

Preliminary investigations into the phylogenetic relationships in the genus *Erica* L.

by

Ntsikelelo Blessings Lester

A thesis presented in partial fulfilment
of the requirements for the degree
of Masters of Science
at the University of Stellenbosch



UNIVERSITEIT
STELLENBOSCH
UNIVERSITY

Supervisor: Prof D. U. Bellstedt

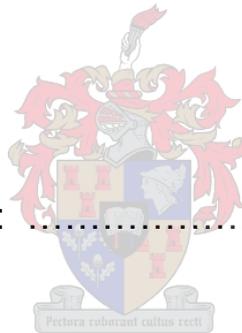
Co-supervisor: Dr. L. L. Dreyer

December 2006

Declaration

I, the undersigned, hereby declare that the work contained in this thesis is my own original work and that I have not previously, in its entirety or in part, submitted it at any university for a degree.

Signature:



Date:

ABSTRACT

Erica is a genus of about 860 species world wide, with 700 of these found in South Africa's southwestern and southern Cape, making it by far the most speciose genus in the Cape Floristic Region. This poses a particular challenge in the construction of a molecular phylogeny of the genus. The choice of suitably variable gene regions is a crucial decision on which the successful phylogenetic reconstruction of this important genus is critically dependent.

The aim of this project was therefore to determine which DNA regions, both chloroplast and nuclear, would be sufficiently variable to give adequate informative characters that may be useful at the species level phylogenetic reconstruction. A subset of 30 species, representing the range of morphological diversity and pollinator preference within *Erica*, was selected for study. For each of these species the variability in eight chloroplast regions (*trnL-F*, *matK*, *trnS-G*, *rps12-rpl20*, *psbA-trnH*, *trnC-D*, *rps4-trnT* and *trnT-L*) and the nuclear ITS region was investigated.

The *psbA-trnH*, *trnC-D*, *rps4-trnT* and *trnT-L* chloroplast regions were found to be problematic to amplify and to possess too few Parsimony Informative Characters to be of use in phylogenetic reconstruction. Four of the chloroplast regions, *trnS-G*, *trnL-F*, *matK* and *rps12-rpl20* and the nuclear ITS region could be amplified and sequenced with success. The ITS region was found to be reasonably variable, with the chloroplast genes showing less variability.

The DNA extraction method employed showed itself to be of critical importance in the success of the study. Two DNA extraction protocols, both modified from the original Doyle and Doyle (1987) method, were tested. The one included double the amount of β -mercaptoethanol and Polyvinylpyrrolidone (PVP) and the other included an extended phenol: chloroform: isoamylalcohol step. These variables, together with the effectiveness of these methods on fresh vs. silica dried plant samples, were investigated to determine which of the two would yield high quantities and qualities of DNA and result in the best method for the extraction of DNA from *Erica* species.

OPSOMMING

Erica is 'n genus van omtrent 860 spesies wêreldwyd, met 700 van hierdie spesies aanwesig in die suidwes en suid Kaap van Suid Afrika, wat dit by verre die mees spesieryke genus in die Kaapse Floristiese Streek maak. Dit stel 'n besondere uitdaging in die konstruksie van 'n molekulêre filogenie van die genus. Die keuse van geskikte variërende geen-areas is 'n belangrike besluit waarvan die suksesvolle filogenetiese rekonstruksie van hierdie belangrike genus krities afhanklik sal wees.

Die doel van hierdie projek was dus om te bepaal watter DNS areas, buide chloroplas en kern, genoegsaam varieer om voldoende informatiewe kenmerke te lewer om bruikbaar te wees in 'n spesie-vlak molekulêre rekonstruksie. 'n Subgroep van 30 spesies, wat die reeks van morfologiese diversiteit en bestuier voorkeure in *Erica* verteenwoordig, is dus vir die studie geselekteer. Vir elk van hierdie spesies is die variasie in agt chloroplast areas (*trnL-F*, *matK*, *trnS-G*, *rps12-rpl20*, *psbA-trnH*, *trnC-D*, *rps4-trnT* en *trnT-L*) en die kern ITS area ondersoek.

Dit was problematies om die *psbA-trnH*, *trnC-D*, *rps4-trnT* en *trnT-L* chloroplast areas te amplifiseer, en daar is gevind dat hulle te min Parsimonie Informatiewe Kenmerke besig om bruikbaar te wees in filogenetiese rekonstruksie. Vier van die chloroplas areas, *trnS-G*, *trnL-F*, *matK* en *rps12-rpl20* en die kern ITS kon suksesvol geamplifiseer word en die basisvolgordes kon suksesvol bepaal word. Daar is gevind dat die ITS area redelik variërend is, terwyl chloroplas areas minder variasie getoon het.

Die DNS ekstraksie metode wat gebruik is het die kritiese belang van die ekstraksie metode in die sukses van die studie bewys. Twee DNS protokolle, beide gemodifiseer van die oorspronklike Doyle en Doyle (1987) metode, is getoets. Die een het dubbel die hoeveelheid β -mercaptoetanol en Polyvinylpyrrolidone (PVP) bevat, en die het 'n uitgebruide fenol: chloroform: isoamylalkohol stap ingesluit. Hierdie veranderlikes, saam met die effektiwiteit van hierdie metodes op vars teenoor silika-gedroogde plant monsters, is ondersoek om vas te stel watter een van die twee die hoogste kwaliteit en kwantiteit DNS sou lewer en dus sal lei tot die beste DNS ekstraksie metode vir *Erica* spesies.

ACKNOWLEDGEMENTS

I am deeply grateful to my supervisors, Prof. Dirk. U. Bellstedt and Dr Léanne L. Dreyer, for giving me the opportunity to increase my understanding of molecular systematics. The support, encouragement, and advice given throughout this project were invaluable and are treasured.

The University of Stellenbosch and the National Research Foundation, through whose financial injection both my research and upkeep were financed, are acknowledged. This project would not have been possible without this support.

Many other people have also contributed to this project in significant ways:

- I am very grateful to Dr Ted Oliver and Benny Bytebier who collected most of the specimens, included in this project and provided me with advice concerning the morphology and ecology of the various taxa. Your input was invaluable and the project would not have been a success without your help.
- Mrs. Coral de Villiers for her help and support on the technical side of the project. My fellow laboratory colleagues, Margaret de Villiers and Dr. Annelise Botes, who were always very helpful, hospitable and always willing to give a hand. May the Lord reward you with more blessings.
- Chris Visser and Adri Rothmann for the great times spent together even through tough times. Have chips with that.....
- To Thato Moses and Gosiamé, my baby, for being understanding, patient and supportive throughout the duration of my studies. It was well worth it and we'll find happiness in it.
- My friend Dimpho Molale for always being supportive and offering advice through good and bad times. May the Lord bless you and all of your efforts.

Finally, my family, for showing me how invaluable education is and for always being there, what ever it might have taken. Your love, help, support and advice have brought me through it all from the beginning to the end I will always be grateful.

‘Majestic is the power of letters when their order's changed

But atoms bring to bear wider variety,

And so create,

A rich and infinitely wide diversity of things.

Lucretius

TABLE of CONTENTS

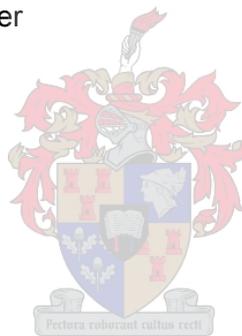
DECLARATION	i
ABSTRACT	ii
OPSOMMING	iii
ACKNOWLEDGEMENTS	iv
CHAPTER 1: INTRODUCTION	1
CHAPTER 2: BACKGROUND AND SIGNIFICANCE (LITERATURE OVERVIEW)	2
2.1. INTRODUCTION	2
2.2 <i>ERICA</i> L.	5
2.2.1 BOTANICAL CLASSIFICATION	5
2.2.2 DISTRIBUTION AND ORIGIN	6
2.3 PLANT SYSTEMATICS	8
2.4 MOLECULAR SYSTEMATICS (OVERVIEW)	11
2.5 OVERVIEW OF TECHNIQUES USED IN MOLECULAR SYSTEMATICS	11
2.5.1 DNA EXTRACTIONS	11
2.5.2 MOLECULAR TECHNIQUES	12
2.6 THE SIGNIFICANCE OF MOLECULAR SYSTEMATICS IN PLANT SYSTEMATICS	14
2.7 ANGIOSPERM CLASSIFICATION AND PHYLOGENY	17
2.8 <i>ERICA</i> SYSTEMATICS	20
CHAPTER 3: PRELIMINARY INVESTIGATIONS INTO THE PHYLOGENETIC RELATIONSHIPS IN THE GENUS <i>ERICA</i> L.	28
3.1 INTRODUCTION	28
3.2 MATERIALS AND METHODS	29
3.2.1 PLANT SAMPLING	29
3.2.2 DNA EXTRACTION	31
3.2.3 PCR AND DNA SEQUENCING	32
3.2.4 SEQUENCE ASSEMBLY AND PHYLOGENETIC ANALYSES	34
3.2.5 DATA MATRICES AND ALIGNMENTS FROM GENBANK	34
3.2.6 PARSIMONY ANALYSES	35
3.3 RESULTS	36

3.4 DISCUSSION	53
3.4.1 DNA EXTRACTION, PCR AMPLIFICATION AND SEQUENCING	53
3.4.2 DNA REGIONS, SEQUENCE ASSEMBLY AND PHYLOGENETIC ANALYSES	54
3.4.4 TOPOLOGIES AND INFERENCE FROM PHYLOGENETIC TREES	56
3.5 GENBANK DOWNLOADED AND EXPERIMENTAL DATA TREES	58
3.6 CONCLUSIONS	58
CHAPTER 4: ASSESSMENT OF AN OPTIMAL PROTOCOL FOR THE EXTRACTION OF DNA IN <i>ERICA L.</i> SPECIES.	59
4.1 INTRODUCTION	59
4.2 MATERIALS AND METHODS	60
4.2.1 PLANT SAMPLING	60
4.2.2 DNA EXTRACTION	60
4.2.3 PCR AMPLIFICATION	62
4.3 RESULTS	62
4.3.1 DNA EXTRACTION METHODS	62
4.3.2 PCR REACTIONS	62
4.4 DISCUSSION AND CONCLUSIONS	65
CHAPTER 5: CONCLUSIONS AND FUTURE PERSPECTIVES	67
LITERATURE	68

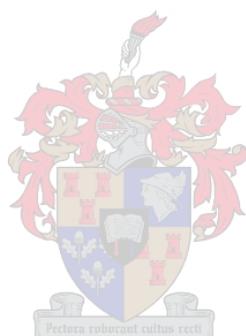


ABBREVIATIONS AND SYMBOLS USED IN THESIS

ABI:	Applied Biosystems Inc.
bp:	base pairs
BSA:	bovine serum albumin
ca.:	circa, approximately
CFR:	Cape Floristic Region
CI:	consistency index
CTAB:	cetyltrimethylammonium bromide
DNA:	deoxyribonucleic acid
dNTP:	deoxyribonucleotide triphosphate
indels:	insertions and deletions
<i>in prep.</i> :	in preparation
ITS:	internal transcribed spacer
KCl:	potassium chloride
M:	molar concentration
MgCl ₂ :	magnesium chloride
Myr:	million years
NH ₄ Ac:	ammonium acetate
(p.).	page(s)
PAUP*:	Phylogenetic Analysis Using Parsimony and other methods
PCR:	polymerase chain reaction
PIC	Parsimony Informative Characters
rDNA:	ribosomal DNA
RI:	retention index
s.s.:	<i>sensu stricto</i>
sp.:	species
ssp.:	subspecies
Taq:	<i>Thermus aquaticus</i>
TBR:	tree bisection reconnection



v/v: volume per volume

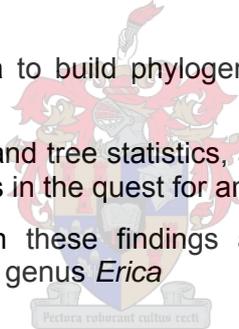


Chapter 1: Introduction

Erica includes about 860 species world-wide, with 700 of these found in the Fynbos vegetation peculiar to South Africa's southwestern and southern Cape regions. The systematics of several Fynbos genera have been analysed, but not that of *Erica*. Phylogenetic analyses of this genus were delayed mainly due to the enormous number of included species. Careful planning is crucial before a molecular phylogenetic study on such a speciose group is to be started. The main objective of this study was to determine which gene regions, both chloroplast and nuclear, would be sufficiently variable to yield enough informative characters that may be useful at the species-level in a phylogenetic analysis of *Erica*.

This aim was achieved by using a subset of 30 species representing a range of morphological diversity and pollinator preferences within the genus. In the process, the following sub-objectives were identified:

- to assess the phylogenetic effectiveness of gene regions that have previously been reported as informative at the species-level and establish their usefulness in *Erica*.
- to use these gene regions to determine which region has the highest number of Parsimony Informative Characters (PIC's)
- to use the DNA sequence data to build phylogenetic trees based on the selected gene regions
- to compare the tree topologies and tree statistics, and thus, determine which regions could be used for further investigations in the quest for an *Erica* phylogeny
- to draw conclusions based on these findings and thus to set the stage for further phylogenetic research within the genus *Erica*



This thesis is divided into four main chapters, the first is an introduction and explanation of the thesis layout, indicating the aims and objectives of the study and outlining each section of the thesis. The second is a literature overview covering all aspects leading to this study. The information deemed as relevant to the study is presented in this section. The background and overview of the study are also dealt with in this chapter. The evolution of plant systematics is further explained in a stepwise manner, from the description of the angiosperm phylogeny to the systematics of *Erica*, showing the path and studies undertaken to lead to the questions that are addressed in this study.

The third chapter presents experimental work undertaken in this study in the form of a scientific journal paper. The results obtained from the experimental work performed in this study are presented and discussed. The fourth chapter is also presented in the form of a scientific journal paper. It assesses an optimal protocol for the extraction of DNA in *Erica*. The final chapter (Chapter 5) presents conclusions and future perspectives; this section summarises the three previous chapters along with future recommendations and suggestions.

Chapter 2: Background and Significance (Literature review)

2.1. Introduction

On the southern-most tip of Africa lies a patch of land which covers about 90 000 km² stretching from the Bokkeveld escarpment in the Northern Cape to Port Elizabeth in the Eastern Cape, known as the Cape Floristic Region (CFR) (Goldblatt, 1978). This region is unique in terms of its flora, which is different from any other flora in the world and from the floras of the areas surrounding it. These floristic differences entail taxonomic composition, affinities of its taxa, plant structures and the ecology of the region at large (Linder, 2003).

The exclusivity of the CFR has led to many studies which have dealt with various aspects of this region. The CFR is recognised as the smallest of the six Floral Kingdoms of the world (Goldblatt, 1978). Its high levels of species endemism (70%) are similar to those of islands and the CFR has been reported to harbour about 9000 species, comparable to the most diverse equatorial regions (Linder, 2003). The Cape Floral Kingdom (CFK) is the only Floral Kingdom contained within the borders of a single country, with the others being the Antarctic Boreal or Paleoarctic, Neotropic, Paleotropic and Australasian.



Figure 2.1: The Cape Floristic Region as a biodiversity hotspot

The most striking features of the CFR are its species diversity and endemism, with 70% of species being endemic to the region. It is one of the 25 richest and most threatened reservoirs of plant and animal life on Earth. This very unique vegetation was once spread throughout the Western Cape

and the southwestern portion of the Eastern Cape as shown in Figure 2.1. The CFR and the Fynbos Biome are often confused, but these are not the same, though they are roughly coincidental (Van Wyk and Smith, 2001). The Fynbos Biome comprises the bulk of the CFR (Rutherford, 1986). The CFR itself includes some small patches of other biomes, for example Sub-Tropical Thicket in the Eastern Cape Province, the Succulent Karoo elements in the far northern region of the Western Cape Province, and both Nama Karoo and the Afromontane Forest in the Eastern Cape Province. Rutherford (1986) reported that the Fynbos stretched beyond the Western Cape through to Port Elizabeth in the Eastern Cape. Today Fynbos is mainly restricted to the mountainous and inaccessible parts of these regions due to spread of agriculture (Cowling, 1995).

Fynbos is a fire-adapted and drought-resistant shrubland largely confined to nutrient-poor soils in the winter rainfall areas of the southwestern Cape. Fynbos is mostly characterised by the presence of three main components: the restioids, ericoids and proteoids. The components may vary in abundance across Fynbos landscapes. The presence of wiry, reed-like restios is the uniquely distinguishing feature of Fynbos, with some vegetation being classified as fynbos on the basis of only 5% presence of restioids (Cowling, 2002).

The Fynbos includes four main growth forms, i.e. the proteoids, ericoids, restioids and the geophytes (the bulbous herbs). *Erica* is the largest ericoid genus, comprising of close to 700 species (7% of the total flora). Other large ericoid genera in the Fynbos include *Aspalathus* (245 species), *Agathosma* (130), *Cliffortia* (106), *Muraltia* (106) and *Phyllica* (133). Although *Erica* appears to be the largest genus in the CFR, it is the least studied genus phylogenetically and it has not been revised in its entirety (Cowling, 1995).

This great diversity of the CFR has led to many studies aimed at characterising its composition and understanding the evolution and radiation patterns of the included taxa. In 2003, Linder in one such study, reported on the forces that might have caused the high endemism and radiation patterns. In his study he identified groups of species that he defined as “Cape Clades”, which refers to those clades which have had most of their evolutionary history in the CFR, and that are thought to have been in the CFR since the Pliocene (Linder, 2003).

A few factors are thought to have affected or caused the radiation of the Cape species. These include, for example, patterns of climate change, the geomorphological evolution of southern Africa, sea-level fluctuations, with the two main factors appearing to be the sea-level fluctuations and climate changes. These have caused numerous extensions and contractions of the flora (Linder, 2003). This scenario leads us to the conclusion that the isolation of the Fynbos from the rest of southern Africa was caused by factors that were already present by the beginning of the Tertiary Period, which suggests that the high species diversity present in the CFR today is a result of the flora having had a long time to evolve under relatively stable conditions, with the possibility of a “crown group” being found in the CFR (Linder, 2003).

Linder and Hardy (2004) note that the generally accepted explanation for the massive diversity in the CFR was that most species evolved in the Pliocene and late Miocene (2 to 10 Myr ago) in response to the aridification of the region at that time. They used the African Restionaceae, which is known to constitute the third largest clade in the Cape flora (Linder, 2003), to try and date the diversity of the flora. Their results showed that the origin of the Restionaceae radiation ranges from 20 to 42 Myr ago, which is older than expected. The comparison of data from many different clades led them to conclude that these clades may have radiated in response to Fynbos vegetation increasing its extent in the Cape as a result of climatic change.

With this in mind, Linder (2003) used two decisive factors to define “Cape Clades” and to determine which groups should be included and which not. The first was to test whether the crown-group was in the CFR, that is, that 50% of the species of that clade occur in the CFR. The second criterion was to test if the clade originated in the CFR. He concluded that only 33 clades make up 50% of the 8888 species of the flowering plants of the CFR and the first nine clades contribute 9% of the flora. Table 2.1 (after Linder (2003)) is a tabular representation of the Cape Flora clades arranged according to the number of species present in the CFR for the top ten taxa. The percentage values for species in the CFR and cumulative percentage of the total flora are also supplied.

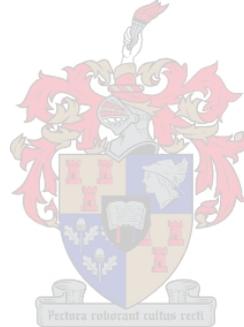


Table 2.1: Ten most speciose Cape floral clades arranged according to the number of species in the CFR, following Linder (2003).

Clade	Total species	CFR species	Percentage (%) in the CFR	Cumulative number of species	Cumulative % of total flora
<i>Erica s.l.</i>	860	658	77	658	7
Ixioideae & Nivenioideae	900	516	57	1174	13
African Restionaceae	350	340	97	1514	17
Crotalariaeae p.p	297	291	98	1805	20
Diosmeae (<i>Agathosma</i> , <i>Diosma</i> , etc)	276	268	97	2073	23
Proteae	340	264	78	2337	26
<i>Pelargonium</i>	250	148	59	2485	28
Irideae	226	136	60	2621	29
Phylliceae (<i>Nesiota</i> , <i>Phyllica</i> , <i>Noltia</i>)	152	134	88	2755	31
Relhaniinae	170	131	77	2886	32

2.2 *Erica* L.

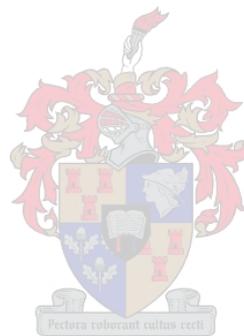
2.2.1 Botanical classification

Erica forms woody shrubs to trees varying in height from a few centimetres to three meters or more in CFR and higher in tropical Africa and Europe. The leaves are narrow with margins enrolled along the lower surface (they are rarely flat). Each flower has one bract and 2 bracteoles (rarely fewer) on the pedicles. There are four, mostly green, sepals that are either small or large to large and petaloid more conspicuous than the corolla. There are mostly 4 fused petals. *Erica* possesses 4 to 8 stamens (mostly 8) and they are free from the corolla. Anthers open through subterminal pores and often have appendages at the base of each theca. The ovary is mostly 4-chambered and often has a nectariferous disk at its base. Each locule includes one to mostly many ovules. Fruits are mostly dehiscent capsules that contain several to many very small seeds but can be few seeded indehiscent types. *Erica* species can be found flowering throughout the year (Trinder-

Smith, 2002). A few *Erica* species from southern Africa, southern Europe and northern Africa are illustrated in Figure 2.2.

2.2.2 Origin and Distribution

Erica has a distribution that ranges from the southern-most tip of Africa to the northern-most tip of Norway (Schumann and Kirsten, 1992). In South Africa the vast majority of species are confined to the CFR. Within the CFR, the highest concentration of species is found in the Caledon district, where more than 235 species co-occur. Elsewhere in southern Africa, *Erica* is found in the Eastern Cape, Kwa-Zulu Natal, Gauteng and Mpumalanga provinces of South Africa. It also occurs in Lesotho and Swaziland. Lowland regions also contain several species, for example in the Natal Midlands. Montane species are found in the non-CFR Drakensberg Mountains, Soutpansberg and Magaliesburg (Schumann and Kirsten, 1992).



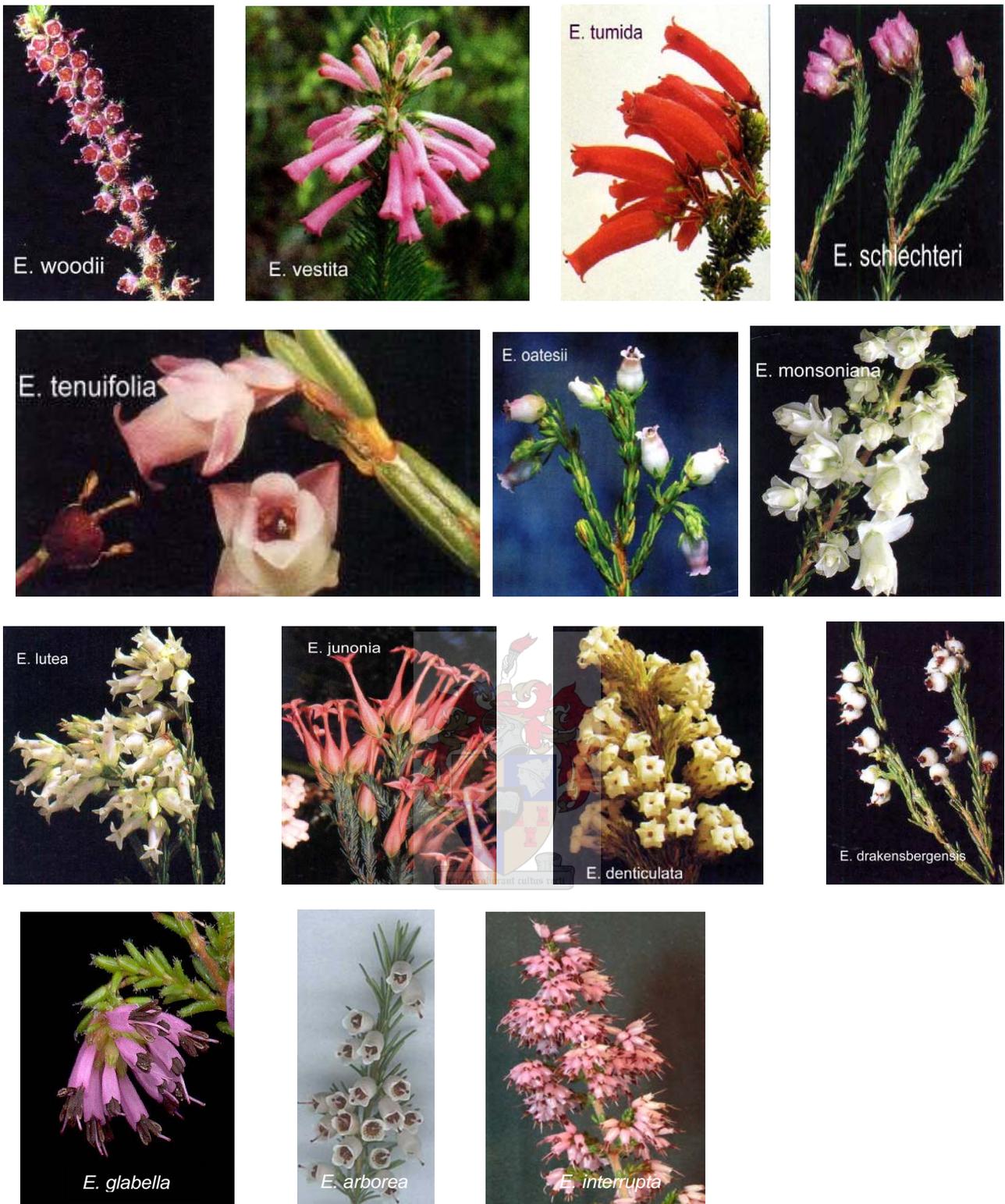


Figure 2.2: Some of the *Erica* species included in the study representing different morphologies, pollinator preferences and geographical regions. Pictures were acquired from the *Erica* interactive key with kind permission from Dr. E. G. H. Oliver.

The notion that Africa must have been the cradle of *Erica* origin is based on the subfamily Ericoideae being restricted to Africa, Europe and the islands on both the East and West of Africa, which suggests that the genus originated after the Gondwanaland break-up. The change in temperatures around the continent over time and through the epochs appears to have influenced the distribution of *Erica*. The changing topography of the continent, especially in the southwest and south could also have been an important factor that influenced the distribution patterns of *Erica* (Schumann and Kirsten, 1992).

