Molecular phylogenetics of *Widdringtonia* Endl. (Cupressaceae) based on ITS1 sequences of nuclear ribosomal DNA.

Kristal Maze

Honours thesis, Department of Botany

University of Cape Town, 1993

Supervisors: Mr N.P. Barker & Prof H.P. Linder

Widdringtonia cedarbergensis near Sneeuberg Peak in the Cedarberg range.

ABSTRACT

Phylogenetic relationships within the genus *Widdringtonia* are poorly understood due to weak morphological and allozymatic characters. The suitability of the internal transcribed spacer (ITS1) region of the nuclear ribosomal DNA for phylogenetic studies in the genus *Widdringtonia* was examined. This region of the genome has been sequenced in four representatives of *Widdringtonia*, and has been shown to be unsuitable for phylogenetic studies in the genus. This is because the ITS1 region does not have sufficient synapomorphies on which a phylogeny can be based. Distance measures have confirmed the fact that the variability in the sequences is stochastic.

INTRODUCTION

Widdringtonia is a small genus of African conifers known as the "African Cypresses" of essentially montane distribution, ranging from the Cape Peninsula in the south, to Mount Mulanje of Malawi in the north (Fig. 1). Today four species are most commonly recognised; *Widdringtonia nodiflora* (L.) Powrie (the most widespread), *W. cedarbergensis* Marsh, *W. schwartzii* Marloth and *W. whytei* Rendle. The boundaries between these species have been, and still are much debated, this is probably due to the genus being poorly known, as well as having relatively few distinctive morphological characters on which to base a classification. Molecular phylogenetics has the potential to clarify evolutionary patterns and polarities that are difficult to establish using other characters (Clegg & Durbin, 1990).

In a number of Gymnosperms, particularly the conifers, phylogeny construction at generic level can be problematic due to the lack of obvious character variation between taxa. This appears to be the case in the genus *Widdringtonia*. The species boundaries are informationally weak in terms of morphological characteristics (Marsh, 1966). For instance Marsh (1966) splits the genus into two groups on the basis of the valve structure of the female cone and the shape of the adult leaves. One of the groups consists of *W.cedarbergensis* and *W.schwartzii*, which differ in their seed characteristics only. The other group consists of *W.nodiflora* sensus lato (in which she includes *W.whytei*). I strongly suspect that these species boundaries may have been inferred largely from their distributions (eg. *W. whytei* & *W. nodiflor*a on Mt. Mulanje, previously thought to have been one species, ie. *W. nodiflora* (Pauw, unpublished)). Pauw's separation of *W.whytei* is based on the ability of W. nodiflora to resprout after fire and other cone characteristics.

Population genetics studies have been conducted using allozymes for a number of *Widdringtonia* populations and were found to be unsuitable for species level studies in this genus (Janet Thomas, unpublished data, Pauw unpublished). Allozymes are isozyme characters which are variants of polypeptides representing different allelic alternatives of the same gene locus. The allozymes showed such a large degree of variation within each species that no unique locus or allele was consistently present in all of the individuals of one species. Thus the species could not be distinguished from each other.

The aim of this study is to construct a phylogeny of *Widdringtonia* based on molecular characteristics obtained from direct PCR sequencing of an internal transcribed spacer (ITS) region of the nuclear DNA. It is hypothesized that this region of the nuclear genome contains enough variation for phylogenetic studies at the interspecific level (Baldwin, 1993) however this is one of the first phylogenetic studies in higher plants that makes use of this region of the genome. It is thus uncertain at this stage how phylogenetically useful ITS sequences will be for all higher plants.

Regions of the 18-26S nrDNA have been used in the past for phylogenetic studies at both family and higher levels by DNA sequencing (Martin & Dowd, 1991), as well as among closely related genera or species by restriction site analysis (Jorgansen & Cluster, 1988). Only three published studies are known to have successfully used the ITS regions in phylogenetic reconstruction; protoctistans (Lee and Taylor, 1991) and apes and humans (Gonzalez *et al*, 1990), Calycadenia (Compositae) (Baldwin, 1993).

Not only is it important to systematics to resolve the phylogeny of *Widdringtonia*, but economically and conservationally the members of the genus are important. Trees of the

species *W. whytei*, commonly known as Mulanje cedars, are a valuable source of soft wood (Chapola, 1990). Seed collected from the cedars of Mt. Mulanje have been grown in silviculture around the world, as the wood of this species is regarded as one of the most useful of all coniferous timbers. In plantations problems are encountered in that many of the trees appear dwarfed. Descriptions of these dwarf trees fits Pauw's (unpublished) description of *W. nodiflora*. Pauw found that the *Widdringtonias* on Mnt Mulanje comprise two distinct forms, and proposed their separation into two species. He provided a key for distinguishing between these species at all stages in the life-history to enable seed collectors to differentiate between the plants and thus collect the appropriate seed.

At one time the small stands of *W. schwartzii* and *W. cedarbergensis* provided a valuable source of high quality softwood (Ward, 1958; Smith, 1955). The much depleted populations of *W. cedarbergensis* now face extinction due to the high fire-frequency management of the area in the recent past (Manders, unpublished). The wood is commonly referred to as cedar, *W. cedarbergensis* is known as the Clanwilliam cedar and *W. schwartzii* the Willowmore cedar.

Origins

It is important to bear in mind the antiquity of the Cupressaceae, which is inferred from its wide but frequently discontinuous range and numerous endemics. Palaeomagnetic and other evidence indicate that Africa finally separated from the remainder of Gondwanaland during the early Cretaceous period. Very little is known about the early distribution of *Widdringtonia*. The earliest authentic fossil material is of Tertiary age but from within its

present area of distribution (Kerfoot, 1975). From the available evidence, it is likely that *Widdringtonia* descended from a cool-climate ancestor which already existed in the Tertiary period and not as has been suggested, from warm-climate ancestors similar to the modern genus *Callitris*. The genus is most probably a representative of an ancient Austral flora. *Callitris* and *Actinostrobus* are normally considered to share common ancestry. However from rbcL sequence data it is apparent that *Disselma* (an Austral genus) is the closest relative of *Widdringtonia* (Gadek & Quinn, 1993).

Taxonomic Background

The genus *Widdringtonia* was first described as *Pachylepis*, by Brongniart in 1833. This name had been used the previous year for a genus of Compositae so in 1841 the name was changed to Parolinia by Endlicher, but this name had also been used before, by Webb for a genus of Cruciferae. Finally in 1842, Endlicher renamed the genus *Widdringtonia*, after Edward Widdrington. The genus is distinguished from other African Cupressaceae (*Tetraclinis* and *Callitris*) by having opposite and decussate leaves, and four cone-scales all of the same size.

The family Cupressaceae is sub-divided in various ways by different authors. The most popular view is that the Cupressaceae should be divided into at least two sub-families; the Callitroideae and the Cupressoideae. *Widdringtonia* falls into the subfamily Callitroideae (the southern hemisphere family) and the tribe Libocedreae (Li, 1953).

The first detailed account of the genus was compiled by Masters (1905) who also described

a new species, *W.mahoni* from the eastern highlands of Zimbabwe, making six *Widdringtonia* species (ie. *W. juniperoides* Endl., *W. schwartzii* Marloth., *W. cupressoides* Endl., *W. whytei* Rendle, *W. mahoni* Mast. and *W. equisetiformis* Mast.). Since this description, the genus has undergone numerous taxonomic changes. For instance *W.mahoni* was later included as *W. whytei* by Rendle (1911). Masters also describes the numerous other species names that have been mentioned in books for example *W. natalensis* (*W. cupressoides*) and *W. wallichii* (*W. juniperoides*) both are names that were proposed by Endlicher but have since change to the names in brackets.

In 1933 Stapf maintained six species *W. whytei*, *W. cupressoides*, *W. stipitata*, *W.dracomontana*, *W. schwartzii* and *W. juniperoides*. Stapf stated that all of these species were distinctive except for *W. stipitata*.

It appears that local geographical forms were in the past recognised as species, however the degree of variation appears such that there are no constant differences separating them as species. On the basis of this Marsh (1966), in the most recent revision of *Widdringtonia*, reduced the number of *Widdringtonia* species to three. *W. whytei* Rendle was included with *W.dracomontana* in *W. cupressoides* (L.) Endl. *W. cupressoides* was originally described from the Cape Mountains where it occurs with *W. cedarbergensis* and *W. schwartzii*, both of which have very restricted distribution ranges. Marsh (1966) could find no distinguishing character to separate the range of variants of *W. cupressoides*. She did however recognise geographical races.

