INTRASPECIFIC GENETIC VARIATION AMONG POPULATIONS OF KENYAN AFRICAN VIOLET. *SAINTPAULIA RUPICOLA* B. L. BURTT: A MOLECULAR APPROACH

BY

SIMON MAINA MACHARIA
B.Ed Science (Egerton)

A thesis submitted in partial fulfilment of the award of the degree of Master of Science in Botany

DEPARTMENT OF BOTANY
UNIVERSITY OF NAIROBI

FEBRUARY, 1998

DECLARATION

Declaration by the candidate

I hereby declare that this thesis has not been submitted, either in the same or different form, to this or any other university for a degree. No part of this thesis may be reproduced without the prior permission of the author and/or University of Nairobi.

Simon Maina Macharia — Date 25/4/98

Declaration by the supervisors

This thesis has been submitted with our approval as university supervisors.

Prof. T. K. Mukiama — Date 13/5/98

1st supervisor

Chairman

Department of Botany

University of Nairobi

P. O. Box 30197, Nairobi

Dr. Rashid A. Aman — Date 10/6/98

2nd supervisor

Director, Research and Scientific Affairs

National Museums of Kenya

P. O. Box 40658, Nairobi

DEDICATION

To my mother, for your endless support and encouragement. To my daughter, Rachel and my son, Vickie, who have provided the driving force

Acknowledgements

A number of people deserve special acknowledgment in view of the invaluable support they gave me during my M.Sc course. In particular, I would like to thank Dr. Rashid Aman of the Department of Molecular Genetics, National Museums of Kenya (NMK) for providing most of the supervision for this work, and in whose lab its bulk has been carried out. I am also greatly indebted to Mr. Timothy Pearce of the Plant Conservation Programme (PCP), NMK for providing the initial funding for this project. Special thanks also go to the Board of Postgraduate Studies, University of Nairobi, for recommending me for a DAAD scholarship, and DAAD for providing the scholarship. I would also like to thank Prof. T. K. Mukiama of the Department of Botany, University of Nairobi for supporting my scholarship offer, and providing part of the supervision for this project especially through the final stages of preparation of this thesis. I am greatly indebted to Mr. Benny Bytebier of the PCP, NMK, who assisted me in the development of this project and personally guided me successfully through the field survey and provided part of the supervision for this project.

Special thanks also go to Prof. Dr. Wolfgang Hachtel of the Botanisches Institut der Universitat Bonn, Germany, for offering me a chance to work in his lab for six months. I am also greatly indebted to Prof. Eric Knox of Rutgers University, USA, during his tenure at the NMK, for a lot of invaluable advice throughout the course of the project, and for providing me with tobacco chloroplast DNA probes used in the initial survey of this project. A big thank you to the members of the Department of Molecular Genetics, NMK for the bulk of technical training and support and

providing a conducive and enjoyable environment in which I successfully completed this project. Last but by no means least, I would like to convey my special thanks to my parents, relatives, friends and all those who contributed in any way towards the successful completion of this study for their unwavering love, patience and encouragement during the course of this study. And to my friend Ngige, thank you for assisting me in printing!

TABLE OF CONTENTS

Table of contents continued

Table of contents continued

Table of Contents continued

Table of Contents continued

List of **Figures**

List of Tables

List of Plates

SYMBOLS AND ABBREVATIONS

Symbol/Abbreviation	Full name
APS	Ammonium persulphate
ATP	Adenosine triphosphate
bp	Base pair
BSA	Bovine Serum Albumin
BME	ß-mercaptoethanol (2-mercaptoethanol)
$CaCl_2$	Calcium chloride
°C	Degree celcius
CTAB	Cetyltrimethylammonium bromide
CsCl	Caesium chloride
ddH_2O	Double distilled water
DTT	Dithiothreitol
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonulcleotide triphosphates
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytosine triphosphate
dGTP	Deoxyguanosine triphosphate
dTTP	Deoxythyamidine triphosphate
EDTA	Ethylenediaminetetra-acetic acid (disodium salt)
EtBr	Ethidium bromide
EtOH	Ethanol

Abbreviations continued

Abbreviation	Full name
GPS	Geographical Positioning System
HCl	Hydrochloric acid
hr	Hour(s)
IPTG	Isopropylthio-beta-D-galactoside
kb	Kilobase(s)
KOAc	Potassium acetate
LB	Lauria broth
M	Molar
mg	Milligram
μ	Micro
μg	Microgam
μl	Microlitre
μM	Micromolar
min	Minute(s)
ml	Millilitre
mM	Millimolar
$MgCl_2$	Magnesium chloride
$MnCl_2$	Manganese chloride
NaCl	Sodium chloride
ng	Nanogram(s)
NaI	Sodium iodide

Abbreviations continued

Abbreviation	Full name
NH_4OAc	Ammonium acetate
nm	Nanometer(s)
NMK	National Museums of Kenya
OD	Optical Density
PCP	Plant Conservation Programme
PCR	Polymerase Chain Reaction
pmol	Picomole(s)
PNK	Polynucleotide kinase (from T4 phage)
RF	Replicative form (double-stranded)
DNA	Deoxyribonucleic acid
RNA	Ribonucleic Acid
RNAse	Ribonuclease A
rpm	Revolutions per minute
SDS	Sodium dodecyl sulphate
SSC	Saline sodium citrate (appendix 2)
TAE	Tris-acetate-EDTA (see appendix 2)
TBE	*Tris-borate-EDTA buffer (see appendix 2)*
TE	Tris-EDTA buffer (see appendix 2)
TEMED	N,N,N',N',-tetramethylethylenediamine

Abbreviations continued

Abbreviation	Full name
Tris (base)	Tris(hydroxy-methyl)aminomethane
TTP	Thymidine triphosphate
UPGMA	Unweighted Pair Group Method with Arithmetic mean
UV	Ultraviolet light
v/v	Volume by volume
w/v	Weight by volume
X-gal	5-bromo-4-chloro-3-indoyl-beta-D-galactopyranoside

ABSTRACT

Saintpaulia rupicola B. L. Burtt (Gesneriaceae) is one of the African Violet species under high threat of extinction from its habitat. It is endemic to coastal Kenya where it occurs in Kilifi and Kwale districts in four distinct populations. The Cha Simba and Mwarakaya populations contain plants that correspond to the typical *Saintpaulia rupicola*. However, the other two populations, Kacharoroni and Mwache, consist of plants which are morphologically different, and which may qualify to be treated as separate species, *Saintpaulia* sp. nov.

In order to design a conservation plan for this species, an insight in the genetic diversity among these populations is required. Furthermore, a study at the molecular level could help to clarify the taxonomic position of *Saintpaulia* sp. nov. with regard to *S. rupicola* and other taxa in the genus *Saintpaulia*.

The Internal Transcribed Spacer region (ITS) of the nuclear ribosomal DNA was analyzed as a genetic marker to determine the genetic variation among these populations. DNA sequencing of this region revealed its conserved nature, with only four (0.63 % of the whole region) base substitutions among the four populations. Low genetic distances (between 0.0000 to 0.0048) were observed among these populations, indicating that the four are closely related, the most closely related being Kacharoroni and Mwache populations (0.0016).

Cladistic analysis using Geneworks Release 2.3 and PHYLIP analysis programs generated a total of five phylogenetic trees which cluster Mwarakaya with Cha Simba and Kacharoroni with Mwache.

CHAPTER 1

INTRODUCTION, LITERATURE REVIEW AND STATEMENT OF THE PROBLEM

1.0 Introduction

Species characterization based on genotypic characters and determination of genetic variation in individuals and species are essential tenets for effective conservation and enhancement of threatened genetic resources. As may often happen, concert of phenotypic and genotypic studies may provide a more complete understanding of species delimitation especially in cases where classification by morphological characters alone is doubtful.

Morphological traits alone may sometimes provide inadequate or misleading cues to phylogenetic distinctions at the sub-species and species levels. Moreover, phylogenetic analysis based on morphological traits alone may represent only a small portion of an organism's genetic make-up and is influenced by environmental factors, thereby limiting its utility in describing potentially complex genetic relationships which may exist within taxa. However, knowledge of the underlying level, structure and origin of genetic variation is essential for the development of scientific approaches for the sustainable utilization, taxonomic delimitation and conservation of species. Molecular analysis of genetic variation overcomes these constraints and thus, in concert with morphological analysis, taxonomic delimitation is made clearer.

A proper species conservation plan seeks to conserve a maximum but a manageable number of different genotypes. However, before this can be done, it is essential to know the degree of genetic variation among species and populations for the implementation of a sound conservation strategy. This study seeks to address determination of the degree of genetic variation among the naturally occurring populations of the Kenyan African Violet, *Saintpaulia rupicola* B. L. Burtt (Burtt, 1964) by molecular techniques as a contribution to better understanding of their taxonomic delimitation and the development of a species conservation plan.

1.0.1 The African Violets

The African Violets belong to the genus *Saintpaulia* Wendland of family Gesneriaceae. They are indigenous to East Africa, with two recognised species, *Saintpaulia rupicola* B. L. Burtt and *S. teitensis* B. L. Burtt (Burtt, 1960), endemic to Kenya. They have a tremendous value as indoor ornamental plants, being one of the best known and most widely grown of all house plants throughout the world. Several species have a direct ornamental value in the wild form but most are used in breeding programs to create the more popular *Saintpaulia* hybrids. Some societies in the USA and UK are solely devoted to the breeding and improvement of the African Violets.

1.0.2 The African violet as an endangered species

The African violets have faced considerable destruction of their natural habitat since the late nineteenth century, and have almost been eradicated from the wild. Natural populations of *Saintpaulia* now exist as small, isolated populations in the remaining

forests of coastal Tanzania and Southern Kenya (Sheil, 1993).

Wild populations of African violets grow in localized patches on well drained surfaces in deep shade and high humidity, in near soil free environments offered by rock faces (e.g. limestone outcrops) and rough tree boles. They rarely survive when exposed to the sun (Sheil, 1993).

The main threat to the African violets has been the destruction of their natural habitat. Land pressure has been severe in these remnant forests. This, in addition to demand for firewood leads to the clearing of the forests thus exposing them to direct sunlight. Limestone extraction in some of these areas is also a major threat which is almost wiping out some populations.

1.0.3 *Saintpaulia rupicola*

S. rupicola is the most geographically isolated species of the genus *Saintpaulia*. From Kaloleni (Kilifi district), which is the locality most often mentioned (Burrt, 1964), it is nearly 150 km in a direct line with the next species, whether to the Taita hills (Kenya), the Usambara mountains (Tanzania), or down the coast to Tanga (Tanzania). *S. rupicola* was first identified in 1964 from samples collected by J. Lavranos at Chonyi, 4 km north of Kaloleni town in Kilifi district (Burtt, 1964). The type specimen describes the plant as growing on limestone cliffs, which ''form an isolated outcrop in the otherwise smooth and undulating, deeply soil-covered hills''.

S. rupicola is known from two localities near Kaloleni, namely Cha simba and Mwarakaya. Through a concerted effort of the Coastal Forest Survey team (Robertson and Luke, 1993), two more localities were identified. Kacharoroni, near the Rare (Vitengeni) river in Kilifi district and Mwache (Mwache Forest Reserve) in Kwale district. These plants are so unlike *S. rupicola* (especially the size of leaf hairs, leaf size colour and succulence) or any other *Saintpaulia* species that Robertson and Luke (1993) classified them as *Saintpaulia* sp. nov. However, for the purposes of this work, we shall maintain reference of all these populations by their original (locality) names, Cha Simba, Mwarakaya, Mwache and Kacharoroni.

