.* 2000b; Persson 2000b; Yen and Olmstead 2000a, b; Bradford and Barnes 2001; Kores *et al.* 2001; Muasya *et al.* 2001a, b; Roalson *et al.* 2001; Liede *et al.* 2002). The single-copy *trn*L–F region is composed of a *trn*L intron, a 3′ exon (56 base pairs), and an intergenic spacer (IGS) amid the *trn*L (UAA) 3′ exon and the *trn*F (Taberlet *et al.* 1991) (Fig. 2.4). The *trn*L–F IGS section (101–422 bp) and the *trn*L intron, together with the *mat*K gene, evolve more rapidly than other regions in the chloroplast genome (Neuhaus and Link 1987; Gielly and Taberlet 1994a; Hilu and Liang 1997) and are usually used to investigate relationship among genera (Taberlet *et al.* 1991; Johnson and Soltis 1994; Plunkett *et al.* 1997), among species (Gielly and Taberlet 1994; Kajita *et al.* 1998), and also within species (Fujii *et al.* 1997). With these regions/genes, one can obtain a consistent phylogeny (Kusumi *et al.* 2000). In *Pelargonium*, using *trn*L–F data had been valuable in enlightening patterns of phenotypic and cytological variation and generating new hypotheses (Bakker *et al.* 1999). The size of the *trn*L intron varies from about 350 base pairs (*Avena* in Poaceae) to 600 base pairs (*Euphorbia* in Euphorbiaceae). This is unexpected; given the features of the *trn*L intron, which must be more stable because of its secondary structure and its catalytic properties (Gielly and Taberlet 1994a). Taberlet *et al.* (1991) showed that the sequences of the intergenic spacer could be very useful for evolutionary studies at interspecific and intraspecific levels. As Gielly and Taberlet (1994a) mentioned, the *trn*L intron evolves at the same rate as that of the IGS. By the study of those regions, they could clearly distinguish three close genera including *Hordeum*, *Triticum*, and *Aegilops*. A phylogeny from *trn*L intron alignments is inferred for a few species of *Gentiana* L. too, apparently illustrating the phylogenetic utility of these zones below the generic level (Gielly and Taberlet 1994a). Taberlet *et al.* (1991) also suggested that the *trn*L (UAA) region (a type I intron) is probably less variable due to the fact that it has catalytic properties and forms secondary structures. Therefore, it would be more useful for phylogenetic studies at higher taxonomic levels (Taberlet *et al.* 1991).

Briggs *et al.* (2000) concluded that most of the insertions and deletions in the *trn*L–F region appear to be phylogenetically informative. In their study sequences of the *trn*L intron and the *trn*L–F spacer gave more resolved trees than did the longer *rbc*L data. In contrast, Persson (2000) argued that *trn*L–F sequences are not informative enough for resolution of the relationships among closely related genera and at intertribal level in Rubiaceae.

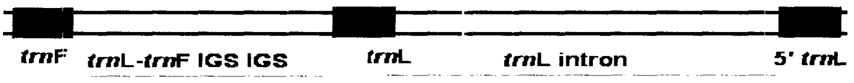

Fig. 2.4. A scheme of different parts of the non-coding *trn*L–F region including the *trn*L–F intergenic spacer and *trn*L intron.

2.7.2 Nuclear genes (regions)

2.7.2.1 Nuclear ribosomal DNA (nrDNA)

According to White *et al.* (1989), by directly sequencing amplified fragments of genes such as the nuclear ribosomal DNA genes, both the resolving power and phylogenetic range of comparative studies can be extended. They believe that a DNA sequence, amplified with many evolutionarily preserved or 'universal' primers provides a useful source of information about various ranges (from populations to phyla).

2.7.2.1.1 Characteristics

Nuclear ribosomal DNA (nrDNA) arrays should be especially useful for evolutionary studies because as Hillis and Davis (1986) mention:

1. The rDNA sequence is moderately to highly repetitive therefore multiple copies are usually present and sequenced. In other words, because of high copy quantities, they have been technically fairly easy to work with, thus letting systematists collect data from a wide range of taxa.
2. rDNA contains both the regions that evolve slowly (the 18S, 5.8S, and 28S rRNA genes) and the regions that evolve faster (the internal transcribed and non-transcribed spacers) (fig. 2.5). So they provide information for various taxonomic levels.