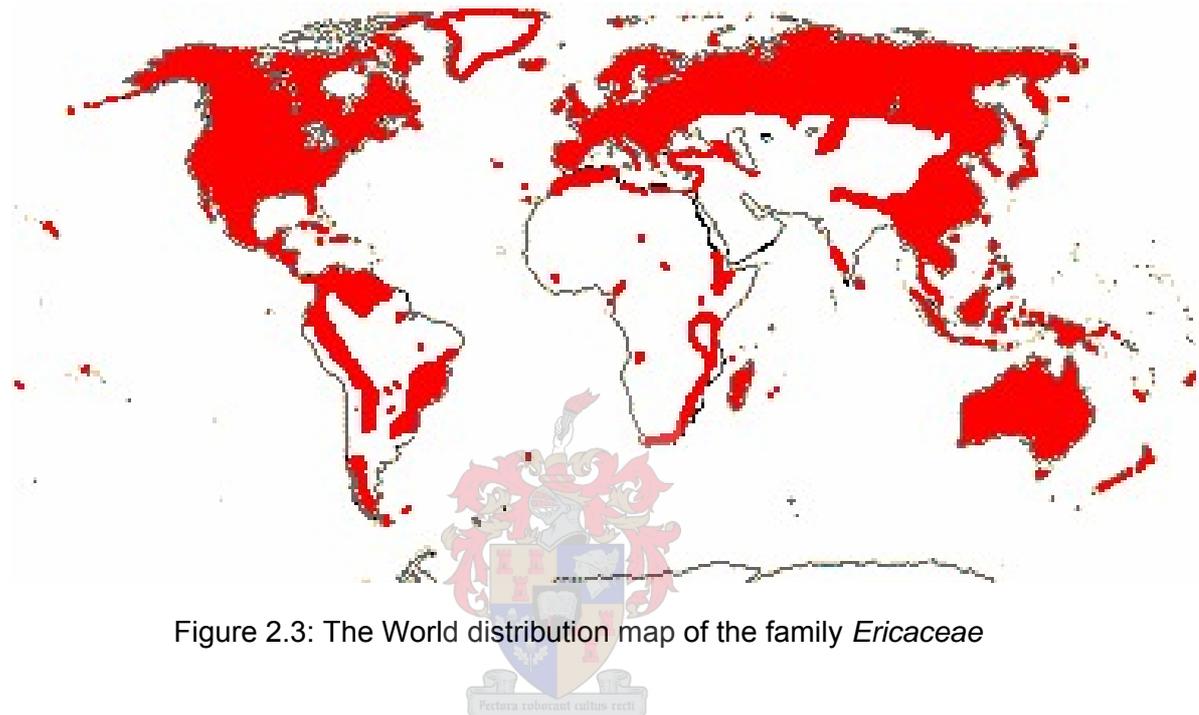


Figure 2.3: The World distribution map of the family *Ericaceae*

2.3 Plant systematics

Cronquist (1981, 1988) proposed morphological classifications of angiosperms in which he divided the dicots into 6 subclasses (Magnoliidae, Rosidae, Hamamelidae, Caryophyllidae, Dilleniidae and Asteridae). The subclass of interest in this study, Dilleniidae, had 13 orders with 78 families and 25 000 species, with three quarters of these species belonging to only five orders namely Malvales, Theales, Ericales, Capparales and Violales (Figure 2.4) (Cronquist 1981, 1988).

Another classification of flowering plants by Takhtajan (1997) treats flowering plants as a division or phylum (Magnoliophyta) with two classes (monocots and dicots), which are organized into subclasses. The higher-level organization is similar to the Cronquist system (1981, 1988). These classifications, along with many others, were subject to personal opinion, leading to discussions on the inadequacies of previously available classification systems for flowering plants (Kuzoff and Gasser, 2000).

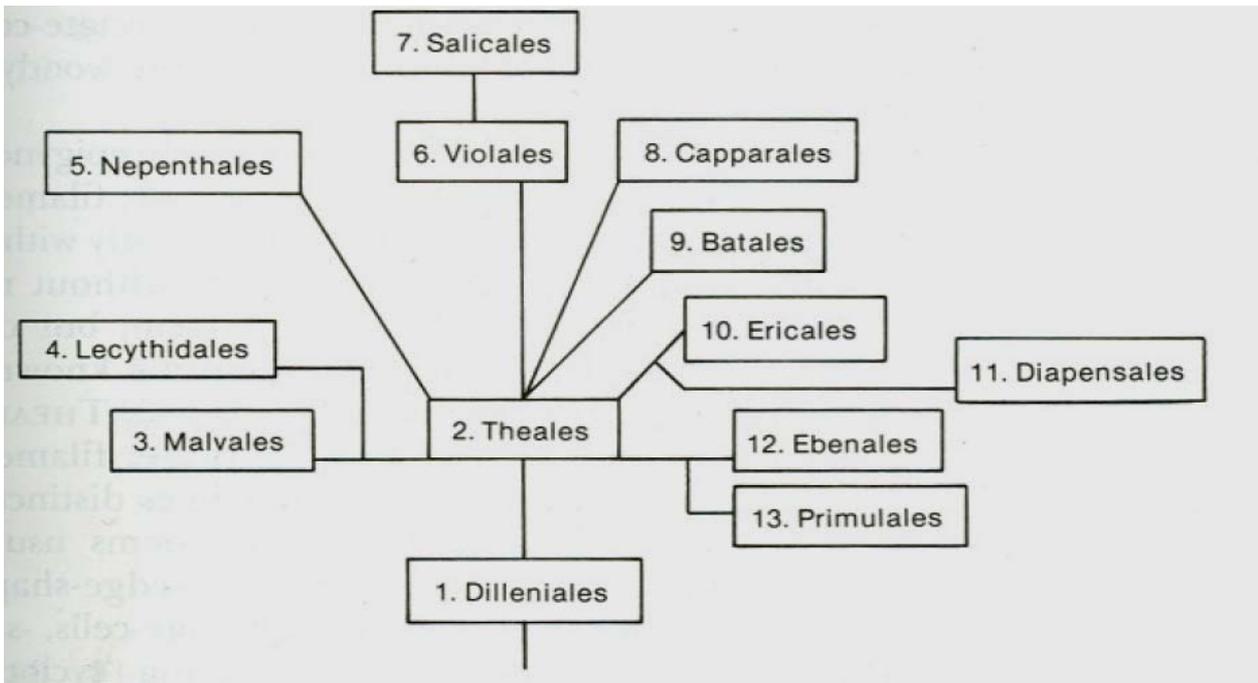


Figure 2.4: The 13 orders within the subclass Dilleniidae *sensu* Cronquist 1981, 1988.

A major break-through in our understanding of angiosperm classification came with the publication of the angiosperm phylogenetic tree (figure 2.5), published by the Angiosperm Phylogeny Group (APG, 1998) and subsequently revised (APG, 2003). This tree supports the monophyly of many major groups above the family level, and classified the 462 flowering plant families into 40 putatively monophyletic orders and a small number of monophyletic, informal higher groups. The APG (2003) proposed a basic classification into the major clades monocots, commelinoids, eudicots, core eudicots, rosids including eurosids I and II and asterids including euasterids I and II. There were, however, a number of families without assignment to order or any resolved phylogenetic placement.

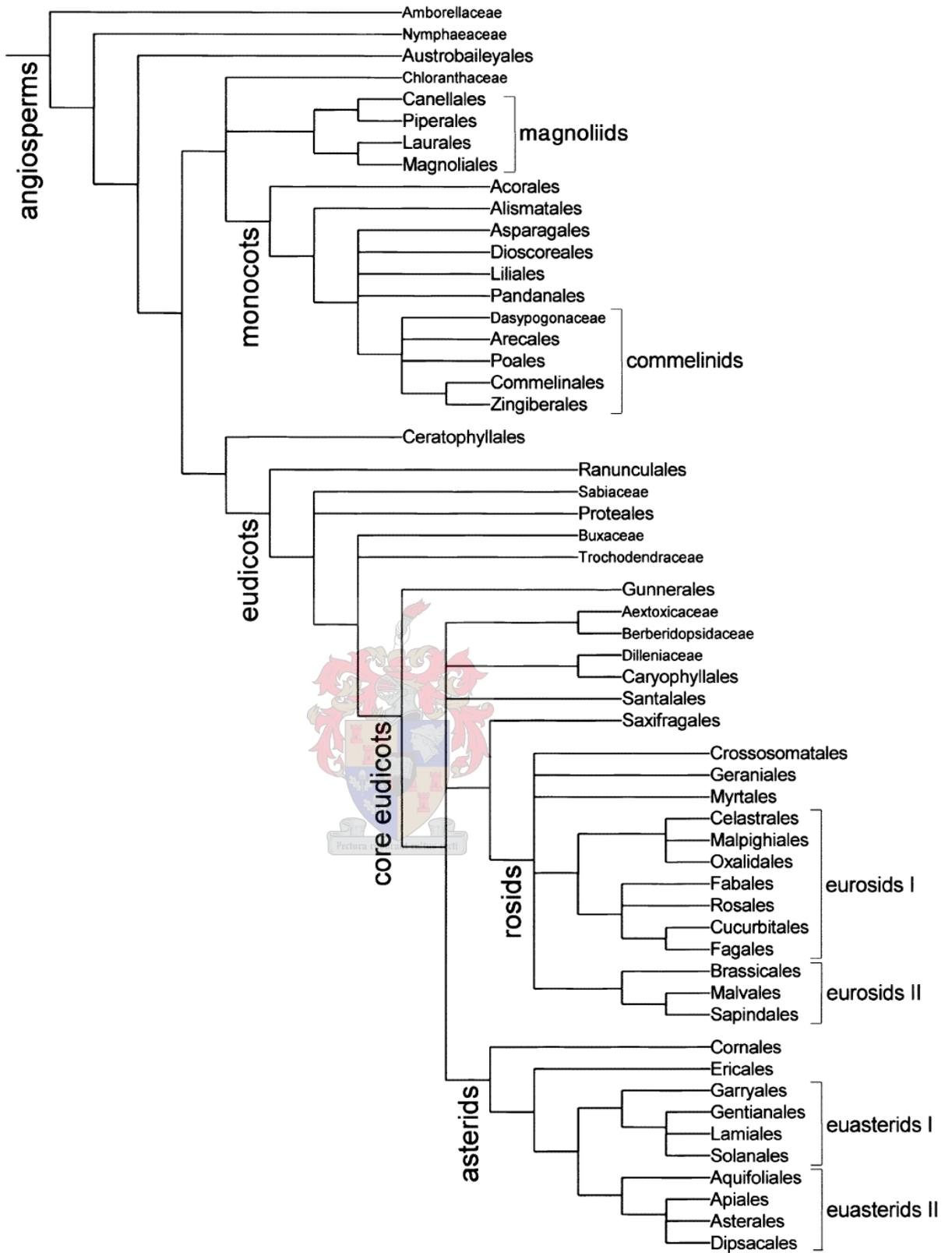


Figure 2.5: The angiosperm phylogeny as proposed by the APG (2003).

2.4 Molecular systematics (overview)

Molecular systematics is the science that uses data captured from DNA or RNA sequencing to find the evolutionary relationships among various groups of organisms. This study is also known as phylogenetic systematics; phylogenetics treats a species as a group of lineage-connected individuals over time. Techniques used in the field of molecular biology have become useful and very crucial in the execution of phylogenetic analyses. Due to automated sequencing technologies, advances in phylogenetic analysis software and access to increasingly powerful personal computers, the ease with which phylogenies can be generated has improved considerably over the past decade. These studies use data matrices, aligned gene sequences and statistical analyses. Building trees from such data matrices and assessing the statistical support for recovered clades have been facilitated by innovations in phylogenetic analysis software. A number of major relationships have been revealed through such analyses, which have given insight into the justification of relationships that had been previously thought to be true (Kuzoff and Gasser 2000).

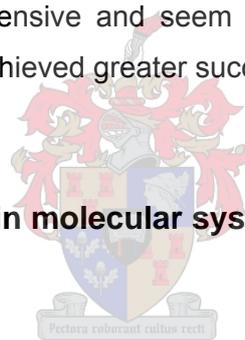
The use of molecular data has taken precedence over the other methods used in systematics in that most recent studies seem to rely more on these methods. Advantages include the fact that these methods are not too labour intensive and seem more reliable over a wide spectrum of organisms. Molecular methods have achieved greater success over many years now (Judd, 1999).

2.5 Overview of techniques used in molecular systematics

2.5.1 DNA extractions

DNA extraction forms a major part of the analyses in molecular biology studies. Over time, the development of molecular methods has brought about rapid advances in methods employed for DNA extraction. A few DNA extraction procedures, which can be used in isolating genomic DNA from various plant sources, have been developed. These include, for example, the salt extraction method and the cetyltrimethyl ammonium bromide (CTAB) method and its modifications.

Although there are two main procedures that are generally followed, i.e. the CTAB and the SDS based methods, most plant studies used the CTAB DNA extraction method published by Doyle and Doyle (1987). Although both of these methods have been known to give high yields in genomic DNA, there are a few shortcomings associated with extracting DNA in this manner. Several plants are known to contain high amounts of polysaccharides, polyphenols, tannins and hydrocolloids (sugars and carragenans). The problems associated with the presence of these compounds in the isolation and purification of DNA may result in the DNA being degraded. Causes of degradation include the activity of endonucleases, the co-isolation of polysaccharides along with the DNA and the presence of inhibitory compounds (e.g. polyphenols and other secondary metabolites) as a result of this co-isolation. Another crucial factor in genomic DNA isolation is that during the DNA



precipitation some of the contaminating RNA precipitates along with DNA, which causes many problems including suppression of PCR amplification (Padmalatha and Prasad, 2006).

A few studies have been aimed at finding a procedure that is genus independent, efficient, inexpensive, simple, rapid and yields pure DNA amplifiable by PCR. In these the CTAB method was modified through following the same basic principles. An example is the work of Padmalatha and Prasad (2006), in which they optimised the DNA isolation and PCR protocol for RAPD analysis of selected aromatic plants of conservation importance in India. Their modifications made use of PVP while grinding, used successive long-term small chloroform: isoamylalcohol extractions and proposed an overnight RNase treatment with all steps carried out at room temperature. Their results showed that the quality of the extracted DNA was improved by modifying some of the steps in the Doyle and Doyle (1987) method. The cetyltrimethylammonium bromide (CTAB) method was used to keep polysaccharides in solution and although SDS, sorbitol and glucose were added initially in the extraction buffer, they did not seem to exhibit any effect (Padmalatha and Prasad, 2006).

Chakraborti *et al.* (2006) also applied some modifications to the CTAB method, in which they determined a small and large-scale genomic DNA isolation protocol for chickpea (*Cicer arietinum* L.), suitable for molecular marker and transgenic analyses. In this study the key steps were considered to be based on four main changes, namely (1) extraction with high salt CTAB buffer to remove polysaccharides, (2) use of β -mercaptoethanol (0.2%) and PVP (1%) to remove polyphenolic compounds, (3) phenol: chloroform extraction to remove proteins (4) chloroform: isoamylalcohol extraction to remove the remaining phenol. These changes and modifications also gave better yields in the quality of the DNA produced. They concluded that their protocol could produce rapid, reliable DNA within 3-4 h. The small-scale method can be used for PCR based marker studies and screening of transgenics, whereas the large scale method is ideal for Southern hybridization analysis.

One other study to consider when deciding which protocol to use is that of Narayanan *et al.* (2006), which compares the SDS and CTAB methods used for DNA extraction. Their results show that by modifying the CTAB method you may lose in quantity, but end up with a purer template (Narayanan *et al.* 2006). The aim of all of these studies was to find a method that can isolate a purer template at the cost of yield.

2.5.2 Molecular techniques

There are a number of molecular techniques that can be employed in the assessment of phylogenies and biodiversity. All of these techniques require that the DNA is isolated from plant material as outlined in the previous section. Restriction fragment length polymorphism (RFLP), arbitrary primed DNA, amplified fragment length polymorphism (AFLP), microsatellites, sequence-

tagged simple sequence repeats (SSRs) and polymerase chain reaction (PCR) sequencing are DNA based techniques commonly used in plant molecular systematics. These techniques vary in that the methods used in each affects the template DNA in various ways. Variables to consider when selecting an appropriate technique include the ease with which they are performed, how they resolve genetic differences, the type of data that they generate, labour intensity, time frame involved in the technique and the taxonomic levels at which they may be most appropriately applied (Karp *et al.*, 1996).

- In the Restriction Fragment Length Polymorphism (**RFLP**) technique, the DNA is digested with restriction enzymes; the resultant fragments are separated by gel electrophoresis and blotted onto a filter. Probes are hybridized to the DNA. RFLPs give highly reproducible patterns, but variations in fragment lengths between individuals or species can arise either when mutations alter restriction sites or result in insertions and/or deletions between them (Karp *et al.*, 1996).
- Randomly Amplified Polymorphic DNA (**RAPDs**) analysis is a technique in which amplification products are produced by short synthetic primers of random sequence by PCR and are separated on agarose gels in the presence of ethidium bromide and visualized under ultraviolet light (Karp *et al.*, 1996).
- Amplified Fragment Length Polymorphism (**AFLPs**) is a technique in which DNA fragments are obtained from endonuclease restriction, followed by ligation of oligonucleotide adapters to the fragments and selective amplification by PCR (Karp *et al.*, 1996).
- **PCR sequencing** involves the determination of the nucleotide sequence within a DNA fragment amplified by Polymerase Chain Reaction (Karp *et al.*, 1996).
- **Microsatellites** are molecular marker loci consisting of tandem repeat units of very short (1-5 basepairs) nucleotide motifs which are amplified by PCR using suitable primers (Karp *et al.*, 1996).

Table 2.2 (A) summarizes the differences between the different techniques discussed above. Various characteristics of the techniques, ranging from abundance, level of polymorphism, locus specificity, reproducibility of the results, co-dominance and labour intensity are compared. Table 2.2 (B) outlines the technical aspects of each technique.

Table 2.2: (A). A summary of the variable differences between molecular systematics techniques.

	Yield	Level of polymorphism	Locus specificity	Co-dominance of alleles	Reproducibility	Labour intensity
RFLPs	High	Medium	Yes	Yes	High	High
Sequencing	Low	Low	Yes	Yes	High	High
RAPDs	High	Medium	No	No	Low	Low
Microsatellites	High	High	Yes	Yes	High	Low
AFLPs	High	Medium	No	Yes/No	High	Medium

Table 2.2 (B). A comparison of the technical aspects of various molecular systematics techniques.

	Technical demands	Operational costs	Development costs	Quantity of DNA required	Amenability to automation
RFLPs	High	High	Medium-high	High	No
Sequencing	High	High	High	Low	Yes
RAPDs	Low	Low	Low	Low	Yes
Microsatellites	Low-medium	Low-medium	Low	Low	Yes
AFLPs	Medium	Medium	Medium	Medium	Yes



2.6 The significance of molecular systematics in plant systematics

The development of these techniques has led to major advances in molecular systematic studies, from which a few key studies resulted. One such study is that by Savolainen *et al.* (2002) on phylogeny reconstruction and functional constraints in organellar genomes, with a comparison of plastid and mitochondrial genes. They shed light on which genes to target when investigating specific groups for amplification and Potentially Informative Characters (PICs). Subsequently studies such as the assessment of the utility of low-copy nuclear gene sequences in plant phylogenetics by Sang (2002) led to a better understanding and inference of phylogenies.

The usage of molecular techniques and evidence in the quest to unravel the history and evolutionary patterns of plants has increased massively in recent years. This is manifested in the large increase in the number of articles published annually (+/-4000) that include a phylogenetic tree (Savolainen, 2000). Molecular systematics has not only proven to be advantageous in terms of reducing labour, but has also succeeded in the identification of important new characters which are not easily identified or that might not be identified using other methods. Another advantage of this

method is the number of characters that can be gathered for a selected taxon or data set (Soltis *et al.*, 1998).

Though advantageous in many ways, there are a few shortcomings associated with this method of inferring phylogenies. DNA does not stay unaltered for all time, since events such as hybridization, mutations and introgression affect the actual make-up of the DNA. This may cause a particular sequence retrieved from a given locus not necessarily to reflect the true phylogeny. Another limitation is that when using DNA, only a very small portion of the genome is sequenced at a time; while other systematics methods have the advantage of being based on a more complete set of characters (Dixon and Hills, 1993).

Lee (2001) mentions the inclusion of insertion and deletion (indels) characters can result in the interpretation of separate occasions as one evolutionary event. Such problems can, in part, be overcome by collecting data from different sources (genomes). Soltis (1998) suggested the comparison of trees derived from both chloroplast and nuclear DNA data in order to capture information that could be a close reflection of the actual phylogeny when assessing relationships at low taxonomic levels. The informativeness of chloroplast and mitochondrial data at this level has led to certain questions. Data from chloroplasts were, for example, found to show too little variation between taxa to allow robust phylogenetic reconstruction at the species level in the *Lampranthus* group (Klak *et al.*, 2003) and *Protea* (Reeves, 2001).

One advantage that can be associated with using the plastid genome is its size, which is usually small, ranging between 120 and 200 kb long. It is therefore easier to sequence a large part of the genome and capture all the necessary information needed to infer phylogenies. The fact that chloroplast DNA constitutes a single copy region is also advantageous, in that gene regions do not belong to multi-gene families such as genes in the nuclear genome. Gene families are known to cause problems when inferring phylogenies (Soltis 1998).

The mitochondrial genome has been reported to rearrange itself frequently in plants, which renders the order of the genes to be variable and hence unreliable as the rearrangements do not necessarily differentiate or characterize the groups of species. Chloroplasts, on the other hand, are known to be stable with little and very rare rearrangements, so that they can be used to mark evolutionary changes and be used as a means of determining the different groupings in plants (Judd *et al.*, 1999). The complete chloroplast genome of tobacco, for example, has been determined (Wakasugi *et al.*, 1998). Figure 2.6 is an adaptation from Shaw *et al.*, (2005) of the scaled map of the 21-noncoding cpDNA regions surveyed in this investigation (based on the *Nicotiana* chloroplast genome (Wakasugi *et al.*, 1998)).

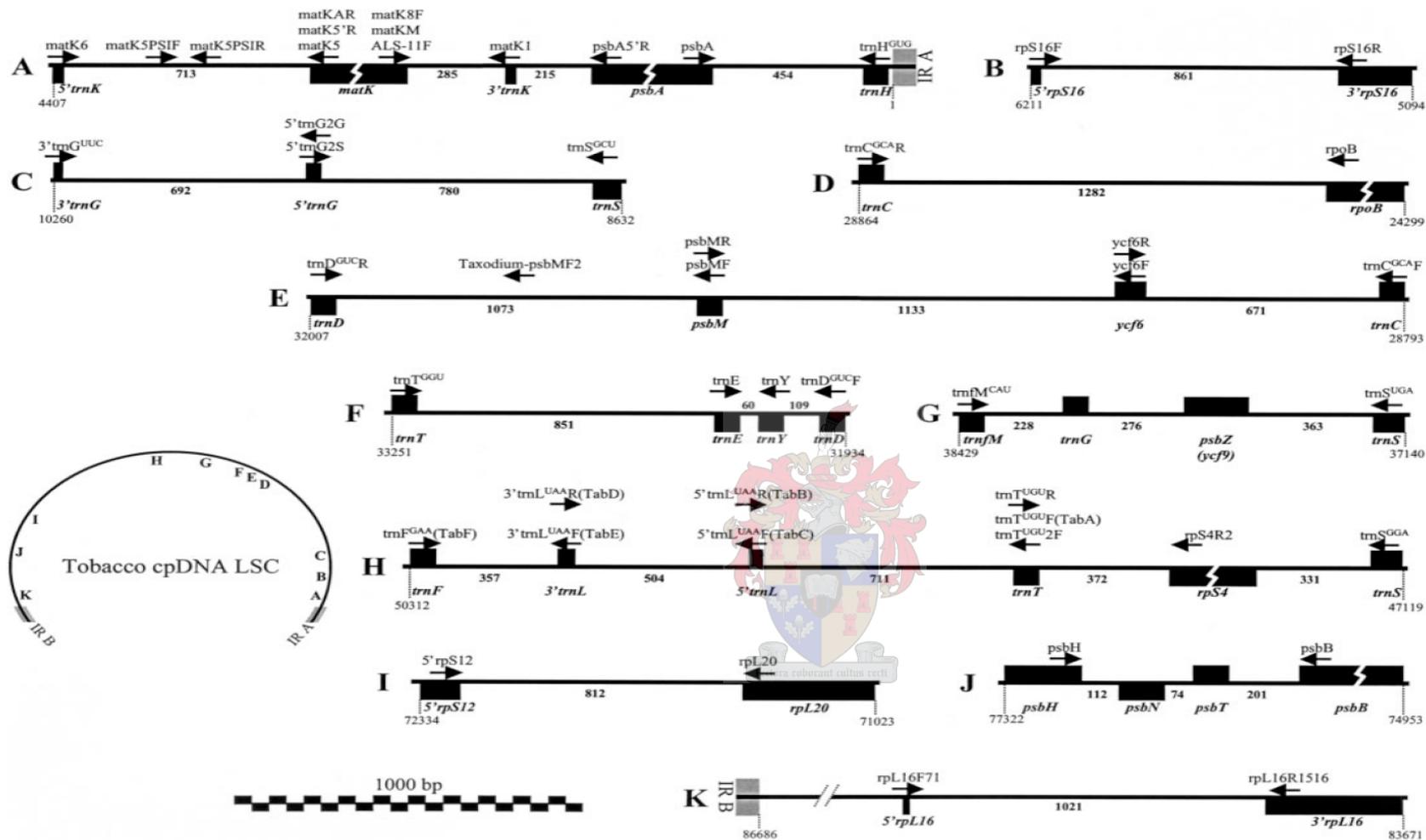


Figure 2.6: Scaled map of the 21-noncoding cpDNA regions surveyed by Shaw *et al.* (2005) (based on the *Nicotiana* chloroplast genome [Wakasugi *et al.*, 1998]). The orientation and relative positions of the genes are identified (A–K) along the Large Single Copy (LSC) portion with specific positions denoted by offset numbers at the beginning and end of each region. Gene names are italicized below and amplification and sequencing primer names are in roman typeface above with directional arrows. Lengths of non-coding regions are centered below.

The complete sequencing of chloroplast genomes increased the dependability on chloroplast data to produce phylogenies. The vulnerability of the plastid genome to events such as hybridisation and introgression, however, led to doubts concerning the inferred topologies. The inclusion of other data sources (e.g. nuclear and/or mitochondria regions) thus became important for comparison and an assessment of the usefulness of chloroplast data in phylogeny reconstruction (Soltis *et al.*, 1998).

Due, in part, to these shortcomings in the chloroplast data, nuclear data (e.g. 18S, 26S and the internal transcribed spacers between them and the 5.8S rDNA) have also been extensively used in molecular systematics of plants. Though the nuclear data cannot exclusively be relied upon, the Internal Transcribed Spacers (ITS1 and ITS2) and 5.8S rDNA spacer have been extensively used and they have added valuable information in groups at lower levels of relatedness (Soltis *et al.*, 1998).

Molecular data using DNA have been successful in the reconstruction of phylogenies of various organisms and have revealed the evolutionary history over a diverse spectrum of biological groups. Among plants, various phylogenies have been reconstructed at both the family and species level with varying rates of success.

