Systematics Project

Phylogenetic investigations of the African Restionaceae using *rbc*L data

Carl Morrow
Botany Honours 1994
Supervisor: Mr. Nigel Barker

ABSTRACT

DNA sequences for the *rbc*L gene were determined for four representatives of the African Restionaceae. These data were analysed along with six *rbc*L sequences (representing the rest of the Poales with *Carex* as the outgroup) from GenBank. Both parsimony analysis and distance measures were used. The topologies of the resultant trees were similar with the Restionaceae being a well supported group. Conflict in genus groupings, within the Restionaceae, was noticed when the trees from the morphological data set and the *rbc*L data set were compared. Reasons for this conflict are discussed. Times of divergence between different taxa were calculated and these compared closely to the ages published elsewhere.

INTRODUCTION

The recent developments in the field of molecular biology have made it possible for systematists to collect phylogenetic information from a wide range of plant species. An example of this information gathering can be seen in Chase *et al.* (1993). Here, a collaboration of 42 scientists made it possible to create a data set containing the DNA sequences of the *rbc*L gene from 499 species of seed plants. The size of the data set meant that it was impossible to exhaustively analyse the information and so the result of the study was a first approximation of the phylogeny of the seed plants. At the end of the paper the authors encourage other workers to take up the challenge and perform more restricted *rbc*L based studies on smaller groups of plants so that, ultimately, a better approximation of the phylogeny can be obtained. The project presented here is performing this important function for the African Restionaceae.

THE AFRICAN RESTIONACEAE

The Restionaceae are a group of evergreen, rush-like plants with erect photosynthetic culms and leaves reduced to sheaths (Linder, 1984). All of the species are wind pollinated and, with a few exceptions, the flowers are dioecious. The flowers are small and aggregated into spikelets.

The family is predominantly distributed in the southern hemisphere. In Africa, there are approximately 320 species of Restionaceae grouped into 19 genera (Linder, 1984; Linder pers. comm.) The African Restionaceae are primarily restricted to the South Western Cape where they can dominate the vegetation of a particular area (Linder, 1984; Linder, 1991[b]). There are about 100 species of the Restionaceae in Australia, 3 in New Zealand, 1 in Malaysia and South-east Asia and one in Chile (Linder, 1984). This distribution is shown Figure 1.

The African Restionaceae are an old group of plants. The antiquity of the group was demonstrated by the discovery of restiod pollens 60 - 65 million years old (Linder, 1991[b]).

Recently, the taxonomy of this group along with its allied families has received considerable attention, although knowledge about biological aspects of the family is poor (Linder, 1991[b]). It is generally accepted that the African Restionaceae are a monophyletic group within the family (Linder, 1986; Linder 1991[b]). The affinities of this family with related groups are still somewhat unclear. The Restionaceae have been placed in the order

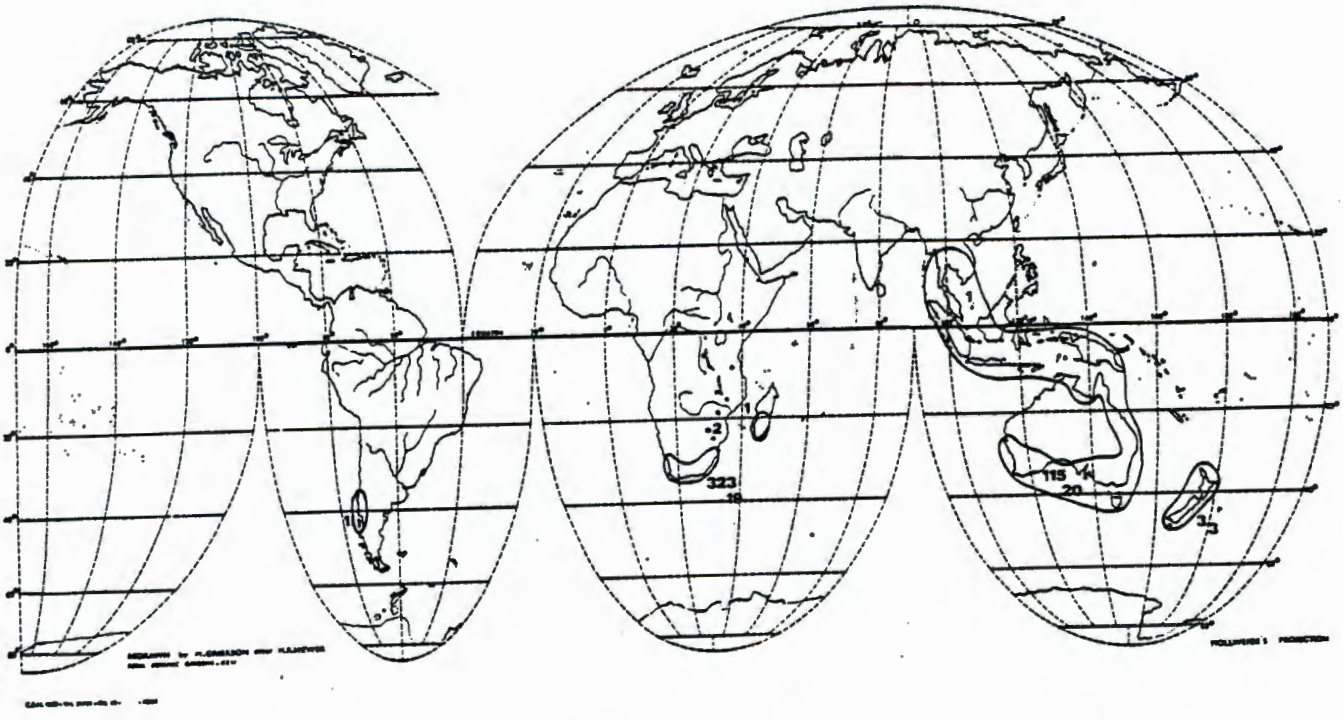


Figure 1: The worldwide distribution of the Restionaceae. The number of genera and species in each area are indicated on the map with single numbers indication only one genus present. (from Linder, 1991[b])

the Restionales along with the Ecdeiocoleaceae, Anarthriaceae, Centrolepidaceae, Joinvilleaceae and Flagellariaceae. Dahlgren and Clifford (1982)(cited in Linder, 1991[b]) contend that the order Restionales is the sister to the Poales and that these two groups should be combined. It is generally agreed that the Poaceae and the Restionaceae are related to one another but the precise positioning of the various sister groups is not clear (Linder, 1991[b]). Kellogg and Linder (in press) analysed the relationships within the Poales by combining data from 6 data sets (Morphology, cpDNA structure, *rbc*L, *rpo*C2, cpDNA restriction analysis and ribosomal RNA). Analysis of this combined data set revealed the following: the Centrolepidaceae may be derived from within the Restionaceae, the Restionaceae and Ecdeiocoleaceae are a monophyletic group and the position of the Anarthriaceae in relation to the other families has yet to be convincingly resolved (Figure 2). The complex of families comprising the Poales are quite distinct from the Cyperales. For this reason, *Carex hostiana* (a member of the Cyperales) has been used as the outgroup in the analyses that were performed in this study.