2.7.2.1.2 ITS

Internal transcribed spacer (ITS) sequences obtained from nuclear ribosomal RNA are reliable sources of information at the intergeneric level (e.g. Baldwin *et al.* 1995). According to Oxelman *et al.* (1999) and Smith (2000), poorly resolved or unresolved relationships based on only the data obtained from *ndh*F are resolved further by ITS data. In tribe Cariceae (family Cyperaceae) the length of the ITS 1 region ranges from 228-249 base pairs (bp) whereas the ITS 2 part varies from 196 to 245 bp. The whole length of the region varies from 425 to 653 bp with about 370 variable sites, of which 280 are potentially parsimony-informative (Roalson and Friar 2000)

Smith (2000) recommends the use of ITS sequences in systematic works as an informative tool for resolving phylogenetic relationships at the specific and generic levels. The transition/transversion ratio is 1.65 for ITS 1 and 1.17 for ITS 2, and 1.45 for the combined data matrix (Moller and Cronk 1997a). The two internal transcribed spacers (ITS1 and ITS2) of the nuclear rDNA evolve much more rapidly than the genes that flank them and are used in studies of closely related species and genera (Baldwin 1992, 1993; Suh *et al.* 1993; Wojciechowski *et al.* 1993). ITS also evolves faster than rbcL and ndhF genes (Oxelman *et al.* 1999). Both ITS and 5S-NTS can show sequences with various evolutionary histories (Doyle 1992; Baldwin *et al.* 1995), which can differ from the phylogeny of the organism (Persson 2000). On the other hand, although, the primers designed for sequencing ITS region are generally able to sequence the ITS region, they are not easy to use in all taxa or study groups. There are many references suggesting modifications in protocols using the "universal" primers introduced by White et al. (1989). Even so, sometimes the "universal" primers do not work properly for all the taxa being studied and extra primers are needed to complete the sequence data set. Also, the "universal" primers have been primarily designed for fungi. This suggests that the ITS loci are not always easy to amplify using the products of PCR by a standard protocol. Whereas in the majority of the cases evolutionary topologies concluded from internal transcribed spacer data seem sensible biologically, one has to scrutinise results carefully once a number of taxa vary greatly in the PCR yield.

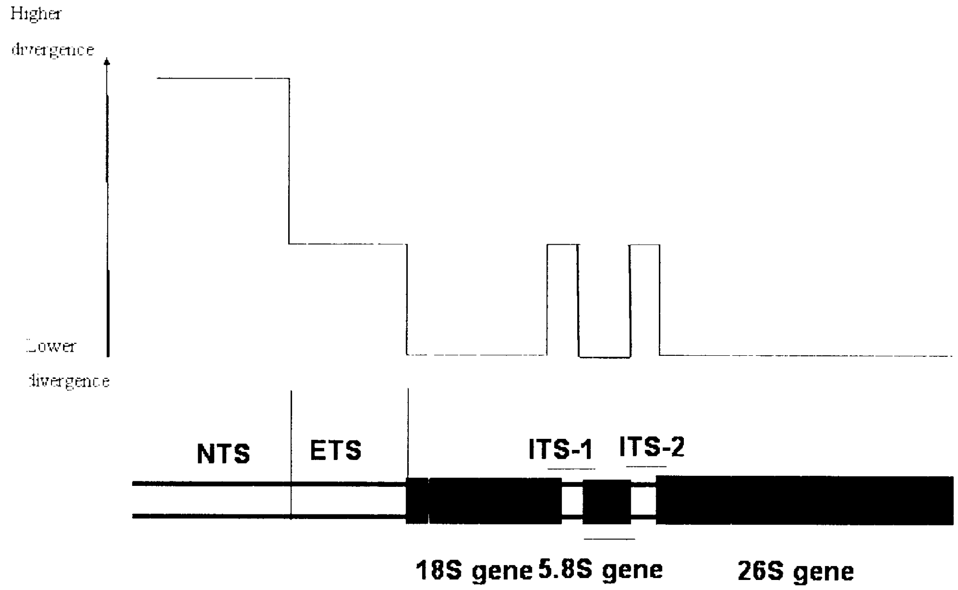

Fig. 2.5. Variable versus conserved sites in the nrDNA. The arrow shows relative degree of divergence (modified from Appels and Honeycut 1986). ETS = external transcribed spacer, ITS = internal transcribed spacer, NTS = non-transcribed spacer.

