CHAPTER 3

Phylogenetic relationships within the Abildgaardieae–Arthrostylideae group: *trn*L intron and *trn*L–*trn*F Intergenic spacer

3.1 Introduction

It is not straightforward, for several reasons, to establish a rule for choosing a specific region on cpDNA for resolution of phylogenies (Gielly and Taberlet 1994a). On the one hand, sequencing *rbc*L implies a difficult task that sometimes uncovers little information (Palmer 1987). Therefore, it has limitations for examining related genera, as stated elsewhere for Cyperaceae (Muasya *et al.* 1998), *Hordeum*, *Triticum*, and *Aegilops* (Doebley *et al.* 1990; Gaut *et al.* 1992) and the Asteraceae (Kim *et al.* 1992). On the other hand there are several reasons for not choosing other cpDNA genes/regions as mentioned in chapter 2. In general, the *trn*L–*trn*F (*trn*L–F) intergenic spacer (IGS) and *trn*L intron non-coding region show numerous base substitutions and indels that seem to be very informative at the intertribal level particularly if the indels are scored and considered carefully (Richardson *et al.* 2000).

The single-copy *trn*L–F region of the chloroplast genome contains the *trn*L (UAA) intron, the 3' exon, and IGS between this exon and *trn*F (GAA) (Taberlet *et al.* 1991). This is an appropriate region to use to study the taxa of Abildgaardieae because as mentioned in chapter two, this region is not under selection pressure and also the sequences of this region help to resolve the phylogenetic relationships at lower levels of taxonomy than family level.

Two non-coding sections of the cpDNA were selected for this study: nucleotide sequences of a chloroplast intron and IGS. These data were used to evaluate evolutionary relationships among members of the Abildgaardieae and five species of the closest tribes to this tribe.

I expect this region to address questions about monophyly of Abildgaardieae, monophyly of the genera of Abildgaardieae, evolutionary relationships among these genera, the efficiency of a non-coding chloroplast region in delimitation of monophyletic groups and for classification at different taxonomic levels, and consistency of the previous infrageneric classifications of the genus *Fimbristylis*.

3.2 Materials and methods

3.2.1 Plant material

Taxa from four tribes (38 species including three outgroup species) were analysed using *trn*L intron and *trn*L–F intergenic spacer sequence variation. Leaf cells predictably contain a few thousand chloroplast DNA molecules more than other cells in a plant (Doyle and Doyle 1999). Some DNA samples were collected from fresh material, freeze-dried material, and CTAB-conserved material (saturated brine containing cetyl trimethylamonium bromide / disinfectant; Rogstad 1992 modified by Thomson 2002), but mostly leaves dried by silica gel were used. Tests with different methods of tissue preservation show that quickly dried tissues of plants, if stored in a dry environment, results in a high quality, relatively undegraded cpDNA (Olmstead and Palmer 1994). The identification of voucher specimens was verified by either Karen Wilson or Jeremy Bruhl (my supervisors) except specimens collected by Van Klaphake, which were identified by me and Van Klaphake. Voucher specimens and GenBank accession numbers are listed in Appendix 1.

Outgroup taxa should be taxonomically relatively close to the taxa under study to permit sequence alignment and at the same time distant enough to allow the tree to be rooted easily (Moller and Cronk 1997). In this study, three outgroup taxa, rather than one, were selected for more certain and reliable rooting of trees. One species each of *Eleocharis*, *Bolboschoenus*, and *Schoenoplectus* (all in Cyperaceae tribe Scirpeae according to Bruhl's 1995 classification or Eleocharideae (*Eleocharis*) and Fuireneae (the latter two genera) as defined by Goetghebeur (1986, 1998) were used as outgroup taxa based on the results of Goetghebeur (1986), Bruhl (1995) and Muasya *et al.* (1998, 2000a).

3.2.1.1 Sampling

Sampling was aimed to include a wide range of Australian species within four genera in Abildgaardieae and two genera in Arthrostylideae and 1–3 species of each section of the sections of *Fimbristylis* recognised by Kern (1974), depending on the availability of species within each section, to correspond to a broad range of non-molecular diversity. Because this study aimed to explore generic as well as higher-order relationships, one accession per species was deemed sufficient except where subspecies are geographically separated or morphologically divergent.

However, to examine infraspecific variation, I used two or three samples of several species with a wide geographic range (Appendix 1).

I looked at several samples of some species, which were ecogeographically different particularly if there were even a little morphologic differences to assess variation. Multiple sampling of species in most cases, and particularly in *Fimbristylis* resulted in more stable positioning of species across the trees. Conversely, in *Crosslandia setifolia* this approach to sampling usefully revealed diversity among accessions.

3.2.2 DNA extraction, amplification and sequencing

DNA extractions were done by standard methods (Dellaporta *et al.* 1983) modified by A. Marchant (my co-supervisor at Royal Botanic Gardens, Sydney) (Appendix 5). DNA purification followed the Dicalite method of Gilmore et al. (1993) modified according to A. Marchant (pers. comm.) (see Appendix 6).

3.2.2.1 DNA amplification

Double stranded molecules of *trn*L intron and *trn*L–F intergenic spacer were amplified by the PCR with A50272 and B49317 primers as PCR primers (Table 3.1). Taberlet *et al.* (1991) originally proposed both primers for *trn*L intron and *trn*L–F IGS amplification. The more preserved regions are, the higher chance there is that the primers work in a higher range of taxonomic groups. So the *trn*L–F target sequences are easy to amplify in a different range of taxa such as tobacco, *Marchantia*, and rice using the same primers (Taberlet *et al.* 1991). As accurate amplification of *trn*L-F region for some specimens was not possible using only the primers introduced by Taberlet *et al.* (1991), I made some modifications to previously designed primer AdTabB2 ($^{5'}$AGAGTCCCATTCTACATGTC$^{3'}$, Briggs *et al.* 2000) by adding bases A and C in the 5′ end, removing a C in the middle (underlined) and replacing an A with G (underlined). This was achieved by alignment and comparison of the sequences of the first five species sequenced in this study. The new and modified primer was named AbilB2. Primer sequences are shown in Table 3.1 and Fig. 3.1.

PCR reaction combination for *trn*L intron and *trn*L–F IGS consisted of 5 µl of Bioloine 10% buffer [25 mM TAPS (tris- (hydroxymethyl)-methyl-amino-propanesulfonic acid, sodium salt) pH. 9.3, 50 mM KCl], 2 mM MgCl2, 0.2 mM

dNTPs, 0.1 µl of 5 units Bioline Taq polymerase, 10 µl of each primer (2 µM) and 1 µl total DNA template in a 50 µl reaction volume.

The reaction outline for *trn*L intron and *trn*L–F IGS amplification includes: a primary 94°C for 5 min followed by 30 cycles of 94°C for 30 sec; 60°C for 30 sec; 72°C for 1 min. Double-stranded amplified products of polymerase chain reaction have been purified by the 'Concert Rapid PCR purification system' (Life Technologies Inc., UK) following the producer's specifications. For each cleansed amplification outcome, both strands were cycle-sequenced using an ABI Prism automated sequencer (PE Biosystems, Inc., Foster City, USA) with BigDye Terminator Cycle Sequencing Ready Reaction Kit (PE Biosystems, Inc., Foster City, USA) and the following primers: A50272, B49317, A49855, B49873, ATB1, AbilB2, ATA2#2, ATA3 according to the provider's instruction. I determined the sequences in both directions. This approach produced highly useful sequence data and made it possible for me to get unambiguous categorisations.

Table 3.1. Sequences of the eight primers used for the amplification of two non-coding regions of cpDNA.

The code refers to the 3'-most base pair in the published tobacco cpDNA sequences (Shinozaki *et al.* 1986). The B and A appearing in the code refer to each strand of DNA

Code	Sequence 5'–3'
B49317	GGAAATCGGTAGACGCTACG
A49855	GGGGATAGAGGGACTTGAAC
AdTabB1≅	AAGTGGTAACTTCCAAATC
AdTabA2#2≅	ATTGACATGTAGAATGGGACTC
AbilB2*	ACAGAGTCCATTCTGCATGTC
AdTabA3≅	TTCCGTTGAGTCTCTGCACCTATC
B49873	GGTTCAAGTCCCTCTATCCC
A50272	ATTTGAACTGGTGACACGAG

* Modified primer by the author

≅ Designed and provided by A. Marchant (personal communication)

Failures of polymerase chain reactions to increase the required product were mostly attributed to insufficiency of DNA extraction (e.g. DNA degradation; high salt concentration; precipitation of inhibiting compounds) or PCR conditions (e.g. annealing temperature and time, buffer composition) and were mostly overcome by elimination of these inadequacies. The rest of the failures were due to my practical errors and worked after repeating the experiments.

3.2.3 Alignment and gap coding

Automated DNA chromatograms have been proofed and revised, and contigs were assembled using Sequencher 3.1.1 (Gene Codes Corporation, Ann Arbor, MI). The sequences were truncated to contain only the target region. Identification of the two ends of each non-coding region was based on comparison with other species of Cyperaceae on the GenBank website (http://www.ncbi.nlm.nih.gov).

Multiple alignments of the sequences of chloroplast *trn*L–F region were first obtained using the program ClustalX (Thompson *et al.* 1994) using a gap cost:gap extension cost of 10:5 followed by manual revising of the alignment (Baum *et al.* 1994) and Dialign (Morgenstern 1999) using a threshold T=7 with sensible slight manual modifications to minimise the number of gaps and subsequently adjusted by eye for the intron-intergenic spacer region. DNA sequences were assembled using Sequencher version 3.1.1 (Gene Codes Corporation, Ann Arbor, MI). In positions 410–485 of the *trn*L–F IGS matrix my base assembling was ambiguous or many alternative arrangements were possible. Therefore, these positions have been deleted from my analyses.

It has been proposed that parsimony's potential ability to resolve the topological status of a group of taxa decreases significantly if among-site rate heterogeneity exists in the data (Tateno *et al.* 1994; Kuhner and Felsenstein 1994; Huelsenbeck 1995). Such rate heterogeneity would arise from the structured sequence styles patterns described in Kelchner (2000) in non-coding cpDNA. Although it has been proposed that the reliability of parsimony estimates is strengthened by increasing the number of taxa included in an analysis (e.g. Wakeley 1993; Sullivan *et al.* 1995; Yang 1996), it is unclear if this effect is independent of among-site rate variation (Kelchner 2000).

Parsimony specifies no particular probabilistic evolutionary model; however, like every phylogenetic estimation method, parsimony is affected by non-independence of characters. This problem would be eased to some degree by recognition of mutations

like inversions or inserted and deleted repetitions as conditional events and either exclusion of them from analyses or coding them separately as described below. Any non-independent evolution in adjacent nucleotides of a sequence would generate an erroneous weighting consequence for these positions in a parsimony analysis that considers each base to be an autonomously evolving feature (Kelchner 2000).

Insertions and deletions are believed to have considerable evolutionary value (e.g. Golenberg *et al.* 1993; Mes and Hart 1994; Natali *et al.* 1995; Downie *et al.* 1996a; Kelchner and Clark 1997; Oxelman *et al.* 1997; Sang *et al.* 1997; Liden *et al.* 1997; Downie *et al.* 1998; Bayer and Starr 1998), so one must consider counting gaps as having the status of coded characters added in the sequence matrix (e.g. Hodges and Arnold 1994; Kelchner and Clark 1997; Sang *et al.* 1997; Downie *et al.* 1998; Hoot and Douglas 1998; Bayer and Starr 1998). The exclusion of gaps and elimination of coded gap characters from a non-coding sequence could be an interesting approach to assess the resolution of point substitutions alone (e.g. Kelchner 1996; Kelchner and Clark 1997). I conducted a similar study considering coded gaps only and excluding all other characters in the matrix as Kelchner (2000) has recommended.

Minute inversions must be recognised and deleted from analysis, to be added as present/absent characters at the end of the matrix (Kelchner and Wendel 1996; Kelchner and Clark 1997). This eliminates possible scoring of frequent non-homologous similarities that might be artefacts of the inversion mutation (Kelchner 2000). There were no inversions in my data.

3.2.4 Phylogenetic analysis

Bootstrap (Felsenstein 1985) and jackknife (Farris *et al.* 1997) analyses, have both been misunderstood as showing the direct phylogeny whereas they actually estimate statistical error in situations where the underlying sampling distribution is unknown or difficult to derive analytically (Efron 1982; Efron and Gong 1983). Since it is impractical to sample repeatedly from the underlying distribution of taxonomic characters itself, these methods offer useful ways to approximate the distribution by resampling from the original dataset. They are basically as consistent as the underlying study procedures. With the same character sampling, using more taxa in an analysis is expected to result in lower bootstrap values (Olmstead *et al.* 1992; Chase *et al.* 1993; McDade *et al.* 2000; Meimberg *et al.* 2000a; Muasya *et al.* 2000a; Persson 2000; Yen and Olmstead 2000).

To confirm the results of bootstrap and jackknife analyses a non-resampling method is needed to permit assessment of data support for the individual clades. The Bremer Support measure (BS; also known as decay analysis; Bremer 1988, 1994; Donoghue *et al.* 1992; for application to large data sets, see Baum *et al.* 1994; Morgan 1997) is a method that has this advantage. Bremer support enables the user to sidestep the effect of character definition issues discussed above if the model used in the phylogeny estimation considers the variable character definition in the sequence. BS values are not viewed being probabilistic assessments themselves (Oxeman *et al.* 1999). A failure to match the quantities to an accepted scale, which is commonly applicable, causes the method to be mistrusted by some systematists. However, the improvement in method by Oxelman *et al.* (1999) has increased comparative information capacity of the measure. This modification includes minimum branch length values accompanied by each Bremer Support value and results in the non-standard mode. This procedure may make more sense and give more information than bootstrap and jackknife values for non-coding cpDNA data (Kelchner 2000).

Values of consensus efficiency that go over 1 show that a consensus tree is wrongly interpreted as it allows fewer cladograms than should actually appear. For example, as Nielsen *et al.* (1996) explained, there is an Adams consensus of evenly parsimonious cladograms for metazoan phyla (their Fig. 1) as if it was a strict component consensus tree (Wilkinson and Thorley 2001). Adams consensus trees are developed by repositioning those groups that appear in inconsistent positions on different fundamental cladograms to the nearest node they have in common. Adams trees accordingly contain every intersecting taxon common to each fundamental tree in all given series of trees.

Where the strict component consensus is employed for source trees of different groups of data, lack of useful outcome may be explained as resulting from strong incongruity between those trees and even between all the separate data sets (e.g. Jenner and Schram 1999).

These steps were undertaken after PCR (mostly based upon Doyle 1993):

1. Finding the sequence alignment. The 'Sequencher 3.1.1' software package of Gene Codes Co. was used to edit and proof the automated DNA sequencing chromatograms. This software also has the ability to assemble contigs. Sequences of the whole *trn*L-F region (*trn*L intron, *trn*L-*trn*F IGS) were

constructed from the output files (electropherograms), with the aid of this program. Accurate sequence alignment is absolutely vital as it sets the character homologies on which the whole study is based. ClustalX (Thompson *et al.* 1994) and Dialign (Morgenstern 1999) were used for sequence alignment.

2. Making character choices and recognising character states. Some DNA characters are expected to be less informative compared to others, usually because these characters experience many changes, which obscure original mutations, blurring character relationships, and causing underestimates of the branch lengths. I either excluded these characters entirely from analyses or down-weighted them using weighting methods using 'successive approximation weighting' (SAW).
3. Choosing analytical method. There are several major classes of methods for inferring a phylogenetic tree, each with its own strengths and weaknesses. These are: parsimony, distance, and maximum likelihood (ML) methods (reviewed in Swofford and Olsen 1990; Felsenstein 1988). The most parsimonious trees are sought, i.e. those that minimise the number of character state changes or tree length. Parsimony has become the method of choice for most DNA-sequencing phylogenetic works in plants (Olmstead and Palmer 1994). ML has many useful characteristics (Felsenstein 1988), but as a methodological option for obtaining the tree with the best estimation of the most reliable relationships, the number of taxa that can be included in an analysis restricts this method. With large data sets, maximum likelihood may play an important role in comparing alternative cladograms (Doebley *et al.* 1990) derived from another phylogenetic analysis method, that is, comparing equal–length trees found by parsimony analysis (Olmstead and Palmer 1994). The emergence of Bayesian methods during the time of my study has overcome some of the processing time limitations of ML. Distance matrix methods can handle larger data sets in a fraction of the time needed for parsimony analysis, suggesting the worth of distance matrix analyses for investigating large data sets. However, philosophical objections to this approach are strong.
4. Assessing confidence in the trees obtained. Bootstrap and jackknife

approaches re-sample the data set to arithmetically measure how robustly the data support a particular topology (Kitching *et al.* 1998). Bootstrap is the most frequently used method for assessing the results of a parsimony analysis of sequencing data (Olmstead and Palmer 1994). The phylogenetic bootstrap method (Felsenstein 1985) has many proponents and is now so widely employed that this is the minimal necessity for a phylogenetic study (Penny and O'Kelly 1991). Bootstrapping methods are a common set of approaches for assembling pseudoreplicate information in circumstances where proper resampling could be impractical if not impossible. The name of this method refers to pulling one's self up by the bootstraps in this statistically difficult situation. The pseudoreplicate data sets formed by this method allow the investigator to determine if the stochastic consequences are probable to influence the findings (for phylogenetic context, the branching order of the tree). In parsimony, a useful index of support for a monophyletic group may be obtained by calculating the difference in tree lengths between the shortest trees that contain versus lack that group (Bremer 1988). This statistic is referred to as the 'decay index' (Donoghue *et al.* 1992) or the 'support index' (Bremer 1994). A difficulty with the decay index can be that it is not clear how large a value must be for the group to be considered well supported. The bootstrap and Jackknife methods differ in the approach of resampling. In the former, data are picked randomly, while replacement is being achieved, from an original data set until another data set with the same number of observations is obtained. Thus, some data will be excluded in bootstrapping; others will be included at least once. For each resampling, the value of interest will be computed. The jackknife, however, resamples an original data set by reducing a number of data at a time and calculating the statistic from the rest of the observations (see Miller 1974).