The taxonomy of the Restionaceae has been difficult due to the shortage of convenient macro-morphological characters (Linder, 1984). A phylogeny of the Restionaceae was

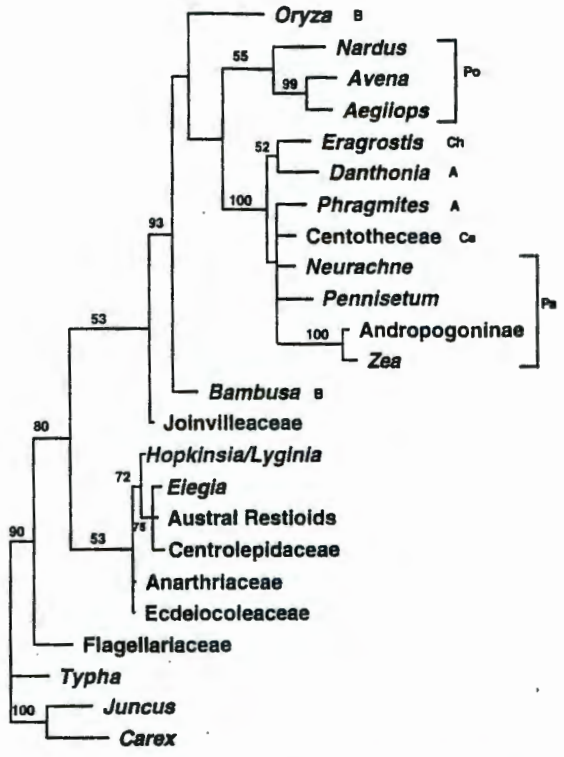


Figure 2: The strict consensus tree showing the phylogeny of the Poales based on the 6 data sets outlined above. (from Kellogg and Linder, in press)

ascertained (Linder, 1984) and subsequent to this initial publication other characters were studied and added to the data set. An updated phylogeny was published (Linder, 1991[a]) and this is the phylogeny (Figure 3) that is used as the basis for the discussion of the results achieved in this study.

AIM OF THIS PROJECT

The existing phylogeny for the African Restionaceae (see Figure 3) is weak at nodes 2,3,5 and 7. This is due the fact that there are not many macro-morphological characters for the Restionaceae (Linder 1984) The aim of this project was to construct a phylogeny of the African Restionaceae using *rbc*L data. The resultant tree was then compared to the existing phylogeny. The data generated in this study is also a useful addition to the established *rbc*L sequence data base.

USING MOLECULAR DATA IN SYSTEMATICS

The use of nucleotide sequence data in systematics has received much attention. A key advantage of nucleotide sequences is the 4 distinct states (G, A, T or C) that each character can hold. The separateness of the characters means that they can be coded simply and unambiguously. This simplicity can be problematic in that if two sequences are the same

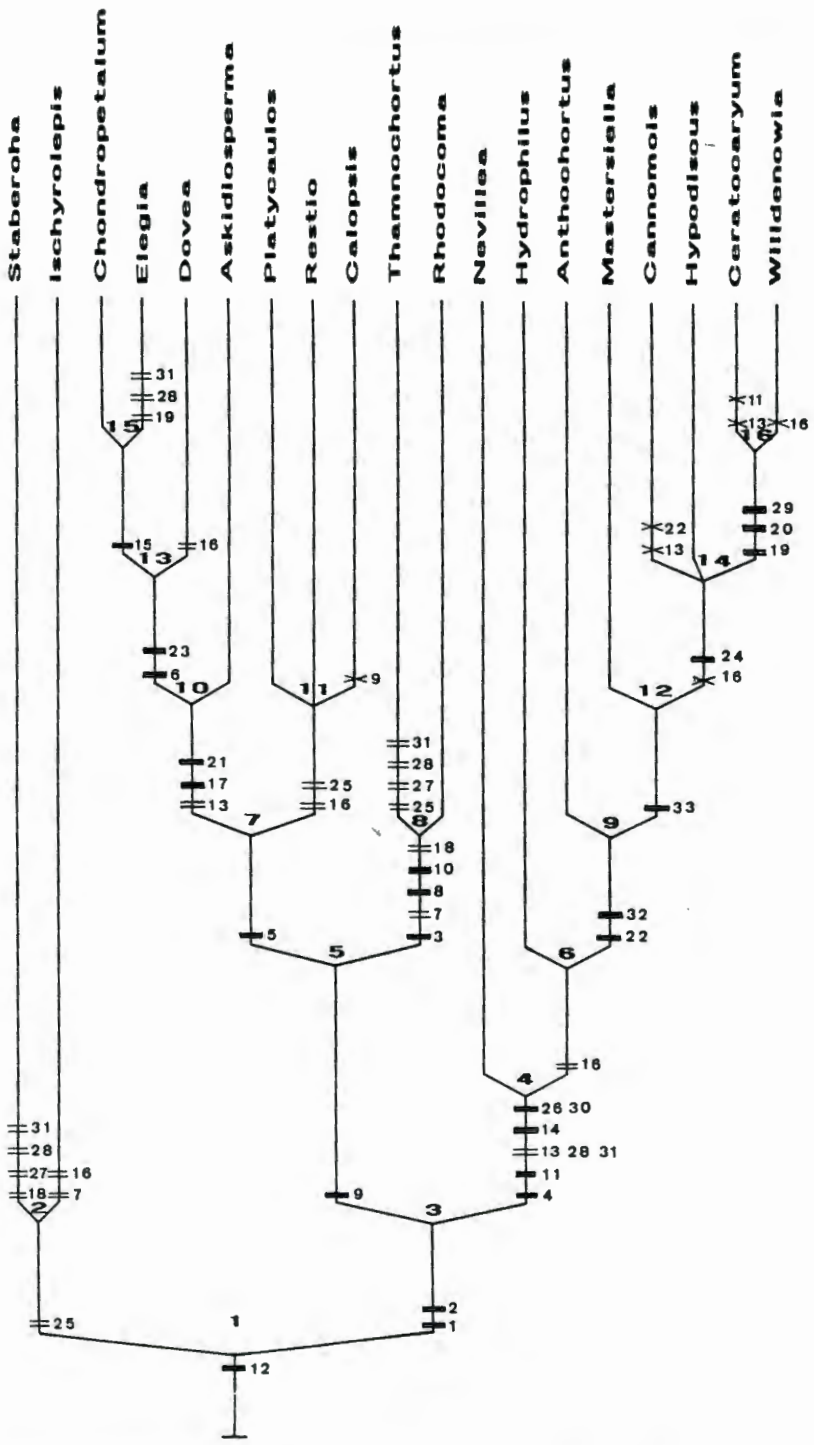


Figure 3: The consensus cladogram for the African Restionaceae. The large numbers are labelling the nodes while the smaller numbers relate to the characters presented in the original paper. (from Linder, 1991[a])

at a particular position one cannot be certain that they have always been the same. Reversals of state are undetectable. That is to say: a change from an A to a G and back to an A is undetectable. To alleviate this problem, one has to use long stretches of sequence so that one can sense the average change over time and a sufficiently conserved study molecule is chosen so that none of the codon positions are saturated.

Other advantages of molecular data are as follows (Penny *et al.*, 1990): Sequences evolve at different rates and so they can be specifically chosen to allow for the resolution of a particular historical event. Along with this, sequences often have a wide scope or domain meaning they occur in a wide range of organisms which makes it possible to analyse the relationships between widely diverged lineages. With the correct choice of study molecule, one can be reasonably sure that phylogenetically useful characters will be obtained from the organisms in a particular study. Finally, Penny *et al.* (1990) state that the cost per character from a DNA sequence is relatively low and it is getting progressively cheaper. This makes sequence data an attractive option in phylogenetic studies but the limitations of this data must also be taken into account.

There are numerous problems in weighting transitions (purine to purine or pyrimidine to pyrimidine changes), transversions (purine to pyrimidine or vice versa) and whether or not there is such a thing a neutral changes in the codons considering the possibility of transfer RNA (tRNA) biases (Clegg, 1993; Chase *et al.*, 1993). West and Faith (1990) report that transversions are rarer than transitions and so, during certain analyses, the data should be coded as purines or pyrimidines without specifying the actual base in order to prevent dilution effects by the number of transitions. This is undesirable because one loses half the information content of the sequence one has and Penny *et al.* (1990) point out the importance of maintaining the maximum amount of information possible.