The rDNA gene copies, consisting of the internal transcribed spacer regions, appear to be arranged into multicopy repeat pieces in specific nucleolar regions that organise these genes, and have been thought to be regulated by influences such as 'concerted evolution' (Hillis and Dixon 1991), thought to be the result of unequal crossing over (Smith 1976) or gene conversion events (Arnheim 1983). However, recent studies have revealed that this 'molecular drive' appears to be less efficient than previously thought: the variation of individual copies in particular cases seems to be now undoubted, and can exceed the variation among species (Karvonen *et al.* 1994; Smith and Klein 1994; Oxelman and Liden 1995). The high sequence difference between the copies of internal transcribed spacers in amplifiable replicates of the genus *Zea* may provide a warning of possible differences between the genes of ribosomal DNA in a particular nucleolar organiser region (NOR) locus. Out of 60 duplicated copies only two have been identical (those belonging to different species!) (Buckler and Holtsford 1996).

In certain instances it does not seem possible to gain normal internal transcribed spacer amplicon harvests under any PCR protocol, although many amplifications of other DNA extractions (e.g. chloroplast DNA; cpDNA) seem to be unimpaired (Moller 2000). A worrying but likely reason could be a partial amplification of the total number of the ITS copies present. There can be many reasons for these incomplete amplifications, for instance a change in one of the PCR primer sites together with imperfect homogenisation by concerted evolutionary forces. Apparently conserved sections flanking ITS1 as well as ITS2 could have huge variation, adequate to permit possible changes in primer sites; e.g. in the 5.8S ribosomal DNA gene, the location of internal primer sites has up to 3.7% difference between *Brassica* species (Suh *et al.* 1992). This has had evidently happened also for the *Alpinia* species, where the specific internal sequencing primer had to be designed (Moller 2000).

The addition of denaturants like dimethylsulphoxide has been recommended to increase the specific amplification of high stability internal transcribed spacer copies as opposed to low stability products (assumed non-functional paralogues); however, this approach was found to provide inconsistent outcomes in competition amplifications (Buckler *et al.* 1997). The solution of using several other primer series in an assemblage of a matrix raises the challenge of accurate homology, since the sequences may arise from differential sampling of ITS gene from the whole gene pool and thus may not evaluate homologous items (Moller 2000).

A disadvantage for ITS, is that sometimes there is not any difference at subfamilial, tribal, and subtribal levels. Even these sequences lack the ability of resolving relationships among very closely related taxa, as has been shown in the case of *Saintpaulia* species, where seven species had identical ITS sequences. Thus, the internal transcribed spacer sequence study failed in resolving the relationships amongst members of the *Saintpaulia ionantha*-complex (Moller and Cronk 1997b). At this level the extent of equal substitutions helps single-copy genes, for instance *Gcyc*, to be phylogenetically more useful (Doyle and Doyle 1999). For example, in Brassicaceae sequence differences between *Sinapis alba* L. and *Arabidopsis thaliana* (L.) Heynh. are 24.3% for ITS 1 and 18.9% for ITS 2 (Rathgeber and Capesius 1989); whereas Asteraceae subtribe Madiinae shows a range of sequence divergence from 0.4 to 19.2%in ITS 1 and 0 to 12.9% in ITS 2 (Baldwin 1992); in Apiaceae subfamily Apioideae from 0.5 to 33.2% in ITS 1 and 0 to33.2% in ITS 2 (Downie and Katz-Downie 1996).

