Species-level phylogenetic reconstruction of the African cycad genus *Encephalartos* (Zamiaceae).

Makhegu Amelia Mabunda

A thesis submitted in partial fulfilment of the requirements for the degree of Magister of Scientiae in the department of Biodiversity and Conservation Biology of the University of the Western Cape.

UNIVERSITY *of the*
WESTERN CAPE

August 2007

Species-level phylogenetic reconstruction of the African cycad genus *Encephalartos* (Zamiaceae).

Makhegu Amelia Mabunda

Keywords

Species-level

Phylogenetic reconstruction

Relationships

Encephalartos

DNA sequencing

Amplified fragment length polymorphism markers (AFLPs)

Maximum parsimony

Biogeography

Habitat optimization

Origin

ABSTRACT

SPECIES-LEVEL PHYLOGENETIC RECONSTRUCTION OF THE AFRICAN CYCAD GENUS *ENCEPHALARTOS* (ZAMIACEAE).

Makhegu Amelia Mabunda

M. Sc. Thesis, Department of Biodiversity and Conservation Biology, University of the Western Cape.

This thesis explores species-level phylogenetic relationships of the African cycad genus *Encephalartos*, which is one of the eleven genera of cycads. The genus is confined to Africa and comprises approximately 65 species, 38 of which are found naturally in South Africa. The phylogenetic studies on *Encephalartos* to date still result in many unresolved polytomies so it is not possible to fully understand the relationships between different taxa. In this study, AFLPs were used together with DNA sequencing to reconstruct the phylogenetic relationships of the genus. This study is the first to be presented with aims of resolving the relationships of *Encephalartos* using AFLPs together with DNA sequences. Total DNA was extracted from accessions sampled from the Kirstenbosch Botanical Garden and the Montgomery Collection, representing the majority of *Encephalartos* species listed in the most recent world list of cycads. Sequences of the *trnL* intron, *rpoC1*, *ITS* 1, *ITS* 2, and AFLP profiles from two sets of selective primers were used to reconstruct the phylogenetic relationships within the genus using maximum parsimony methods. As in earlier studies, unresolved polytomies were recovered from the sequencing data. The AFLP trees have some resolution but CI and RI indices were low indicating high levels of homoplasy in the data. The relationships resolved by this study for all the data sets separately and combined were different to those previously suggested for the genus. The biogeography of *Encephalartos* is also investigated by habitat optimization of the genus to estimate the origin of the genus with respect to its current distribution.

August 2007

DECLARATION

I declare that "***Species-level phylogenetic reconstruction of the African cycad genus Encephalartos (Zamiaceae)***" is my own work, that it has not been submitted for any degree at any university. It has only been submitted to the University of the Western Cape, and that all the sources I have used or quoted have been indicated and fully acknowledged by complete references.

Makhegu Amelia Mabunda August 2007

Signed:

ACKNOWLEDGEMENTS

I would firstly like to acknowledge my supervisors (Mr. Frans Weitz, Dr Gail Reeves and Dr John Donaldson) for their help and seeing the potential in me when I didn't. Many thanks as well to Dr Krystal Tolley for her input and advice.

I am also thankful to SANBI for providing me with all the facilities to do my masters, including funding, lab space and also accommodation!

Thank you to all the people at SANBI who contributed to keeping me sane. At times it was hard and frustrating and everyone was there to listen. Thank you!

Many thanks to my friends; you know yourselves…I will need another thesis to list all of you in!!!!

To the people who I dedicate this thesis to my loving and lovely parents, family (immediate and extended). I know you didn't have to, but you did it anyway. I dedicate this to your unshakable commitment for wanting to see me succeed… Ndzi mi rhandza ngopfu hinkwenu ka n'wina!

Of course, all this was made possible by the omniscient, omnipotent and omnipresent One who makes the final decisions, my God. All glory is to Him!

TABLE OF CONTENTS

Keywords.....i
ABSTRACT.....ii
DECLARATION.....iii
ACKNOWLEDGEMENTS.....iv
TABLE OF CONTENTS.....v
LIST OF ABBREVIATIONS.....vi
INDEX OF TABLES.....viii
INDEX OF FIGURES.....ix

CHAPTER 1 – INTRODUCTION 1

1. General background of cycads.....1
1.1. ***Diversity and geographical distribution***.....1
1.2. ***Threats to cycads***.....6

2. Sequencing data and AFLP markers, uses, advantages and disadvantages.....7

3. Phylogenetic reconstruction of cycads.....9
3.1. ***Phylogenetic reconstruction in genera of cycads.***.....9
3.2. ***Phylogenetic reconstruction in Encephalartos..***.....10

4. Habitat types.....14

CHAPTER 2 – MATERIALS AND METHODS.....16

2.1. Plant Material.....16

2.2. DNA Extraction.....16

2.3. PCR and DNA Sequencing.....16

2.4 . Amplified Fragment Length Polymorphism (AFLP) procedure.....18
2.4.1. ***Restriction Ligation***.....18
2.4.2. ***Pre-selective Amplification***.....18
2.4.3. ***Selective Amplification***.....18

2.5. Data Analysis.....19
2.5.1. ***AFLP markers***.....19
2.5.2. ***Parsimony analyses***.....20
2.5.3. ***Enforcing monophyly***.....21
2.5.4. ***Habitat optimization***.....21

CHAPTER 3 – RESULTS…….22

3.1. Analysis of DNA sequence data……22
3.2. Analysis of AFLP data……23
3.3. Results of partition homogeneity tests……24
3.4. Combined Analyses……25
3.5. Testing the monophyly of Cooper and Goode's groupings……26

CHAPTER 4 – DISCUSSION……35

LITERATURE CITED……47

APPENDIX 1……61

APPENDIX 2……65

APPENDIX 3 ……66

SUMMARY……68

LIST OF ABBREVIATIONS

BSA:	Bovine serum albumin
CTAB:	Cetyltrimethyl ammonium bromide
DNA:	Deoxyribonucleic acid
dNTP:	Deoxynucleotide triphosphates
g:	Gram(s)
HCl:	Hydrochloric acid
KlC:	Potassium Chloride
$MgCl_2$:	Magnesium chloride
ml:	Milliliter
mM:	Millimolar
ng:	Nanogram
PAUP*:	Phylogenetic analysis using parsimony
PCR:	Polymerase chain reaction
PVP:	Poly vinyl pyrrolidone
U:	Unit(s)
vs:	Versus
µM:	Micromolar

INDEX OF TABLES

TABLE 3.1 – Summary statistics for partition homogeneity tests……24

TABLE 3.2 – Summary table of tree statistics for each of the analyses. ..……...27

TABLE 3.3 – Results of the Kishino-Hasegawa tests ……..……26

INDEX OF FIGURES

FIGURE 1.1 – Distribution map of *Encephalartos* in Africa and Madagascar (taken directly from Donaldson, 2003)……...…4

FIGURE 1.2 – Pictures of *Encephalartos* species showing the variation in growth and cone form taken from the Kirstenbosch Collection, National Botanical Garden, Kirstenbosch…….5

FIGURE 1.3 – Cladogram of extant cycad genera and outgroups based on molecular data (reproduced directly from Chaw *et al.*, 2005)……...……....10

FIGURE 1.4 – Strict consensus tree of the recent phylogenetic reconstruction of *Encephalartos* based on *ITS* 1 & 2 data by Treutlein *et al.* (2005)..…….....13

FIGURE 3.1 – One of the equally most parsimonious trees found from analysis of 605 ITS sequences. Bootstrap percentages are shown in bold beneath branches. Those nodes collapsed in the strict consensus of 258 trees are marked by a green circle. The highlighted branches represent groups that were also found in the *ITS* tree of Treutlein *et al*. (2005) ……..……..……..……..……..……..28

FIGURE 3.2 – One of the equally most parsimonious trees found from analysis of 384 AFLP markers (using primer combination B4). Bootstrap percentages are shown in bold beneath branches. Those nodes retained in the strict consensus of 190 trees are marked by a red circle. The highlighted branches represent groups that were also found in the *ITS* tree of Treutlein *et al*. (2005)..……..29

FIGURE 3.3 – One of the equally most parsimonious trees found from analysis of 394 AFLP markers (using primer combination B13). Bootstrap percentages are shown in bold beneath branches. Those nodes collapsed in the strict consensus of two trees are marked by a green circle. The highlighted branches represent groups that were also found in the *ITS* tree Treutlein *et al*. (2005)…….……...…..30

FIGURE 3.4 – One of the equally most parsimonious trees found from analysis of 778 AFLP markers (using primer combinations B4 and B13). Bootstrap percentages are shown in bold beneath branches. Those nodes retained in the strict consensus of 30 trees are marked by a red circle. The highlighted branches represent groups that were also found in the *ITS* tree Treutlein *et al*. (2005).…..……..……..…...31

FIGURE 3.5 – One of the equally most parsimonious trees found from analysis of 778 AFLP markers (using primer combinations B4 and B13) and 605 *ITS* sequences. Bootstrap percentages are shown in bold beneath branches. Those nodes collapsed in the strict consensus of three trees are marked by a green circle. The highlighted branches represent groups that were also found in the *ITS* tree Treutlein *et al*. (2005)……...32

FIGURE 3.6 – Habitat optimization tree of *Encephalartos* using the parsimony tree from the *ITS* analysis and the habitat matrix (Appendix 3).……...…….……...33

-CHAPTER 1-

INTRODUCTION

1. General background of cycads

Cycads are the most primitive living seed plants with a fossil record dating back to the Permian 248-290 million years ago (MYA) and possibly the Carboniferous 290-354 MYA (Gonzales-Astorga *et al.*, 2003). They are one of five cone-bearing groups of seed plants (Rai, 2003), but they may be mistaken for unrelated palms and ferns when not in cone (Jones, 1993). Paleontological studies place cycads within the earliest diverging group of seed plants, Cycadophytes. This group includes the seed ferns (also unrelated to ferns) and the cycads but cycads are the only living survivors of this group (Brenner, 2003). Previous studies have grouped cycads with conifers and Ginkgo as gymnosperms but more recent studies based on molecular and morphological data indicate that the gymnosperms are paraphyletic and thus the cycads are currently placed in a single order Cycadales, in a separate class Cycadophyta (Hill, 1999).

1.1. *Diversity and geographical distribution*

Cycads have been shown to be monophyletic by both morphological and molecular phylogenetic studies. Synapomorphies in structural features such as girdling leaf traces, a specialized pattern of vascular bundles in the petiole, the presence of mucilage canals, distinctive meristems and poisonous glycosides (methylating compounds synthesized in cycad seeds that act as mutagens) called cycasins are evidence of their monophyly (Campbell *et al.,* 1999). All living genera of cycads were originally classified within a single family, the Cycadaceae. Later, the cycads were reassigned into three families (Chaw *et al.,* 2005) with *Cycas* belonging to the suborder Cycadineae and Stangeriaceae and Zamiaceae belonging to the suborder Zaminiineae (Stevenson, 1992). The three extant families and their genera are as follows: Cycadaceae (*Cycas*); Stangeriaceae (*Stangeria*, *Bowenia*) and Zamiaceae (*Dioon, Lepidozamia, Encephalartos, Ceratozamia, Macrozamia*, *Microcycas, Zamia*, and *Chigua*) (Hill *et al*., 2004).

All three families are represented in Australia and Africa, whereas the New World and Asia each have only one family, Cycadaceae and Zamiaceae respectively. The greatest diversity of cycads is found in South Africa, Australia and Mexico, although the diversity varies at family, genus and species levels. There has been wide evolutionary radiation within *Dioon, Encephalartos*, *Macrozamia*, *Zamia*, *Ceratozamia* and *Cycas*. A probable explanation as to why cycads have survived to the present day is that they are well equipped to overcome environmental stresses. They are able to survive drought and fires and are resistant to many pathogens and predators. It is thought that their secondary chemical compounds may contribute to their tolerance of these extreme conditions (Brenner *et al.*, 2003).

For much of the Earth's history, there existed a single large continent which broke apart, forming two large fragments, Gondwana and Laurasia. These subsequently underwent further fragmentation to produce the present continents (McCarthy and Rubidge, 2005). Studies on fossil cycads suggest that these plants may have originated before the split of the super continent and this also explains their present distribution around the five continents (Jones, 1993). Presently, the Cycadaceae has only one genus (*Cycas*), which is distributed along the West African coast, Madagascar, Asia, India, China, Japan, Northern Australia as well as a large number of oceanic islands of the Pacific. Stangeriaceae has two genera, *Stangeria* with one species in Africa and *Bowenia* with two species that are endemic to Australia. Zamiaceae is the largest family with eight genera and a total of 202 species (Hill *et al.,* 2004), of which only the genus *Encephalartos* occurs in Africa (Goode, 1989), see Figure 1.1. *Encephalartos* is the second largest cycad genus with 65 species (Hill *et al*., 2004), of which 38 are found naturally in South Africa (Donaldson, 2003). Recently, Cooper and Goode (2004) recognized 70 species of *Encephalartos,* but their interpretation of cycad taxonomy remains controversial and largely invalid (see later discussion). There is considerable structural and morphological diversity in growth form, cone structure, pubescence and leaf shape within this genus (see Figure 1.2).

The Cycadales have a substantial fossil record that dates back at least 70 MYA (Rai *et al.,* 2003). However, the fossil record of the extant Cycadales extends only to the Tertiary. It has been suggested that the earliest relatives of cycads arose as early as the Pennsylvanian ((Norstog and Nicholls, 1997; Hermsen *et al.,* 2006)), approximately 300MYA (Chaw *et al.,* 2005). These fossil cycads are doubtful in terms of identity and many seem to be intermediates in evolutionary stage (Osborne, 2002). Cycads have been present long enough to have experienced and survived several mass extinction events including the ones that took place at the beginning and end of the Mesozoic era (Cáceres, 1998). Cycads represent the oldest surviving lineage of seed plants and hence play a significant role in our understanding of the evolution of morphological characters in angiosperms and gymnosperms (Brenner *et al.*, 2003).

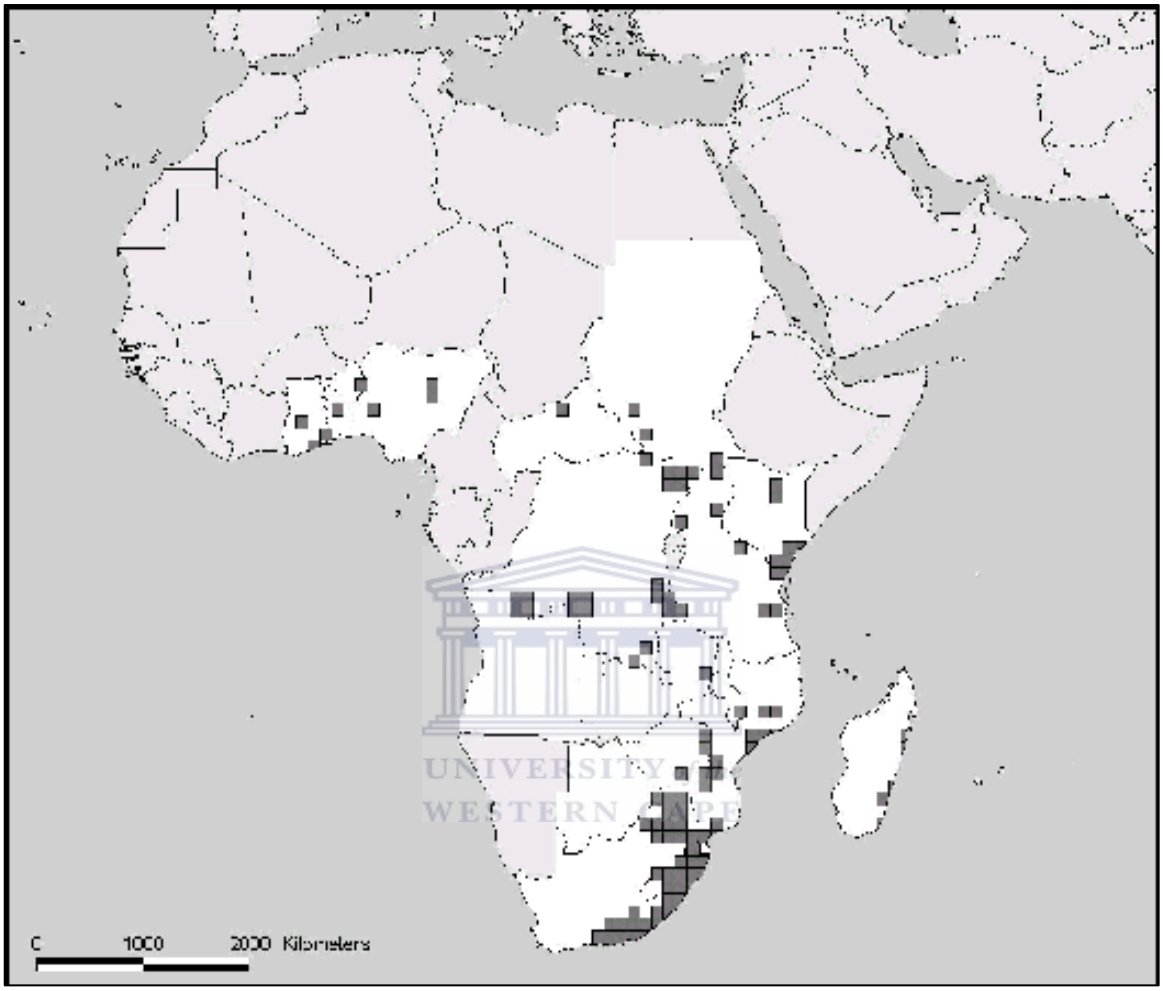

Figure 1.1 – Distribution map of *Encephalartos* in Africa and Madagascar (copied directly from Donaldson, 2003) illustrated by dark grids.

E. altenstenii Lehm.

E. horridus (Jacq). Lehm.

Suckers of *E. friderici-guilielmi* Lehm.

E. gratus Prain

Woolly cones of *E. friderici-guilielmi* Lehm.

Cone of *E. Horridus* (Jacq).

Figure 1.2 – Pictures of *Encephalartos* species showing the variation in growth and cone form taken from the Kirstenbosch Collection, National Botanical Garden, Kirstenbosch. Photos: Amelia Mabunda

1.2. *Threats to cycads*

There are still new species of cycads being discovered in tropical Africa, especially north of the Limpopo River (which forms a border between South Africa, Botswana and Zimbabwe). In this area, many species of *Encephalartos* are on the brink of extinction in their natural habitat (Donaldson *et al.*, 2003 and Goode, 2001) largely due to a combination of reproductive failure and illegal collecting. Other factors include the combined effects of man's domestic and agricultural demands. Due to over-exploitation, conservation efforts of these plants are hindered by their slow growth rates - they only produce cones when they reach maturity and this can take up to 15 years (Norstog and Nicholls, 1997), coupled with their limited potential for vegetative propagation (Osborne, 1989).