Each of these populations is only about 1 km^2 in area. All of them grow on limestone outcrops in shaded areas especially deep in the forest or under deep shade provided by caves. All populations are under threat of extinction, mainly because of clearing of forest cover as the demand for cultivated/arable land and firewood continues to increase. In the Mwache population, an additional serious threat is quarrying of the limestone on which the plants grow. This population has almost been wiped out (see plate 2, chapter 3). Although clearing of the forest continues at Kacharoroni, this might still be the biggest population.

1.0.4 Conservation efforts at the National Museums of Kenya

The African Violets have been propagated widely. As such, the wild type African Violet, a valuable heritage of nature, has been overlooked and ignored as the horticulturalist works to create a hybrid more beautiful than the original parents. This has been happening as the wild-type African Violets lose their suitable habitat

as discussed in section 1.0.2. In addition, possible accidental escape of the hybrids back to the wild may cross-fertilize with the wild-type African Violets. Thus, there exists a danger that the wild type African Violet may become extinct. The Plant Conservation Programme (PCP) of the National Museums of Kenya (NMK) has embarked on the propagation and conservation of threatened plant species, including the African Violets. This has started with the species endemic to Kenya, *S. rupicola* and *S. teitensis*. During the propagation of these plants in the nursery, crossing between species and populations may result in the merging of gene pools, the very battle that the conservationist is trying to fight. An understanding of the genetic variation between such gene pools would lead to better sampling in order to maintain the genetic diversity that exists in these populations.

Knowledge of the underlying level, structure and origin of genetic variation within and among plant populations is essential for the development of scientific approaches for their sustainable utilisation and conservation. Although morphological characters and various production traits have been used traditionally to characterise levels and patterns of diversity, these traits alone represent only small portion of the genome of an organism and are also influenced by environmental factors, thereby limiting their utility in describing the potentially complex genetic relationships which may exist within taxa. To overcome these constraints, various molecular approaches for genetic analysis have been devised (see section 1.2).

In concert with morphological analysis which is going on at the NMK, this molecular analysis may contribute to the delimitation of species boundaries and the

determination of the units of conservation in this species. In addition, there remains an almost complete lack of knowledge about the variation of the natural populations of the African Violets.

1.0.5 Statement of the problem

1.0.6 Specific objectives

- To collect and analyse samples of *Saintpaulia rupicola*

- To study the extent of genetic variation within and among populations of *Saintpaulia rupicola*

1.0.7 Questions to be addressed

- Does genetic variation exist that clearly separates the four populations of *Saintpaulia rupicola* as separate genetic entities?

- Does this variation, if it does exist, necessitate classification of *Saintpaulia* sp. nov as separate species?

- What recommendations would be made from the genetic variation regarding conservation strategies of this species?

1.1 Literature review

1.1.1 Classification status of the African Violets

Very little has so far been done on the classification and nomenclature of the African Violets. Of the species that have been described, some are closely allied (Johansen, 1978) and the status of 20 to 30 or so recognised wild forms remains controversial (Sheil, 1993). Several studies on taxonomic delimitation have been done. These include morphology (Burtt, 1947, 1956, 1958, 1960, 1964, 1976), Ecology (Johansson, 1978), *in-vitro* propagation (Harvey and Knap, 1979), carpological analysis (Ivaniva, 1966), interspecific hybridization (Arisumi, 1964) and chromosome numbers (Lee, 1962, Fussell, 1958; Milne, 1975; Ratter, 1975), cytology (Ehrlich, 1958) and Karyomorphology (Sera and Karasuwa, 1984).

Classification of *S. rupicola* is based on morphological analysis (Burtt, 1960). As noted in section 1.0.3, Robertson and Luke (1993) classified the Kacharoroni and Mwache populations as *Saintpaulia* sp. nov. using morphological analysis until a more thorough classification is done. This molecular analysis of genetic variation is meant to clarify these taxonomic differences.

1.1.2 Genomes available for DNA analysis in plant and animal taxa

Genomes available for analysis of DNA variation in animal and plant taxa are the nuclear genome (e.g. nuclear ribosomal DNA or nrDNA), the mitochondrial genome (mitochondrial DNA or mtDNA) and the chloroplast genome in plants (chloroplast DNA or cpDNA). Most phylogenetic studies in plants have been based on cpDNA and nrDNA (Gielly and Taberlet, 1994; Olmstead and Palmer, 1994; Soltis *et al*,

1991b), while mtDNA has been a primary target of molecular variation in evolutionary studies of animals (Olmstead and Palmer, 1994). The respective genomes have been discussed in the following sections.

1.1.2.1 Mitochondrial DNA (mtDNA)

The generally maternal inheritance of mtDNA, combined with its ease of isolation, simple structure, small size, active recombination and rearrangement, and high rate (about 10^{-8}) of nucleotide substitution have made the molecule amenable to population genetic and phylogenetic analysis in animals (Nei, 1975; Wolfe *et al,* 1987; Harrison, 1989; Sperling, 1993). The non-recombining inheritance of the genome and its extensive intraspecific polymorphisms have allowed estimation of gene trees at the intraspecific level (Avise, 1989) and the analysis of microevolutionary processes (Avise *et al,* 1987; Soltis *et al.*, 1991a).

Analysis and comparison of nucleotide substitutions in plant mtDNA, cpDNA and nrDNA sequences has revealed that the substitution rate in mtDNA is less than one-third that in cpDNA, which in turn evolves only half as fast as nrDNA (Wolfe *et al,* 1987).

The analysis of mtDNA has revolutionized the study of phylogeny and population genetics in animals (Harrison, 1989; Palmer, 1992). In plants, it is cpDNA rather than mtDNA which has been the genome of choice for molecular phylogenetic and systematic studies (Palmer, 1988, 1992). This is because animal mtDNA is highly conservative in size (15.7 kb to 19.5 kb) and structure but changes rapidly in

primary sequence (Zurawski *et al*, 1984).

In contrast, the mitochondrial genome in plants is characterised with a number of complications which limit its applicability. Plant mtDNA is much harder to purify than cpDNA and is much less abundant in leaves. It is also relatively less characterised in the sense that only a few genes have been sequenced and relatively few clone banks are available (Palmer, 1992). Although the lower substitution rate in plant mtDNA as compared to cpDNA and nrDNA may make its comparative sequencing a higher resolution tool in phylogenetics, it has limitations. Thus, its wide variation in size (200 kb to over 2000 kb), structure (Palmer, 1987), gene content and variable amounts of post-transcriptional sequence change (due to RNA editing) present difficulties in whole genome restriction mapping studies (Olmstead and Palmer, 1994).

Although the mitochondrial genome varies widely in size and structure as noted above, it changes slowly in primary sequence as compared to cpDNA and nrDNA. This low evolutionary rate is regarded to be too scant in restriction site variability to be useful at the interspecific level and below, where cpDNA and rDNA can provide informative data. High rates of rearrangement in plant mtDNA and its effects also make the molecule not very suitable for studies of species relationships at higher taxonomic levels as might theoretically be expected (Palmer, 1992). Another factor which complicates the analysis of plant mtDNA is that it has many exogenous cpDNA sequences, some as large as 12 kb in length, which might interfere with probe specificity and hence analysis (Palmer, 1992).

1.1.2.2 The chloroplast genome

The chloroplast genome is inherited clonally, usually uniparentally through the maternal parent in angiosperms and paternally in some gymnosperms (Palmer, 1987). It is abundant in leaf cells, is small in size (a factor that makes restriction mapping and gene cloning relatively easy), has a simple, highly conservative mode of evolution (Palmer, 1987) and varies little in size and gene content. The angiosperm chloroplast genome is also characterised by structural rearrangements such as inversions and gene or intron deletions (Downie and Palmer, 1992) which can be informative markers in studies involving monophyletic groups. There is also extensive background of molecular information on the chloroplast genome. For example, complete cpDNA sequences are known for the liverwort (*Marchantia polymorpha)*, tobacco (*Nicotiana tabacum)* and rice *(Oryza sativa)*. Its constituent protein encoding genes potentially provide a large database for phylogenetic analysis. All these factors make cpDNA particularly suitable for phylogenetic analysis.

Restriction site analysis of cpDNA has proven to be a suitable method for phylogenetic reconstruction and systematic studies at the intraspecific, interspecific and intergenic levels (reviewed by Palmer, 1987; Soltis *et al,* 1992; Downie and Palmer, 1992).

1.1.2.3 Ribosomal DNA

Eukaryotic ribosomal RNA genes, also known as the nuclear ribosomal DNA (nrDNA) is a set of nuclear genes which code for three of the rRNA components of cytoplasmic ribosomes (Saghai-Maroof *et al*, 1984). These genes are transcribed as a single unit and occur as clusters of tandemly repeated segments. Adjacent transcription units are separated by a non-transcribed spacer region known as the intergenic spacer (IGS). A single transcription unit and an IGS comprise a single repeat unit (Learn and Schaal, 1986). This is shown in figure 1.

The (nrDNA) consists of tandem repeat units varying from 500 to 40,000 copies per genome (Rogers and Bendich, 1987), factors that make it relatively easy to detect or clone in the laboratory. It also undergoes rapid concerted evolution (Zimmer et al, 1990; Arnheim et al, 1980; Arnheim, 1983), within and even between loci (Arnheim et al, 1980; Arnheim, 1983; Appels and Dvorak, 1982; Hillis et al, 1991). This promotes its usefulness in phylogenetic reconstruction (Sanderson and Doyle, 1992).

In plants, rDNA units: 18S, 5.8S and 26S are separated by intergenic spacers (Suh *et al,* 1993) as shown in figure 1. The 5.8S coding region is located between the 18S and 26S coding regions and is flanked by internal transcribed spacers 1 and 2 (ITS1 and ITS2). These repeat units of rDNA have different rates of sequence change, therefore different regions of the molecule can be used to

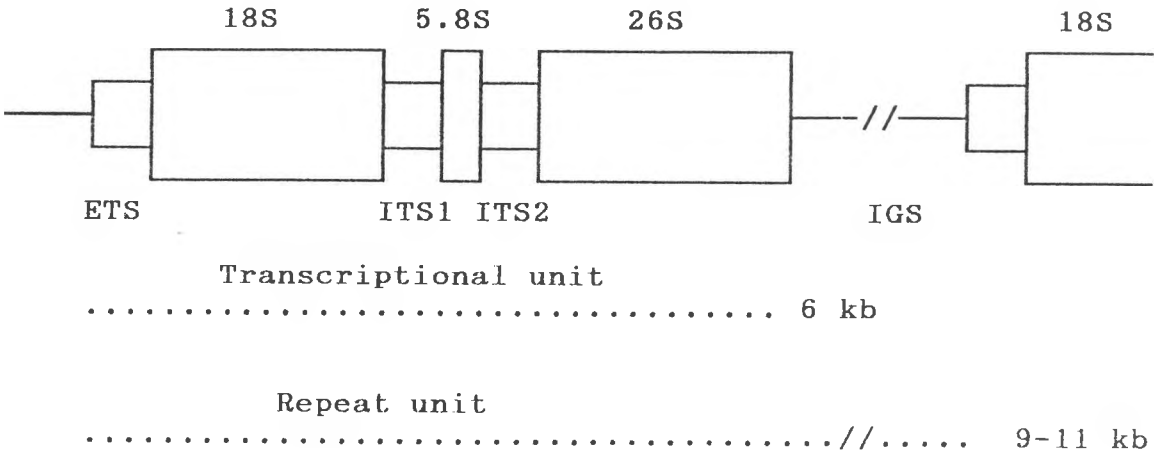


Figure 1: Typical plant rDNA repeat unit drawn to scale (from Hamby and Zimmer, 1992). The IGS region is made of sub-repeat regions that are highly variable in number, causing a high level of length variation in the region at the population level, or even within single individuals (reviewed by Baldwin, 1992). ETS is the External Transcribed Spacer; ITS1 and ITS2 are the Internal Transcribed Spacers. The boxes are the coding regions.

examine lineages with different levels of divergence (Suh *et al,* 1993). The 18S and 26S coding regions have been useful in resolving phylogenetic questions at the family level and above while the ITS sequences have potential value in assessing genetic relationships at lower taxonomic levels such as genera and species because the sequences of spacer regions generally evolve more rapidly than coding regions (Suh *et al,* 1993).