2.7.2.1.3 5S-NTS

The 5S non-transcribed spacer is superior to ITS for inferring relationships of close taxa in various reasons. For instance, as compared to the very high GC content of ITS, the amount of GC in 5S-NTS makes it very easy for amplification. The 5S-NTS provided approximately double the informative data compared with the data obtained from ITS in the genus *Alibertia* (Rubiaceae) studied by Persson (2000). The 5S non-coding region is better than ITS region for resolving clades at the infrageneric group in the *Alibertia* group, and the support for the resolved internal clades is strong in many cases. Likewise the consistency index and retention index are higher in the 5S-NTS analyses (Persson 2000). Indels are usually more frequent in non-coding regions than coding regions and could be useful in finding phylogenetic relationships.

2.7.2.1.4 Value of high copy nuclear regions

There are many reasons for more frequent use of high-copy nuclear regions rather than low-copy genes, apart from methodological considerations or the accessibility of comparable sequences from other data sets. First, low-copy genes have been poorly known. Knowledge of the quantity of copy and the degree of joint evolution between copies might be critical for an affordable within-individual sampling strategy. Second, there may not be adequate sequences on hand across a taxonomic group to offer a preliminary estimate of phylogenetic utility. Primer development appears to be therefore very difficult; also sequencing genes of a large number of taxa is a potentially risky waste of money (Mason-Gamer *et al.* 1998).

2.7.2.2 Other nuclear genes

2.7.2.2.1 *waxy* gene

Granule-bound starch synthase (GBSSI) or waxy gene exists as a solitary copy in almost all species in which it has been studied (e.g. Shure *et al.* 1983; Klosgen *et al.* 1986; Rohde *et al.* 1988; Clark *et al.* 1991; van der Leij *et al.* 1991; Denyer and Smilth 1992; Dry *et al.* 1992; Salehuzzaman *et al.* 1993; Wang *et al.* 1995), although it seems to be duplicated in Rosaceae (Mason-Gamer *et al.* 1998).

2.7.2.2.2 *Chs*

Chs is a single copy nuclear gene that plays a central role in secondary metabolism of flavonoids (Koch *et al.* 2001). It has been revealed to be extremely useful in phylogenetic analyses (Koch *et al.* 2000) particularly at deeper levels in cruciferous plants (Koch *et al.* 2001).

2.7.3 Review of analytical methodology

There has been considerable discussion in the literature on the subject of whether data from different kinds of data sources should be analysed together or separately (Huelsenbeck *et al.* 1996; Persson 2000) Although the relative value of morphological and molecular data continues to be debated, most systematists believe that every character, whatever its category, has a potential for providing researchers with somewhat different information about the species or individual that possesses it (Donoghue *et al.* 1989; Kluge 1989; Doyle 1993). Concerns about dependence of characters within data sets would be the main reason for gathering information from independent sources (e.g. cpDNA and morphology) and for analysing them separately (Swofford 1991; de Quieroz 1993). The number of characters will vary depending on the total homoplasy. The model of divergence among the taxa studied is another factor that is effective in selecting the number of characters (Olmstead and Palmer 1994). Recent analyses indicate the value of combined morphological and molecular analyses (Hansen *et al.* 2000; Muasya *et al.* 2000a; Weiblen 2000).