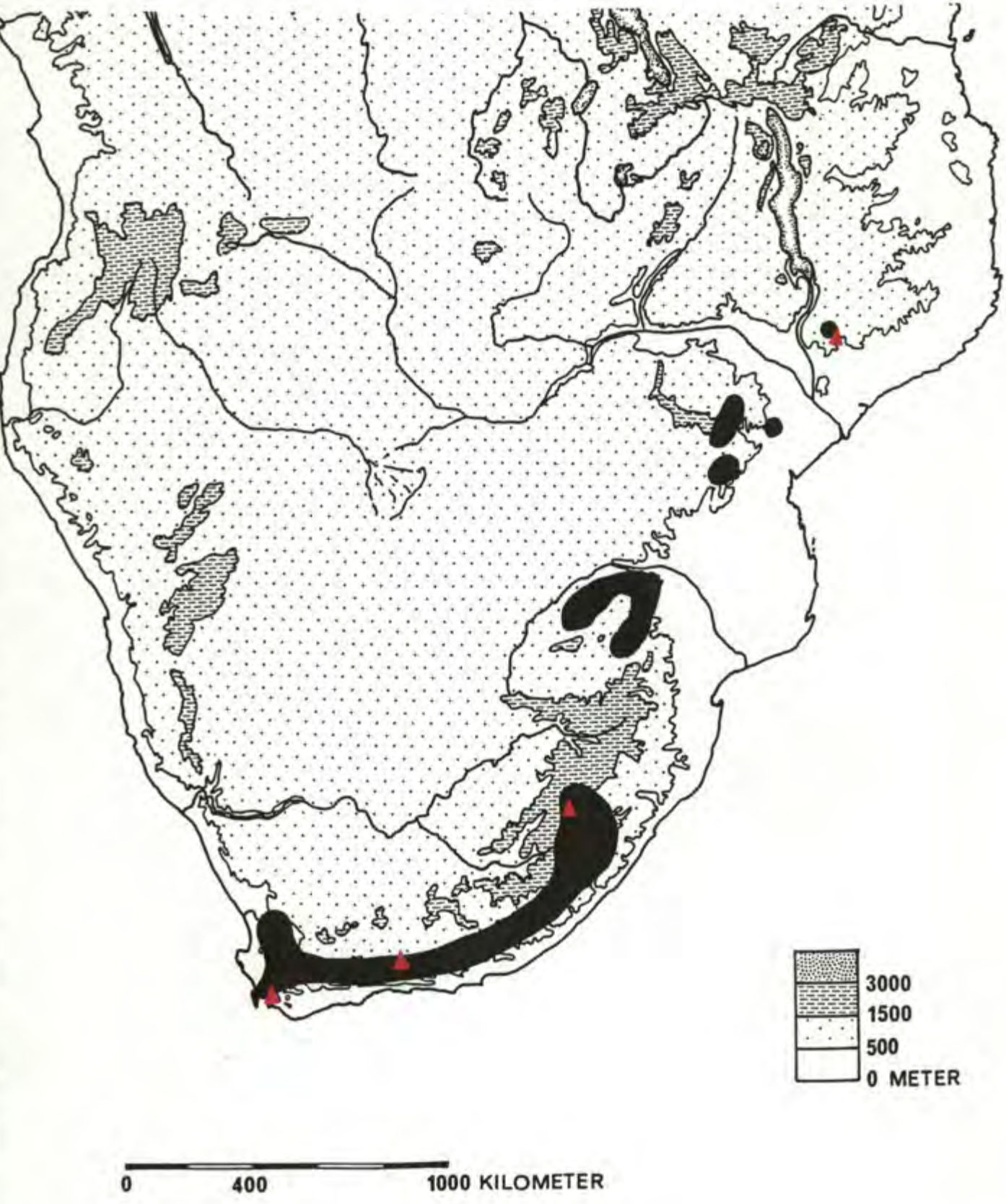

Figure 1. Map of southern Africa showing the distributions of the four *Widdringtonia* species and the sample areas(▲).

W. cupressoides was renamed as *W. nodiflora* by Powrie in 1972 when it was discovered that the type specimen of *Brunia nodiflora* L. (Bruniaceae) was a twig of *W. cupressoides*, but this did not involve a change in the species concept put forward by Marsh (1966).

The collapse of *Widdringtonia* into three species has generated much interest in the systematics of the group. This is mainly in conjunction with the wide ranging species, *W. nodiflora*. Many botanists have remarked in particular on the presence of two very different forms on Mt. Mulanje (Rendle, 1893). This is the Mulanje cedar *W. whytei*, and its dwarf or glaucous form as described by Chapman (1961). This dwarf form is multiple stemmed, 15ft high, often compared to *W. nodiflora* that has larger cones, the ability to vigorously coppice, and is restricted to rugged terrain. In this study *W. whytei* Rendle is recognised as a fourth species after Pauw's (unpublished) morphological and electrophoretic diagnosis/findings which strongly suggest that there are two species on Mt. Mulanje. Pauw's findings also fit the original description of *W. whytei* by Rendle (1893).

Considering the history of controversy over the classification of the *Widdringtonias*, it is not surprising that no attempt has been made to construct a phylogeny for the genus. The morphological and allozyme characters are simply too weak. Thomas (unpublished data) failed to find allozymatic autapomorphies for *W. cedarbergensis* and *W. schwartzii*, just as Pauw (unpublished) failed to find allozymatic differences between *W. whytei* and *W. nodiflora*. It has been proposed that allozymes are sensitive to genetic drift and thus differences within each species become obscured by the variation, rendering allozymes unsuitable for species level studies in this genus (Hamrick, 1987).

Why a molecular approach?

Although the use of DNA in plant systematics is relatively recent (ten years), it has assumed a major role in phylogeny reconstruction. This is particularly the case with taxa that have very few morphological characters on which phylogenetic relationships can be based. Here mutational differences at the level of DNA are useful (Clegg & Durbin, 1990). The genus *Widdringtonia*, typical of most of the conifers, is poorly understood phylogenetically. This is possibly due to the antiquity of the group as well as many of its members being missing today and from the fossil record. This renders phylogenetic reconstruction difficult in most conifers. As previously mentioned the situation is further compounded in that *Widdringtonia* species are based on weak, indistinct characters. From the taxonomic history above it is clear that systematists have long battled with the classification of the *Widdringtonias*. This is one of the reasons for this study, as characters at the sequence level may be more informative for phylogenetic inference in *Widdringtonia*.

Choosing a suitable sequence.

It is crucial to choose a sequence that is suitable for the taxonomic level under study. Sequences should display sufficient variation to enable phylogenetic analysis, but not so much that there is substantial homoplasy. Initially a non-coding region of chloroplast DNA was chosen for this species level study, as it is potentially more variable than other coding regions of the genome. It is assumed that non-coding chloroplast DNA evolves three times as fast as coding DNA (Soltis *et al.*, 1992). Because this non-coding region of the chloroplast DNA is not as highly conserved as coding areas of the chloroplast genome, it was initially expected

that this region was more suitable than sequences such as rbcL, for assessing phylogenetic relationships at the interspecific level. This region of the chloroplast DNA has been sequenced and no differences were found between two *of the Widdringtonia* species. The fact that there was no difference between the sequences of the two species at this level could mean that the species have diverged relatively recently, although the genus is relatively ancient.

The ITS1 region (internal transcribed spacer) of 18-26S nuclear ribosomal DNA was chosen for this study as it has been found to have a higher nucleotide sequence divergence relative to non-coding regions of cpDNA (ie. it is less conserved than the non-coding regions of the chloroplast genome) (Baldwin, 1993). Other regions of rDNA have been shown to be useful such as the 16S subunit that evolves relatively slowly and is useful for studying distantly related organisms. The internal transcribed spacer region and intergenic spacer of the nuclear rRNA repeat units evolve fastest and may vary among species within a genus or among the populations (White *et al.*, 1990).

The problem with using the nuclear genome is in identifying regions that are useful for phylogenetic comparisons with cpdna. The nuclear DNA regions must be (1) evolutionarily conservative- evolving primarily by point mutations at a level comparable to that in cpdna; (2) phylogenetically interpretable- evolving in a manner that allows resolution of speciation events; (3) easily examined in the laboratory; and (4) sufficiently large to offer enough potentially useful characters for phylogeny reconstruction (Baldwin, 1993).

The nuclear genes of higher eukaryotes which code for 5.8S RRNA and 17-18S small-subunit and 25-28S large-subunit occur as DNA sequences repeated (500 to more than 40 000 copies

per genome) in tandem (Takaiwa *et al.*, 1985; Schaal & Learn, 1988; Yokota *et al.*, 1989) (Figure 2). These repeating units are composed of a number of regions that vary in functional constraint and consequently in evolutionary rate rendering some regions suitable for phylogenetic inference at the species level (Schaal & Learn, 1988). The ITS1 region is one of these regions and according to Schaal & Learn (1988) it shows substantial length conservation among moderately closely related species. Each repeating unit consists of a transcribed region and a non-transcribed spacer. The internal transcribed spacers separate the 17S, 5.8S, and 25S RNA coding sequences.

The ITS nuclear gene region is particularly attractive for phylogeny reconstruction at the interspecific level for a number of reasons; the region has a high copy number (Rogers & Bendich, 1987), rapid concerted evolution (Zimmer *et al.*, 1980), and diverse rates of evolution within and among component subunits and spacers (Appels & Dvorak, 1982).

Gene differentiation and divergence is opposed by a homogenizing effect known as concerted evolution (Zimmer *et al.*, 1980) which involves various mechanisms such as unequal crossing over, leading to duplication and deletion of sequences, gene conversion, and reciprocal exchange (Sanderson & Doyle, 1992). Concerted evolution is considered to be useful for phylogeny reconstruction in producing almost completely homogenized families such as nuclear ribosomal loci. It has however been pointed out by Hillis and Davis (1988) that homogenization may not be complete in the spacer regions of different chromosomes.

The ITS1 region of 18-26S nuclear ribosomal DNA (Figure 2) from only 4 representatives of the genus *Widdringtonia* were successfully sequenced. Originally 3 more representatives and two outgroup species, *Disselma archeri* Hook and *Cupressus gluaca* were to be

sequenced to explore the utility of this region for phylogenetic reconstruction in *Widdringtonia*.