2.7 Angiosperm classification and phylogeny

In the APG, (2003), (Figure 2.5) five new orders were recognized, namely Austrobaileyales, Canellales, Celastrales, Crossosomatales and Gunnerales. These represent well-supported monophyletic groups of families that were not classified to order in APG (1998). Results of the new APG (2003) did not propose any particularly new circumscriptions, except for the addition of these previously unclassified groups to the tree. It resulted from recent analyses that showed that these so-called unclassified families were nested inside the APG orders or well-supported as sister groups to some of the former orders. None of the initial APG orders were changed in APG (2003).

The APG tree serves as our basis of understanding the phylogenetic relationships among the principal lineages (or clades) of angiosperms. It partially clarifies the evolutionary events that triggered the diversification and rise of this ecologically dominant plant group. For these and other reasons, reconstructing the angiosperm phylogeny has been a milestone in plant systematics (Kuzoff & Gasser, 2000). The APG tree was reconstructed from sequence data from the *rbcl*, *atpB* and 18S rDNA genes (APG, 2003). This led to certain gene regions being viewed as universal regions, which promoted a certain dependence on specific gene regions to resolve all phylogenetic problems.

The quest to unravel phylogenies of the main components of the CFR has not been left behind in this chase. Several studies have attempted to reconstruct phylogenies from a broad range of genera and families, including for example *Pelargonium* (Geraniaceae; Bakker *et al.*, 1999),

Aspalathus (Hawkins, 2000), *Protea* (Proteaceae; Reeves, 2001), *Phyllica* L. (Rhamnaceae; Richardson *et al.*, 2001), *Moraea* (Iridaceae; Goldblatt *et al.*, 2002), *Cliffortia* L. (Rosaceae; Whitehouse, 2002), *Ehrharta* Thunb., (Poaceae; Verboom *et al.*, 2003) and *Disa* (Orchidaceae; Bytebier *et al.*, 2006). Within some families such as the Proteaceae, DNA sequence data have been applied to elucidate relationships both among genera (Barker *et al.*, 2002; Hoot & Douglas, 1998) and within genera e.g. *Protea* (Reeves, 2001).

2.7.1 Variability assessments in plant molecular phylogenetic studies

Although there have been immense advances in molecular phylogenetic studies, plant phylogenetic studies have greatly relied only on a few chloroplast and nuclear gene regions as a source of data. The magnitude of information derived from using these regions cannot be disputed, but the use of other markers is crucial in order to improve results in terms of better resolution and level of accuracy (Sang, 2002).

Several studies have assessed relative utility of different gene regions at various levels. With most of the phylogenies being resolved at order and family level (APG, 1998), the next hurdle was to find gene regions informative and variable at the species level. In one such study, Sang (2002) reported on the relative utility of low-copy nuclear gene sequences in plant phylogenetics. He discusses how low-copy nuclear genes in plants are a rich source of information with potential to recover the phylogenetic information at all taxonomic levels (Sang, 2002). He also stated that the use of low-copy nuclear genes increased the amount of work, since it requires techniques such as cloning and Southern blotting, but concluded that evolving introns of the low-copy nuclear genes can provide much-needed phylogenetic information around the species boundary and allow us to address fundamental questions concerning processes of plant speciation (Sang, 2002).

Another study by Aoki *et al.* (2003) attempted to compare levels of variation among several different non-coding cpDNA regions across a wide range of lineages. They investigated the usefulness of cpDNA in detecting intraspecific variation in plant species from the evergreen broad-leaved forests in Japan, including 41 component species. They used 16 non-coding cpDNA regions, and found that 14 of the species showed intraspecific variations in these regions. *trnL-F*, *petD-rpoA* and *rp16A* were the regions that gave relatively large amount of intraspecific variation. In 15 species and one species group the amount of intraspecific variation was compared and *rps16* was found to be the region with the highest variation. They concluded that these cpDNA regions could be useful in studies aimed at finding intraspecific relationships among other angiosperm groups (Aoki *et al.* 2003).

A few subsequent key studies followed in the search for DNA regions that could be informative at the specific level. Aoki *et al.* (2003) argued that regions such as the *trnL* intron, the *trnL-trnF* spacer and the *trnK* intron/*matK* gene would be useful at these levels. In another attempt to

investigate the levels of variation among several different non-coding cpDNA regions across a wide range of lineages, Shaw *et al.*, (2005) ruled out the results of Aoki *et al.* (2003) as being equivocal due to insufficient data.

Shaw *et al.* (2005) set out to determine whether there is any predictable rate heterogeneity among 21 non-coding cpDNA regions, identified as phylogenetically useful at low taxonomic levels. To test for rate heterogeneity among the different cpDNA regions, they used three species from each of 10 groups representing eight major phylogenetic lineages of phanerogams.

They acknowledged chloroplast DNA as a primary source of data in phylogenetics. In their results they divided the regions into three tiers based on their quantitative values. This categorised the gene regions from the most informative to the least informative. The results of this study did not necessarily reflect a certain number or combination of regions that can be used universally at low levels. Instead, their results suggest several regions (non-coding), which may be utilised at this level.

In their study, the number of nucleotide substitutions, indels, and inversions were referred to collectively as potentially informative characters and included both synapomorphies and autapomorphies. In parsimony only synapomorphies are scored as parsimony informative characters. Shaw *et al.*, (2005) included autapomorphies as part of what they referred to as potentially informative characters, as autapomorphies have the potential to become synapomorphies (and therefore parsimony informative characters) when taxon sampling is increased. Figure 2.7 is a graphical illustration of the varying levels of potentially informative characters across the various taxonomic groups they studied. It can be observed that there are common trends amongst taxonomic groups with regards to the informative characters in each region, thus making it easier to identify regions that may be useful when working on a particular group (Shaw *et al.*, 2005).

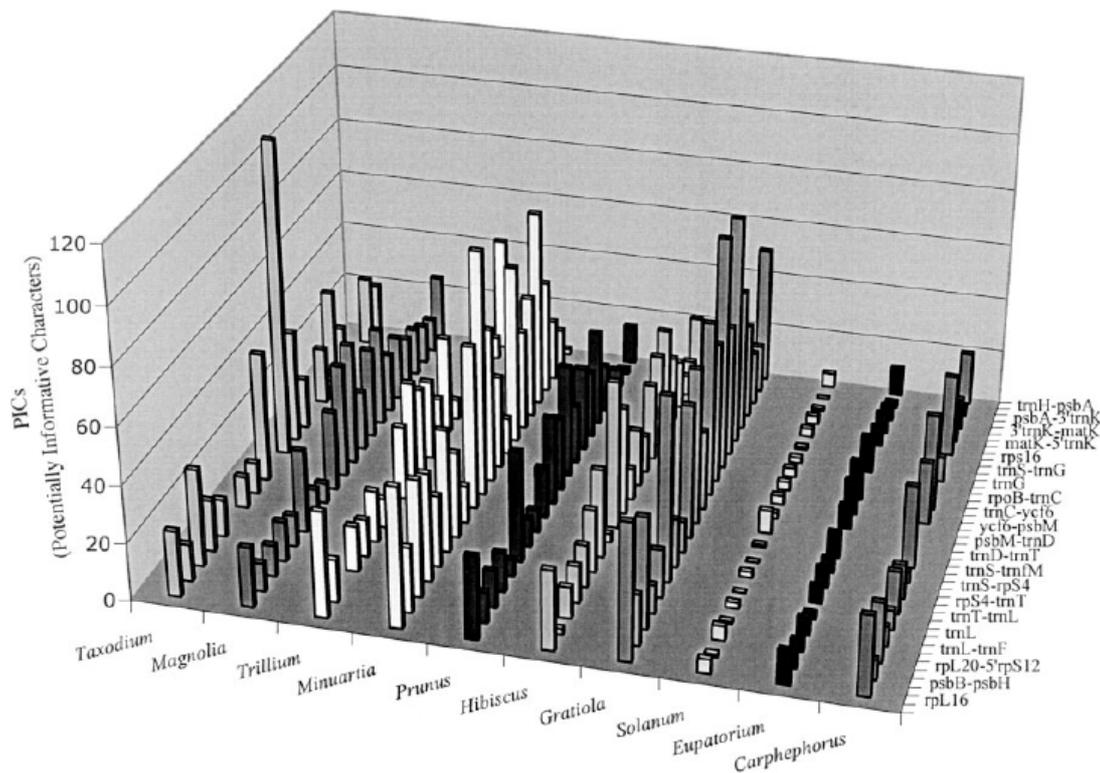


Figure 2.7: Graphical illustration of the varying levels of potentially informative characters across the various taxonomic groups from Shaw *et al.* (2005). Various plant groups are shown on the x-axis with the various gene regions on the z-axis.

Other important findings from the Shaw *et al.* (2005) study were that rate heterogeneity exists among non-coding cpDNA regions and that a survey using as few as three representative taxa can be predictive of the amount of phylogenetic information offered by a cpDNA region. The outcome of such a pilot study can be used to gain preliminary insights into the likely outcome of a full-scale test. This can have major cost implications with regards to time and resources. The results of their study clearly indicate the importance and applicability of performing pilot studies to identify appropriate regions for further, more inclusive, study. They also show how crucial the choice of an appropriate gene region can be before trying to run a phylogenetic analysis of any group. The outcome of their study set the stage for further analyses and phylogenetic reconstruction of larger groups like the CFR mega genus *Erica*.

2.8 *Erica* systematics

According to Cronquist (1981, 1988) the order Ericales contained 8 families with the family Ericaceae including 125 genera. Takhtajan (1997) proposed that the Ericales comprised largely of Sarracenianae, Ericanae, Primulanae and some families in Theanae, all adjacent groups in the Dilleniidae. Efforts by the Angiosperm Phylogeny Group (APG, 1998) have resulted in the subclass Dilleniidae being completely dissolved and the orders that were included in this subclass being

spread between the Rosids and the Asterids. Ericales were placed in the Asterids. The age of the Asterid stem is believed to be *ca.* 128 myr old, dating to the mid-early Cretaceous, with Ericales diverging soon afterwards (Bremer *et al.* 2004). Ericales now include 26 families and about 9450 species, with Ericaceae as one of the major families based on morphological characters (Judd, 1999). The APG (1998), though having resolved many other relationships, reported three families thought to be Ericales as unclassified, namely Lissocarpaceae, Pentaphragmataceae and Sladeniaceae.

The poor understanding of relationships in the order Ericales prompted further research. Anderberg *et al.* (2002) reconstructed the phylogenetic relationships in the order Ericales and tried to increase the basal resolution within the order by using five gene regions from the chloroplast genome. This study included all families thought to be closely related to Ericaceae. Their results yielded new circumscriptions in the Ericaceae in terms of the tree at large, found support for a number of clades and determined the relationships amongst various families including Ericaceae. The combined tree provided a more resolved topology (Figure 2.8) and the plastid data gave better resolution than the mitochondrial data. This led to the courageous assumption that several of the included groups had evolved rapidly and simultaneously, resulting in the observed difficulties in retrieving well-supported relationships. Anderberg *et al.* (2002) also reported that Ericaceae is sister to the rather poorly known Cyrillaceae and that these two families are in turn sister to Clethraceae.



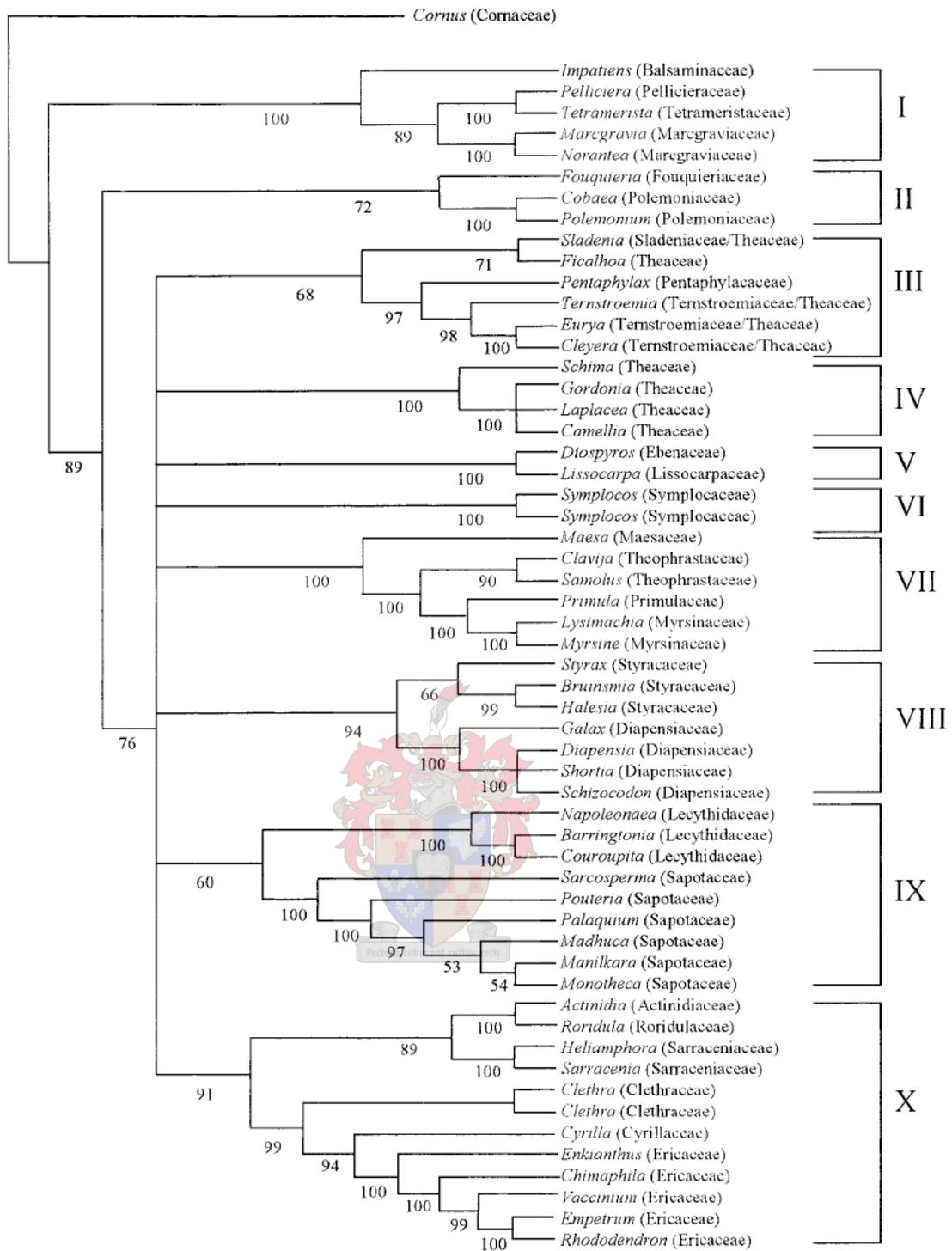


Figure: 2.8: The Ericales tree based on analysis of a combination of sequences from the five genes *atpB*, *ndhF*, *rbcL*, *atp1* and *matK*. Roman numerals correspond to numbering of clades in the text from Anderberg *et al.* (2002).

Kron *et al.* (2002) undertook a molecular systematic study to resolve the relationships within the family Ericaceae, using the *rbcL* and *matK* gene regions. Combining this information with morphological data they proposed a new classification in which 4 sub-families were recognized. Only monophyletic groups were recognized within Ericaceae. These results confirmed that well-

known genera like *Rhododendron*, *Erica*, *Calluna*, *Vaccinium* and *Gaultheria* were included in the subfamily *Ericoideae* (Figure 2.9).

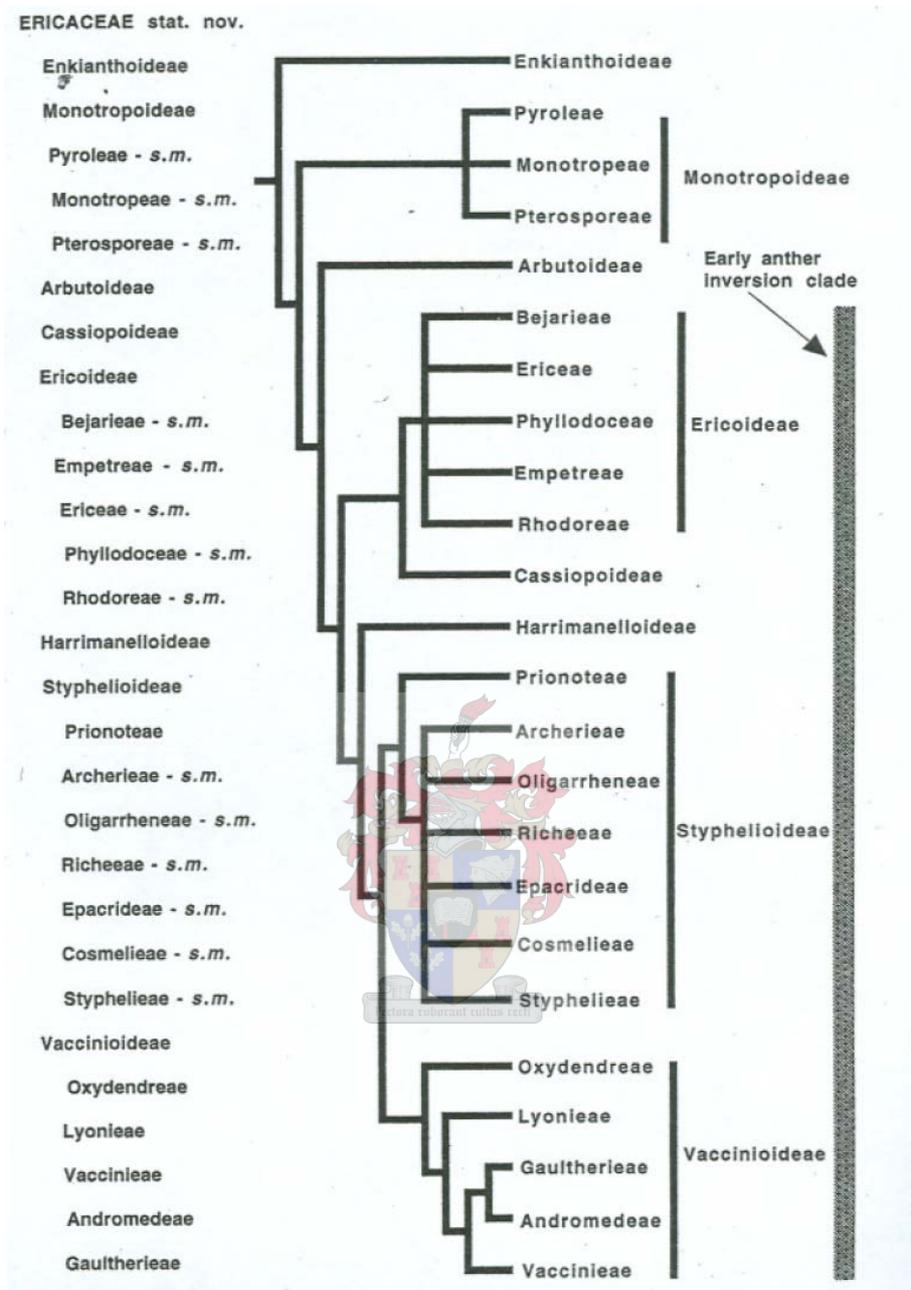


Figure 2.9: The new classification of Ericaceae based on *rbcl* and *matK* sequence data adapted from Kron *et al.* (2002).

Oliver (2000) tried to resolve the relationships within the genera of the tribe Ericaceae using morphological characters. He revised the minor genera of Ericaceae and found that the main groups of minors containing the putatively more derived species are all restricted to the Cape Floristic Region. Only two species are known to have localities outside this region, *E. rigidula* that extends into the Kamiesberg region of Namaqualand and *E. zeyheriana* extending to Alexandria in the east.

He reduced the overall number of *Erica* minor genera species in the genus from 123 to 84, including 15 new species.

Oliver's (2000) classification reduced 23 minor genera to synonymy under *Erica* and now include only Ericaceae 3 genera: the mega-genera *Erica*, *Calluna* and *Daboecia* (Oliver's 2000). These results also supported the reduction of *Philippia*, *Blaeria* and *Ericinella* under *Erica* (Oliver 1987d, 1988a, 1993b, 1993c, 1994). Oliver (2000) also emphasized that through his work and other studies, a solid base has been laid upon which research on the phylogenetic relationships and evolution of Ericaceae can be initiated using additional morphological and molecular data.

The phylogenetic tree of *Erica* as presented by Oliver (2000) is illustrated in Figure 2.10. This shows the new circumscriptions of the genus *Erica sensu* Oliver. There are problems that arose during this cladistic analysis. The trees were mostly very poorly resolved. This is manifested by the mixed aggregation of genera in the consensus tree. Oliver (2000) put forward a few conclusions upon which he based the reduction of minor genera into *Erica*. He stated that some of the minor genera are actually groups of species that may be large or comprised of only one species that could have evolved at different points from within *Erica*. The paraphyletic state of *Erica* as compared to the polyphyletic state of the minor genera appears to support the merger of the minor genera in *Erica*.

This study by Oliver (2000) clearly showed that morphological data and the other traditional characters used are not sufficiently informative as tools for inferring phylogeny.



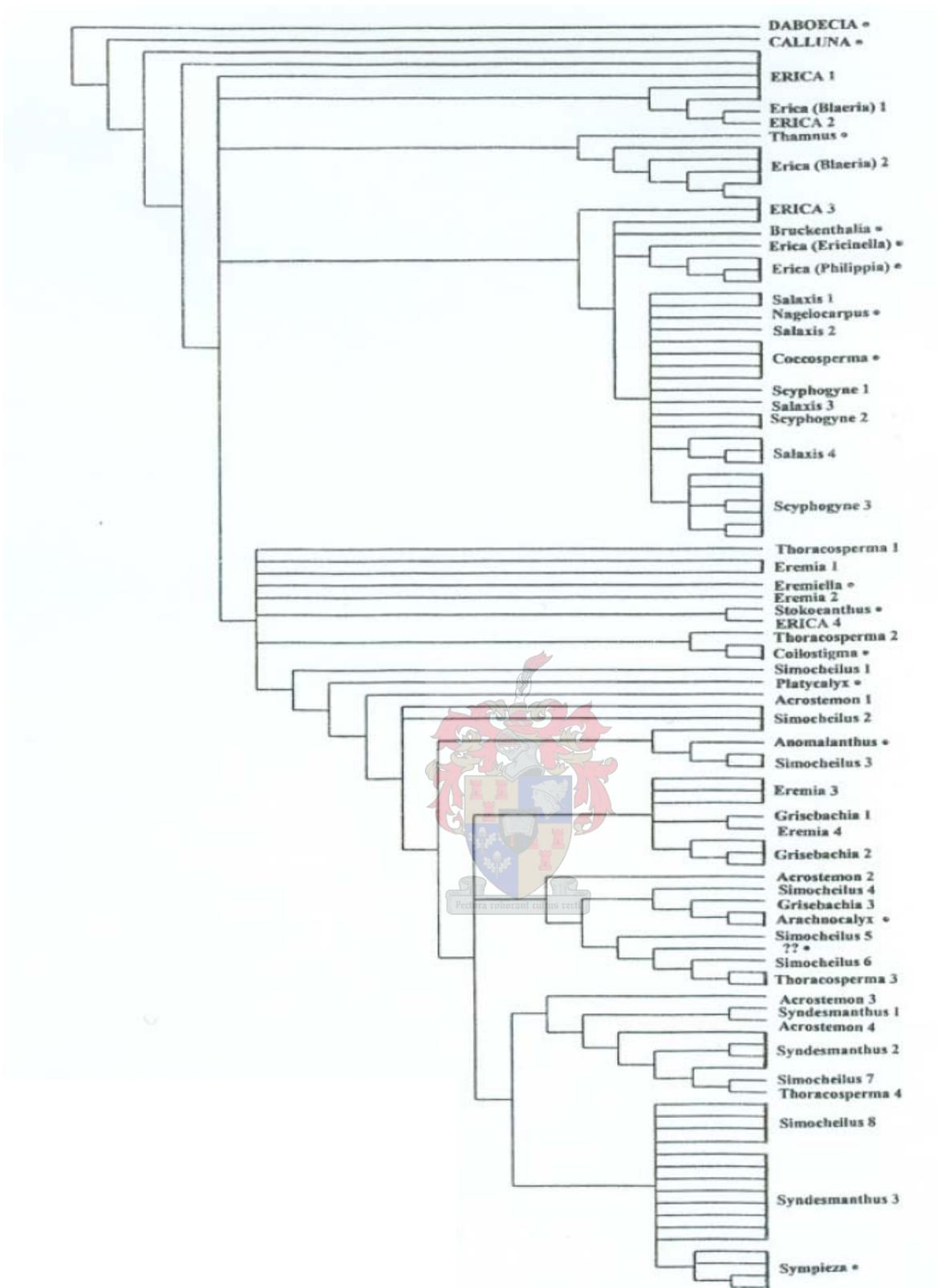


Figure 2.10: The Ericaceae phylogenetic tree by Oliver (2000), which shows the considerable degree of homoplasy in the group.

Studies into the molecular phylogenetic relationships amongst *Erica* species based on molecular studies have thus far only been preliminary. McGuire and Kron (2005) assessed the relationships between the African taxa and the European taxa. In this study the phylogenetic relationships were based on information from the nuclear ITS region, *matK* and the *rbcL-atpB* spacer. The results of this study (Figure 2.11) showed that the African taxa are descendents of a European ancestor. The African taxa form a monophyletic group nested in a grade of European taxa. Though the combined tree shows some resolution in the African clade, very few nodes are significantly supported.