In this thesis, I studied the concept of phylogenetic construction from two viewpoints. The first is intended to find the level of confidence for my most parsimonious cladogram in general, while the other examines individual support assigned to each clade taking part in the most parsimonious cladogram and asks which clades are consistently strongly supported and which are only weakly supported.

The aligned *trn*L–F sequences were analysed using PAUP* version 4.0b10 (Swofford 2001). The phylogenetic relationships within the Abildgaardieae–d Arthrostylideae group using their *trn*L intron and *trn*L–F IGS data were analysed using three methods: neighbor-joining (NJ) (Saitou and Nei 1987), maximum parsimony (MP), and maximum likelihood (ML). The analysis was run under the following conditions: HEURISTIC (approximate searching method, which generally use 'hill-climbing' techniques of search), TBR (tree bisection and reconnection branch-swapping method, which clips the cladogram into two or more subcladograms and then re-roots subcladograms before reconnecting them to each branch of the remnant main cladogram), and STEEPEST DESCENT (ACCTRAN or accelerated transformation mode; 100 random addition cycles) due to the size of the data set under study. For each couple of species, I performed the ratio of observed mutational events employed by O'Donnel (1992): ratio of observed mutational events = [(TS + TV + ID) / L] % 100, where TS = number of observed transitions, TV = number of observed transversions, ID = number of observed insertions / deletions (multibase length variations are considered as 1), and L = chain length (TS + TV + ID + number of sites showing the same nucleotide). To evaluate the evolutionary utility of all insertions/deletions, two distinct tests were performed, based on nucleotide replacements and both base substitutions and insertions/deletions. In the second case, both mutational outcomes (replacements and insertions/deletions) were programmed in the matrix of unordered multistate characters, the base stretch matching to an insertion/deletion or two overlapping insertions/deletions being considered as a single site.

The bootstrap option (Felsenstein 1985) of the program and also decay analysis (Donoghue *et al.* 1992) were used to measure relative support in the unweighted analysis. For the parsimony analyses, I performed heuristic search. Weighted parsimony analysis was performed as indicated by the norm of Albert and Mishler (1992) following estimation of a transition/transversion proportion (using a few commands within PAUP* 4.0b10 version: DSET DISTANCE = ABS SUBST = TRATIO; SHOWDIST, SAVEDIST FORMAT = TABTEXT FILE = KIOUMARS121). Gaps were introduced in my data sets to align the sequences and were indicated separately in the same way as binary characters (Fig. 3.2) and all characters and character step conversions were weighted equally. This means that each indel was given the same weight as a single base change, as well as the same

weight as each other. For the neighbour-joining method (Saitou and Nei 1987), the bootstrap support for each clade was estimated based on 1000 replicates (Felsenstein 1985).

Heuristic tree searches were conducted employing the Fitch criterion (unordered, equal weights; Fitch 1971) with 1000 replicates of tree bisection-reconnection (TBR) swapping, but permitting only 10 trees to be held each time saving time in tree swapping. Successive approximation weighting (Farris 1969) was performed using the rescaled consistency index until an equal shortest tree length was obtained in two successive rounds (Fig. 3.8).

Internal support was estimated through 1000 replicates (Felsenstein 1985). Trees were reconstructed using simple taxon addition and TBR swapping, retaining groups with more than 50% support in the final bootstrap consensus tree. I use these descriptions to categorise bootstrap values: weak, 50–70%; moderate, 70–80%, robust, 80–100%.

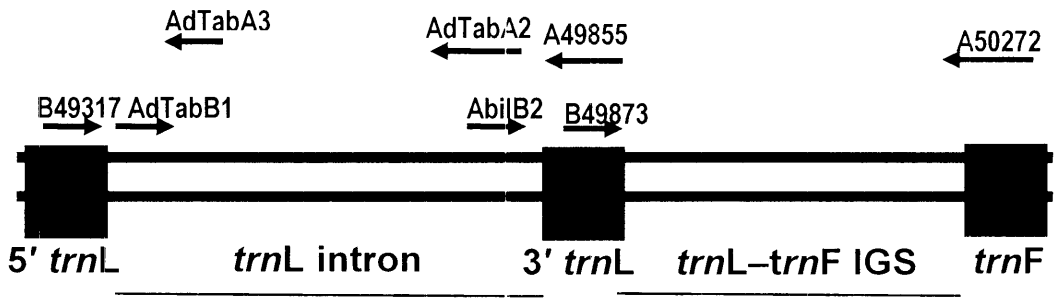

Fig. 3.1. The primers used in this study to amplify *trn*L–F region.

Decay analysis is a method of assessing strength of clades on a tree through searching for all trees of successively greater lengths than the most parsimonious to determine how much longer a tree must be to dissolve the ambiguity of the strength of the clades (Donoghue *et al.* 1992; Hamby and Zimmer 1992; Doyle and Doyle 1993). Decay indices for individual clades were obtained by comparing the strict consensus of all equal-length trees, using SIMPLE addition sequence and TBR in PAUP. Correct estimation of branch lengths linking species is a necessity not only in some procedures of evolutionary reconstruction but also to use molecular data to date evolutionary events (Doyle 1993).

The Hasegawa-Kishino-Yano model with rate heterogeneity was used for maximum likelihood (ML) analysis with a few taxa selected from different clades of the parsimony analysis to see if the results confirmed the results obtained from the other analyses. The ML model was optimised to the data set (transition/transversion ratio estimated; base frequency estimated; percentage of constant sites estimated; variable sites set as independent approximation to the gamma distribution with the shape parameter estimated).

3.2.5 Measures of character fit and character weighting

Standard measures of the cladograms were obtained from the PAUP analyses including tree length, consistency index and retention index.

```
Eleo. cylindro.    ::::::::::AATATATAT::ATTT:ATTTATAAAT:::AATATATAA
Schoen. litoralis  ::::::::::::::::::::::::::::::::::::::::::::::::
Bolbos. caldwellii:::::::::::::::::::::::::::::::::::::::::::::::::
Abild. oxystachya  :TC::::TTAAATATATATTAA:GAATATT::::::::::::ATATTT
Abild. vaginata    :AT::::TATA:TATATATATATTATT:::::::::::::::::::
Abild. schoenoides:TC::::TTAAAAATATATTAA:G::::::::::::ATGAATATATAT
Cross. setifolia   :AT:::::::::: TATATATATTAAT:::::::::::::::::::
Fimb. cephalophora:TT:::::::::: ::TATCTATTAAAAAAAAATAGATATATATATAT
Fimb. cinnamom.    :AT:::::::::: ::::::::::::::::::ATATATATATATATAT
```

Fig. 3.2. Alignment of a small part of the intergenic spacer of the *trn*L–F region for the six ingroup species and the three outgroups showing indels. Ambiguous alignment of bases 410–483 of IGS region has been deleted. 86 gaps were scored as binary characters and added to the end of the data matrix (these ambiguous gaps in *trn*L–F IGS region were not scored). *Eleo. cylindro.* = *Eleocharis cylindrostachys*; *Schoen.* = *Schoenoplectus*; *Bolbos.* = *Bolboschoenus*; *Abild.* = *Abildgaardia*; *Cross.* = *Crosslandia*; *Fimb.* = *Fimbristylis*; *cinnamom.* = *cinnamometorum*.

3.2.5.1 Character weighting

Character weighting can potentially increase resolution of relationships among taxa. This approach also has appeal as fewer most parsimonious trees are usually produced.

Kitching *et al.* (1998) recommended two possible types of character weighting:

1. Characters weighted equally—As all the equally most parsimonious trees are similarly 'true', we should not try to choose between them, although

there is only one 'true tree' (if evolution has been 'well-behaved', with dichotomous speciation and no inter-specific hybridization).

2. The analysis could be repeated incorporating some justified approach to character weighting (giving more weight to some characters than the others, for example weighting characters based on mutation frequency); to obtain one or fewer shorter trees.

There are many different consensus methods and various contexts in which they may be used (Swofford 1991; Wilkinson 1994; Leclerc 1998). Majority-rule consensus trees (Margush and McMorris 1981; Wilkinson 1996) can be useful when we use bootstrap or jackknife analyses but can be problematic when used for representing a group of evenly optimal cladograms from an analysis of a particular data collection (Wilkinson and Benton 1996). Strict consensus trees give information through permitting or prohibiting a subgroup of likely trees (Page 1992; Wilkinson 1994; Thorley *et al.* 1998).

In general, a consensus cladogram is less resolved than any of the equally most parsimonious cladograms, and so is regarded as a more uncertain summary of our data compared with any of the original trees, although it is less likely to be positively wrong, as at least all but one of the equally most parsimonious trees must be. Because of the uncertainty of the consensus tree, Carpenter (1988) disapproved of consensus mode as a final extraction and recommended that before consensus cladogram construction, character weighting be undertaken. Although I agree with weighting some characters in favour of a bias to the characters that are obviously distinct features of a taxon or a group of taxa (for example large indels or distinct synapomorphies), I disagree with Carpenter's view about the consensus mode. Even with character weighting, the likelihood of a consensus of all trees resulted from a weighting approach to be closer to the 'true tree' is theoretically much more than any other trees except one, which is the so called 'true tree', although we would not know if it is practically always the case. This does not mean that the consensus tree is a phylogeny, it is just a summary of how well (congruent) the original trees agree. All original trees are equally probable and none has privilege to others. On the contrary, the consensus tree has all the information gathered on one tree and although it is mostly less resolved than each single tree, in cases where weighted characters are just slightly different from other characters,

the resulting trees will bias in favour of selective advantages given to some lovely characters chosen by the researcher rather than by natural selection. This, I believe, is another argument for the reliability of the consensus mode.

According to Kitching *et al.* (1998), weighting of data has been split into *a priori* and *a posteriori* approaches, that is, weighting before or after the cladogram construction. While *a priori* weighting is hypothesis- and cladogram-independent, *a posteriori* weighting is hypothesis and cladogram dependent. In other words, these two categories are distinguished by the fact that the *a posteriori* methods forbid distortion of the sister group relationships depicted in the phylogenetic trees. *A priori* methods, by contrast, allow pruning and adding of taxa and taxon relationships and thereby may alter the sister group relationships depicted in the cladograms.

For *a priori* weighting, two contradictory approaches have been suggested: "character analysis" and "character compatibility". There are two modes to *a posteriori* weighting too, both based on the models of cladistic consistency: successive approximations weighting (Farris 1969, 1989), (henceforth called 'successive weighting'); and 'implied weighting' (Goloboff 1993). While there are more approaches for weighting characters, these have the greatest relevance to cladistic practice (Kitching *et al.* 1998). In *a priori* weighting, character analysis means the re-evaluation of initial data to find out whether any mistakes have happened, such as inappropriate coding or the wrong way of interpreting homology (similarity) (Kitching *et al.* 1998). Molecular data show a nonflexible number of character states: nucleotide data are symbolised by merely four features (the nucleotides) plus probably another one for 'gaps'. Hence, homology is more easily discovered than in morphological data, as permutations in the characters and differential weighting of any practical 'imbalances' regarded as worthwhile can be calculated precisely (Stevens 2001 onwards). I tried all above-mentioned possible weighting options.

3.3 Results

Successful double-stranded cpDNA amplifications were obtained from all species sampled (Appendix 3). Addition of a couple of primers to a PCR reaction, intensifying the *trn*L–F IGS and its next intron on cpDNA (*trn*L) using primers

A50272 and B49317 (Taberlet *et al.* 1991) resulted in a strong band in the species of Abildgaardieae and outgroups (Fig. 3.3).

3.3.1 Characteristics of the *trn*L–F intergenic spacer region and the *trn*L intron

Originally it was intended to analyse the trnL intron and *trn*L–F spacer regions separately, but because the intron was so highly conserved, and its tree was largely unresolved, it was decided to combine these two regions to maximise the information available. The length of the *trn*L–F IGS section and the *trn*L intron sequences obtained in this study varied from 942 bp (*F. sericea*) to 2301 bp (*F. polytrichoides*). The aligned sequences after deletion of the unmatched sequences of *F. polytrichoides* were 1952 bp. In the IGS region, the nucleotide sites varied (333–486 bp) due to a large number of indels. *Fimbristylis polytrichoides* has the largest *trn*L–F IGS ever recorded with 1600 base pairs. This bit is neither a duplication nor insertion from a known region on cpDNA or nuclear DNA. It is located between basepairs 1817 and 1818 in the original alignment of 3100 bp (before cutting this insertion). *Abildgaardia vaginata* and *Crosslandia setifolia* (K.L. Clarke 115) had a variant dinucleotide repeat [d(A–T)]n/[d(T–A)]n for that region (Table 3.2). The nucleotide sequences of the *trn*L intron include a very variable section at 410–490.

The mean nucleotide makeups of my *trn*L intron and *trn*L–F IGS fragments were 0.36 (A), 0.31 (T), 0.16 (C), and 0.17 (G). These averages were 0.27 (A), 0.41 (T), 0.15 (C), and 0.17 (G) in *trn*L–F IGS and 0.35 (A), 0.31 (T), 0.15 (C), and 0.19 (G) in *trn*L region. While the transition:transversion ratio was 1.74, it was 1.88 for the spacer region and 1.70 for *trn*L. The total G/C content was 33%.

Of the 1952 bp of an aligned *trn*L–F region 1292 positions were unvarying (68.61%); of a remaining 578 variable positions 328 positions (17.42%) were potentially parsimony informative. *Fimbristylis polytrichoides* has an additional large segment (all extra basepairs contiguous) that has apparently derived from somewhere else on the cpDNA but the rest of this large segment did not match with any existing segments at GenBank.

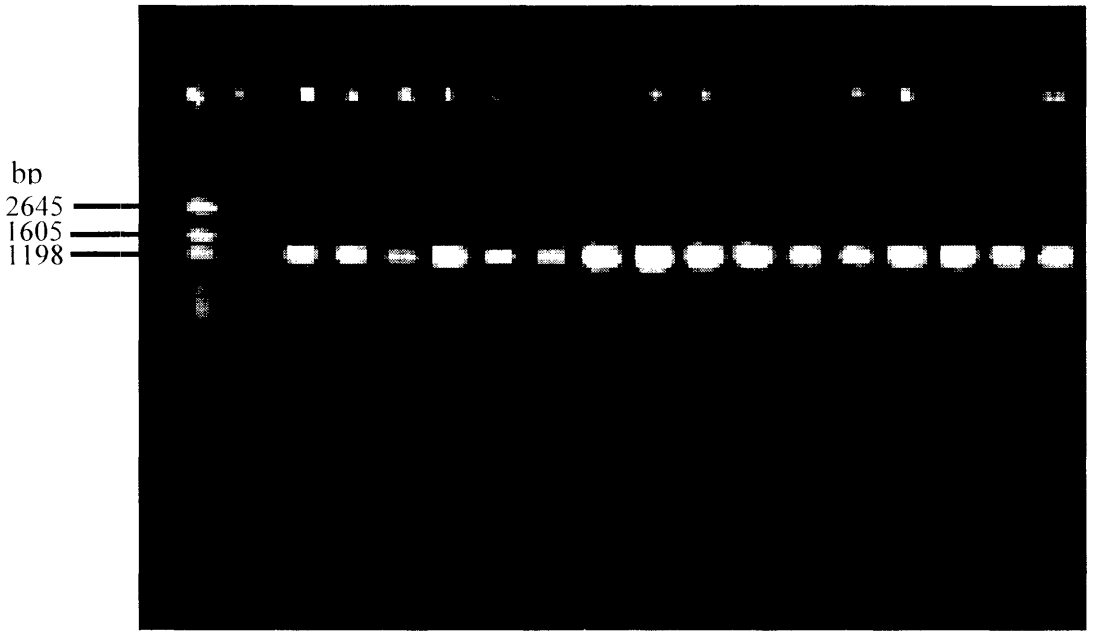

Fig. 3.3. Simultaneous PCR amplification of chloroplast DNA (*trn*L–F IGS and *trn*L intron) used for *Fimbristylis bisumbellata* (lanes 3 and 4), *F. punctata* (lanes 5, 6, 9, and 10), *Abildgaardia schoenoides* (lanes 7 and 8), *Crosslandia setifolia* (lanes 11 and 12), *F. microcarya* (lanes 13 and 14), *F. pterygosperma* (lanes 15 and 16), *Abildgaardia pachyptera* (lanes 17 and 18), *A. ovata* (lanes 19 and 20). Negative control is in lane 2. Lane 1 is a Promega 'pGEM DNA Markers' (catalogue number: G1741).