Primers

In order to sequence a region of any genome, primers are needed to recognise and anneal to the sites that flank the region to be sequenced. Primers are short lengths of DNA (in this case 20 & 25 bases) in which the sequences are known. White *et al.* (1990) in their fungal studies, have developed primers to allow sequencing of the entire ITS region. The ITS primers make use of conserved regions of the 18S, 5.8S, and 28S RRNA genes to amplify the noncoding regions between them (White *et al.*, 1990). The first successful study exploring the use of these primers for PCR amplification and sequencing of the ITS region in angiosperms and the value of these sequences as a source of phylogenetic data was conducted for the subtribe Madinae of Compositae (Baldwin, 1992). The results were highly concordant with a chloroplast based phylogeny. More information about the function of the primers is discussed in the detailed methods section (Appendix) under the polymerase chain reaction (PCR).

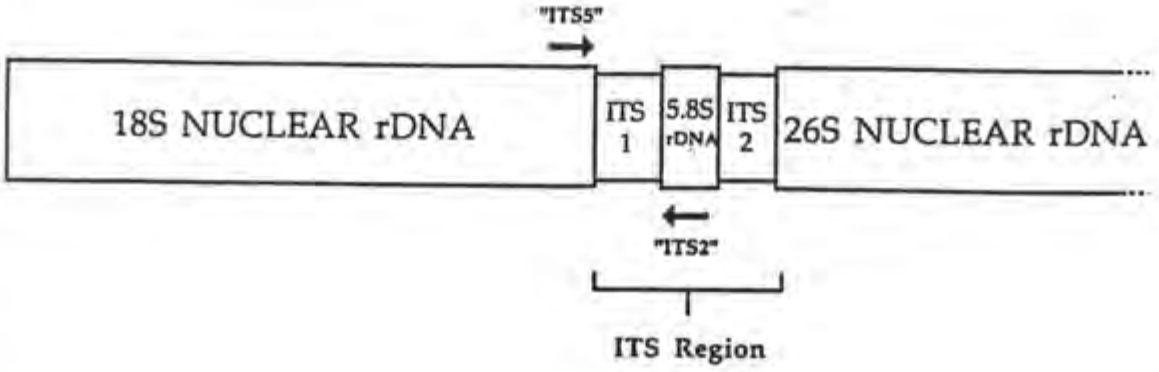

Figure 2. Repeat unit of 18-26S nuclear ribosomal DNA, minus the intergenic spacer and much of the 26S subunit. ITS Region =internal transcribed spacer region. Arrows indicate approximate positions of primers used to amplify single-stranded DNA for sequencing. Primers were designated and named by White *et al.* (1990). Primer sequences (5' to 3'): "ITS2" = GCTGCGTTCTTCATCGATGC; "ITS5" = GGAAGTAAAAGTCGTAACAAGG. (Modified from Baldwin, 1993).

METHODS AND MATERIALS

Plants used and their locations

Total nuclear DNA was isolated from one seedling of each "species" except for W. nodiflora, where one seedling from each of three distant populations was sampled (Table 1). The reason for sampling three distantly separated populations of *W. nodiflora*, is that they have a much larger distributional range than the other species that are very much restricted, and thus may show a considerable degree of intraspecific genetic variation. Figure 1 shows distributions of the four species as well as the rough locations of the populations sampled.

Table 1. *Widdringtonia* collections examined for RDNA ITS1 variation.

Species	Locality
Widdringtonia nodiflora	1. Koggelberg
	2. Cathedral Peak
	3. Tuchila (Mt. Mulanje)
W. cedarbergensis	Cedarberg
W. schwartzii	1. Baviaans Kloof
	2. Sandvlakte
W. whytei	Lichenya (Mt. Mulanje)

Outgroup choice

Outgroups are used to determine which character groups shared within the ingroup are

pleisiomorphic relative to other states (Wheeler, 1990). Topologies are most robust when the most exclusive sister group is used to root the topology. From a phylogeny of the relationships within the Cupressaceae, based on rbcL sequences (Gadek & Quinn, 1993), the outgroup to *Widdringtonia* was assumed to be *Disselma archeri*.

Disselma is a monotypic genus and consists of shrubs or small trees with scale like decussate foliage, small globose cones with two pairs of scales. The single species *Disselma archeri* is an alpine shrub that grows in cool moist sites at the treeline in Tasmania (De Laubefels, 1965). The second outgroup, *Cupressus glauca* was collected on the U.C.T. campus. *Cupressus* is thought to be one of the more primitive members of the Cupressaceae (De Laubenfals, 1965).

Outline of sequencing strategy.

The seeds collected from individuals from these areas were germinated in germination chambers in Benlate solution. Double-stranded DNA was extracted from the seedlings once the cotyledon leaves had grown to 1cm (approximately 8 weeks old) from which double stranded DNAs from the ITS1 region of each sample were generated by polymerase chain reaction (PCR) using the primers ITS2 and ITS 5 (White *et al.*, 1990) to amplify the ITS1 region (5'5.8S-3'18S). PCR reactions, PCR product purification, and direct dideoxy sequencing of resultant single-stranded DNAs were conducted by the methods detailed in Appendix 1. The sequence data was analyzed using both parsimony (Wagner) and distance (UPGMA and Neighbour-joining) methods.

RESULTS

Sequences using the ITS2 primer were obtained for *W. nodiflora* for the Koggelberg population, *W. schwartzii* and *W. cedarbergensis*. Two sequences were available for *W. schwartzii* for different populations. Sequences using the ITS5 primer were obtained for *W. schwartzii* and *W. cedarbergensis* only. These sequences were manually aligned by sequential pairwise comparisons using the DAPSA programme (E. Harley). The aligned sequences are presented in Table 2.

Aligning the sequences involves inserting gaps in some positions in the sequences which correspond to insertions or deletions in other sequences, so that the positions inferred to be homologous are in the same column of the data matrix. Mismatches correspond to substitution events. The alignment is done by choosing one sequence arbitrarily as a "reference" and one by one the other sequences are aligned to the reference sequence. If one of the sequences causes a gap to be inserted into the reference sequence, the other sequences also change correspondingly. It is easiest to choose a reference sequence from the pair of sequences that are most similar.

Because only two sequencing loadings were carried out for each representative, the entire ITS1 region was not sequenced and thus the length variation cannot be determined. However from the PCR checking gels, it appears that the ITS1 sequence lengths are invariant between the *Widdringtonia* species. Over 400 base pairs have been counted for the ITS1 region from these sequences. This number of base pairs is much greater than has been found in other species, for example in the genus *Calycadenia*, the total lenghth of the ITS1 region ranged from 254 to 257 base pairs (Baldwin, 1993).

Table 2. Aligned DNA sequences of the ITS1 region in 18-26S nuclear ribosomal DNA from four representatives of *Widdringtonia*. A = *W.schwartzii*(1) B = *W.shcwartzii* (2) C = *W.cedarbergensis* D = *W.nodiflora*; - =gaps, * = homologous positions, ⇓ = the only phylogenetically informative character.

ITS2

```
                                                                ⇓
A TATTTGATTA GCAATTTTTA -TTGTATTTT C--G----CG AGGC-GAACA
B ---------- ---------- ------TTTT C--G--G-CG AGGC-GAACA
C --TTTGATT- GCAATTTTTA CTGGTAMTCT TT-GTCGGAG AGGC-GGACG
D --TTT--TT- GCAATTTTTA -TTGTCTTTT C--G--G-CG AGGC-GGACA
                        * *    *     * **** * **

A CGCAAGGTGC TCG-AA--G- -CC-AAAGTG TTTGGATTTC -----GTAGG
B CGCAAGGTGC TCGCAAACG- -CC-AAAGTG TTTGGATTTC A-----AAGG
C CGCAAGGTGC TCGCAAACA- --CCAAAGCT ATTGTATTTG AGGGCAAG--
D CGCAAGGTGC TCGCAAACG- -CC-AAAGTG TTTGGATTTC A----GAAGG
  ********** *** **       * ****    *** ****

A C-CNGTGGCA -CCTA-
B C-C--TTGCA -CCTAT
C -----TTGCA -CCTAT
D GCG-GTTGCA -CCTAT
      * ***  ****
```

ITS 5

```
B AGGTATGAGG GA--TGCGCT ACGC------ -AA-CTT-CA AAC---TCTC
C ATGCATTTN- -AGCTGCGCT GCGCGCGCTN NAAGCTTACA AAAAAATATA
  * * **      *  ******  ***        ** *** ** **    * *

B ----CGCCCT GC--GC---A AGACGACCTC CGCTCGAACG SGCTAA-CAG
C AATRCGCA-T GCAAGCTTVA A---G--CT- --CTCGAATG -GCCAAACAG
     ***  * **  **   * *   *  **    ****** *  ** ** ***

B CTGCAACATC
C CTGCAAC-TC
  ******* **
```

Sequence analysis: parsimony and distance.

Parsimony

Phylogenetic inference is based on the inheritance of ancestral characters and on the existence of an evolutionary history defined by these characters. The Wagner parsimony method (PAUP, Swofford, 1989) was used to analyze the sequence data, which infers phylogenies directly from the character data, based on the principle of maximum parsimony. Parsimony methods for inferring phylogenies operate by selecting trees that minimize the total tree length (ie. the number of evolutionary steps (character state changes) needed to explain a given data set). By minimizing the total number of steps also minimizes the amount of homoplasy needed to explain the data. Wagner parsimony imposes minimal constraints on permissible character state changes and assumes that characters are measured on an interval scale. Wagner parsimony also assumes that any transformation from one character state to another also implies a transformation through any intervening states, as defined by the ordering relationship (ie. additive). The exhaustive option of PAUP was used to search for the most optimal tree by evaluating every possible tree.