It has also been demonstrated that pollinator survival is linked to the size and composition of the cycad flora and that decreasing cycad populations could lead to decreasing pollinator populations and vise versa (Donaldson, 2004). Thus the extinction of these pollinators is also one of the factors that may be contributing to the decline of cycad populations. Evidence is accumulating that cycads have symbiotic relationships with specific insects (weevils), and it is these insects that have been recorded to play a major part in cycad pollination (Oberprieler, 2004; Jones, 2002; Norstog and Nicholls, 1997; Oberprieler, 1995(a, b); Norstog and Fawcetti, 1995 and Vorster, 1995). There are 60 species of weevils belonging to 14 genera that have been recorded to have a close affinity to cycads, and thus a close association in diversity between cycads and their pollinators is perhaps to be expected. Therefore, these weevils may also shed light into the evolution of the cycads (Oberprieler, 1995a). An interesting example of co-evolution of plants and their pollinators is that between figs (*Ficus*) and fig wasps (Hymenoptera, Agaonidae). The figs comprise of about 750 species and the wasps of more than 700 species-specific couples. Co-evolution of the fig and fig wasps has been used as an aid to explore the evolutionary pathways of *Ficus*. However, it was found later than there may be independent evolution of some *Ficus* traits, associated with differences in the pollination behaviors of the wasps (Michaloudi *et al*., 2005 and Jousselin *et al.,* 2003).

Michaloudi *et al.* (2005) confirmed that there may be host-switching of wasps between the fig species. These sorts of relationships help to clarify taxonomic uncertainties in species that lack distinct morphological, ecological, geographical etc. characters. The nature of plants and their pollinators fosters new awareness into the organism's evolution. The biogeography of pollinators may also provide substantial evidence on the evolution of their host species when molecular data on its own fails.

2. Sequencing data and AFLP markers, uses, advantages and disadvantages

DNA sequence data have played a fundamental role in reconstructing phylogenies of a wide range of organisms. In plants, the most commonly used DNA markers are from the plastid genome. They are generally easy to amplify and have high levels of variation at higher taxonomic levels (Bailey *et al.,* 2004). Plastid DNA sequences have been shown to be powerful at resolving family and generic-level relationships using regions such as *rbcL*. There has however been a common problem with low variation and resolution when attempting to resolve relationships within closely related or morphologically diverse species representing large genera (Bailey *et al.,* 2004; Crawford, *et al.,* 2002). The most common alternative region for reconstruction of species-level phylogenies is the internal transcribed spacer (*ITS*) region from the nuclear genome. However, due to its presence in multiple repeat units in the genome, this region can suffer from problems associated with paralogy (Bailey *et al.,* 2003) and still in many cases it fails to provide sufficient phylogenetically informative characters at low taxonomic levels due to low variation.

Sequencing data has been used by many authors to infer species-level relationships of various plant genera (*e.g.* Ahmed *et al.,* 2006; Tsai *et al.,* 2006; Lee and Wen 2004; Shaw and Small, 2004; Pelser *et al.,* 2002; and Richardson *et al.,* 2001), in animal studies (*e.g.* Duftner *et al.,* 2006; Tolley *et al.,* 2006; Baker *et al.,* 2005 and Weins and Pankrot, 2002), inferring relationships at higher taxonomic levels (*e.g.* Chaw *et al.,* 2005; Bogler and Francisco-Ortega, 2004; Simões *et al.,* 2004; Hill *et al.,* 2003; Rai *et al.,* 2003 and Klak *et al.,* 2003) and also hybridization studies (*e.g.* Baumel *et al.,* 2002).

In cases where DNA sequence data has been shown to be inadequate for species-level phylogenetic reconstruction, many investigators (*e.g.* Mekanawakul *et al.,* 2004; Després *et al.*, 2003; Badr *et al*., 2002; Gimenes *et al.,* 2002 and Koopman *et al.,* 2001) have turned to alternative markers such as AFLPs which have been shown to have the potential to resolve relationships at lower taxonomic levels.

The AFLP technique developed was by Vos *et al.* (1995) for DNA fingerprinting. It is based on the selective PCR amplification of restriction fragments from a total digest of genomic DNA. It is used to visualize hundreds of amplified DNA restriction fragments simultaneously and can generate fragments of 50 to 500 base pairs in size. The dependence on sequence knowledge of the target genome is eliminated by the use of adapters of a known sequence that are aligned to the restriction fragments (Muluvi, 1999; Palacios *et al.*, 1999; Vos *et al.,* 1995). One of the advantages of using multilocus markers like AFLPs over DNA sequencing is that the numerous fragments generated come from different linkage groups in the nuclear genome. Therefore, the potential problem of producing a phylogeny with a small region of the genome is diminished (Crawford *et al.,* 2004). Because of the numerous numbers of markers that are generated, some AFLP markers will be located in variable regions and therefore revealing minor differences between the species studied. The AFLP markers are also treated as dominant since the identity of homo/heterozygotes cannot be established unless breeding/pedigree studies are carried out to determine the inheritance patterns of each band. They are therefore scored as either present or absent. However, the large number of bands gives an estimate of variation across the entire genome, thus giving a good general picture of the level of genetic variation (Mueller and Wolfebarger, 1999). The technique has also been shown to have a high rate of polymorphism in a single assay when compared with other fingerprinting techniques such as RFLPs (restriction fragment length polymorphisms) or RAPDs (randomly amplified polymorphic DNAs) (Federici *et al.*, 2001 and Mueller and Wolfebarger, 1999).

AFLPs have recently been found to have a broad taxonomic applicability in studying relationships within members of various taxa. The technique has been used in plants for

species-level phylogenies (*e.g.* Mekanawakul *et al.,* 2004; Després *et al.*, 2003; Badr *et al*., 2002; Gimenes *et al.,* 2002 and Koopman *et al.,* 2001), population genetics studies of both endangered and cultivated plants (*e.g* Da Silva, 2005; Lin *et al.,* 2004 and Palacios *et al.*, 1999). The technique has also been applied increasingly in animal studies (*e.g* Sullivan *et al.,* 2004; Allender *et al.,* 2003; Ogden and Thorpe, 2002; Parsons and Shaw, 2001 and Albertson *et al.,* 1999). All the studies highlighted the usefulness and reproducibility of the AFLP technique in resolving phylogenetic relationships within the various studied taxa.

3. Phylogenetic Reconstruction of cycads

3.1. *Phylogenetic reconstruction in genera of Cycads*

One of the first molecular systematic studies of cycads by Caputo *et al.* (1991) showed the potential of using molecular markers to resolve relationships among the genera in this group. Phylogenetic analyses using parsimony showed that *Dioon* was sister to the rest of the American genera. The topology of the trees was also congruent with that of previous morphological data, making the Zamioideae a monophyletic unit and *Dioon* their sister group. Caputo *et al.* (1993) extended the study by including Old World genera and the results did not conform to previous morphological results when *Macrozamia* was included. The past few years have seen several phylogenetic relationships hypothesized using morphological characters (de Laubenfels, 1999 and Schutzman *et al.,* 1993) and molecular data (Chaw *et al.,* 2005; Bogler and Francisco-Ortega, 2004; Hill *et al.,* 2003; Rai *et al.,* 2003; Treutlein and Wink, 2002). Most if not all the studies used the formal classification of the extant cycads by Stevenson (1992) as a platform to compare their current findings. These studies are however not congruent with each other in terms of tree topologies in that they depict different phylogenetic relationships between the genera. The African genus *Encephalartos* was found by the studies to be closely related to *Macrozamia* and *Lepidozamia*, which are endemic to Australia. The sister relationships between *Encephalartos* and *Lepidozamia* remains unclear though due to lack of fossil

record (Chaw *et al.,* 2005). A simplified phylogeny of the genera is illustrated in Figure 1.3. below.

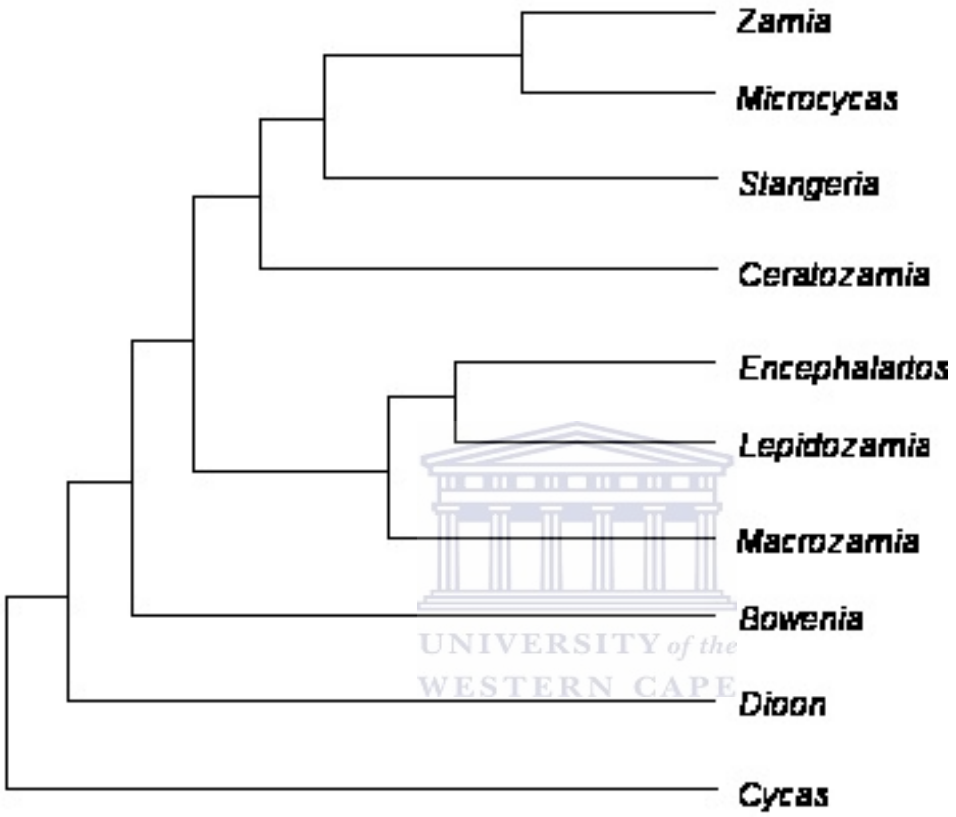

Figure 1.3 – Cladogram of cycad living genera and outgroups based on molecular data (reproduced directly from Chaw *et al.,* 2005).

3.2. *Phylogenetic reconstruction in Encephalartos*

The taxonomy of the South African *Encephalartos* species was well defined by Dyer (1965b) and the description only included 26 species of the genus. Until recently, most research into the evolutionary relationships within the genus *Encephalartos* has been based upon morphological and biochemical characters (e.g. Osborne *et al.,* 1993). From a morphological perspective, Vorster (1993) reviewed the macroscopical external characteristics of the species of *Encephalartos* in South Africa in search of characters with taxonomic value to describe the species. However, he did highlight that these

characters did not render a full circumscription of phylogenetic groupings of species within the genus. In 1999, the same author presented a revision of the tropical species of *Encephalartos*. His conclusion was that the species have very little diversity in both vegetative and cone characteristics, providing very few distinguishing features.

A preliminary non-molecular study of *Encephalartos* by Osborne *et al.* (1993) used 86 morphological, vegetative and biochemical characters to reconstruct relationships among 52 species of the 65 species currently described (Hill *et al.,* 2004). The relationships between species were analyzed using phenetic methods and a phenogram was produced to reflect evolutionary relationships. The outcome of this analysis was the identification of five groups within the genus, with five of the 52 species in an unresolved position. This study has created a platform from which patterns of relationships can be tested in future phylogenetic reconstructions within the genus.

The most recent publication on the molecular phylogenetics of the African genus *Encephalartos* is by Treutlein *et al.* (2005) (Figure 1.4). Their study of 51 species distributed from Nigeria through to South Africa, and using *ITS* 1 and *ITS* 2 sequences as well as *rbcL* and genomic ISSR fingerprinting techniques, showed that most of the recovered relationships agreed with previous morphological data and three major clades within the genus were identified. These clusters have overlapping distribution ranges in the eastern parts of South Africa. The study recovered similar groupings to those of Osborne *et al.,* (1993), although the relationships between some terminal taxa were different. There was distinctive clustering which agreed with both morphology and geographical distribution, although there was very little resolution within the *rbcL* and ISSR data. This study confirmed the work by Van der Bank *et al.* (2001) that the genus displays maximum genetic diversity in the mountain regions of the eastern parts of South Africa as it contains the largest diversity of *Encephalartos* genotypes.

In a controversial contribution to the taxonomy of *Encephalartos*, Cooper and Goode (2004) split the genus into eleven genera, which they regard as natural groupings. No formal phylogenetic analysis was used to derive the groupings and the nomenclatural changes are largely invalid, as they do not follow the international code for botanical nomenclature. Although Treutlein *et al.* (2005) resolved three main clades, there were still many unresolved polytomies (Fig. 1.4). Lack of resolution hinders our understanding of the basic taxonomy as well as the evolutionary relationships among taxa, and makes the controversial taxonomy of Cooper and Goode difficult to test.

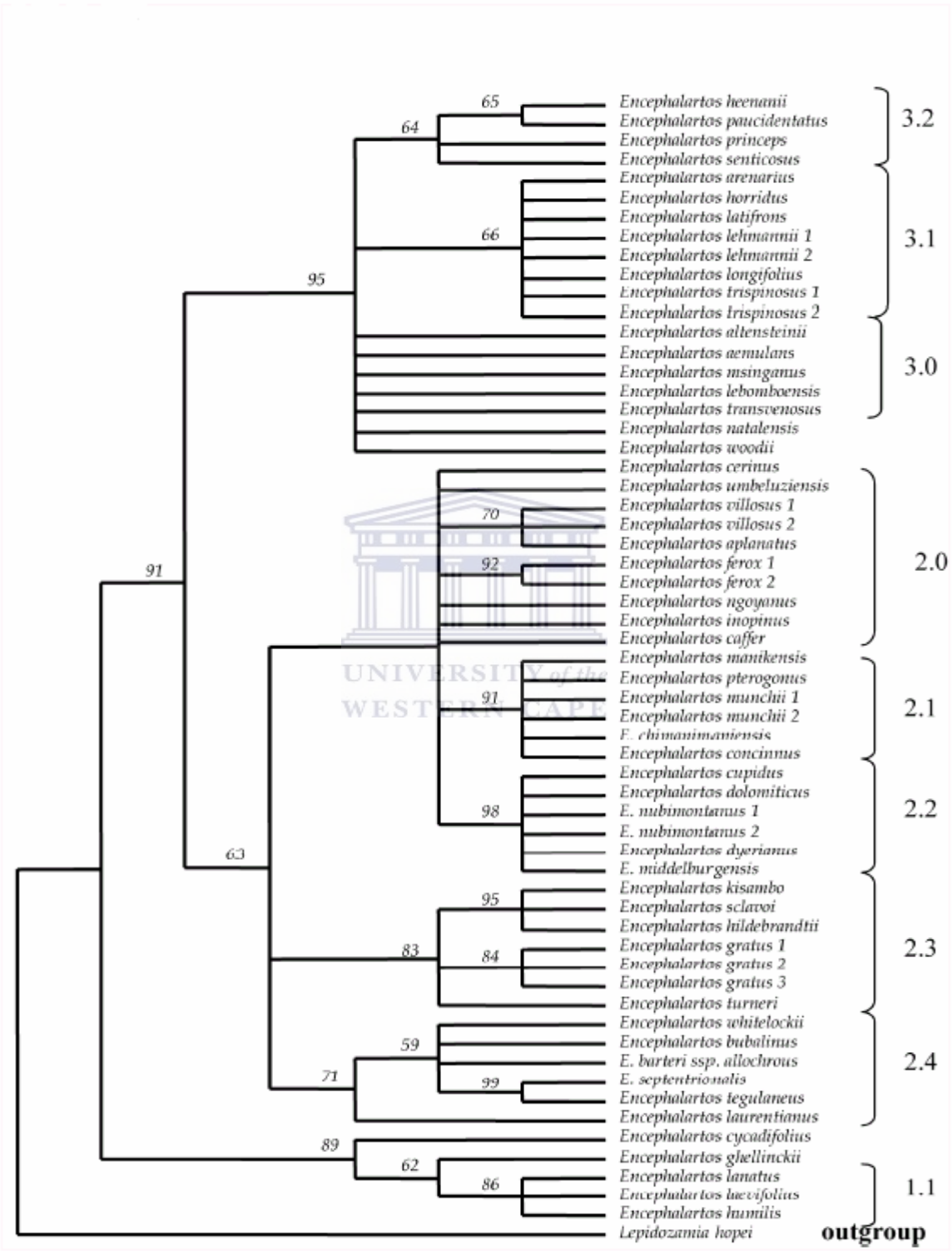

Figure 1.4 – A strict consensus tree of the recent phylogenetic reconstruction of *Encephalartos* based on *ITS* 1 & 2 data by Treutlein *et al.*, (2005). The three clusters are labeled as *ITS* groups 1, 2 and 3 with their subgroups labeled as 1.1, 2.1 etc. Bootstrap values are indicated above the branches. Reproduced directly from Treutlein *et al.,* 2005.