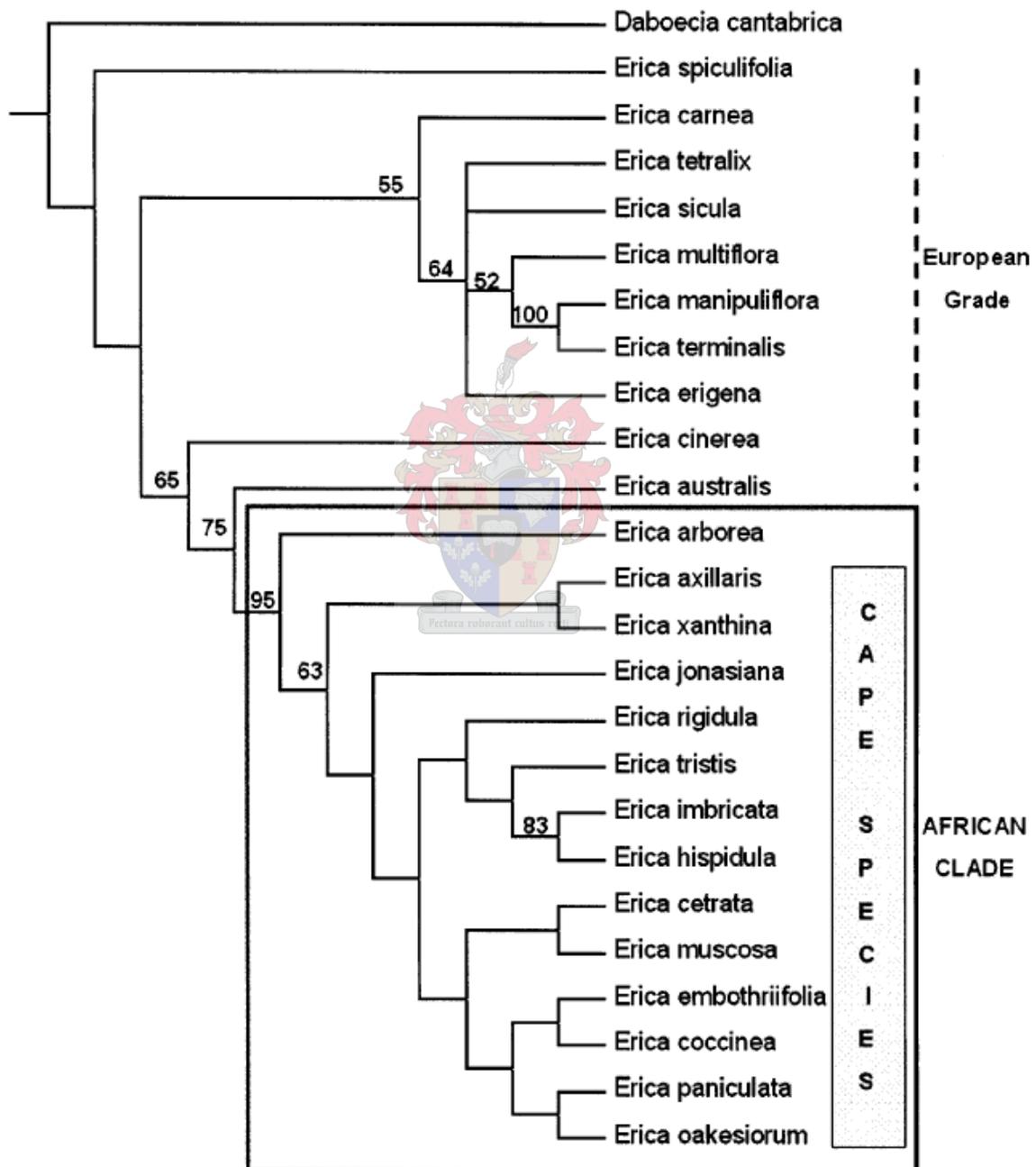


Figure 2.11: The most parsimonious tree of the combined chloroplast and ITS data sets from McGuire and Kron (2005).

The establishment of a phylogeny of the Cape mega-genus *Erica*

Though rapid advances have been made in molecular phylogenetics, the goal of accurate reconstruction of phylogenies remains a challenge. A major difficulty stems from the incongruence between gene phylogenies and the underlying organismal phylogeny. Many studies have reported on the phylogenies of various plants at different levels. Recently many such studies have also been completed for many CFR taxa.

The investigation into phylogenetics and relationships of *Erica* poses a particular challenge. The size of the genus itself in the CFR is super-large, making it seem to be a very ambitious task to embark upon, which will take forever to complete. A few key factors mentioned above play a pivotal role when embarking on studies of such calibre. The choice of suitably variable gene regions, amongst others, will be a crucial decision on which the successful phylogenetic reconstruction of the genus will be critically dependent.

The present study utilises DNA sequences from several non-coding plastid regions and the ITS region of the nuclear genome to determine whether there is sufficient variability in these gene regions to give adequate phylogenetically informative characters that may be useful at the species level assessment of the CFR mega-genus *Erica*. The study itself is not aimed at answering questions relating to the relationships within the genus, but rather tries to answer which regions should preferentially be used to reconstruct the *Erica* molecular phylogeny.

Therefore, the resulting phylogenetic trees that will be presented here will potentially give a clear idea about the informativeness of the gene regions based on their tree topologies and statistics. Ultimately the regions that appear to have the highest PIC levels may be utilised in further studies in the quest to determine the phylogenetic relationships of the mega-genus *Erica* both in the CFR and worldwide.

Chapter 3: Preliminary investigations into the phylogenetic relationships in the genus *Erica* L

3.1 Introduction

Erica L. is a genus of about 860 species world-wide, with 700 of these found in South Africa's south western and southern Cape, making it by far the most speciose genus in the Cape Floristic Region (CFR). Recent extensive morphological studies aimed at resolving the relationships within the genera of the *Ericaceae* in Southern Africa (Oliver, 2000). This resulted in the recognition of a single genus *Erica*. Unfortunately morphological characters have proved to be uninformative in a phylogenetic framework (Oliver, 2000). This lack of phylogenetically informative morphological characters and the evidence of widespread homoplasy (Oliver, 2000) may also be exacerbated by the size of the genus. These pose a particular challenge in the construction of a phylogeny of the genus.

As a result of similar shortcomings in phylogenies of various other plant groups based on morphological classifications, molecular systematics has been employed successfully in many studies that aimed to unravel phylogenetic relationships of plant genera. This is manifested by the large amount of literature produced that deals with phylogenies and by the production of an angiosperm phylogenetic tree (APG, 1998, 2003). Molecular systematics has not only proven to be advantageous in terms of reducing labour, but has revealed new informative morphological characters that are not easily identified or that might not be identified using traditional methods. Another advantage of the molecular phylogenetic approach is that large numbers of characters can be generated for a selected taxon or data set (Soltis *et al.*, 1998).

There are a number of factors that need to be considered when working on a group as large as the genus *Erica*. In order to be certain which approach to take, certain conditions have to be optimised. These include the DNA extraction method, PCR conditions, the informativeness of the chosen gene region(s) and the choice of primers to be used. A few key studies have focused on trying to optimise these conditions. Padmalatha *et al.* (2006) and Narayanan *et al.* (2006) tried to optimise the DNA extraction protocol and PCR conditions. Similarly Sang (2002), Aoki (2003) and Shaw *et al.* (2005) tried to identify suitable DNA regions that are informative at the specific level. These studies have been crucial in recent phylogenetic studies, as their results provided us with alternatives to use at species level. This avoided the use of the same old DNA regions that had become almost universally used due to their successes at resolving higher level phylogenies, despite them often being essentially uninformative at the species level.

The importance of such studies cannot be over-emphasised. Shaw *et al.* (2005) showed that when working with a large group or genus, one does not necessarily need to use large sample numbers

to determine the informativeness of a region. The use of up to three species can be representative enough to reveal the usefulness of a particular nuclear or chloroplast region. Obviously this is very advantageous (and cost effective) in large scale studies or in genera with high species numbers. Shaw *et al.* (2005) identified a number of regions that can be used to assess the informativeness and relative utility in plants at low taxonomic levels. They mentioned a number of chloroplast gene regions that have not commonly been employed in the past, which may be used successfully to reconstruct species level phylogenies and thus determine relationships at lower levels. These studies laid down the necessary groundwork needed to serve as guidelines when working with large groups.

The aim of the present study was therefore to use a small selection of taxa from the genus *Erica* and selected DNA regions, both chloroplast and nuclear, to determine which of these regions would be sufficiently variable to give adequate phylogenetically informative characters that may be useful when trying to reconstruct the phylogenetic relationships within the genus *Erica*. Eight chloroplast regions, *trnL-F*, *matK*, *trnS-G*, *rps12-rpl20*, *psbA-trnH*, *trnC-D*, *rps4-trnT* and *trnT-L*, and the nuclear ITS region were thus amplified for a subset of 30 *Erica* species. The informativeness and utility of these gene regions were assessed by running parsimony analyses and comparing the resultant tree topologies.

3.2 Materials and Methods

3.2.1 Plant sampling

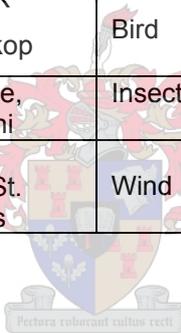
A subset of 30 species representing the range of morphological diversity, geographical spread and pollinator preference within the genus were chosen. All of the species included in this study were collected as either fresh plant material or dried in silica gel from plants growing in their natural habitats. Most samples were collected and all identified by Dr. E. G. H. Oliver. Two species that were included as potential outgroups, namely *Erica arborea* and *Erica trimera* were also sampled. They are both tropical African species. The included taxa and their pollinator preference are tabulated in Table 3.1. No herbarium voucher specimens were made as most of the samples had already been deposited in the Compton Herbarium, Kirstenbosch National Gardens, South Africa.



Table 3.1: The taxa included in the comparative molecular phylogenetic study of *Erica*, along with their geographical locations, pollinator preference and collection numbers

Species name	Location	Pollinator preference	Collection number
<i>E. amatolensis</i> L.	E. Cape Katberg	Insect/Wind	EGH Oliver 12111
<i>E. amidae</i> E.G.H.Oliv.	CFR Gordon's Bay	Bird	EGH Oliver 12352 Type
<i>E. annalis</i> E.G.H.Oliv.	CFR Kammanasie Mtns	Bird	EGH Oliver 11929
<i>E. arborea-1</i> L.	Kenya, Nakuru, Hell's gate National Park	Insect	Bytebier B 2335
<i>E. arborea-2</i> L.	Zurich Botanical Garden	Insect	Bytebier B 2686
<i>E. atherstonei</i> Diels ex Guthrei & Bolus	Mpumalanga Lydenberg	Insect	EGH Oliver 122262
<i>E. calycina</i> L.	CFR Jonaskop	Insect	EGH Oliver 7564
<i>E. cerinthoides</i> L.	CFR Hout Bay	Bird	EGH Oliver 12271
<i>E. cereris</i> E.G.H.Oliv.	CFR Ceres	Insect	EGH Oliver 9794
<i>E. chrysocodon</i> Guthrei & Bolus.	CFR Franschhoek Pass	Insect	EGH Oliver 12391
<i>E. cooperi</i> Bolus.	E. Cape Naude's nek	Insect	EGH Oliver 12144
<i>E. curviflora</i> L.	CFR	Bird	Turner 773
<i>E. denticulate</i> L.	CFR Jonaskop	Insect	EGH Oliver 11424
<i>E. dracomontana</i> E.G.H.Oliv.	E. Cape Barkly east	Wind	EGH Oliver 12130
<i>E. drakensbergensis</i> Guthrei & Bolus.	KZN, Tugela gorge	Insect	EGH Oliver 12181
<i>E. evansii</i> E.G.H.Oliv.	KZN, Bulwer	Wind	EGH Oliver 12171
<i>E. glabella</i> Thunb.	CFR Steenberg	Insect	EGH Oliver12278a
<i>E. globiceps</i> E.G.H.Oliv.	CFR Bredasdorp	Insect	Bytebier B 2652
<i>E. hillburtii</i> E.G.H.Oliv.	E. Cape Elliott	Wind	EGH Oliver 12125
<i>E. interrupta</i> E.G.H.Oliv.	CFR Pearly Beach	Insect	EGH Oliver 8755
<i>E. junonia</i> Bolus.	CFR Cold Bokkeveld	Insect	EGH Oliver 12109
<i>E. lutea</i> P.J.Bergius.	CFR Jonaskop	Insect	EGH Oliver 7588

<i>E. monsoniana</i> L.	CFR Jonaskop	Insect/ Bird	EGH Oliver 11372
<i>E. oatesii</i> Rolfe var.	KZN, Tugela gorge	Bird	EGH Oliver 12183
<i>E. plumose</i> Thunb.	CFR Mamre	Insect	EGH Oliver 12393
<i>E. rosacea</i> (L.Guthrie) E.G.H.Oliv.	CFR Swartberg Pass	Insect	EGH Oliver 12394
<i>E. schlechteri</i> Bolus.	E. Cape Barkly east	Insect	EGH Oliver 12116
<i>E. solandri</i> Andrews.	CFR George Outeniqua mnts.	Insect	EGH Oliver 11895
<i>E. tenuifolia</i> L.	CFR Jonaskop	Insect	EGH Oliver 7562
<i>E. trimera</i> L.	Ethiopia	Wind	Meihe s.n.
<i>E. tristis</i> Bartl.	CFR Kalk Bay	Wind	EGH Oliver 12271a
<i>E. tumida</i> Ker.Gawl.	CFR Cold Bokkeveld	Bird	EGH Oliver 12110
<i>E. vestita</i> Thunb.	CFR Jonaskop	Bird	EGH Oliver 8982
<i>E. woodii</i> Bolus.	E. Cape, Ntsikeni	Insect	T. A. Oliver 6
<i>E. zeyheriana</i> E.G.H.Oliv.	CRF Cape St. Francis	Wind	Turner 781



3.2.2 DNA Extraction

Total genomic DNA was extracted from half of the plant samples (+/- 15 samples) using the CTAB method of Doyle and Doyle (1991). Fresh leaf material (0.5 – 1.0 g) or 0.2 g of silica dried tissue was ground with the traditional mortar and pestle method in liquid nitrogen to snap-freeze the tissue. The ground material was placed in a 1.5 ml Eppendorf tube and 500 µl of CTAB extraction buffer (+ 0.2% β-mercaptoethanol) was added to the mix and placed in a 60°C heating block for 45 minutes. An equal volume of chloroform: isoamylalcohol (24:1 v/v) was added and mixed properly by inversion for 10 min and centrifuged at 3500 x g for 5 min. The supernatant was carefully decanted and transferred to a new tube, precipitated with equal volumes of cold isopropanol and gently mixed to produce fibrous DNA. This was incubated at -20°C for a minimum of 30 min. The samples were centrifuged at 650 x g for 5 min. The pellet was washed with wash buffer (1 part ammonium acetate and 3 parts ethanol), air dried and resuspended in 500 µL of TE buffer.

Subsequently, the DNA was precipitated for a second time. To this end, 200 µl of the TE buffer resuspended sample was mixed with 400 µL of water and 300 µL of 7.5 M of NH₄Ac and then 2.25 mL of ice cold ethanol. The mixture was incubated at -20°C for 30 min followed by centrifugation at

10,000 x g for 10 min. The pellet was air dried and resuspended in TE buffer. All the centrifugation steps were carried out at room temperature to avoid precipitation with CTAB, DNA degradation and to obtain good quality DNA.

A modified procedure of the above was used as an alternative DNA extraction method. For the other half of the samples (+/- 15 samples) plant tissue was ground up with the aid of the Qiagen tissue lyser in 2 X CTAB extraction buffer to which double the amount of β -mercaptoethanol (0.4%) and a spatula tip of Polyvinylpyrrolidone (PVP) (0.5 mg) was added. The subsequent steps of the DNA extraction were performed as above. The second precipitation using NH_4Ac mixed with H_2O and cold ethanol was also done. Purified DNA samples were finally resuspended in 1 \times TE and stored at 4 °C.

3.2.3 PCR and DNA Sequencing

Several chloroplast DNA regions, including *trnL-F*, *matK*, *trnS-G*, *rps12- rpl20*, *psbA-trnH*, *trnC-D*, *rps4-trnT* and *trnT-L* were amplified. The nuclear ITS region was also amplified. All PCR reactions were done using the Thermohybrid® PX2 thermal cycler and the Labnet International, Inc Multigene II PCR systems. Various PCR parameters were set for the varying gene regions as suggested in Shaw *et al.* (2005) in agreement with the calculated annealing temperatures of the various gene regions, which were determined in some cases by running a gradient of annealing temperatures in PCR. The PCR parameters are described below. A key to the shorthand for these PCR parameters is as follows: initial denaturing step (temperature, time); number of repetitions of the amplification cycle X (denaturing temperature, time; primer annealing temperature, time; chain extension temperature, time)]; final extension step (temperature, time). All reactions ended with a final 15°C hold step.

trnL-trnF: was amplified using primers ***trnL5^{UAA}*** (**TabC**) (CGA AAT CGG TAG ACG CTA CG) (Taberlet *et al.*, 1991) and ***trnF^{GAA}*** (**TabF**) (ATT TGA ACT GGT GAC ACG AG) (Taberlet *et al.*, 1991) with the parameters: No denaturing temperature; 35 cycles X (94°C, 1 min; 50°C, 1 min; 72°C, 2 min); 72°C, 5 min.

matK: was amplified using the primers ***matK (1F)*** (ATG GAG GAA TTC AAA AGA AAT TTA G), and the ***matK (1600R)*** (CCT CGA TAC CTA ACA TAA TGC) (McGuire and Kron, 2005), with the PCR parameters 80°C, 5 min; 35 cycles (95°C, 1 min; 50°C, 1 min; 65°C, 5 min); 65°C, 10 min.

trnS-G: was amplified using ***trnS*** (AGA TAG GGA TTC GAA CCC TCG GT) and ***trnG*** (GTA GCG GGA ATC GAA CCC GCA TC) (Hiratsuka *et al.*, 1989; Jansen and Palmer, 1987), with the parameters 80°C, 5min, 35 cycles X (96°C, 45 sec; 52°C, 45 sec; 72°C, 1 min); 72°C, 10 min.

rps12- rpl20: was amplified and sequenced using primers ***rps12*** (ATT AGA AAN RCA AGA CAG CCA AT) and ***rpl20*** (CGY YAY CGA GCT ATA TAT CC) (Hamilton, 1999). Amplification

parameters were 96°C, 5 min; 35 cycles X (96°C, 1 min; 50–55°C, 1 min; 72°C, 1 min); 72°C, 5 min.

psbA-trnH: The PCR parameters for this region were 80°C, 5 min; 35 cycles X (94°C, 30 s; 50–56°C, 30 s; 72°C, 1 min); 72°C, 10 min with primers **trnH^{GUG}** (CGC GCA TGG TGG ATT CAC AAT CC) (Tate and Simpson, 2003) and **psbA** (GTT ATG CAT GAA CGT AAT GCT C) (Sang *et al.*, 1997).

trnC-D: This region was amplified and sequenced using primers **trnC^{GCA}** (CCA GTT CRA ATC YGG GTG) and **trnD^{GUC}** (GGG ATT GTA GYT CAA TTG GT) (both primers modified from Demesure *et al.* (1995)). The PCR parameters for this region were 95°C, 1 min; 35 cycles X (95°C, 30 sec; 57°C, 45 sec; 72°C, 2 min); 72°C, 10 min.

rpS4-trnT: The primers that were used for this region are **rpS4R2** (CTG TNA GWC CRT AAT GAA AAC G), **trnTR** (AGG TTA GAG CAT CGC ATT TG) (both primers by Saltonstall (2001)), with the parameters 80°C, 5 min; 35 cycles X (94°C, 1 min; 50°C, 1 min; 72°C, 2 min); 72°C, 5 min.

trnT-L was amplified using primers **trnT (TabA)** (CAT TAC AAA TGC GAT GCT CT) (Taberlet *et al.*, 1991) and **trnL (TabB)** (TCT ACC GAT TTC GCC ATA TC) (Taberlet *et al.*, 1991) with the parameters 35 cycles X (94°C, 1 min; 55°C, 1 min; 72°C, 1 min); 72°C, 10 min.

The nuclear **ITS** was first amplified using **ITS 5p** (GGA AGG AGA AGT CGT AAC AAG G) and **ITS4** (TCC TCC GCT TAT TGA TAT GC) (Baldwin, 1992) with the parameters 94°C for 5 min; 30 cycles (94°C, 1 min; 55°C, 1 min; 72°C, 1 min); 72°C, 10 min, which amplified with success, but could not be sequenced. **AB101** (ACG AAT TCA TGG TCC GGT GAA GTG TTC G) (White, 1990) and **8P** (CAC GCT TCT CCA GAC TAC A) (Baldwin, 1992) were subsequently used to amplify the region and the combination of **8P** and **2G** (GTG ACG CCC AGG CAG ACG T) (Yokota *et al.*, 1989) was used to sequence the region as a result of AB101 also failing to sequence the region.

The amplification of all the regions was done in 100 µl PCR reactions with the following reaction components: 4 µl of template of total genomic DNA, 10X buffer (JMR-Holdings, USA), 2.5 mM of MgCl₂, (JMR-Holdings, USA) 4 µl dNTPs (BioLine), 1 unit of *Taq* (Supertherm) (JMR-Holdings, USA) and 100 ng of each primer. Some reactions included Bovine Serum Albumin with a final concentration of 0.2 mg/mL to improve amplification. The amplified PCR products were purified using Promega Wizard[®] SV gel and PCR cleaning system according to the manufacturer's protocol. The samples were then concentrated from 80 µl to 20 µl of the sample volume using the Savant[®] speed vacuum concentrator.

Cycle sequencing reactions were carried out in the forward and reverse direction in a Multigene II PCR system using 10 µl reactions with the following components: the sequencing buffer (BigDye[®]), Terminator mix (BigDye[®] Terminator V3.1. Cycle sequencing Kit) (Applied Biosystems), the designated sequencing primers and the purified DNA template strand. Each cycle consisted of

96°C denaturation for ten seconds, 52°C annealing for thirty seconds and 60°C extension for four minutes. The products were purified and detected on an automated capillary sequencer (ABI 3100 Genetic Analyser) at the Central Analytical Facility, Stellenbosch University.

An additional part to this study was done by Angeline Khunou of the Leslie Hill Molecular Systematics Laboratory at the Kirstenbosch Research Centre, as a comparative measure to determine the informativeness of the nuclear ITS region, in determining the level of informativeness amongst closely related *Erica coccinea* complex. This part of the study followed the same protocol as outlined above and the sequences were also included in the matrix that was used to build the phylogeny tree based on the nuclear ITS.

3.2.4 Sequence Assembly and Phylogenetic Analyses

Sequence assembly and editing were performed using Chromas version 2.23 (Technelysium) and BioEdit Sequence Alignment Editor. All matrices were aligned using the ClustalW multiple alignment tool in BioEdit and further edited by eye. All the matrices that were produced were analysed using parsimony analyses in PAUP* version 4.01 b 10 (Swofford 2002).

3.2.5 Data matrices and alignments from GenBank

A number of sequences of the *matK* and the nuclear ITS regions that were generated by McGuire and Kron (2005) were downloaded from GenBank. These were included in the present study to assess the informativeness and the variability in different regions over the spectrum of *Erica* samples from the present study and those that were previously investigated. Table 3.2 lists these samples and their accession numbers, collectors and their various localities adapted from McGuire and Kron (2005).

Table 3.2: Specimens analysed and sequenced by McGuire and Kron (2005) that were downloaded from GenBank and included in the present study.

Species	GenBank accession number		Specimen/voucher	Geographical location
	ITS	<i>matK</i>		
<i>Daboecia cantabrica</i> C. Koch		AY520786	RBGE 1975–1770	Europe
<i>E. arborea</i> L.	AY520788	AY517907	David Small.s.n.	S. Africa
<i>E. australis</i> L.	AY520789	U61329	RBGE 781912	Europe
<i>E. axillaris</i> Salisb.		AY517925.	E.G.H.Oliver s.n.	Africa
<i>E. carnea</i> L.	AY520790	AY517908	David Smalls.n.	Europe
<i>E. cetrata</i> E.G.H.Oliver	AY520787	AY517906	E.G.H.Oliver s.n	Africa
<i>E. cinerea</i> L.	AY520791	AY517909	David Small s.n.	Europe

<i>E. coccinea</i> L	AY520792	AY517910	E.G.H.Oliver s.n.	Africa
<i>E. embothriifolia</i> (Salisb). E.G.H.Oliver.	AY520793	AY517911	E.G.H.Oliver s.n.	Africa
<i>E. erigena</i> R.Ross	AY520794	AY517912	David Small s.n.	Europe
<i>E. hispidula</i> (L.) E.G.H.Oliver	AY520795	AY517913	E.G.H.Oliver s.n.	Africa
<i>E. imbricata</i> L.	AY520796	AY517914	E.G.H.Oliver s.n.	Africa
<i>E. jonasiana</i> E.G.H.Oliver.	AY520797	AY517915	E.G.H.Oliver s.n.	Africa
<i>E. manipuliflora</i> Salisb.	AY520798	AY517916	David Small s.n.	Europe
<i>E. multiflora</i> L.	AY520799	AY517917	David Small s.n.	Europe
<i>E. muscosa</i> (Aiton) E.G.H.Oliver.	AY520810	AY517926	E.G.H.Oliver s. n.	Africa
<i>E. oakesiorum</i> E.G.H.Oliver	AY520800	AY517918	E.G.H.Oliver s.n.	Africa
<i>E. paniculata</i> L.	AY520801	AY517919	E.G.H.Oliver s.n.	Africa
<i>E. plukenetii</i> L.	AY520802	AY517920	E.G.H.Oliver s.n.	Africa
<i>E. rigidula</i> (N. E. Br.) E.G.H.Oliver	AY520811	AY517927	E. G. H. Oliver s.n.	Africa
<i>E. sicula</i> Guss.	AY520804	U61341	RBGK 1985–2612	Europe
<i>E. spiculifolia</i> Salisb.	AY520785	U61387	RBGK 1969–52067	Europe
<i>E. terminalis</i> L.	AY520805	AY517921	L David Small s.n.	Europe
<i>E. tetralix</i> L.	AY520806	U61340	AA 195–79	Europe
<i>E. tristis</i> Bartl.	AY520809	AY517924	E. G. H. Oliver s.n.	Africa
<i>E. xanthina</i> Guthrie & Bolus.	AY520808	AY517923	E. G. H. Oliver s.n.	Africa
Key: AA = Arnold Arboretum, RBGE = Royal Botanical Garden at Edinburgh, RBGK = Royal Botanical Garden at Kew, ITS = internal transcribed spacer.				

3.2.6 Parsimony Analyses

In the analyses of all the matrices, the nucleotides were treated as unordered characters with equal weighting (Fitch parsimony; Fitch, 1971). Heuristic searches used 1 000 replicates of random taxon addition and tree bisection-reconnection (TBR) branch swapping with a limit of five trees saved during each replicate to reduce the time spent swapping on islands of equally parsimonious trees.

The consistency (CI) and retention (RI) indices, which provide an indication of the measure of fit between the data and tree topologies, were calculated for each analysis. Nonparametric bootstrap analyses (Felsenstein, 1985) were performed in PAUP* to assess the internal support for each of the nodes, as well as a heuristic search strategy with 1000 replicates (with one of the shortest trees saved per replicate), simple taxon addition and TBR branch swapping. In a phylogeny that is completely resolved with n number of taxa, the potential number of nodes is $n-1$. As a measure of comparison of the informativeness of the various gene regions, the number of nodes with $\geq 75\%$ bootstrap support was expressed as a fraction of the potential number of nodes in the phylogeny.

3.2.7 Tree statistics, topologies, the combined tree and Partition heterogeneity test

A log file of each data set was stored that included the tree statistics for all regions on which parsimony analyses were run. The tree statistics included the consistency (CI) and retention (RI) indices, the number of trees built, the number of variable characters, the number of constant characters, the parsimony informative characters and the characters used in each analysis. For all the trees that were built, an assessment of the tree topologies and the measure for congruency between the various data sets were done.

3.3 Results

3.3.1 DNA extraction

The efficiency of DNA extraction was assessed through the subsequent success of PCR amplifications of the *trnLF* gene region. These results showed that the original extraction that was done using the Doyle & Doyle (1987) method with mortar and pestle grinding was largely unsuccessful. The modified extraction method yielded better results with the yield in DNA improved as compared to extraction yields with the Doyle & Doyle (1987) method. The addition of double the amount of β -mercaptoethanol and PVP improved the DNA extraction results.