There are two assumptions for using characters data. Firstly there is an assumption of independence among characters, and secondly that the characters are homologous. In other words characters must be defined so that all states observed over taxa for that character must have been derived, maybe with modification, from a corresponding state observed in a common ancestor of those taxa.

Sequence data is treated as unordered multistate characters. The characters are the positions on the sequences and the character states are the actual nucleotides at those positions. However the situation is complicated in that not only must homologous molecules be used but positional homology is required for the phylogenetic analysis of the sequence data. In other words the nucleotides found at one position in the taxa being studied should all trace their ancestry to a single position that was present in a common ancestor of those taxa. Insertions and deletions lead to the assumption that nucleotides found in the same positions in the sequences of the taxa being studied are homologs.

The unrooted tree (one in which the common ancestor is not identified) produced from this analysis is shown in Figure 3 but is based on only one informative site. *W.cedarbergensis* and *W.nodiflora* are grouped by this informative site. The remaining sites are either constant or autapomorphic.

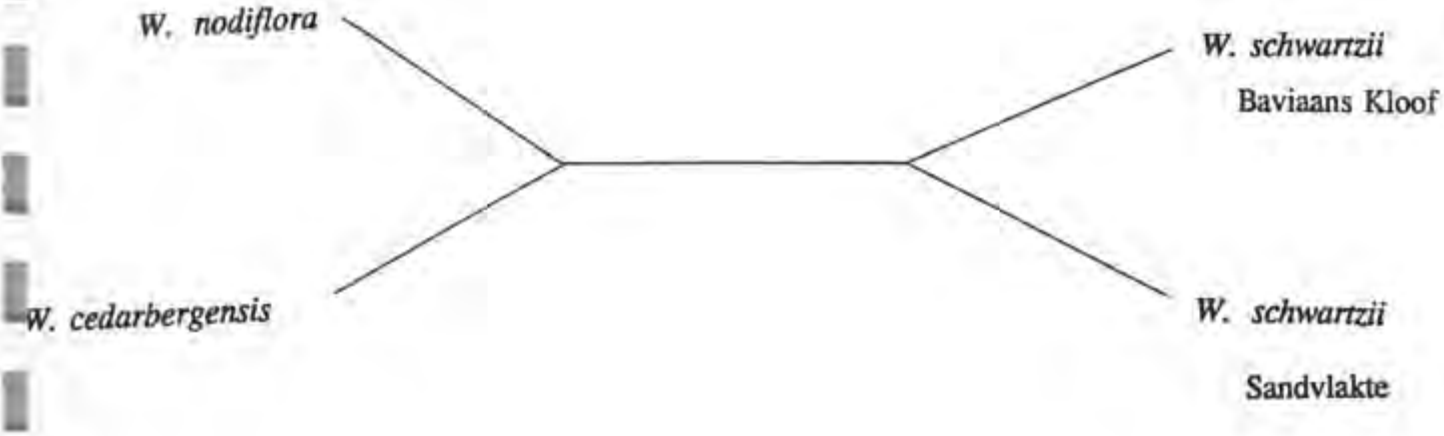

Figure 3. Phylogenetic unrooted tree generated from Wagner parsimony for four representatives of the genus *Widdringtonia*.

Distance

Distance data differs from character data in that distance data specify a relationship between pairs of taxa or sequences (Sneath & Sokal, 1973). Distance measures of zero indicate that objects are identical, values above zero separate increasingly different objects. To use sequence data for distance analysis, the sequences have to be transformed to distance values relating all pairs of sequences. An grouping analysis is then carried out on the pairwise distances. However there is a disadvantages to using pairwise distance measures for sequence data in that information is lost in the transformation. Penny (1982) showed that a number of different sequences result in the same distance matrix, but from the distances only it is impossible to go back to the original sequences.

Little is known about how the apparent information loss affects the accuracy of the subsequent phylogenetic analyses. Some sequence data sets result in identical branching patterns with character-based and distance-based analyses (Olsen, 1987). Distance does not allow the incorporation of different kinds of data as does character-based analyses. Only character-based analyses can enable the identification of informative characters in order to limit future studies to those characters (Swofford & Olsen, 1990).

Other disadvantages of this simple cluster analysis are that it is an algorithm with no objective definition of what constitutes an optimal tree when the data are not ideal, specifically because genes do not diverge uniformly in all organisms. Cluster analysis is sensitive to highly divergent sequences. Another problem with distance methods is that if there is missing data in one of the sequences, it can force one or more members out of the group in the inferred tree dependent on random noise. Of the 116 nucleotides sequenced 26

are missing at the start of the sequence for the Baviaans Kloof W. schwartzii sequence. The first 26 nucleotides were therefore excluded from the analysis. However there were still sections in some of the sequences where it was too faint to read.

The UPGMA (unwieghted pair group method using arithmetic averages) cluster analysis programme was used to produce a tree based on distance data. There is one assumption in this method and that is that the data is approximately ultrametric (ie. the pairwise distance values precisely fit a rooted tree with a constant molecular clock). The tree is constructed by linking the least distant pairs of taxa, followed by successively more distant taxa or groups of taxa. As the clustering process proceeds, the clusters of taxa are merged until the last two clusters are merged into a single cluster containing all of the original taxa. The tree in Fig 4 was produced showing that the *W.cedarbergensis* sequence is most similar to the Sandvlakte *W. schwartzii* sequence, then the Koggelberg *W. nodiflora*, and the most divergent sequence is from *W.schwartzii* from the Baviaans Kloof sequence.

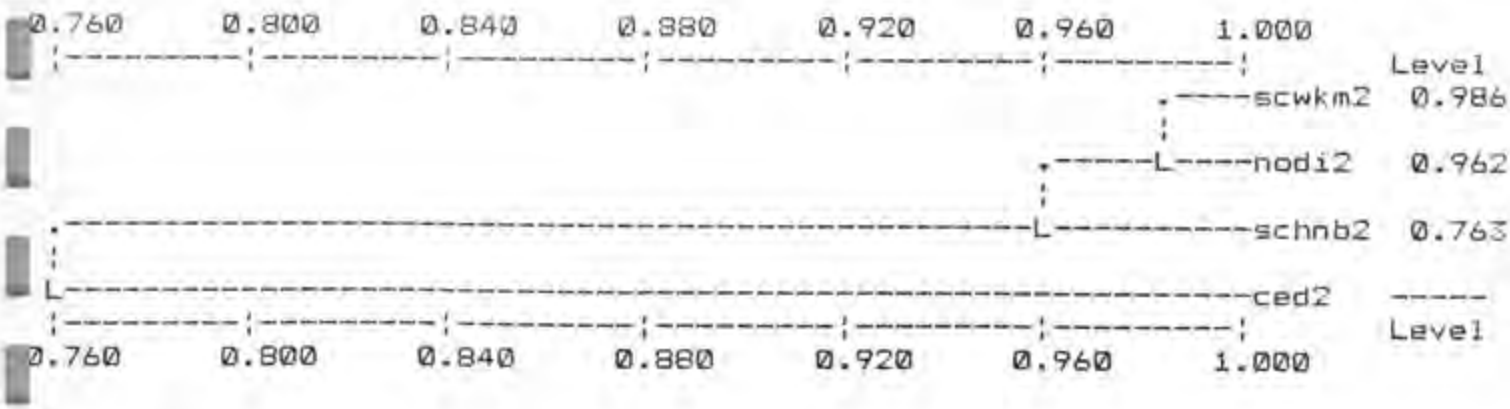

Figure 4. Tree obtained form the UPGMA cluster analysis based on sequence data for four representatives of the genus *Widdringtonia*.

Similarity

Another method for illustrating the relationship between two sequences is by their percentage similarity. This is equal to the number of aligned sequence positions that have identical nucleotides divided by the number of sequence positions compared. The neighbour-joining method (Satou & Nei, 1987 in Swofford & Olsen) was used to analyze the sequence data. This is an additive tree technique that rests on the assumption that the lengths of the branches lying on the path between any pair of taxa can be summed to yield a meaningful quantity (ie. amount of evolution). It is conceptually related to cluster analysis, but it removes the assumption that the data are ultrametric, that is that it does not assume that all lineages have diverged equal amounts. Instead it assumes that the data come close to fitting an additive tree, so correction for superimposed substitutions is important for data that might include lineage-to-lineage differences in average rate. This method differs to cluster analysis in that it keeps track of nodes on the tree rather than taxa or clusters of taxa.

The distances between each of the sequences for each pairwise comparison are calculated and are then modified so that the separation between each pair of nodes is adjusted on the basis of their average divergence from all other nodes (ie. this normalises the divergence of each sequence for its average clock rate). The tree is then constructed by linking the least distant (most similar) pair of nodes as defined by the modified distances. When two nodes link, the common ancestor is then added to the tree and the terminal nodes with their respective branches are removed from the tree. At each step two terminal nodes are replaced by one new node until two nodes remain, separated by a single branch. The tree in figure 5 was obtained by the neighbour-joining analysis and shows that *W. cedarbergensis* and the Sandvlakte *W.shcwartzii* are most similar and share a common ancestral node. The other

W.schwartzii sequence is the most evolutionarily different to the Sandvlakte *W. schwartzii*.