1.1.2.3.1The Internal Transcribed Spacer (ITS)

The ITS region of the nrDNA (about 650 bp long) has been used extensively at the species level for assessment of genetic variation by DNA sequencing and restriction site analysis (Sytsma and Schaal, 1985; Rieseberg *et al*, 1988; Smith and Sytsma, 1990; Crisci *et al*, 1990; Kim and Mabry, 1991; Swensen *et al*, 1995). Baldwin (1993) has revealed within-species variation in the ITS sequences with potential for resolving infraspecific relationships among different populations in *Calycadenia*, and for addressing questions of species boundaries. Other studies (Hsiao *et al,* 1993; Hsiao *et al,* 1994; Hsiao *et al,* 1995; Wojciechowski *et al,* 1993) of ITS sequence variation have also highlighted the usefulness of ITS sequences in resolving phylogenetic relationships at the species and population levels. Swensen *et al* (1995) have revealed restriction site variation in the ITS region among populations and varieties of *Malacothamnus fasciculatus* (Malvaceae). These factors make the ITS region an ideal genetic marker for the present study.

None of the above studies revealed any genetic variation among individuals of one population. In all the studies cited above, between 50 to 61 % of this variation came

from ITS1, 35 to 40 % in ITS2 and 2 to 4 % in the 5.8S region.

1.1.2.3.2Structure of the ITS region

The Polymerase Chain Reaction (PCR, section 1.2.2) is a powerful *in vitro* technique which allows the amplification of a specific DNA region that lies between two regions of known DNA sequence. This technique, which was invented in 1985 by Kary Mullis (Saiki *et al*, 1985), permits DNA amplification by using short, single stranded DNA molecules which are complementary to the ends of a defined sequence of DNA template. These short single stranded molecules are known as oligonucleotide primers.

A schematic representation of the ITS region together with the primers prescribed for this region (Baldwin, 1992) is shown in figure 2.

1.2 DNA techniques for determining genetic variation in plants

1.2.1 Restriction site analysis

Restriction site analysis, traditionally called Restriction Fragment Length Polymorphism (RFLP) analysis, is a DNA-based technique that has widely been used for studying genetic variation within and between populations. In this technique, DNAs from a cohort of closely related species are digested by restriction enzymes and the pattern of restriction fragments resulting from the digests resolved on the same gel and compared. The restriction enzymes are very specific in the DNA sequence they recognize and cut. Thus a change in the sequence which

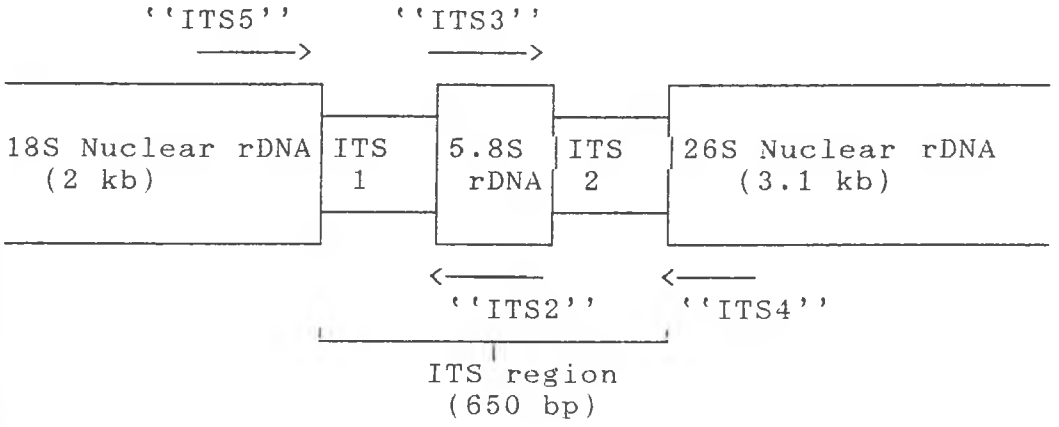


Figure 2: Repeat unit of 18-26S nuclear ribosomal DNA (nrDNA), minus the intergenic spacer (after Baldwin, 1992). ITS region = Internal Transcribed Spacer region. Arrows indicate the approximate positions of primers designed and named for the ITS region by White *et al*, 1990 (Table 1). Only ''ITS4'' and ''ITS5'' were used to amplify and sequence the whole ITS region.

Table 1: Primers designed and named (White *et al*, 1990) for the ITS region. Sequences are given in the 5'-3' direction

PRIMER	SEQUENCE
ITS2	GCTGCGTTCTTCATCGATGC
ITS3	GCATCGATGAAGAACGCAGC
ITS4	TCCTCCGCTTATTGATATGC
ITS5	GGAAGTAAAAGTCGTAACAAGG

produces an addition or deletion of (a) site(s) recognized by the enzyme may change the pattern of restriction fragments seen on a gel. Differences between fragment patterns are usually ascribed to gain or loss of restriction sites resulting from single base substitutions within the site (Palmer, 1987). By cutting DNAs from different individuals using restriction enzymes, patterns of variation may be observed and analyzed to estimate the amount of genetic variation in the DNA sequences.

While RFLP analysis has proven a suitable tool in assaying genetic variation at the species level and below, its resolving power remains lower compared to DNA sequencing techniques. In some cases, the restriction site analysis may not reveal differences between or among closely related species. For example, Kim *et al* (1992) found no restriction site variation in the chloroplast genomes of *Krigia dandelion, K. wrightii,* and *K. occidentalis* (Asteraceae) using 22 restriction enzymes. Similar observations have been made in *Microceris* (Wallace and Jansen, 1990) and *Viguiera* (Schilling and Jansen, 1989) of the family Asteraceae. RFLP analysis requires use of large amounts of genomic DNA (5-10 μg), thus may not be feasible if the DNA amount is limiting. Handling and analysis of restriction site variation data is largely manual and scoring for variation is cumbersome. Above all, the major shortcoming of this technique is that a method is yet to be found for storing restriction site mapping data in a way that would allow retrieval, addition of new data sets and comparative scanning (as say in a computer) for new versus old accessions. To circumvent this problem, some researchers such as Doyle *et al* (1990) and Rieserberg *et al,* (1991) have resorted to assaying new accessions for previously marked informative restriction site variation. This method, however, precludes

identification of a relationship which may have been undetectable in the previous uninformative restriction site analyses.

In contrast to RFLP analysis, procedures for the identification of DNA polymorphisms based on PCR and sequencing are technically simple and require only small amounts of genomic DNA (10-20 ng). These provide higher resolution and precise measure of the genetic variation within and between species at the base pair level. Such data are highly informative with regard to microevolutionary, systematic and phylogenetic studies of species.

1.2.2 Randomly Amplified Polymorphic DNA sequences (RAPDs)

This genetic marker uses the Polymerase Chain Reaction (PCR), a technique that multiplies copies of a particular region of DNA from minute amounts of DNA material (Russel *et al,* 1993). The template DNA is denatured by heating, and a short, single-stranded nucleotide fragment (primer) annealed to this template. The primer is then extended, producing a strand complementary to the template strand. This reaction is catalysed by a thermostable DNA polymerase, *Taq* polymerase. By repeating this cycle many times, the targeted DNA is multiplied exponentially.

The RAPD technique identifies polymorphism in plants without having any prior knowledge of a DNA sequence (Welsh and McClelland, 1990; Williams *et al*, 1990). It requires a fast and simple analytical process involving amplification of unknown DNA sequences using single, short, random oligonucleotide primers. These primers anneal to the regions of the genome to which they have homology and any

bands subsequently observed may be used as raw data for the comparison of plant genotypes. Most RAPD markers are dominant and polymorphisms will be detected. Thus, a DNA sequence will be amplified from one individual but may not be amplified from another. This simple technique requires small amounts of DNA and requires no radioactive material. RAPDs provide a cost-effective method for the precise and routine evaluation of variability, and reveal the extent and distribution of genetic diversity. Chalmers *et al* (1992) revealed extensive genetic variability between *Gliricidia sepium* and *G. maculata* using RAPD analysis. This variation was partitioned into between and within population components.

1.2.3 Polymerase Chain Reaction (PCR) and DNA sequencing

The Polymerase Chain Reaction (PCR) (Section 1.2.2) enables specific amplification and sequencing of regions of interest. Identification of conserved stretches of flanking sequence across a wide range of taxa permit the design of 'universal' primers (Kocher *et al.*, 1989; Tarberlet *et al,* 1991; Hillis and Dickson, 1991), which have proven value in PCR amplification and DNA sequencing. This technique can be employed to study any section of the three genomes (nuclear, mitochondrial and chloroplast, see Section 1.2) available for analysis. The PCR product can then be sequenced directly or by first cloning into a vector.

DNA sequencing is done by the dideoxynucleotide chain-termination method described by Sanger *et al* (1977). This method involves the synthesis of a DNA strand by Klenow fragment of DNA polymerase *in vitro* using a single strand DNA template.

Synthesis is initiated at only one site where an oligonucleotide primer is annealed to the single stranded or denatured template, and the synthesis reaction is terminated by the incorporation of a nucleotide analog that does not support continued DNA elongation (Sambrook *et al*, 1989). These analogs, 2'3'-dideoxy-nucleotide triphosphates (ddNTPs), lack the 3'-hydroxyl group necessary for DNA chain elongation. When proper mixtures of dNTPs and one of the ddNTPs are used, the enzyme-catalysed polymerization is terminated wherever a ddNTP is incorporated. Since the incorporation of the ddNTPs will assume a normal distribution, representative strands differing by a single nucleotide will be found in the reaction mixtures. When electrophoresed in high resolution denaturing acrylamide gels, these appear as sequencing ladders (Sambrook *et al*, 1989).