The silica dried plant material seemed not to give enough DNA to be used for amplification. The fresh plant material gave better yields of DNA, thus implying that the use of fresh material takes preference over the use of silica dried material. The usage of the Qiagen® Tissue lyser to extract samples improved the sample DNA yield and was used in further experiments with both silica dried and fresh material. Table 3.2 shows all the samples from which DNA was extracted, the various gene regions that were employed in this study and the results of the extraction methods, amplifications and sequencing reactions. Most of the samples that gave successful results represented fresh plant material.

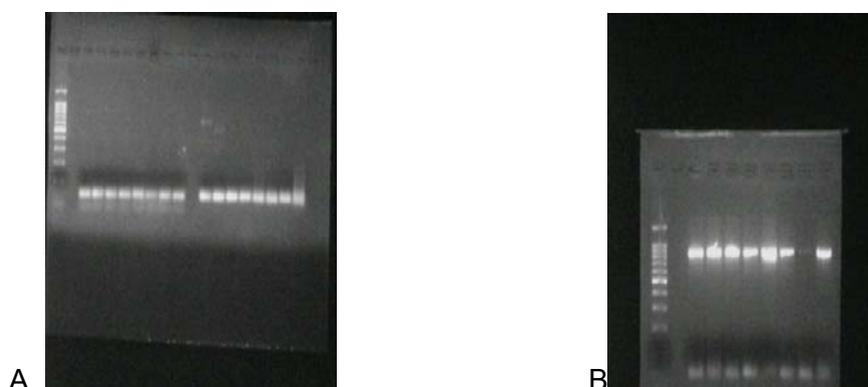


Figure 3.1: (A) 2% agarose gel of DNA extracted using original extraction method with mortar and pestle (B) 2% agarose gel of some DNA extracted in (A) using the tissue lyser, PVP and double the amount of β -mercaptoethanol, both after PCR with the same primer pair under identical conditions.

3.3.2 The chloroplast regions

Only four of the chloroplast gene regions that were used (*trnL-F*, *matK*, *trnS-G* and *rps12-rpl20*) gave amplification products in all the taxa that were studied. The remaining regions could only be amplified in some taxa, and then only gave very faint bands when viewed on a 2% agarose gel. This did not yield enough DNA product to be sequenced or further analysed. The regions that amplified and were sequenced with success are discussed in detail below:

***trnL-F* data**

The *trnL-F* region (consisting of the *trnL* intron and the *trnL-F* spacer region) could be sequenced without ambiguities and aligned with success, with the data set including 21 taxa and a total of 1103 characters. Sequence data were trimmed to eliminate autapomorphic ends. Of the 799 characters that were included in the analysis, 713 were constant, 86 (10.76%) were variable of which 27 (3.38%) were parsimony informative. Using the parsimony algorithm to find the tree scores showed that the number of trees that were generated was 7515 and the tree length was 433, with CI = 0.861 and RI = 0.717. One of the 7515 most parsimonious trees is shown in Figure 3.2 (p. 44). The bootstrap values of over 50% (bold and italicised) and branch length values (bold) are shown above and below the branches, respectively. Only four of the possible 20 nodes were retrieved with bootstrap support values of over 75%.



***matK* data**

Two data sets from this region were analysed, first the data that were produced from the results of the present study followed by the second which included the accessions from GenBank to form a data matrix that could be used as a measure of the variability and informativeness of the region over a wider spectrum of *Erica* samples. The results from the two data sets are presented below:

In the first data matrix and analysis, which included only accessions from the present study, the *matK* dataset was sequenced without ambiguities and aligned with success, with the data set including 18 taxa and a total of 958 characters in the analysis. Sequence data were trimmed to eliminate autapomorphic ends. 657 characters were included in the analysis, of which 554 characters were constant, 103 (15.67%) were variable of which 27 (4.09%) were parsimony informative. When using the parsimony algorithm to find the tree scores 15 trees were generated, each with a tree length of 120, a CI = 0.883 and a RI = 0.674. The strict consensus tree of the 15 trees is shown in Figure 3.3 (p. 45). The bootstrap values of over 50% (bold and italicised) and

branch length values (bold) are shown below and above the branches, respectively. Only 2 of the possible 17 nodes were retrieved with bootstrap support values of over 75%.

In the analysis with the combined data matrix of the accessions from the present study and those from GenBank, the *matK* data set included 42 taxa and a total of 1690 characters. Sequence data were trimmed to eliminate autapomorphic ends. 701 characters were included in the analyses, of which 504 were constant, 197 (28.10%) were variable and 58 (8.27%) were parsimony informative. Also using the parsimony algorithm to find the tree scores, 255 trees with a tree length of 288 were generated with CI = 0.809 and RI = 0.613. The strict consensus tree of the 255 trees is shown in Figure 3.4 (p. 46). The bootstrap values of over 50% (bold and italicised) and branch length values (bold) are shown below and above the branches, respectively. Only four of the possible 41 nodes were retrieved with bootstrap support values of over 75%.

***trnS-G* data**

The *trnS-G* amplification products could also be sequenced without ambiguities and aligned with success, with the data set including 15 taxa and a total of 821 characters. Sequence data were trimmed to eliminate autapomorphic ends, resulting in 483 characters included in the analysis, of which 423 were constant, 60 (12.42%) were variable and 17 (3.52%) were parsimony informative. Also using the parsimony algorithm to find the tree scores 2107 trees with a tree length of 72 were generated with CI = 0.889 and RI = 0.652. The strict consensus tree of the 2107 trees is shown in Figure 3.5 (p. 47). The bootstrap values of over 50% (bold and italicised) and branch length values (bold) are shown below and above the branches, respectively. Only one of the possible 14 nodes had bootstrap support values over of 75%.

Table 3.2: Summary of results from amplification and sequencing experiments, √√ = Good results, √ = Fair results and 0 = no results

SPECIES NAME	DNA EXTRACTION	<i>trn</i> L-F				ITS							<i>matK</i>			
		PCR	Cycle sequencing		Aligned	PCR		cycle sequencing			Aligned	PCR	Cycle sequencing		Aligned	
			C & F	C		F	5P & 4	AB101 & 8P	8P	2G			5P	4		1f & 1600R
<i>E. arborea</i>	√	√√	√√	√√	√√	√	√√	√√	√√	0	0	√√	0	0	0	0
<i>E. atherstonei</i>	√	√	0	0	0	√	√	√√	√	0	0	√	√√	√√	√√	√√
<i>E. woodii</i>	√	√√	√√	√√	√√	√	√	0	√	0	0	0	√√	√√	√√	√√
<i>E. hillburtii</i>	√	0	0	0	0	√	√	0	0	0	0	0	√√	√√	√√	√√
<i>E. cooperi</i>	√	0	0	0	0	√	√	√√	√√	0	0	√√	0	0	0	0
<i>E. dracomontana</i>	√	0	0	0	0	√	√	√√	√√	0	0	√√	√√	√√	√√	√√
<i>E. tumida</i>	√	0	0	0	0	√	√	0	0	0	0	0	√√	√√	√√	√√
<i>E. cereris</i>	√	√	0	0	0	√	√	0	0	0	0	0	√√	√√	√√	√√
<i>E. trimera</i>	√	√√	√√	√√	√√	√	√	0	0	0	0	0	√√	0	0	0
<i>E. zeyheriana</i>	√	0	0	0	0	√	√	0	0	0	0	0	0	0	0	0
<i>E. drakensbergensis</i>	√	0	0	0	0	√	√	√√	√√	0	0	0	√√	0	0	0
<i>E. amatolensis</i>	√	√√	√√	√√	√√	√	√	√√	√√	0	0	√√	√√	0	0	0
<i>E. solandri</i>	√	0	0	0	0	√	√	√√	√√	0	0	0	√√	0	0	0
<i>E. oatesii</i>	√	0	0	0	0	√	√	√√	√√	0	0	√√	√√	0	0	0
<i>E. schlechteri</i>	√	0	0	0	0	√	√	0	√	0	0	0	√√	0	0	0
<i>E. junonia</i>	√	0	0	0	0	0	0	0	0	0	0	0	√√	0	0	0
<i>E. annalis</i>	√	√√	√√	√√	√√	√	√	√√	√√	0	0	0	0	0	0	0
<i>E. evansii</i>	√	√√	√√	√√	√√	√	√	√√	√√	0	0	√√	0	0	0	0
<i>E. glabella</i>	√	√√	√√	√√	√√	√	√	√√	√√	0	0	√√	√√	√√	√√	√√
<i>E. chrysocodon</i>	√	√√	√√	√√	√√	√	√	√√	√√	0	0	√√	√√	√√	√√	√√
<i>E. plumosa</i>	√	√√	√√	√√	√√	√	√	√√	√√	0	0	√√	0	0	0	0
<i>E. interrupta</i>	√	√√	√√	√√	√√	√	√	√√	√√	0	0	√√	√√	0	√√	√√
<i>E. curviflora</i>	√	√√	√√	√√	√√	√	√	√√	√√	0	0	√√	√√	√√	√√	√√
<i>E. globiceps</i>	√	√	√	√√	√√	√	√	√√	√√	0	0	√√	0	0	0	0
<i>E. rosacea</i>	√	√	√	√√	√	√	√	√√	√	0	0	√	√√	√√	√√	√√
<i>E. tristis</i>	√	√	√	√√	√√	√	√	√√	√√	0	0	√√	√√	0	√√	√√
<i>E. cerinthoides</i>	√	√	√	√√	√√	√	√	√√	√√	0	0	√√	√√	0	√√	√√
<i>E. monsoniana</i>	√	√	√√	√√	√√	√	√√	√√	√√	0	0	√√	√√	0	√√	√√
<i>E. vestita</i>	√	√	√√	√√	√√	√	√√	√√	√√	0	0	√√	√√	0	√√	√√
<i>E. tenuifolia</i>	√	√	√√	√√	√√	√	√√	√√	√√	0	0	√√	√√	√√	√√	√√
<i>E. denticulata</i>	√	√	√	√√	√	√	√	√√	√	0	0	√	√√	√√	√√	√√
<i>E. calycina</i>	√	√	√√	√√	√√	√	√√	√√	√√	0	0	√√	√√	0	0	0
<i>E. lutea</i>	√	√	√√	0	0	√	√√			0	0	0	0	0	0	0
<i>E. amidae</i>	√	√	√√	√√	√√	√	√√	√√	√√	0	0	√√	√√	√	√	√

Table 3.2 cont.

SPECIES NAME	DNA EXTRACTION	<i>trnS-G</i>				<i>rpS12-rpL20</i>				<i>trnC-D</i>			
		PCR	Cycle sequencing		Aligned	PCR	cycle sequencing		Aligned	PCR	Cycle sequencing		Aligned
		S & G	S	G		<i>rpS12-rpL20</i>	<i>rpS12</i>	<i>rpL20</i>		C & D	C	D	
<i>E. arborea</i>	√	0	0	0	0	√√	√√	√√	√√	0	0	0	0
<i>E. atherstonei</i>	√	0	0	0	0	√√	√√	√√	√√	0	0	0	0
<i>E. woodii</i>	√	0	0	0	0	√√	√√	√√	√√	0	0	0	0
<i>E. hillburtii</i>	√	0	0	0	0	√√	√√	√√	√√	0	0	0	0
<i>E. cooperi</i>	√	0	0	0	0	√√	0	0	0	0	0	0	0
<i>E. dracomontana</i>	√	0	0	0	0	√√	0	0	0	0	0	0	0
<i>E. tumida</i>	√	0	0	0	0	√√	0	0	0	0	0	0	0
<i>E. cereris</i>	√	0	0	0	0	√√	0	0	0	0	0	0	0
<i>E. trimera</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. zeyheriana</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. drakensbergensis</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. amatolensis</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. solandri</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. oatesii</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. schlechteri</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. junonia</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. annalis</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. evansii</i>	√	0	0	0	0	√√	√√	0	√√	0	0	0	0
<i>E. glabella</i>	√	√√	√√	√√	√√	√√	√√	0	√√	0	0	0	0
<i>E. chrysocodon</i>	√	√√	√√	√√	√√	√√	√√	√√	√√	0	0	0	0
<i>E. plumosa</i>	√	0	0	0	0	√√	0	0	0	0	0	0	0
<i>E. interrupta</i>	√	√√	√√	√√	√√	√√	√√	√√	√√	0	0	0	0
<i>E. curviflora</i>	√	√√	√√	√√	√√	√√	√√	√√	√√	0	0	0	0
<i>E. globiceps</i>	√	0	0	0	0	√√	√√	√√	√√	0	0	0	0
<i>E. rosacea</i>	√	√√	√√	√√	√√	√√	√√	√	√	0	0	0	0
<i>E. tristis</i>	√	√√	√√	√√	√√	√√	√√	√√	√√	0	0	0	0
<i>E. cerinthoides</i>	√	√√	√√	√√	√√	√√	√√	√√	√√	0	0	0	0
<i>E. monsoniana</i>	√	√√	√√	√√	√√	√√	√√	√√	√√	0	0	0	0
<i>E. vestita</i>	√	√√	√√	√√	√√	√√	√√	√√	√√	0	0	0	0
<i>E. tenuifolia</i>	√	0	0	0	0	√√	√√	√√	√√	0	0	0	0
<i>E. denticulata</i>	√	√√	√√	√√	√√	√√	√√	√	√	0	0	0	0
<i>E. calycina</i>	√	√√	√√	√√	√√	√√	√√	√√	√√	0	0	0	0
<i>E. lutea</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. amidae</i>	√	√√	√√	√√	√√	√√	√√	√√	√√	0	0	0	0

Table: 3.2 cont.

SPECIES NAME	DNA EXTRACTION	<i>trnT-L</i>				<i>rps4-trnT</i>				<i>psbA-trnH</i>			
		PCR	Cycle sequencing		Aligned	PCR	cycle sequencing		Aligned	PCR	Cycle sequencing		Aligned
		Tab a & b	Tab a	Tab b		<i>rps4 & trnT</i>	<i>rps4</i>	<i>trnT</i>		<i>psbA & trnH</i>	<i>psbA</i>	<i>trnH</i>	
<i>E. arborea</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. atherstonei</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. woodii</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. hillburtii</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. cooperi</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. dracomontana</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. tumida</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. cereris</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. trimeria</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. zeyheriana</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. drakensbergensis</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. amatolensis</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. solandri</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. oatesii</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. schlechteri</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. junonia</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. annalis</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. evansii</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. glabella</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. chrysocodon</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. plumosa</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. interrupta</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. curviflora</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. globiceps</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. rosacea</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. tristis</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. cerinthoides</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. monsoniana</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. vestita</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. tenuifolia</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. denticulata</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. calycina</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. lutea</i>	√	0	0	0	0	0	0	0	0	0	0	0	0
<i>E. amidae</i>	√	0	0	0	0	0	0	0	0	0	0	0	0

rps12- rpl20

The *rps12-rpl20* amplification products could be sequenced without ambiguities and aligned with success, with the data set including 19 taxa and a total of 709 characters. Sequence data were trimmed to eliminate autapomorphic ends, resulting in 709 included characters, of which 675 were constant, 34 (5.03%) were variable and 9 (1.3%) were the parsimony informative. Also using the parsimony algorithm to find the tree scores, 6 trees with a tree length of 40 were generated, with CI = 0.950 and RI = 0.913. The strict consensus tree of the 6 trees is shown in Figure 3.6 (p. 48). The bootstrap values of over 50% (bold and italicised) and branch length values (bold) are shown below and above the branches, respectively. Only 1 of the possible 18 nodes had bootstrap support values of over 75%.

3.3.3 The nuclear ITS data

This region, like the *matK* region, was also analysed with two data matrices, namely the original data set from the experimental data including the *E. coccinea* complex and a second data set obtained through a combination of the sequences retrieved from GenBank and the sequences produced in the present study. The results of both these analyses follow below:

The nuclear ITS region of the accessions from the present study, including the *E. coccinea* complex, could be amplified and sequenced without ambiguities and aligned with success. The matrix included a total of 31 taxa and 1050 characters. Sequence data were trimmed to eliminate autapomorphic ends, resulting in 859 characters used in the analysis of which 694 were constant, 104 (18.5%) were variable and 61 (7.1%) were parsimony informative. The use of the parsimony algorithm in PAUP* to find tree scores resulted in 8200 trees with tree length of 259, with CI = 0.710 and RI = 0.694. As with the chloroplast data, the strict consensus tree of the 8200 equally parsimonious trees is presented in Figure 3.7 (p.49). The bootstrap values of over 50% (bold and italicised) and branch length values (bold) are shown below and above the branches, respectively. Only five of the possible 30 nodes had bootstrap support values of over 75%.

The nuclear ITS data of the accessions from the present study, including the *E. coccinea* complex, and the sequences downloaded from GenBank were used to build a complete data matrix as was done with the *matK* data. The combined data matrix could be aligned with success, resulting in a matrix of 55 taxa, including 1048 characters. Sequence data were trimmed to eliminate autapomorphic ends, resulting in 859 characters of which 569 were constant, 290 (33.7%) were variable and 138 (16.06%) were parsimony informative. The use of the parsimony algorithm in PAUP* to find tree scores resulted in 2140 equally parsimonious trees, with a tree length of 592, with CI = 0.601 and RI = 0.669. The strict consensus tree of the 2140 parsimonious trees is presented in Figure 3.8 (p.50). The bootstrap values of over 50% (bold and italicised) and branch

length values (**bold**) are shown below and above the branches, respectively. Only 13 of the possible 54 nodes had bootstrap support values of over 75%.



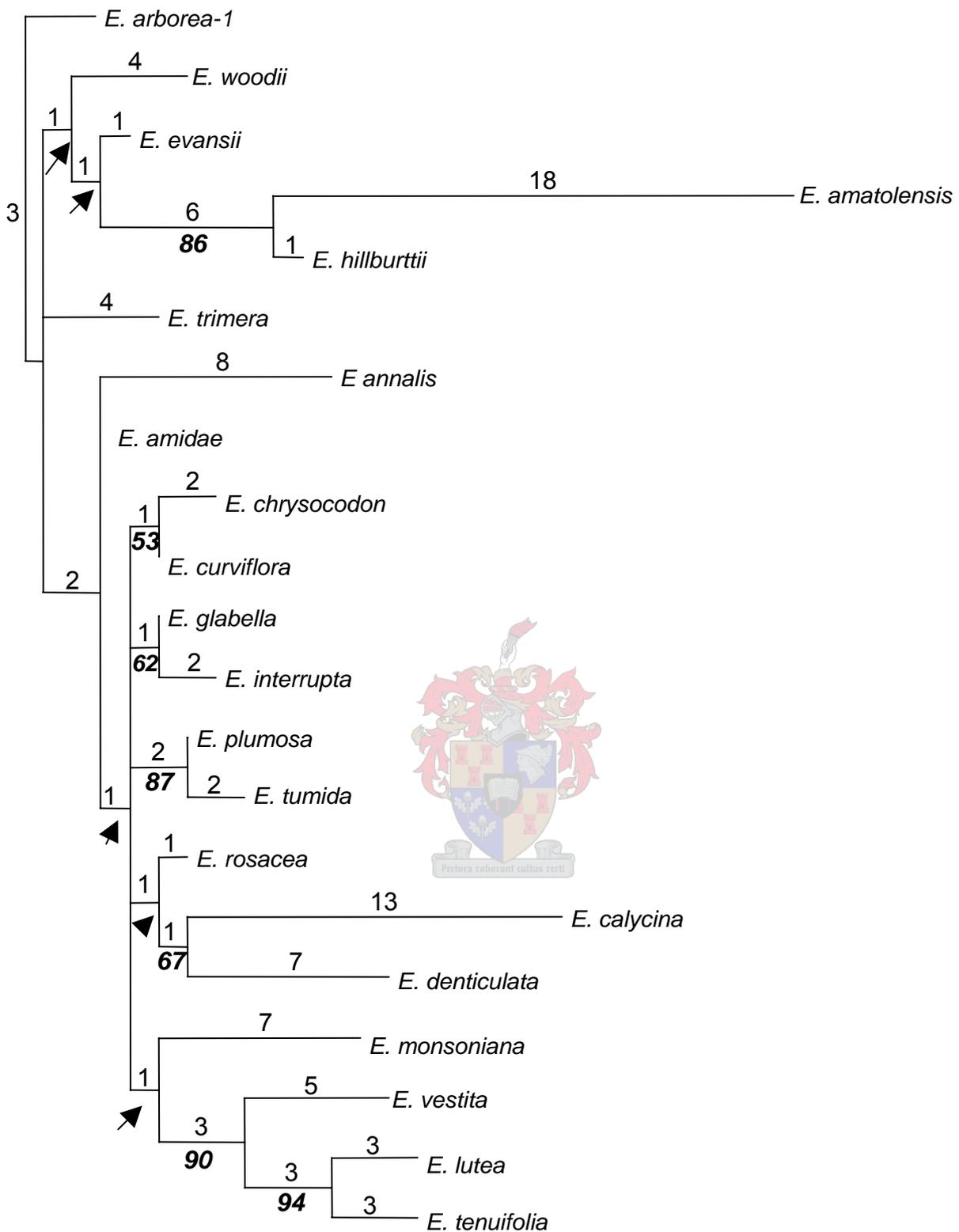


Figure 3.2: The *trnL-F* tree, represented by one of the equally parsimonious trees. Numbers above the branches indicate the branch lengths, while the numbers below (bold and italicised) indicate the bootstrap support values greater than 50%. Black arrows indicate nodes that collapsed when the strict consensus tree of the 7515 equally parsimonious trees was built. The total number of nodes is 20 and 4 of these have bootstrap support of over 75%.

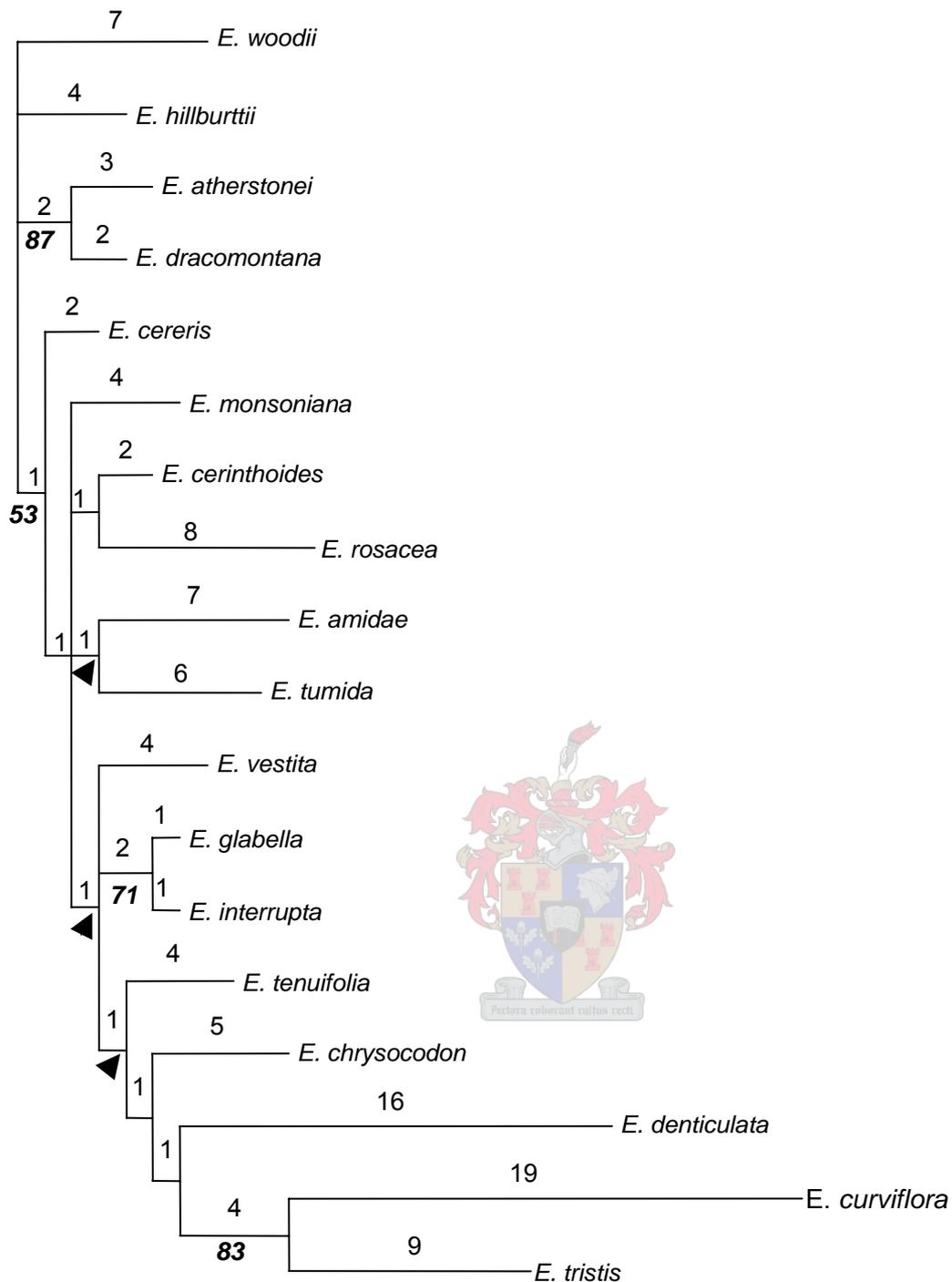


Figure 3.3: The *matK* tree built from samples used in the present study, represented by one of the equally parsimonious trees. Numbers above the branches indicate the branch lengths, while the numbers below (bold and italicised) indicate the bootstrap support values greater than 50%. Black arrows indicate nodes that collapsed when the strict consensus tree of the 15 equally parsimonious trees was built. The total number of nodes is 17 and 2 of these have bootstrap support of over 75%.