4. Habitat types

All cycads are considered to be xerophytes (inhabit a variety of dry habitats). They reached their peak in the late Triassic and inhabited various ecological habitats (riverine forests to grassland) along with other species (McCarthy and Rubidge, 2005). The family Zamiaceae is more diverse in terms of its geographical distribution as it is not confined to one continent (Donaldson, 2003), implying that it may have been present before the break-up of the super continent Pangaea. The current distribution of cycads may also be due to radiations from a few ancestral types isolated on Laurasia and Gondwana, or could be explained by genetic drift following the separation of already evolved genera. Optimizing the habitat types of Cycads would also open up new possibilities into interpreting the history of the genus *Encephalartos*. The separation of sister taxa *Encephalartos* (Africa) and *Lepidozamia* (Australia) could be correlated with the separation of Africa and Australia approximately 80MYA (Hill *et al*., 2003) as they appear to have evolved from a common ancestor perhaps in Gondwana (200-135MYA), before Africa and Australia split (Bogler and Francisco-Ortega, 2004). Treutlein *et al.* (2002) suggested that the divergence between these three genera could be explained more easily by Miocene long distance dispersal than by continental drift. Hermsen *et al.* (2006) found, using minimum age mapping techniques that *Encephalartos* appears to have split from the *Lepidozamia-Macrozamia* lineage at approximately 33 MYA and no later than the Eocene. Because these genera are thought to have split during the breakup of Gondwana which started approximately 140MYA (McCarthy and Rubidge 2005), it would be feasible to criticize the latest date by Hermsen *et al.* (2006) for their split in that the date contradicts what the previous authors have concluded. Therefore, optimizing the habitats of *Encephalartos* in Africa may give clarity as to which habitat is most ancestral and thus giving insight into how and when the habitats evolved. Dating the divergence of *Encephalartos* and the diversification of different groups within the genus would provide valuable insights into the evolution of these groups in relation to major geomorphological and climatic events. However, this remains difficult when several clades remain poorly resolved in existing phylogenies.

To date, phylogenetic relationships within *Encephalartos* remain poorly understood. This is largely due to the problem of unresolved polytomies and weakly supported clades within the existing phylogenetic studies and thus there remains a gap to look for additional molecular markers to resolve the relationships of the genus.

Therefore, the aims of this study are:

1. To reconstruct and investigate phylogenetic relationships among species of the genus *Encephalartos* in Africa using AFLP and sequencing data to complement previous studies that used DNA sequencing and morphological characters.

2. To test the phylogenetic hypothesis from this study against the taxonomic relationships for *Encephalartos* proposed by Cooper and Goode (2004), and to compare the phylogenetic hypothesis from this study with that suggested by previous studies using DNA markers.

3. To map habitat types on the obtained phylogeny to make inferences regarding ancestral habitat types.

4. To evaluate the utility of the AFLP and DNA sequencing methods in reconstructing the phylogeny of the genus and make recommendations for follow-up studies.

The chapter outline for the rest of the thesis is as follows:

- Chapter 2- Materials and Methods.
- Chapter 3- Results of all analyses done.
- Chapter 4- Discussion of results, conclusions and technical recommendations for future studies.

-CHAPTER 2-

MATERIALS AND METHODS

2.1. Plant material

The plant material used in this study was collected from the National Botanical Garden in Kirstenbosch (KBG), South Africa and from the Montgomery Botanical Center (MBC) in Miami, United States of America (Appendix 1). In total, 47 species of *Encephalartos* were sampled, 29 from Kirstenbosch and 18 from the MBC. The following taxa *Dioon edule*, *Dioon merolae*, *Macrozamia riedlei* and *Zamia furfuracea* were obtained from the Conservation Biology greenhouse in Kirstenbosch and sampled as outgroups. Dr. J. S. Donaldson, from the South African National Biodiversity Institute (SANBI) verified the identification of the plant samples collected from the Kirstenbosch collection. Fifty one internal transcriber spacer (*ITS* 1 and 2) sequences of *Encephalartos* previously published by Treutlein *et al.* (2005) were also extracted from GenBank (see Appendix 1).

2.2. DNA Extraction

Total genomic DNA was extracted from 0.5-1.0 g silica dried or fresh leaf material using the 2XCTAB protocol described by Doyle and Doyle (1987) modified by the addition of 2 % PVP to the extraction buffer. All DNA extracts were purified using QIAquick silica columns (QIAGEN) and the remaining supernatant suspended in 100 % ethanol and allowed to precipitate in a –20 ºC freezer. Further purification was carried out using a cesium-chloride ethidium-bromide density gradient (1.55 g/ml) (Savolainen *et al*., 2006).

2.3. PCR and DNA Sequencing

DNA amplification was performed using 100 µl reactions containing Promega magnesium thermophilic buffer (50 mM KCL, 10 mM Tris-HCL and 0.1% Triton X-100), 3 mM $MgCl_2$, 0.004% BSA (Savolainen *et al*., 1995), 0.2 mM of each dNTP, 10µM of each primer and 2.5 U *Taq* polymerase. The plastid *trnL* intron was amplified using

primers '*c*' (5'-CGA AAT CGG TAG ACG CTA CG-'3) and '*d*' (5'-GGG GAT AGA GGG ACT TGA AC-'3) (Taberlet *et al*., 1991). A partial sequence of the plastid *rpoC1* intron[♣] was amplified using primers 1F (GTGGATACACTTCTTGATAA) and 3R (TGAGAAAACATAAGTAAACG). The *trnL-F* intergenic spacer ('*e*' and '*f*' primers of Taberlet *et al.*, 1991) proved impossible to sequence for most taxa due to several long homopolymer (A or T) regions.

Amplification was carried out in a Gene Amp PCR System 9700 (Applied Biosystems) using the following program for the *trnL* intron: initial denaturing at 94 °C, two minutes; then 30 cycles of denaturing, 94 °C, one minute; annealing 52 °C, one minute; extension, 72 °C, one minute and final extension 72 °C, seven minutes. The following program was used for the *rpoC1* intron: initial denaturing at 94 °C, one minute; then 32 cycles of denaturing, 94 °C, 30 seconds; annealing 48 °C, 40 seconds; extension, 72 °C, 40 seconds and final extension 72 °C, five minutes. The amplified fragments were verified by electrophoresis on a 1% agarose gel stained with ethidium bromide and visualized under ultraviolet light. Purification of successfully amplified fragments was carried out using GFX^{TM} columns (Amersham Bioscience) and sequenced using PCR primers as sequencing primers and Big Dye terminator mix (Applied Biosystems) following the manufacturer's protocol. The products were separated on a denaturing polyacrylamide gel and run on an ABI 377 automated sequencer.

Sequences were edited and assembled using Sequencher 4.1 (Gene Codes Corporation). All assembled sequences were transferred to PAUP* version 4.02b (Swofford, 2000) and aligned manually.

♣ Special thanks to the DNA barcoding of land plants project as part of the Consortium for the barcoding of life funded by the Sloan and Betty Moore foundations for allowing me to use the sequences as part of my thesis.

2.4. Amplified fragment length polymorphism (AFLP) procedure.

AFLP markers were generated according to the AFLP™ Plant Mapping Protocol (Applied Biosystems). The three main steps are outlined below:

2.4.1. *Restriction ligation*

Genomic DNA was digested with two restriction enzymes: *Eco*RI (a rare cutter with a six base pair recognition site) and *Mse*I (a frequent cutter with a four base pair recognition site) to generate a large number of fragments. *Eco*RI - *Mse*I fragments are preferentially amplified (rather than *Eco*RI - *Eco*RI or *Mse*I - *Mse*I fragments). In the same reaction, the digested DNA was ligated using adapters (*Mse*I and *Eco*RI adapters) to generate template DNA for amplification and also to prevent a second restriction from taking place after the ligation has occurred. These adapter sequences served as primer-binding sites to amplify the restriction fragments.

2.4.2. *Pre-selective amplification*

The purpose of this step was to increase the amount of template available for mapping and to reduce the number of amplified fragments by 16-fold. The pre-selective primers used consisted of a core sequence, an enzyme specific sequence and a selective single-base extension at the 3'-end. The adapters from the previous reaction served as primer-binding sites for the restriction fragments, making it possible to amplify many DNA fragments without prior knowledge (Vos *et al.*, 1995).

2.4.3. *Selective amplification*

The AFLP™ Plant Mapping protocol from Applied Biosystems provides eight *Eco*RI and eight *Mse*I primers, which yield a possible 64 primer combinations to use for selective amplification.

This step was performed by using the pre-selective amplification products as a template for a second amplification step using fluorescently labeled primers that consisted of an identical sequence to the pre-selective primers, with three selective nucleotides at the 3'-end. From the large amount of fragments generated by the restriction enzymes, only the subset of fragments with the matching nucleotides were amplified, thereby reducing the complexity of the products (Vos *et al*., 1995). AFLP profiles were generated with two selective PCR primer pairs *Mse*I-CAT/*Eco*RI-ACA (FAM) and *Mse*I-CTA/*Eco*RI-ACT (FAM). From now on these primer combinations are referred to as B4 and B13 respectively. Only the EcoR1 primers are fluorescently labeled.

The final PCR products were denatured in a mixture containing deionised formamide, loading dye and GeneScan 500 ROX^{TM} size standard at 95 ºC for 3 minutes. The samples were then run on a 5% denaturing polyacrylamide gel according to manufacturer's protocols on an ABI 377 automated sequencer.

The generated fragment data were analyzed using GeneScan version 3.1.2 and Genotyper version 2.5 (Applied Biosystems). The latter program allowed for the visualization of band patterns as a fingerprint trace that could be further inspected by eye for any possible misinterpretation from automated procedures. Fragments of 50-500 base pairs (bp) were scored as either present or absent. The use of internal size standards in each lane permitted exact calibration of different fragments against each other and made possible separation of non-homologous fragments that were nearly equal in length. Additional bands that were not automatically scored by Genotyper were scored where the presence of fragments was obvious as distinct shoulders of more intense bands of an adjoining size class. Additional bands for samples with weak signals were also scored. The data were extracted as a table scored as either present (1) or absent (0) to produce a binary matrix.

2.5. Data analysis

2.5.1. *AFLP markers*

The primer combinations used for *Encephalartos* did not work well for the outgroup taxa (this is not surprising as AFLP primer combinations are often species specific and in this

case specific to this genus). As a result the AFLP trees were rooted using information from the *ITS* phylogeny by Treutlein *et al*. (2005) and also Osborne *et al*. (1993) and Oberprieler (1995). The latter author concluded that *E. cycadifolius*, *E. friderici-guilielmi*, *E. ghellinckii*, *E. humilis* and *E. lanatus* belong to a complex that appears to have split from the rest of the group at an earlier age, this split is also evident in the *ITS* tree of Treutlein *et al*. (2005).

2.5.1. *Parsimony analyses*

In total five data matrices were available for analysis - two AFLP (B4 and B13), two plastid (*trnL* intron and *rpoC1* intron) and one nuclear (*ITS*1&2 were treated as one data set). Topological congruence of the data sets was evaluated using the partition homogeneity test (Farris *et al.*, 1995) implemented in PAUP*. One hundred partition homogeneity replicates with 100 replicates of random taxon addition and tree bisection reconnection (TBR) branch swapping were used. The congruence of the phylogenetic signal was evaluated between each possible pair of data matrices that were analyzed (AFLP B4 versus AFLP B13, AFLP B4 versus *ITS*, AFLP B13 versus *ITS* and combined AFLP versus *ITS* sequences). The statistical P-value was recorded as an indication of significant incongruence between the data sets.

All phylogenetic analyses for the following data matrices were performed using the parsimony algorithm PAUP* version 4.02b (Swofford, 2000) using the maximum parsimony algorithm:

- *trnL* intron for 49 taxa
- *rpoC1* intron for 35 taxa
- ITS 1&2 for 51 taxa
- AFLP B4 for 44 taxa
- AFLP B13 for 44 taxa
- AFLP B4 and B13 combined for 42 taxa
- AFLP B4 and B13 and *ITS* 1&2 combined for 46 taxa

Heuristic searches were conducted with 1000 replicates of random taxon addition and TBR branch swapping, saving 10 trees per replicate. All characters were treated as unordered, weighted equally and gaps were treated as missing data. Internal nodal support was assessed using 1000 bootstrap replicates with simple taxon addition saving 10 trees per replicate. Only those groups with bootstrap values above 50% were reported.

2.5.3. *Enforcing monophyly*

The *ITS* tree published by Treutlein *et al.* (2005) was used to evaluate the groupings within *Encephalartos* (Appendix 2) proposed by Cooper and Goode (2004). Using MacClade v4.01 (Maddison and Madison, 2001) the monophyly of the species-groups in Appendix 2 was enforced onto one of the equally most parsimonious trees obtained in this present study from the analysis of *ITS* sequences. In order to examine the statistical significance of the tree topologies, the Kishino-Hasegawa test (Kishino and Hasegawa, 1989) was performed using both likelihood and parsimony criteria to assess whether there was substantial difference in tree length between the original tree and that with Cooper and Goode's groups enforced. The number of additional steps and likelihood scores were used as an indication as to whether Cooper and Goode's groupings represent a significant deviation from those recovered in the *ITS* tree.

2.5.4. *Habitat optimization*

The habitat in which each species occurs was mapped onto one of equally most parsimonious trees (chosen at random) obtained from the analysis of *ITS* sequences. A habitat matrix was drawn-up in MacClade v4.01 (Maddison and Maddison, 2001). Seven habitat types were determined from the literature and were defined as follows: forest, grassland, woodland/savannah, shrub land, granite outcrop, quartzite hills and cliffs/river gorges (Appendix 3). Most species were allocated their primary habitats; however, some of them occur in places in which the surrounding habitat is not necessarily their primary habitat. These species were therefore given multiple habitat scores.

-CHAPTER 3-

RESULTS

3.1. Analysis of DNA sequence data:

trnL intron – *trnL* intron sequences were generated for 49 ingroup and four outgroup taxa. Of the 480 characters included in the parsimony analysis 413 were constant, 67 were variable and of these 15 were potentially parsimony informative (however most of these characters represented informative variation between *Encephalartos* and the outgroup taxa). One thousand replicates of random taxon addition resulted in 10 000 equally parsimonious trees of length 74 and consistency index (CI) = 0.973 and retention index (RI) = 0.926. Very few variable characters (most of which were autapomorphic) were found within *Encephalartos* and therefore this region provided very little information to infer species-level relationships. Due to the lack of variability in this region the trees have not been shown. However, the tree statistics corresponding to this analysis and those to follow are summarized in Table 3.1.

rpoC1 intron – rpoC1 intron sequences were generated for 35 taxa. Of the 509 characters included in the analysis 489 were constant, 20 were variable and of these 3 were potentially parsimony informative. The analysis using 1000 replicates of random taxon addition yielded 7020 equally parsimonious trees with tree length of 32 steps, CI=1.000 and RI=1.000. As for the *trnL* intron, levels of sequence variability were extremely low and yielded no information regarding species-level relationships. For this reason the trees have not been shown here.

ITS region – ITS 1 & 2 sequences for 51 taxa were downloaded from GenBank. A total of 605 characters were analyzed, of which 525 were constant, 80 were variable and of these 45 were potentially parsimony informative. Parsimony analysis using 1000 replicates of random taxon addition yielded 258 equally parsimonious trees with tree length of 96 steps, CI=0.885 and RI=0.956. The groups retrieved in this analysis were the same as those presented by Treutlein *et al.* (2005). One of the equally most parsimonious trees

was chosen at random and is shown in Figure 3.1. The groups and subgroups as presented by Treutlein *et al.* (2005) are indicated on Figure 3.1. Bootstrap percentages >50% are shown below branches, and those nodes not recovered in the strict consensus of 258 trees are indicated by green circles.

3.2. Analysis of AFLP data:

B4 data set – In total 44 taxa were analyzed for which a total of 384 fragments were scored. Of the 384 characters included in the analysis eight characters were constant, 376 characters were variable and of these 305 characters were potentially parsimony informative. A heuristic parsimony search using 1000 replicates of random taxon addition yielded 190 equally parsimonious trees with tree length 1500, CI= 0.251 and RI = 0.362. One of the equally most parsimonious trees was chosen at random and is shown in Figure 3.2. Only three nodes were retained in the strict consensus tree and these are indicated with red circles. The bootstrap analysis recovered six nodes with bootstrap support >50% -these values are indicated below branches.

Taxon relationships that are highlighted according to their similarity in the *ITS* tree of Treutlein *et al.* (2005) are those of *E. inopinus* and *E. umbeluziensis* representing subgroup 2.0; *E. heenanii, E. longifolius, E. natalensis and E. princeps* representing group 3; *E. concinnus* and *E. pterogonus* representing subgroup 2.1; *E. arenarius*, *E. senticosus* and *E. trispinosus* representing group 3 and *E. cupidus, E. laurentianus, E. manikensis* and *E. tegulaneus* representing group 2. While some groupings were recovered that were similar to those found in the *ITS* tree, the split between the grassland species (*E. cycadifolius*, *E. friderici-guilielmi, E. ghellinckii*, *E. humilis* and *E. lanatus*) belonging to a complex that appears to have split from the rest of the group at an earlier age was not evident in the trees recovered from analysis of this AFLP data set and the data sets that follow.

B13 data set – A total of 44 taxa were included for which 394 fragments were scored. Of the 394 characters included in the analysis 20 characters were constant, 374 characters

were variable and of these 307 were potentially parsimony informative. One thousand replicates of random taxon addition gave two equally parsimonious trees with tree length 1567, CI= 0.234 and RI = 0.381. One of the two equally parsimonious trees was chosen at random and is shown in Figure 3.3. Only one node collapsed in the strict consensus tree and this is indicated by a green circle. Six nodes received bootstrap support of >50% and these are indicated below branches.

Taxon relationships that are highlighted according to their similarity in the *ITS* tree of Treutlein *et al.* (2005) are those of *E. altensteinii* and *E. horridus* representing subgroup 3.1; *E. concinnus, E. gratus, E. latifrons, E. longifolius* and *E. sclavoi* representing subgroup 2; *E. manikensis* and *E. pterogonus* representing subgroup 2.1 and *E. paucidentatus* and *E. woodii* representing group 3.

3.3. Results of partition homogeneity tests:

This method tests for congruence between trees derived from different sources of evidence. If the comparison between two tree topologies receives a P-value >0.05 the difference in tree topologies is considered not-significant, and suggests that the two data sets are congruent with respect to their phylogenetic signal and can be combined into a single analysis (Farris *et al.,* 1995).

Of the four incongruence tests performed all comparisons resulted in p-values showing significant incongruence (all p=0.01; see Table 3.2).

Table 3.1: Summary statistics for the partition homogeneity test showing the data partitions and their corresponding p-values.

Partition	P-value
AFLP B4 vs AFLP B13	0.01
AFLP B4 vs *ITS*	0.01
AFLP B13 vs *ITS*	0.01
AFLP B4 and B13 combined vs *ITS*	0.01

Despite the negative results of the partition homogeneity test there were no strongly (>90%) supported and incongruent clades in the separate analyses, and thus the AFLP data sets and *ITS* data were combined directly.