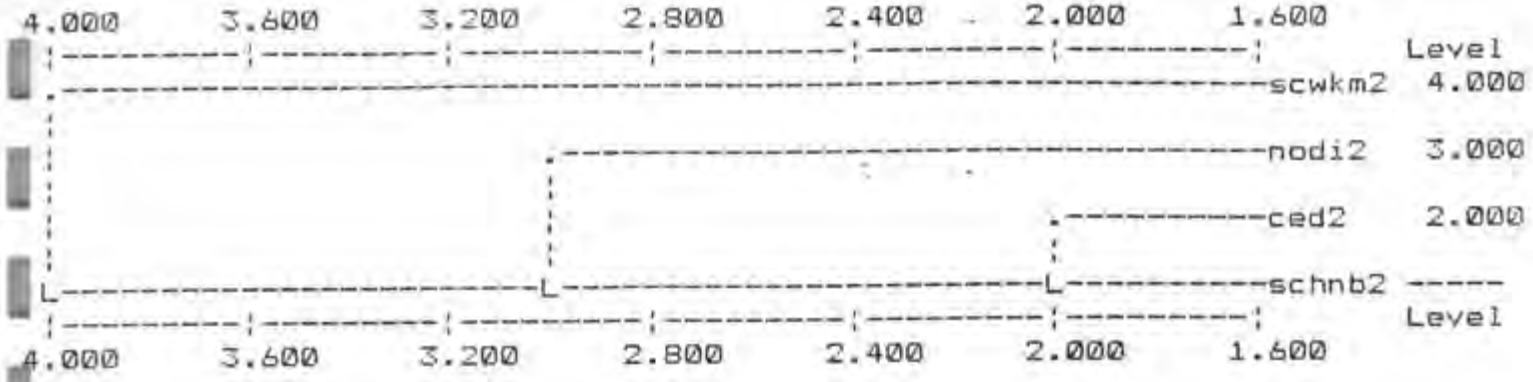

Figure 5. Tree obtained from the Neighbour-joining analysis based on sequence data for four representatives of the genus *Widdringtonia*.

The pairwise comparison of the ITS 5 sequences from *W. cedarbergensis* and *W.schwartzii* shows that they have a 45% divergence. The sequence divergence of the ITS2 sequences of the same sample is 40%. This small difference is probably due to the difficulty in reading faint (sequence B) and compressed (sequence C) areas of the sequences.

DISCUSSION

Unfortunately the fact that there was only one informative character in the Wagner parsimony analysis means that a DNA sequence-based phylogenetic tree cannot be constructed on a single informative character simply because molecular character states, unlike good morphological characters, can easily change to one of only three other character states.

From the cluster analysis and the neighbour-joining analysis, it is clear that the two sequences of *W.schwartzii* are more dissimilar to each other than to either of the other two species. If two members of a species (that is one of the two most morphologically distinct species) shows higher sequence divergence than between species then there are two possible explanations; either one of the W. schwartzii sequences has a mistaken identity, or the ITS1 region is too variable to be used for an interspecific study in this genus. The latter is probably the case judging from the variability in the other sequences.

Therefore as was the case with allozymes, sequences of the ITS1 region of the nuclear genome appear to be too variable for phylogenetic use at species level in this genus, as evidenced by the single informative character. The sequence variation, larger size of the ITS1 region in this genus compared with other taxa, and bands across more than one lane in the sequences, all suggest that concerted evolution may not be occurring in the ITS1 region in *Widdringtonia*. This would make the ITS1 region interesting from a gene evolution perspective.

Methodological failings

The problems encountered at various stages will be discussed briefly.

Problems with DNA amplification - DNA amplification was low *Disselma archeri* (the proposed sister taxa) and *Cupressus*. It should be borne in mind that the primers used in this study were developed for fungus and that this might be an area where problems could arise.

The chance of a primer annealing to sequences other than the chosen target is very low. This is because there is a 1 in 4 chance (4^{-1}) of finding an A, G, T or C in any given DNA sequence, similarly there is a 1 in 16 chance (4^{-2}) of finding any dinucleotide sequence (eg. GC), there is a 1 in 256 chance for finding a 4-base sequence, therefore in this study where the primers are 20 (ITS2) and 22 (ITS5) bases long, there is a 1 in 1 099 500 000 (over 1000 billion) and 1 in 17 592 000 (4^{-22}) chance respectively of finding the correct sequence. The fact that these primers are so long means that even with high annealing temperatures, mismatching and non-specific priming cannot be ruled out (Rybicki, unpublished).

One of the possible problems, considering that the primers used were originally developed for fungi, is that the primers may not be complimentary to the corresponding region on the *Disselma* genome. The reason for low yield then could be that mismatches could have occurred, resulting in low DNA production especially at low annealing temperature. To determine whether or not the primers matched the *Disselma* genome or not, it is proposed that a PCR is run using a different enzyme called VENT. Unlike a taq polymerase this enzyme not only causes polymerization in the 5' - 3' direction, but reverses every so often

to check the sequence is correct. If on reversal it finds a mismatch, this area will be cut out. It is expected that the 3' end of the primer is the region that is not complementary to the *Disselma* genome and is thus limiting polymerization in the 5' - 3' direction. This non-complementary region would be cut out by the VENT enzyme so that polymerization can continue in the 5' - 3' direction.

Too many secondary compounds in DNA mixture also result in poor amplification and/or purification. For instance DNA from *W.whytei* and *W.nodiflora* from Mulanje, were initially extracted from six month old seedlings which contained more secondary compounds than the younger seedlings with cotyledon leaves only.

Other problems encountered at the PCR stage were that a second band of DNA was produced in *Disselma* (increasing the annealing temperature can help to get rid of double bands of DNA), and in *Cupressus* and *W.whytei* the DNA yield was low (the number of cycles was increased to 40 to increase the yield).

Problems at the QIAEX stage - DNA can be lost if precise measurements are not made for the isopropanol precipitation.

Sequencing problems - Many of the sequencing attempts were unsuccessful, resulting in the following:

- Bands appearing across all four lanes. This could be because the reagents were not mixed thoroughly during the reactions; the annealing step was run too long or too hot; labelling temperature was too hot or long resulting in too many pause sites in the first 100 bases from the primer; termination step too cool or too long.

- Dirty, uninterpretable sequences resulting from impure DNA (ie. excess primers etc.).

- Concerted evolution is supposed to keep each tandem repeat homologous, but there has been a recent report of ITS areas varying between the repeats (Hillis & Davis). This probably caused bands across more than one lane in some of the sequences (eg. *W.cedarbergensis*).

CONCLUSION

The sequences the ITS1 region of the nuclear genome in *Widdringtonia* do not possess sufficient informative characters (synapomorphies) to be useful for species level studies in Widdringtonia. The bulk of the characters in the four sequences are either constant or autapomorphic. However this region shows potential for studies of the evolution in the gene. In particular it provides an opportunity for testing the concerted evolution hypothesis. There is evidence in this study to suggest that concerted evolution may not be occurring.

The sequence divergence of the two *W.schwartzii* sequences does not mean that they are distantly related as the distance measure suggests. The distance analysis makes use of autapomorphies which are phylogenetically uninformative. These differences are can thus be attributed to stochastic variation, not phylogenetic relatedness.

The phylogenetic relationships in the genus *Widdringtonia* remain unresolved at this stage. The ITS1 sequence for phylogenetic reconstruction in *Widdringtonia* is not appropriate, based on the single informative character found in a sequence of 116 base pairs. At this stage no other region of the genome can be proposed for molecular studies at the species level in *Widdringtonia*, however molecular systematics is still a relatively new field and the bulk of the nuclear genome is unexplored.

APPENDIX 1

DETAILS OF METHODS SECTION

DNA Extraction

The CTAB method (Doyle & Doyle, 1987) was used to isolate total genomic DNAs (nuclear, chloroplast and mitochondrial) from the fresh 8 week old seedling tissue.

Procedure:

1. Preheat 5 ml of CTAB isolation buffer (2% CTAB, 1.4 M NaCl, 0.2% 2-mercaptoethanol, 20 mM EDTA, 100 Mm Tris-HCl, pH 8.0) in a glass centrifuge tube to 60°C in a water bath.

2. Powder 0.1-1 g of the fresh tissue (the entire seedling; cotyledon leaves and radical was used) in liquid nitrogen in a chilled mortar and pestle. Scrape the powder into the preheated buffer and swirl gently to mix.

3. Incubate the sample at 60°C for 30 minutes with occasional swirling. This breaks down the cell walls and dissolves the DNA from its tightly coiled form in the nucleus.

4. Extract once with chloroform-isoamyl alcohol (24:1), mixing gently but thoroughly.

5. Spin in a clinical centrifuge at room temperature to concentrate phases.

6. Remove the aqueous phase (consisting of the DNA and the salt) with a pipet, transfer to clean glass centrifuge tube, add 2/3 volumes cold isopropanol, and mix gently to precipitate the nucleic acids.

7. Spin in clinical centrifuge for 1-2 minutes. The DNA should have precipitated as a diffuse and very loose pellet. Gently pour off as much of the supernatant as possible without losing the precipitate. Add 10-20 ml of wash buffer (76% EtOH, 10 Mm ammonium acetate) to pellet and swirl gently to resuspend nucleic acids.

8. Spin down nucleic acids for 10 minutes after a minimum of 20 minutes of washing. The wash step is a convenient stopping point.

9. Pour off supernatant carefully (some pellets are still loose even after this longer spin) and allow to air dry briefly at room temperature.

10. Resuspend nucleic acid pellet in 3 ml TE (10 Mm Tris_Hcl, 1mM EDTA, Ph 7.4) + 1ml 10M ammonium acetate mix and 10 ml EtOH (cold).