Nucleotide sequences generate a large number of characters that can be used in the phylogeny reconstruction. The combination of the morphological (non-molecular) and molecular data has caused systematists to consider the informativeness of the different characters. Morphological characters generally encompass the product of a series of genetically controlled steps that ultimately results in the structure being examined. This would indicate that the information content of this character far outstrips the information content of a single nucleotide position on a sequence. Systematists are concerned that the large number of individually weaker characters from molecular data would override the richer morphological characters. This leads to the subjective institution of a weighting system to resolve this conflict. It should be remembered, however that assumptions are still being made if on does not use a weighting scheme.

THE GENE OF INTEREST: *rbc*L

The chloroplast encoded ribulose-1,5 bisphosphate carboxylase/oxygenase (*rbc*L) gene has been of interest to botanists for a considerable time as it is a component of Rubisco, a crucial enzyme responsible for the fixation of carbon dioxide in the Calvin cycle (Clegg, 1993) The Rubisco holoenzyme consists of 8 large and 8 small subunits. The large subunit (*rbc*L) is encoded on the chloroplast DNA while the small subunit's code is present in a multigene family on the nuclear genome (Clegg, 1993).

Originally the chloroplast genome received a lot of attention because it was easier to manipulate cpDNA than nuclear DNA (Zurawski and Clegg, 1993). Standardised procedures were developed that allowed researchers to study aspects of the chloroplast genome. This resulted in a large body of information on the molecular structure, evolution and organisation of the chloroplast genome which provided a suitable framework in which systematists could start their phylogenetic studies (Clegg and Durbin, 1990).

The *rbc*L gene is suitable for resolving higher level relationships within the seed plants because the gene is highly conserved (Clegg, 1993). The alignment of the *rbc*L sequences derived from different organisms is generally unambiguous although insertions and deletions of codons occasionally occur in the extreme 3' end of the gene (Clegg, 1993). Furthermore the gene is not interrupted by introns. Normally the gene is present as a single copy on the chloroplast. It was found that most of the other copies identified were non-functional pseudogenes although it was discovered that *Canella* and members of the Ulmaceae contained two, reading frame intact copies of the *rbc*L gene (Chase *et al.* 1993). Successful alignment of the gene is necessary to obtain the correct positional homology.

Phylogenies derived from molecular sequences are tracking the history of the particular sequence and this history can be significantly different to the whole organism's phylogeny (Li and Graur, 1991). Duvall *et al.* (1993) found that the tree topologies obtained from *rbc*L sequences were fundamentally similar to morphologically based phylogenies.

Clegg (1993) discusses the suitability of applying the molecular clock hypothesis to the *rbc*L gene. The theory states that molecules evolve in a regular, clock like manner over time. If one is able to calibrate this "clock" through the use of supporting evidence such as fossils or the use of relative rate tests that utilise the geometry of three species trees (and are thus independent of the hypotheses present in palaeontology) then one would be able to

precisely age all of the internal nodes present on the cladogram. Clegg (1993) argues that the time scale that is superimposed on the phylogeny should be based on generation time and not physical time. This is impractical, however, because of the great differences in generation time between the different organisms. Clegg (1993) cites a study that showed significant differences in the rates of change in the *rbc*L gene amongst the monocots. Interestingly the rate of substitutions were correlated to the minimum generation time of the taxa in question.

MATERIALS AND METHODS

SOURCES OF THE SEQUENCES

Fresh material of a number of species from the Restionaceae was collected and dried in packets of silica gel (Chase and Hillis, 1991).

The *rbc*L sequence of *Elegia* along with five outgroup sequences were taken from GenBank. The species used in this study can be seen in Table 1.

Table 1: Table showing the origin of the *rbc*L sequences used. Where applicable, the GenBank accession numbers appear in parenthesis.

Species	Source
Askidiosperma sp.	Kirstenbosch Botanic Gardens
Ischyrolepis subverticilata	Kirstenbosch Botanic Gardens
Hypodiscus aristatus	Morrow 05; Table Mountain
Nevillea sp.	Morrow 08; Kogelberg
Elegia sp.	GenBank (L12675)[1]
Joinvillea plicata	GenBank (L01471)
Oryza sativa	GenBank (D00207)
Flagellaria indica	GenBank (L12678)
Juncus effusus	GenBank (L12681)
Carex hostiana	GenBank (L12672)

Note:
1: The GenBank accession numbers were obtained from the appendix of the Annals of the Missouri Botanic Garden 80(3).

DNA MANIPULATIONS

Total DNA was extracted using the CTAB method of Doyle and Doyle (1987). Amplification of the *rbc*L gene was achieved by means of the polymerase chain reaction (PCR) using primers Z-1 and Z-1375R (primer sequences supplied by G. Zurawski, DNAX Corp.). The PCR parameters were 35 cycles of 10 seconds at 94°C (denaturing), 5 seconds at 35°C (annealing) and 2 minutes at 72°C (extension) using a Techne PHC2 thermocycler. The thermocycler was water cooled using water from a Techne chiller. The amplified

product was electrophoresed in 0.8% agarose, stained using ethidium bromide and visualised using a UV transilluminator. The desired PCR product was cut from the gel and the DNA extracted using the Qiagen-Qiaex system.

Sanger-dideoxy sequencing (Sanger *et al.*, 1977) was carried out using the Sequenase kit (USB) and α^{35}SdATP. Only one strand of the DNA was sequenced using primers supplied by G. Zurawski (DNAX Corp.). The sequencing products were electrophoresed on a 6% denaturing PAGE (Polyacrylamide Gel Electrophoresis) gel. The sequences were visualised using autoradiography.

SEQUENCE ALIGNMENT AND DATA MATRIX GENERATION

The DNA sequences were entered into **DAPSA** version 2.8 (E.H. Harley, 1994) and aligned using the multiple manual sequence alignment option. The new Restionaceae sequences were aligned against the *Elegia* sequence. Differences between any of the five Restionaceae sequences were noted and the original autradiographs were then consulted to see if these differences were real. This process was repeated twice ensuring confidence in the four new sequences.

The sequences were checked further by converting them into an amino acid sequence. The absence of unknown codons and stop codons meant that the sequences were fundamentally correct.

Once the data set of all ten sequences was properly aligned with the primer regions and non-overlapping ends removed, phylogenetically informative sites were obtained using **DAPSA** (E.H. Harley, 1994). All of the base positions were used in the search for informative sites.

A morphological data matrix was derived from the matrix presented in Linder (1991[a]). This matrix contained only those phylogenetically informative characters applicable to the genera used in this study. The matrix and characters used are presented in Appendix 1.

DATA ANALYSIS

The phylogenetically informative site from the *rbc*L data were subject to parsimony analysis using **Hennig86** version 1.5 (J.S. Farris, 1988) and bootstrap analysis using **Ra** (J.S. Farris, 1994)

The data was also analysed using distance methods. The neighbour joining method of Saitou and Nei (1987) was performed using **Mega** (Kumar *et al.*, 1993). Both the Jukes & Cantor one parameter model and Kimura's two parameter model were used in the analysis. A bootstrap analysis of the resultant trees was also performed.