3.3.2 Phylogenetic analysis

A preliminary phylogenetic analysis was done via sequencing information of *trn*L intron and *trn*L-F IGS regions, with *Abildgaardia oxystachya*, six *Fimbristylis* species, and two *Bulbostylis* species; in addition, two species from Arthrostylideae (*Arthrostylis aphyla* and *Actinoschoenus* sp.) and one from Scirpeae were selected as outgroup taxa. The trees were 652 steps in length (CI = 0.76; CI without ignoring uninformative characters = 0.65; RI = 0.77). The *trn*L–F data was distributed in two islands. The strict consensus cladogram and its bootstrap quantities for separate clades are shown in Fig. 3.4. The genera *Arthrostylis* and *Actinoschoenus* constitute a clade with 100% support in bootstrap analysis.

To ascertain the functional outgroup of the tribe Abildgaardieae, the initial analyses included representatives of all tribes Scirpeae, Eleocharideae, plus Arthrostylideae. The *trn*L–F data establish representatives of Scirpeae and Eleocharideae as the functional outgroups for the tribe Abildgaardieae. Highest

resolution was achieved with base substitutions plus indel matrix.

Table 3.2. Number of the TA/AT simple-sequence replicates appearing in the *trn*L–*trn*F IGS among species of Abildgaardieae

Species	Voucher	Locality	Replicates
Abildgaardia ovata	V. Klaphake 4	NSW	9
A. pachyptera	K. L. Clarke 181	NT	5
A. schoenoides	K. L. Clarke 70	QLD	9
A. vaginata	J. J. Bruhl 2057	NSW	Variable
Crosslandia setifolia	K. L. Clarke 115	WA	Variable
Fimbristylis cinnamometorum	J. J. Bruhl 2058	NSW	10
F. neilsonii	K. L. Wilson 10051	WA	4

Figure 3.5 presents multiple alignments of my *trn*L intron base arrangements of ten species of the Abildgaardieae–Arthrostylideae group showing indels as homologies among the members of these two close tribes. A few deletions including several nucleotides occur at the similar position of the sequences of some species, implying that those events would be regarded as evolutionarily informative. Using substitutions only, a heuristic search calculation of PAUP* resulted in 280 most parsimonious trees. One of these trees is shown in Figure 3.6 (consistency index 0.67; tree length 1273, and 1010 ignoring autapomorphies). Strict consensus of these trees is a highly resolved tree at the intertribal level but shows a polytomy among the species of *Fimbristylis*, *Abildgaardia*, and *Crosslandia* within Abildgaardieae. However, Arthrostylideae are nested between *Bulbostylis* and the rest of the Abildgaardieae (Fig.3.7).

Successive weighting using indels in addition to base substitutions, in general, confirmed the main branches particularly at the genus level (Fig. 3.8). The *trn*L–F phylogeny for species of Abildgaardieae has two main clades (I and II). Clearly, most of the range of morphological variation in Abildgaardieae is accommodated by species represented by clade I. Clade II, in contrast, comprises species assigned to genus *Bulbostylis*. Current generic delimitations between *Fimbristylis* and

Abildgaardia are not confirmed by *trn*L–F data indicating that these genera may be paraphyletic. Also, maximum likelihood analysis of selected taxa confirmed the maximum parsimony results (Fig. 3.9). The maximum parsimony and maximum likelihood trees mainly differed in the placement of *Fimbristylis littoralis*. *Bulbostylis* was never placed inside the rest of the Abildgaardieae *sensu stricto* clade by any of the methods or trees inspected

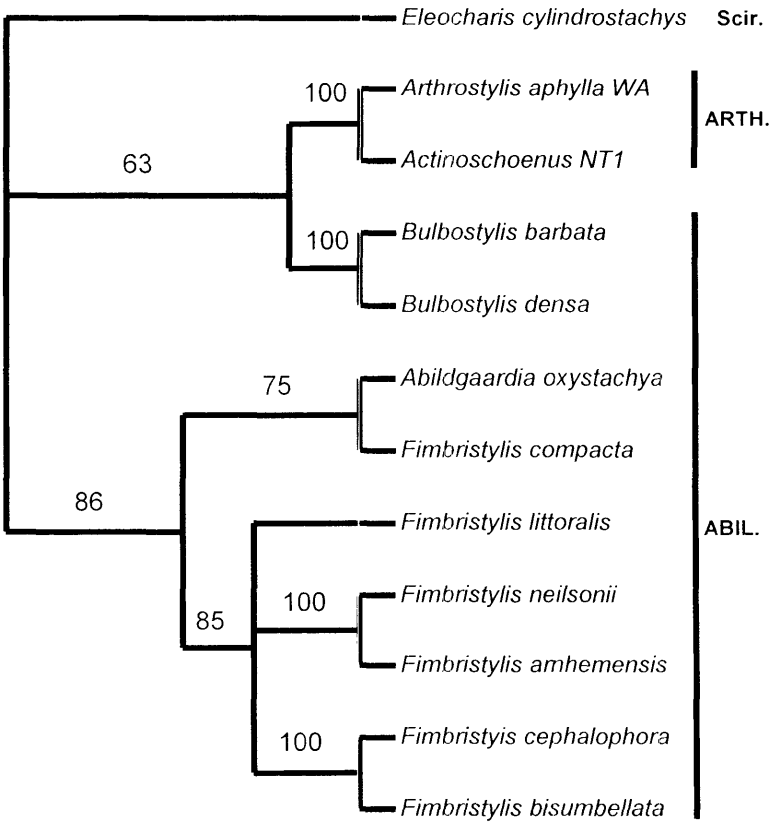

Fig. 3.4. Strict consensus tree derived from 1000 most parsimonious trees resulting from a preliminary study of the *trn*L–F region sequences by unweighted Fitch parsimony. Percentage bootstrap values are given above branches. Abbreviations denote non-Abildgaardieae and Abildgaardieae taxa sensu Bruhl (1995). These are ABIL = Abildgaardieae; ARTH = Arthrostylideae; and SCIR = Scirpeae.

The largest clade (Clade I: bootstrap 100%, decay value of 13) includes all sampled species of *Fimbristylis*, *Abildgaardia*, and *Crosslandia* (Fig. 3.8). Within this clade, 12 separate sublineages form an unresolved polytomy in the strict consensus tree (Fig. 3.7). One sublineage encompasses 12 species of *Fimbristylis*, supporting the

monophyly of these species. Within this group, the close relationship of *F. lanceolata* and *F. cymosa* on one hand and *F. cephalophora* and *F. sericea* on the other hand is noteworthy. *Crosslandia setifolia* together with *Abildgaardia vaginata* showed a sister group relationship to the species of *Fimbristylis* in all most parsimonious, maximum likelihood and neighbour-joining trees, but this resolution collapsed in the strict consensus tree.

```
Arthr. aphylla   -----AGCCTAACAAAAAATTGATTCAA----------AAAAAAA
Abil. pachyptera -----AGCCTAACAAAAAATTGATTCAAAAAAAAAAGTCAAAAAAA
Bulb. barbata    AATTGAGCCTAACAAAAAATTGATTCAA----------AAAAAAA
Cross. setifolia -----AGCCTAACAAAAAATTGATTCAA----------AAAAAAA
Fimb. dichotoma  ----------AACAAAAAATTGATTCAA----------AAAAAAA
Fimb. polytri.   ---------AATCAAAAAATTGATTCA-----------AAAAAAA
Fimb. velata     ---------AATCAAAAAATTGATTCAA----------AAAAAAA
Fimb. sieberiana -----AG-----CAAAAAATTGATTCA-----------AAAAAAA
Fimb. schultzii  -----AGCCTAACAAAAAATTGATTCAAAAAAAAAAGTCAAAAAAA
Fimb. tetragona  ------------CAAAAAATTGATTCA-----------AAAAAAA
```

Fig. 3.5. Part of the nucleotide arrangements of the *trn*L intron, in eight species of Abildgaardieae and Arthrostylideae. Abreviations: *Arthr.* = *Arthrostylis*, *Abil.* = *Abildgaardia*, *Bulb.* = *Bulbostylis*, *Cross.* = *Crosslandia*, *Fimb.* = *Fimbristylis*, and *polytri.* = *polytrichoides*. A dash indicates the absence of a base.

The grouping of Arthrostylideae into a single clade corroborates the views of Goetghebeur (1986) and Bruhl (1995) about a monophyletic origin of this tribe.

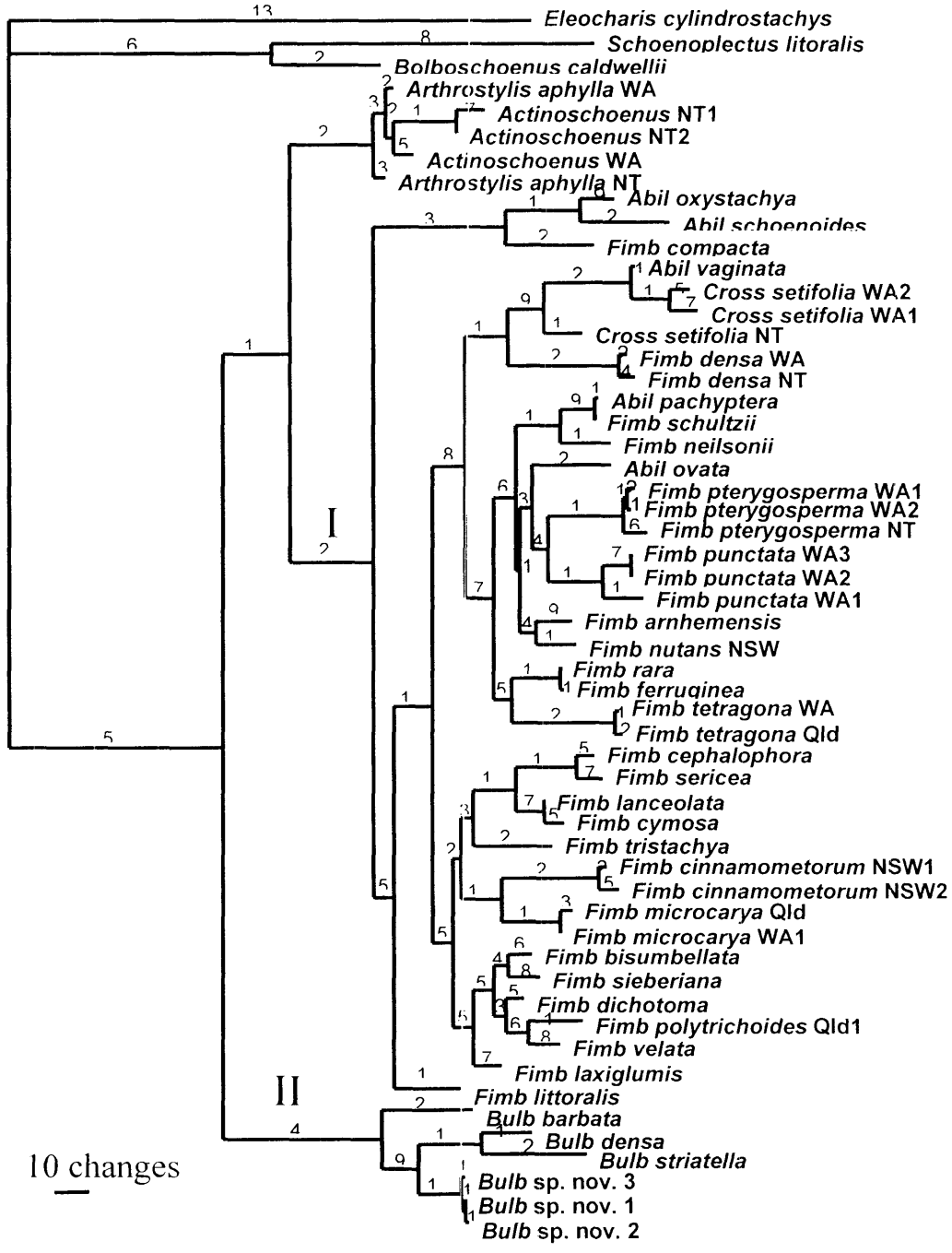

Fig. 3.6. Phylogram of one of the most parsimonious trees for a range of taxa from Abildgaardieae and Arthrostylideae by a maximum parsimony approach based on DNA sequences obtained from the *trn*L–F intergenic spacer region, and its next intron (*trn*L) estimated using 1000 replicates. Number above each branch shows the branch length. *Eleocharis* rooted this tree. Tree length is 1273. Consistency index is 0.667 (0.565 without uninformative data). Retention index is 0.767. WA = Western Australia; NT = Northern Territory; NSW = New South Wales; Qld = Queensland. I and II are the main clades of Abildgaardieae.

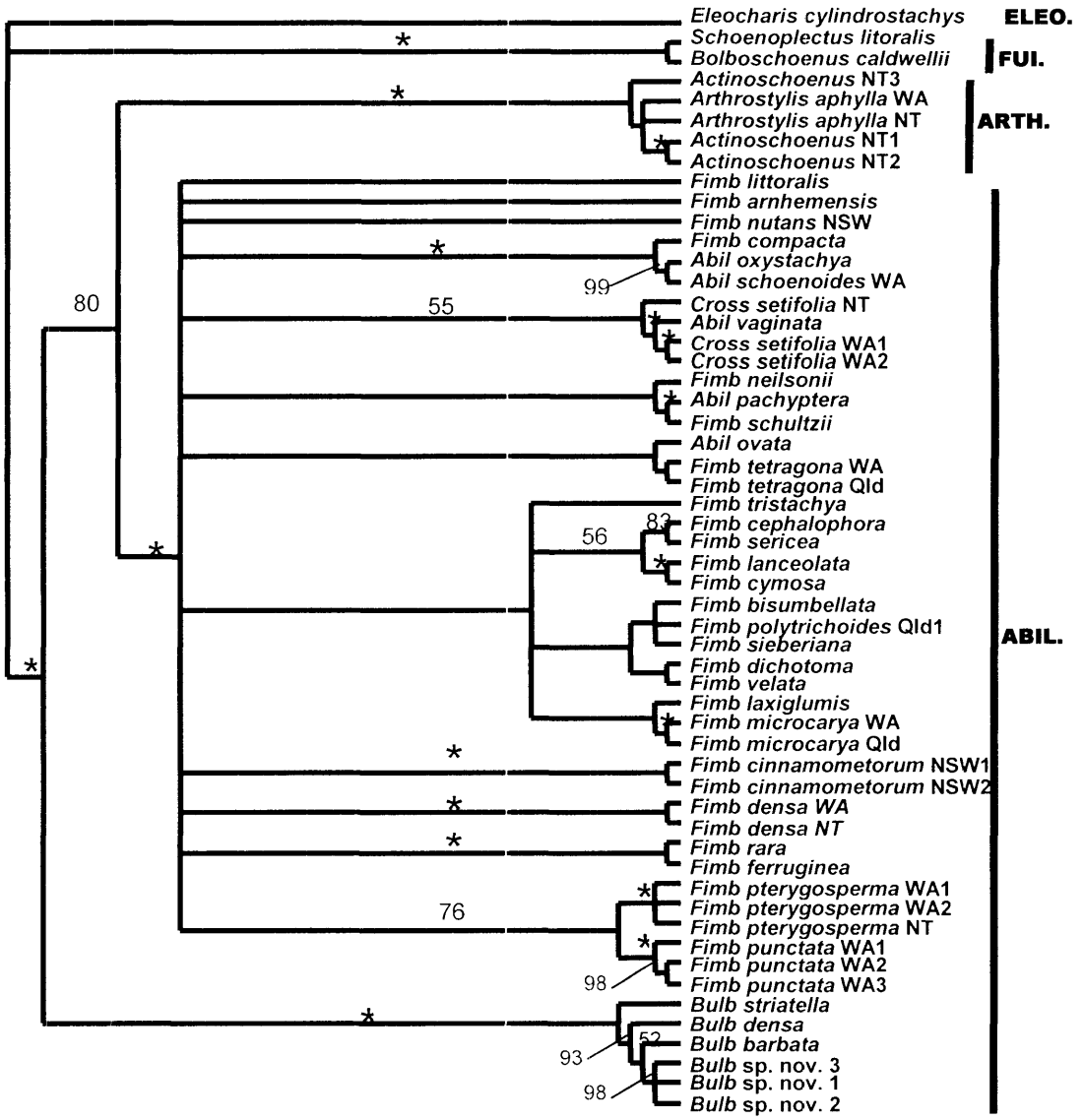

Fig. 3.7. Strict consensus tree derived from 280 most parsimonious trees resulting from study of *trn*L intron and *trn*L–F IGS base arrangements. Numbers above branches are bootstrap values over 50%. Abbreviations denote tribes of Cyperaceae sensu Goetghebeur (1986, 1998) because of the more tribes recognised in his studies: ABIL. Abildgaardieae, ARTH. Arthrostylideae, FUI. Furineae, and ELEO. Eleocharideae. WA = Western Australia; NT = Northern Territory; NSW = New South Wales; * represents 100.