11. Spin down DNA at high speed (10 000 * g for 10 minutes in refrigerated centrifuge for 10mins).

12. Final wash in 70% ethanol. Spin for 10 minutes.

13. Air dry sample and resuspend in 100 μl TE.

Double stranded polymerase reaction (PCR)

The PCR is a rapid procedure for in vitro enzymatic amplification of a specific segment of DNA. In this study PCR is used to amplify the DNA of the ITS1 region of the internal transcribed spacer region of 18-26S nuclear ribosomal DNA (Figure 2).

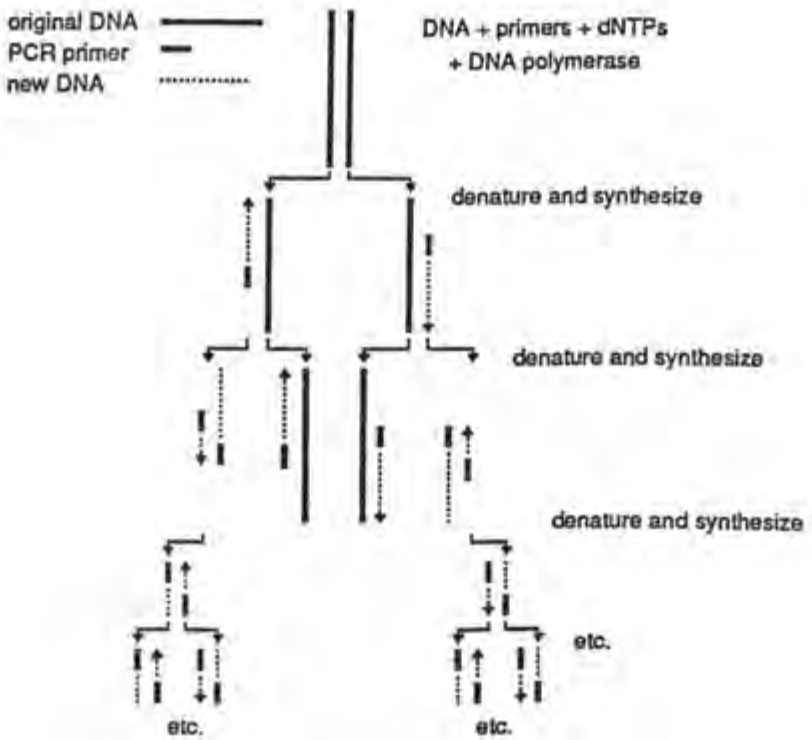

Figure 6. The polymerase chain reaction. DNA to be amplified is denatured by heating the sample. In the presence of DNA polymerase and excess deoxynucleotide triphosphates, oligonucleotides that hybridize specifically to the target sequence can prime new DNA synthesis. The first cycle is characterised by a product of indeterminate length; however, the second cycle produces the discrete "short product" which accumulates exponentially with each successive round of amplification. This can lead to the many million-fold amplification of the discrete fragment over the course of 20 - 30 cycles. (From Current Protocols, 1991.)

The theory behind PCR is illustrated in Fig 6. Present in the solution to be amplified there are three nucleic acid components: the double-stranded DNA (template) to be amplified, to which two single-stranded oligonucleotide primers anneal. Also present in the reaction are the enzyme (a DNA polymerase), deoxyribonucleoside triphosphates (dNTPs), a buffer, and salts ($MgCl_2$) (see table 1).

The primers in the reaction greatly exceed the amount of DNA to be amplified. They hybridise to opposite strands of the DNA with their 3' ends facing each other so that the synthesis by DNA polymerase (which catalyses strand growth in the 5'→3' direction) extends across the segment of DNA between them. After this first cycle, new strands of intermediate length are produced which, like the parental strands, can hybridise to the primers upon denaturation and annealing. These products accumulate only arithmetically with each subsequent cycle of denaturation, annealing to the primers, and synthesis.

The second cycle of denaturation, annealing, and synthesis produces two single-stranded products that together form a discrete double-stranded product which is exactly the length between the primer ends. The amount of this product doubles with every subsequent cycle of synthesis, denaturation, and annealing, accumulating exponentially so that 30 cycles should result in a 2^{28} -fold (270- million fold) amplification of the discrete product.

The following are required for incorporation into each of the PCR reactions:

Sterile distilled water	DNTPS at 100 μM each
10X reaction buffer	$MgCl_2$ (1 to 7mM)
Primer 1 at 100pmol	Enzyme (taq) at 2 units
Primer 2 at 100pmol	Template

Table 3. The following volumes (in μl) of each component are required for a range of magnesium concentrations of 0-4. Higher magnesium concentrations may be needed in some instances for optimal PCR.

Mg conc	H_2O	10X	Primer 1	Primer 2	dNTP	$MgCl_2$	Temp-late	Enz
0	81	10	2	2	3	0	2	0.5
1	77	10	2	2	3	4	2	0.5
2	73	10	2	2	3	8	2	0.5
3	69	10	2	2	3	12	2	0.5
4	65	10	2	2	3	16	2	0.5

Once the above reagents have been pipetted into each of the 5 reaction tubes, mixed and spun down. Reaction mixtures were sealed with 200 μl mineral oil to prevent evaporation during the PCR cycles. The reaction is then loaded into the PCR machine and the following temperature cycles mentioned below are conducted (these parameters can be varied if problems are experienced for instance in this experiment a higher annealing temperature (45 dec C) was used for *Disselma* to get rid of a second band of DNA, and 10 extra cycles were run for *W. cedarbergensis* and *W. schwartzii* to increase the DNA yields).

The PCR cycles were: 15 seconds at 96 °C to denature the double stranded template DNA, 15 seconds at 40 °C to anneal the primers to single-stranded template DNA, and 2 minutes at 72 °C to extend the primers.

The reason for running the PCR at various Mg concentrations is to determine the optimal concentration for DNA production. When the cycles have been completed the amplification products are checked on a 1% agarose gel. 20 μl of the reaction is loaded into the gel and run at 150V for approximately 2 hours. The products are then stained with ethidium bromide for 5mins and visualised with U-V light. The optimal concentrations for DNA production should be apparent as bright bands (Figure 7). Two or three of the $MgCl_2$ concentrations should give efficient synthesis of what should be a single predominant product.

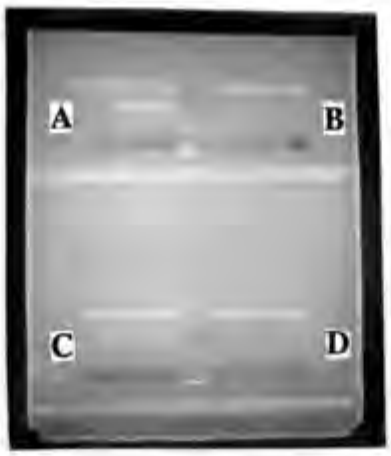

Figure 7. A photograph of an ethidium stained agarose gel under UV light showing the magnesium concentrations at which DNA was amplified. A - *Cupressus*; B - *W. schwartzii*; C - *W. cedarbergensis*; D - *Disselma archeri*. From this knowledge DNA can then be bulked up for sequencing at the optimal magnesium concentration. For W. scwartzii amplified well at a magnesium concentration of 8 μl whereas Disselma did not amplify well a any of the magnesium concentration.

Bulking up.

DNA is bulked up for sequencing at the optimal conditions determined from the PCR checking gel (Plate 5.). 10 reaction tubes each containing 100 μl are prepared for this bulking up step at the appropriate magnesium concentration. After the PCR reactions the DNA from these 10 reaction tubes is pooled and concentrated to a volume of 20 to 30 μl by an isopropanol precipitation. The total volume (V) of the pooled PCR reactions is measured. Half of this volume (1/2 V) of 8M ammonium acetate and 1.5 times the volume of isopropanol are added to the pooled PCR reactions. This is then spun in a clinical centrifuge for 10 minutes, the supernatant is then poured off and then washed in 70% alcohol and centrifuged for 5 minutes and the supernatant poured off. The DNA pellet is then dried in a centrifuge and vacuum for 5 minutes. 20 μl of TE is added to resuspend the DNA.

The concentrated product (with 2 μl of the xylene cyanol marker added) is then run on a 1% agarose gel, at 100 V for 3-5 hours (the length of time needed for the xylene cyanol marker to run 2/3 of the gel depends on the agarose content of the gel). The gel is then ethidium bromide stained and visualised with a UV lamp. The amplified fragment is quickly cut out of the gel (to avoid radiation damage of the DNA under the UV light) with a sharp clean scalpel and transferred to a 1.5 ml reaction tube.

PCR Product purification - QIAEX

Before sequencing, the PCR fragment must be purified from the contaminating primers and PCR reaction reagents. The QIAEX purification method was used.

The QIAEX purification method is designed to extract and purify DNA from tris-acetate-EDTA agarose gels or aqueous solutions. DNA dissolved from the agarose gel is bound to silicagel particles in the presence of high salt. These are spherical 3μm particles with an activated surface. The salt solution causes the modification of the structure of water thereby forcing the nucleic acids in solution to adsorb to the surface of the silicagel particles. The DNA can then be washed to remove all non-nucleic acid impurities such as agarose, proteins, salts, and ethidium bromide. The elution of the DNA from the silicagel is accomplished in low salt solutions such as in TE buffer at Ph 8.0.

Main aims of this step:

1. to separate excess primers, DNTP'S and template from the dsPCR product
2. to allow the concentration of the double stranded PCR product into a volume suitable for sequencing.