Parsimony analysis of the morphological data was performed using **Hennig86** version 1.5 (J.S. Farris, 1988)

RESULTS

Four *rbc*L sequences were obtained from: *Ischyrolepis subverticilata*, *Hypodiscus aristatus*, *Askidiosperma sp.* and *Nevillea sp.*. These new sequences appear in appendix 2. Two of the sequences were not complete. The *Askidiosperma* sequence was unreadable in two regions; from position 442 to 471 and position 705 to 740. The *Nevillea* sequence contains a gap from position 696 to 723.

A total of 145 phylogenetically informative sites were found in the DNA sequences of this 10 species data set. The matrix of the characters used are presented in appendix 3.

PARSIMONY ANALYSIS

A most parsimonious trees, using either the *rbc*L or morphological data, were derived using the 'ie' option in **Hennig86** version 1.5 (Farris, 1988) and are shown in Figure 4 and 5. Figure 6 shows the bootstrap analysis performed on the *rbc*L tree. The analysis was carried out using **Ra** (Farris, 1994). Interestingly the topology of the bootstrap tree is different to that obtained from the 'ie' option.

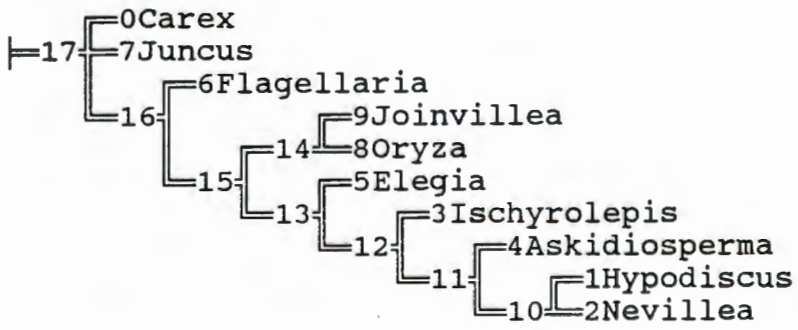


Figure 4: Shortest rooted tree, using the *rbc*L data, of the groups under study using Wagner parsimony. The outgroup is *Carex* and the length of the tree is 253 steps with a ci of 73 and ri of 76.

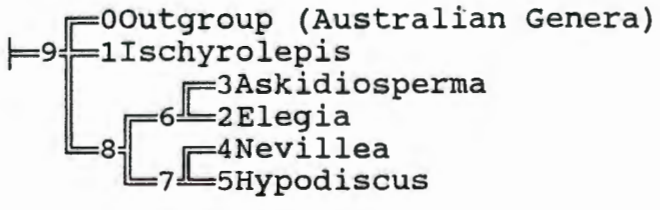


Figure 5: The shortest rooted tree using the morphological data derived from Linder (1991[a]). The length of the tree is 15 steps with a ci of 86 and a ri of 86

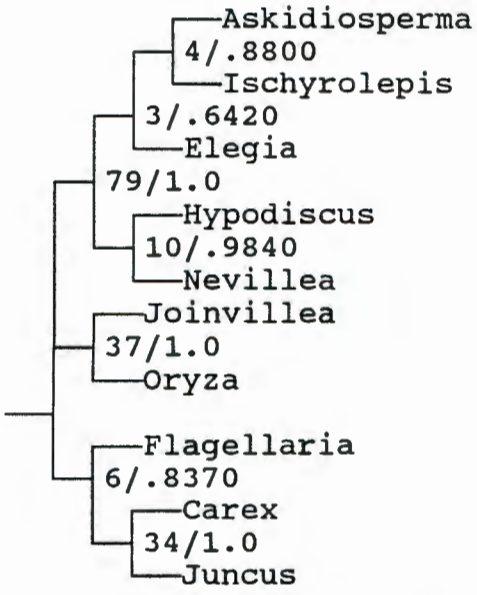


Figure 6: The results from the bootstrap analysis of the shortest tree derived from the *rbc*L data using Ra (Farris, 1994). 1000 replicates were performed and the bootstrap values are quoted in the second position at each node. The values are probabilities with a maximum possible value of 1. Nodes with values of less than 0.5 have been collapsed. The first value is a decay value which gives a measure of number of steps that the parsimony must be relaxed by before that node will collapse.

In order to see the influence that the other groups have on the topology of the tree, an analysis was performed on a data set containing only the Restionaceae with *Joinvillea plicata* as the outgroup. Only 18 phylogenetically informative sites were identified. The resulting tree, Figure 7, has the same topology as the corresponding section of the tree presented with the bootstrap analysis (Figure 6). A search for informative sites for the Restionaceae alone resulted in a matrix with 14 characters.

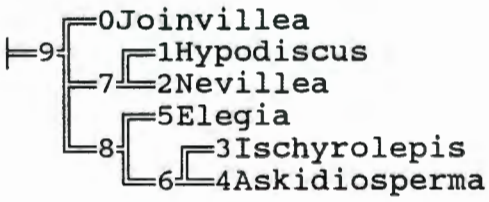


Figure 7: Shortest rooted *rbc*L tree of just the Restionaceae using Wagner parsimony and *Joinvillea sp.* as the outgroup. The length of the tree is 29 steps with a ci of 79 and ri of 75.

The topology of the *rbc*L tree was modified to reflect that of the morphological tree (Figure 8). The resultant tree is only 5 steps (2%) longer than the shortest *rbc*L tree (Figure 4)

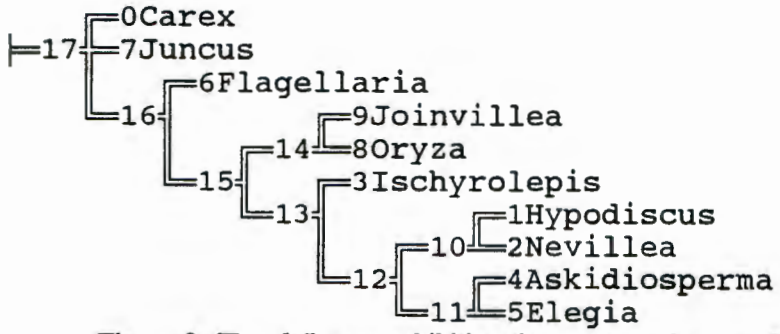


Figure 8: The *rbc*L tree exhibiting the same topology as the morphologically based tree presented in Linder (1991*). The length is 258 steps.

DISTANCE METHODS

Distance methods were also used in the analysis. The neighbour joining method of Saitou and Nei (1987) was used utilising both the Jukes & Cantor one parameter model and Kimura's two parameter model. Bootstrap analysis of 500 replicates was performed on and the results of all of these tests can be seen in Figures 9 to 12.

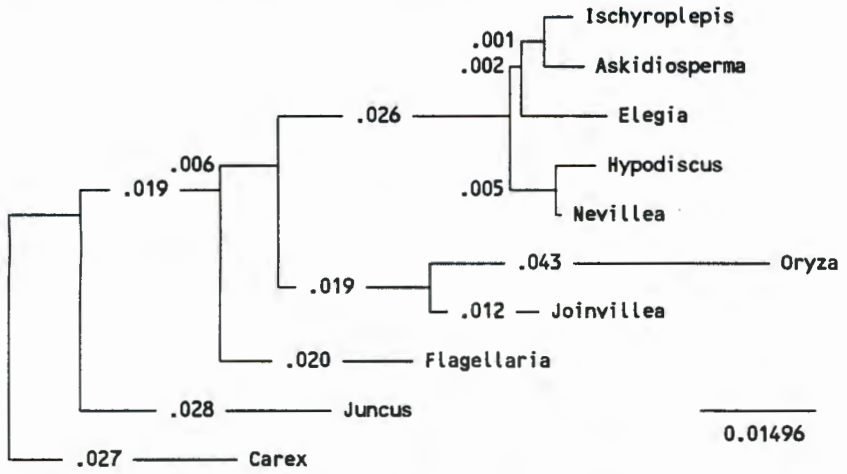


Figure 9: Tree obtained from Saitou and Nei's (1987) Neighbour-Joining method using the Jukes and Cantor one parameter model.