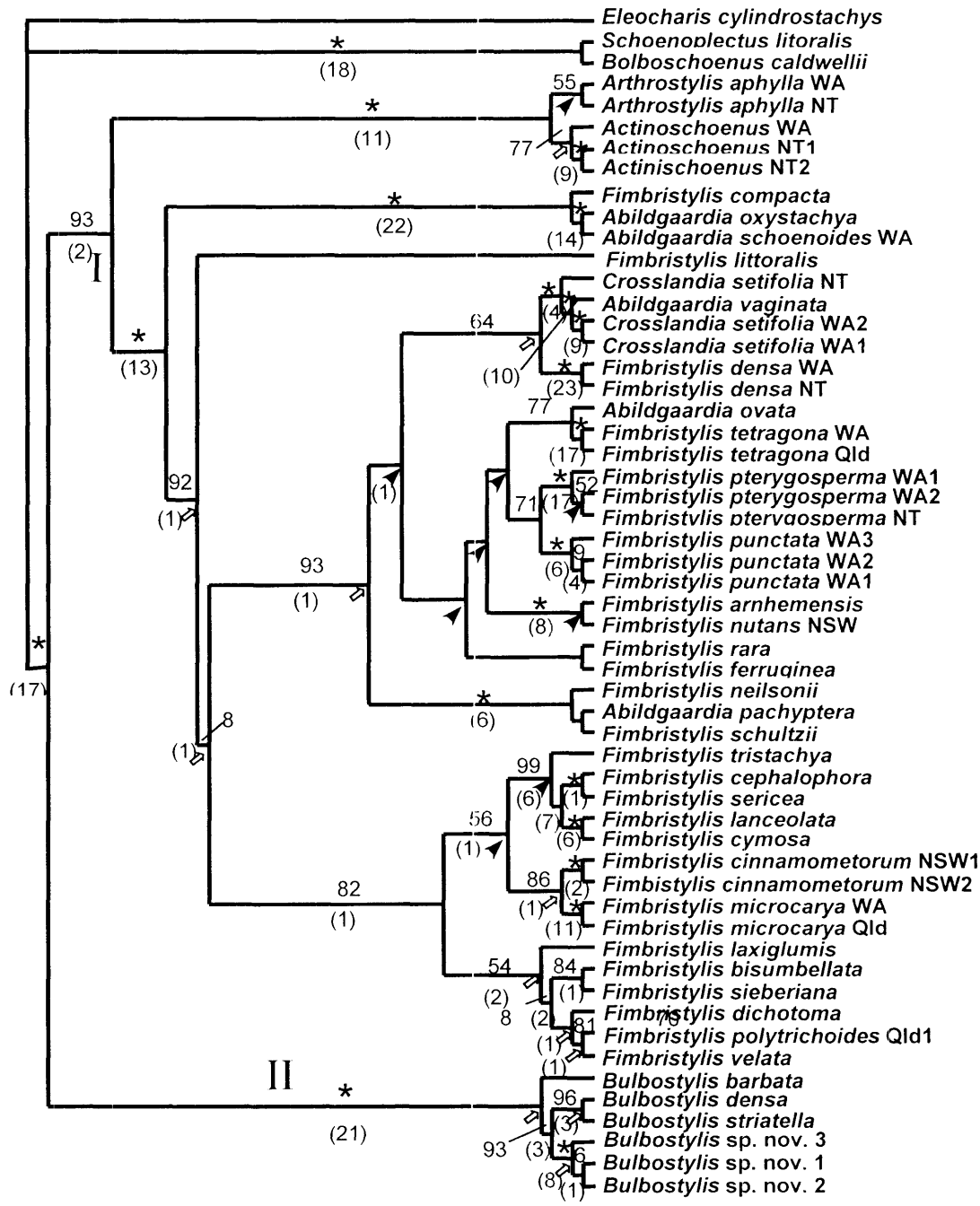

Fig. 3.8. One of the 9 cladograms found by successive approximation weighting. Arrows mark clades absent in the strict consensus tree of the Fitch (open ones) or both Fitch and SW (solid arrows) analyses. Numbers above these branches show SW bootstrap percentages; Fitch decay values have been shown below the branches (in bracts); star represents 100. I and II are the main clades of Abildgaardieae.

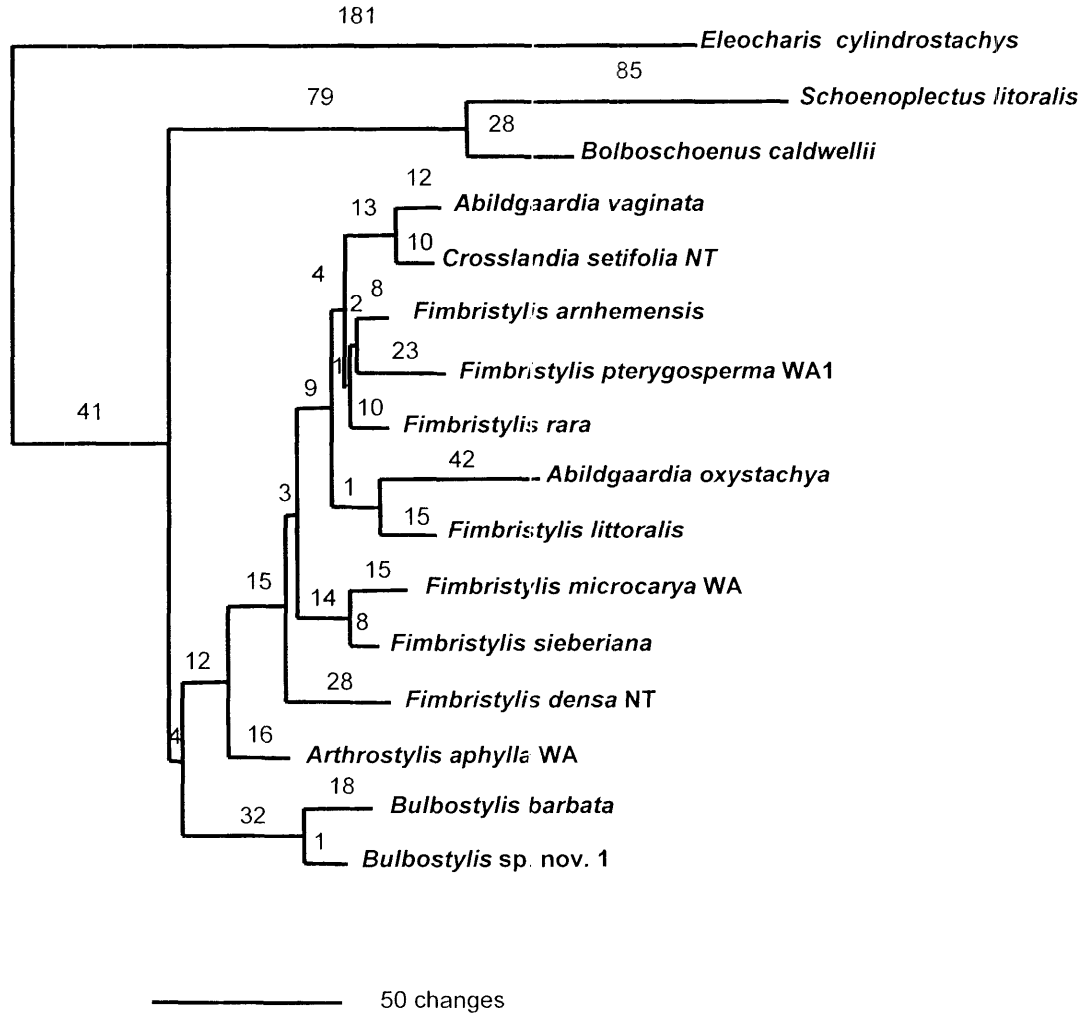

Fig. 3.9. Maximum likelihood tree of the Abildgaardieae–Arthrostylideae complex using a limited number of taxa from the study group (–ln = 6046.43276).

3.4 Discussion

While support for Arthrostylideae nested within Abildgaardieae is strong (80–93% bootstrap support in different analyses, Figs 3.7 and 3.8), this has been consolidated by the synapomorphy of an insertion common to the members of Arthrostylideae available in this study and species of *Bulbostylis*. My *trn*L–F trees do not fully confirm Goetghebeur's (1986) and Bruhl's (1995) suggested position of *Bulbostylis* (Figs 3.6–3.9). They both suggested a monophyletic Abildgaardieae including *Bulbostylis*. Although Bruhl (1995) referred to the ambiguous position of *Abildgaardia* in different analyses, he generally regarded C_3 species of *Abildgaardia* as the closest members of Abildgaardieae to Arthrostylideae while C_4 species of *Abildgaardia* were found to make a monophyletic group with *Bulbostylis*,

Crosslandia, and *Fimbristylis*. This is obviously not the case here. The affinity of some C_4 species of *Abildgaardia* (all the species of *Abildgaardia* in this study are C_4 species of *Abildgaardia*) to Arthrostylideae and their role in connecting the two tribes to each other is not deniable (Figs. 3.6–3.9). According to my cpDNA data, the species of *Abildgaardia* should be regarded as members of *Fimbristylis*. This agrees with Clarke (1902) and other earlier works. In agreement with Taberlet *et al.* (1991), I conclude that the sequencing of *trn*L–F IGS regions of cpDNA can be utilised for evolutionary studies of closely related species although it was not able to resolve all the close relationships within Abildgaardieae.

Genera with basal spikelets such as *Bulbostylis*, *Crosslandia*, and *Abildgaardia vaginata* (K Clarke, pers. comm.) that might have been considered as a monophyletic clade appear within separate and distant clades. The *Bulbostylis* clade is robustly supported (100/21), whereas *Fimbristylis*, *Abildgaardia* and *Crosslandia* are not verified as different groups. In *Fimbristylis–Crosslandia–Abildgaardia vaginata* clade, the species of this genus are paraphyletic (Figs 3.6, 3.8, and 3.9).

In terms of the relationships among the tribes Abildgaardieae and Arthrostylideae there will be a detailed discussion in chapter 6 but in general, results gained from this cpDNA study provide more resolution of relationships between the two tribes than the previous broader scale morphological studies have (cf. Goetghebeur 1986, 1998; Bruhl 1995). That is also the case for relationships within *Fimbristylis* (Figs 3.7 and 3.8).

Two non-exclusive factors could explain the low degree of sequence divergence in *trn*L–F. First, because a limited range of species of *Fimbristylis* and *Abildgaardia* (and other taxa) has been studied, the trees obtained could represent a non-representative evaluation of species covering minimal diversity. However, this is not likely to be the case given the obvious morphological variation sampled across the species.

The second idea is built on the assumption that the molecular clock has variable rates for the considered genera or lineages. Sequencing of *trn*L intron and *trn*L–F IGS (cpDNA) in numerous taxa of monocotyledons and dicotyledons allows a suggestion for possible differences in the rate of the molecular clock along lineages (Bishop and Friday 1985; Li and Graur 1991; Felsenstein 1993; Clegg *et al.* 1994; Sang *et al.* 1995; Moller *et al.* 1999). In the Abildgaardieae, the genus *Bulbostylis*

corresponds to the clade, which diversifies more slowly than the clade of *Fimbristylis*, *Abildgaardia*, and *Crosslandia*. My results show the potential of the explanation of the practical variations at the intratribal level, mainly, by variations in the rate of the molecular clock along lineages.

Taxa defined by *trn*L–F data in *Fimbristylis* conflict with the traditional classification (Kern 1974). Sections would not be identified, and distantly related taxa, as proposed by Kern (1974), may show very close relationships. Moreover, my cpDNA analyses may not resolve the conflict about the distinctness of *Abildgaardia* from *Fimbristylis*.

In summary, extreme care must be carried out for the arrangement and evaluation of intron and IGS sequences. Although, prior to evolutionary analysis, the entire matrices have been precisely reviewed to find out any obscured mutational event, such as minute inversions or misaligned repeat units as Kelchner (2000) suggested, it is still possible that some mutational events have not been detected.

Unexpected or yet to be known phenomena might be responsible for much of the non-homologous gap characters in molecular systematics data at any taxonomic level and my study group is not an exception; because, again according to Kelchner (2000), those who use any non-coding region for their molecular systematics should be aware of the effects and occurrence of these, sometimes difficult to manage, regions.

In conclusion, Abildgaardieae are not monophyletic using *trn*L–F region while the only monophyletic genera within Abildgaardieae are *Bulbostylis* and *Abildgaardia* (ignoring of the inclusion of *Abildgaardia vaginata* within *Crosslandia*). *Fimbristylis* and *Crosslandia* are both paraphyletic. Further, the efficiency of *trn*L–F region as a non-coding chloroplast region in delimitation of monophyletic groups and for classification at genus and tribe levels is satisfactory, but it is not informative enough at the infrageneric level. Finally, the previous infrageneric classifications of the genus *Fimbristylis* (Bentham 1878; Kern 1974) were not consistent with the results of this study.

CHAPTER 4

Internal transcribed spacers of the nuclear genome (ITS) and combined ITS/*trn*L–*trn*F data for phylogenetic studies in the Abildgaardieae

4.1 Introduction

In Cyperaceae, sequence diversity in the ITS region of nrDNA has proven fruitful in constructing phylogenetic hypotheses at lower taxonomic levels (Roalson and Friar 2000; Roalson *et al.* 2001). Hence this region was chosen not only for the primary purpose of exploring the relationship between Abildgaardieae and Arthrostylideae, but also to investigate relationships within *Fimbristylis*.

Despite strong support for the primary clades in the *trn*L–*trn*F (*trn*L–F) trees (Chapter 3, Fig. 3.7), relationships within *Fimbristylis* are scantily resolved. In this chapter a phylogenetic reconstruction of Abildgaardieae is presented, using nrDNA internal transcribed spacer (ITS) sequences to test the results obtained from *trn*L–F data and to seek more resolution for unresolved clades. After that, both *trn*L–F and ITS data are combined to see if resolution of unresolved clades can be obtained from combined data.

4.2 Materials and methods

4.2.1 Plant material

Specimens for DNA analysis collected by J. Hodgon, V. Klaphake and K. L. Clarke *et al.* were preserved in silica gel. Specimens collected by K. L. Wilson and S. Jacobs were preserved in CTAB solution (Thomson 2002). A few fresh samples were collected by me but only one was used here to save time and money and also avoid repetition. Voucher specimens are listed in Appendix 1.

4.2.2 nrDNA region analysed

Three molecules of ribosomal DNA are transcribed as a single precursor. Internal Transcribed Spacers 1 and 2 (ITS1 and ITS2) are intron-like sequences, which are excised from the full-length ribosomal DNA transcript, during the maturation of the small sub-unit (18S) and large sub-unit (28S and 5.8S) rRNAs (see Fig. 2.5) (Appels and Honeycut 1986).

4.2.3 DNA extraction, amplification and sequencing

To obtain sufficient quantities from the genes for sequencing, double-stranded DNA were amplified by PCR. I amplified ITS, using my newly designed ITSZ and ITS8 primers and in some samples the universal ITS1 (White *et al.* 1990) and my ITS8. For sequencing reactions all these primers plus ITS2 and ITS3 (White *et al.* 1990) were used (Table 4.1).

4.2.3.1 Laboratory procedures

Templates of ITS were prepared using sequencing primers: ITSZ as the forward primer and ITS8 as the reverse primer (Table 4.1). PCR reaction combination for ITS consisted of 5 µl of Bioline 10% buffer [25 mM TAPS (tris- (hydroxymethyl)-methyl-amino-propanesulfonic acid,sodium salt)] pH. 9.3, 50 mM KCl, 2 mM $MgCl_2$, 0.2 mM dNTPs, 0.1 µl of 5 units Bioline Taq polymerase, 2 µM of each primer, 10% DMSO (dimethylsulfooxide), and 1 µl total DNA template in a 50 µl reaction volume. Polymerase chain reaction amplifications followed a primary 95° C for 5 min followed by 35 cycles of 95° C for 40 sec; 63° C for 40 sec; 72° C for 80 sec and then another 72° C for 5 min. For the first 10 cycles, the annealing temperature was reduced 1° C for each cycle while for the rest (25 cycles), it was kept at 53° C.

Electrophoresis of the PCR products was done in a 1% agarose gel with a 0.5× TBE buffer [(0.045 M) Tris-borate, (0.001 M) EDTA, pH 8.0] (Sambrook *et al.* 1989). The gel was subsequently dyed with ethidium bromide and visualised using a UV light source to confirm a single product. Double-stranded amplified products of polymerase chain reaction were purified by the "Concert Rapid PCR purification system" (Life Technologies Inc., UK) following the producer's specifications. For each cleaned PCR product, both strands were cycle-sequenced using an ABI Prism automated sequencer (PE Biosystems, Inc., Foster City, USA) with BigDye Terminator Cycle Sequencing Ready Reaction Kit (PE Biosystems, Inc., Foster City, USA).