3.4. Combined Analyses:

Analysis of B4 and B13 AFLP markers combined – One thousand replicates of random taxon addition gave 30 equally parsimonious trees with tree length 3138, CI= 0.233 and RI = 0.303. One of the equally most parsimonious trees was chosen at random and is shown in Figure 3.4. Nine nodes were retained in the strict consensus tree and these are indicated with a red circle. Only eight nodes received bootstrap support >50% and these are indicated below branches.

Taxon relationships that are highlighted according to their similarity in the *ITS* tree of Treutlein *et al.* (2005) are those of *E. cupidus* and *E. manikensis* representing group 2; *E. altensteinii* and *E. horridus* representing group 3; *E. natalensis* and *E. princeps* representing group 3; *E. concinnus, E. gratus*, *E. ituriensis* and *E. pterogonus* representing group 2.

Combined ITS and AFLP analysis – In total 1383 characters were analyzed for 58 taxa. Of the 1383 characters included in the analysis 558 were constant, 825 were variable and of these 658 were potentially parsimony informative. The heuristic tree search using 1000 replicates of random taxon addition recovered 3 equally parsimonious trees with tree length of 3557, CI=0.233 and RI=0.304. One of the equally most parsimonious trees was chosen at random and is shown in Figure 3.5. Sixteen nodes collapsed in the strict consensus tree and these are indicated by green circles; five nodes received bootstrap support of >50% and these are shown below branches.

Taxon relationships that are highlighted according to their similarity in the *ITS* tree of Treutlein *et al.* (2005) are those of *E. altensteinii* and *E. horridus* representing group 3; *E. laurentianus* and *E. tegulaneus* representing subgroup 2.4; *E. cerinus, E. cupidus*, *E.*

inopinus and *E. manikensis* representing group 2, *E. natalensis* and *E. princeps* representing group 3 and *E. concinnus* and *E. pterogonus* representing subgroup 2.1.

3.5. Testing the monophyly of Cooper and Goode's groupings

Tree 1 represents the parsimony tree that was generated using the downloaded *ITS* sequences (Treutlein *et al.*, 2005) as described earlier in the chapter, and Tree 2 was the *ITS* tree constrained to reflect monophyletic groupings as described by Cooper and Goode (2004). A P-value of <0.05 was considered to indicate that the two trees being compared were significantly different from each other, whereby the tree with the shortest branch length and lowest –lnL score is considered to be the best hypothesis to describe relationships in the genus (Table 3.3). On this basis the monophyly of the groups as suggested by Cooper and Goode (2004) can be rejected (using both parsimony and likelihood scores).

Table 3.3: Results of the Kishino-Hasegawa tests using both parsimony and likelihood scores for two trees – the first representing the shortest *ITS* tree, and the second representing the *ITS* tree with Cooper and Goode's grouping enforced.

Phylogenetic Hypothesis	**-lnL**	**Tree Length**
ITS * (tree 1)	3484.24	106
ITS tree with Cooper & Goode's groups enforced (tree 2)	3846.73	165
P value	<0.001	<0.001

Table 3.2: Summary table of tree statistics for each of the analyses.

	# of characters	Constant characters		Variable characters		Informative characters^		# of equally parsimonious trees	Trees length	Consistency index (CI)	Retention index (RI)
		No.	~ %	No,	~ %	No.	~ %				
trnL intron	480	413	86	67	14	15	22	10 000	74	0.973	0.926
rpoC1 intron	509	486	95.4	23	4.5	3	15	7020	34	1.000	1.000
ITS 1 & 2	605	525	87	80	13	45	56	258	96	0.885	0.956
B4	384	8	2	376	98	305	81	190	1500	0.251	0.362
B13	394	20	5	374	95	307	82	2	1567	0.239	0.381
B4 + B13 combined	778	38	5	740	95	594	80	30	3181	0.233	0.303
AFLP B4, AFLP B13 + *ITS* combined	1383	558	40	825	60	658	80	3	3557	0.233	0.304

^Informative characters are calculated as a percentage of the variable characters and as a percentage of the overall characters.

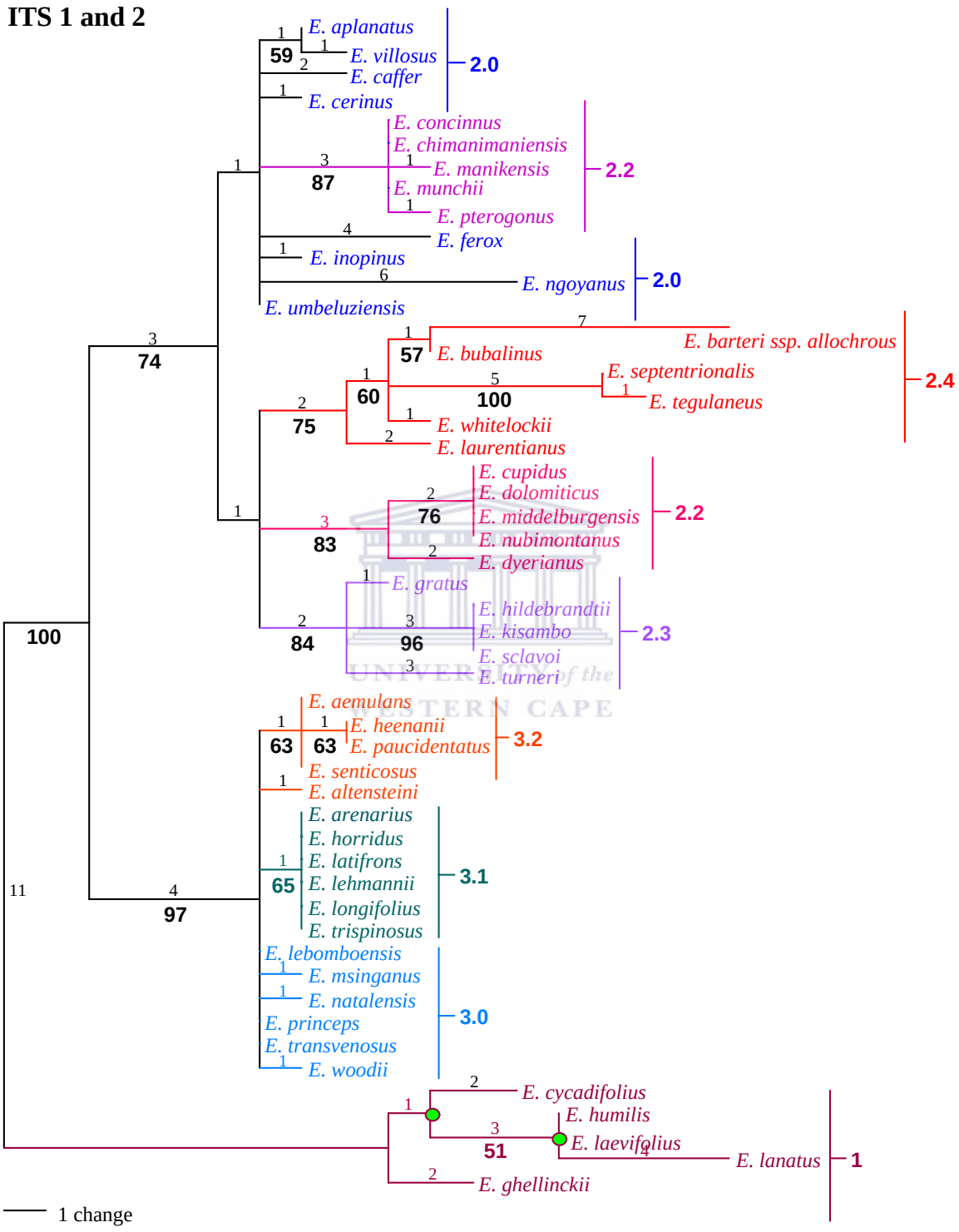

Fig. 3.1 – One of the equally most parsimonious trees found from analysis of 605 *ITS* characters. Bootstrap percentages are shown in bold beneath branches. Those nodes which collapsed in the strict consensus of 258 trees are marked by a green circle. The highlighted branches represent groups that were also found in the *ITS* tree of Treutlein *et al.* (2005).

AFLP B4

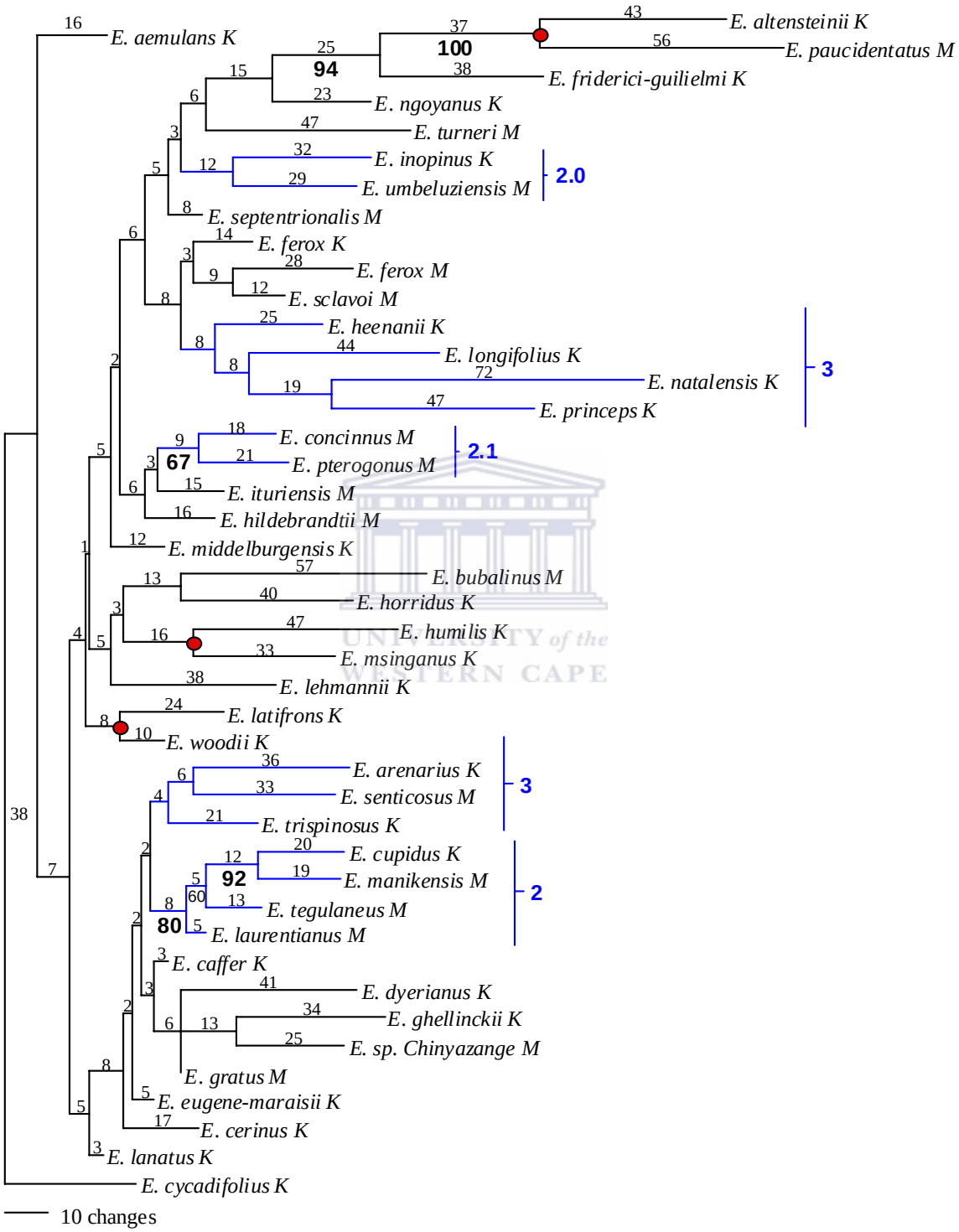

Fig. 3.2 – One of the equally most parsimonious trees found from analysis of 384 AFLP markers (using primer combination B4). Bootstrap percentages are shown in bold beneath branches. Those nodes retained in the strict consensus of 190 trees are marked by a red circle. The highlighted branches represent groups that were also found in the *ITS* tree of Treutlein *et al.* (2005).

AFLP B13

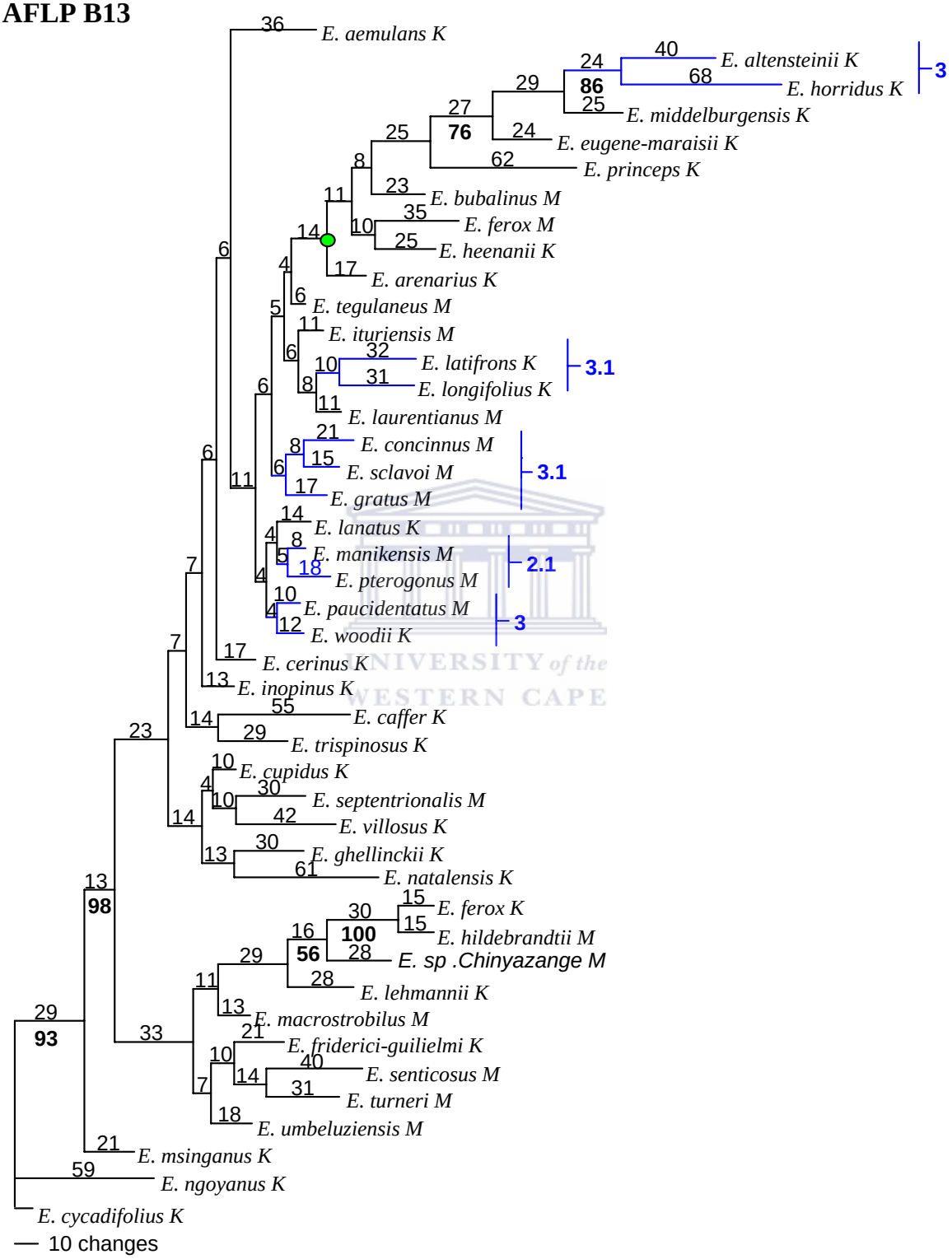

Fig. 3.3 – One of the equally most parsimonious trees found from analysis of 394 AFLP markers (using primer combination B13). Bootstrap percentages are shown in bold beneath branches. Those nodes collapsed in the strict consensus of two trees are marked by a green circle. The highlighted branches represent groups that were also found in the *ITS* tree of Treutlein *et al.* (2005).

Combined AFLP B4 and B13

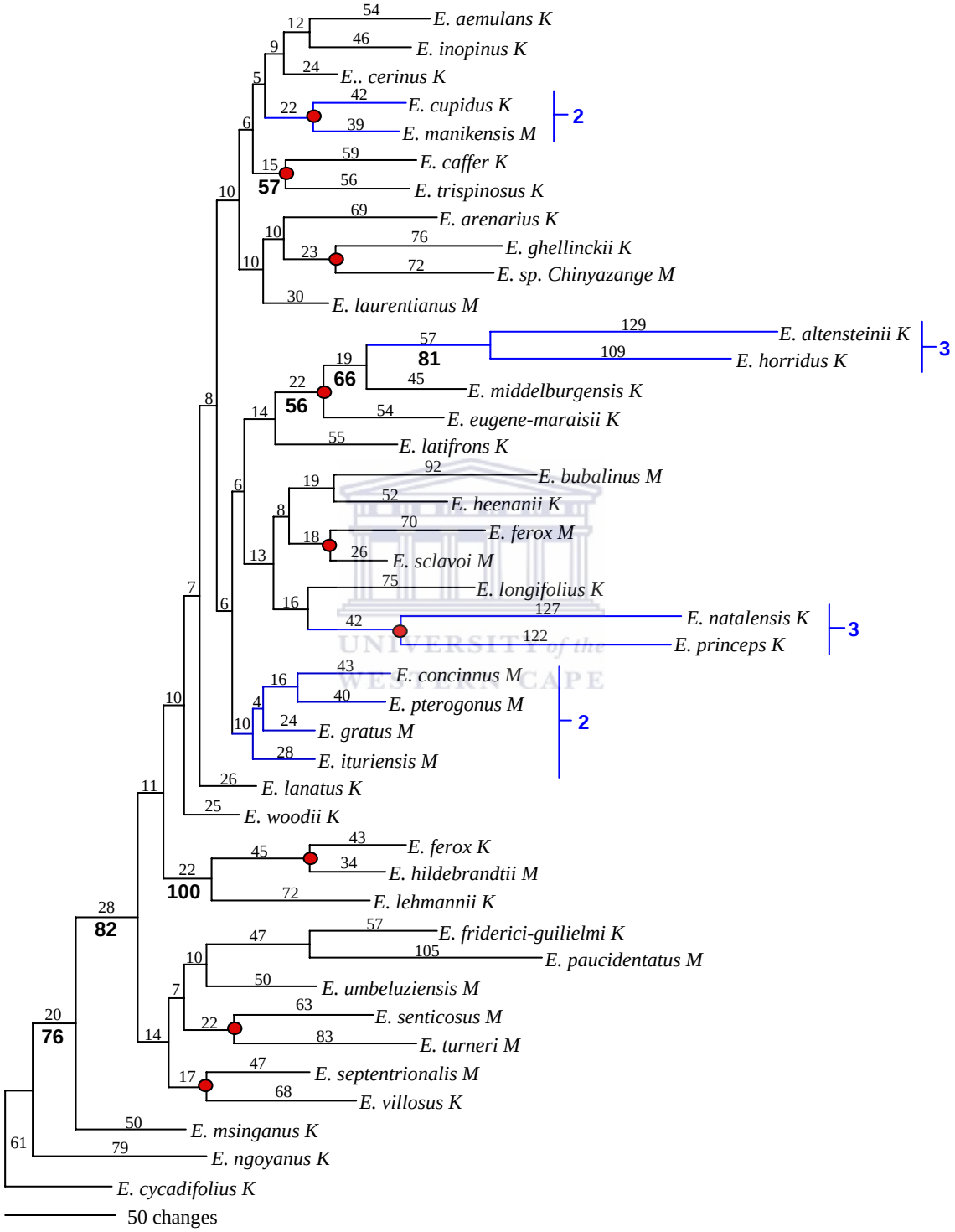

Fig. 3.4 – One of the equally most parsimonious trees found from analysis of 778 AFLP markers (using primer combinations B4 and B13). Bootstrap percentages are shown in bold beneath branches. Those nodes retained in the strict consensus of 30 trees are marked by a red circle. The highlighted branches represent groups that were also found in the *ITS* tree of Treutlein *et al.* (2005).