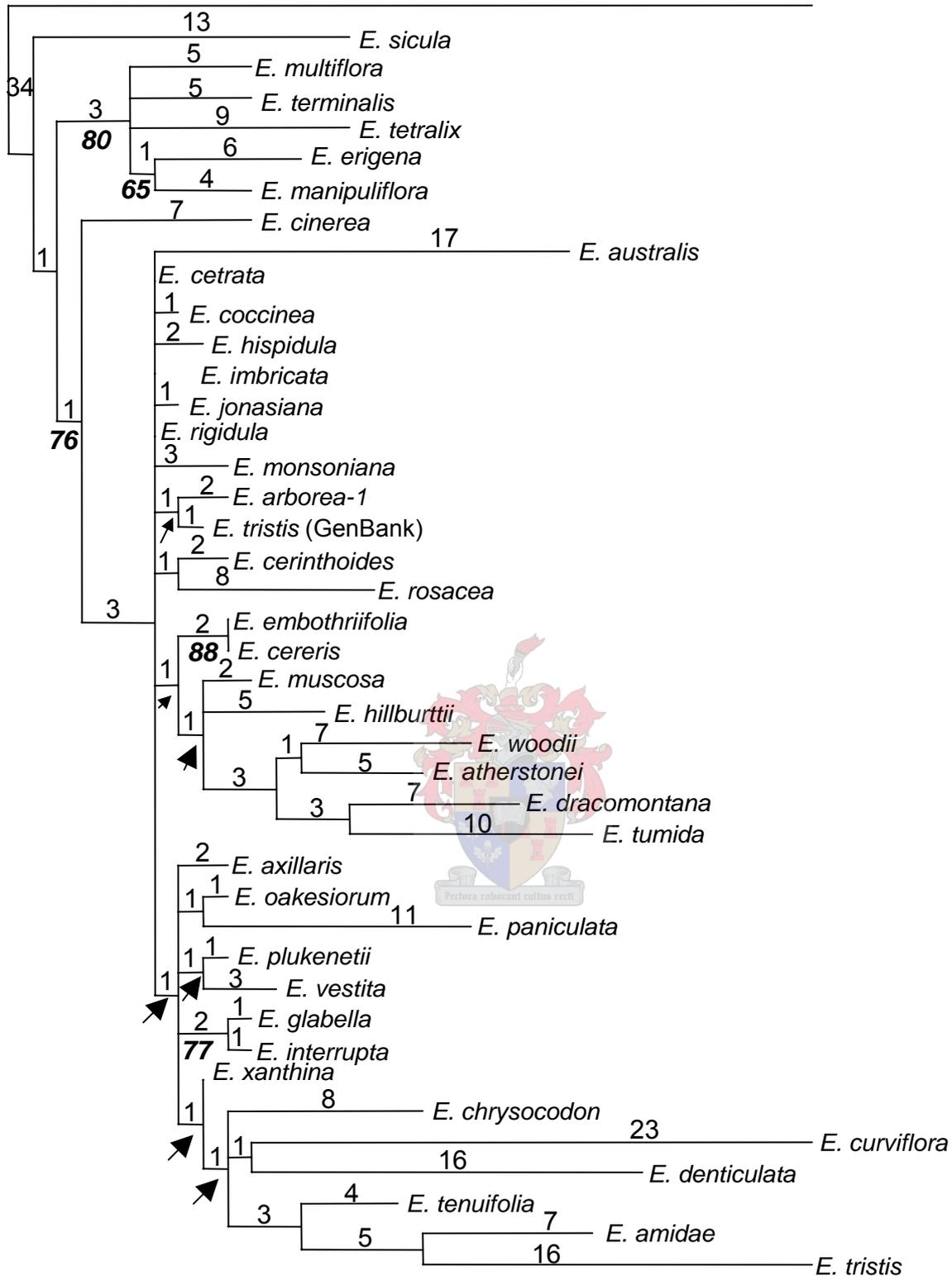


Figure 3.4: The *matK* tree built from samples used in the study plus the McGuire and Kron, (2005), accessions retrieved from GenBank, represented by one of the equally parsimonious trees. Numbers above the branches indicate the branch lengths, while the numbers below (bold and italicised) indicate the bootstrap support values greater than 50%. Black arrows indicate nodes that collapsed when the strict consensus tree of the 15 equally parsimonious trees was built. The total number of nodes is 41 and 4 of these have bootstrap support of over 75%.

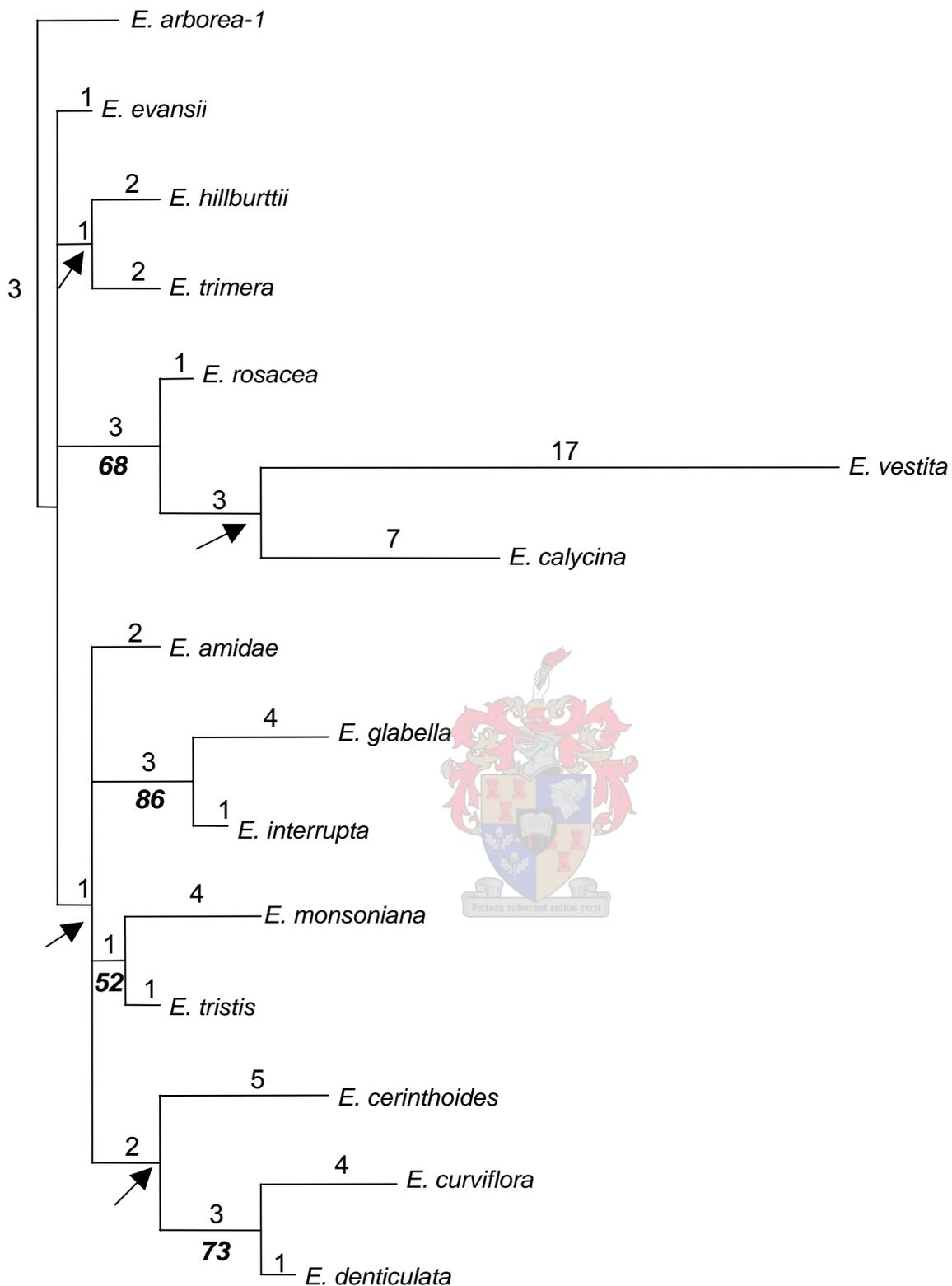


Figure 3.5: The *trnS*-G tree, represented by one of the equally parsimonious trees. Numbers above the branches indicate the branch lengths, while the numbers below (bold and italicised) show the bootstrap support values greater than 50%. Black arrows indicate nodes that collapsed when the strict consensus tree of the 2107 equally parsimonious trees was built. The total number of nodes is 14 and only 1 of these has bootstrap support of over 75%.

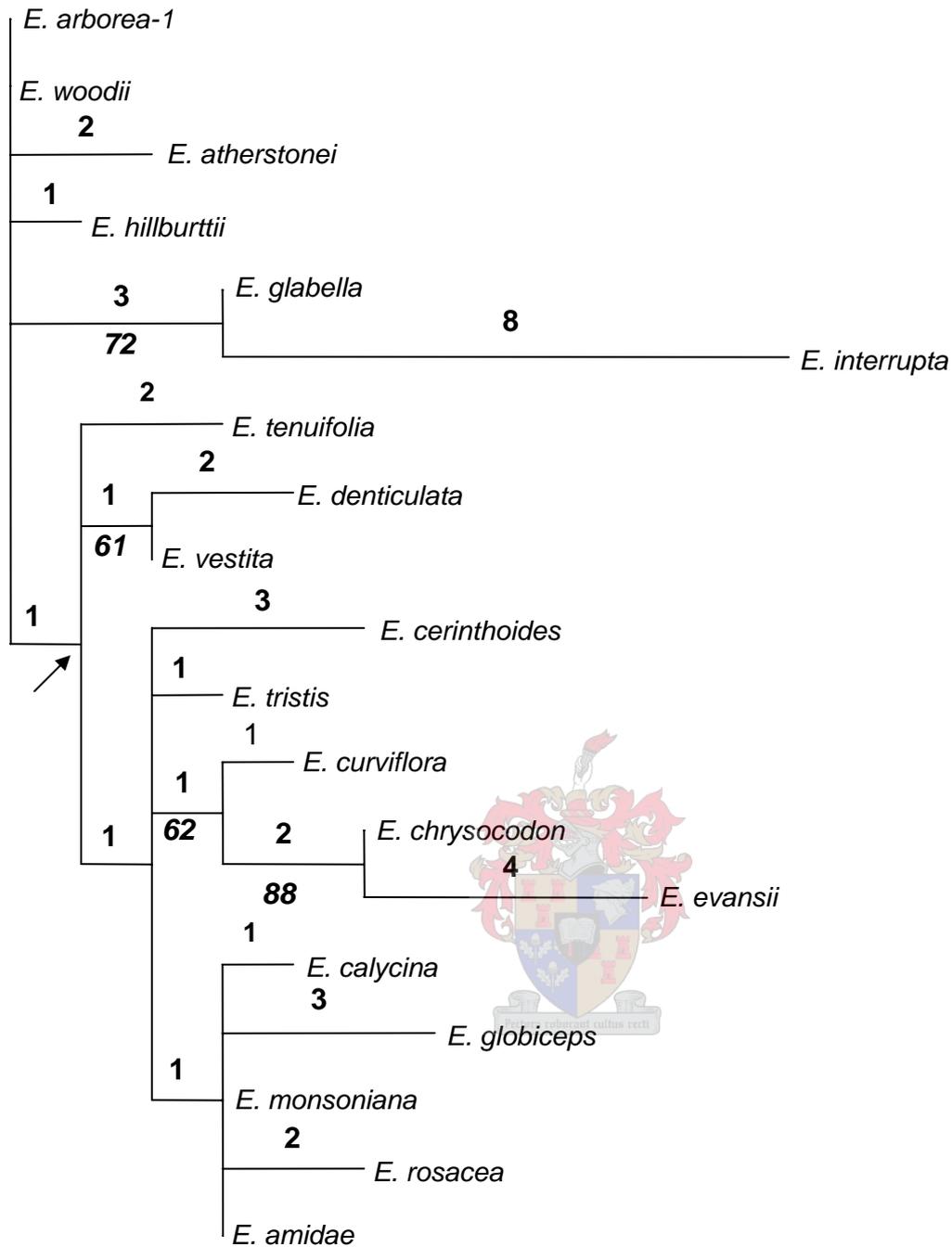


Figure 3.6: The *rpS12-rpL20* tree, represented by one of the equally parsimonious trees. Numbers above the branches (bold) indicate the branch lengths, while the numbers below (bold and italicised) indicate the bootstrap support greater than 50%. Black arrows indicate nodes that collapsed when the strict consensus tree of the 3 equally parsimonious trees was built. The total number of nodes is 18 and only 1 of these has bootstrap support of over 75%.

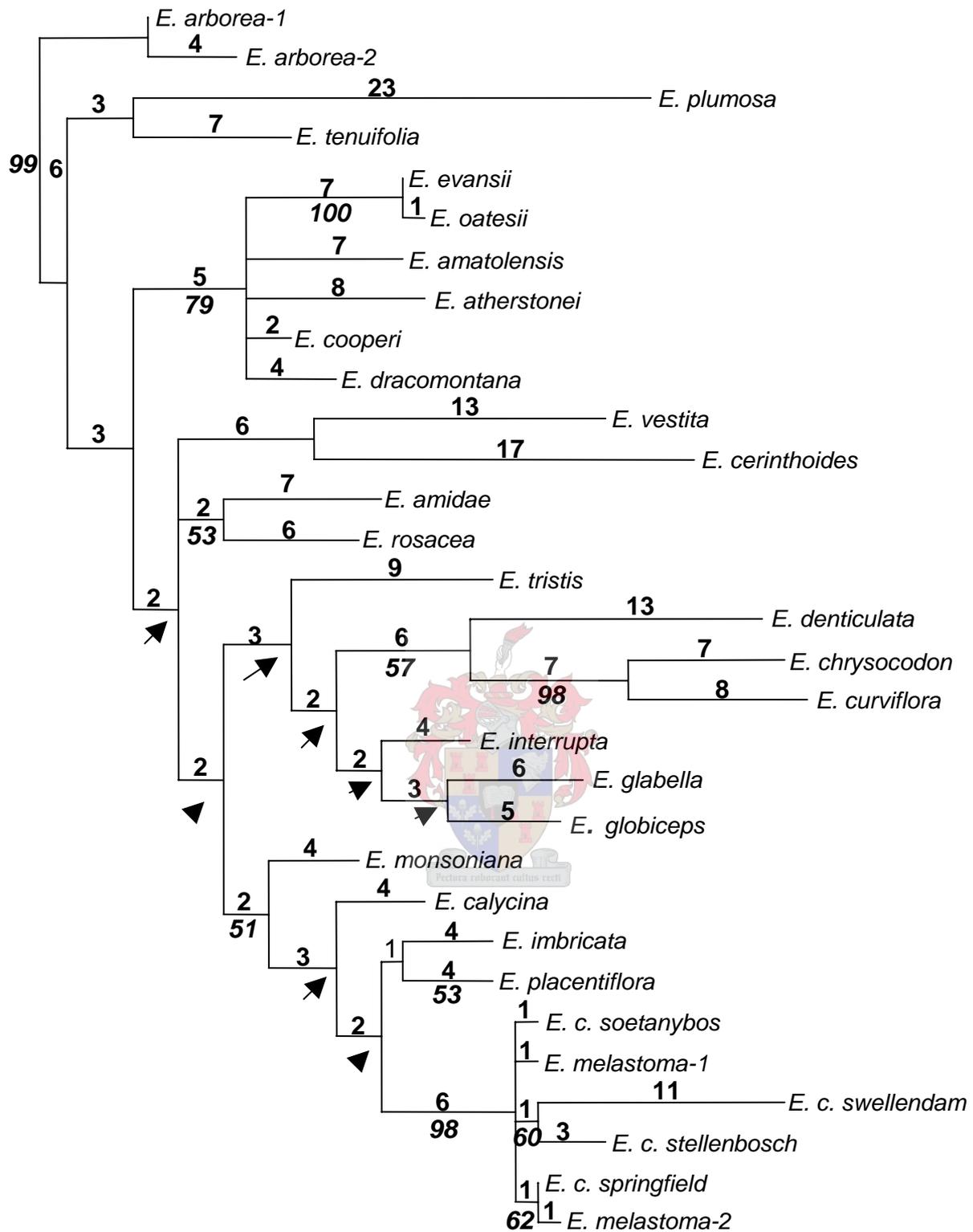


Figure 3.7: The ITS tree built from samples used in the study including the *E. coccinea* complex (*E. c.*), represented by one of the equally parsimonious trees. Numbers above the branches (bold) indicate the branch lengths, while those below (bold and italicised) show the bootstrap support values greater than 50%. Black arrows indicate nodes that collapsed when the strict consensus tree of the 8200 equally parsimonious trees was built. The total number of nodes is 30 and 5 of these have bootstrap support of over 75%.

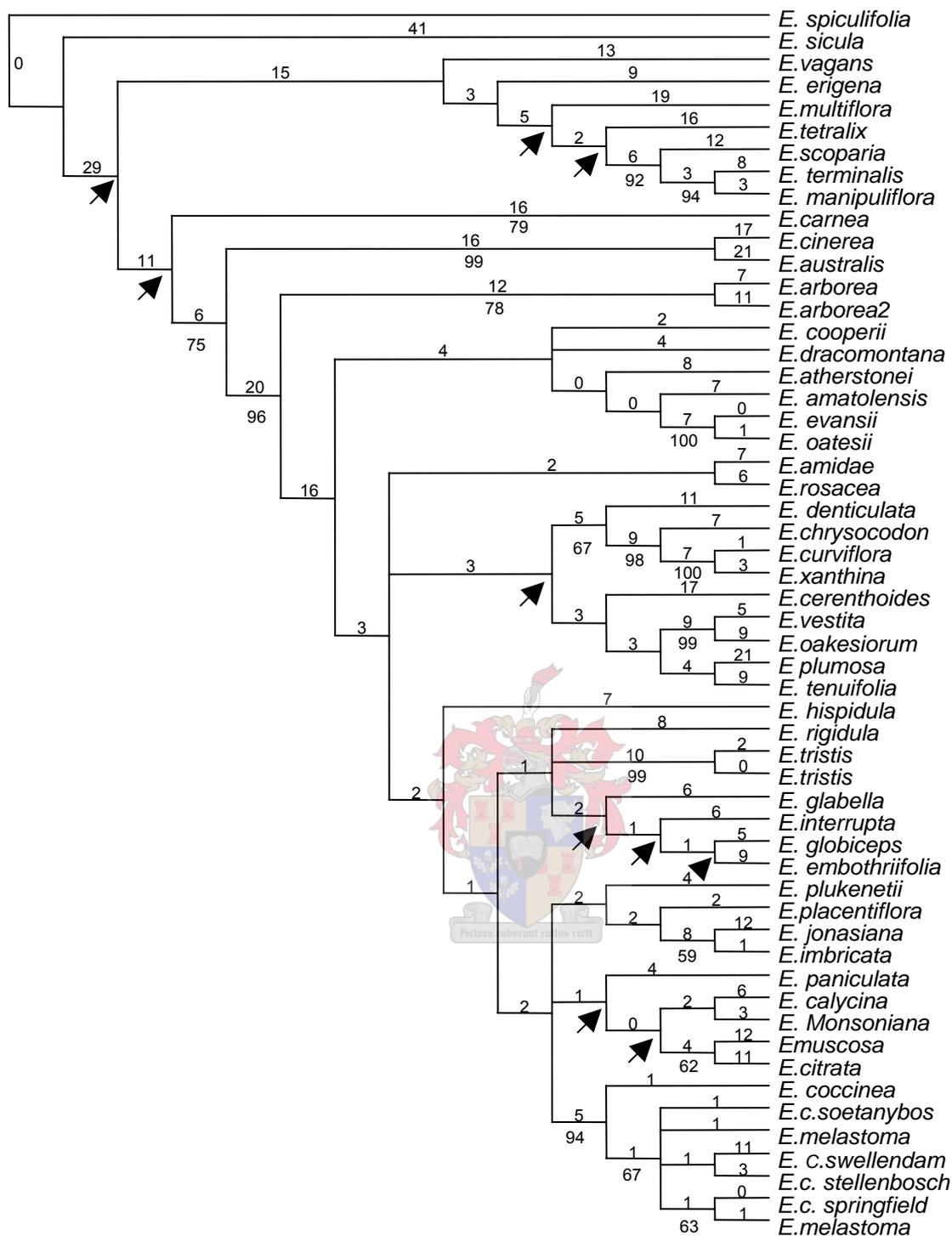


Figure 3.8: The ITS tree built from samples used in the study including the *E. coccinea* complex plus McGuire and Kron (2005) accessions retrieved from GenBank, represented by one of the equally parsimonious trees. Numbers above the branches indicate the branch lengths, while those below show the bootstrap support values greater than 50%. Black arrows indicate nodes that collapsed when the strict consensus tree of the 2140 equally parsimonious trees was built. The total number of nodes is 54 and 13 of these have bootstrap support of over 75%.

The results from the parsimony analyses and resultant trees show varying amounts of difference in the informativeness of each region and the comparative value of each of these regions. The trees inferred from the various data sets differ in terms of topology, despite the fact that they were built from data retrieved from the same organisms. When dealing with such data one should be careful, because some of these topological differences may be caused by lineage sorting and hybridization, which could be a reflection of the differences between phylogenies underlying the respective data sets (Hipp *et al.*, 2004).

From our combined data sets of the nuclear ITS region and the *matK* regions, it can be seen that an increase in the number of taxa analysed increases the informativeness of the trees and resolves the nodes to give better topologies in the regions involved. Table 3.4 summarizes all of the statistical variations in all the data sets used for the various parsimony analyses.

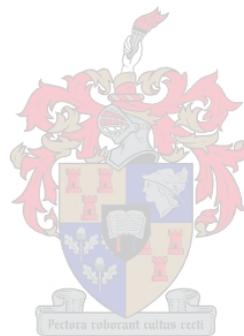


Table 3.4: The summary of tree statistics obtained through the parsimony analyses of the various data sets.

Gene region	Total number of characters	Total number of included characters	Variable characters (V)		Parsimony informative characters (P)		Number of equally parsimonious trees	Length of most parsimonious trees	Consistency index (CI)	Retention index (RI)	Nodes with bootstrap support values of $\geq 75\%$
			No.	% variability	No.	% P					
ITS	1050	859	104	18.5	61	7.1	8200	259	0.710	0.694	5/30
ITS combined data	1048	859	290	33.7	138	16.06	2140	592	0.601	0.669	13/54
<i>trnL-F</i>	1103	799	86	10.76	27	3.38	7515	433	0.861	0.717	4/20
<i>matK</i>	958	657	103	15.67	27	4.09	15	120	0.883	0.674	2/17
<i>matK</i> combined	1690	701	197	28.10	58	8.27	255	288	0.809	0.613	4/41
<i>trnS-G</i>	821	483	60	12.42	17	3.52	2107	72	0.889	0.652	1/14
<i>rpS12-rpL20</i>	882	709	34	5.03	9	1.83	6	4	0.950	0.913	1/18

Although the four chloroplast regions gave results, they are fairly uninformative except for *matK*, which showed a high percentage of potential informative characters. Of these four regions *rpS12-rpL20* was the least informative. In contrast, the nuclear ITS region had a relatively high percentage of informative characters and the data set underpinned a rather well-resolved tree. The informativeness of all regions (including both the chloroplast and the nuclear ITS) is graphically compared in Figure 3.7

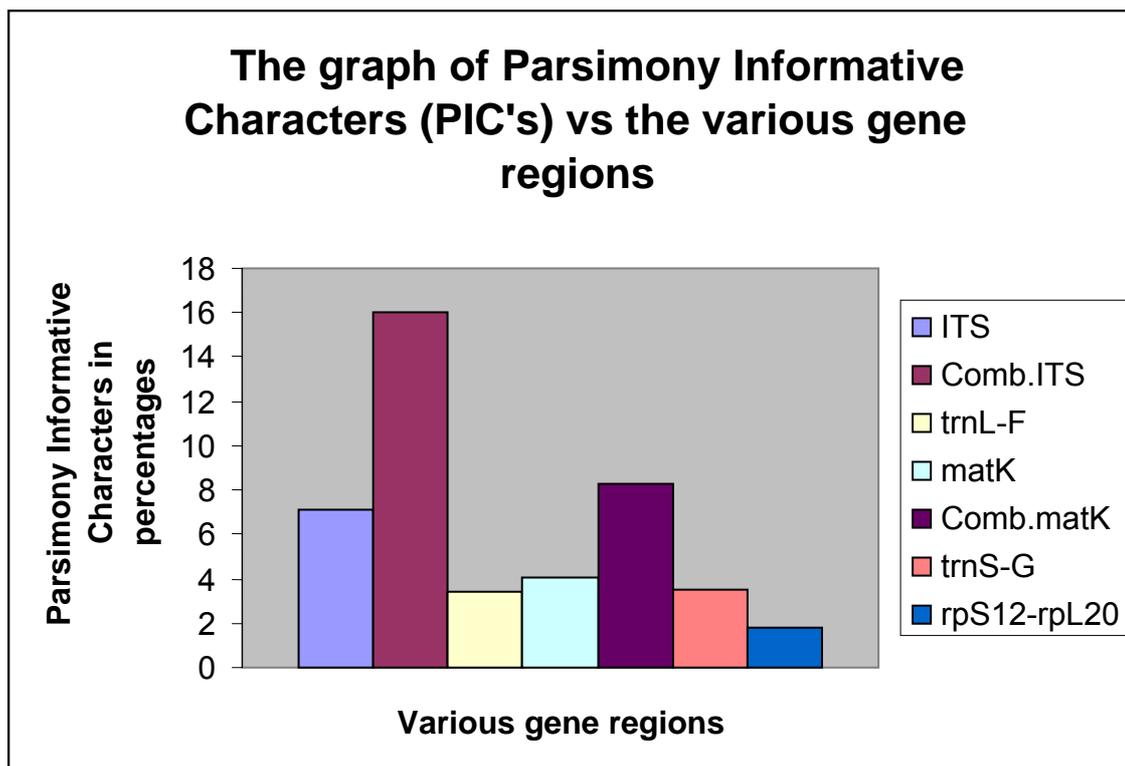


Figure 3.9: Graphical illustration of the Parsimony Informative Characters in the various gene regions for which results could be obtained in the present study.

These results show the variability (expressed as a percentage value) of Parsimony Informative Characters amongst the various gene regions. They show that, of the chloroplast regions that were sequenced, *matK* has the highest informativeness, followed by *trnS-G* and then *trnL-F*. The *rpS12-rpL20* regions were uninformative, with by far the lowest percentage of informative characters (1.83%) (Table 3.4). The combined data sets showed an increased percentage of informativeness in both the ITS and *matK* regions.

3.4 Discussion

3.4.1 DNA extraction, PCR amplification and Sequencing

In the current study both the original CTAB method and a slightly modified CTAB method were followed. Across all the taxa and gene regions that were amplified and sequenced, only approximately half could be sequenced with success. Of the eight chloroplast regions used in this study, only four could be sequenced successfully. The remaining regions either did not give amplification products or if they did amplify, the PCR products could not be sequenced without sequence ambiguities. The results showed that the addition of double the amount of β -mercaptoethanol plus PVP during extraction led to a large improvement in the quality of the

extracted DNA, as most of the PCR amplifications performed on these DNA extracts worked successfully. The reason for this may be that the modified CTAB method apparently reduces the amount of contaminants and PCR suppressors which could have been co-isolated in the original CTAB extraction method.

The superior PCR amplifications obtained from fresh samples suggest that the use of fresh plant material will give higher yields of DNA than silica dried material. It appears that prolonged storage of silica dried plant material increases the chances of DNA degradation and the acquisition of increased amounts of contaminants (Sharma *et al.*, 2002). However, this does not necessarily hold true for all samples, as some of the silica dried samples extracted with the first method gave amplification products in the PCR and could be sequenced successfully without ambiguities.

There are a number of other possible reasons for the negative PCR results, such as the primer specificity or primer binding sites having mutated. These two factors are equally responsible for the success of a PCR reaction. If the primer cannot recognize its binding site on the DNA strand, no amplification will result from the reaction. The presence of a mutation in the template strand as a result of either internal or external factors may cause the sequence of the primer binding region to change, thus resulting in either no binding or non-specific binding.