Table 4.1. Sequences of the five primers used for amplification of two internal transcribed spacers and a 5.8 S coding region between these regions of nrDNA. The code refers to the 3′-most base pair in the published *Saccharomyces cerevisiae* nrDNA sequences (White *et al.* 1990).

Code	Sequence 5′–3′
ITS1	TCCGTAGGTGAACCTGCGG
ITS2	CCTGCGTTCTTCATCGATGC
ITS3	GCATCGATGAAGAACGAGC
ITSZ*	GGAAGTAAAAAGGCGTAACAA
ITS8*	CGCCTGACCTGGGGTAT

*Designed by the author

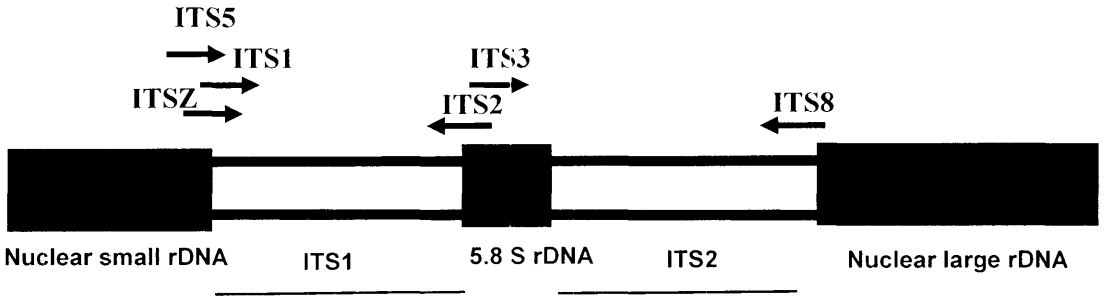

Fig. 4.1. Primers used in this study to amplify ITS region.

4.2.4 Sequence analysis, methods of analysis and testing

After alignment and gap coding of indels (following the same protocol as described in Chapter 3, section 3.2.3) the arranged sequences were used as unordered character states for PAUP* version 4.0b10 (Swofford 2001) to generate phylogenetic trees. Both bootstrap and decay (or Bremer support) analyses (Felsenstein 1985; Bremer 1988; Donoghue *et al.* 1992) were performed using PAUP* to determine comparative support for various clades

Heuristic parsimony assessments are much faster than maximum likelihood (ML) analyses, but they are much slower than distance methods. Initial analyses were carried out with the characters equally weighted. When the quantity of trees exhausted available computer memory, 100 replicate searches for optimal trees were performed using random taxon addition and saving only 100 trees per replicate to sample the remaining tree space. The trees found were subsequently used as starting trees for a simple search that was aborted when available memory was exhausted.

The consensus tree constructed by this strategy probably includes clades that would have collapsed if the searches had gone to completion. In an effort to detect such groups the strict consensus option was employed as a reverse constraint (Cantino *et al.* 1998) in a subsequent heuristic search with 100 replicates of random taxon addition swapping on 500 trees in each replicate. Subsequent analyses were carried out on subsets of the study group, applying 'successive approximation weighting' (SAW) according to the 'rescaled consistency index' (RC) to reduce the weight of highly homoplastic sites.

Bootstrap analysis was used as a quantitative measure of the fit of my data to the results of a parsimony analysis thereby providing a relative measure of the strength of the clades in a tree (Moller and Cronk 1997).This assessment was done using the bootstrap option (bs) in PAUP* employing 5000 replicates, saving only 100 trees per replicate (Mort *et al.* 2000). Bootstrap percentages are not precisely comparable across different studies and must not be accepted as confidence degrees but rather as the way to categorise the corroboration for the clades retrieved in a single analysis (Hillis and Bull 1993). However, comparing the bootstrap values with decay values and clade significance (Schulte *et al.* 1998; Macey *et al.* 1999; Lee 2000), bootstrap values < 70% have been interpreted as indicating no or poor internal support for clades; values from 70% to 80% are considered to indicate moderate support; values > 80% are regarded as having acceptable support. Pairwise distances between species, percent G/C content, and the numbers of variable and parsimony informative characters were also calculated in PAUP*.

Construction of the constraint trees using MacClade 4.03 (Maddison and Maddison 2001) was the next step. To investigate further the support within main clades of the resulted trees, subsets of taxa were constructed by reducing the representation of each main clade in turn. Deletion of the complete subclades except one species of each of these clades was then tried.

Finally, a ML analysis was performed on the ITS dataset under the HKY85 model (the unequal base-frequency generalisation of the simple K2P model; Hasegawa *et al.* 1985), to allow unequal equilibrium base frequencies using the standard settings as implemented in PAUP* 4.0b10 (Swofford 2001). Another reason for selecting the HKY85 model was to join the K2P and F81 models together by allowing transitions and transversions to occur at different rates. This seemed to me the most realistic model available among the best known and most commonly applied models in use

even though realism always costs and the cost here is adding parameters and therefore adding uncertainty. In addition to allowing transitions and transversions to vary, the HKY85 + Γ model allows for variation in rates of substitution among sites, increasing the number of free parameters from four in the HKY85 model to five in the HKY85 + Γ model. Although incorporating gamma (Γ) distribution to HKY85 is a more recommended model (Yang *et al.* 1994) than HKY85 alone, using either model resulted in the same maximum likelihood tree in my analyses despite the difference in branch lengths. Here the trade-off is more parameters against more sampling errors and less statistical power because first, using fewer parameters, does not guarantee getting an accurate estimate (Page and Holmes 1998) and second, the ML analysis is used as a confirmation test of the maximum parsimony analysis. The ML analysis ran for 38 h on an Intel PIII 1.6 GHz with 128 Mb RAM running Windows 2000, completing 23500 rearrangements.

Computational limits (due to the exhaustion of the computer's memory if all specimens were included in the analysis) allowed only one species from each clade in the maximum parsimony analyses to be included, resulting in 27 taxa for the rDNA dataset. Ten replicate searches were carried out for the ITS dataset with a heuristic search algorithm as implemented in PAUP* 4.0b10 with random addition sequence and 'tree bisection reconnection' (TBR, in which the clipped subcladogram is re-rooted before it is reconnected to each branch of the remnant cladogram) options invoked.

4.2.4.1 Combined ITS/*trn*L–F data

As a partition homogeneity test (PHT; an approach that uses a resampling method to estimate the degree to which two datasets or their subsets are in agreement; also known as incongruence length difference test or ILD test) (Mickevich & Farris 1981; Farris *et al.* 1994; see also Swofford 1991) refuted significant conflict between my *trn*L–F dataset and the ITS dataset, a combined analysis was therefore performed. I used this model because it is by far the most widely used model for the examination of the level of disagreement among data partitions (Sanderson and Shaffer 2002). The PHT is based on the difference in tree length (that is, the total number of inferred changes under parsimony) between a single tree in which two data partitions are combined, compared to the sum of the tree lengths of each partition on its own maximum parsimony tree. When

significant incongruence is found, it may be due to differences in the phylogenetic signal of the two data partitions, differences in the level of noise (that is, the randomness of a dataset, or a combination of the two (Dolphin *et al.* 2000). Although PHT is not universally accepted (Yoder *et al.* 2001), it remains the test of choice for the moment.

Several other models have been proposed recently (Chapter 5, section 5.2.4.1.1). They include another randomization assessment (Rodrigo *et al.* 1993), the application of congruence between trees (Miyamoto and Fitch 1995), Partitioned Bremer Support, which calculates the constructive or unconstructive contribution of data partitions to the character support for a particular node in a combined data analysis (Baker and Desalle 1997), and a likelihood ratio test that evaluates differences in probability with and without the constraint that the same phylogeny underlies all data partitions (Huelsenbeck and Bull 1996). None of these latest alternatives is more easily implemented than the partition homogeneity test, although they all have merit. Further, some researchers presume that 'they will enjoy less widespread use than the PHT test in the near future' (e.g. Sanderson and Shaffer 2002, p. 58).

All bases plus length mutations for both cpDNA as well as rDNA datasets were added. This is because the trees obtained from the combined base and gap data were much more resolved and showed higher level of both bootstrap and decay support compared with each of the two components being analysed separately.

The seven accessions which were not sequenced in rDNA studies [*Actinoschoenus* sp. (NT2), *Bulbostylis* sp. nov. (NT2), *Bulbostylis* sp. nov. (NT3), *Fimbristylis pterygosperma* (WA2), *F. punctata* (WA), *F. punctata* (NT1), and *F. punctata* (NT2)] and the five accessions which were not sequenced in cpDNA experiments [*F. acuminata*, *F. pauciflora*, *F. polytrichoides* (Qld2), *Crosslandia setifolia* (WA1), and *C. setifolia* (WA2)] were treated as unknown in the cpDNA data set then included in the combined analyses. Sequence divergences amongst taxa were calculated by typing the SHOWDIST command for the *trn*L–F region, as well as for the ITS region in PAUP*.

For parsimony testing, heuristic searches were conducted on the combined dataset (addition sequence random, 100 replicates, TBR branch swapping, 'MulTrees' on, 'steepest descent' off). A bootstrap search was performed with tree bisection-reconnection swapping option, 'MulTrees' on, 'steepest descent' off, and

'maxtrees' set to 1000. Jackknife resampling was adjusted to 50% deletion, and the other settings were identical to the bootstrap settings. After running the analyses in parsimony criterion using unweighted characters, successive approximation weighting (SAW) was implemented to get more details about the infrageneric groupings. Strict, majority-rule, and semi-strict consensus trees were obtained for the combined data analysis.

Maximum likelihood analyses were run as set out in section 4.2.5. Neighbour-joining (distance) method was used to compare with the results of MP amd ML analyses. To test the results of the neighbour-joining tree, I hypothesised the positions of *Bulbostylis*, Arthrostylideae, and *Abildgaardia ovata* in a constraint tree and ran a maximum parsimony analysis as set out in section 4.2.5.

4.3 Results

4.3.1 ITS structure, size, and composition

The unaligned ITS chain varied between 588 bp in *Abildgaardia ovata* and 660 bp in *Fimbristylis neilsonii*. Variation within *Fimbristylis* is less, from 603 base pairs in *Fimbristylis acuminata, F. arnhemensis*, and *F. ferruginea* to 660 in *F. neilsonii*, with ITS2 exceeding ITS1 in length: 237–279 bp (ITS2) compared with197–223 bp (ITS1). Percent G/C content varies (50.3–57.5%) among all species, and shows a similar order within ITS1 and ITS2: 46.3–59.8% and 47.7–57.3%, respectively. The 5.8S gene is very uniform in length: 162–165 base pairs.

The aligned data for 53 ITS sequences were 824 bp long (Appendix 4). The alignment of ingroup sequences required numerous indels, most of which involved the loss or gain of 2 or 3 bp; one deletion of 2 and two deletions of 3 bp characterise all five samples of *Crosslandia* (indels 1 and 5, Table 4.2). All ingroup taxa were readily arranged and compared with each other, but matching them with all outgroups was problematic in a few sections of the whole sequenced region and also in a few cases the match was almost not achievable. In particular, *Fuirena ciliaris* could not be aligned; hence I excluded this species from my analyses.

The aligned ITS sequences contained 435 variable characters of which 359 were informative, and among these, 312 were informative concerning *Fimbristylis* (87% of whole informative positions; 71% of whole variable positions, 35.5% of whole sequence array) (Table 4.3). Alignment resulted in 58 gaps with more than one base pair length, ranging in length from 2 to 15 bp. Seven gaps could be assigned to polyT

tracts. Insertions/deletions (only if of two or more base pairs length) were coded as additional binary data. Forty-six indels were demonstrated to be potentially valuable within the ingroup; a further four indels in ITS1 and four in ITS2 distinguished outgroups from the ingroup, while another eight were informative only amongst the outgroups (Tables 4.2 and 4.4).

Uncorrected pairwise distances of species within the whole dataset ranged from 0.2% (*Fimbristylis lanceolata* and *F. compacta*) to 32.1% (*Bolboschoenus caldwellii* and *Crosslandia setifolia*). Within Abildgaardieae, divergence ranged from 0.2% (*Fimbristylis lanceolata* and *F. compacta*) to 19.2% (*Bulbostylis barbata* and *Abildgaardia vaginata*). Among the species of *Fimbristylis,* divergences varied highly (12.9% between *Fimbristylis nutans* and *F. compacta*) suggesting enough divergence to provide resolution of infrageneric relationships.

4.3.2 ITS phylogenetic analysis

The initial search applying equal weights found 24 most-parsimonious trees of 1283 steps. The consistency index was 0.49 (0.46 excluding uninformative characters) and the retention index (RI) 0.72. The strict consensus cladogram is shown in Fig. 4.2. No shorter trees were found, nor were equally parsimonious trees that were inconsistent with Fig. 4.2 found in any searches from the reverse constraint heuristic search.

Arthrostylideae (Fig. 4.2, clade 1) is well separated from Abildgaardieae (Fig. 4.2, clade 2). The representatives of Abildgaardieae form two main clades indicated as I and II in Fig. 4.2. There is strong bootstrap support for both clades (100% and 91%, respectively). All sampled species of *Abildgaardia* group together except *Abildgaardia vaginata*, which is grouped with *Crosslandia setifolia* in clade A. In this study, where specimens of the same species from different provenances were included, they grouped together (*Fimbristylis densa*, *F. tetragona*, *F. cinnamometorum*, *F. microcarya*, and *F. pterygosperma*) except for *Crosslandia setifolia*.

Species assigned to *Fimbristylis* appeared only in Abildgaardieae clade I, within clades X and Y, and one composed of *F. littoralis* (Fig. 4.2). Relationships between a few lineages observed within clade X appear to be mostly uncertain. Within clade X the species form four poorly resolved clades labelled A–D. *Fimbristylis acuminata* and *F. nutans* plus *F. pauciflora* are weakly supported in clade A with *Crosslandia*

and *Abildgaardia vaginata*. *Crosslandia* was strongly supported (together with *A. vaginata*) within clade A. *Abildgaardia* clade followed by *Fimbristylis littoralis* branch early in main clade I. *Bulbostylis* is well supported in clade II , with *Bulbostylis* sp. nov. sister to all the other taxa and with strong support for the placement of all these taxa.

Table 4.2. Potentially informative indels for relationships within the ingroup.

Indel	Affected sites*
1	71–73
2	122–125
3	153–168
4	175–178
5	255–258
6	255–259
7	259–261
8	282–284
9	304–305
10	306–307
11	306–308
12	306–309
13	625–628
14	641–643
15	648–649
16	694–701
17	709–710

* Counted on the basis of the base numbers of the contig of a I sequences in Sequencher

Fimbristylis cephalophora, *F. compacta*, and *F. lanceolata* form a robust clade (100%, Figs. 4.2 and 4.3) within the main clade I. *Fimbristylis rara* is the robust sister of *F. ferruginea*. *Fimbristylis littoralis* diverges early from the rest of *Fimbristylis* (including *Abildgaardia vaginata* and *Crosslandia setifolia*) (Figs 4.2 and 4.3). Hence the nrDNA topology shows that the species of *Fimbristylis* are not

widely polyphyletic. However, this genus is not monophyletic, given the relationship between *F. acuminata*, *F. nutans*, and *F. pauciflora* (Figs 4.2 and 4.3) and the *Crosslandia–Abildgaardia vaginata* group. The sister relationship between *Fimbristylis littoralis* and the rest of the sampled *Fimbristylis* species, suggests that these taxa are congeneric.

When character state changes were mapped on one of the most parsimonious trees, 31 (7.1%) of the total of 435 variable sites had a CI < 0.25 with up to 16 changes being required per site. Heuristic search using 100 replicates of random taxon addition and equal character weighting yielded 12 trees of 1099 steps in two islands (CI = 0.53; RI = 0.71; RC = 0.38). The strict consensus constructed from the new trees is consistent with the consensus tree of the complete data. Three rounds of heuristic search using successive weighting of the complete ITS data set gave three most parsimonious trees of 484.5 steps (CI = 0.73; RI = 0.84; RC = 0.61), the strict consensus of which is given in Fig 4.3.

Inner relationships in clade I are resolved in the successive approximation weighting (SAW) tree (Fig. 4.3): *Fimbristylis sieberiana* is sister to the rest of clade G. The three trees obtained by SAW, were identical except for the positions of *Fimbristylis compacta*, *F. lanceolata*, and *F. cephalophora* relative to each other; however, they are sister to each other, making a consistent clade in all trees. While there is little or no bootstrap support for the earlier diverging clades inside clade X, there is good support for a sister relationship of clade H and the clade comprising clade G and *F. sieberiana* (83% bs). Support for the clades earlier recognised inside foremost clade II is higher than in the initial analysis.