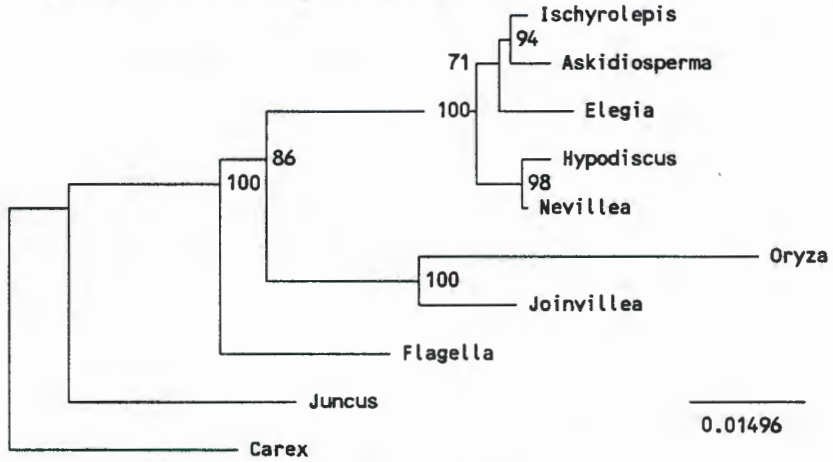


Figure 10: Results of the bootstrap analysis (%) performed on the tree above (Figure 9).

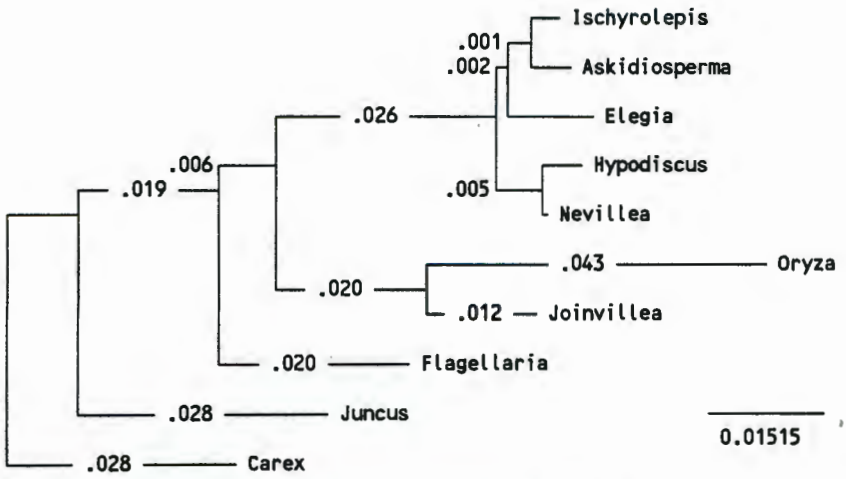


Figure 11: Tree obtained from Saitou and Nei's (1987) Neighbour-Joining method using Kimura's two parameter model.

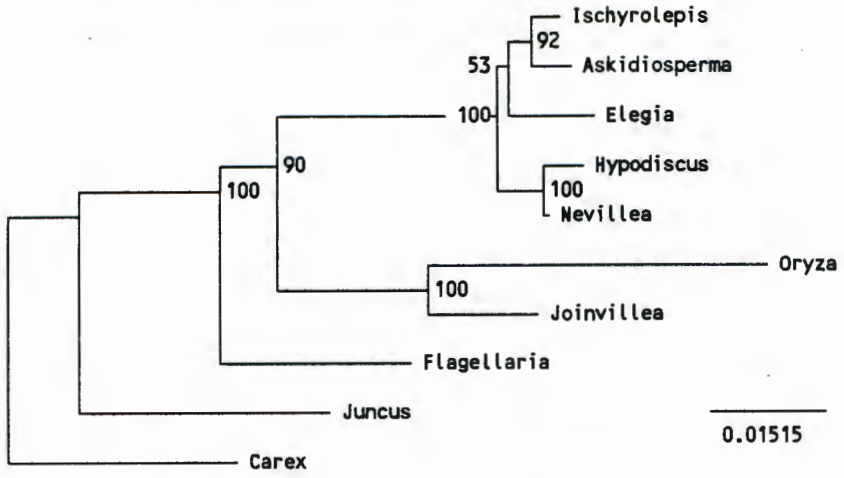


Figure 12: Results of the bootstrap analysis (%) performed on the tree above (Figure 11).

DISCUSSION

COMPARISON OF THE DIFFERENT ANALYSES

The trees obtained using the neighbour-joining method (Saitou and Nei, 1987) (Figures 9 and 11) were topologically identical and the bootstrap values were very similar. The only notable difference is the support for *Ischyrolepis*, *Askidiosperma*, *Elegia* clade. The Jukes and Cantor assumptions result in a well supported (71%) clade while the analysis using Kimura's 2 parameter model results in a weakly (53%) supported clade. Both of these values are significantly lower than the bootstrap values associated with the other nodes.

Comparison between the distance analyses and parsimony analysis show that the tree generated in the bootstrap analysis of the most parsimonious tree has an identical topology to the distance trees. It is strange that the most parsimonious *rbc*L tree has a different topology to the tree resulting from the bootstrap analysis of the *rbc*L tree.

Missing data for one or more of the taxa can influence the topology of the tree derived from parsimony analysis. The bootstrap analysis can also be affected if the characters present in the gap are repeatedly sampled during the analysis. In this study, the *Nevillea* and *Askidiosperma* sequences were not complete, however, only 1 informative site for Restionaceae (10 in the whole *rbc*L data set) was present in these regions and so it is unlikely the tree topology was influenced by these gaps in the data.

COMPARISONS BETWEEN THE *rbc*L AND MORPHOLOGICAL TREES

The Restionaceae group is very strongly supported in all of the analyses, evident by the bootstrap values of 100%. The *Hypodiscus* - *Nevillea* clade is also strongly supported by all of the analyses performed. Comparison with complete morphological data set (Figure 3) shows that this grouping (*Hypodiscus* - *Nevillea*) corresponds with the extremely well supported node 4 (Linder, 1991[a]). The other three Restionaceae genera (*Elegia*, *Ischyrolepis* and *Askidiosperma*) have been grouped in a manner that contradicts the phylogeny presented in Linder (1991[a]). If the *rbc*L gene is tracking the morphological lineage closely, as is indicated by the affinity between *Hypodiscus* and *Nevillea* as well as the statement by Kellogg and Linder (in press) that, in the Poales, the molecular trees are similar to organismal trees, then *Elegia* should be grouped with *Askidiosperma*. This is not the case and reasons need to be sought to explain this discrepancy.

It is possible that the "Elegia" sequence is not from an *Elegia*, but another restio. The voucher specimen of *Elegia* was collected in a botanical garden and it is only identified as *Elegia sp.* (Chase *et al.*, 1993). Another possibility is that the sequence was incorrectly read from the autoradiograph. The only way to check for this error would be to study the original autoradiograph or repeat the sequencing with a different, correctly identified, specimen.

The *Ischyrolepis* and *Askidiosperma* specimens where collected from Kirstenbosch Botanic Gardens. The *Ischyrolepis* plants were correctly labelled but the *Askidiosperma* plants were labelled as *Chondropetalum ebracteatum* although there were identical plants nearby that were labelled *Askidiosperma andreaeanum*. Unfortunately the plants were not flowering and so a positive identification was not possible. Both *Chondropetalum* and *Askidiosperma* are in the same clade, stemming from node 10 (Figure 3) and so the results of the analyses should theoretically be unaffected by this confusion (H.P. Linder, pers comm.)