AFLP and ITS

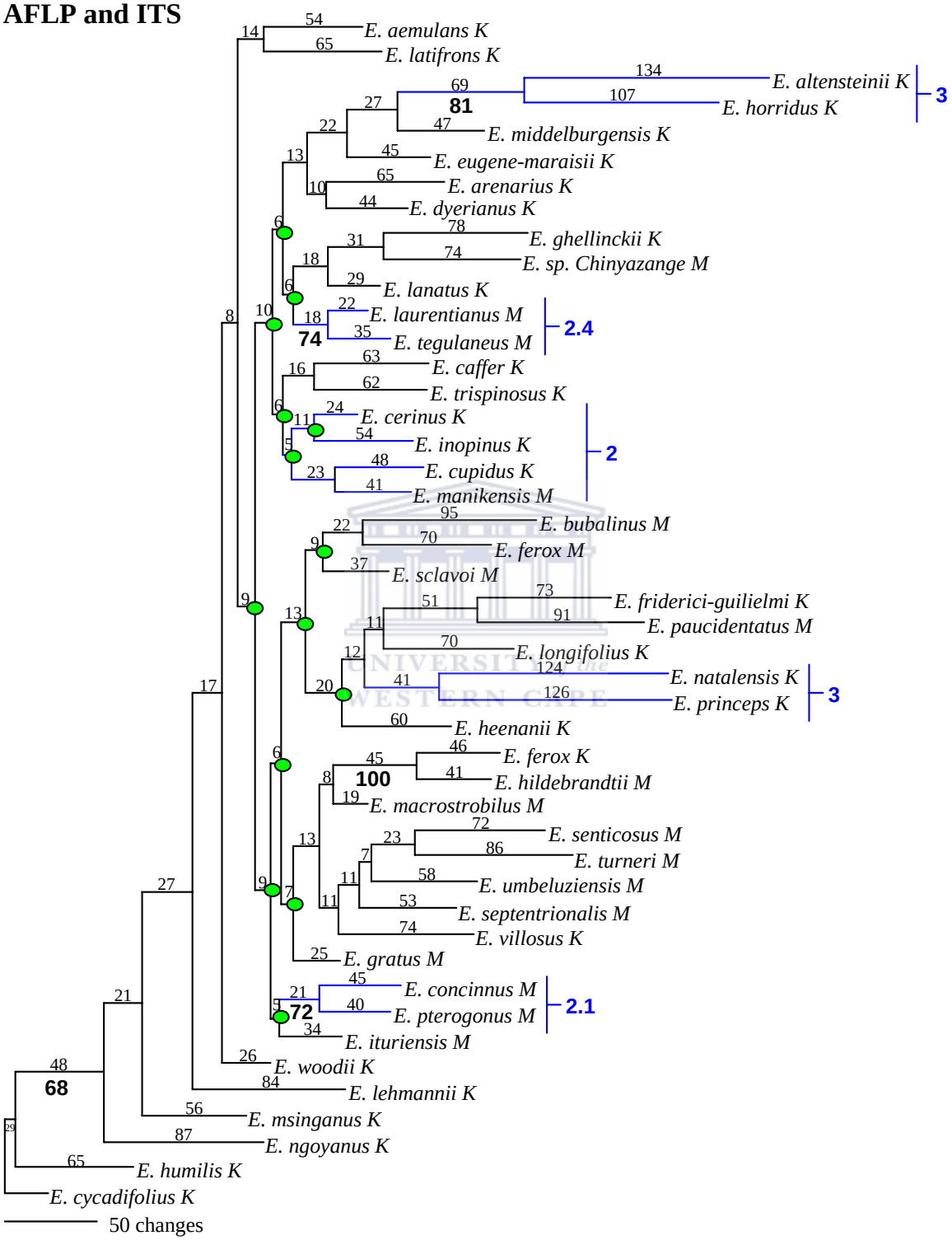

Fig. 3.5 – One of the equally most parsimonious trees found from analysis of 778 AFLP markers (using primer combinations B4 and B13) and 605 *ITS* characters. Bootstrap percentages are shown in bold beneath branches. Those nodes collapsed in the strict consensus of three trees are marked by a green circle. The highlighted branches represent groups that were also found in the *ITS* tree of Treutlein *et al.* (2005).

Habitat Optimization of *Encephalartos*

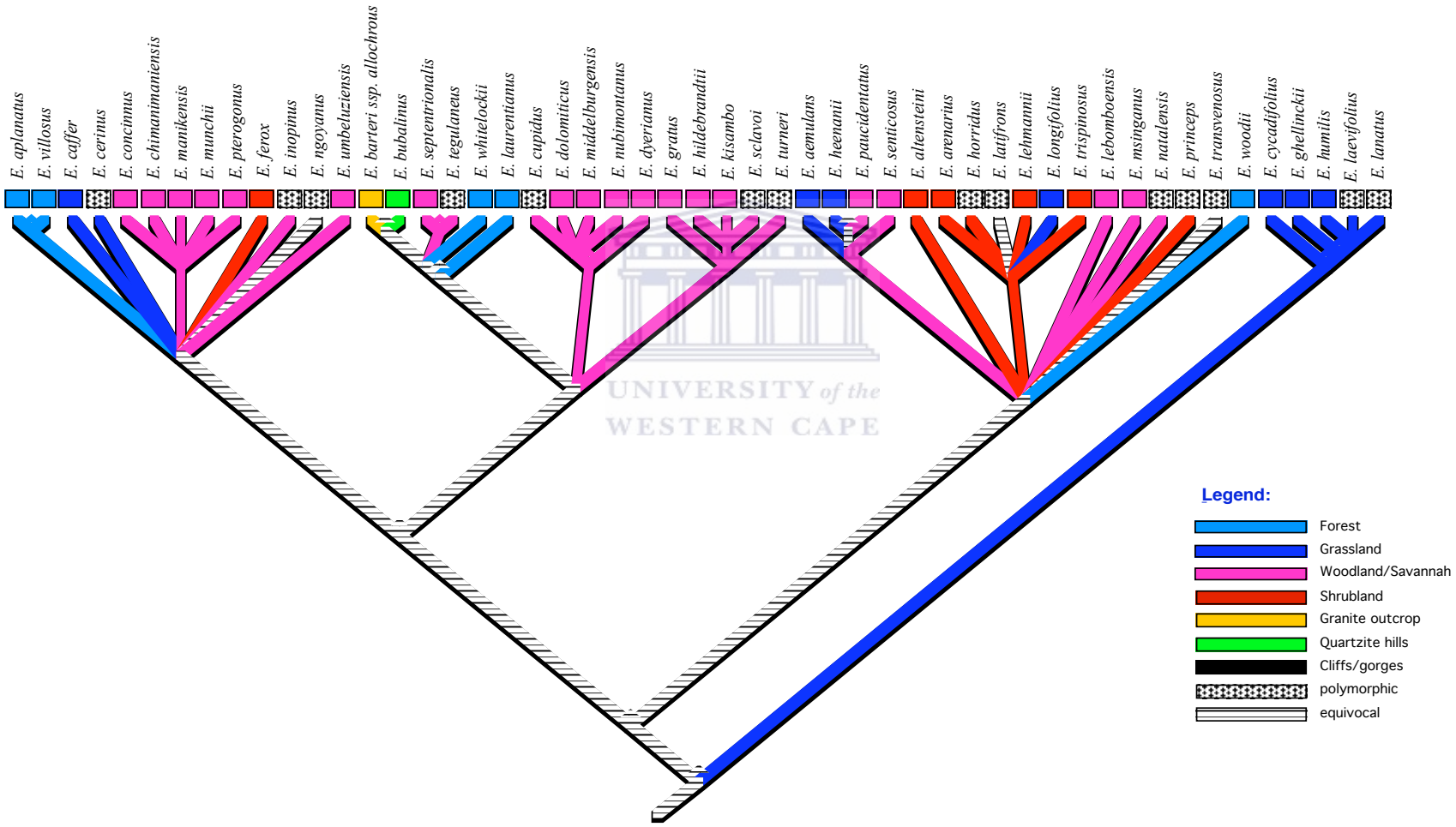

Fig. 3.6 –One of the equally most parsimonious trees found from analysis of *ITS* sequences with habitat optimized using McClade v4.01 (Appendix 2). The colors represent the different habitats as indicated on the legend. Polymorphic in this instance indicates that the species occurs in multiple habitats and equivocal indicates that the habitat for that species is uncertain.

The habitat optimization procedure illustrates that woodland/savannah habitat is occupied by the majority of species indicating that the spread into these habitats is associated with the most substantial diversification. There is a basal split in the *Encephalartos* cladogram that separates the grassland specialists, i.e. *E. cycadifolius, E. friderici-guilielmi, E. ghellinckii, E. humilis* and *E. lanatus*, from species that occur in either forests or grassland/savannah. These species seem to have diversified after they had occupied grassland habitat and there has been no secondary reversion to either forest or savannah. There are a few other grassland species (*E. aemulans, E. caffer, E. heenanii* and *E. longifolius*) but these seem to have evolved independently from the specialists with limited subsequent diversification. These results also suggest that there has been almost no diversification within forest habitats.

-CHAPTER 4-

DISCUSSION

The main aim of this study was to reconstruct evolutionary relationships between species of the genus *Encephalartos* using AFLP and sequencing data. The intention was to complement earlier studies based on morphological characters (*e.g.* Osborne *et al.,* 1993), molecular data (*e.g.* Van der Bank *et al.,* 2001) or both (*e.g.* Treutlein *et al.,* 2005), and which provided only limited resolution of evolutionary relationships. Two plastid non-coding regions, the *trnL* intron and *rpoC1* intron, were evaluated for genetic variation. The *rpoC1* intron sequences for *Encephalartos* were generated by the author as part of the DNA barcoding project and were reproduced and incorporated with permission from the funders. The sequences of both these regions contained very little genetic variation for phylogenetic analysis and, therefore, will not be discussed in isolation. The difficulty in finding DNA sequences with sufficient variability was partly anticipated at the start of this thesis and was the justification for investigating the use of AFLP data to infer species-level relationships. This discussion will compare the current results with previous phylogenetic inferences in the genus *Encephalartos*. Possible explanations for difficulties experienced in building a species-level phylogeny for *Encephalartos* will also be discussed and recommendations will be given for future studies on the genus.

Phylogenetic analysis

The overall relationships resolved by the AFLP results for both the data sets separately and combined were different to those relationships previously suggested. A conspicuous feature in the analyses was the lack of support for the internal branches in the trees. The tree topologies for the AFLP analyses had short internal and long terminal branches. Lack of internal branch support is one of the challenges in phylogenetic reconstruction as a high frequency of short internal branch lengths makes it difficult to infer relationships. Whether this pattern results from insufficient data, species extinction, or rapid radiation is often not clear (Morrison *et al.,* 2004). The AFLP data sets had high percentages of potentially phylogentically informative characters, but the CI and RI indices were

extremely low. It would be important to determine whether this is as a result of homoplasies or if it reflects true synapormorphies. The high level of potentially informative characters displayed by the AFLP data can give the impression that the data sets are comparatively information rich (Cronn *et al.*, 2002). However the low CI and RI indices rather suggest that the characters are not in agreement with respect to their phylogenetic signal, and these data do not represent a very reliable source of informative characters.

The levels of phylogenetic signal were so poor that the plastid DNA sequence data produced here were not even able to separate the chosen outgroups from the in-group. Thus the regions sampled were not even sufficiently variable to discriminate evolutionary relationships at generic level (*Dioon edule, Dioon merolae, Macrozamia riedlei* and *Zamia furfuracea* sampled as outgroups).

ITS 1 and *ITS* 2 sequences were downloaded from GenBank and added to the existing data set in order to determine whether the added sequences would influence the current results and if there would be greater resolution of phylogenetic relationships, additional support for nodes within the phylogeny, or changes to the taxa constituting these clades when compared to previous phylogenetic hypotheses. The topology of the AFLP phylogeny after addition of the *ITS* results was similar to the phylogeny before addition of these sequences in terms of short internal branches, long terminal branches as well as little bootstrap support. The relationships resolved by the addition of these sequences do not support previous phylogenetic relationships. For instance, a group of cycads from the Eastern Cape province with blue-green and strongly dentate leaflets form a distinct cluster from the rest of the group (*e.g.* Treutlein *et al.*, 2005, Van der Bank *et al.*, 2001 and Osborne *et al.*, 1993), however, in this current study this cluster is not evident from any of the analyses. The lack of sequence variability for all the DNA sequence regions may be suggestive of a recently and rapidly speciating group, diverging into new habitats. Even though hints of clades appear on all the phylograms, the different genes give different results and are all weakly supported.

Another important aspect of this study was to test the phylogeny proposed by Cooper and Goode (2004) by enforcing monophyly on the groups (Appendix 1) they suggested. This was done by enforcing their groups onto the already published *ITS* phylogeny by Treutlein *et al.* (2005) (which despite the DNA sequencing and AFLP efforts made in this study still appears to be the best available hypothesis of evolutionary relationships in the genus to date). The groups proposed by Cooper and Goode are not backed-up by detailed phylogenetic/taxonomic analyses or motives; they also provide no evidence on how they came to their conclusions for the different groups. Therefore, their groupings have raised substantial controversy and curiosity among taxonomists in terms of evaluating if the groups warrant recognition. Enforcing the monophyly on the groups suggested by Cooper and Goode onto the *ITS* phylogeny by Treutlein *et al.* (2005) showed that the new classification cannot be justified based on the available data. The groups as suggested by Cooper and Goode (2004) are not supported (using both parsimony and likelihood scores) because of the significance in the additional steps and likelihood scores of the tests performed. The values showed that that the proposed groups by Cooper and Goode represented a significant deviation compared to those recovered by Treutlein *et al.,* (2005).

The utility of molecular markers in reconstructing the phylogeny of *Encephalartos*

AFLP techniques - The key feature of this technique is the ability to assay many different DNA regions throughout the genome simultaneously (Crawford *et al.,* 2004; Després *et al.,* 2003; Mueller and Wolfenbarger, 1999 and Vos *et al.,* 1995). Amplified fragment length polymorphisms have been found to be widely distributed throughout the nuclear genome (Fay *et al.,* 2005) and this gives them several more advantages for resolving phylogenetic relationships of closely related species (Hodkinson *et al.,* 2000). This technique has also been a common alternative approach to resolve phylogenetic relationships between taxa where both cpDNA and *ITS* sequencing have failed. Several studies (Sullivan *et al.,* 2004; Allender *et al.,* 2003; Després *et al.,* 2003; Badr *et al.,* 2002; Gimmens *et al.,* 2002 Koopman *et al.,* 2001 and Albertson *et al.,* 1999) have used

AFLPs to answer various questions and have proven them to be successful, particularly among closely related species or at the intraspecific level.

For the current study, the generation of profiles for the *Encephalartos* species was challenging and inconsistent. The PCR reactions were replicated several times to evaluate if the results could be optimized to gain consistency in the quality of the profiles. Each time the reactions were optimized, the profiles were different from the previous profiles in terms of band size or the quality of the AFLP fingerprint. This made it difficult to interpret which fingerprints were the "true" ones for specific taxa. This predicament could have been solved if multiple individuals were sampled for the different species. The weak signal in the profiles may also be attributed to the genomic size of the taxa. Another factor is that the primer combinations used for *Encephalartos* did not work well for the outgroup taxa. The failure to get profiles from the outgroups could be because many of the loci were not amplified strongly enough to be scored. This was evident from the noisy baselines of the profiles which are an indication of weakly amplifying loci that do not reach the threshold for detection (Fay *et al.*, 2005).

The profiles generated for *Encephalartos* species in this study differed in terms of their signal. This may be attributed to the genomic size for some of the taxa. Suggestions have been made that genome size is likely to have an effect on multilocus DNA fingerprinting that assays the whole nuclear genome (Fay *et al.*, 2005). There is also a concern as to whether co-migrating bands are truly homologous when scoring AFLP bands (Goldman *et al.*, 2004). However, the overall errors, including mispriming and scoring error generally amount to less than 2% (Mueller and Wolfenbarger 1999). The sizing of profiles using automated methods however is said to reduce the probability of two differently sized peaks or bands being scored as one. This also reduces the concern for homology associated with AFLP data whereby it is assumed that the homology of the peaks cannot be determined. Noise (uncertain peaks) in AFLP data has also been shown to pose no problem if the species are distantly related (Goldman *et al.*, 2004). It cannot be assumed that the noise in the AFLP data did not have an affect on the results due to the fact that the species of *Encephalartos* are very closely related. Though all the profiles

were sized equally, this did not reduce the likelihood that uncertain peaks were scored as present or absent and this introduces homoplasies in the data sets, even though the affects may be minor.