There are other factors that are also important in that they can influence the success of a PCR reaction. Such factors include denaturing time and temperature; the primer design, the primer length; elongation temperature and time; reaction buffers and cycle numbers (Rybicki *et al.*, 2001). However, because these were optimised for all the PCR's performed, they are unlikely to have played a role here. The components that are commonly altered in most reactions are the MgCl₂ concentration (which is important in the binding of the polymerase to the template strand) and the annealing temperature. The effect of the annealing temperature is such that, when it is low, binding is less specific and may allow the primer to bind to any possible sequence, even those that are not targeted. This results in mismatches and undesired results. High annealing temperatures, in contrast, are more rigorous and search for the closest match between the primer and the target DNA. This results in reduced non-specific binding and thus promotes specificity in primer binding.

This indicates that the specificity, PCR yield and successful amplification are heavily influenced by every component of the reaction and emphasizes the importance of optimizing the conditions for each reaction and the components involved.

3.4.2 DNA regions, Sequence Assembly and Phylogenetic Analyses

The cpDNA regions chosen in this study were largely based on the Shaw *et al.*, (2005) study, which used 21 chloroplast regions to assess their informativeness over a range of species. They classified the regions as ranging from the most to the least informative. The nuclear ITS region was chosen because McGuire and Kron (2005) had used it in a phylogenetic reconstruction of the

genus. The nuclear ITS region appeared to be the most informative of all the regions analysed in the study. The *matK* region also showed a high percentage informativeness (Table 3.4). The remaining regions ranged between 1% and 3% informativeness, which render them rather uninformative within *Erica*. When comparing our results with those of Shaw *et al.* (2005), our results are mostly in agreement with their results. The *trnL-F* and *rpS12-rpL20* regions were reported to be of low informativeness with regard to the plant groups used in their study, which is also seen in the employment of these regions in *Erica*. The *trnS-G* region, on the other hand, was found to be informative in the groups that Shaw *et al.* (2005) studied, but was found to be rather uninformative in *Erica*.

As the *matK* and ITS regions were the most informative with regard to building the *Erica* phylogeny, the remainder of our discussion from this point onwards will focus only on these two regions.

The *matK* gene region (*trnK-matK-trnK*), or some portion of it, was first employed in intrafamilial phylogenetic studies by Steele and Vilgalys (1994) and Johnson and Soltis (1994). This region has been employed in many studies and has been reported to be putatively the most variable coding region found within chloroplast DNA (Neuhaus and Link, 1987; Olmstead and Palmer, 1994). This region has been used in many studies to evaluate informativeness, build phylogenies and various other applications (McGuire and Kron, 2005) even including an assessment of the angiosperm phylogeny based on this region (Hilu *et al.* 2003).

Shaw *et al.* (2005), however, found the *matK* region to be amongst the least informative in their study. In contrast, results from the present study found the *matK* region to yield the highest number of variable characters of all the chloroplast regions that were sequenced, and is the most informative chloroplast region overall. The informativeness of the *matK* region in *Erica* species is further shown by the addition of sequences from GenBank to the data matrix derived from (McGuire and Kron, 2005). The combination of data through increasing the sample size two-fold proves that the region is very informative in *Erica* species. These results give *matK* preference over the other chloroplast DNA regions with regards to informativeness within *Erica* and suggest the effective utility of the region in inferring the phylogeny of the genus *Erica*. This is in direct contrast to the findings of Shaw *et al.* (2005).

Sequences of the nuclear ITS region have been used in numerous phylogenetic reconstructions, and proved the region to be informative at various taxonomic levels. In addition, this region has also been reported to yield constant results over a spectrum of species (McGuire and Kron, 2005). Sang (2002) reported that the use of nuclear genes, which usually have variable divergence rates due to their high nucleotide substitution rate, can potentially provide markers that will be

phylogenetically useful at various taxonomic levels. This was confirmed by our results in *Erica* in terms of the ITS region. The region gave a high variability of 18.5% (Table 3.4). This value is relatively high when compared to the chloroplast regions included in this study, of which the highest rate that was found in *matK* was only 4.09%. The combination of the experimentally derived sequences and the GenBank downloaded sequenced further supports the informativeness of this region.

From assessing the results of the data from the experimental ITS and *matK* sequences, we can conclude that these two regions are the most informative. The combination of our experimental data with the downloaded sequences aligned well with each other and the increase in samples had a positive influence on the informativeness of the individual gene regions. The informativeness of both ITS and *matK* were increased by the addition of samples to the matrix, which shows the positive effect of a larger sample size when trying to infer a phylogeny. The increase in the sample size may increase the number of Parsimony Informative Characters, thus increasing the informativeness of the region. The nuclear ITS data from our analyses show clearly that when the sample size was increased, the number of characters included in the parsimony analyses stayed the same, but the number of Parsimony Informative Characters increased. This confirms the proposal by Shaw *et al.* (2005) that as few as three samples of a group of species may be used to determine informativeness of a particular gene region. Our results further prove that sample size does not affect the variability or informativeness of a certain gene region; this is seen in the data from both the nuclear ITS region and the *matK* region with the inclusion of sequences downloaded from GenBank. It shows that the regions are informative and that this inclusion of more samples resulted in an increase in the variability and thus an increase in the potentially informative characters.

3.4.4 Topologies and inference from phylogenetic trees

In both the nuclear ITS and *matK* trees it was evident that each data set analysed revealed some well-supported clades and a fair amount of resolution in terms of positioning of the nodes in each tree. The *Erica arborea* samples from both Europe and Africa were treated as outgroups for most of the analyses following McGuire and Kron (2005), except in the cases where the samples did not sequence (*matK*) and where there was a combination of the experimental data and sequences downloaded from GenBank.

3.4.4.1 The *matK* region

The *matK* tree (p.45) that was built based on my experimental data matrix includes 18 taxa and is well-resolved at most of the nodes. The basal resolution is also resolved with only a few polytomies present. The region has been reputed to be very informative at these levels (Shaw *et al.* 2005).

Although our tree built from this region is not totally resolved, the amount of resolution and the tree statistics drawn from the analyses vouch for the informativeness of this region in *Erica*. The tree includes nodes that are retained in the strict consensus tree. The branch lengths indicate the amount of variability present in the trees, with some long and some short, indicating the relatedness of the included species. The tree also has reasonable bootstrap support values at the indicated nodes. With the amount of resolution present and the statistical data derived from the analyses used to build the trees, the region can be considered for future phylogenetic use in *Erica*.

The *matK* tree (p.46) that was built from our experimental data plus the sequences downloaded from GenBank includes 42 taxa, and displays a fair amount of resolution ranging from the internal nodes to the base of the tree. This tree includes *Daboecia cantabrica* as the outgroup. The tree is resolved from the base and the in-group is clearly monophyletic. The branch lengths indicate the amount of variability present in the trees, with some long and some short, indicating the relatedness of species. The tree also has bootstrap support values at the indicated nodes that show the support for the nodes with clear resolution. Although there are still a number of polytomies present in the tree, it is encouraging to see the European and African taxa resolved into two distinct clades confirming the basic results of McGuire and Kron.

The trees built from the analysis of the nuclear ITS region were the best resolved. The first tree was built from our experimental data only, and included 31 taxa. The tree depicted in Fig 3.6 (p. 49) shows that most of the nodes of this tree collapse in the strict consensus tree, thus rendering certain nodes into polytomies. One of the interesting observations is the resolution of the *E. coccenia* complex into a well supported monophyletic clade, sister to the closely related *E. imbricata* and *E. placentiflora*. This further confirms the informativeness of the region as derived from the statistical analysis.

The second ITS tree was built with a combination of our experimental data and the Shaw *et al.* (2005) sequences obtained from GenBank. This tree includes 55 taxa, and one of the resultant equally most parsimonious trees is depicted in Fig.3.6 (p. 50). This tree is rather well-resolved, and most nodes were retained when the strict consensus tree was built, resulting in fewer polytomies. The bootstrap support values also show the improved support for the nodes. Both the topology of this tree and the statistical data derived from analyses, support the notion that the nuclear ITS region is viable for future phylogenetic assessments of the genus *Erica*.

3.5 GenBank downloaded and experimental data trees

There are a number of common traits that can be observed from the phylogenies produced by data from the nuclear ITS region and the *matK* regions. The informativeness of both trees can be used to infer certain characteristics that may be associated with biogeographical patterns in *Erica*. The nuclear ITS tree displays long branches, indicating the distant relationships between individual species. *E. arborea* (which was suggested to be the link between European and African species (McGuire and Kron, 2005)) is sister to all the African taxa. This may in future be used to determine the patterns of radiation among *Erica* species, but needs to be confirmed by more inclusive analyses. It may be misleading to do so here, based on the under-sampling within our study in relation to the size of the genus.

The *matK* tree displays similar trends to those inferred from the ITS tree. In contrast to the nuclear ITS tree, the *matK* tree does not have many long branches and it is not as well-resolved as the ITS tree. This could be because *matK*, as a chloroplast region, may be more prone to hybridisation and introgression than the slowly evolving nuclear regions, or it may indicate that the group has experienced recent radiations. In this tree *E. arborea* resolves in an internal node in the African clade, with *E. australis* (another European species) also resolving into this clade, and with *E. cinerea* appearing as sister to the whole group.

Discordance between nuclear and chloroplast-based phylogenies has been detected within several plant groups (e.g. Soltis and Kuzoff (1995); Soltis, Johnson and Looney (1996); Kellogg, Appels and Mason-Gamer (1996)). The data presented here give the idea that this can be avoided if the same taxa are analysed with no missing data. This is evident from the topologies of the combined data of the nuclear ITS region and that of the *matK* region. Although these two trees may be incongruent, they give a similar message with regards to relationships amongst the taxa involved.

3.6 Conclusions

The results of this study revealed a few key factors which will affect future molecular phylogenetic studies in the *Erica*. It was found that the DNA extraction protocol employed is very important and is a crucial determinant of whether a study will be a success or not. The choice of primers employed and the conditions used to sequence these as important. The results of this study showed a number of DNA regions, both nuclear and chloroplast, can be deemed as informative and may be utilised in the investigation of the phylogenetic reconstruction. It is likely that *matK* will be of considerable phylogenetic importance in such a future study

The utilisation of other regions may also be a fruitful exercise. These may not be the only regions that may be of use in unravelling the *Erica* phylogeny. There are several other non-coding chloroplast and nuclear DNA regions that were not investigated in this study, which may be as informative as or better than those reported here.

Chapter 4: Assessment of an optimal protocol for the extraction of DNA in *Erica L.* species.

4.1 Introduction

DNA extraction is an essential part of molecular biology and plant molecular systematics. It has been shown that the quality and quantity of the extracted DNA is a crucial determinant of whether the study will be a success or not. Doyle and Doyle (1987) proposed a plant DNA extraction method which is based on the employment of CTAB as a detergent in the extraction method. This method has been largely successful in many plant groups, but in some instances inadequate. For this reason, various other methods and modifications to the CTAB methods followed.

One such method was the one proposed by Karakousis *et al.* (2003), which suggested the usage of Sarkosyl as a detergent instead of the traditional CTAB. This method does not vary much from the CTAB method, the only difference being the application of an alternative detergent. The same applies for the method proposed by Lin *et al.* (2001), which was based on SDS instead of CTAB. These methods differ with significant additions of reagents, which are viewed to affect the DNA yield in some manner or other such as the addition of high salt concentrations and increased PVP (Porebski *et al.*, 1997) and the addition of sodium bisulfite in the method of Plaschke *et al.* (1995).

Other methods of DNA extraction use phenol. Traditionally, phenol-based methods were used for DNA extraction from bacteria (Maniatis, 1982 & 1986). Prior to the common use of the CTAB method for the extraction of plant DNA, most methods were based on phenol extraction (Dellaporta *et al.*, 1983; Rogers and Bendich, 1985; Porebski *et al.*, 1997; Sharma *et al.*, 2002). However, due to the hazardous nature of phenol, most plant DNA extraction protocols are based on the CTAB method with minor modifications. However, where DNA extraction is problematic, phenol-based methods may have distinct advantages which have been exploited recently by some researchers (Karakousis *et al.*, 2003; Chakraborti *et al.*, 2006; Padmalatha *et al.*, 2006). Both Padmalatha *et al.* (2006) and Chakraborti *et al.* (2006) concluded that the addition of a phenol extraction step significantly improved the quality of the extracted DNA, but resulted in a slight reduction in DNA yield in plant species containing high levels of phenolic compounds and polysaccharides.

The DNA extraction method used in this study, i.e. a modified CTAB method in which double the amount of β -mercaptoethanol and PVP has already led to a large improvement in the yield of DNA extraction from *Erica* species and a considerable decrease in the rate of DNA degradation thereafter during long storage periods of DNA samples. However, a further improvement in DNA isolation from *Erica* specimens may be valuable in increasing the success rate and efficiency of PCR amplifications. For this reason, the extraction of DNA using a phenol-based method and the previously employed modified CTAB method were used to extract DNA from a number of *Erica* samples, both fresh and silica dried. The efficiency of the DNA extraction methods was assessed by comparing these DNA samples in PCRs of the *trnLF*, *trnS-G*, *matK* and the nuclear ITS regions.

4.2 Materials and methods

4.2.1 Plant sampling

A total of ten species representing the range of morphological diversity, geographical spread and pollinator preference, six of which were silica-dried and four of which were fresh, were chosen for inclusion in this study. Most samples were collected and all identified by Dr. E. G. H. Oliver. These taxa, geographical locations and their storage status, that is whether silica dried or fresh samples, are tabulated in Table 4.1. No herbarium voucher specimens were made as most of the samples had already been deposited in the Compton Herbarium, Kirstenbosch National Gardens, South Africa.

Table 4.1: The taxa used in the comparative extraction protocol experiments, along with their geographical locations, pollinator preferences, storage status and collection numbers.

Species	Pollinator preference	Geographical location	Storage status	Collection number
<i>E. arborea</i> L.	Kenya, Nakuru, Hell's Ggate National Park	Insect	Silica dried	Bytebier B 2335E.
<i>E. cereinthoides</i> L.	CFR Hout Bay	Bird	Fresh	EGH Oliver 12271
<i>E. evansii</i> EGH Oliv.	KZN, Bulwer	Wind	Silica dried	EGH Oliver 12171
<i>E. hillburtii</i> EGH Oliv.	E. Cape Elliott	Wind	Silica dried	EGH Oliver 12125
<i>E. interrupta</i> EGH Oliv.	CFR Pearly Beach	Insect	Fresh	EGH Oliver 8755
<i>E. junonia</i> Bolus.	CFR Cold Bokkeveld	Insect	Silica dried	EGH Oliver 12109
<i>E. rosacea</i> (L.Guthrie)EGH Oliv.	CFR Swartberg Pass	Insect	Fresh	EGH Oliver 12394
<i>E. trimera</i> L.	Ethiopia	Wind	Silica dried	Meihe s.n.
<i>E. tristis</i> Bartl.	CFR Kalk Bay	Wind	Fresh	EGH Oliver 12271a
<i>E. tumida</i> Ker.Gawl.	CFR Cold Bokkeveld	Bird	Silica dried	EGH Oliver 12110

4.2.2 DNA extraction

4.2.2.1 CTAB method

A modified version of the CTAB method (Doyle & Doyle, 1987) was used to extract DNA from the *Erica* species listed above. Plant material (0.5 – 1.0 g of fresh leaf material or 0.2 g of silica dried material) was ground with the aid of the Qiagen tissue lyser in 2 X CTAB extraction buffer with the addition of double the amount of β -mercaptoethanol (0.4%) and a spatula tip of polyvinylpyrrolidone (PVP) (+/- 0.5 mg) added to a 2 ml Eppendorf tube. The tubes were incubated

at 65°C in a heating block for 60-90 min with intermittent shaking and swirling. Subsequently an equal volume of chloroform: isoamylalcohol (24:1 v/v) was added and mixed properly by inversion for 10 min and centrifuged at 3500 x g for 5 min. The supernatant was carefully decanted, transferred to a new tube, precipitated with equal volumes of cold isopropanol and gently mixed to produce fibrous DNA which was incubated at -20°C for a minimum of 30 min. The samples were centrifuged at 650 X g for 5 min. The pellet was washed with wash buffer, air dried and resuspended in 500 µL of TE buffer.

A second precipitation of the DNA followed in which 200 µl of the TE buffer resuspended DNA sample was mixed with 400 µL of water and 300 µL of 7.5 M of NH₄Ac and then 2.25 mL of ice cold ethanol. The mixture was incubated at -20°C for 30 min followed by centrifugation at 10 000 x g for 10 min. The pellet was air dried and resuspended in TE buffer. All the centrifugation steps were carried out at RT to avoid precipitation with CTAB, DNA degradation and to obtain good quality DNA.

4.2.2.2 The Padmalatha method

The second method which was used to extract DNA was the Padmalatha *et al.* (2006) method. Plant material (0.5 – 1.0 g of fresh leaf material or 0.2 g of silica dried material) was also ground with the aid of the Qiagen tissue lyser in 2 X CTAB extraction buffer with the addition of double the amount of β-mercaptoethanol and a spatula tip of Polyvinylpyrrolidone (PVP) added to a 2 ml Eppendorf tube as in the method described above.

The tubes were incubated at 65°C in a heating block for 60-90 min with intermittent shaking and swirling. An equal volume of chloroform: isoamylalcohol (24:1) was added and mixed by inversion for 30 min and centrifuged at 10,000 x g for 15 min at RT (room temperature) to separate the phases. The supernatant was carefully decanted, transferred to a new tube and precipitated with equal volumes of cold isopropanol. This was gently mixed to produce fibrous DNA and incubated at -20°C for a minimum of 30 min. The samples were centrifuged at 10,000 x g for 15 min. The pellet was washed with 70% ethanol; air dried and resuspended in 3 ml of TE buffer and incubated overnight at 37°C.

The dissolved DNA was then extracted with equal volumes of phenol: chloroform: isoamylalcohol (25:24:1, v/v/v) after which it was centrifuged at 4,500 x g for 15 min. The aqueous layer was transferred to a fresh 15 ml tube and re-extracted with equal volume of chloroform and isoamylalcohol (24:1) and then centrifuged at 10,000 x g for 15 min.

The supernatant was transferred to a fresh tube and 1/10 volume of sodium acetate was added, followed by an equal volume of absolute alcohol and incubated at -20°C for 30 min followed by centrifugation at 10,000 x g for 15 min. The pellet was air dried and resuspended in TE buffer.

4.2.3 PCR amplification

DNA extracts from all the plant samples isolated with both of the above-mentioned methods were subsequently used in PCRs. This was done by using some of the primers that were used to investigate the relative utility of different gene regions in *Erica* species (Chapter 3). Primers c and f were used to amplify the *trnL-F* region (Taberlet *et al.*, 1991), primers 1F and 1600R to amplify the *matK* region (Taberlet *et al.*, 1991 and McGuire *et al.*, 2005), primers S and G to amplify the *trnS-G* (Hiratsuka *et al.*, 1989 and Jansen and Palmer, 1987) and the primers 8P and AB101 to amplify the nuclear ITS region (Baldwin, 1992 and White, 1990). The results of the two DNA isolation methods were compared by running the products on 2% agarose gels.

4.3 Results

4.3.1 DNA extraction methods

In both of the extraction methods employed, there was a clear pellet in the final precipitation steps indicating the presence of DNA. Comparative DNA yield and purity were not assessed by agarose gel electrophoresis, but the extraction efficiency was assessed by the efficiency of amplification reactions through PCRs.

4.3.2 PCR reactions

The results of the various PCR reactions are presented below for each gene region amplified. In each gel presented in this section, lane 1 is the DNA ladder, used to determine the size of the DNA amplification product from the PCR. Lanes 2-6 of each gel contained extracts from silica dried plant material that were used in PCRs and lanes 7-10 contained extracts from fresh plant material that were used in PCRs.

trnL-F region

The PCR amplification products of DNA samples prepared using both extraction methods using the *trnLF* gene region are shown in Figure 4.1. DNA samples isolated with the modified CTAB method were successfully amplified with the exception of *E. junonia* (Lane-6). Lanes 2-5 of this gel gave bright bands indicating the presence of large quantities of amplification product. All the DNA samples isolated with the Padmalatha *et al.* (2006) method gave lower amounts of amplification products. Lanes 1,7,8,9 and 10 showed successful amplification, but lane 2-5 showed very faint bands with lane 6 having no bands at all.

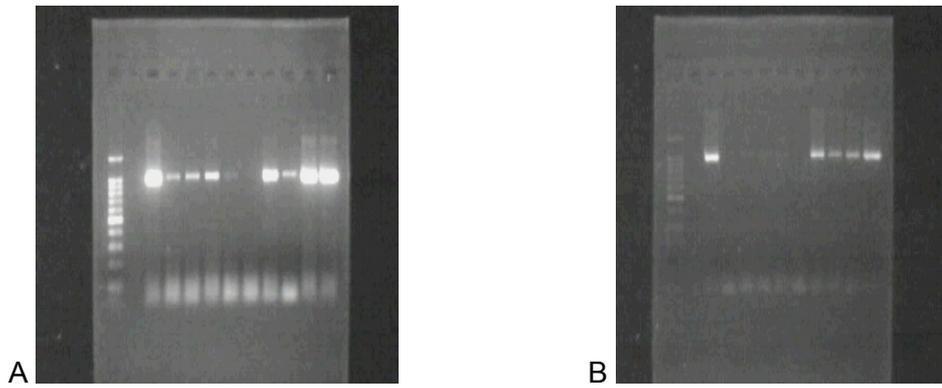


Figure 4.1: (A) The 2% agarose gel of PCR amplifications of the *trnL-F* region using DNA extracted following the original extraction method using the tissue lyser, PVP and double the β -mercaptoethanol; (B) 2% agarose gel of PCR amplifications of the *trnL-F* region of the same plant samples extracted using the tissue lyser, PVP and double the β -mercaptoethanol with addition of phenol: chloroform: isoamylalcohol.

The *matK* region

The PCR amplification products of DNA samples prepared using both extraction methods for the *matK* gene region are shown in Figure 4.2. All the DNA samples isolated with the modified CTAB method were amplified successfully, with the exception of *E. junonia* (Lane-6). Lanes 2-5 of this gel gave reasonably bright bands, indicating the presence of significant quantities of amplification product as in the amplification with the *trnL-F* shown above. In the extraction method which employed the Padmalatha *et al.* (2006) method (Figure 4.2b) only samples in lanes 1 and 10 amplified with some success, while the rest of the lanes only showed smears.

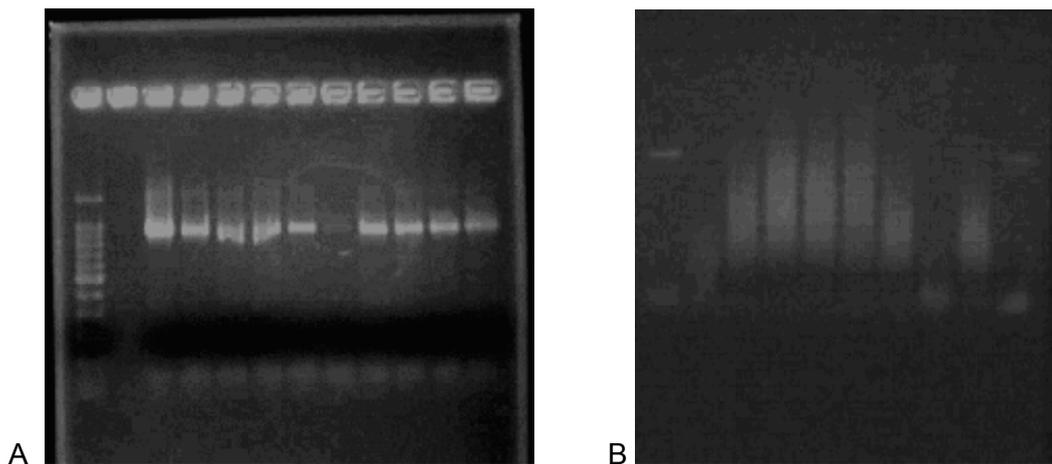


Figure 4.2: (A) The 2% agarose gel of DNA extracted with the original extraction method using the tissue lyser, PVP and double the β -mercaptoethanol; (B) 2% agarose gel of the same plant samples extracted using the tissue lyser, PVP and double the β -mercaptoethanol and addition of the phenol: chloroform: isoamylalcohol, both after PCR with the same *matK* PCR primer pair.

The trnS-G region

The PCR amplification products of DNA samples prepared with both extraction methods using the *trnS-G* gene are shown in Figure 4.3. All the DNA samples isolated with the modified CTAB method were amplified successfully, with the exception of *E. junonia* (Lane-6). Lanes 2-5 of this gel gave bright bands, indicating the presence of large quantities of amplification products. In the extraction method which employed the Padmalatha *et al.* (2006) method, (Figure 4.3b) only six of the ten samples amplified successfully. Lane 1, 7 and 10 showed the strongest bands, while lanes 2, 4, 5, 7 gave much weaker bands and the remaining lanes produced only smears or no bands at all.

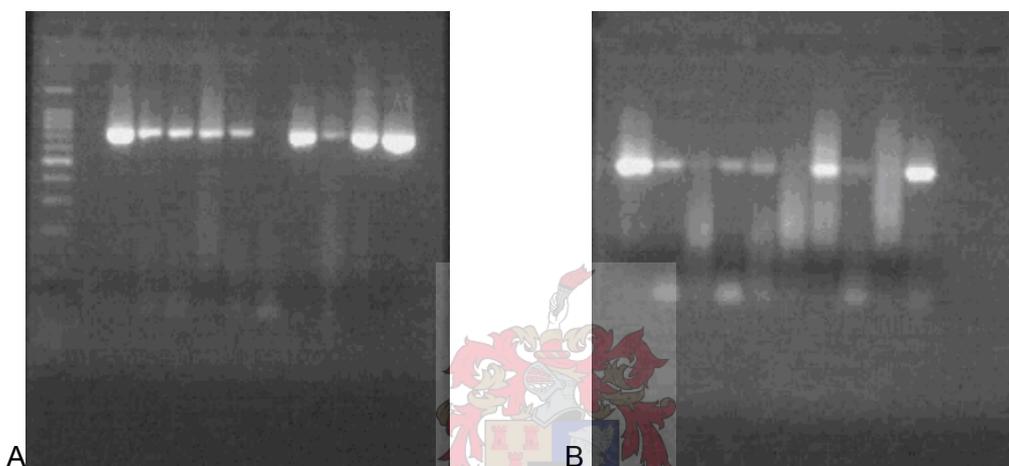


Figure 4.3: (A) The 2% agarose gel of DNA extracted with the original extraction method using the tissue lyser, PVP and double the β - mercaptoethanol; (B) 2% agarose gel of the same plant samples extracted using the tissue lyser, PVP and double the β - mercaptoethanol and added phenol: chloroform: isoamylalcohol, both after PCR with the same *trnS-G* primer pair.

The nuclear ITS region

The PCR amplification products of DNA samples prepared with both extraction methods using the nuclear ITS gene region are shown in Figure 4.4. All the DNA samples isolated with the modified CTAB method were amplified successfully, with the exception of *E. junonia* (Lane-6). Lanes 2-5 of this gel gave bright bands, indicating the presence of large quantities of DNA in the amplification products. In the extraction method which employed the Padmalatha *et al.* (2006) method (Figure 4.2b) all the samples amplified, except for the sample in lane 6 (*E. junonia*). In this gel all bands are bright and strong, with samples in lane 1, 7 and 10 giving the strongest bands.