CHAPTER 2

MATERIALS AND METHODS

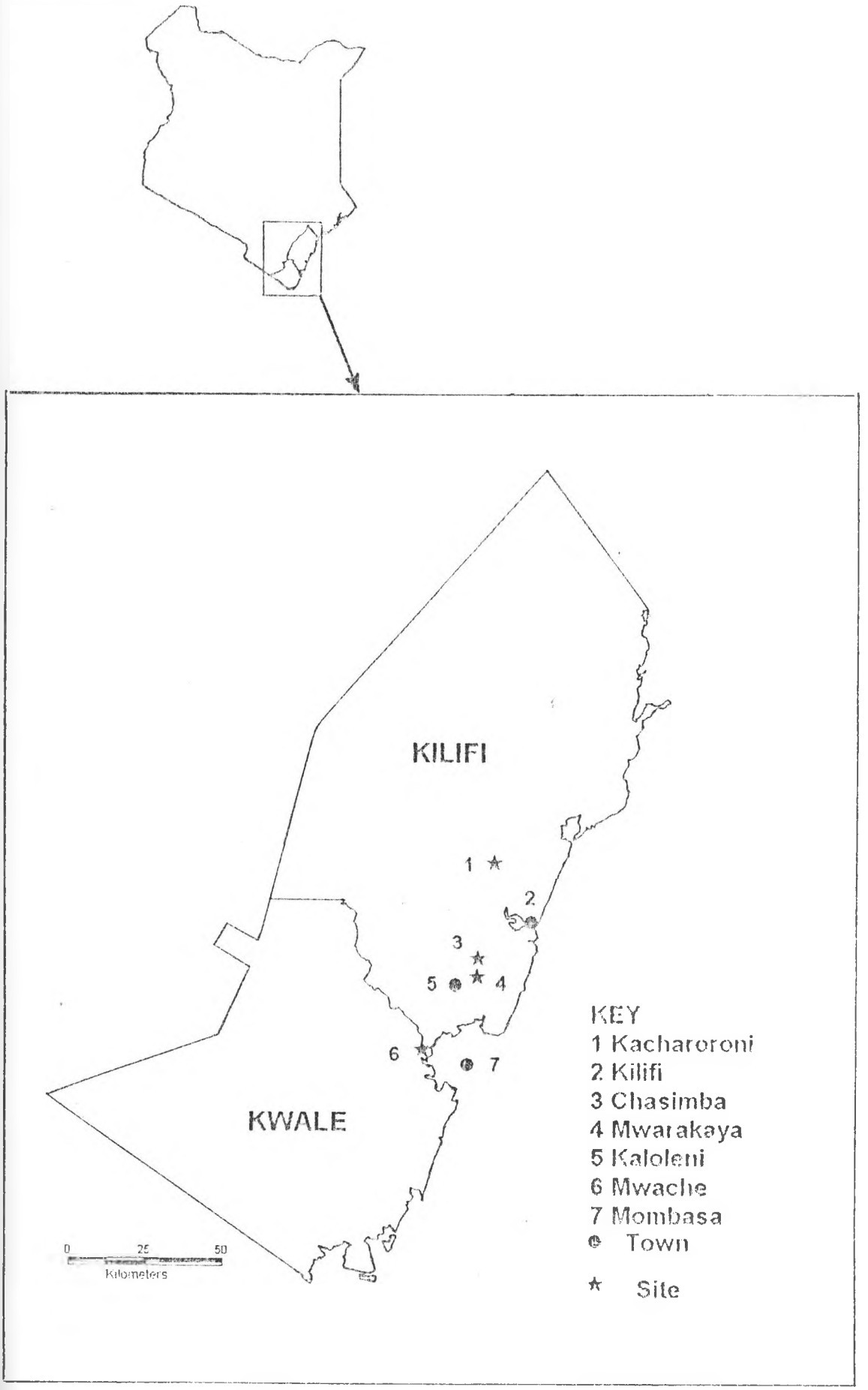
2.0 Collection of plant material from the field

All the four sites, namely Mwarakaya, Cha Simba, Mwache and Kacharoroni were visited and samples of *Saintpaulia rupicola* and *Saintpaulia* sp. nov. collected. Where possible at least three leaves of one plant were collected. One or two leaves were frozen in dry ice for DNA extraction. Another leaf was kept in wet cotton wool for propagation.

Leaves from different plants were given different collection numbers and botanical passport data like date of collection, locality, GPS co-ordinates, habit and habitat details. In most cases, photographs and slides of the plants and the sites from which they were collected were taken. All leaves collected were successfully propagated into plants at the PCP, NMK. They are kept as an *ex-situ* reference collection. Whenever necessary, more material for DNA analysis could be obtained here.

The plant samples collected from the populations were given collection numbers (e.g. Pearce 101 to 159). Each number represents a different plant chosen at random from the total population. Leaves were propagated to full plants and are kept *ex-situ* at the nursery of PCP, NMK, Nairobi. Collections from Kacharoroni were given collection numbers 101 to 114, Cha Simba 115 to 130, Mwarakaya 131 to 145 and Mwache 146 to 159. Figure 3 shows the location of the sites from which samples were collected.

Figure 3: Map of Kenya showing sites from which samples were collected.



2.0.1 Collection of *Saintpaulia* sp nov. samples from Kacharoroni

GPS location: S3° 27.56'; E 39° 44.61, Altitude, 85 m

The plants grew on the right hand slopes (looking downstream) of the Rare (Vitengeni) river in the Kacharoroni forest. In places where the plants were well shaded, they were quite plentiful on the limestone outcrops.

2.0.2 Collection of *S. rupicola* samples from Cha Simba

GPS location:

Main outcrop: S 3° 44.37'; E 39° 41.69', Altitude, 190 m

Smaller outcrop: S 3° 44.22'; E 39° 41.83', Altitude, 240 m

The Cha Simba area has three limestone outcrops which lie about 500 m apart. Samples were collected from the main outcrop which occurs in the middle of the other three and the one which occurs to the north. This latter outcrop, which had not been previously explored by the PCP team has only a few African Violets as compared to the main outcrop.

2.0.3 Collection of *Saintpaulia rupicola* samples from Mwarakaya

GPS location: S 3° 47.49'; E 39° 41.69', Altitude, 140 m

The Mwarakaya population is located about 2 km south of the Cha Simba population. It has three main limestone outcrops arranged in a northerly-southerly direction. Samples were collected from the two northernmost outcrops which had not been previously visited. The main threat to this population is cultivation and clearing of trees for building and firewood. A comparison of *S. rupicola* and *Saintpaulia* sp. nov. is shown in Plate 1.

2.0.4 Collection of *Saintpaulia* sp. nov. samples from Mwache

GPS location: S 4° 00.15; E 39°32.08', Altitude 100 m

The Mwache population occurs in the Mwache forest in Kilibole East, Kwale district. This population had not been visited previously by the PCP team. It is the most threatened of all the populations since the limestone on which the African Violets grow is being quarried away. Clearing of forest cover exposes the plants to a harsher micro climate than they can endure (Plate 2). Furthermore, due to clearing of the natural forest cover, new invasive species (e.g. *Lantana camara*) overgrow the *Saintpaulia*.

Plate 1: Field photographs illustrating morphological differences between *Saintpaulia rupicola* and the *Saintpaulia* sp. nov. **A**: *Saintpaulia rupicola* (Mwarakaya population) growing on a well shaded, moist limestone outcrop. **B**: *Saintpaulia* sp. nov. (Mwache population) partly exposed to sunlight due to cutting down of trees.

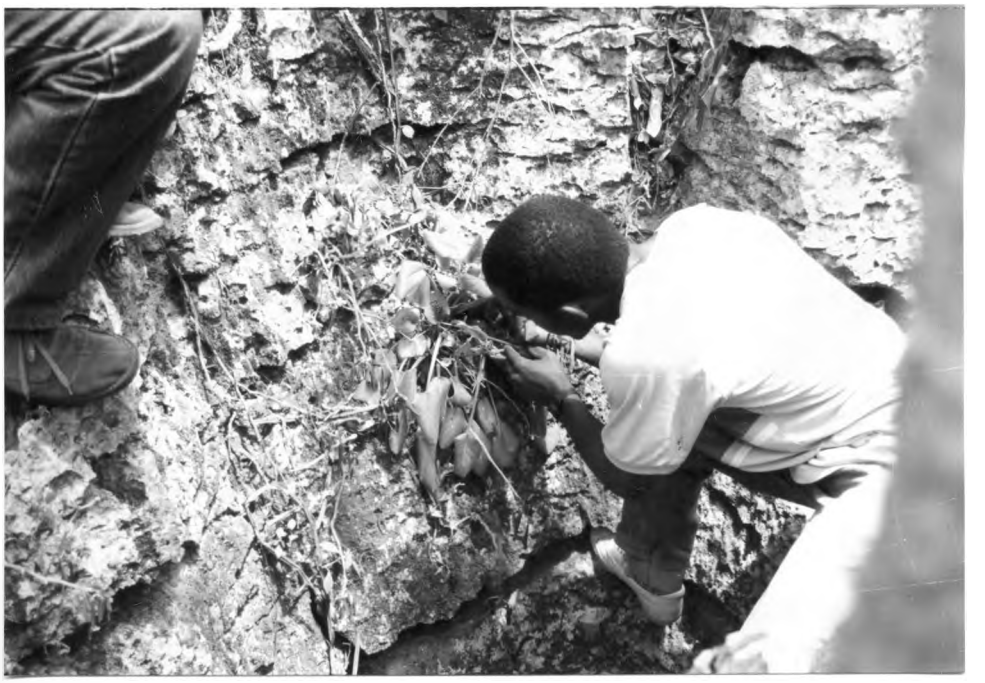


Plate 2: Field photograph of *Saintpaulia* sp. nov. (Mwache population) growing on an exposed limestone outcrop. Notice the shrivelled leaves. At the time of taking this photograph, this area was being cleared and the limestone quarried away for building purposes.

2.1 Extraction of DNA

2.1.1 Isolation of total (genomic) DNA: The CTAB procedure

Total (genomic) DNA was extracted using Doyle and Doyle's (1987) fresh tissue modification of the CTAB isolation procedure of Saghai-Maroof *et al* (1984).

About 3 g of the leaf tissue (fresh or fresh frozen), or flowers and flower buds if available, were ground in a mortar (pre-heated to 65°C) using 18 ml of 3x CTAB extraction buffer (see appendix 2). CTAB in the extraction buffer disrupts the membrane so that the DNA is released into the extraction buffer. EDTA protects the DNA from nuclease degradation by binding magnesium ions generally considered a necessary cofactor for most nucleases. This mixture was poured into a 50 ml falcon tube. The mortar and pestle were rinsed with 2 ml of the CTAB buffer, and this added to the 50 ml tube. This mixture was then incubated at 65°C for about 3 hr, with occasional gentle mixing and extracted once with ice-cold chloroform: isoamyl alcohol (24:1) to remove chlorophyll and other hydrophobic components and to denature and separate proteins from the DNA.

To resolve the resulting phases, the extract was centrifuged at 3000 rpm for 5 min in a Heraeus Sepatech Omnifuge, and the top, clear aqueous phase removed with a 5 ml pipette and transferred to two 15 ml falcon tubes each filled to about 8 ml.

The DNA in the aqueous phase was precipitated by adding about two third volumes (5 to 6 ml) of cold (-20°C) 2-propanol and mixing gently, then incubated overnight at -20°C or for 15 min at -70°C. The precipitated DNA was pelleted by

centrifugation at 4000 rpm for 10 min and pouring off the supernatant. The DNA pellet was washed by adding 1 ml wash buffer (appendix 2) and leaving it for 20 min having dislodged the pellet. It was then centrifuged at 3000 rpm for 5 min, the supernatant poured off and drained upside down for a few minutes until the pellet was dry.

The dry pellet was resuspended for 30 min in TE (appendix 2) at 65°C. RNA was digested using DNAse-free RNAse (20 μg/ml) at 37°C for 30 min. The quality of the DNA was tested by running it on 0.8 % agarose gel alongside a high molecular weight lambda DNA marker.

Extra DNAs from the genus *Saintpaulia* were obtained from the Royal Botanic Gardens, Kew, to be used as outgroups for comparison. However, for the purposes of this work, only one outgroup, *S. brevipilosa,* has been used.

2.1.2 Extraction of replicative form (RF) plasmid DNA from *E. coli*

RF plasmids were extracted from transformed *E. coli* as described by Maniatis *et al* (1989). The bacterial colony containing the plasmid was grown in 3 ml of LB medium and grown overnight with vigorous shaking at 37°C. The resulting medium was centrifuged in 1.5 ml tube at 14000 rpm for 2 min. The supernatant was poured out and more medium added and centrifuged. The supernatant was again removed by gentle aspiration, leaving the pellet as dry as possible.