Analysis of molecular data also can provide a way to compare results of classifications obtained from morphological data in many cases (Yen and Olmstead 2000), the interpretation of which has led to confusing or contradictory suggestions of the evolutionary relationships in the tribe Abildgaardieae. For instance, in some other families molecular analyses place some genera in a classification that is not similar to the classification resulting from morphological data. In these cases, gathering more DNA data including data on additional genera and/or species is suggested or adding results obtained from nuclear genes (Briggs *et al.* 2000). Because of a reduced flower structure and a uniform vegetative morphology in Cyperaceae, polarisation of characters based on outgroup comparison is difficult (Bruhl 1995). Therefore, the suggested ways are to collect as much taxa and characters (either non-molecular, micromolecular, or macromolecular) as possible and analyse them all separately, in groups of characters/taxa, and in combination altogether, then compare the results and find out how reliable (based on support value) they are to get a closer view to what is really going on within a taxonomic unit or among different taxonomic units. This, indeed, does not mean that any sort of radical methods such as process partitioning (Bull *et al.* 1993) must be considered, because it seems absolutely irrelevant to consider that subsets of characters evolve according to different sets of rules. Further, in all the analyses, any selection of outgroup must be delayed until unrooted trees have been reconstructed and based on these unrooted trees, outgroup must be selected, unless there is a high support for a (few) taxa as outgroup in previous studies. Sometimes exploratory experiments using different approaches may help. One possibility might be to take advantage of midpoint rooting, which assigns the root to the midpoint of the longest path between two terminal taxa in the tree. Under the hypothesis that the two most divergent lineages in the ingroup tree have evolved at an equal rate, this method correctly roots a tree without reference to outgroup (Swofford *et al.* 1996). More generally, restructuring a cladogram under an assumption of the molecular clock will also infer the root, although this is a slightly stronger assumption. Alternatively, a variety of outgroups might be tried and assessed for the effect that they might have on the ingroup topology and placement of the root together. If all construct similar result, then there is some confirmation to that root stance. This type of testing has achieved a variety of conclusions ranging between constancy in outgroups (Hutcheon *et al.* 1998; Dalevi *et al.* 2001) and considerable heterogeneity (Milinkovitch and Lyons-Weiler 1998).

For a consistent resolution of the suprageneric taxonomy of the Cyperaceae much more sampling is required (Muasya *et al.* 2000a). Increased taxon sampling will increase the phylogenetic data derived from the sequence in two ways: first, more phylogenetically informative characters will be collected and second, more homoplasious characters will be interpreted than using fewer taxa. With sufficient taxon sampling, the total topology is more likely to be well resolved, given adequate computer time, than when taxon sampling is limited, in which case the erroneous placement of one or more taxa is likely to carry significant phylogenetic implications. Finding the balance involving sufficient data and enough analysis is an important element for designing a study in DNA-based systematics. Systematic results may be modified as more data become available (Yen and Olmstead 2000). The principle of character congruence will assist in determining the non-homologous characters, resulting in increased overall resolution (Olmstead and Palmer 1994). Conclusions based on only one single source have often resulted in major problem at all taxonomic levels (Muasya *et al.* 2000a).

Parsimony analysis is particularly sensitive to the issue of taxon sampling because of the difficulty of long branches attracting on a tree. However, data sets with many taxa present serious computational problems that might result in the inability to achieve maximum parsimony or to find all the shortest trees (Olmstead and Palmer 1994). The generic sample should be as complete as possible to increase the effectiveness of parsimony analysis (Swofford and Olsen 1990) and multiple genes must be sequenced (Muasya *et al.* 2000a) to avoid the problems related to using one gene (Doyle 1993; Nadot *et al.* 1995). On the other hand, the latter approach can make the researcher face other problems dealing with mixed data (e.g. Kluge 1989).

The availability of rapid, high-yielding, and relatively cheap DNA separation methods (e.g. Doyle and Doyle 1987) has been important for molecular systematists who must work on huge numbers of specimens to assess the natural history of taxa (Doyle 1993).

Soltis *et al.* (1998) suggested that one solution to the computational dilemmas caused by big data sets is the increase of nucleotides together with taxa. We should try to get trees as much as we can, using different statistical methods, but remembering that all statistical analyses might have incidental errors (Chase *et al.* 2000). The goal of any systematist, whether working on plants or animals, and whether using molecular or non-molecular data is to exclude all trees but the one that best fits available data, although it is not easy to do even by informative molecular data (Doyle 1993). One of the most important questions could be the number of informative sites that we need to achieve a robust phylogenetic tree (Hilu and Liang 1997).