An examination of the allocation of the 54 potentially informative indels on the successive weighting consensus cladogram (Fig. 4.3) shows that 26 are consistent with a single origin. The remaining 28 indels require reversals. Indel 54 shows a two base addition that happens in the starting point of the clade X, but seems to be subsequently removed within the *Fimbristylis rara—F. ferruginea* clade (Fig. 4.2). Four indels need more than two separate origins within the ingroup. Indel 23 is the two base indel that needs four origins: within the Arthrostylideae clade and the Abildgaardieae clade in outer clades, and in the clade made up of the Northern Territory specimens of *Crosslandia* and *Abildgaardia vaginata* as well as in *F. littoralis* in inner parts of the tree (Fig. 4.3). Indel 26 is a four base indel that occurs

at the starting point of the *F. compacta*–*F. lanceolata* clade and also in *F. arnhemensis* and *Bulbostylis* sp. nov. (Fig. 4.3).

Three trees were obtained from the ML analysis; one is shown in Fig. 4.4. These trees appear to be congruent with my parsimony strict consensus (Fig. 4.2). The differences among the trees are minimal and do not involve well-supported branches. Well-supported branches corroborate many relationships seen in Fig. 4.2.

Resolution of the ITS strict consensus tree (Fig. 4.2) is much greater than that of the *trn*L–F strict consensus tree (Fig 3.7). In most of the cases when the sequences of two taxa were compared, sequence differences of *trn*L intron and *trn*L–F IGS sequences were more than those of ITS (Fig. 4.5), even if only substitutions are included: for example, the number of insertions/deletions and substitutions in ITS, the *trn*L intron, and the spacer for *A. ovata*–*B. barbata* is 4, 2, and 7.5, respectively. It means that the rate of evolution in the ITS region is double that of the *trn*L intron, and the rate in the spacer is 3.75 times faster than the intron.

Some aspects of genetic evolution for the matrices studied in ITS analysis using Fitch weighting are shown in Table 4.3. The percentage of variable characters is least for the *trn*L intron (13.33%; 135 per 1012), high for the *trn*L–F IGS (34.16%), and greatest for ITS (49.54%). Each variable position changed 2.9 times (tree length divided by variable characters) in all three regions. Rescaled consistency index (RI) was greatest in *trn*L (0.76), halfway in the intergenic spacer (0.74) while lowest in ITS (0.72). In this study, the frequency of informative base substitutions is lower in ITS1 (39.2%) than in ITS2 (47.2%), and the frequency of potentially informative indels is markedly different (22 and 38, respectively Table 4.4).

4.3.3. The phylogenetic analysis of the ITS/*trn*L–F combined dataset

The combined data set has 2858 characters, 783 of which are potentially phylogenetically informative. Fitch analysis generated 12 cladograms, each 2709 steps long, CI = 0.55, RI = 0.71 (Fig. 4.5). Reweighting yielded 1 cladogram of 1169.86 weighted steps (CI = 0.81 and RI = 0.85) (Fig. 4.7.)

The results of the combined analysis show that Abildgaardieae, excluding *Bulbostylis*, constitutes a strongly supported clade (bootstrap value = 100%, Fig. 4.6), but support is poor for many major branches within it. There is strong support for *Bulbostylis* as a genus separate from the rest of Abildgaardieae given that Arthrostylideae is nested between the *Bulbostylis* clade and the rest of

Abildgaardieae. There is strong support for the monophyly of Arthrostylideae (Figs 4.6 and 4.7) but it is nested inside Abildgaardieae in MP and ML analyses. Analyses of distance method confirm the location of *Bulbostylis* much closer to Arthrostylideae than to the rest of the tribe Abildgaardieae. *Fimbristylis* is paraphyletic; the *Fimbristylis* sensu stricto clade (Fig. 4.6, clade Z) (Fig. 4.7, clade Z) includes species in two genera: *Abildgaardia* pro parte max. and *Crosslandia* with *Abildgaardia vaginata* embedded in it.

Figure 4.8 shows the neighbour-joining tree, which differs from the cladograms particularly in the position of *Bulbostylis*, Arthrostylideae, and *Abildgaardia ovata*. The results of parsimony analysis of the neighbour-joining constraint tree were less parsimonious than for the cladistic analysis with tree lengths 43 (for new position of *A. ovata*) and 57 (for new position of *Bulbostylis* and Arthrostylideae) steps longer than the shortest trees from unconstrained parsimony analysis.

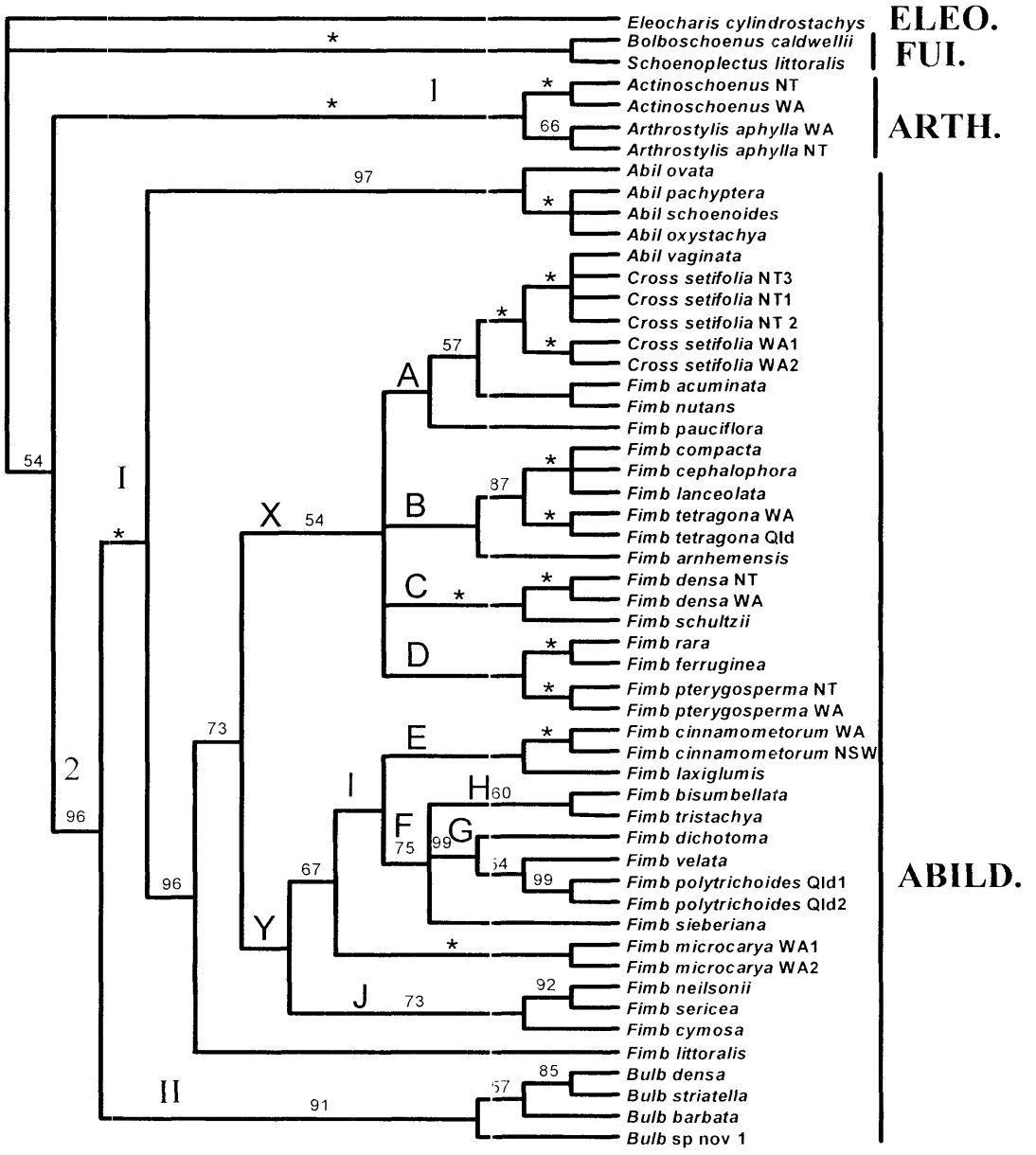

Fig. 4.2. Strict consensus tree from 24 most parsimonious trees for the ITS data set unweighted Fitch-parsimony. Bootstrap values are in percent on top of branches from 5000 replicates. Asterisk represents 100% bootstrap. As there was no obvious difference among the sequences of different accessions of *Bulbostylis* sp. nov., they have been represented here by one accession name. Abbreviations on the right denote tribes of Cyperaceae sensu Goetghebeur (1986) because of the larger number of tribes suggested in that study than, for example, in Bruhl (1995). *Abil* = *Abildgaardia*, *Cross* = *Crosslandia*, *Fimb* = *Fimbristylis*, *Bulb* = *Bulbostylis*. ABILD = Abildgaardieae, ARTH = Arthrostylideae, FUI = Fuireneae, and ELEO = Eleocharideae. WA = Western Australia; NT = Northern Territory; NSW = New South Wales; Qld = Queensland. I and II refer to the main clades of Abildgaardieae; A-H, X, and Y refer to clades discussed in the text.

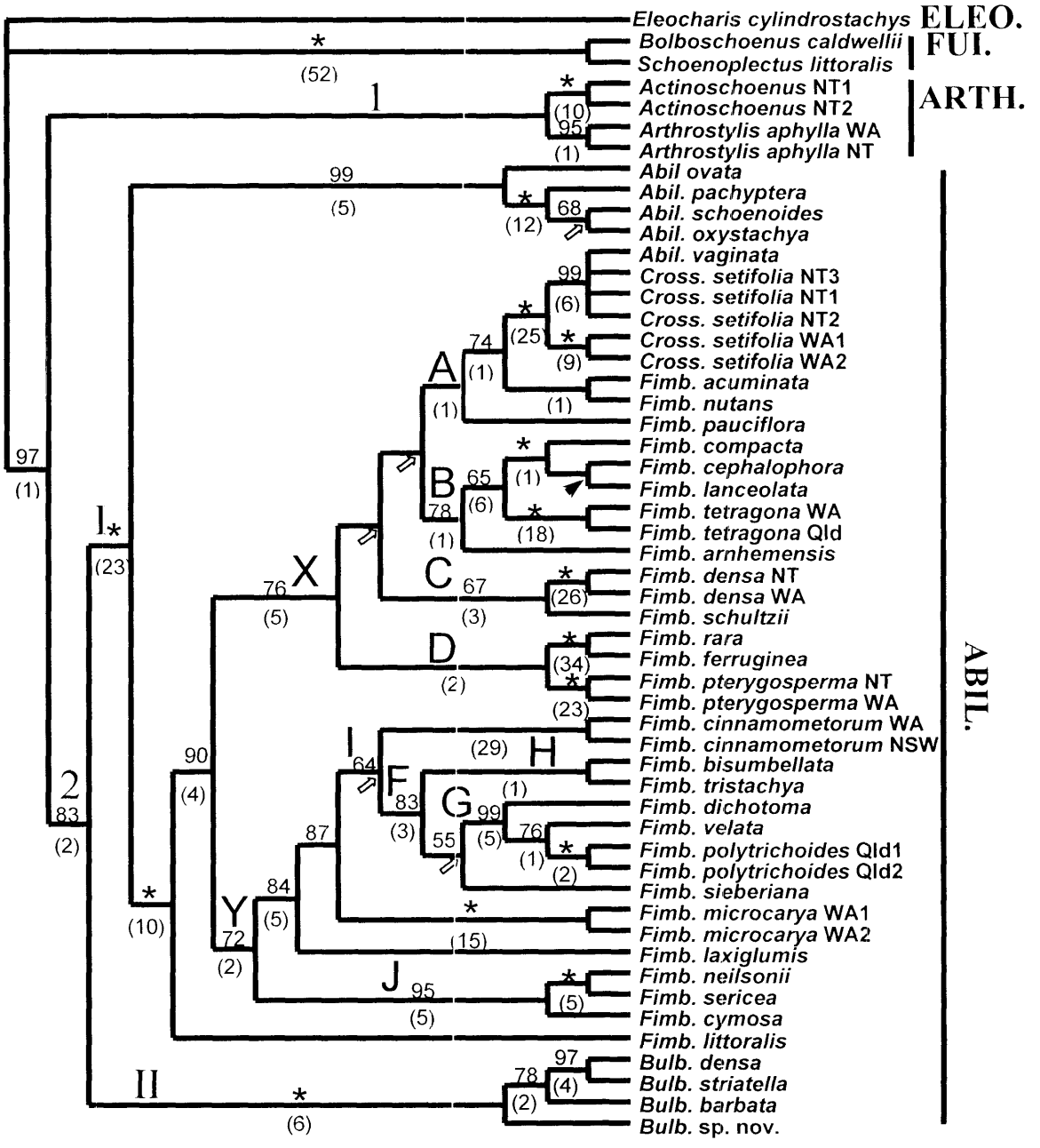

Fig. 4.3. One of the three cladograms found by successive approximation weighting of ITS data. Arrows mark clades absent in the strict consensus tree of the Fitch (open arrows) or both Fitch and successive weighting (solid arrows) analyses. Numbers above these branches are Fitch bootstrap percentages; Fitch decay values are shown below the branches (in brackets). Other labels and abbreviations are as in Fig. 4.2

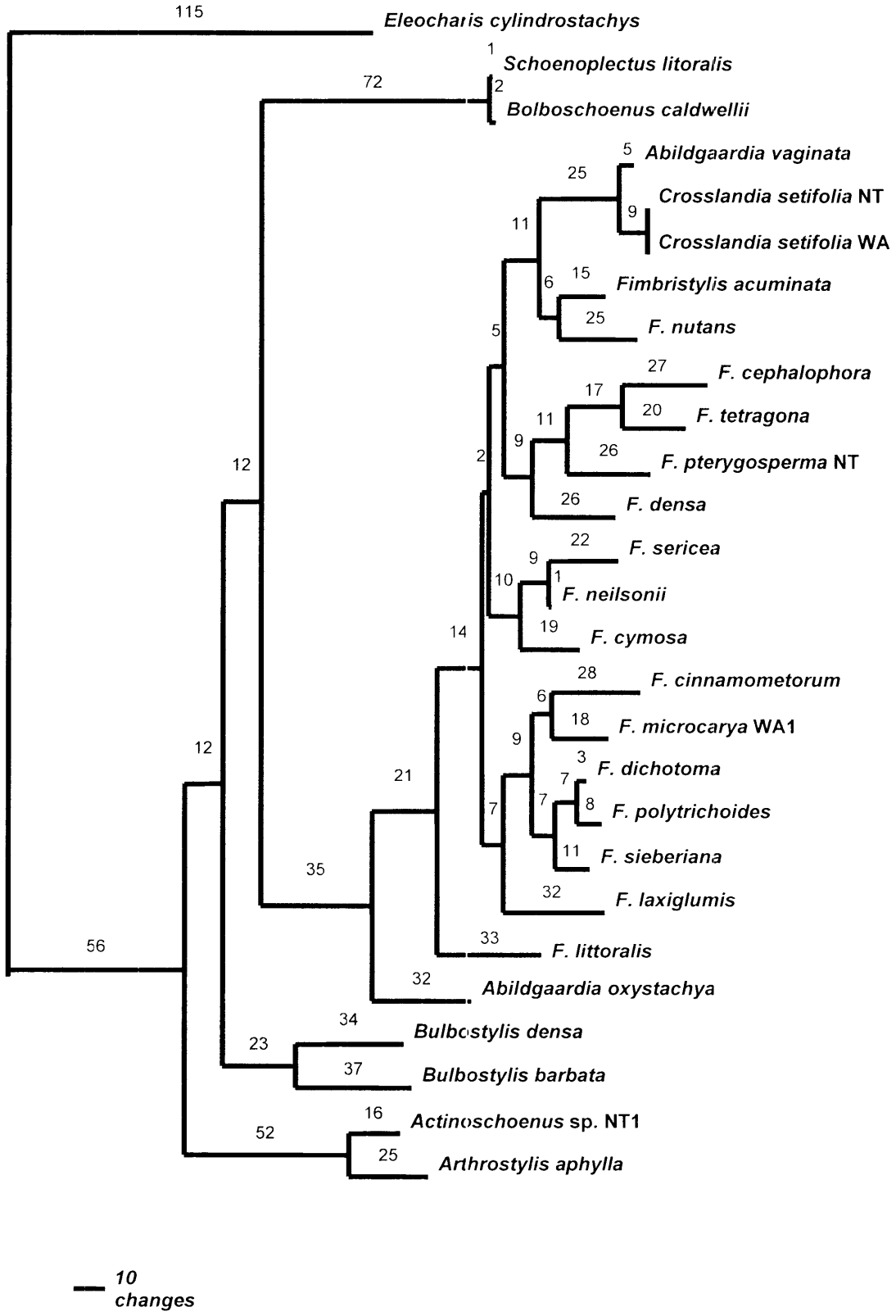

Fig. 4.4. Maximum likelihood tree (–lr = 5545.65996) of the Abildgaardieae–Arthrostylideae complex using a limited sample of taxa from the study group.