DNA sequence utility – The challenge of acquiring DNA sequences that were variable enough to infer phylogenetic relationships from *Encephalartos* was evident in both the plastid *trnL* and *rpoC1* introns. While nuclear and chloroplast DNA sequences routinely provide sufficient characters to resolve higher level and generic relationships, low variation and resolution have been common problems encountered when attempting to resolve relationships among closely related plant species (Bailey *et al.,* 2004). In this study, DNA sequence variation has been the major limiting factor. Individuals of *Encephalartos* have also been sequenced as part of the barcoding project for five primers (including the *rpoC1* intron used in this study). All the primer combinations that were tested proved to be difficult in generating sequences for *Encephalartos*. Where regions have worked, the sequence variation has been too small to infer a robust phylogeny of *Encephalartos*. Therefore, it is clear from this study that a gene region that evolves at a reasonable rate for species-level phylogeny of the genus may need to be identified in order to get sufficient variation to resolve the relationships. The low evolutionary rate of chloroplast DNA sequences limits the power of these sequences for species-level phylogenetics (Després *et al.,* 2003).

Habitat optimization of *Encephalartos*

Taxon phylogenies based on molecular data can be translated into area cladograms for the analysis of relationships among geographic areas (Samuel *et al.,* 2003). In this case, habitat preferences of *Encephalartos* species were analyzed using the phylogeny derived by Treutlein *et al*. (2005). This was an attempt to understand the evolution of the genus with respect to its current distribution in different habitats and to interpret this in the context of divergence times of the different clades as suggested by previous authors (e.g. Treutlein *et al.,* 2005).

The habitat optimization procedure showed that a basal split in the *Encephalartos* cladogram separates the grassland specialists, i.e. *E. cycadifolius*, *E. friderici-guilielmi*, *E. ghellinckii*, *E. humilis* and *E. lanatus*, from species that occur in either forests or grassland/savannah. Savannah habitats are estimated to have emerged approximately 30MYA (McCarthy and Rubidge, 2005), and grasses may only have become widespread as recently as 5 MYA. A lot of climatic fluctuations and long-term shifts towards cooler and drier conditions in east and central Africa were taking place during this time, which resulted in the gradual replacement of forest habitats by savannah and woodlands (McCarthy and Rubidge, 2005). Treutlein and Wink (2002) hypothesized that the extant species of the genus spread over southern and central Africa in the late Miocene and Pliocene 1.6-5MYA. This implies that the major ancestral split in *Encephalartos* between grassland species and forest/savannah species only took place relatively recently and that diversification in these habitats may have occurred within the past 5-10 million years. The greatest diversification has occurred in the clades associated with savannah and shrublands (often arid savannah) and it seems that expansion into grassland habitat has also occasionally occurred independently in several of these lineages (*E. aemulans, E. caffer, E. longifolius and E. heenanii*). Based on the vegetation history of South Africa, and the current habitat of the related genus *Lepidozamia* (Chaw *et al.*, 2005), the ancestral habitat for *Encephalartos* is probably forests. However, whereas *Lepidozamia* still occurs in forest habitat, there is little retention of the forest habitats within *Encephalartos*. It is not known whether the reduction in forest habitat within the past 30 MYA has resulted in lineage extinctions that may explain some of the unresolved polytomies in the existing phylogenies (this study, Treutlein *et al.*, 2005).

Phylogenetic studies of cycad weevils may assist in the dating of the divergence of cycads in different habitats. A number of species complexes within *Encephalartos* were discussed by Oberprieler (1995) in conjunction with their weevil fauna and suggestions were made that this weevil fauna could aid in clarifying the taxonomic status and relationships of the species. There are 14 genera of weevils (Coleoptera: Curculionoidea) that are associated with cycads and of these only five genera are regarded as sufficiently specific to be suitable candidates for resolving relationships and species-limits among the

cycads. The oldest diversification of these weevils dates back to the Cretaceous (140MYA) when the extant cycad genera were already established (Oberprieler, 1995b). Because these insects are dependent on the existence of *Encephalartos* for survival, their diversification may be correlated with the diversification of *Encephalartos*.

The weevil genus *Platymerus* is only associated with the complex comprising of *E. cycadifolius, E. humilis, E. lanatus, E. friderici-guilielmi* and *E. ghellinckii* that appears to have split from the rest of the group at an earlier age. This *Encephalartos* complex is totally avoided by the weevil genus *Antliarhinus* (Oberprieler, 1995b). These two weevil genera may give insight into the divergence of *Encephalartos* in terms of the phylogenetic position of its "most ancient" species complex. This split is also evident in the *ITS* tree of Treutlein *et al*. (2005) and partial study by Osborne *et al.* (1993) and it would appear that the divergence of this weevil genus may be correlated with this division. The split has been dated to have occurred perhaps during the Tertiary (65-2MYA) judging from the age of the weevils (Oberprieler, 1995b). During this time, a lot of climatic fluctuations took place that may have influenced the distribution of some African plant taxa. For instance, Galley *et al.* (2006) showed that the migration of flora from the Cape to the Afrotemperate regions is congruent with the formation of the uplands in tropical Africa which dates back to the Miocene (25MYA), with further uplifts later in the Pliocene and the Pleistocene. The authors showed that because of this, there have been numerous migrations from the Cape to north of the Limpopo river. This may support the hypothesis that *Encephalartos* shows its maximum diversity in the Eastern Cape (Treutlein *et al*., 2005), and thus making this region its evolutionary center. The radiation might be correlated with the subsequent habitat changes in the Miocene. Dating the divergence times and exploring the historical biogeography of *Encephalartos* would therefore give clarity on these suggestions. The technique will also help to reveal the environmental conditions and changes that have modeled the evolutionary processes producing the present genetic structure.

Biological phenomena that may lead to conflicting phylogenetic inferences

There have been debates surrounding the combination of multiple data sets for phylogenetic inference (Soltis *et al.,* 1998). It has been recognized however that there is a benefit in analyzing multiple data sets. The application of combining multiple data sets to a common group of taxa is becoming increasingly common. One of the consequences of analyzing multiple data sets is that the phylogenies inferred may differ in certain details, and this many be incongruent with respect to their phylogenetic signal (Soltis *et al.,* 1998 and Huelsenbeck, *et al.,* 1996). When different phylogenetic estimates are in agreement with each other, there is strong probabilistic evidence supporting the phylogeny. However, this may not always be the case when the estimates disagree with each other interesting patterns of evolution may be learned from such incongruence. Pursuing the matter makes it possible to identify possible information on the biological processes that may have lead to the incongruence (McCracken *et al.,* 1999). The underlying causes of incongruence remain unknown but may include introgressive hybridization among allopatric species, rapid diversification and lineage sorting to name a few (Cronn *et al.,* 2002). Irrespective of the source of incongruence, the underlying results is that the data will fail to accurately resolve the evolutionary history of the species being studied. This section therefore suggests possible events or evolutionary phenomena that may have taken place within *Encephalartos* that have influenced its evolutionary pattern and thus the resulting conflicting tree topologies in this study.

Hybridization and introgression - It would be easy to hypothesize that the relationships illustrated by the results may be a result of hybridization within the genus. This phenomenon is common in plants and phylogenetic analyses of some plant genera have shown incongruence between data sets, suggesting that hybridization has been prevalent (Álvarez and Wendel, 2006). The term hybridization is usually employed in a broad sense to refer to crosses between genetically differentiated forms regardless of their taxonomic status (Avise, 2004). This phenomenon may have several consequences ranging from hybrids of limited adaptive value to successful introgression involving repeated backcrosses with parental genotypes (Baumel *et al.,* 2002). Plant species are often

geographically isolated from their relatives due to past vicariance or long-distance dispersal. These geographically dispersed species may have gained contact with their related species due to changes in climate or geology, opening corridors for migration. Some species may occur in geographic proximity to their relatives but be ecologically isolated from them. Under these circumstances, the opportunity for hybridizing would increase between plants that were previously isolated from each other. The likelihood for contact between previously isolated species has increased due to human activities, resulting in the breaking down of ecological barriers (Abbott *et al.*, 2003).

Although interspecific hybridization is recorded in the plant kingdom, it tends to be restricted to specific taxa (Abbott *et al.,* 2003). Natural hybrids can only occur if the two species share the same pollination vector and if their cones reach maturity at the same time (Jones, 2002). There have been reported instances of natural hybridization to date where two species occur sympatrically (Treutlein, *et al.,* 2005). A number of natural hybrids in *Encephalartos* have been recorded but they are relatively rare because most species either don't overlap or they cone at different times. One well-known hybrid is between *E. altensteinii* and *E. trispinosus*. The parent species occur together in some areas and appear to hybridize freely. Intermediate forms of the hybrids also occur, suggesting backcrossing with the parent species. There are a number of other suspected hybrids that have not been confirmed in species that occur close together (Vorster, 1987). Hybridization and introgression may influence the interpretation of phylogenetic trees (Wendel and Doyle, 1998) and their effects cannot be ruled out as one source of the conflicting results for this current study. Hybrids and introgressants may be misidentified as the true parental species due to a mixture of phenetic characteristics that make it difficult to distinguish between species. A majority of the consensus trees resulted in polytomies, possibly suggesting a mixture of genes due to possible hybridization. The separate and combined analyses of the data sets were incongruent with each other. According to Álvarez *et al.* (2006), hybridization may result in incongruence between two or more data sets for the same species. If each morphologically defined group formed a discrete clade, it would suggest that hybridization is not a frequent occurrence in this genus. However, the phylogenetic structure of the results illustrates the possibility that

hybridization may have been widespread. One way in which phylogenetic trees can be used to uncover potential hybridization events is to compare trees derived from nuclear and plastid data sources. Due to the difference in inheritance patterns in these two genomes hard incongruence can identified by well supported conflicting tree topologies. However, although hybridization is a real possibility within the genus *Encepalartos*, is has not been possible to thoroughly evaluate incongruence between the nuclear and plastid trees due to the lack of signal, particularly in the plastid data sets. Even the *ITS* tree of Treutlein *et al*. (2005) is not particularly well supported by the bootstrap, and thus it is hard to differentiate between hard incongruence and soft incongruence (i.e. differences in tree topology caused by sampling error – too few informative characters or taxa).

Rapid/recent diversification or evolution - One of the problems in phylogenetic analysis is the occurrence of short internal branches. Because short internal nodes often have weak support when measured by jackknife and bootstrap re-sampling or other indicators of relative confidence and subsequent phylogenetic analyses, using additional molecular markers often fails to yield the same topology. Short internal nodes are frequently observed in phylogenetic trees and are thought to be a common cause of misleading phylogenetic inference as well as topological incongruence among data sets. The short internal branches in the sequencing results may reflect the rapid diversification of the genus shortly after its origin (Cronn *et al.,* 2002). However, this representation of rapid radiation is often not clear (Morrison *et al.,* 2004). In this study, short internal branches followed by long terminal branches in the phylograms do indicate possible radiation at some point, which is also supported by incongruence between the different genomic compartments. This may be true since the extant lineages of *Encephalartos* occurred in the Pliocene/Pleistocene (approximately 5-20MYA) and species of *Encephalartos* show very little genetic variation, suggesting recent speciation (Treutlein *et al.,* 2005). Van der Bank *et al.* (2001) showed also in their allozyme study on nine Eastern Cape *Encephalartos* species that there may have been recent speciation within the group, which was illustrated by small genetic differences between the species.

Extinction - Species are lost by extinction and subsequent phenotypic evolution might obscure the ecological pattern of speciation events. Extinction can leave a gap on the shape of a phylogenetic tree and to obtain an accurate phylogeny, nearly all the species from the particular group need to be sampled (Barraclough and Nee, 2001). There have been recorded extinct species within *Encephalartos*, but there are also possibly more species that have not been collected from the war stricken countries of Africa, and which may be the missing link between the long and short branches. These may help to resolve relationships within the genus once sampled.

Different genomic compartments - The possibility of obtaining the same or even similar tree topologies for the datasets is very small for any given number of species. If they give similar topologies, it shows that they reflect the same underlying evolutionary history (Page and Holmes, 1998). Combining data sets from different phylogenetic analyses and genomic compartments is a very common approach to increase phylogenetic signal. The disadvantage however lies in that different genomic compartments evolve at different rates and therefore have differential patterns of inheritance. Incongruence between the data sets has therefore been the cause of inaccurately estimated phylogenetic hypotheses. Several authors have however suggested that combining the data sets irrespective of the incongruence will allow those clades for which there is congruence to gain increased support (Goldman *et al.*, 2004). This however cannot be said for this current study. The combined analysis of the sequencing and AFLP data had no significant support for the nodes in the analyses; the relationships resolved by the study also do not support previous phylogenetic relationships of the genus.

Conclusions

Species-level phylogenies are important and necessary for understanding the evolution of particular groups and for interpreting and understanding ecological adaptations. However, this may be impossible if the relationships of the species are not fully resolved. Systematists generally seek fully resolved trees, which may yield stronger inferences about character evolution or biogeographic history and usually view polytomies a

reflecting uncertainty about relationships (Hewitt, 2004). The results of this study are a significant contribution to the attempts in resolving the phylogenetic relationships of *Encephalartos*. This study is the first to use AFLPs together with DNA sequence data in an attempt to resolve the relationships within *Encephalartos* and to reduce the uncertainty that remains from previous molecular and morphological studies. Despite the positive results obtained using AFLPs in other groups, the data obtained from *Encephalartos* were inconsistent and incongruent with DNA sequence data. The species level relationships within *Encephalartos* therefore remain problematic and unresolved. These results show that even though there was sufficient signal for analysis, there may be incomplete lineage sorting within the group due to recent speciation. This phenomenon of incomplete lineage sorting does pose a challenge to the phylogenetic reconstruction of recently derived species. Further phylogenetic studies of *Encephalartos* should include the combination of molecular and morphological data with additional DNA sequence data that has sufficient variability for phylogenetic analysis. This may provide greater resolution and clarity on the phylogenetic relationships.

LITERATURE CITED

Abbott, R. J., James, J. K., Milne, R. I. and Gillies, A. C. M. 2003. Plant introductions, hybridization and gene flow. Philosophical Transactions of the Royal Society of London. 358: 1123-1132.

Ahmed, S. M., Verma, V., Qazi, P. H. Ganaie, M. M., Bakshi, S. K. and Qazi, G. N. 2006. Molecular phylogeny in Indian *Tinospora* species by DNA based molecular markers. Plant Systematics and Evolution. 256: 75-87.

Albertson, R. C., Markert, P. D. and Kocher, T. D. 1999. Phylogeny of a rapidly evolving clade: The cichlid fishes of Lake Malawi, East Africa. Proceedings of the National Academy of Science. 96: 5107-5110.

Allender, C. J., Seehausen, O., Knight, Turner, G. F. and Maclean, N. 2003. Divergent selection during speciation of Lake Malawi cichlid fishes inferred from parallel radiations in natural coloration. Proceedings of the National Academy of Science. 100: 14074-14079.

Álvarez, I. and Wendel, J. F. Cryptic interspecific introgression and genetic differentiation within Gossypium aridum (Malvaceae) and its relatives. Evolution. 60: 506-517.

Avise, J. C. 2004. Molecular Markers, Natural History and Evolution. 2nd Edition. Sinauer Associates, Inc. Publishers, USA.

Badr, A., Shazly, H., Rabey, H. and Watson, L. E. 2002. Systematic relationships in *Lathyrus* sect. *Lathyrus* (Fabaceae) based on amplified fragment length polymorphism (AFLP) data. Canadian Journal of Botany. 80: 962-969.

Bailey, C. D., Hughes, C. E. and Harris, S. A. 2004. Using RAPDs to identify sequence loci for species level phylogeny reconstruction: An example from *Leucaena* (Fabaceae). Systematic Botany. 29: 4-14.

Bailey, C. D., Carr, T. G., Harris, S. A. and Hughes, C. E. 2003. Characterizing of angiosperm nrDNA paralogy and pseudogenes. Molecular Phylogenetics and Evolution 23: 435-455.

Baker, A. J., Huynen, L. J., Haddrath, O., Millar, C. D. and Lambert, D. M. 2005. Reconstructing the tempo and mode of evolution in an extinct clade of birds with ancient DNA: the giant moas of New Zealand. Proceedings of the National Academy of Science. 102: 8257-8262.

Barraclough, T. G. and Nee, S. 2001. Phylogenetics and speciation. Trends in Ecology and Evolution. 16: 391-399.

Baumel, A., Ainouche, M. L., Bayer, R. J., Ainouche, A. K. and Misset, M. T. 2002. Molecular phylogeny of hybridizing species from the Genus *Spartina* Schreb. (Poaceae). Molecular Phylogenetics and Evolution. 22: 303-314.

Bogler, D. J. and Francisco-Ortega, J. 2004. Molecular systematic studies in cycads, evidence from *trn*L and ITS2 rDNA sequences. The Botanical Review. 70: 260-273.

Brenner, D. E., Stevenson, W. and Twigg, R. W. 2003. Cycads: Evolutionary innovations and the role of plant-derived neurotoxins. Trends in Plant science. 8: 446-452.

Cáceres, L. J. M. S. “The Cycads: Fossils of the Past.” Palm and Cycad Societies of Florida, Inc. 1998.
<http://www.plantapalm.com/vce/intro/fossilspast.htm>(2003)

Campbell, S. C., Judd, S., Kellogg, E. A. and Stevens, F. P. 1999. Plant Systematics: A Phylogenetic Approach. Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts, USA.

Caputo, P., Masquiis, C., Wurtzel, T., Stevenson, D. W. and Wurtzel, E. L. 1993. Molecular Biology in Cycad Systematics. In: The biology, Structure and Systematics of the Cycadales. Proceedings of Cycad 90, the Second International Conference on cycad biology (Eds. Stevenson, D.W., Norstog, K. J. 1993.). Palm & Cycad Societies of Australia Ltd. p213-219.

Caputo, P., Stevenson, D. W. and Wurtzel, T. E. 1991. A phylogenetic analysis of American Zamiaceae (Cycadales) using chloroplast DNA restriction fragment length polymorphisms. Brittonia. 43: 135-145.

Chaw, S-M., Walters, T. W., Chang, C-C., Hu, S-H. and Chen, S-H. 2005. A phylogeny of cycads (Cycadales) inferred from chloroplast matK gene, trnK, and nuclear rDNA ITS region. Molecular Phylogenetic and Evolution. 37: 214-234.

Cooper, M. R. and Goode, D. 2004. The cycads and cycad moths of Kwa-Zulu Natal. Peroniceras Press, New Germany.

Crawford, D. J. and Mort, M. E. 2004. Single-locus markers for inferring relationships at lower taxonomic levels: observation and comments. Taxon. 53: 631-635.