The plasmid was then extracted from the bacterial pellet using a modified version

of the alkaline lysis method (Birnboim and Doly, 1979; Isc-Horowicz and Burke, 1981). The pellet was resuspended in 100 μl ice-cold solution I (see appendix 2) by vigorous vortexing followed by addition of 200 μl of solution II (appendix 2). The contents were mixed by inverting the tube rapidly for five times. One hundred and fifty μl of ice-cold solution III (appendix 2) was then added, and the tube vortexed gently in an inverted position to disperse solution III through the viscous bacterial lysate.

The resulting solution was centrifuged at 14000 rpm for 5 min at 4°C in a microfuge, and the supernatant transferred to a fresh tube. DNAse-free pancreatic RNAse (appendix 2) was added at a final concentration of 20 μg/ml, and incubated at 37°C for 30 min to digest all the RNA. This was then extracted twice with phenol-chloroform and twice with chloroform. The double stranded DNA was precipitated by adding two and a half volumes of ethanol and incubating for 15 min at -70°C then spinning at 14000 rpm for 10 min at 4°C in a microfuge. The supernatant was removed by gentle aspiration, the pellet washed with 1 ml of 70 % cold (-20°C) ethanol and resuspended in 30 μl TE. Two μl of this was analysed by electrophoresis on 1 % agarose gel.

2.2 Determination of the amount of the DNA

The amount of genomic DNA extracted was determined by measuring the Optical Density (OD) at 260 nm (OD_{260}). The OD at 280 nm (OD_{280}) was also taken to measure the amount of protein contaminants. The OD_{260}/OD_{280} ratio provided an estimate of the purity of the nucleic acid. Pure preparations of DNA have an

OD_{260}/OD_{280} value of 1.8. If there is contamination with protein or phenol, this value is less, and accurate quantitation of the DNA is not possible (Sambrook *et al*, 1989).

The samples that yielded sufficient quantities of genomic DNA were then selected for CsCl purification. These were quantified by taking the spectrophotometric measurements of the OD at 260 and 280 nm and by comparing their intensity under UV with that of a high molecular weight lambda DNA marker after electrophoresis as explained in section 2.3. Where more than one extraction was done per sample, the DNAs were pooled before purification.

2.3 CsCl purification of genomic DNA

The DNA extracted by the CTAB method was further purified on a CsCl gradient. Several DNA samples were thus selected for further purification on CsCl gradients as described by Sambrook *et al* (1989). CsCl was dissolved in TE to give a refractive index of 1.386 and EtBr added to a final concentration of 200 μg/ml. For each ml of the DNA suspension, 1 g of freshly powdered Cscl was added in SW55Ti centriguge tubes. These tubes were then filled with the CsCl/EtBr solution and covered with a drop of glycerol.

They were then centrifuged for 24 hr in a Beckman Ultracentrifuge using an SW55Ti rotor at 55000 rpm. The band of genomic DNA, evident by its luminiscence under UV, was removed by piercing the tube with a hypodermic needle and sucking it out in a volume of 0.5-1 ml.

The EtBr was removed by first diluting the DNA two-fold with ddH_2O, followed by three extractions with 2 ml isopropanol. Each time, the isopropanol was mixed gently with moderate shaking and left to stand for 5 min for phase separation. Isopropanol containing EtBr forms the upper purple layer. On the last extraction, the tubes were centrifuged at 3000 rpm for 5 min to enhance phase separation. The CsCl was removed by precipitation of the DNA with six volumes of 96 % ethanol (-20°C), pelleted and washed in 70 % cold ethanol. It was then air dried by inverting the tube for 20-30 min and resuspended in 200 µl of TE. The DNA was reprecipitated with two and a half volumes of 96 % ethanol (-20°C), washed, air dried and resuspended in 200 µl of TE. The quality of this DNA was checked by electrophoresis of 5 µl on 1 % agarose gel.

2.4 Agarose gel electrophoresis

Agarose gel electrophoresis was carried out as described by Sambrook *et al* (1989). EtBr at a final concentration of 0.5 µg/ml was routinely included in the gel as well as in the running buffer in order to facilitate visualizing the DNA under UV light.

The correct amount of powdered agarose was added to a measured quantity of 1X TBE or 1X TAE buffer (appendix 2), to give a final concentration of 1 % agarose gel then heated in a microwave oven until the agarose dissolved. The solution was cooled to 50°C, EtBr added and poured into a mould prepared with the right comb. After the gel had completely set (after about 20 to 35 min), the comb was removed, and the gel mounted in the electrophoresis (submarine) tank (33 cm X 17.5 cm X 5.5 cm, Hoefer Scientific Instruments) containing 1.2 liters of the electrophoresis

buffer. Samples were mixed with DNA loading buffer (appendix 2) and loaded into the gel. An appropriate DNA marker was routinely included alongside the samples. Electrophoresis was carried out at 160 volts for 30 min.

The DNA was visualized using a Photo/Prep I UV trans-illuminator. The gel was photographed using a Polaroid MP-4 camera loaded with Polaroid 667 (ASA 3000 / 36°c) Coaterless Instant Film (Sigma).

2.5 PCR amplification of ITS region

The whole ITS region (section 1.1.1.3.1, chapter 1) was amplified from genomic DNA by the Polymerase Chain Reaction (PCR; Saiki, 1989) using the ''ITS4'' and ''ITS5'' primers (table 1) designed by White, *et al*, (1990). Amplifications were performed in 0.5 ml tubes in 40 μl reactions containing about 100 ng of genomic DNA, 2.5 μl of 200 mM dNTPs in equimolar ratio, 4 μl of 10 X *Taq* polymerase buffer (appendix 2), 3 μl of 8 pmol/μl of both primers and 0.5 μl of 2.5 U/ml *Taq* DNA polymerase (Perkin Elmer). The final volume of each reaction was made up with sterile triple distilled, water.

Each reaction mixture was sealed with a drop of mineral oil, and the contents briefly centrifuged to bring them to the bottom of the tube.

PCR reactions were carried out in a Hybaid thermal reactor, programmed to denature the DNA at 94°C for 1 min, anneal the primer to the single stranded DNA at 48°C for 1 min and extend the annealed primer at 72°C for 2 min for 35 cycles,

with the final extension at 72°C for 10 min to allow complete synthesis of unfinished DNA strands.

The PCR products were analyzed by electrophoresis of 4 μl of the PCR product on a 1 % EtBr-stained agarose mini-gel. The ''ITS4'' and ''ITS5'' primers should amplify a product of approximately 650 bp (see section 3.2, chapter 3 for results).

2.6 Purification of the PCR product

The PCR product was purified by the gene-clean procedure according to the manufacturer's protocol (Promega). The DNA band was excised from the Etbr-stained gel under long wave UV light and the gel slice then transferred to a 1.5 ml eppendorf tube. Six hundred μl of NaI stock solution (appendix 2) was added and the tube placed in a waterbath at 55°c for about 1 hr with intermittent vortexing. When the gel had completely dissolved, the GLASSMILK suspension was vortexed for about 1 min until it was all in suspension and 5 μl of this added to the dissolved agarose. This was then incubated on ice for 5 min for the GLASSMILK (Silica 325 mesh) to bind the DNA, vortexing every 1 min to ensure the GLASSMILK stayed in suspension.

The suspension of bound DNA was then pelleted by spinning for about 20 sec and the supernatant saved in case the binding was not complete. This supernatant was later discarded. The pellet was washed three times with ethanol wash solution (appendix 2), spinning down the pellet after every wash. After the last wash, all the ethanol wash solution was carefully removed using a pipette.

The DNA was eluted from the GLASSMILK by adding 30 μl of TE and vortexing to resuspend the GLASSMILK. This suspension was incubated at 55°C for 10 min with intermittent vortexing and GLASSMILK pelleted by centrifugation at 14,000 rpm for 30 sec. The upper 25 μl was removed taking care not to pick the GLASSMILK and put into a fresh eppendorf tube. Five μl of this purified DNA was then analyzed by electrophoresis on 1 % agarose (section 3.2).

2.7 Cloning of purified PCR product

The purified PCR product was cloned into the pGEM-T vector (Promega) using the pGEM-T vector Systems and following the manufacture's protocol. This vector accepts inserts of fresh PCR products in a region which contains a mutation which in complement with the host vector allows for blue/white colour selection when the plasmid and cells are grown in the presence of IPTG and X-gal (appendix 2).

Ligation reactions were set up in 10 µl reactions containing 2 µl of the purified PCR product, 1 µl (50 ng) of the pGEM-T vector, 1 µl of the T4 DNA ligase 10X buffer, 1 µl (1 unit) of T4 DNA ligase and 5 µl of ddH_2O. After a 3 hr incubation at 15°C, the reactions were stopped by heating for 10 min at 72°C and allowed to cool down to room temperature.

The ligation mixture was then transformed into high efficiency JM 109 *E. coli* competent cells provided in the cloning system. The competent cells were thawed on ice for about 5 min. Two μl of the ligated PCR product:pGEM-T vector was added to 50 μl of the competent cells in a sterile test tube and incubated on ice for

20 min, heat-shocked at 42°C for 1 min then incubated on ice for 2 min. Five hundred μl of SOC medium (appendix 1) was then added, and the cells grown with shaking at 37°C for 1 hr. Fifty μl of these cells were spread on YT/amp/IPTG/X-gals plates (appendix 1) and incubated overnight at 37°C.

2.7.1 Selection of recombinant vector

2.7.2 Colour selection

The resultant colonies were selected by their colour; white colonies indicate the presence of the insert, whereas blue colonies indicate a vector without an insert.

2.7.3 Selection by PCR amplification

Colour selection of recombinant vector may sometimes be inefficient in case of mutations in the ß-galatosidase gene which may hinder production of active ß-galactosidase, and the consequential formation of a blue colour when the phage and cells are grown in the presence of IPTG and X-gal. A second screening method was thus necessary.

The white colonies were picked using a wire loop, put into 100 μl of TE and vortexed to distribute the colony into the solution. Twenty μl of this plasmid suspension was boiled for 10 min and centrifuged for 5 min to pellet any impurities present. Five μl of this was used as the template in a 10 μl PCR reaction containing 0.2 μl of 10 mM dNTPs in equimolar ratio, 1 μl of 10 X *Taq* polymerase buffer (appendix 2), 1 μl of 10 mM each of M13 forward and reverse primers and 0.1 ml of *Taq* DNA polymerase (2.5 U/μl). The final volume of each reaction was made

up with triple distilled, autoclaved water.

PCR cycles were set at an initial denaturation temperature of 94°C for 3 min, followed by 30 cycles of a denaturation temperature of 94°C for 30 sec, annealing temperature of 50°C for 30 sec and extension temperature of 72°C for 1 min, with a final extension of 10 min at 72°C. The PCR products were analysed on 1 % agarose gel.

2.7.4 Small-scale preparation of replicative form (RF) plasmid DNA

RF plasmids were extracted from colonies already identified to be positive by colour and PCR selection as described in section 2.1.2.

2.8 DNA sequencing

DNA was sequenced by the dideoxynucleotide chain-termination method described by Sanger *et al* (1977) using the Sequenase™ Version 2.0 DNA Sequencing kit from United States Biochemical and following the protocol detailed therein. Two individuals per population were sequenced, one through the direct sequencing of the PCR product and the other from a positive clone.