Procedure

1. To the each gel slice (approximately 300mg) add 900μl solubilization buffer QX1 and 15 μl previously vortexed QIAEX suspension. Incubate at 50 °C for 10 mins. This is to solubilize the agarose. The sample should be mixed briefly every 2 minutes by flicking the tube to keep the QIAEX in suspension and to avoid shearing the DNA. The high

concentration of sodium perchlorate in buffer QX1 disrupts hydrogen bonding between sugars in the afarose polymer, allowing the solubilization of the agarose gel slice. At the same time the high salt concentration disassociates DNA binding proteins from the DNA ligand. Incubating the DNA solution in a strong electrolyte with large anions (eg.ClO_4^-) causes a modification in the structure of water, forcing the DNA to adsorb to silicagel particles.

2. Centrifuge the sample, which now contains the DNA bound to the QIAEX, for 30 seconds in a microcentrifuge. Remove the supernatant with a pipette.

3. Add 500 μl of buffer QX2 and resuspend the pellet carefully by flicking the tube. Centrifuge for 30 seconds and discard the supernatant. Repeat this wash step. These two high salt washes with buffer QX2 remove agarose and contaminants from the DNA solution. After each step all traces of the supernatant should be removed with a pipette to eliminate impurities and reduce buffer carryover.

4. Add 500 μl of buffer QX3, resuspend the pellet carefully by flicking the tube. Centrifuge for 30 seconds and discard the supernatant. In this step residual salt is quantitatively removed. Repeat this wash step but centrifuge the sample for an 30 seconds on the last wash step and immediately remove any traces of ethanol with a pipette.

5. Air-dry the pellet for 5-10 minutes to remove all traces of ethanol. Ethanol may reduce enzymatic activity in subsequent reactions. Do not vacuum dry.

6. For elution of the DNA from the QIAEX add 100-150 μl TE-buffer (10mM Tris/Hcl, 1 Mm EDTA, Ph 8.0), thus lowering the salt concentration. Resuspend the pellet by flicking

the tube, and incubate for 5 minutes at 50 °C. Periodically flick the tube to keep the QIAEX in suspension.

7. Centrifuge for 2 minutes and transfer the supernatant containing the DNA to a clean tube. The QIAEX particles sediment and the supernatant contains the DNA.

8. Combine elutes and isopropanol precipitate the DNA (previously described). Resuspend in 20 μl TE and store at -20 °C until needed.

Sequencing step

The Sanger (dideoxy) method was used (Sanger *et al.*, 1977). The sequencing step involves the synthesis of a DNA strand by a DNA polymerase in vitro using a single-stranded DNA template (general strategy illustrated in Figure x). The synthesis is initiated at the site where the oliginucleotide primer anneals to the template. This synthesis reaction is terminated when a 2',3'-dideoxynucleotide 5'-triphosphate (ddNTP) is incorporated into the chain. These ddNTPs lack the 3'-OH group that is required for DNA chain elongation. The ratio of DNTPS to one of the four ddntps is 10 to 1 which ensures that enzyme-catalysed polymerization will be terminated in a fraction of the population of chains at each site where the ddntp can be incorperated. There are four separate reactions, each with a different ddntp to give complete sequence information. A radioactively labelled nucleotide, S35 is also included in the synthesis, so the labelled chains of various length can be visualized by autoradiography after separation by high-resolution electrophoresis.

The DNA synthesis is carried out in three steps. First the primer is annealed to the single-stranded DNA template. The labelling step follows in which the annealed template-primer is then extended by DNA polymerase in the presence of four limiting concentrations of the four deoxyribonucleoside triphosphates (DNTPS), including (alpha- ^{35}S) dATP. This labelling step continues until the labelled nucleotide has been completely incorperated into the DNA chains which are distributed randomly in length from several nucleotides to hundreds of nucleotides. The third step synthesis continues in the presence of additional DNTPS and one ddntp (didoexynucleotide). DNA synthesis occurs until all growing chains are terminated by a dideoxynucleotide. In this step the chains are extended by the addition of EDTA and formamide, denatured by heating and run on electrophoresis gels.

Preparation

The gel:

- 2l of buffer 10X NNB buffer is required. It is made of 324g Tris base, 55 g Boric acid, 18.6g EDTA and 800 ml distilled water.
- stock acrylamide solution, 50% acrylamide, 3.5% N,N' bisacryl. For 500 ml: 250g acrylamide, 12.5g N,N' bismethyleneacrylamide and 200ml distilled water. W
- to make up 200 ml of 6% gel solution (enough for two gels):

mix 20ml 10X NNB, 96g urea and 84ml distilled water and heat to dissolve the urea (do not boil). When the urea has dissolved add 24 ml of the acrylamide stock solution and filter. This removes impurities from the urea that may be hazardous for sequencing.

Pouring the gel

The two glass plates, between which the gel will be situated, must be very clean before the gel can be poured. 70% ethanol can be used for this. The top plate (the shorter of the two) must be thinly coated with Sigmacote (0.5 ml sigmacote spread evenly, especially the ends where the gel may stick when prising the plates apart at a later stage) over the surface. The side of the shorter plate with the sigmacote becomes the inside surface. Once the sigmacote has dried wash the excess sigmacote off with 70% ethanol.

100ml of the acrylamide working solution is used for one gel. 350 μl of 20% Ammonium persulphate and 80 μl TEMED must be added to the working solution. This bottom plate should be placed on a polystyrene block so the it is about 5cm off the work bench. Before adding the ammonium persulphate and TEMED to the acrylamide solution, clamps should be laid out next to the plates to hold them in place once the gel has been poured; two 0.35mm thick spacers must be placed one along each of the long sides of the bottom plate (the longer plate) and clamped in position with one clamp each.

Once the of ammonium persulphate and TEMED have been added to the acrylamide solution and swirled briefly, the polymerisation commences. Pour the solution onto the bottom plate, starting at the bottom so that a third to half of the plate is covered. Ensure that no bubbles or islands form. Ensuring that the side with the sigmacote is inner most, lower the bottom edge of the top plate onto the bottom plate. The solution should run along the junction between the two plates right to the edges. Ensure that bubbles are not formed. Once the top plate has been completely lowered rapidly align the plates and clamp them in place. Insert the blank at the top, and allow the gel to set for approximately and hour. If it is to be left

overnight placed pieces of filter paper dampened with buffer over each end and wrap each end in cling wrap. This prevents the gel from dehydrating.

Sequencing reactions

(The Pharmacia sequencing kit was used for the sequencing reactions which are carried out on ice).

Add 2.5 μl of each ddntp (ie G, A, T & C) and 0.5 μl of the detergent each of two reaction tubes, so that there are in total eight reaction tubes. Label four of these G, A, T & C respectively and 2 on each to indicate the ITS2 primer, similarly label the other four but with 5 to indicate the ITS 5 primer. Place these on ice for later use.

1. The annealing reactions

For each set of four sequencing lanes, a single annealing reaction is carried out. To two sterile reaction tubes labelled 2 and 5 respectively, add 7 μl template DNA, 1 μl primer (0.5 pmol/μl), 2 μl reaction buffer and 2 μl detergent. This mix is then heated to 95 °C for 2mins in the heating block and is then snap cooled in ethanol in ice for 30 seconds.

2. The labelling reactions

During the labelling reactions, the enzyme extends the primers and consumes the available DGTP, dCTP, dTTP and labelled DATP. The extensions range in size, from several to 80-

100 nucleotides, the longer extensions being far less common. Because the longer extensions contain proportionately more label they still appear on the gels.

Add 3 μl of undiluted labelling mix, 1 μl alpha S35 DATP, 2 μl enzyme (1:4 diluted with buffer) and 1.75 μl of detergent. Spin these down briefly to mix. This is carried out for 30 secs at 12 °C.

3. Termination reactions

In this step the addition of high levels of DNTPS and the ddntps at 37 °C results in further elongation of each chain at a very rapid rate until a dideoxy nucleotide is incorporated.

Pipette 4.5 μl of the labelling mix into each of the previously prepared ddntp tubes (labelled G, A, T & C) which are then left at 37 °C for 3 minutes, then add 5 μl stop solution. This mix can be stored at 4 °C until required, or loaded onto a gel immediately.

Running the gel

1. Remove the clamps and the blank from the poured gel; clean up the edges of the plates. Put the gel rig together by clipping the sides onto the glass and backing. Make sure that the backing has been vaselined to prevent the buffer leaking. Add the running buffer (1X NNB), 1 l in each compartment. Place the rig in place. Clean out the comb-well by squirting buffer into it with a needle and syringe. Pre-run the gel for 30 minutes at 75W for it to warm up to approximately 50 °C.

2. Insert the comb until the teeth penetrate the gel to a depth of about 1mm.

3. Wash out the wells again. Heat the samples to 95 °C for 1 min.

4. Load 3 μl of each sample into the wells. Close the reaction tube lids as the mixture is highly evaporative.

5. Run at 75 W for the required length of time at 50-55 °C. In this case normally about 2 hours per loading. Two markers will be visible. The dark blue is the lower marker and will run to approximately the 30bp mark. Two loadings are required for this study so steps 3-5 are repeated.

6. Before the gel has finished running, a gel fixing solution is made which consists of: 300 ml methanol, 300 ml acetic acid, 60 ml glycerol and 2340 ml water. Pour this solution into a big tray.

7. When the gel has finished running, remove the buffer from the top chamber, take the gel rig apart and lay the glass plates (with the gel) on the bench. Remove the comb and the spacers on each side of the gel carefully, and gently prise the two plates apart. The gel should stick to the bottom plate, as the top plate was coated with sigmacote to prevent the gel sticking to it.