Single genes may reflect gene trees but investigating more than one gene can show a taxon tree. Although there are instances where data from various genes are not congruent, most of the studies show congruence or at least partial congruence (Smith 2000). In Doyle's (1992) opinion the fundamental difference between gene trees and species trees has been recognised by molecular evolutionists and population geneticists (e.g. Nei 1987) but systematists are slower in accepting this difference (Doyle 1992). Relevant to these issues is also the point that a gene is not a homogeneous complex of bases since various parts along a gene undertake different functions and other sections possibly lack a function. Some sections might represent noise, so are better not considered in the analysis. Section(s) of the gene that are extremely variable might represent regions with saturation of substitution sites and multiple nucleotide substitutions and are, thus, phylogenetically less informative. It is therefore imperative for a gene proposed for use in phylogenetic investigations to be tested thoroughly to find patterns of variability and to assess the usefulness of various sections. To address these points, an investigator must divide a gene to a few sectors of the same size (roughly), and run parsimony analysis on informative sites from each sector. The objective will be to hunt for sectors, which may cause a great deal of distortions in a phylogenetic tree. Molecular systematic studies rely upon data that vary from partial to complete sequencing of genes. Studies have demonstrated the usefulness of partial sequencing in resolving systematic and evolutionary problems from the tribe to the division level (Hilu and Liang 1997). If progressively larger random subsets of the complete data set come together on a tree, it is considered as proof that adequate data were taken. As a general rule, when studying distantly related organisms using analyses that are parsimonious, as wide a range of taxa as possible should be sampled. Taxonomic orders or classes may be "distantly related" for comparisons by *rbc*L whereas families may be distant for researchers that use the *ndh*F gene, and the subgroups of a tribe may be distant for ribosomal DNA ITS works. However, at any level of divergence, sequencing more DNA will help get a sufficient level of character sampling (Olmstead and Palmer 1994).

Doyle (1993) refers to this point that all of this discussion raises the question of how we can recognise the number of nucleotides, genomes, individuals, or taxa that are sufficient to estimate phylogeny with confidence. This question is important considering the high cost of DNA studies, specifically in relation to maintaining financial sources for this field.

2.8 Objective of this study

Here, I estimate evolutionary relationships within Abildgaardieae and between this tribe and its assumed closest tribe(s), Arthrostylideae (and Scirpeae) using two DNA regions (the cpDNA *trn*L–F IGS and *trn*L intron, and nrDNA ITS). This is to reach the aims that have been mentioned in section chapter one (section 1.2). These DNA regions were chosen because:

1. The application of nuclear DNA sequencing analysis to lower-level phylogenetic questions in Abildgaardieae and its genera is one of the objectives of this project and ITS complies this need as mentioned earlier.
2. The ITS region undergoes quick concerted evolution (e.g. Arnheim *et al.* 1980; Zimmer *et al.* 1980; Apples and Dvorak 1982; Arnheim 1983; Hillis *et al.* 1991) via unequal crossing-over and gene conversion. This property raises intragenomic consistency of replicated fragments, in several cases even between nrDNA loci on non-homologous chromosomes (e.g. Arnheim *et al.* 1980; Arnheim 1983; Hillis *et al.* 1991; Wendel *et al.* 1995), and, in general, promotes accurate reconstruction of relationships at various intrafamilial levels from these sequences.
3. In some plant families, ITS sequences have proven very valuable for assessment of inter- and intrageneric relationships (see section 2.4.2.1.1).
4. The small size of ITS (about 700 base pairs in higher plants) and the existence of very conserved bases flanking both ITS1 and ITS2 spacers makes this region easy to amplify, even from herbarium material.
5. The limitations of genes for determining relationships at the lower taxonomic levels have been proved (Kelchner 2000).
6. Substitution bias in the protein coding regions of chloroplast genome, in general, is highly affected by their neighbouring bases, therefore, these region are not more advantageous than non-coding regions anymore (Morton 1997).
7. Non-coding regions face limited selective influence and might evolve at extents far exceeding those of coding regions (see section 2.4.5.1.1).
8. The performance of the *trn*L–F region in resolving intergeneric relationships within Cyperaceae has been satisfactory in the previous studies (e.g. Yen and Olmstead 2000b; Muasya *et al.* 2001a).

9. Both *trn*L–F and ITS regions are easy in terms of amplification in a broad taxonomic range because the 'un versal' primers designed by Taberlet *et al.* (1991) for the earlier and White *et al.* (1990) for the latter are available.
10. While indels are very significant and informative mutations in resolving the relationships among taxa, they are less frequent in coding regions (Golenberg *et al.* 1993).