2.8.1 Sequencing of recombinant plasmid DNA

Between 3 to 5 μg of the recombinant plasmid DNA was denatured for 30 min at 37°C in 0.1 volumes of 2 M NaOH, 2 mM EDTA. It was then precipitated with NaOAc/absolute ethanol, pelleted, washed and resuspended in 7 μl double distilled water. Two 2 μl of the sequenase (5X) reaction buffer and 1 μl (0.5 pmol) of the

primer was then added.

The primer was annealed to the denatured template by heating for 2 min at 65°C and allowing it to cool slowly to room temperature for 30 min. The DNA synthesis was then carried out in two steps, the labelling and the chain termination steps. In the labelling reaction, the annealed primer was extended using limiting concentrations of dNTPs, including [^{35}S] dATP. This step continues to virtual complete incorporation of labelled nucleotide into DNA chains which are distributed randomly in length from several nucleotides to hundreds of nucleotides. The labelling reaction was followed by the termination reactions, in which the concentration of all the dNTPs was increased and a ddNTP added. Processive DNA synthesis occurs until all growing chains are terminated by a ddNTP. The reaction was then stopped by adding formamide stop solution (see appendix 2).

2.8.2 Direct sequencing of the PCR product

The purified PCR product was directly sequenced as described above, except that the double stranded DNA was denatured by heating at 95°C for 5 min and quickly chilled on ice for 2 min. One μl of the sequencing primer and 2 μl of the 5X sequencing buffer were then added. The labelling and the termination reactions were then done as described in the Sequenase protocol.

2.9 Preparation of sequencing gel and electrophoresis

Electrophoresis was carried out in 6 % denaturing acrylamide gels. This was made by mixing 50 g of ultra-pure urea, 10 ml of 10X TBE and 15 ml of 40 %

acrylamide (38 g acrylamide, 2 g bis-acrylamide, and made to 100 ml with sterile water). The solution was warmed to dissolve the urea, made up to 100 ml and filtered through a filter with a pore size of 0.45 microns. Polymerization of the acrylamide was initiated by the addition of 350 μl 10 % APS and 30 μl TEMED.

The gel was cast between two sequencing glass plates separated by 0.4 mm spacers. Before casting, one plate was siliconized by wiping it with 2 % dichlorodimethylsilane in 1,1,1-trichloroethane, and then with distilled water. Gels were electrophoresed at 2000 V/ 70 W with the buffer chambers filled with 1X TBE. A sharkstooth comb was placed between the plates with the flat side touching the gel, and the gel allowed to polymerize overnight. It was then mounted onto the electrophoresis unit, and the comb placed the other way round in such a way that the combs just touched the gel.

The samples were denatured at 90°C for 2 min and loaded onto the gel. Each set of four sequencing reactions of the same DNA template were loaded into adjacent slots. Each sample was run in three sets, the first for two hours, the second for four hours and the third for six hours. It was routinely possible to read between 360 to 400 bases from the primer.

After completion of electrophoresis, the gel was transferred to 8 mm whatman paper and dried at 80°C under vacuum for 2 hr. The dry gels were then exposed to X-ray film for 12-48 hr at room temperature without intensifying screens.

3.0.0 Alignment and analysis of DNA sequences

The generated sequences were first aligned and analyzed using Geneworks Release 2.3 (IntelliGenetics, 1991) on a Power Macintosh computer.

To find the accurate phylogenetic tree topology, three tree reconstruction methods were used. The sequences aligned using Geneworks Release 2.3 were analyzed using the UPGMA (Unweighted Pair Group Method with Arithmetic mean). This gives calculated relationships of the aligned sequences (Nei, 1987). Further analysis was done using the MUTALIN program. Various phylogenetic programmes within the Phylogenetic Inference Package (PHYLIP) were used for analysis of the aligned sequences. First, a distance matrix was calculated by the pairwise sequence comparisons assuming a Kimura 2-parameter of evolution (Kimura, 1980). A distance-based UPGMA tree was then calculated using the distance matrix. Cladistic analysis was carried out using the DNAPARS (DNA parsimony) and DNAML (DNA Maximum Likelihood Method) programmes in PHYLIP.

CHAPTER 3

RESULTS

3.0 Extraction and quantification of genomic DNA

Plate 3 shows the results of extraction of genomic DNA before and after CsCl extraction. Flowers and flower buds yielded approximately twice as much DNA as leaves. OD measurements before CsCl purification (Table 2) indicated low DNA yield (chapter 4).

3.1 PCR amplification of the ITS region and purification of the PCR product

All the genomic DNA samples selected and purified for analysis responded positively to PCR amplification. This amplification of the ITS region produced an amplification product of approximately 650 bp (Plate 4).

3.1.1 Cloning of the purified PCR product

The method of screening adopted for the identification of positive clones, namely PCR amplification of the ITS region from the recombinant plasmid produced an amplification product of approximately 650 bp which confirmed the presence of the insert (Plate 5). Positive clones from the Mwarakaya, Mwache, and Kacharoroni populations were obtained. It was not possible to obtain a positive clone from the Cha Simba population. The positive colonies were grown to extract the recombinant plasmid DNA for sequencing (Plate 6).

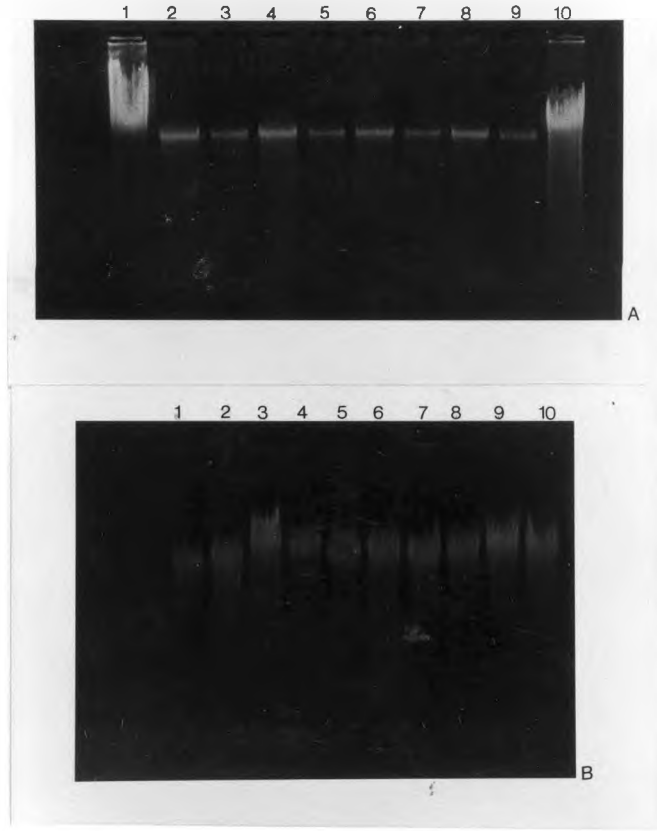


Plate 3: Agarose gels showing extracted genomic (from *S. rupicola*) and chloroplast (from spinach) DNAs. **A**: 1, High molecular weight lambda DNA (500 μg/ml); 2, Kacharoroni (Pearce 461/93/105) flowers and flower buds; 3, Kacharoroni leaves; 4, Mwache (Bytebier 154) flowers and flower buds; 5, Mwache leaves; 6, Mwarakaya (Bytebier 145) flowers and flower buds; 7, Mwarakaya leaves; 8, Chasimba (Pearce 463/93/106) flowers and flower buds; 9, Chasimba leaves; 10, Spinach purified chloroplast DNA. **B:** The selected DNAs after CsCl purification. 1, High molecular weight lambda DNA (500 μg/ml); 2-10, selected DNAs (*S. rupicola* and *Saintpaulia* sp. nov.) purified by CsCl method.

Table 2: Spectrophotometric determination of the amount of DNA before purification.

Extraction no	Collection no	Species	Population/ source	OD_{260}	OD_{280}	OD_{260}/OD_{280}	Concentration (μg/ml)
1	Pearce 101	*Saintpaulia* sp. nov	Kacharoroni	0.009	0.138	0.065	112.5
2	Pearce 102	*Saintpaulia* sp. nov	Kacharoroni	0.012	0.135	0.034	150
3	Pearce 103	*Saintpaulia* sp. nov	Kacharoroni	0.008	0.133	0.060	100
3B	Pearce 463/93 106	*saintpaulia rupicola*	Cha Simba	0.009	0.108	0.083	112.5
4	Pearce 104	*Saintpaulia* sp. nov	Kacharoroni	0.011	0.132	0.083	137.5
5	Pearce 115	*Saintpaulia rupicola*	Cha Simba	0.008	0.132	0.060	100
6	Pearce 129	*Saintpaulia rupicola*	Cha Simba	0.004	0.132	0.033	50
6B	Pearce 131	*Saintpaulia rupicola*	Mwarakaya	0.009	0.11	0.082	112.5
7	Pearce 127	*Saintpaulia rupicola*	Cha Simba	0.029	0.145	0.2	362.5
8	Pearce 116	*Saintpaulia* sp. nov	Cha Simba	0.012	0.134	0.08	150
8B	Pearce 154	*Saintpaulia* sp. nov	Mwache	0.014	0.138	0.03	175
9	Pearce 130	*Saintpaulia rupicola*	Cha Simba	0.014	0.133	0.11	175
10	Pearce 152	*Saintpaulia* sp. nov	Mwache	0.001	0.130	0.008	12.5
11	Pearce 151	*Saintpaulia* sp. nov	Mwache	0.002	0.131	0.015	25
12	Pearce 146	*Saintpaulia* sp. nov	Mwache	0.008	0.132	0.060	100
13	Pearce 149	*Saintpaulia* sp. nov	Mwache	0.006	0.135	0.044	75
13B	Pearce 461/93/ 105	*Saintpaulia* sp. nov	Kacharoroni	0.021	0.124	0.17	262.5
14	Pearce 147	*Saintpaulia* sp. nov	Mwache	0.008	0.133	0.060	100
14B	Pearce 145	*Saintpaulia rupicola*	Mwarakaya	0.017	0.129	0.131	212.5
15	Pearce 159	*Saintpaulia* sp. nov	Mwache	0.013	0.132	0.098	162.5
16	Pearce 144	*Saintpaulia rupicola*	Mwarakaya	0.009	0.129	0.07	112.5
17	Pearce 131	*Saintpaulia rupicola*	Mwarakaya	0.021	0.132	0.159	262.5
18	Pearce 144	*Saintpaulia rupicola*	Mwarakaya	0.013	0.136	0.096	162.5
19	Pearce 141	*Saintpaulia rupicola*	Mwarakaya	0.012	0.134	0.09	150