If the sequences are correct then the difference between the trees needs to be reconciled.

STRICT OR RELAXED PARSIMONY?

As can be seen in Figure 4 the most parsimonious tree is 253 steps long. If the parsimony is relaxed by just five steps, a very minor relaxation considering the number of characters used, one gets the tree in Figure 8 that has the same topology as the morphological cladogram, Figures 3 and 5. The fundamental assumption of parsimony is that character evolution must be explained in the simplest manner possible (Hillis and Moritz, 1990). Using this argument means that one should use the tree presented in Figure 4 as the hypothesised phylogeny for this group, however it should be added that other, near parsimonious, solutions are topologically the same as the trees derived from other data sets.

The relaxation of the parsimony of the morphologically based tree was also studied. The most parsimonious tree derived from the condensed morphological data set (Figure 5) corresponded to the topology of the tree presented in Linder 1991[a]). When the branches of the tree were swapped to yield the topology of the most parsimonious *rbc*L tree, the length of the morphological tree was increased from 15 steps up to 21 steps.

By looking at straight proportion one can see that branch swapping has a significantly smaller effect on the *rbc*L tree than the morphologically based tree. This is a misleading statistic, however, due to the tremendous difference in the number of characters used in each of the analyses. Furthermore, it has been shown that the *rbc*L sequence results in only 14 phylogenetically informative sites within the Restionaceae this would indicate that the five step parsimony relaxation of the molecular tree is also a very significant change. A more detailed analysis would reveal the full extent of the effects that branch swapping has on the parsimony of the two trees (H.P. Linder, pers. comm.)

A total analysis using both data sets could reveal information about the relationship between the taxa as well as conflict between data sets. Unfortunately this was not possible in the present study due to the different outgroups (*Carex hostiana* and the Australian Restionaceae) used in the two data sets.

MONOPHYLY OF THE AFRICAN RESTIONACEAE

Linder (1986; 1991[b]) present evidence for the monophyly of the African Restionaceae. Unfortunately this present study cannot test this hypothesis because only African Restionaceae are represented. Sequences of Australian Restionaceae have been ascertained (H.P. Linder, pers. comm.) but unfortunately it was not possible to obtain them for use in this project.

AGE OF DIVERGENCE

The rate of *rbc*L evolution for the grass family has been calculated to be $\approx 1.3 \times 10^{-9}$ substitutions per synonymous site per year (Clegg, 1993). Using this value an approximate time of divergence was calculated for the split between Cyperales (*Carex hostiana*) and the Poales (2 Restionaceae) as well as between the Poaceae (*Oryza sativa*) and the Restionaceae. *Hypodiscus* and *Ischyrolepis* were used to represent the Restionaceae because the identity of the source material is definite and the sequences are complete meaning that errors resulting from missing data points are avoided.

An approximation of the number of synonymous substitutions was made by counting the total number of nucleotide differences between the two sequences and subtracting the number of differences observed between the amino acid sequences derived from the nucleotide sequences. The number of synonymous changes for the *rbc*L gene per year is $\approx 1.8 \times 10^{-6}$. This was calculated by multiplying the rate of change per site presented by

Clegg (1993) by 1365 (the number of nucleotides used in this study).

Division of the number of synonymous substitutions counted by the rate of change for the gene per year gives an approximate value for the age of the divergence between the two lineages. A summary of the results is presented in Table 2.

Table 2: The calculated age of divergence between different groups in the Poales.

Groups being compared	# of nucleotide differences	# of amino acid differences	Difference	Approximate age of divergence (in millions of years)
Carex & *Hypodiscus*	130	20	110	62.0
Carex & *Ischyrolepis*	131	19	112	63.1
Oryza & *Hypodiscus*	132	19	113	63.7
Oryza & *Ischyrolepis*	130	12	118	66.5

These divergence ages calculated from the *rbc*L substitution rates, are remarkably near to the value of 60 to 65 million years ago presented by Linder (1991[b]) for the age of the earliest known Restiod pollens. It is interesting that both the Poales - Cyperales and the Poaceae - Restionaceae splits are indicated to be at the same time. This indicates a rapid radiation of the different lineages which is consistent with Duvall *et al.* (1993)

MAJOR CRITICISMS

One of the important problems associated with molecular based projects is the extremely thin sampling. In this project a single individual, sometimes a single culm of an individual, was used to represent an entire genus. This assumes that the *rbc*L sequences of all of the plants within the genus are identical. There is strong evidence supporting this assumption. It was shown by Zurawski and Clegg (1993) that there were no sequence differences between different barley cultivars and Clegg (1993) states that measurement of mutation rates of the gene indicate that differences within genera are unlikely. In the present study, it was found that there were only 39 nucleotide changes of which 14 were phylogenetically informative, across the Restionaceae and so infrageneric differences seem highly improbable. It would

still be better to sequence two specimens per genus to preclude the possibility of using a nonrepresentative mutation in the analysis.

Although there are 145 phylogenetically informative sites in this data set of 10 taxa there are only 14 informative sites within the Restionaceae. This is a very low number considering the fact that sequences 1365 nucleotides long were studied. From the morphological phylogeny (Figure 3) one can see that *Nevillea* and *Hypodiscus* are terminal of the clade circumscribed by node 4. The *rbc*L sequences of these two genera contain 6 nucleotide differences meaning they are 99.5% similar, as calculated by **DAPSA**. Indications are that the *rbc*L is actually not variable enough to resolve genus level relationships within the African Restionaceae (H.P. Linder, pers. comm.). This hypothesis is supported by Clegg (1993) where it is shown that the *rbc*L sequence is most suitable for resolving relationships in the range of $4x10^8$ to $1x10^8$ years. The Restionaceae family is thought to be between 60 and 65 million years old (Linder, 1991[b]) and so it can be seen that this age is below the range referred to in Clegg (1993). Duvall *et al.* (1993) cite previous generic level studies within the Arecaceae, Bromeliaceae, Poaceae and Zingerberiales that lacked resolution because of the low substitution rate of *rbc*L. Clegg (1993) criticises the failure of the Poaceae study by saying that a very distant outgroup was used which resulted in the compression of the ingroup. N.P. Barker (unpublished results) has demonstrated that the *rbc*L gene is able to determine genus level relationships within the Poaceae. These differences in resolution may also be because of variable evolution rates of *rbc*L in the different lineages.

A further problem, highlighted by the questionable identity of the *Elegia* sequence, is the faith one must place on the quality of the data taken from GenBank. There is no possibility of checking the sequence unless one repeats the whole sequencing procedure which takes time and is expensive. This could be alleviated to some extent if the following steps are taken. The collection of voucher specimens is very important so that disputes over potential misidentification of the specimens can be satisfactorily resolved. I feel that the original autoradiograph should also be stored with the voucher specimen. Often errors are only highlighted when one is comparing sequences and access to the original autoradiograph would be extremely useful in identifying the genuine phylogenetically informative sites. If a laboratory is using one of the automatic sequencers that are now available then one has to

rely on the integrity of the computer that read the sequence because no "hard" copy of the sequence is available.

METHODOLOGICAL PROBLEMS

It was found that a very low yield of DNA was obtained from some of the specimens. This can be remedied by using more initial plant material to ensure a better yield of DNA and the PCR step can contain a greater number of cycles which results in greater amplification of the gene of interest.