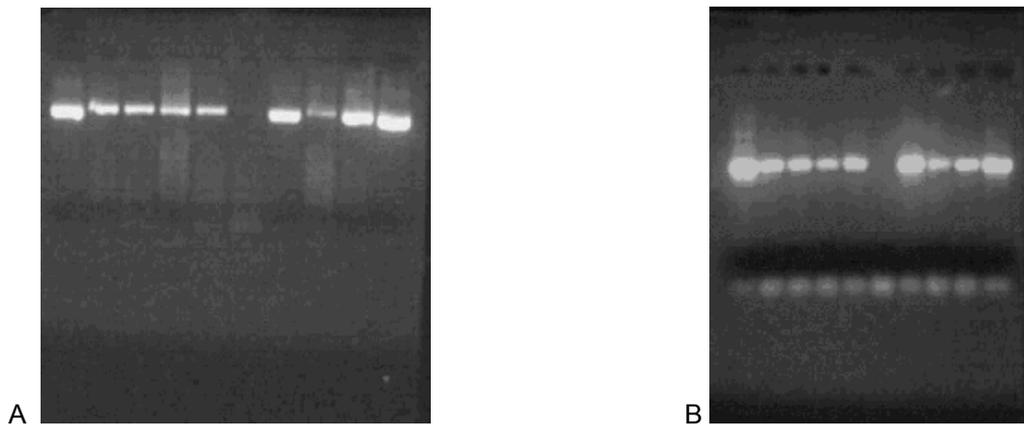


Figure 4.4:(A) The 2% agarose gel of DNA extracted with the original extraction method using the tissue lyser, PVP and double the β - mercaptoethanol; (B) 2% agarose gel of the same plant samples extracted using the tissue lyser, PVP and double the β - mercaptoethanol and added phenol: chloroform: isoamylalcohol, both after PCR with the same nuclear ITS PCR primer pair.

When the results of the amplifications of all four regions are compared, certain general deductions can be made. Firstly *E. junonia* did not amplify with any of the regions, possibly as a result of the plant material being completely degraded due to long storage. Secondly DNA extracted using the modified CTAB method generally gave better results than the Padmalatha *et al.* (2006) method, with the bands from the modified CTAB being much brighter than the ones resulting from the Padmalatha *et al.* (2006) method.

DNA extracted from fresh plant material also as a rule gave better yields of DNA than silica dried plant samples. Sample 1 of each gel represented material of *E. arborea* which was recently collected and stored in silica gel. In contrast samples 2-4 represent species that were collected about two years ago, and which had been stored in silica gel. The contrasting success rates of samples in lane 1 versus samples in lanes 2-4 clearly illustrate the degradation of *Erica* DNA during silica gel storage.

The variation in the results between the different gene regions may be correlated to the sensitivity and extent of match of the different primers used. This would result in some gene regions amplifying better than others.

4.4 Discussion and Conclusions

My focus was to find a DNA extraction protocol which would be optimal for the extraction of DNA from *Erica* species. This was done by comparing various conditions relating to the plant material to be extracted. I could deduce that the modified CTAB method resulted in higher DNA yield than the Padmalatha *et al.* (2006) method and that fresh plant material gives better results than silica dried plant material. Previous studies dealing with DNA extraction from plant samples focused mainly on

improved results obtained with fresher plant samples (Padmalatha *et al.*, 2006; Chakraborti *et al.*, 2006; Narayanan *et al.*, 2006; (Porebski *et al.*, 1997; Plaschke *et al.*, 1995 and Lin *et al.*, 2001).

The use of silica gel to store plant material is, however, a common practice. Our results show that the prolonged storage of *Erica* samples in silica gel is not recommended. Preference should therefore be given to isolation of DNA from fresh material. If fresh material cannot be obtained, our results suggest that DNA isolation should be undertaken as soon as possible after collection, as DNA appears to degrade during silica gel storage. This is confirmed by results achieved with the *E. arborea* samples in the various amplifications, where the sample that had been recently collected and stored in silica gel for a short period gave far superior results.

Based on our results we can conclude that the Padmalatha *et al.* (2006) method that includes a treatment with phenol: chloroform: isoamylalcohol tends to yield purer DNA. Spectrophotometric analyses by Padmalatha *et al.* (2006) showed that the DNA isolated using phenol treatment results in lower levels of protein contamination. Our results do, however, also show that the Padmalatha *et al.* (2006) method does not only remove contaminants, but also substantially reduces the amount of DNA present. This reduction in yield with a resultant reduction in PCR amplification products outweighs the advantages of the method. We thus conclude that the extraction method we applied in the preliminary investigations using double the β -mercaptoethanol plus PVP and no addition of the phenol: chloroform: isoamylalcohol (Chapter 3), gives better results with both fresh and silica dried plant material. As explained above though, fresh material should be given preference. Finally we suggest that the use of the tissue lyser instead of the traditional mortar and pestle will yield superior results when extracting DNA from *Erica* species.



Chapter 5: Conclusions and future perspectives

This study focused on a few important points to consider before embarking on phylogenetic analyses of a large group such as the genus *Erica*, which has received minimal previous phylogenetic analyses. The factors dealt with in this study are summarized below:

The study set out to identify gene regions that had previously been reported to be informative at the species level and to establish their usefulness in the reconstruction of reasonably well-resolved and meaningful phylogenetic trees in *Erica*.

In order to test this, a successful extraction protocol for *Erica* firstly had to be developed. Once in place, it enabled us to evaluate various gene regions for their Parsimony Informativeness.

Based on these analyses, I could reach the following conclusions:

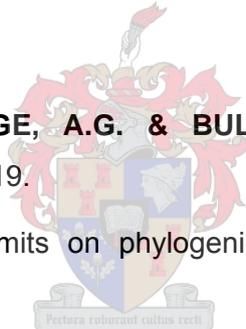
- Five of these nine regions (*trnL-F*, *matK*, *trnS-G*, *rps12- rpl20* and the nuclear ITS region) consistently gave PCR products during amplification and yielded unambiguous sequences.
- The remaining four regions, that is, *psbA-trnH*, *trnC-D*, *rps4-trnT* and *trnT-L* could be amplified in some instances but gave insufficient DNA product to be sequenced with no ambiguities.
- Based on the outcome of these analyses the *trnL-F*, *matK*, *trnS-G* and the nuclear ITS regions were found to be sufficiently informative, with the *matK* and ITS regions found to be the most parsimony informative and promising for further applications in the reconstruction of the *Erica* phylogeny.

The study therefore builds a firm base from which a large-scale phylogenetic analyses of the highly speciose genus *Erica* can be launched in future.

Cited literature

- ANDERBERG A, RYDIN, C. & KALLERSJO, M.** 2002. Phylogenetic relationships in the order Ericales s.l.: analyses of molecular data from five genes from the plastid and mitochondrial genomes. *Am. J. Bot.* 89(4): 677–687.
- ANGIOSPERM PHYLOGENY GROUP (APG).** 1998. An ordinal classification for the families of flowering plants. *Ann. Missouri Bot. Gard.* 85: 531–553.
- ANGIOSPERM PHYLOGENY GROUP (APG).** 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Bot. J. Linn. Soc.* 141: 399–436.
- AOKI, K., SUZUKI, T. & MURAKAMI, N.** 2003. Intraspecific sequence variation of chloroplast DNA among the component species of evergreen broad-leaved forests in Japan. *J. Pl. Res.* 116: 337–344.
- BAKKER, F.T., CULHAM, A., DAUGHERTY, L.C. & GIBBY, M.** 1999. A *trnL-F* based phylogeny for species of *Pelargonium* (Geraniaceae) with small chromosomes. *Pl. Syst. Evol.* 216: 309–324.
- BALDWIN, B.G.** 1992. Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: an example from the Compositae. *Mol. Phyl. Evol.* 1: 3–16
- BARKER, N.P., WESTON, P.H., ROURKE, J.P. & REEVES, G.** 2002. The relationships of the southern African Proteaceae as elucidated by internal transcribed spacer (ITS) DNA sequence data. *Kew Bull.* 57: 867–883.
- BAUM, D.A., SMALL, R.L., & WENDEL, J.F.** 1998. Biogeography and oral evolution of Baobabs (*Adansonia*, Bombacaceae) as inferred from multiple data sets. *Syst. Biol.* 47: 181–207.
- BREMER K, CHASE, M.W. & STEVENS, P.F.** 2004. An ordinal classification for the families of flowering plants., Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond.
- BYTEBIER B., BELLSTEDT D.U. & LINDER, H. P.** 2006. A molecular phylogeny for the large African orchid genus *Disa*. *Mol. Phyl. Evol.* DOI: 10.1016/j.ympev.2006.08.014 article in press.
- CHAKRABORTI D. SARKAR, A., GUPTA, S. & DAS, S.** 2006. Small and large scale genomic DNA isolation protocol for chickpea (*Cicer arietinum* L.), suitable for molecular marker and transgenic analyses. *Afr. J. Biot.* 5(8):. 585–589.
- COWLING, R. M. & RICHARDSON, D.** 1995. Fynbos: South Africa's unique floral kingdom. Pp. 21–40. Vlaeberg: Fernwood Press, Cape Town, South Africa.
- COWLING, R.M. & LOMBARD, A.T.** 2002. Heterogeneity, speciation/ extinction history and climate: explaining regional plant diversity patterns in the Cape Floristic Region. *Div. Distr.* 8: 163 – 179.

- CRONQUIST, A.** 1981. An integrated system of classification of flowering plants. Columbia Univ. Press, New York, USA.
- CRONQUIST, A.** 1988. The evolution and classification of flowering plants. The New York Botanical Garden, Bronx, New York, USA.
- DELLAPORTA S.L., WOOD J. & HICKS J.B.** 1983. A plant DNA mini-preparation: version II. *Plant Mol. Biol. Rep.* 1: 19–21.
- DEMESURE, B., SODZI, N. & PETIT, R.J.** 1995. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Mol. Ecol.* 4: 129–131.
- DIXON, M.T. & HILLIS, D.M.** 1993. Ribosomal RNA secondary structure: compensatory mutations and implications for phylogenetic analysis. *Mol. Biol. Evol.* 10: 256 – 267
- DOYLE J.J. & DOYLE, J.L.** 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19: 11–15.
- DOYLE, J.J.** 1992. Gene trees and species trees: molecular systematics as one-character taxonomy. *Syst. Bot.* 17: 144–163.
- FARIS, J.S., KALLERSJO, M., KLUGE, A.G. & BULT, C.** 1994. Testing significance of incongruence. *Cladistics* 10: 315–319.
- FELSENSTEIN, J.** 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39: 783–791.
- FITCH, W. M.** 1971. Toward defining the course of evolution: minimum change for a specific tree topology. *Syst. Zool.* 20: 406–416.
- GOLDBLATT, P.** 1978. An analysis of the flora of southern Africa: its characteristics, relationships and origins. *Ann. Miss. Bot. Gar.* 65: 369–436.
- GOLDBLATT, P., SAVOLAINEN, V., PORTEOUS, O., SOSTARIC, I., POWELL, M., REEVES, G., MANNING, J.C., BARRACLOUGH, T.G. & CHASE, M.W.** 2002. Radiation in the Cape flora and the phylogeny of peacock irises *Moraea* (Iridaceae) based on four plastid DNA regions. *Mol. Phyl. Evol.* 25: 341–360.
- HAMILTON, M.B.** 1999a. Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Mol. Ecol.* 8: 521–523.
- HAWKINS J.A.** 2000. Evolution and biogeography of *Aspalathus* L. and the South African Flora. PhD. Department of Plant Sciences, University of Reading, London, United Kingdom.
- HILU K. W., BORSCH, T. MULLER, K. SOLTIS, D.E., SOLTIS, P.S., SAVOLAINEN, V., CHASE, M.W., POWELL, M.P., ALICE, L.A., EVANS, R., SAUQUET, H., NEINHUIS, C., SLOTTA,**



- T.A.B., ROHWER, J.G., CAMPBELL, C.S. & L.W. CHATROU.** 2003. Angiosperm phylogeny based on *matK* sequence information. *Am. J. Bot.* 90(12): 1758–1776.
- HIPP A.L., HALL J.C. & SYTSMA K.J.** 2004. Congruence versus Phylogenetic Accuracy: Revisiting the Incongruence Length Difference Test. *Syst. Biol.* 53(1):81–89.
- HIRATSUKA, J.H., SHIMADA, R., WHITTIER, T., ISHIBASHI, M., SAKAMOTO, M., MORI, C., KONDO, Y., HONJI, C.R., SUN, B.Y., MENG, Y., LI, Q., KANNO, A., NISHIZAWA, Y., HIRAI, A., SHINOZAKI, K. & SIGIURA M.** 1989. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Mol. Gen. Evol.* 217: 185–194.
- HOOT, S.B. & DOUGLAS, A.W.** 1998. Phylogeny of the Proteaceae based on *atpB* and *atpB-rbcL* intergenic spacer region sequences. *Aus. Syst. Bot.* 11 (3 – 4): 301 – 320.
- JANSEN, R.K., & PALMER, J.D.** 1987. A chloroplast DNA inversion marks an ancient evolutionary split in the sunflower family (Asteraceae). *P. N. A. S., USA* 84: 5818–5822.
- JOHNSON, L.A. & SOLTIS, D.E.** 1994. *matK* DNA sequence and phylogenetic reconstruction in Saxifragaceae s.s.. *Syst. Bot.* 19: 143–156.
- JUDD, W.S. (eds. C.S. Campbell, E.A. Kellogg & P.F. Stevens),** 1999. Phylogenetic relationships of angiosperms. In *Plant systematics: A phylogenetic approach*. Chapter 8:pp. 161 – 419. Sinauer Associates Inc., Massachusetts, USA.
- KARAKOUSIS, A. & LANGRIDGE, P.** 2003. A High-Throughput Plant DNA Extraction Method for Marker Analysis. *Plant Mol. Biol. Rep.* 21: 95a–95f.
- KARP A.O. SEBERG & BUIATTI, M.** 1996. Molecular Techniques in the Assessment of Botanical Diversity. *Ann. Bot.* 78: 143–149.
- KELLOGG, E.A., APPELS R. & MASON-GAMER, R. J.** 1996. When genes tell different stories: the diploid genera of Triticeae (Gramineae). *Syst. Bot.* 21: 321–247.
- KLAK, C., HEDDERSON, T.A. & LINDER, H.P.** 2003. A molecular systematic study of the *Lampranthus* group (Aizoaceae) based on the chloroplast *trnL-trnF* and nuclear ITS and 5S NTS sequence data. *Syst. Bot.* 28 (1): 70–85.
- KRON, K.A.,** 1997. Phylogenetic relationships of Rhododendroideae (Ericaceae). *Am. J. Bot.* 84: 973–980.
- KRON, K.A, JUDD, W, STEVENS, P.F., CRAYN, D.M., ANDERBERG, A.A., GADEK, P.A., QUINN, C.J. & LUTEYN, J.L.** 2002b. A phylogenetic classification of Ericaceae: Molecular and morphological evidence. *Bot. Rev.* 68: 335–423.
- KUZOFF R.K. & GASSER, C.S.** 2000. Recent progress in reconstructing angiosperm phylogeny. *Trends Plant Sci.* 5: 330–336.

- LEE, M.S.Y.** 2001. Unalignable sequences and molecular evolution. *Trends Ecol. Evol.* 16: 681–685.
- LI J., ALEXANDER, J., WARD, T., DEL TREDICI, P. & NICHOLSON, R.** 2002. Phylogenetic relationships of Empetraceae inferred from sequences of chloroplast gene *matK* and nuclear ribosomal DNA ITS region. *Mol. Phyl. Evol.* 25: 306–315.
- LIN R., DING, Z., BI LI, L. & KUANG, T.** 2001. A Rapid and Efficient DNA Mini-preparation Suitable for Screening Transgenic Plants. *Plant Mol. Biol. Rep.* 19: 379a–379e.
- LINDER, H.P.** 1985. Gene flow, speciation and species diversity patterns in a species-rich area: the Cape Flora. In *Species and Speciation 4* (ed. E.S. Vrba), pp.53–57. Transvaal Museum, Pretoria, South Africa.
- LINDER, H.P.** 2003. The radiation of the Cape Flora, southern Africa. *Biol. Rev.* 78: 597 – 638.
- LINDER, H.P. & HARDY, C.R.** 2004. Evolution of the species-rich Cape Flora. *Phil. Trans. R. Soc. Lond. B* 359: 1623–1632.
- MADDISON, W.P. & MADDISON, D.R.** 1992. MacClade: Analysis of Phylogeny and Character Evolution. Sinauer Associates, Sunderland.
- MANIATIS, T., FRITSCH, E.F & SAMBROOK, J.** 1982; 1986. *Molecular Cloning, A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- MAST, A.R.** 1998. Molecular Systematics of subtribe Banksiinae (*Banksia* and *Dryandra*; Proteaceae) based on cpDNA and nrDNA sequence data: Implications for taxonomy and biogeography. *Aus. Syst. Bot.* 11: 321– 342.
- MCGUIRE F.A. & KRON, K.A.** 2005. Phylogenetic relationships of European and African Ericas. *Int. J. Plant Sci.* 166(2): 311–318.
- NARAYANAN C, DUBEY, S., WALI, S.A., SHUKLA, N., KUMAR, R., MANDAL A.K & ANSARI, S.A.** 2006. Optimization of DNA extraction for ISSR studies in *Tectona grandis* L.f. - an important forest tree species. *Afr. J. Biot.* Vol. 5 (13): 1220–1223.
- NEUHAUS, H. & LINK, G.** 1987. The chloroplast tRNA Lys (UUU) gene from mustard (*Sinapis alba*) contains a class II intron potentially coding for a maturase-related polypeptide. *Cur. Gen.* 11: 251–257.
- OLIVER, E.G.H.** 1987d. Studies in the Ericoideae. VII. The placing of the genus *Philippia* into synonymy under *Erica*; the southern African species. *South African Journal of Botany* 53: 455–458.
- OLIVER, E.G.H.** 1988a. Studies in the Ericoideae (Ericaceae). VI. The generic relationship between *Erica* and *Philippia* in southern Africa. *Bothalia* 18: 1–10.
- OLIVER, E.G.H.** 1993b. Studies in the Ericoideae (Ericaceae). XI. The generic relationship between *Erica* and *Blaeria*. *Kew Bulletin* 48: 771–780.

- OLIVER, E.G.H.** 1993c. Studies in the Ericoideae (Ericaceae). XII. The placing of the genus *Blaeria* into synonymy under *Erica*; nomenclatural and taxonomic changes for the southern African region. *Bothalia* **23**: 1–7.
- OLIVER, E.G.H.** 1994. Studies in the Ericoideae. XV. The generic relationship between *Erica* and *Ericinella*. *Bothalia* **24,2**: 121–126.
- OLIVER, E.G.H.** 2000. Systematics of Ericaceae (Ericaceae: Ericoideae): Species with indehiscent and partially dehiscent fruits. *Contrib. Bolus Herb.* 19: 1–483.
- OLMSTEAD, R.G. & PALMER, J.D.** 1994. Chloroplast DNA systematics: a review of methods and data analysis. *Am. J. Bot.* 81: 1205–1224.
- PADMALATHA K. & PRASAD, M.N.V.** 2006. Optimization of DNA isolation and PCR protocol for RAPD analysis of selected medicinal and aromatic plants of conservation concern from Peninsular India. *Afr. J. Biot.* 5 (3): 230–234.
- PLASCHKE J., GANAL, M.W. & RÖDER, M.S.** 1995. Detection of genetic diversity in closely related bread wheat using microsatellite markers. *Theor. Appl. Genet.* 91: 1001–1007.
- POREBSKI, S., BAILEY, L.G. & BAUM, B.R.** 1997. Modification of CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol. Biol. Reprtr.* 15: 8–15.
- REEVES, G.** 2001. Radiation and macro-evolutionary ecology of the African genus *Protea* L. Ph.D. thesis, Imperial College of Science, Technology and Medicine, University of London, London.
- RICHARDSON, J.E., WEITZ, F.M., FAY, M.F., CRONK, Q.C.B., LINDER, H.P., REEVES, G. & CHASE, M.W.** 2001. Phylogenetic analysis of *Phyllica* L. with an emphasis on island species: evidence from plastid *trnL-F* DNA and nuclear internal transcribed spacer (ribosomal DNA) sequences. *Taxon* 50: 405–427.
- RIESEBERG, L.H. & BRUNSFIELD S.** . 1992 Molecular evidence and plant introgression. In P. S. Soltis, D. E. Soltis, and J. J. Doyle [eds.], *Mol. syst. pl.* 151–176. Chapman and Hall, New York, USA.
- ROGERS, S.O. & BENDICH, A.J.** 1985. Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. *Plant Mol. Biol. Rep.* 5: 69–76.
- RUTHERFORD, M.C. & WESTFALL H. R.** 1996. Biomes of southern Africa- an objective categorization. *Mem. Bot. surv. S. Africa* 54: 98
- RYBICKI, E.** 2005. A Manual of Online Molecular Biology Techniques. Dept Molecular & Cell Biology, University of Cape Town. <http://www.mcb.uct.ac.za/manual/MolBiolManual.htm>
- SALTONSTALL, K.** 2001. A set of primers for amplification of noncoding regions of chloroplast DNA in the grasses. *Mol. Ecol. Not.* 1: 76–78.

- SANDERSON M.J.** 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol. Biol. Evol.* 14: 1218–1231
- SANG, T., CRAWFORD, D.J. & STUESSY, T.F.** 1997. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *Am. J. Bot.* 84: 1120–1136.
- SANG T.** 2002. Utility of Low-Copy Nuclear Gene Sequences in Plant Phylogenetics. *Crit. Rev. Bioch. Mol. Biol.*, 37(3): 121–147.
- SAVOLAINEN V., CHASE, M.W., HOOT, S.B., MORTON, C.M., SOLTIS, D.E., BAYER, C., FAY, M.F., DE BRUIJN, A.Y., SULLIVAN, S. & QIU, Y.L.** 2000. Phylogenetics of flowering plants based on combined analysis of plastid *atpB* and *rbcl* gene sequences, *Syst. Biol.* 49: 306–362.
- SCHUMANN D. & KIRSTEN, G.** 1992. *Ericas of South Africa*, Vlaeberg Fernwood Press, South Africa.
- SHARMA, A.D., GILL, P.K. & SINGH, P.** 2002. DNA isolation from dry and fresh samples of polysaccharide rich plants. *Plant Mol. Biol. Rep.* 20: 415a– 415f.
- SHAW, J., LICKEY, E.B., BECK, J.T., FARMER, S.B., LIU, W., MILLER, J., SIRIPUN, K.C., WINDER, C.T., SCHILLING, E.E. & SMALL, R.L.** 2005. The Tortoise and the Hare II: Relative utility of 21 non-coding chloroplast DNA sequences for phylogenetic analysis. *Am. J. Bot.* 92(1): 142–166.
- SOLTIS, D.E. & KUZOFF, R.K.** 1995. Discordance between nuclear and chloroplast phylogenies in the *Heuchera* group (Saxifragaceae). *Evolution* 49: 727–742.
- SOLTIS, D.E., JOHNSON, L.A. & LOONEY, C.** 1996. Discordance between ITS and chloroplast topologies in *Boykinia* group (Saxifragaceae). *Syst. Bot.* 21: 169–185.
- SOLTIS, D.E., SOLTIS, P.S. & DOYLE, J.J.** 1998. *Molecular systematics of plants II: DNA sequencing*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- STEELE, K.P. & VILGALYS, R.** 1994. Phylogenetic analyses of Polemoniaceae using nucleotide sequences of the plastid gene *matK*. *Syst. Bot.* 19: 126–142.
- SWOFFORD, D.L.** 2002. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4.0b10. Sinauer Associates, Sunderland, Mass.
- TABERLET, P., GIELLY, L., PAUTOU, G. & BOUVET, J.** 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Mol. Biol.* 17: 1105 – 1109.
- TAKHTAJAN, A.** 1997. *Diversity and Classification of Flowering Plants*. In: Columbia University Press, USA.
- TATE, J.A. & SIMPSON, B.B.** 2003. Paraphyly of *Tarasa* (Malvaceae) and diverse origins of the polyploid species. *Syst. Bot.* 28: 723–737.

- TRINDER-SMITH, T. H.** 2003. The Levyns Guide to the Plant Genera of the Southwestern Cape. *Contrib. Bolus Herb.* 21. Bolus Herbarium, Cape Town University, South Africa.
- VAN WYK, A.E. & SMITH G.F.** 2001. Regions of Floristic Endemism in Southern Africa. A review with emphasis in Succulents, Pp. 19-23. Umdaus press, Hatfield, South Africa.
- VERBOOM, A.G., LINDER, H.P. & STOCK, W.D.** 2003. Phylogenetics of the grass genus *Ehrharta*: Evidence for radiation in the summer-arid zone of the South African Cape. *Evolution*: **57** (5): 1008 – 1021.
- WAKASUGI, T., SUGITA, M., TSUDZUKI, T. & SUGIURA, M.** 1998. Updated gene map of tobacco chloroplast DNA. *Plant Mol. Biol. Rep.* 16: 231–241.
- WHITE, T.J., BRUNS, T., LEE, S. & TAYLOR, J.W.** 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Pp. 315-322 In: PCR Protocols: A Guide to Methods and Applications, eds. Innis, M. A., D. H. Gelfand, J. J. Sninsky, and T. J. White. Academic Press, Inc., New York, USA.
- XU, D., SAKAI, H., KANAZAWA, A.J.M., SHIMAMOTO, A. & SHIMAMOTO, Y.** 2000. Sequence variation of non-coding regions of chloroplast DNA of soybean and related wild species and its implications for the evolution of different chloroplast haplotypes. *Theor. App. Gen.*101: 724–732.
- YOKOTA, Y., KAWATA, T., IIDA, Y., KATO, A. & TANIFUJI, S.** 1989. Nucleotide sequences of the 5.8S rDNA gene and internal transcribed spacer regions in carrot and broad bean ribosomal DNA. *Mol. Ev.* 29: 294–301.

[Http://: mobot.org/mobot/rearch.html](http://mobot.org/mobot/rearch.html)



APPENDIX A: Formulas

For stats calculations

$$\text{Percentage variability} = \frac{\text{Total number of variable characters}}{\text{Total number of characters}} \times 100\%$$

$$\text{Percentage Informativeness} = \frac{\text{Number of parsimony informative characters}}{\text{Number of total characters}} \times 100\%$$

For annealing temperatures

$$\text{Annealing temperature} = (C + G) \times 4^\circ\text{C} + (T + A) \times 2^\circ\text{C} \text{ from primer sequence}$$

For primer concentrations in pmol/ μL

$$\text{Concentration (C)} = A_{260} \times 100 / 1,5 \text{ (NA)} + 0,71 \text{ (Nc)} + 1,2 \text{ (NG)} + 0,87 \text{ (N)}$$