8. Slowly lower the bottom plate and gel into the fixing solution. Allow it to fix for 30 minutes. This is to remove the urea.

9. Siphon off the fixer into a bucket (it can be used once more). Lift the plate with the gel out of the tray, place it on the bench and gently dry off the excess fixer with paper towel.

10. Lay a suitable sized piece of filter paper on top of the gel (shiny side down) and gently press it into the gel. Carefully pull the paper away from the plate. The gel should stick to the filter paper and come away cleanly.

11. Cover the exposed side of the gel with cling wrap, cut to size, and place in a vacuum gel dryer. Dry at 80 deg. C for an hour.

12. Remove from the dryer and gently remove the clingwrap ensuring that the gel is dry.

Trim the edges of the filter paper if required and place in an autorad. cassette along with a sheet of X-ray film placed emulsion side down. Leave the cassette at room temperature. The amount of exposure time will depend on the amount of DNA as well as the strength of the S35. A geiger counter can be used to test how much radioactivity is in the dried gel.

ACKNOWLEDGEMENTS

I would like to thank Nigel Barker for technical training, guidance, and computer assistance throughout the project, Eric Harley for use of facilities, Peter Linder and William Bond for helpful discussion and funding, Janet Thomas and Anton Pauw for seed collections, and finally Paul Gadek for unpublished sequences of the Cupressaceae.

REFERENCES

APPELS, R. & DVORAK, J. (1982). Relative rates of divergence of spacer and gene sequences within the RDNA region of species in the Triticeae: implications for the maintenance of homogeneity of a repeated gene family. Theoretical and Applied Genetics. 63:361-365.

BALDWIN, B.G. (1993). Molecular phylogenetics of *Calycadenia* (Compositae) based on ITS sequences of nuclear ribosomal DNA: chromosomal and morphological evolution reexamined. Amer. J. Bot. 80(2):222-238.

CHAPMAN, J.D. (1961). Some notes on the taxonomy, distribution, ecology and economic importance of *Widdringtonia*, with particular reference to W. whytei. Kirkia, ?:138-154.

CHAPOLA, G.B.J. (1990). Wood properties of wide and narrow-crowned variants of *Widdringtonia* nodiflora Powrie (Mulanje Cedar) growing at Zomba Mountain, Malawi. S.Afr.Forestry Journal, 154:47-50.

CLEGG, M.T. & DURBIN, M.L. (1990). Molecular approaches to the study of plant biosystematics. Aust. Syst. Bot. 3:1-8.

DE LAUBENFELS, D.J. (1965). The relationships of Fitzroya cupressoides (Molina) Jonhston and *Disselma* archeri J.D.Hooker based on morphological considerations. Phytomorphology, :414-419.

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

GADEK, P.A. & QUINN, C.J. (1993). An analysis of relationships within the Cupressaceae sensu stricto based on rbcL sequences. Ann. Missouri Bot. Gard. 80(3):581-581.

GONZALEZ, I.L., CHAMBERS, C., GORSKI, J.L., STAMBOLIAN, D., SCHMICKEL, R.D. & SYLVESTER, J. (1990). Sequence and structure correlation of human ribosomal transcribed spacers. J.Mol. Biol. 212:27-35.

HAMRICK, J.L. (1987). Gene flow and distribution of genetic variation in plant populations. In: Urbanski, K.M. Differentiation patterns in higher plants. Academic Press, London.

HILLIS, D.M. & DAVIS, S.K. (1988). Ribosomal DNA: Intraspecific polymorphism, concerted evolution, and phylogeny reconstruction. Syst. Zool. 32:63-66.

JORGANSEN, R.A. & CLUSTER, P.D. (1988). Modes and tempos in the evolution of nuclear ribosomal DNA: new characters for evolutionary studies and new markers for genetic and population studies. Annals Miss. Bot. Gard. 75:1238-1247.

KERFOOT, O. (1975). Origin and speciation of the Cupressaceae in Sub-Saharan Africa. Boissiera 24:145-150.

LEE, S.B. & TAYLOR, J.W. (1991). Phylogeny of five fungal-like protoctistan *Phytophythora* species inferred from the internal transcribed spacers of ribosomal DNA. Mol. Biol. Evol. 8:620-640.

LI, H. (1953). A reclassification of Libocedrus and Cupressaceae. J. Arnold Arbor. 34:17-35.

MANDERS, P.T. (unpublished report). Literature review for the Clanwilliam cedar (*Widdringtonia cedarbergensis* Marsh). Sub-project for the Dept forestry:Jonkershoek forestry research station.

MARSH, J.H. (1966). Cupressaceae 43-48. In: Codd, L.E., de Winter, B. and Rycroft, H.B. eds. Flora of Southern Africa 1. Government Printer, Pretoria.

MARTIN, P.G. & DOWD, J.M. (1991). A comparison of 18s ribosomal RNA and rubisco large subunit sequences for studying Angiosperm phylogeny. J.Mol.Evol. 33:274-282.

MASTERS, M.T. (1905). Notes on the genus *Widdringtonia*. The Journal of the Linnaen Society, Botany 37:267-274.

OLSEN, G.J. (1987). Earliest phylogenetic branchings: Comparing RRNA-based evolutionary trees inferred with various techniques. Cold Spring Harbor Symp. Quant. Biol. 52:825-837.

PAUW, A. (1993). A revision of the genus *Widdringtonia* Endl. (Cupressaceae) occurring in Malawi, Mozambique, Zimbabwe and the Transvaal (South Africa).

PENNY, D. (1982). Towards a basis for classification: The incompleteness of distance measures, incompatibility analysis and phenetic classification. J.Theoret. Biol. 96:129-142.

PHILLIPS, J.F.V. (1927). Fossil *Widdringtonia* in lignite of the Knysna series, with a note on fossil leaves of several other species. S.Afr. J.Sci. XXIV:188-197.

POWRIE, E. (1972). The typification of Brunia nodiflora L. South African J. Bot. 38(4):301-304.

RENDLE, A.B. (1893). Introduction to the plants of Milanji, Nyasaland, collected by Mr Alexander Whyte, F.L.S. Trans.
Linn.Soc.Lond.Bot. 4:1-4, 60-61.

RENDLE, A.B. (1911). Gymnosperms. In: Swynerton, C.F.M. ed. A contribution to our knowledge of the flora of Gazaland. Lin.Soc.J. 40:235-237.

ROGERS, S.O. & BENDICH, A.J. (1987). Ribosomal DNA genes in plants:variability in copy number and in the intergenic spacer. Plant Mol. Biol. 9:509-520.

RYBICKI, E. (unpublished). PCR primer design and reaction optimisation.

SAITOU, N. & NEI, M. (1987). The neighbour-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4:406-425.

SANDERSON, M.J. & DOYLE, J.J. (1992). Reconstruction of organismal and gene phylogenies from data on multigene families: Concerted evolution, homoplasy, and confidence. Syst.Biol. 41(1):4-17.

SANGER, F., NICKLEN, S. & COULSON, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. 74:5463-5467.

SCHAAL, B.A. & LEARN, G.H. (1988). Ribosomal DNA variation within and among plant populations. Ann. Missouri Bot. Gard. 75:1207-1216.

SMITH, C.A. (1955). Early 19th century records of the Clanwilliam cedar (*Widdringtonia* juniperoides Endl). J.S.Afr. For. Assoc. 15:1-8.

SNEATH, P.H.A. & SOKAL, R.R. (1973). Numerical Taxonomy. W.H. Freeman, San Francisco.

SOLTIS, P.S., DOYLE, J.J. & SOLTIS, D.E. (1992). Intraspecific chloroplast DNA variation:systematic and phylogenetic implications. In:Molecular systematics of plants. Eds. P.S. Soltis, D.E. Soltis & J.J. Doyle.

STAPF, O. (1933). Cupressaceae. In: Hill, W. ed. Flora Capensis V(2) supplement: 14-24. L. Reeve & Co. Ltd., Kent.

SWOFFORD, D. (1989). PAUP, Phylogenetic analysis using parsimony. III. Illinois Natural History Survey: Champaign.

SWOFFORD, D.L. & OLSEN, G.J. (1990). Phylogenetic reconstruction. Chapter 11. In: Hillis, D.M. and Moritz, C. eds. Molecular Systematics. Massachusetts, U.S.A.

TAKAIWA, F., OONO, K. & SUGIURA, M. (1985). Nucleotide sequence of the 17S-25S spacer region from rice RDNA. Plant Mol. Biol. 4:355-364.

WARD, J.H. (1958). Notes on the cedar of Willowmore, *Widdringtonia* schwartzii. S.Afr. For.Assoc. 81-84.

WHEELER, W.C. (1990). Nucleic acid sequence phylogeny and random outgroups. Cladistics 6:363-367.

WHITE, T.J., BRUNS, T., LEE,S. & TAYLOR, J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: PCP protocols: A guide to methods and applications. (Eds. M. Innis, D. Gelfand, J. Sninsky and T. White). pp 315-322, Academic Press, San Diego.

YOKOTA, Y., KAWATA, T., IIDA, Y., KATO, A. & TANIFUJI, S. (1989). Nucleotide sequences of the 5.8S RRNA gene and internal transcribed spacer regions in carrot and broad bean ribosomal DNA. J.Mol. Evol. 29:294-301.

ZIMMER, E.A., MARTIN, S.L., BEVERLEY, S.M., KAN, Y.W. & WILSON, A.C. (1980). Rapid duplication and loss of genes coding for the alpha chains of haemoglobin. Proceedings of the National Academy of Sciences, U.S.A. 77:2158-2162.