Cronn, R. C., Small, R. L., Haselkorn, T. and Wendel**,** F. J. 2002. Rapid diversification of the cotton genus (*Gossypium*: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. American Journal of Botany. 89: 707-725.

Da Silva, J. 2005. DNA fingerprinting of the critically endangered cycad, *Encephalartos latifrons*, using AFLP markers. Unpublished MSc. thesis.

de Laubenfels, D. J. 1999. The families of Cycadaceae. *Encephalartos*. 59: 7-9.

deMenocal, P.B. 2004. African climate change and faunal evolution during the Pliocene-Pleistocene. Earth and Planetary Science Letters. 220: 3-24.

Després, L., Gielly, L., Redoute, B. and Taberlet, P. 2003. Using AFLP to resolve phylogenetic relationships in a morphologically diversified species complex when nuclear and chloroplast sequences fail to reveal variability. Molecular Phylogenetics and Evolution. 27: 185-196.

Donaldson, J. S. 2004. Extinction of cycad pollinators - Do generalists or specialists survive cycad decline? In: The Biology, Structure and Systematics of the Cycadales, Proceedings of the Sixth International Conference on Cycad Biology (Ed. Lindstrom A. J.), Thailand. International Academic Publishers. p154.

Donaldson, J.S., Hill, K. D. and Stevenson, D.W. 2003. Cycads of the World: An Overview. In: Cycads (Ed. Donaldson, J. S.). IUCN - The World conservation Union. Charlesworth & Co Ltd., Huddersfield, UK. p3-8.

Doyle, J. J. and Doyle, J. L. 1987. A rapid isolation procedure from small quantities of fresh leaf tissue. Phytochemical Bulletin Society of America. 19: 11-15.

Duftner, N., Sefc, K. M., Koblmüller, S., Nevado, B., Verheyen, E., Phiri§ H. and Sturmbauer, C. 2006. Distinct population structure in a phenotypically homogeneous rock-dwelling cichlid fish from Lake Tanganyika. Molecular Ecology. 15: 2381-2395.

Dyer, R. A. 1965b. The cycads of Southern Africa. Bothalia. 8: 405-515.

Fay, M. F., Cowan, R. S. and Leitch, I. J. 2005. The effects of nuclear DNA content (C-value) on the quality and utility of AFLP fingerprints. Annals of Botany. 95: 237-246.

Federici, M. T., Vaughan, D., Tomooka, N., Kaga, A., Wang, X. W., Doi, K., Frances, M., Zomila, G. and Saldain, N. 2001. Analysis of Uruguayan weedy rice genetic diversity using AFLP molecular markers. Electronic Journal of Biotechnology. 4:1-16.

Farris, J. S., Källersjo, M., Kluge, A.G. and bult, C. 1995. Testing significance of incongruence. Cladistics. 10: 315-319.

Galley, C., Bytebier, B., Bellstedt, D. U. and Linder, H. P. 2006. The Cape element in the Afrotemperate flora: from Cape to Cairo? Proceedings of the Royal Society Biological Sciences. 274: 535-544.

Gimmens, M. A., Lpoes, C. R. and Valls, J. F. S. 2002. Genetic relationships among *Arachis* species based on AFLP. Genetics and Molecular Biology. 25: 349-353.

Goldman, D. H., Jansen, R. K., Van den Berg, C., Leitch, I. J., Fay, M. F. and Chase, M. W. 2004. Molecular and cytological examination of *Calopogon* (Orchidaceae, Epidendroideae): circumscription, phylogeny, polyploidy, and possible hybrid speciation. American Journal of Botany. 91: 707-723.

González-Astorga, J., Vovides, P. A. and Iglesia, C. 2003. Morphological and geographic variation of the cycad Dioon edule Lindl. (Zamiaceae): ecological and evolutionary implications. Botanical Journal of the Linnean Society. 141: 465-470.

Goode, D. 2001. Cycads of Africa. Volume 1. All New Edition. D & E Cycads of Africa Publishers.

Goode, D. 1989. Cycads of Africa. Striuk Winchester.

Hermsen, E. J., Taylor, T. N., Taylor, E. L. and Stevenson, D. W. 2006. Cataphylls of the middle Triassic cycad *Antarcticycas schopfii* and the new insights into cycad evolution. American Journal of Botany. 93: 724-738.

Hill, K. D., Stevenson, D. W. and Osborne, R. 2004. The world list of cycads. In: The Biology, Structure and Systematics of the Cycadales, Proceedings of the Sixth International Conference on Cycad Biology (Ed. Lindstrom A. J.), Thailand. International Academic Publishers. p195-212.

Hill, K.D., Chase, M. W., Stevenson, H. G. and Schutzman, B. 2003. The families and genera of cycads: A molecular phylogenetic analysis of Cycadophyta based on nuclear and plastid DNA sequences. International Journal of Plant Sciences. 164: 933-948.

Hill, K. 1999. *Cycas* - an evolutionary perspective. In: Biology and conservation of cycads. Proceedings of the Fourth International Conference on Cycad Biology (Ed. Chen, C-J). International Academic Publishers. p98-115.

Hodkinson, T. R., Renvoize, S. T., Chonghaile, G. N., Stapelton, C. M. A. and Chase, M. W. 2000. A comparison of ITS nuclear DNA sequence data and AFLP markers for phylogenetic studies in *Phyllostachys* (Bambusoideae, Poaceae). Journal of Plant Research. 113: 259-269.

Huelsenbeck, J. P., Bull, J. J. and Cunningham, C. W. 1996. Combining data in phylogenetic analysis. Trends in Ecology and Evolution. 11: 152-158.

Jones, D. L. 2002. Cycads of the World: Ancient Plants in Today's Landscape. 2nd Edition. Smithsonian Institution Press, Washington, D. C.

Jones, D. L. 1993. Cycads of the World: Ancient Plants in Today's Landscape. Reed, Chatswood NSW, Australia

Jousselin, E., Rasplus, J-Y. and Kjellberg, F. 2003. Convergence and Coevolution in a mutualism: Evidence from a molecular phylogeny of *Ficus*. Evolution. 57: 1255-1269.

Kishino, H. and M. Hasegawa. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. Journal of Molecular Evolution. 29: 170-179

Klak, C., Khunou, A., Reeves, G. and Hedderson, T. 2003. A phylogenetic hypothesis for the Aizoaceae (Caryophyllales) based on four plastid DNA regions. American Journal of Botany. 90: 1433-1445.

Koopman, W. J. M., Zevenbergen, M. J. and Van den Berg, R. G. 2001. Species relationships in *Lactuca* S. L. (Lactucaseae: Asteraceae) inferred from AFLP fingerprints. American Journal of Botany. 88: 1881-1887.

Lin, T., Lin, Y. and Ishiki, K. 2004. Genetic diversity of *Dimocarpus longan* in China revealed by AFLP markers and partial *rbcL* gene sequences. Scientia Horticulturae. 103 (4): 489-498.

Lee, C. and Wen, J. 2004. Phylogeny of *Panax* using chloroplast *trnC-trnD* intergenic region and the utility of *trnC-trnD* in interspecific studies of plants. Molecular Phylogenetics and Evolution. 31: 894-903.

Maddison, D. R. and Maddison, W. P. 2001. MacClade v4.01. Sinauer Associates Inc., Sunderland, Massachusetts.

McCarthy, T. and Rubidge, B. 2005. The story of Earth and Life: A southern African perspective on a 4.6-billion-year history. Struik Publishers, Cape Town.

McCracken, G. K., Harshman, J., McClellan, G. A. and Afton, A. D. 1999. Data set incongruence and correlated character evolution: An example of functional convergence in the Hind-Limbs of Stifftail diving ducks. Systematic Biology. 48: 683-714.

Mekanawakul, M., Thongtham, and Juntawong, N. 2004. Diversity and genetic variation among population of Thai cycads revealed by AFLP markers. In: The Biology, Structure and Systematics of the Cycadales, Proceedings of the Sixth International Conference on Cycad Biology (Ed. Lindstrom A. J.), Thailand. International Academic Publishers. p92-97.

Michaloud, G., Bossu-Dupriez, N., Chevolot, M. and Lasbleiz, C. 2005. Pollen waste and unrelated traits in fig-fig wasp symbiosis: a new behavior suggesting a host shift. Evolution. 328: 81-87.

Morrison, C. L., Rios, R. and Duffy, J.E. 2004. Phylogenetic evidence for an ancient rapid radiation of Caribbean sponge-dwelling snapping shrimps (S*ynalpheus*). Molecular Phylogenetics and Evolution. 30: 563.581.

Mueller, U. G. and Wolfenbarger, L. l. 1999. AFLP genotyping and fingerprinting. Trends in Ecology and Evolution. 14: 389-394.

Muluvi, G. M., Sprent, J. I., Soranzo, N., Provan, J., Odee, D., Folkard, G., McNicol, W. J. and Powell, W. 1999. Amplified fragment length polymorphism (AFLP) analysis of genetic variation in *Moringa oleifera* Lam. Molecular Ecology. 8: 463-470.

Norstog, K. J. and Nicholls, T. J. 1997. The Biology of the Cycads. Ithaca, USA: Cornell University Press.

Norstog, K. J. and Fawcetti, P. K. S. 1995. Insect-pollination of cycads: evolutionary and ecological considerations. In: Proceedings of the Third International Conference on Cycad Biology (Ed. Voster, P.). Cycad Society of South Africa, Stellenbosch. p265-285.

Oberprieler, R. G. 2004. “Evil weevils”- the key to cycad survival and diversification? In: The Biology, Structure and Systematics of the Cycadales, Proceedings of the Sixth

International Conference on Cycad Biology (Ed. Lindstrom A. J.), Thailand. International Academic Publishers. p170-194.

Oberprieler, R. G. 1995a. The weevils (Coleoptera: Curculinoidea) associated with cycads. 1. Classification, relationships and biology. In: Proceedings of the Third International Conference on Cycad Biology (Ed. Voster, P.). Cycad Society of South Africa, Stellenbosch. p295-334.

Oberprieler, R. G. 1995b. The weevils (Coleoptera: Curculinoidea) associated with cycads. 2. Host specificity and implications for cycad taxonomy. In: Proceedings of the Third International Conference on Cycad Biology (Ed. Voster, P.). Cycad Society of South Africa, Stellenbosch. p335-365.

Ogden, R. and Thorpe, R. S. 2002. The usefulness of amplified fragment length polymorphism markers for taxon discrimination across graduated fine evolutionary levels in Caribbean *Anolis* lizards. Molecular Ecology. 11: 437-445.

Osborne, R. 2002. Cycad fossils. Encephalartos. 69: 4-13.

Osborne, R., Grobbelaar, N. and Voster, P. 1998. South African cycad research: progress and prospects. South African Journal of Science. 84:891-896.

Osborne, R., Grobbelaar, N. and Vincent, P. L. D. 1993. A numerical phenetic study of the genus *Encephalartos lehm.* In: The Biology, Structure and Systematics of the Cycadales, Proceedings of CYCAD 90, the Second International Conference on Cycad Biology (ed. Stevenson, D. W. and Norstog, K. J.). Palm and Cycad Societies of Australia Ltd. p279-280.

Page, R. D. M. and Holmes, E. C. 1998. Molecular evolution-a phylogenetic approach. Blackwell Science Pty Ltd.

Palacios, C., Kresovich, S. and González-Candelas, F. 1999. A population genetic study of the endangered plant species *Limonium dufourii* (*Plumbaginaceae*) based on amplified fragment length polymorphism (AFLP). Molecular Ecology. 8:645-657.

Parsons, Y. M. and Shaw, K. L. 2001. Species boundaries and genetic diversity among Hawaiian crickets of the genus *Laupala* identified using amplified fragment length polymorphism. Molecular Ecology. 10: 1765-1772.

Pelser, P. B., Gravendeel, B. and Van Der Meijden, R. 2002. Tackling speciose genera: species composition and phylogenetic position of *Senecio* Sect. *Jacobaea* (Asteraceae) based on plastid and nrDNA sequences. American Journal of Botany. 89: 929-939.

Rai, S. H., O'Brien, H. E., Reeves, P. A., Olmstead, R. G. and Graham, S. W. 2003. Inference of higher-order relationships in the cycads from a large chloroplast data set. Molecular Phylogenetics and Evolution. 29: 350-359.

Reeves, G., Chase, M. W., Goldblatt, P., Rudall, P., Fay, M. F., Cox, A. V., Lejeune, B. and Souza-Chies, T. 2001. Molecular systematics of Iridaceae: Evidence from four plastid DNA regions. America Journal of Botany. 88: 2074-2087.

Richardson, J. E., Weitz, F. M., Fay, M. F., Cronk, Q. C. B., Linder, H. P., Reeves, G. and Chase, M. W. 2001. Phylogenetic analysis of *Phylica* L. (*Rhamnaceae*) with an emphasis on island species: evidence from plastid *trnL-F* and nuclear internal transcribed spacer (ribosomal) DNA sequences. Taxon. 50: 405-427.

Samuel, R., Stuessy, T. F., Tremetsberger, K., Baeza, C. M. and Siljak-Yakovlev, S. 2003. Phylogenetic relationships among species of *Hypochaeris* (Asteraceae, Cichorieae) based on *ITS*, plastid *trnL* intron, *trnL-F* spacer and *matK* sequences. American Journal of Botany. 90: 496-507.

Savolainen, V., Powell, M. P., Davis, K., Reeves, G. and Corthals, A. 2006. DNA and Tissue Banking for Biodiversity and Conservation: Theory, Practice and Uses. Kew Publishing, Royal Botanic Gardens, Kew.

Savolainen V., Cuenoud, P., Spichiger R., Martinez, M. D. P., Crevecoeur, M. and Manen, J-F. 1995. The use of herbarium specimens in DNA phylogenetics: evaluation and improvement. Plant Systematics and Evolution. 197: 87-98.

Schutzman, B. and Dehgan, B. 1993. Computer-assisted systematics in the Cycadales. In: The Biology, Structure and Systematics of the Cycadales, Proceedings of CYCAD 90, the Second International Conference on Cycad Biology (ed. Stevenson, D. W. and Norstog, K. J.). Palm and Cycad Societies of Australia Ltd. p281-289.

Seehausen, O. 2004. Hybridization and adaptive radiation. Trends in Ecology and Evolution. 19: 198-207.

Shaw, J. and Small, R. L. 2004. Addressing the “Hardest puzzle in American pomology:” phylogeny of *Prunus* Sect. *Prunocerasus* (Rosaceae) based on seven noncoding chloroplast DNA regions. American Journal of Botany. 91: 985-996.

Simões, A. O., Endress, M. E., Van Der Niet, T. Kinoshita, L. S. and Conti, E. 2004. Tribal and intergenic relationships of Mesechiteae (Apocynoideae, Apocynaceae): evidence from three noncoding plastid DNA regions and morphology. American Journal of Botany. 91: 1409-1418.

Soltis, D. E., Soltis, P. S. and Doyle, J. J. 1998. Molecular systematics of plants II: DNA sequencing. Kluwer Academic Publishers, Boston.

Sullivan, J. P., Lavoué, S., Arnegard, M. E. and Hopkins, C. D. 2004. AFLPs resolve phylogeny and reveal mitochondrial introgression within a species flock of African electric fish (Mormyroidea: Teleostei). Evolution. 58: 825-841.

Stevenson, D. W. 1992. A formal classification of the extant cycads. Brittonia. 44(2): 220-223.

Swofford, D. L. 2000. PAUP*4.0. Phylogenetic Analysis using Parsimony (*and other methods). Sinauer Associates, Sunderland, Massachusetts.

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

Tolley, K. A., Burger, M., Turner, A. A. and Matthee, C. A. 2006. Biogeographic patterns and phylogeography of dwarf chameleons (*Bradypodion*) in an African biodiversity hotspot. Molecular Ecology. 15: 781-793.

Treutlein, J., Voster, P. and Wink, M. 2005. Molecular Relationships in Encephalartos (Zamiaceae, Cycadales) based on nucleotide sequences of nuclear ITS 1 and 2, Rbcl and genomic ISSR Fingerprinting. Plant Biology. 7:79-90.

Treutlein, J. and Wink, M. 2002. Molecular phylogeny of cycads inferred from rbcL sequences. Naturwissenschaften. 89: 221-225.

Tsai, C. C., Huang, S. C. and Chou, C. H. 2006. Molecular phylogeny of *Phalaenopsis* Blume (Orchidaceae) based on the internal transcriber spacer of the nuclear ribosomal DNA. Plant Systematics and Evolution. 256:1-16.

Van der Bank, H., Wink, M., Voster, P., Treutlein, J., Brand, L., van der Bank, M. and Hurtler, J. 2001. Alloenzyme and DNA sequence comparisons of nine species of *Encephalartos* (Zamiacecae). Biochemical Systematics and Ecology. 29: 241-266.

Van der Wurff, A. W. G., Chan, Y. L., van Straalen, N. M. and Schouten, J. 2000. TE-AFLP: combining rapidity and robustness in DNA fingerprinting. Nucleic Acid Research. 28: 1-5.

Vorster, P. 2005. Book Review: Cooper, M. R. and Goode, D-The cycads and cycad moths of Kwa-Zulu Natal. *Encephalartos*. 80: 10-14.

Vorster, P. 1999. A review of the arborescent tropical species of *Encephalartos*. In: Biology and conservation of cycads. Proceedings of the Fourth International Conference on Cycad Biology (ed. Chen, C-J). International Academic Publishers. p87-91.

Vorster, 1995. Aspects of the reproduction of cycads. 1. Pollinating mechanisms and the association of *Amorphocerus* (Curculionidae) with *Encephalartos*. In: Proceedings of the Third International Conference on Cycad Biology (ed. Voster, P.). Cycad Society of South Africa, Stellenbosch. p367-389.

Vorster, P. 1993. Taxonomy of Encephalartos (Zamiaceae): Taxonomically useful External Characteristics. In: The Biology, Structure and Systematics of the Cycadales, Proceedings of CYCAD 90, the Second International Conference on Cycad Biology (ed. Stevenson, D. W. and Norstog, K. J.). Palm and Cycad Societies of Australia Ltd. p294-299.

Vorster, P. 1987. Hybridization in *Encephalartos*. Encephalartos. 10: 10-15.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. and Zabeau, M. (1995). AFLP: A new technique for DNA fingerprinting. Nucleic Acids Research. 23: 4407-4114.

Weins, J. J. and Penkrot, T. A. 2002. Delimiting species using DNA and morphological variation and discordant species limits in spiny Lizards (*Sceloporus*). Systematic Biology. 51: 69-91.