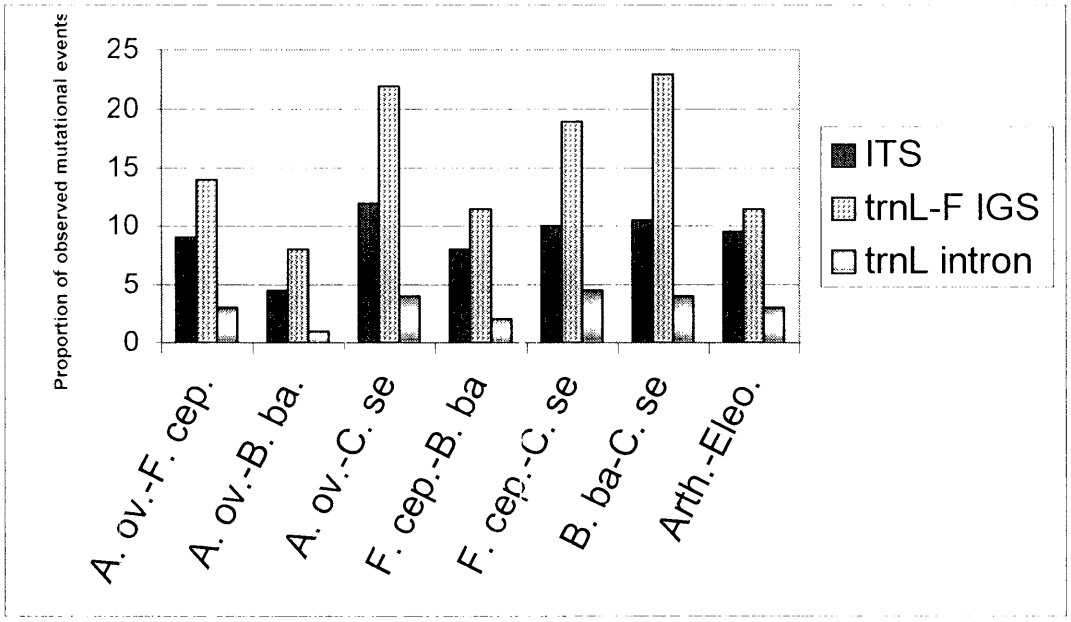

Fig. 4.5. Comparisons, for seven pairs of taxa, of the number of practical mutational events of the ITS gene, the *trn*L intron, and the *trn*L–F intergenic spacer showing the higher sequence differences of *trn*L intron and *trn*L–F IGS regions of chloroplast DNA than of those belonging to ITS. Abbreviations: A. ov. = *Abildgaardia ovata*; F. cep. = *Fimbristylis cephalophora*; B. ba. = *Bulbostylis barbata*; C. se. = *Crosslandia setifolia* NT1; Arth. = *Arthrostylis aphylla* WA; Eleo. = *Eleocharis cylindrostachys*.

Table 4.3 Comparison of genetic evolution in the plastid and nuclear regions. This comparison is based on the Fitch analysis for taxa having a complete set of sequences, showing tree statistics and average number of changes per variable site (tree length/variable characters).

	ITS	*trn*L intron	*trn*L–F spacer
Variable/total	435/878	135/1012	288/843
Tree length	1283	398	849
CI	0.49	0.66	0.69
RI	0.72	0.76	0.74
Changes per site	2.9	2.9	2.9

Table 4.4. Sequence characteristics of the *trn*L intron, *trn*L–F IGS, and the ITS region

	*trn*L–F IGS	*trn*L intron	ITS1	ITS2	5.8S
Length range (bp)	333–1600	576–772	197–223	237–279	162–165
Aligned length (bp)	843*	1012	343	525	166
Number of indels	63	23	22	36	–
GC content	40%	32%	50.9%	51.6%	59.6%

* Excluding *Fimbristylis polytrichoides* because of its extraordinary additional >1000 basepairs insertion (chapter 3, section 3.3.1)

4.3.4 Comparison of exchange rates of the ITS region versus *trn*L–*trn*F region and *trn*L intron

The cpDNA *trn*L–F region has higher variability than the nuclear ITS in this study. The branch lengths indicate that *Bulbostylis* and *Abildgaardia* tend to have higher substitution rates than *Fimbristylis* and *Crosslandia* species. In the ITS and *trn*L–F regions and *trn*L intron, the average quantities of base substitutions in the species of *Abildgaardia* were greater than substitutions within *Fimbristylis* and *Crosslandia*.

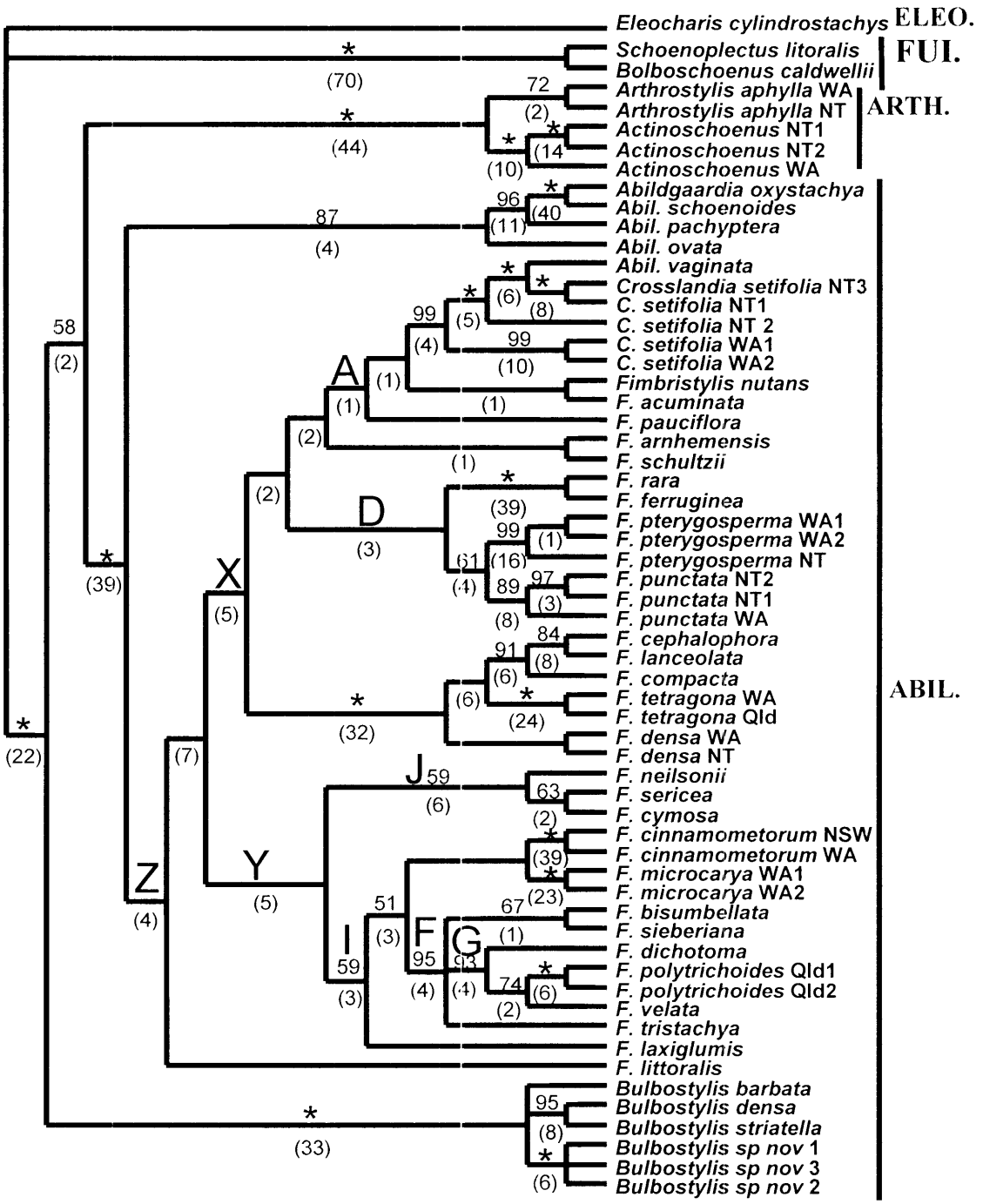

Fig. 4.6. Strict consensus tree from twelve most parsimonious trees for the combined data set from the *trn*L–F and the ITS regions produced by unweighted Fitch-parsimony. Bootstrap values are indicated as percent (on top of branches) after 1000 replicates. Decay values are indicated below branches (in bracts). Abbreviations denote tribes of Cyperaceae sensu Goetghebeur (1986): ABIL = Abildgaardieae, ARTH = Arthrostylideae, FUI = Fuireneae, and ELEO = Eleocharideae. WA = Western Australia; NT = Northern Territory; NSW = New South Wales; Qld = Queensland; * = 100%. A, D, F, G, I, J, X, and Y refer to clades present in Fig. 4.2.

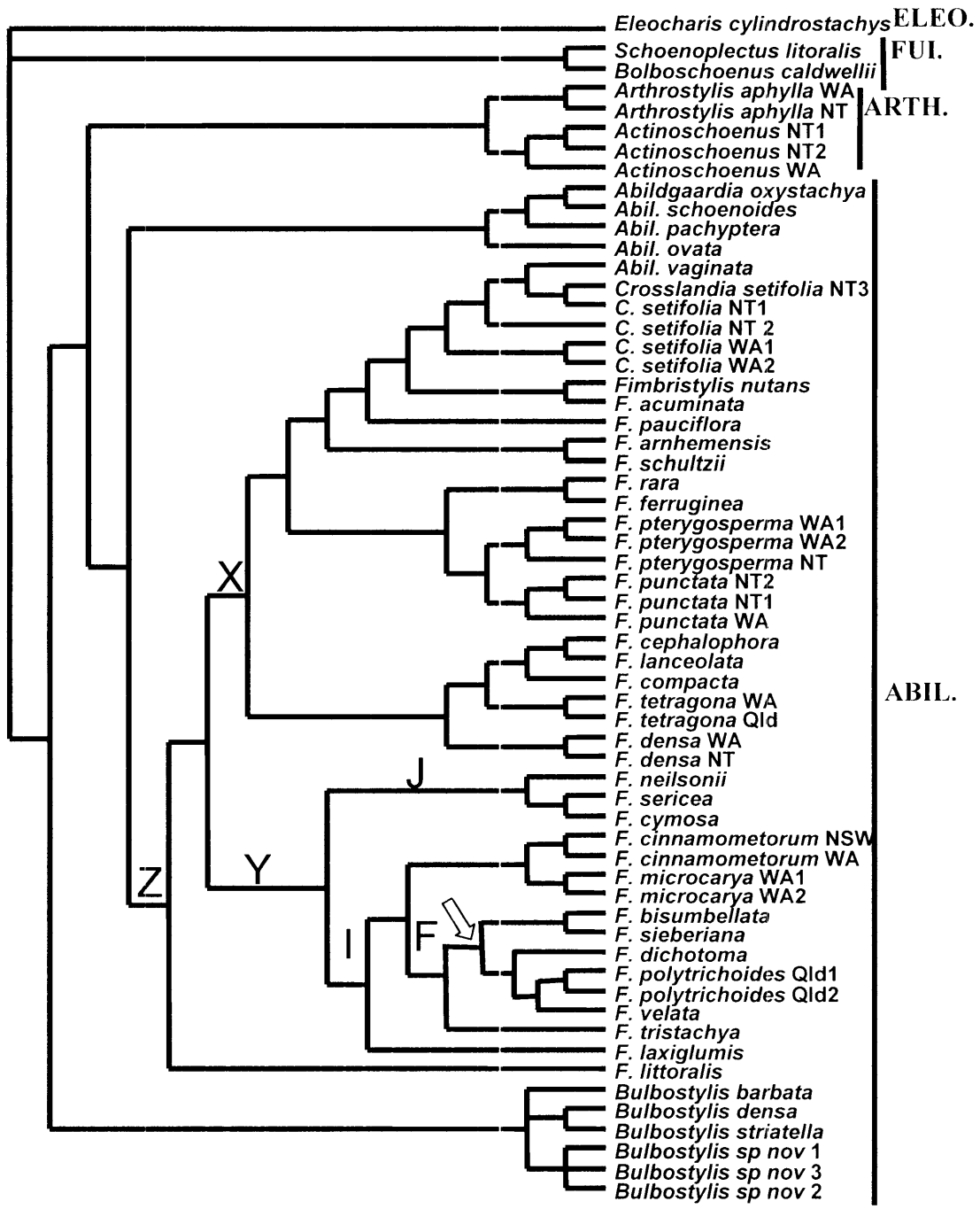

Fig. 4.7. Cladogram of the only tree found by successive approximation weighting of combined data. The arrow indicates the only clade with more resolution than in the strict consensus tree of the Fitch analysis. Clades X and Y are present again. Other abbreviations as in Fig. 4.6.

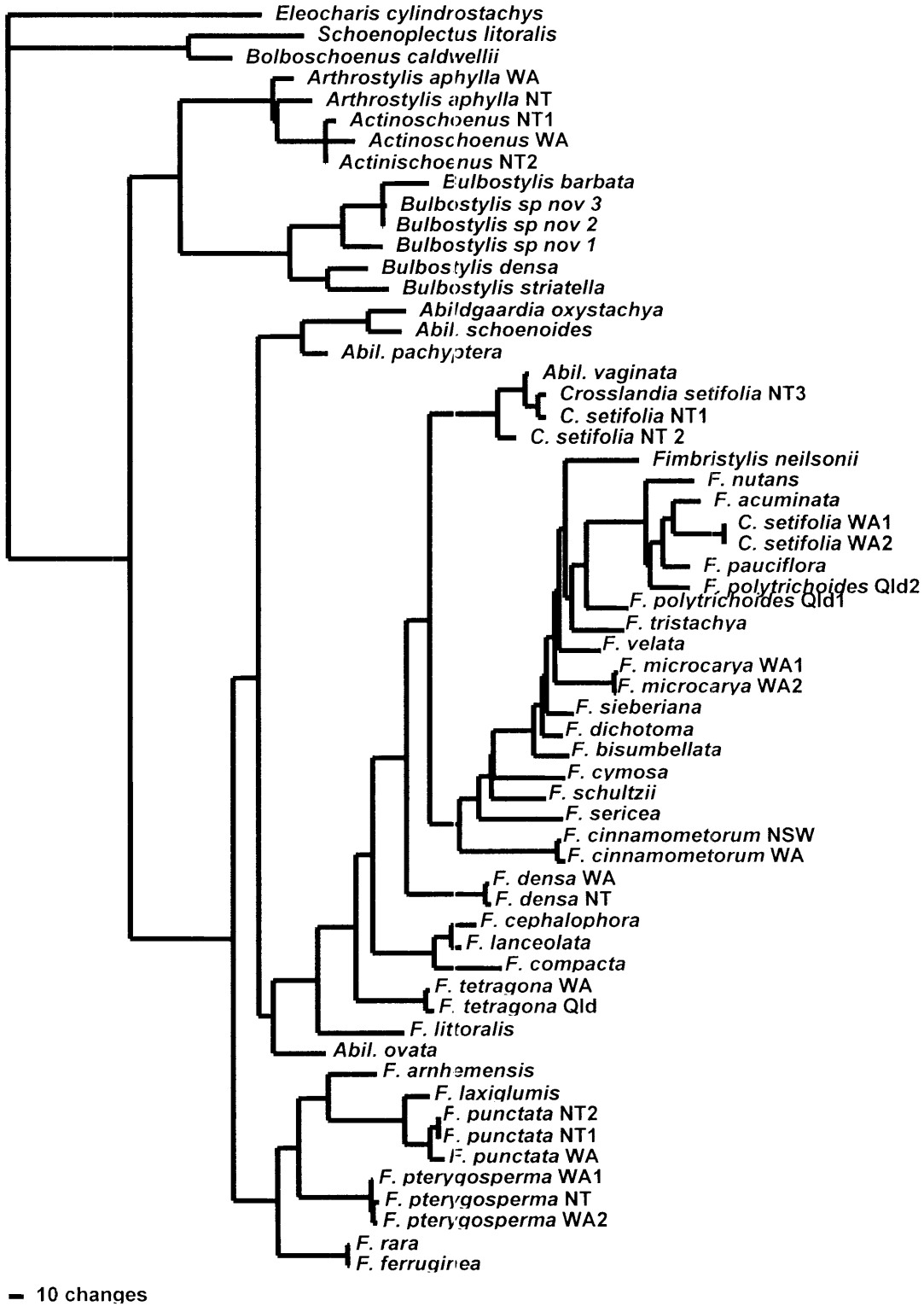

Fig. 4.8. Neighbour-joining distance tree from the chloroplast *trn*L–F and nuclear ITS regions combined data set based on Kimura's two-parameter distance. Branch lengths are proportional to nucleotide substitutions per site. Abbreviations as in Fig. 4.6.