20	Pearce 142	*Saintpaulia rupicola*	Mwarakaya	0.009	0.132	0.07	112.5
21	Pearce 132	*Saintpaulia rupicola*	Mwarakaya	0.014	0.133	0.11	175
22	Pearce 118	*Saintpaulia* sp. nov	Mwarakaya	0.025	0.139	0.18	312.5
23	Pearce 464/93/ 104	*Saintpaulia rupicola*	Mwarakaya	0.028	0.145	0.193	350
24	Pearce 150	*Saintpaulia* sp. nov	Mwache	0.013	0.138	0.094	162.5
25	Pearce 118	*Saintpaulia rupicola*	Cha Simba	0.012	0.139	0.086	150
26	Pearce 126	*Saintpaulia rupicola*	Cha Simba	0.009	0.136	0.066	112.5
26B	Pearce 463/93/ 106	*Saintpaulia rupicola*	Cha Simba	0.032	0.267	0.12	400
27	Pearce 107	*Saintpaulia* sp nov	Kacharoroni	0.014	0.138	0.101	175
28	Pearce 112	*Saintpaulia* sp nov	Kacharoroni	0.007	0.133	0.053	87.5
29	Pearce 158	*Saintpaulia* sp. nov	Mwache	0.015	0.137	0.109	187.5
30	Pearce 156	*Saintpaulia* sp. nov	Mwache	0.009	0.135	0.067	112.5
31	Pearce 148	*Saintpaulia* sp. nov	Mwache	0.008	0.133	0.060	100
32	Pearce 157	*Saintpaulia* sp. nov	Mwache	0.004	0.134	0.03	50
33	Pearce 106	*Saintpaulia* sp. nov	Kacharoroni	0.014	0.132	0.106	175
34	Pearce 111	*Saintpaulia* sp.nov	Kacharoroni	0.017	0.129	0.131	212.5
35	Pearce 113	*Saintpaulia* sp. nov	Kacharoroni	0.061	0.154	0.4	762.5
36	Pearce 119	*Saintpaulia rupicola*	Cha Simba	0.013	0.136	0.1	162.5
37	Pearce 134	*Saintpaulia rupicola*	Mwarakaya	0.024	0.138	0.17	300
38	Pearce 139	*Saintpaulia rupicola*	Mwarakaya	0.019	0.130	0.146	237.5
39	Pearce 153	*Saintpaulia* sp. nov	Mwache	0.006	0.136	0.04	75
40		*Nicotiana tabacum*		0.02	0.019	1.05	250
41		*Streptocarpus caulesens*	Kilibasi (Kwale district)	0.039	0.045	0.8	487.5
42	Luke 3788	*Saintpaulia orbicularis*	Usambara mountains (Tanzania)	0.008	0.133	0,06	100
43		*Saintpaulia magungensis* var. *occidentalis*	Usambara mountains (Tanzania)	0.060	0.158	0.38	750

44	*Saintpaulia shumensis*	Shume, West Usambara (Tanzania)	0.010	0.135	0.074	125

Table 3: *Saintpaulia* DNAs sampled for analysis and their OD measurements after CsCl purification.

Extraction no.	Colection no.	Species	Population / Source	OD_{260}	OD_{280}	OD_{260} / OD_{280}	Concentration (μg/ml)
6B	Pearce131	*Saintpaulia rupicola*	Mwarakaya	0.062	0.043	1.44	775
14B	Pearce 145	*Saintpaulia rupicola*	Mwarakaya	0.071	0.049	1.45	887.5
20	Pearce 126	*Saintpaulia rupicola*	Cha Simba	0.060	0.041	1.47	750
22	Pearce 116	*Saintpaulia rupicola*	Cha Simba	0.065	0.043	1.50	812.5
3B	Pearce 106	*Saintpaulia rupicola*	Cha Simba	0.063	0.042	1.49	787.5
26B	Pearce 463/93/106	*Saintpaulia rupicola*	Cha Simba	0.057	0.038	1.51	712.5
13B	Pearce 461/93/105	*Saintpaulia* sp. nov	Kacharoroni	0.067	0.045	1.50	837.5
17	Pearce 106	*Saintpaulia* sp. nov	Kacharoroni	0.070	0.046	1.51	875
16	Pearce 113	*Saintpaulia* sp. nov	Kacharoroni	0.051	0.035	1.47	637.5
29	Bytebier 158	*Saintpaulia* sp. nov	Mwache	0.071	0.048	1.49	887.5
30	Pearce154	*Saintpaulia* sp. nov	Mwache	0.062	0.041	1.52	775
8B	Pearce 156	*Saintpaulia* sp. nov	Mwache	0.064	0.044	1.46	800
8	Kew 1959-29203	*Saintpaulia tongwensis*	Tongwe, Usambara mts.	0.039	0.028	1.38	487.5
6	Kew 1993 s. n.	*Saintpaulia brevipilosa*	Usambara mts.	0.037	0.026	1.45	462.5
2		*Streptocarpus glandulossissimus* Engl	Gatamaiyu forest	0.072	0.048	1.50	900

Plate 4: PCR amplification products of the ITS region analyzed on 1 % agarose gel. 1. 1 kb ladder; 2, Bytebier 145 (Mwarakaya); 3 Pearce 463/93/106 (Chasimba); 4, Bytebier 154 (Mwache); 5, Pearce 461/93/105 (*Saintpaulia* sp. nov); 6, *Saintpaulia brevipilosa*; (Kew 1993 s.n.): 7, *Streptocarpus glandulossisimus* Engl.; 8, *Nicotiana tabacum*. Sizes are given in bp.

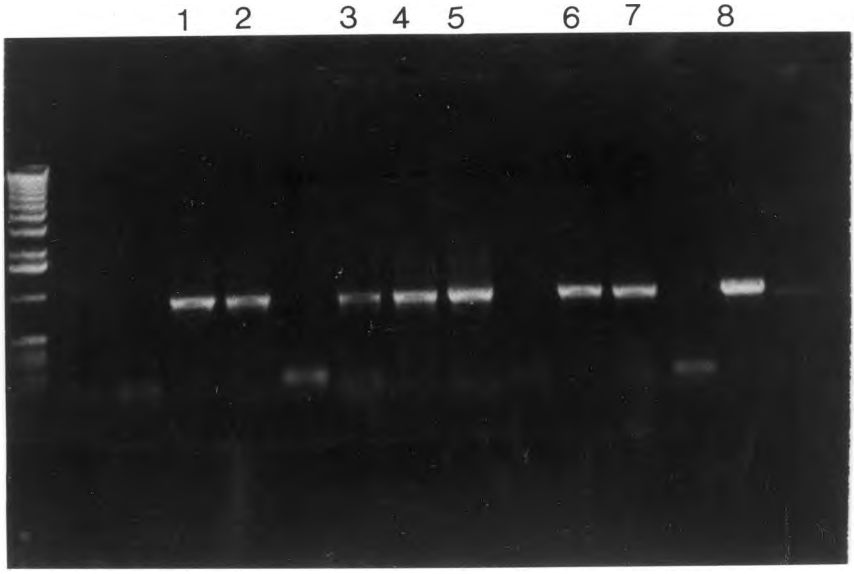


Plate 5: Selection of positive clones by PCR amplification of the insert. 1, Kacharoroni (Pearce 461/93/105), 2 & 3, Mwache (Bytebier 156); 4 & 5, Mwarakaya (Pearce 464/93/104); 6 & 7, Mwarakaya (Bytebier 145); 8, *Saintpaulia brevipilosa*.

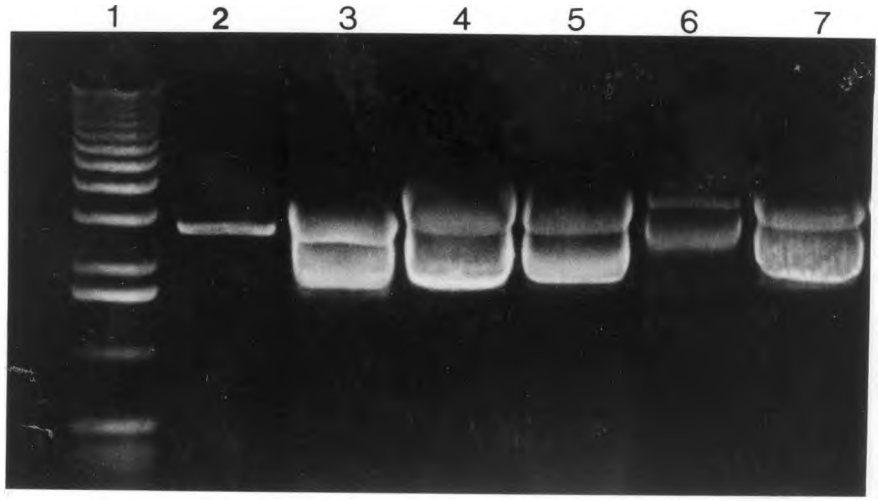


Plate 6: Preparation of purified plasmid for sequencing analyzed alongside pUC18 linearised with *Sma*1 (2.7 kb) and 1 kb ladder. 1, 1 kb ladder; 2 pUC18 linearised with *Sma*1; 3-7, recombinant RF plasmid DNA for sequencing.

3.2 DNA Sequencing

Sequences from pUC 18 positive clones were obtained from one individual each from the Mwarakaya, Kacharoroni and Mwache populations. More sequences were later obtained by direct sequencing of purified PCR products pooled from two or more reactions (see section 4.1). *Saintpaulia brevipilosa* was used as an outgroup and reference taxon against which the four populations were compared. The boundaries of the ITS and ribosomal DNA coding regions were identified by comparison to those of *Astragalus alpinus* (Fabaceae: Wojciechowski *et al*, 1993), *Arnica mollis* (Compositae: Baldwin, 1993) and *Oryza sativa* (Bambusoideae, Takaiwa *et al*, 1985). Alignment with these sequences showed homology especially in the 5.8S rDNA. The boundaries between ITS1, 5.8S rDNA and ITS2 were similarly identified.

Where possible, sequences were obtained both by direct sequencing of the pooled PCR products and sequencing of positive clones. Where sequencing was done through cloning, two clones were sequenced. No variation was observed between individuals of the same population sequenced using these two methods. Due to this and the fact that this data was not representative of all the four populations (no positive clone from the Cha Simba population was obtained), only one sequence is presented here. There was no evidence of repeat-type variation (section 4.1) in those samples sequenced by either of the two methods. Alignment of these sequences (Table 4) provided essential structural information for determination of nucleotide homology and genetic distance among these samples.

Table 4: Aligned sequences of *Saintpaulia rupicola* and *Saintpaulia* sp. nov. The sequences were aligned using Geneworks Release 2.3. Vertical columns represent nucleotide positions, indicated by numbers at the end of a row from 1 to 628 (5'-3'). ; dots denote identity to concensus. Dashes denote gaps. G=Guanine; A=Adenine; T=Thyamine; C=Cytosine. IUPAC nomenclature codes used: Y=C or T; R= A or G; K= G or T; M= A or C.