Wendel, J. F. and Doyle, J. J. 1998. Phylogenetic incongruence: Window into genome history and molecular evolution. In: Molecular systematics of plants II, DNA sequencing (ed. Soltis, D. E., Soltis, P. S. and Doyle, J. J.). Kluwer Academic Publisers, London. p265-296.

APPENDIX 1

Taxa included in this study, along with their accession numbers from both the National Botanical Garden in Kirstenbosch (KBG) and the Montgomery Botanical Center (MBG). *ITS* 1 & 2 sequences imported from Genbank are also given with their Genbank accession numbers and herbarium numbers. DNA regions that were sequenced for each taxon and AFLP profiles generated are indicated by a tick.

Taxon	Collection accession numbers	Genbank accession numbers for *ITS* sequences		*trnL* intron	Partial *rpoC1* intron	AFLP markers	
		ITS 1	*ITS* 2		*1F*	*B4*	*B13*
Dioon edule Lindl	-	-	-	✓	-	-	-
Dioon merolae De Luca, Sabato & Vázq. Torres	KBG Cons. N**	-	-	✓	-	-	-
E. aemulans Voster	KBG 108/1996	-	-	✓	-	✓	✓
E. aemulans Voster	031001	AY335321	AF394427	-	-	-	-
E. aplanatus Voster	031003	AY335266	AF394383	-	-	-	-
E. altenstenii Lehmann	KBG 2497/1916	-	-	✓	-	✓	✓
E. altenstenii Lehmann	031002	AY335319	AF394424	-	-	-	-
E. arenarius R. A. Dyer	KBG 132/1983	-	-	✓	-	✓	✓
E. arenarius R. A. Dyer	031004	AY335314	AF394416	-	-	-	-
E. barteri sp. allocchrous L. E. Newton	031005	AY335310	AF394412	-	-	-	-
E. bubalinus Mellville	MBC 95914*C	-	-	✓	-	✓	✓
E. bubalinus Mellville	031006	AY335309	AF394411	-	-	-	-
E. caffer (Thunb.)Lehmann	KBG 1725/1914	-	-	✓	Partial	✓	✓
E. caffer (Thunb.)Lehmann	031007	AY335296	AF394388	-	-	-	-
E. cerinus Lavranos and Goode	KBG 356/1994	-	-	✓	✓	✓	✓
E. cerinus Lavranos and Goode	031008	AY335288	AF394379	-	-	-	-
E. chimanimaniensis R. A. Dyer & I. Verdoorn	031009	AY335299	AF394393	-	-	-	-
E. concinnus R. A. Dyer & I. Verdoorn	MBC 20011086*A	-	-	✓	-	✓	✓
E. concinnus R. A. Dyer & I. Verdoorn	031010	AY335300	AF394394	-	-	-	-
E. cupidus R. A. Dyer	KBG 705/1997	-	-	✓		✓	✓
E. cupidus R. A. Dyer	031011	AY335277	AF394395	-	-	-	-
E. cyacadifolius (seedling) (Jacquin) Lehm.	KBG (Cons. N**)	-	-	✓	Partial	✓	✓

Taxon	Collection accession numbers	Genbank accession numbers for *ITS* sequences		*trnL* intron	Partial *rpoC1* intron	AFLP markers	
		ITS 1	*ITS* 2		*1F*	*B4*	*B13*
E. cyacadifolius (Jacquin) Lehm	031012	AY335274	AF394374	-	-	-	-
E. dolomiticus Lavranos & D. L. Goode	031013	AY335301	AF394396	-	-	-	-
E. dyerianus Lavranos & D. L. Goode	KBG 695/1997	-	-	✓	✓	✓	-
E. dyerianus Lavranos & D. L. Goode	031014	AY335278	AF394399	-	-	-	-
E. eugene-maraisii I. Verdoorn	KBG 4/1932	-	-	✓	Partial	✓	✓
E. ferox Bertolini f.	MBC 614*X	-	-	✓	-	✓	✓
E. ferox Bertolini f.	031015	AY335269	AF394384	-	-	-	-
E. ferox Bertolini f	KBG 139/1983	-	-	✓	✓	✓	✓
E. friderici-guilielmi Lehmann	KBG 19/1913	-	-	✓	✓	✓	✓
E. ghellinckii Lamaire	KBG 160/1915	-	-	✓	-	✓	✓
E. ghellinckii Lamaire	031017	AY335273	AF394375	-	-	-	-
E. gratus Prain	031018	AY335279	AF394405	-	-	-	-
E. gratus Prain	MBC 64561*C	-	-	✓	-	✓	✓
E. heenanii R. A. Dyer	KBG 964/1986	-	-	✓	-	✓	✓
E. heenanii R. A. Dyer	031021[2]	AY335283	AY335262	-	-	-	-
E. hildebrandtii A. Braun & C. D. Bouché	MBC 64575*S	-	-	✓	✓	✓	✓
E. hildebrandtii A. Braun & C. D. Bouché	031022	AY335287	AF394404	-	-	-	-
E. horridus (Jacquin) Lehmann	KBG 669/1915	-	-	✓	-	✓	✓
E. horridus (Jacquin) Lehmann	031023	AY335315	AF394417	-	-	-	-
E. humilis I. Verdoorn	KBG 270/1984	-	-	✓	-	✓	-
E. humilis I. Verdoorn	031024	AY335263	AF394378	-	-	-	-
E. inopinus R. A. Dyer	KBG 41/1983	-	-	✓	✓	✓	✓
E. inopinus R. A. Dyer	031025	AY335284	AF394387	-	-	-	-
E. ituriensis Bamps & Lisowski	MBC 20020330	-	-	✓	-	✓	✓
E. kisambo Faden and Beentjie	031026	AY335280	AF394402	-	-	-	-
E. laevifolius Stapf & Burtt Davy	031027	AY335281	AF394377	-	-	-	-
E. lanatus Stapf & Burtt Davy	KBG 269/1984	-	-	✓	✓	-	✓
E. lanatus Stapf & Burtt Davy	031028	AY335275	AF394376	-	-	-	-
E. latifrons Lehmann	KBG 111/1983	-	-	✓	✓	✓	✓
E. latifrons Lehmann	031029	AY335290	AF394418	-	-	-	-
E. laurentianus De Wildeman	MBC 96259	-	-	✓	-	✓	✓
E. laurentianus De Wildeman	031030	AY335313	AF394415	-	-	-	-

Taxon	Collection accession numbers	Genbank accession numbers for *ITS* sequences		*trnL* intron	Partial *rpoC1* intron	AFLP markers	
		ITS 1	*ITS* 2		*1F*	*B4*	*B13*
E. lebomboensis I. Verdoorn	031060	AY335324	AF394430	-	-	-	-
E. lehmannii Lehmann	KBG 573/1982	-	-	✓	✓	✓	✓
E. lehmannii Lehmann	031031	AY335276	AF394419	-	-	-	-
E. longifolius (Jacquin) Lehmann	KBG 574/1982	-	-	✓	✓	✓	✓
E. longifolius (Jacquin) Lehmann	031033	AY335265	AF394421	-	-	-	-
E. macrostrobilus S. Jones & J. Wynants	MBC 981917	-	-	✓	✓	-	✓
E. manikensis (Gilliland) Gilliland	031034	AY335285	AF394389	-	-	-	-
E. manikensis (Gilliland) Gilliland	MBC 64531*H	-	-	✓	✓	✓	✓
E. middelburgensis Voster	LHMS 4019	-	-	✓	-	✓	✓
E. middelburgensis Voster	031035	AY335304	AF394400	-	-	-	-
E. msinganus Voster	KBG 13/2000	-	-	✓	-	✓	✓
E. msinganus Voster	031036	AY335323	AF394429	-	-	-	-
E. munchii (R. A. Dyer) and I. Verdoorn	031037	AY335286	AF394391	-	-	-	-
E. natalensis (R. A. Dyer) and I. Verdoorn	KBG 129/1983	-	-	✓	✓	✓	✓
E. natalensis (R. A. Dyer) and I. Verdoorn	031039	AY335268	AF394432	-	-	-	-
E. ngoyanus I. Verdoorn	KBG 86/1945	-	-	✓	-	✓	✓
E. ngoyanus I. Verdoorn	031040	AY335264	AF394386	-	-	-	-
E. paucidentatus Stapf & Burtt Davy	031042	AY335322	AF394428	-	-	-	-
E. paucidentatus Stapf & Burtt Davy	MBC 97199*A	-	-	✓	✓	✓	✓
E. princeps R. A. Dyer	KBG 571/1982	-	-	✓	-	✓	✓
E. princeps R. A. Dyer	031043	AY335272	AF394425	-	-		-
E. pterogonus R. A. Dyer & I. Verdoorn	031044	AY335297	AF394390	-	-	-	-
E. pterogonus R. A. Dyer & I. Verdoorn	MBC 84207*C	-	-	✓	-	✓	✓
E. sclavoi Moretti, D. W. Stevenson & De luca	MBC 95940*A	-	-	✓	✓	✓	✓
E. sclavoi Moretti, D. W. Stevenson & De luca	031045	AY335306	AF394403	-	-	-	-
E. senticosus Voster	031046	AY335270	AF394426	-	-	-	-
E. senticosus Voster	MBC 95947*C	-	-	✓	✓	✓	✓
E. septentrionalis Schweinfurth	MBC 961303	-	-	✓	-	-	✓
E. septentrionalis Schweinfurth	031047	AY335311	AF394413	-	-	-	-
E. sp.(Chinyazange)	MBC 951026	-	-	✓	✓	✓	✓

Taxon	Collection accession numbers	Genbank accession numbers for *ITS* sequences		*trnL* intron	Partial *rpoC1* intron	AFLP markers	
		ITS 1	*ITS* 2		*1F*	*B4*	*B13*
E. tegulaneus Mellville	MBC 95939*A	-	-	✓	-	✓	✓
E. tegulaneus Mellville	031048	AY335312	AF394414	-	-	-	-
E. tegulaneus Mellville	MBC 591116*A	-	-	✓	-	✓	✓
E. transvenosus Staff and Burtt Davy	031059	AY335325	AF394431	-	-	-	-
E. trispinosus (Hook)R. A. Dyer	KBG 110/1983	-	--	✓	✓	✓	✓
E. trispinosus (Hook)R. A. Dyer	031049	AY335317	AF394422	-	-	-	-
E. turneri Lavranos & Goode	MBC 97219*B	-	-	✓	✓	✓	✓
E. turneri Lavranos & Goode	031051	AY335308	AF39408	-	-	-	-
E. umbeluziensis R. A. Dyer	031052	AY335289	AF394380	-	-	-	-
E. umbeluziensis R. A. Dyer	MBC 9867	-	-	✓	-	✓	✓
E. villosus Lehmann	KBG 2687/1915	-	-	✓	Partial	✓	✓
E. villosus Lehmann	031053	AY335267	AF394381	-	-	-	-
E. whitelockii P. J. H. Hunter	031055	AY335290	AF394409	-	-	-	-
E. woodii Sander	KBG 1895/1916	-	-	✓	-	✓	✓
E. woodii Sander	031057	AY335271	AF394433	-	-	-	-
Macrozamia riedlei (Gaudich) C. A. Gardner	Cons. Bio**	-	-	✓	-	-	-
Dioon edule Lindl	Cons. Bio**		-	✓	-	-	-
Dioon merolae De Luca, Sabato & Vázq Torres	Cons. Bio**	-	-	✓	-	-	-
Zamia furfuracea L. F.	Cons. Bio**	-	-	✓	-	-	-

** Conservation Biology green house, Kirstenbosch; KBG- Kirstenbosch Botanical Garden; MBC- Montgomery Botanical Center; LHMS-Leslie Hill Molecular Systematic Laboratory, DNA Bank; - Missing data (profiles/sequences); ✓ Profiles and sequences generated from this study.

APPENDIX 2- Groupings by Cooper and Goode (2004)

Genus	Species in the genus	Geographical distribution
Encephalartos	*caffer, cerinus, ngoyanus, umbeluziensis, aplanatus and villosus*	*Eastern Cape, *Kwa-Zulu Natal and ^Swaziland
Dracostrobus	*ghellinkii, lanatus, humilis, laevifolius, dedekindii, brevifoliatus, cycadifolius, mkomaasiana* sp. nov. *and friderici-guilielmi*	*Eastern Cape to *Mpumalanga Drakensberg
Rugostrobus	*longifolius, latifrons, and arenarius*	*Southern and *Eastern Cape
Acanthozamia	*lehmannii, trispinosus and horridus*	*Southern and *Eastern Cape
Pyrrostrobus	*ferox*	*Eastern Cape to ^Mozambique
Xanthostrobus	*natalensis, aemulans, altenstenii, msinganus, woodii, lebomboensis, curachii* sp. nov., *pietretief* sp. nov. *and senticosus*	*Eastern Cape, *Kwa-Zulu Natal, *Mpumalanga, ^Swaziland and ^Mozambique
Inezamia	*transvenosus, heenanii, relicta and paucidentatus*	*Limpopo and *Mpumalanga
Glaucostrobus	*dyerianus, eugene-maraisii, dolomiticus, hirsutus, nubimontanus, middelburgensis and cupidus*	*Gauteng, *Mpumalanga, *Limpopo
Dyerstrobus	*inopinus*	*Limpopo
Viridestrobus	*manikensis, concinnus, munchii, chimanimaniensis, and pterogonus*	^Southern Mozambique, ^Malawi and ^Zimbabwe
Tanzamia	*hildebrandtii, kisambo, equatorialis, tegulaneus, whitelockii, powysii, septentrionalis, mackenziei, gratus, ituriensis, macrostobilus, laurientianus and turneri*	^Kenya, ^Tanzania, ^Uganda, ^Sudan, ^Angola, ^Mozambique
Congostrobus	*bubalinus, schmitzii, poggei, delucans, marunguesis, schaijesii, allochrous and barteri*	^Kenya, ^Zaire, ^Angola, ^Ghana and ^Nigeria.

*Province; ^Country

APPENDIX 3- Habitat matrix

Taxa Name	Habitats*
E. aplanatus	1
E. aemulans	2
E. altensteinii	4
E. arenarius	4
E. barteri ssp. allochrous	5
E. bubalinus	6
E. caffer	2
E. cerinus	2,7
E. concinnus	3
E. cupidus	3,7
E. chimanimaniensis	3
E. cycadifolius	2,7
E. dolomiticus	3
E. dyerianus	3
E. ferox	4
E. ghellinckii	2
E. gratus	3
E. heenanii	2
E. hildebrandtii	3
E. horridus	4,6
E. humilis	2
E. inopinus	3,7
E. kisambo	3
E. laevifolius	2,3
E. lanatus	2,3
E. latifrons	2,4
E. laurentianus	1
E. lebomboensis	3
E. lehmannii	4
E. longifolius	2
E. manikensis	3
E. middelburgensis	3
E. msinganus	3
E. munchii	3
E. natalensis	3,7
E. ngoyanus	2,3
E. nubimontanus	3
E. paucidentatus	3
E. princeps	2,4
E. pterogonus	3

E. sclavoi	2,3
E. senticosus	3
E. septentrionalis	3
E. tegulaneus	3,7
E. transvenosus	1,3
E. trispinosus	4
E. turneri	3.5
E. umbeluziensis	3
E. villosus	1
E. woodii	1
E. whitelockii	1

*1-forest, 2-grassland, 3-woodland/savannah, 4-shrubland, 5-granite outcrop, 6-quartzite hills, 7-cliffs/gorges.

SUMMARY

Species-level phylogenetic reconstruction of the African cycad genus *Encephalartos* (Zamiaceae).

The main aim of the study was to reconstruct the phylogenetic relationships of the genus *Encephalartos* using DNA sequencing and AFLP data. The taxonomy of this genus remains uncertain and there has not been enough comprehensive molecular studies undertaken to resolve the surrounding uncertainties. This study is the first to use AFLPs together with DNA sequence data in an attempt to resolve the relationships within *Encephalartos* and to reduce the uncertainty that remains from previous molecular and morphological studies. The motivation for this study was therefore to clarify the uncertainties that still remain when it comes to the phylogenetic relationships pf the genus. Initially, sequencing data was generated but there were insufficient variable characters for phylogenetic reconstruction. Therefore, the four main objectives of the study were to:

1. Reconstruct and investigate phylogenetic relationships among species of the genus *Encephalartos* in Africa using AFLP and sequencing data to complement previous studies that used DNA sequencing and morphological characters.
2. Test the phylogenetic hypothesis from this study against the taxonomic relationships for *Encephalartos* proposed by Cooper and Goode (2004), and to compare the phylogenetic hypothesis from this study with that suggested by previous studies using DNA markers.
3. Map habitat types on the obtained phylogeny to make inferences regarding ancestral habitat types.
4. Evaluate the utility of the AFLP and DNA sequencing methods in reconstructing the phylogeny of the genus and make recommendations for follow-up studies.

Sequencing and AFLP primers were evaluated thoroughly to find those that would give data that was sufficiently variable for phylogenetic reconstruction. This led to my

proficiency in generating and analyzing DNA and AFLP data. The results generated from this study were compared with previous phylogenetic hypotheses of the genus. In terms of the relationships that were obtained, there was a high level of incongruency between different data sets and also several unresolved polytomies. Habitat optimization was also performed using downloaded *ITS* sequences to make inferences regarding ancestral habitat types. This was an attempt to understand the evolution of the genus with respect to its current distribution in different habitats and to interpret this in the context of divergence times of the different clades as suggested by previous authors. This analysis showed that there is a basal split in the *Encephalartos* cladogram that separates the grassland specialists, i.e. *E. cycadifolius, E. friderici-guilielmi, E. ghellinckii, E. humilis* and *E. lanatus,* from species that occur in either forests or grassland/savannah. This analysis shed some light into the possible divergence times of the genus and with further studies including dating the divergence for the genus, more accurate inferences could be made.

Regardless of the inconclusive results, the thesis initiates a foundation for future studies using AFLP data which can include more individuals per species of *Encephalartos* and also the combination of molecular and morphological data with additional DNA sequence data that has sufficient variability for phylogenetic analysis. This may provide greater resolution and clarity on the phylogenetic relationships.

Future phylogenetic studies on *Encephalartos* should include the combination of molecular and morphological data with additional DNA sequence data that has sufficient variability for phylogenetic analysis. This may provide greater resolution and clarity on the phylogenetic relationships and resolve the uncertainties surrounding the phylogenetic relationships between the species.