The positions of the primers were plotted onto the sequences obtained from the plants studied. It was found that there were a number of small differences between the *rbc*L sequences and the primer sequences. Primer Z-674 was the most affected with a difference of 4 nucleotides. This could account for the observation that the sequence resulting from this primer was faint and difficult to read. Better sequences can be achieved by either using more α^{35}S dATP or designing a Restionaceae-specific oligonucleotide primer for this region.

CONCLUSIONS

The resolution of the different genera in the African Restionaceae is possible using *rbc*L, but the variation found in the gene within the group is low. Considering the age of the Restionaceae and the number of phylogenetically informative characters, it is apparent that this study is operating near the limits of the resolution of *rbc*L. A test for this resolution would be to sequence *Staberoha*, *Thamnocortus* or *Rhodocoma* and *Mastersiella* or *Wildenowia* and see what the resultant tree topologies are.

Calculation of divergence times between different lineages resulted in ages similar to those supported by fossil data. This is indicating that the *rbc*L molecular clock is keeping time within the Restionaceae.

If the initial data is considered correct, some interesting groupings have been resolved. The relationship of *Ischyrolepis* with *Askidiosperma* and *Elegia* is in conflict with the relationship found using morphological data. Clearly this is a region that needs to be more thoroughly studied to resolve this conflict.

ACKNOWLEDGEMENTS

I would like to thank Nigel Barker for all the assistance he has supplied throughout the duration of this project. His enthusiastic and unending help, combined with the occasional drink in the Club made this project an enjoyable and fulfilling endeavour. A sincere thankyou also needs to go to Peter Linder for his advise and tolerant support. Further thanks need to be directed to Professor E. Harley for the use of his laboratory and to the Flora Conservation Committee of the Botanical Society of South Africa for financial support.

REFERENCES

Chase M.W. and Hillis H.H. (1991) **Silica gel: an ideal material for field preservation of leaf samples for DNA studies.** Taxon **40:215-220**.

Chase M.W., Soltis D.E., Olmstead R.G., Morgan D., Les D.H., Mishler B.D., Duvall M.R., Price R.A., Hills H.G., Qiu Y., Kron K.A., Rettig J.H., Conti E., Palmer J.D., Manhart J.R., Sytsma K.J., Michaels H.J., Kress W.J., Karol K.G., Clark W.D., Hedrén M., Gaut B.S., Jansen R.K., Kim K., Wimpee C.F., Smith J.F., Furnier G.R., Strauss S.H., Xiang Q., Plunkett G.M., Soltis P.S., Swensen S.M., Williams S.E., Gadek P.A., Quinn C.J., Eguiarte L.E., Golenberg E., Learn G.H. Jnr, Graham S.W., Barrett S.C.H., Dayanandan S. & Albert V.A. (1993) **Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbc*L** Annals of the Missouri Botanical Garden **80:528-580**.

Clegg M.T. and Durbin M.L. (1990) **Molecular approaches to the study of plant biosystematics.** Australian Systematic Botany **3:1-8**.

Clegg M.T. (1993) **Chloroplast gene sequences and the study of plant evolution** Proc. Nat. Acad. Sci. USA **90:363-367**.

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

Duvall M.R., Clegg M.T., Chase M.W., Clark W.D., Kress W.J., Hills H.G., Eguiarte L.E., Smith J.F., Gaut B.S., Zimmer E.A. & Learn G.H. Jnr. (1993) **Phylogenetic hypotheses for the Monocotyledons constructed from *rbc*L sequence data** Annals of the Missouri Botanical Garden **80:607-620**.

Hillis D.M. and Moritz C. (Eds.)(1990) **Molecular systematics** Sinauer Associates Inc.

Kellogg E.A. and Linder H.P. (in press) **Phylogeny of the Poales.**

Kumar S., Tamura K. and Nei M. (1993) **Mega: Molecular evolutionary genetics analysis, version 1.0.** The Pennsylvania State University, University Park, PA 16802.

Li W.H. and Graur D. (1991) **Fundamentals of molecular evolution.** Sinaur Associates Inc., Sunderland, Mass.

Linder H.P. (1984) **A phylogenetic classification of the genera of the African Restionaceae** Bothalia **15 (1 & 2) :11-76**.

Linder H.P. (1986) **The evolutionary history of the Poales/Restionales - A hypothesis** Kew Bulletin **42(2):297-318**.

Linder H.P. (1991[a]) **Confidence limits in phylogenies: an example from the African Restionaceae.** Taxon **40:253-256**.

Linder H.P. (1991[b]) **A review of the South African Restionaceae** Contributions from the Bolus Herbarium **13:209-264.**

Penny D., Hendy M.D., Zimmer E.A. and Hamby R.K. (1990) **Trees from sequences: panacea or pandora's box?** Aust. Syst. Bot. **3:21-38.**

Saitou N. and Nei M. (1987) **The neighbour-joining method: a new method for reconstructing phylogenetic trees.** Mol. Biol. Evol. **4:406-425.**

Sanger F., Nicklen S. and Coulson A.R. (1977) **DNA sequencing with chain-termination inhibitors.** Proc. Natl. Acad. Sci. U.S.A. **74:5463-5467.**

West J.G. and Faith D.P. (1990) **Data methods and assumptions in phylogenetic analysis.** Aust. Syst. Bot. **3:9-20.**

Zurawski G. and Clegg M.T. (1993) ***rbc*L sequence data and phylogenetic reconstruction in seed plants: foreword.** Annals of the Missouri Botanical Garden **80:523-525.**

APPENDIX 1

The morphologically based matrix derived from the matrix presented in Linder (1991[a]).

Outgroup (Australian genera)	0000?00000000
Ischyrolepis	0000000000000
Elegia	1101100110101
Askidiosperma	1101100110000
Nevillea	1110011001111
Hypodiscus	1110011001111

The characters used are as follows:

1 Pollen annulus footlayer thickened / not thickened

2 pollen aperture without scattered fragments / with scattered exine fragments

4 Pollen grass-like / pollen type 2b

5 Pollen grass like / pollen type 2c

9 Central ground tissue with vascular bundles in a ring / scattered

11 Ribs alternating with the vascular bundles absent / present

14 Silica bodies present / absent

17 Leaf sheaths persistent / caducous

21 Anthers exserted at anthesis / included at anthesis

26 All carpels fertile / carpel 1 fertile, carpels 2 and 3 fused

28 Ovary cortex with scattered tannin idioblasts / without scattered tannin idioblasts

30 Seed coat tanniferous / not tanniferous

31 Fruit dehiscent / indehiscent

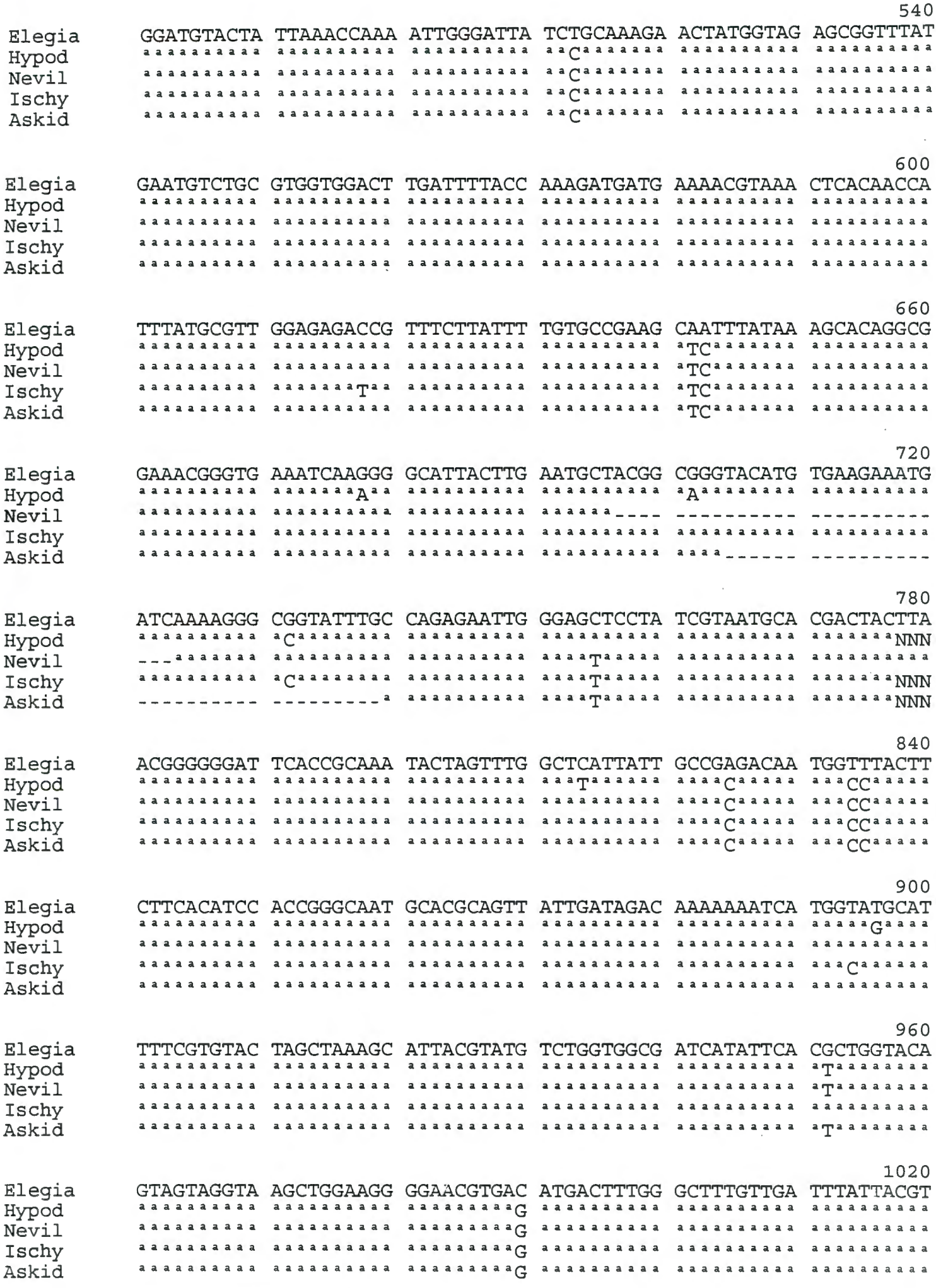
The first state is coded 0 and the second 1.

The numbers correspond to the characters listed in Linder (1991[a])

APPENDIX 2

The *rbc*L sequences of *Hypodiscus aristatus* (Hypod), *Nevillea sp.* (Nevil), *Ischyrolepis subverticilata* (Ischy) and *Askidiosperma sp* (Askid) as compared to the previously published *Elegia sp. rbc*L sequence.

```
                                                                                                    60
Elegia      GTTGGATTTA AAGCTGGTGT TAAAGATTAT AAATTGACTT ATTACACTCC TGATTACGAA
Hypod       aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaCaaaaaa
Nevil       aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaCaaaaaa
Ischy       aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
Askid       aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa

                                                                                                   120
Elegia      ACCAAAGATA CTGATATCTT GGCAGCATTC CGAGTAACTC CTCAACCAGG GGTTCCCCCT
Hypod       aaaaaaaaaa aaaaaaaaaa aaaaaaaaaT aaaaaaaaaa aaaaaaaTaa aaaaaaaGaa
Nevil       aaaaaaaaaa aaaaaaaaaa aaaaaaaaaT aaaaaaaaaa aaaaaaaTaa aaaaaaaGaa
Ischy       aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
Askid       aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa

                                                                                                   180
Elegia      GAGGAAGCGG GAGCTGCGGT AGCTGCCGAA TCTTCTACGG GTACTTGGAC AACGGTTTGG
Hypod       aaaaaaaaaa aCaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaAaaaaa aaaaaaaaaa
Nevil       aaaaaaaaaa aCaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaAaaaaa aaaaaaaaaa
Ischy       aaaaaaaaaa aTaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
Askid       aaaaaaaaaa aTaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa

                                                                                                   240
Elegia      ACTGATGGAC TTACCAGCCT TGATCGTTAC AAAGGACGAT GCTATCACAT CGAGCCCGTT
Hypod       aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
Nevil       aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
Ischy       aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
Askid       aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa

                                                                                                   300
Elegia      GTTGGGGAGG AAAATCAATA TATTGCTTAT GTAGCTTATC CTTTAGACCT TTTTGAAGAG
Hypod       aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
Nevil       aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
Ischy       aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
Askid       aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa

                                                                                                   360
Elegia      GGTTCTGTTA CTAACATGTT TACTTCCATT GTGGGTAATG TATTCGGTTT CAAAGCGCTA
Hypod       aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaTaaaaa aaaaaaaaaa
Nevil       aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaTaaaaa aaaaaaaaaa
Ischy       aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
Askid       aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa

                                                                                                   420
Elegia      CGAGCTCTAC GTCTGGAGGA TCTTCGCATC CCCCCTTCTT ATTCAAAAAC TTTCCAAGGC
Hypod       aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
Nevil       aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
Ischy       aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
Askid       aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa

                                                                                                   480
Elegia      CCACCCCATG GTATCCAAGT GGAAAGAGAT AAGCTGAACA AGTATGGTCG TCCCCTATTG
Hypod       aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
Nevil       aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
Ischy       aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
Askid       aaaaaaaaaa aAaaaaaaaa aaTGaaaaaa a---------- ---------- -aaaaaaaaa
```


540
Elegia GGATGTACTA TTAAACCAAA ATTGGGATTA TCTGCAAAGA ACTATGGTAG AGCGGTTTAT
Hypod
Nevil
Ischy
Askid
600
Elegia GAATGTCTGC GTGGTGGACT TGATTTTACC AAAGATGATG AAAACGTAAA CTCACAACCA
Hypod
Nevil
Ischy
Askid
660
Elegia TTTATGCGTT GGAGAGACCG TTTCTTATTT TGTGCCGAAG CAATTTATAA AGCACAGGCG
Hypod
Nevil
Ischy
Askid
720
Elegia GAAACGGGTG AAATCAAGGG GCATTACTTG AATGCTACGG CGGGTACATG TGAAGAAATG
Hypod
Nevil
Ischy
Askid
780
Elegia ATCAAAAGGG CGGTATTTGC CAGAGAATTG GGAGCTCCTA TCGTAATGCA CGACTACTTA
Hypod
Nevil
Ischy
Askid
840
Elegia ACGGGGGGAT TCACCGCAAA TACTAGTTTG GCTCATTATT GCCGAGACAA TGGTTTACTT
Hypod
Nevil
Ischy
Askid
900
Elegia CTTCACATCC ACCGGGCAAT GCACGCAGTT ATTGATAGAC AAAAAAATCA TGGTATGCAT
Hypod
Nevil
Ischy
Askid
960
Elegia TTTCGTGTAC TAGCTAAAGC ATTACGTATG TCTGGTGGCG ATCATATTCA CGCTGGTACA
Hypod
Nevil
Ischy
Askid
1020
Elegia GTAGTAGGTA AGCTGGAAGG GGAACGTGAC ATGACTTTGG GCTTTGTTGA TTTATTACGT
Hypod
Nevil
Ischy
Askid