4.4 Discussion

4.4.1 ITS in Abildgaardieae

Organisation of the internal transcribed spacer (ITS) regions in *Abildgaardia*, *Bulbostylis*, *Crosslandia*, and *Fimbristylis* resembles those that are found in other groups of the family, including *Scirpus* (Roalson and Friar 2000), *Eleocharis* and *Fuirena* (as sequenced by Roalson and Friar 2000 and also Appendix 1), *Arthrostylis* and *Actinoschoenus* (Appendix 1), and *Carex* (Starr *et al.* 1999), in that the ITS1 is shorter than ITS2 (197–223 bp and 237–279 bp, respectively).

The preponderance of substitutions over segment mutations in the ITS region, and the short length of the indels (mostly 1–2bp; Table 4.3), are in line with the conclusion (Baldwin *et al.* 1995) that these spacers are under structural constraint due to their role in the maturation of nuclear RNAs. Most indels might be the result of slippage at some stage in DNA replication (Levinson and Gutman 1987; Stephan 1989). By contrast, the large deletions in all sequenced species except *Abildgaardia ovata* and *Bulbostylis* sp. nov. (indel 3; 52 bp) probably result from unequal crossing-over of rDNA repeat units (Smith 1976). Indels of the latter kind may bare useful apomorphies in intensive investigation of this tribe.

The extent of homoplasy is illustrated by indel 39, a two-base insertion that arises twice in main clade I (in *Abildgaardia ovata* and *Abildgaardia vaginata*; Fig 4.3), once in main clade II (in *Bulbostylis densa*) and twice in the outgroups (in *Eleocharis cylindrostachys* and *Bolboschoenus caldwellii*). This phenomenon is widely reported for spacer regions in general (e.g. Golenberg *et al.* 1993), and ITS in particular (e.g. Kim and Mabry 1991; Cross *et al.* 2002)

Studies of ITS in other families indicate the difference between the phylogeny based on ITS1 and ITS2 spacers for different genera (Baldwin *et al.* 1995). In *Krigia* (Lactuceae, Asteraceae), for example, ITS1 pairwise divergence values are, on average, twice those of ITS2. However, Baldwin *et al.* (1995) stated that the difference of average pairwise distance values is not considered to be an adequate measure of relative evolutionary rates. In *Epilobium* (Onagraceae) and *Gossypium* (Malvaceae) ITS1 includes 50 to 100% more potentially informative characters than ITS2 (Baldwin *et al.* 1995). Similar rates of information and capacity for phylogeny reconstruction were obtained for both ITS1 and ITS2 regions in the Pooideae (Hsiao *et al.* 1995). There are more informative base substitutions and indels in ITS1 than

ITS2 (Table 4.4), which is an indication that ITS2 is under more structural restriction than ITS1 in Abildgaardieae.

The ITS region appears fairly conserved among a wide range of species within the tribes Abildgaardieae and Arthrostylideae, with a maximum divergence observed between two species within Abildgaardieae of 19.2% (see section 4.3.1). These values suggest that the members of this tribe are relatively recently developed (Bremer 2000, 2002). Intrageneric rates of divergence vary greatly from one genus to another. Notably, *Fimbristylis* has considerable variation within my sample and the infrageneric relationships should be investigated further with a larger sample (Chapter 6).

4.4.1.1 Relationships and monophyly

ITS data disagree with *trn*L–F data except in the relationships between *Fimbristylis* and *Crosslandia*. The nrDNA data not only did not support the paraphyly of Abildgaardieae inferred from *trn*L–F data (Chapter 3) but recognised Arthrostylideae as a separate clade. However, this contradiction was not highly supported, with 54% bs for Arthrostylideae as sister clade to Abildgaardieae. Therefore, ITS did not give any more resolution to my previously well-resolved outer clades of the cladogram (Figs 3.7, 3.8, and 4.6).

Species of *Abildgaardia* (except *A. vaginata)* were grouped together in the ITS tree indicating a strong support for the monophyly of that genus. This was not the case in the *trn*L–F tree, where *Abildgaardia* was polyphyletic (Fig. 3.7; Chapter 3). One explanation for this might be that trees are tracking different phylogenies detected through biparental inheritance in nDNA and maternal inheritance in cpDNA, resulting from hybridisation between some ancestral species of *Abildgaardia* and *Fimbristylis*. For instance, *Fimbristylis compacta* was grouped with *Abildgaardia oxystachya* and *A. schoenoides* in the *trn*L–F tree (Figs 3.7 and 3.8) whereas it makes a clade with *F. cephalophora* and *F. lanceolata* in the ITS tree (Fig. 4.6). This suggests a possible hybridisation among *Fimbristylis cephalophora*, *F. lanceolata*, *Abildgaardia oxystachya* and *A. schoenoides*. Further, while *Abildgaardia pachyptera* and *A. ovata* have been located within the subclades of *Fimbristylis* in the *trn*L–F tree, they both make a clade with *A. oxystachya* and *A. schoenoides* in the ITS tree. These further support the hybridisation theory. The two trees are not consistent in the indication of the sister

group of *Crosslandia* within the *Fimbristylis* clade leaving the matter open for further work

The results of the present study indicate a closer and more robust relationship between *Abildgaardia vaginata* and the Nothern Territory populations of *Crosslandia setifolia* (100%, Figs. 4.2 and 4.3) than between the Northern Territory and Western Australian populations of *C. setifolia*. *Abildgaardia* is not monophyletic as currently circumscribed, and *Crosslandia* is paraphyletic in all analyses. This indicates the need to move *Abildgaardia vaginata* from *Abildgaardia*. Goetghebeur (1986) proposed recognition of a second species in *Crosslandia* for the single-headed specimens and my results suggest *Abildgaardia vaginata* to be considered as a third *Crosslandia* species. However, the fact that *Crosslandia setifolia* is not monophyletic also suggests another possibility, that is, that the two accessions of *Crosslandia setifolia* and also *Abildgaardia vaginata* are simply species of *Fimbristylis*. The *Crosslandia–Abildgaardia vaginata* complex seems to be a real finding as the morphological studies by Kerri Clarke (pers. comm.) has shown the same affinity between these taxa

The species of other genera of Abildgaardieae and Arthrostylideae are in groups that are given very good or reasonable support: *Actinoschoenus* (100%); *Abildgaardia* excluding *A. vaginata* (97%); *Bulbostylis* (91%); *Arthrostylis* (66%) (Fig. 4.2)., these indicate the monophyly of the genera *Actinoschoenus*, *Abildgaardia* (excluding *Abildgaardia vaginata)*, *Arthrostylis*, and *Bulbostylis*.

Knowing that *Crosslandia setifolia* and *Abildgaardia vaginata* are nested within *Fimbristylis*, they should be included in *Fimbristylis*. However, wider sampling is necessary to find out the detailed relationships among these taxa and also to investigate where *F. littoralis* stands within *Fimbristylis*.

The *Bulbostylis* clade is sister to the rest of the tribe. This result is broadly in line with the findings of Bruhl (1995) who suggested *Bulbostylis* as the basal genus within Abildgaardieae. ITS data also corroborate some aspects of infratribal classification as proposed by other authors (Goetghebeur 1986; Muasya *et al.* 2000a) including *Abildgaardia* (except *A. vaginata*) being basal to *Fimbristylis*, although my results disagree with Muasya *et al.* (2000) in the position of *Bulbostylis* relative to *Fimbristylis* and *Abildgaardia*. Based on the ITS analysis the sister group of Abildgaardieae is Arthrostylideae.

4.4.2 Combining data

4.4.2.1 Resolution of clades from combined cpDNA and nrDNA data

The results from analysis of the combined dataset provide better support for most clades than each separate analysis of sequence data (Figs 3.7, 4.2, and 4.6). Sampling error or functional constraints are among the factors that might cause different results for analysis of individual DNA regions. Both causes can be overcome by combining two separate data sets (Qiu *et al.* 1999).

The trees from analysis of combined ITS and *trn*L–F data was in agreement with that from *trn*L–F data with regard to the position of *Arthrostylis* and *Actinoschoenus*. These trees do not support monophyly of Abildgaardieae, although not surprisingly with a much lower bootstrap support (58%) than the *trn*L–F MP trees because the ITS tree supports this monophyly. This suggests that although higher resolution for the position of *Arthrostylis* and *Actinoschoenus* was not obtained, the *trn*L–F data are strong enough to be confirmed in the combined analyses. This is not the case for monophyly of *Abildgaardia*, as combined analyses support the grouping of all the species of this genus (except *A. vaginata*) confirming the ITS analyses with more resolution but slightly less bootstrap support (87% in combined MP tree versus 93% in ITS tree).

ITS and *trn*L–F trees introduced different positions for *Bulbostylis* sp. nov. within the *Bulbostylis* clade (basal in ITS tree and lately derived in *trn*L–F tree). The analysis of the combined data showed a polytomy among the species of *Bulbostylis*, leaving the position of *B.* sp. nov. unknown.

My phylogenetic analysis of the combined matrix generally corroborates previous molecular works (Muasya *et al.* 1998, 2000a), although Muasya *et al.* (1998) showed that, using *rbc*L data only, *Schoenoplectus lacustris* is the closest taxon to Abildgaardieae while the two other species of *Schoenoplectus* in their sample form the closest group to *Eleocharis*. Most branches of that analysis had low support, including the clades placing *Schoenoplectus* and *Eleocharis*, and *Schoenoplectus* and *Bulbostylis* together.

In my analysis, the *Bulbostylis* clade is sister to the Arthrostylideae clade and the rest of Abildgaardieae (Figs 4.6 and 4.7). These clades cover two tribes but do not contain outgroup genera such as *Eleocharis*, *Schoenoplectus*, *Fuirena*, and *Bolboschoenus*. These genera have been previously suggested to be the members of tribes close to Abildgaardieae in other works on Cyperaceae (e.g. Goetghebeur

1986, 1998; Bruhl 1995). These genera were outgroups in this study and remained as such after analyses were undertaken.

Abildgaardieae is paraphyletic in MP, ML and distance analyses, although the distance method does not distinguish symplesiomorphy from synapomorphy. Some groupings had little support. Relationships extracted by the MP and ML methods are partly in agreement with earlier assessments based on morphology and molecular studies with regard to the closeness of Arthrostylideae to Abildgaardieae (e.g. Bruhl 1995) and the more remote position of *Bulbostylis* with respect to *Abildgaardia* and *Fimbristylis* (e.g. Muasya *et al.* 1998).

The genus *Abildgaardia* (excluding *A. vaginata*), which was considered as one of the sections of *Fimbristylis* by Kern (1974), makes a monophyletic group with strong support in ITS (100%) and combined (87%) analyses. However, it is not monophyletic according to the *trn*L–F analysis, but using more data/genes/regions may result in more corroboration for the infrageneric classifications suggested by Kern (1974) based on morphological data. The strong support for section (genus) *Abildgaardia* is not surprising as it has been already considered as a separate genus by many authors (Chapter 1).

The consensus tree for species of *Fimbristylis* was better resolved in the combined analysis than the trees resulting from the individual analyses of ITS (Fig. 4.2) and *trn*L–F (Fig. 3.7) and provides a reasonable basis for assessing infrageneric classification of this genus (Table 4.5; also see chapter 6). Species from 16 out of 18 sections within *Fimbristylis*, as recognised by Kern (1974), have been sequenced. However, only nine sections have more than one species sequenced in this study. In general, the infrageneric classification suggested by Kern (1974) is not supported in either separate (Figs 3.7 and 4.2) or combined molecular analyses in this study with only three out of the nine sections with more than one sequenced species making monophyletic clades in the combined analysis (Fig. 4.6, Table 4.5).

Section *Cymosae* is monophyletic according to the combined tree while it is not monophyletic in any of the individual DNA region analyses (Figs 3.7, 4.2, and 4.6).Only two species of section *Cymosae* have been sequenced and to be sure whether the section is strongly supported more species within it should be sequenced. The weak support for this monophyly (only 63%) also suggests the need for more sampling within the section.

Table 4.5 provides:

1. a list of sections used in this study, which helps further understanding of the relatively satisfactory number and range of sections that have been used to test Kern's sections.
2. the grouping of the species of *Fimbristylis* (and *Abildgaardia* and *Arthrostylis*) sampled in this study, into these sections.
3. a quick and summarised view of the fact that according to the results of this study it is very unlikely that the sections introduced by Kern have been defined correctly.

Table 4.5. The placement of the species sampled in this study, in sections according to Kern (1974), with interpolation of Australian endemics by K.L. Wilson (pers. comm.)

➳ = monophyletic. — = non-monophyletic. n.a. = not available

Species	Section no.	Section name	Monophyletic*
Fimbristylis microcarya	1	*Trichelostylis*	n.a.
F. littoralis	2	*Miliaceae*	n.a.
F. cymosa	3	*Cymosae*	➳
F. sericea	3	*Cymosae*	➳
F. cephalophora	4	*Tenerae*	—
F. compacta	4	*Tenerae*	—
F. schultzii	4	*Tenerae*	—
F. lanceolata	5	*Leptocladae*	—
F. laxiglumis	5	*Leptocladae*	—
F. neilsonii	5	*Leptocladae*	—
F. rara	5	*Leptocladae*	—
F. densa	6	*Heleocharoides*	—
F. pauciflora	6	*Heleocharoides*	—
F. pterygosperma	6	*Heleocharoides*	—
F. arnhemensis	7	*Signatae*	n.a.
Abildgaardia macrantha	8	*Abildgaardia*	➳
A. ovata	8	*Abildgaardia*	➳

Table 4.5. *(continued)*

Species	Section no.	Section name	Monophyletic*
A. oxystachya	8	*Abildgaardia*	➳
A. pachyptera	8	*Abildgaardia*	➳
A. schoenoides	8	*Abildgaardia*	➳
A. vaginata	8	*Abildgaardia*	—
F. cinnamometorum	9	*Fuscae*	n.a.
Fimbristylis ferruginea	10	*Dichelostylis*	—
F. sieberiana	10	*Dichelostylis*	—
F. tristachya	10	*Dichelostylis*	—
F. bisumbellata	11	*Fimbristylis*	—
F. dichotoma	11	*Fimbristylis*	—
F. velata	13	*Pogonostylis*	n.a.
F. polytrichoides	14	*Neodichelostylis*	n.a.
F. acuminata	15	*Nutantes*	—
F. punctata	15	*Nutantes*	—
F. tetragona	16	*Mischospora*	n.a.
Actinoschoenus spp.	18	*Actinoschoenus*	➳
Crosslandia setifolia	n.a.	n.a.	n.a.
Arthrostylis aphylla	n.a.	n.a.	n.a.
Arthrostylis sp.	n.a.	n.a.	n.a.

*According to the results of this study on combined analysis

4.4.3 Conclusion

The usefulness of ITS and the *trn*L–F IGS and *trn*L intron for answering questions about the close relationships as well as deep branches in the Abildgaardieae is limited. The relatively little resolution at the terminal and basal branches might be explained by the suggestion that Abildgaardieae has evolved recently (Bremer 2002). Another justification for the low resolution might be peripheral segregation and disintegration of populations (Soltis and Soltis 1995). This process is a form of allopatric speciation, where subpopulations on the periphery of the range occupied by

the main population may become isolated, and begin to diverge genetically until reproductive isolation occurs. This process occurs because peripheral populations are likely to be small, so that genetic drift, as well as inbreeding effects, will enhance their new genetic identity (Page and Holmes 1998).

The information content of *trn*L–F is greater than that of ITS even though both regions are considered to be relatively variable. Used together, the overall resolution of the tree was higher than the *trn*L–F tree alone but bootstrap support was generally lower, and there was more indication of Abildgaardieae being paraphyletic.

To test the support for the sections within *Fimbristylis* a combined analysis of molecular and carefully scored morphological data with high bootstrap support in individual analyses is worth considering. Pollen morphological data are studied in Chapter 5 to investigate these sections as well as the relationships within Abildgaardieae and Arthrostylideae.

Higher resolution as well as stronger support of terminal branches can possibly be achieved by using faster evolving genes. One of the considerable problems can be undersampling, which is a good reason for collecting samples extensively. Using combined molecular and non-molecular information along with more intense sampling also might be a useful method to explain the correlations of the basal branches of Abildgaardieae and close genera like *Arthrostylis* and *Actinoschoenus*. A separate study of morphological and anatomical characters is currently being carried out by another UNE PhD student, Ms Kerri Clarke, and the ultimate aim is to use all of our data for such a broad combined analysis. It will be essential to concentrate on genera that show paraphyly in my study such as *Abildgaardia*, *Crosslandia*, and the most important and diverse genus of Abildgaardieae, *Fimbristylis*, and on the genera not included in this study (*Nemum*, *Nelmesia*, *Tylocarya*, and *Trachystylis*). It is also important to avoid single-taxon sampling. Even if the group in question is monotypic, an effort should be made to sequence different individuals (preferably from different geographic areas) to establish a strong clade that will remain stable in the tree. The inclusion of a few species from tribe Cypereae and more species of Schoeneae is also of interest.

In conclusion, our understanding of phylogenetic relationships within the Abildgaardieae is steadily improving. Careful taxon selection along with emphasis on the right gene, non-coding region and morphological characters should produce a comprehensive view in future.