ITS1

```
Consensus       TCGAAACCTG CAAAGCAGAC CAGTGAACCT GTTCCACACC AGATYGTGGT   50
Kacharoroni     .......... .......... .......... .......... ....C.....   50
 Mwarakaya      .......... .......... .......... .......... ....C.....   50
Mwache          .......... .......... .......... .......... ....C.....   50
Chasimba        .......... .......... .......... .......... ....C.....   50
S. brevipilosa  .......... .......... .......... .......... ....T.....   50

Consensus       TGAGATGCTG GAYGCCTTTT GGGGTCGAGC ATCACTGGCT CCGACCCCAA   100
Kacharoroni     .......... ..C....... .......... .......... ..........   100
 Mwarakaya      .......... ..C....... .......... .......... ..........   100
Mwache          .......... ..C....... .......... .......... ..........   100
Chasimba        .......... ..C....... .......... .......... ..........   100
S. brevipilosa  .......... ..T....... .......... .....-.... ..........    99

Consensus       RCGATGAAAG TCGYTTGGGT GAGTAATAAC CTCTCGGCGC ACGAAGCGAA   150
Kacharoroni     G......... ...T...... .........- -......... ..........   148
 Mwarakaya      G......... ...T...... .......... .......... ..........   150
Mwache          G......... ...T...... .......... .......... ..........   150
Chasimba        G......... ...C...... .......... .......... ..........   150
S. brevipilosa  A......... ...T...... .......... .......... ..........   149

Consensus       GGAAAACCAT ACCGGATATC TCTCCATCTT GGTGCTGTCT KCKGTATCCA   200
Kacharoroni     .......... .......... .......... .......... T.T.....-.   197
 Mwarakaya      .......... .......... .......... .......... T.T.......   200
Mwache          .......... .......... .......... .......... T.T.......   200
Chasimba        .......... .......... .......... .......... T.T.......   200
S. brevipilosa  .......... .......... .......... .......... G.G.......   199
```

5.8S

Consensus	AGACGTGATG GGGAACATCT ATTGARCACA RTATACATGT ATAACGACTC	250
Kacharoroni	....=..G.... G..... ..-.....	245
Mwarakaya	....G.... G.....	250
Mwache	....G.... G.....	250
Chasimba	....G.... G.....	250
S. brevipilosa	....A.... A.....	249
Consensus	TCGGCAACGG ATATCTCGGC TCTCGCATCG ATGAAGAACG TAGCGAATGC	300
Kacharoroni	....	295
Mwarakaya	....	300
Mwache	....	300
Chasimba	....	300
S. brevipilosa	....	299
Consensus	GATACTTGGT GTGAATTGCA GAATCCGTGA ACCATCGAGT CTTGAACGCA	350
Kacharoroni	....	345
Mwarakaya	....	350
Mwache	....	350
Chasimba	....	350
S. brevipilosa	....	349
Consensus	AGTTGCGCCS AAGCCATCTG GTTGAGGCAM CGTGCTGCCT GGGCGTCACG	400
Kacharoroni	.....G ....A ...=.....	394
Mwarakaya	.....C ....A ...=.....	399
Mwache	.....G ....-.C	399
Chasimba	.....C ....A	400
S. brevipilosa	.....C ....A ...=.....	398

ITS2

```
Consensus       CATCTCGTTG CCATCCCTTG GGTCATCYGT TCCCACCAAG TGTCGGGGGC  450

Kacharoroni     .......... ...-...... .......T.. .......... ..........  443
 Mwarakaya      .......... ...-...... .......T.. .......... ..........  448
Mwache          .......... ...-...... .......T.. .......... ..........  448
Chasimba        .......... ...-...... .......T.. .......... ..........  449
S. brevipilosa  .......... .......... .......C.. .......... ..........  448

Consensus       GGATGCTTGC AGKGGAGTGT GGATATTGAC CTCCCGTCAC CTAGTGTAGC  500

Kacharoroni     .......... ..G....... .......... .......... ..........  493
 Mwarakaya      .......... ..G....... .......... .......... ..........  498
Mwache          .......... ..G....... .......... .......... ..........  498
Chasimba        .......... ..G....... .......... .......... ..........  499
S. brevipilosa  .......... ..T....... .......... .......... ..........  498

Consensus       GGCTGGCCCA AATAGTATAC CGTGTCGATG TATCACACTA TACGTGGTTG  550

Kacharoroni     .......... .......... .......... .......... ..........  543
 Mwarakaya      .......... .......... .......... .......... ..........  548
Mwache          .......... .......... .......... .......... ..........  548
Chasimba        .......... .......... .......... .......... ..........  549
S. brevipilosa  .......... .......... .......... .......... ..........  548

Consensus       TTGTATTCAC GACTTGGCAA CTGTCGTTCG AACATCGAGC CACGGGCACG  600

Kacharoroni     .......... .......... .......... .......... ..........  593
 Mwarakaya      .......... .......... .......... .......... ..........  598
Mwache          .......... .......... .......... .......... ..........  598
Chasimba        .......... .......... .......... .......... ..........  599
S. brevipilosa  .......... .......... .......... .......... ..........  598
```

Consensus	ACCCAATAGG CACAAGCTGT CTTCGGTTGC G
Kacharoroni	.... .T..... .
Mwarakaya	.... .C..... .
Mwache	.... .T..... .
Chasimba	.... .T..... .
S. brevipilosa	.... .C..... .

631

624
629
629
630
629

3.3 Analysis of the ITS sequences

A total of five phylogenetic trees were obtained. A distance-based UPGMA tree calculated using the distance matrix was similar to the one obtained using Geneworks Release 2.3. with exclusion of indels for analysis (Figure 4A). The MUTALIN programme produced a genetic distance matrix (Table 6). one tree using the DNA Maximum Likelihood Method (Figure 5) and two most parsimonious trees, (Figure 6) using DNA parsimony.

3.3.1 ITS length variation and C-G content

The sequences obtained displayed similarity with the published sequences of *Oryza sativa*. *Astragalus alpinus* and *Arnica mollis* as noted in section 3.3. However, none of the sequences could provide a good outgroup because they displayed a very distant genetic variation which distorted analysis of the African Violets. The entire ITS region varied in length among all the samples analysed from 624 bp in the Kacharoroni population to 630 bp in the Cha Simba population (table 4). ITS1 ranged in length from 238 to 241 bp. The 5.8S rDNA varied from 179 to 181 bp. No size variation was found in ITS2. Alignment of all the sequences required insertion of 5 (0.8 %) gaps in ITS1, 4 (0.6 %) in 5.8S rDNA and none in ITS2. The small variations in length were due to insertion/deletion (indel) events. These are analyzed in Table 5.

Table 5: Sequence variation in the ITS region in *Saintpaulia rupicola*, *Saintpaulia* sp. nov. and *S. brevipilosa*.

Region	Average length	Constant sites (%)	Variable sites (%)	Un-informative sites (%)	Informative sites (%)
ITS1	240	227(94.6)	13(5.4)	13(100)	0(0)
5.8S	180	174(96.7)	6(3.3)	4(66.7)	2(33.3)
ITS2	208	205(98.6)	3(1.4)	2(66.7)	1(33.3)
Total ITS	628	606(96.5)	22(3.5)	19(86.4)	3(13.6)

3.3.2 Genetic and evolutionary distances

The genetic distances observed (Table 6) were low, ranging from 0.0000 to 0.0177. The greatest genetic distances were observed between *S. brevipilosa* and the Kenyan populations of the African Violet. The low genetic distances observed among *S. rupicola* and *Saintpaulia* sp. nov. (between 0.0000 to 0.0048) indicate that the four populations are closely related, the most closely related being Kacharoroni and Mwache populations.

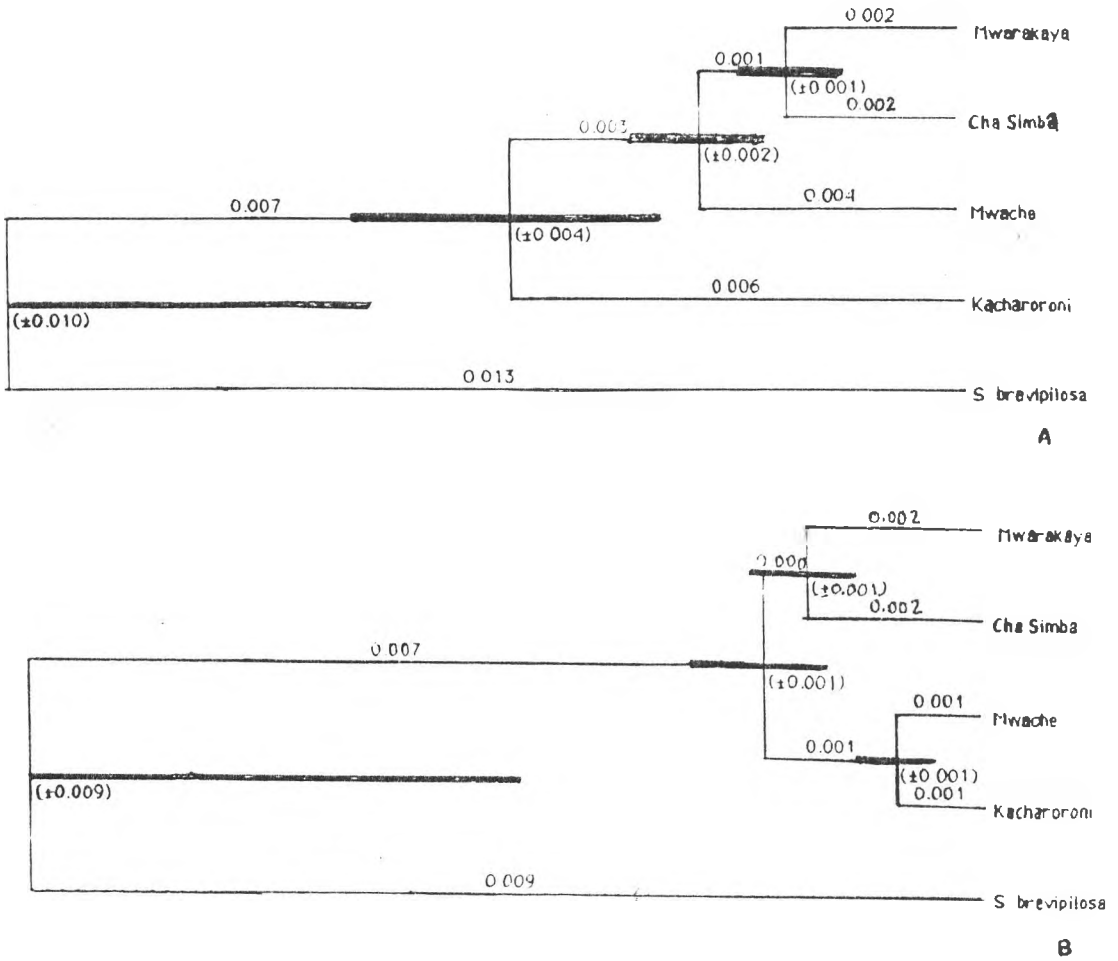


Figure 4: Unweighted Pair Group Method with Arithmetic Mean (UPGMA) evolutionary tree generated using Geneworks Release 2.3. The length of the horizontal lines connecting one sequence to another is proportional to the estimated genetic distance between the sequences. The lines at branch points in the tree are error bars showing the standard error of the branch position. **A**, including indels, and **B**, excluding indels, in phylogenetic analysis

Table 6: Pairwise comparisons of nucleotide substitutions of the five ITS sequences using Kimura 2- parameter.

	S. brevipilosa	Cha Simba	Mwarakaya	Mwache	Kacharoroni
S. brevipilosa	0.0000	0.0177	0.0145	0.0194	0.0179
Cha Simba	0.0177	0.0000	0.0032	0.0048	0.0032
Mwarakaya	0.0145	0.0032	0.0000	0.0048	0.0032
Mwache	0.0194	0.0048	0.0048	0.0000	0.0016
Kacharoroni	0.0179	0.0032	0.0032	0.0016	0.0000

All the trees obtained (figures 4, 5 and 6) reveal that *S. brevipilosa* is the most evolutionally distant among the five individuals analyzed, with *S. rupicola* and *Saintpaulia* sp. nov. being monophyletic. The nucleic acid sequence Maximum Likelihood method and the DNA parsimony algorithm reveal that the Mwarakaya population evolved earlier than the others. The DNA parsimony analysis produced two most parsimonious phylogenetic trees. These differ only in the positioning of the Mwarakaya population. In the first tree (figure 6A), although distant, the Mwarakaya population still clusters with the others as monophyletic. This tree agrees with that generated by the Maximum Likelihood method. In the second tree (figure 6B), the Mwarakaya population is excluded from the cluster of the other populations. In both cases, the Mwache population is revealed to be the most evolutionary recent and most closely related to the Kacharoroni population. These observations are in agreement with the UPGMA analysis excluding indels (figure 4A).

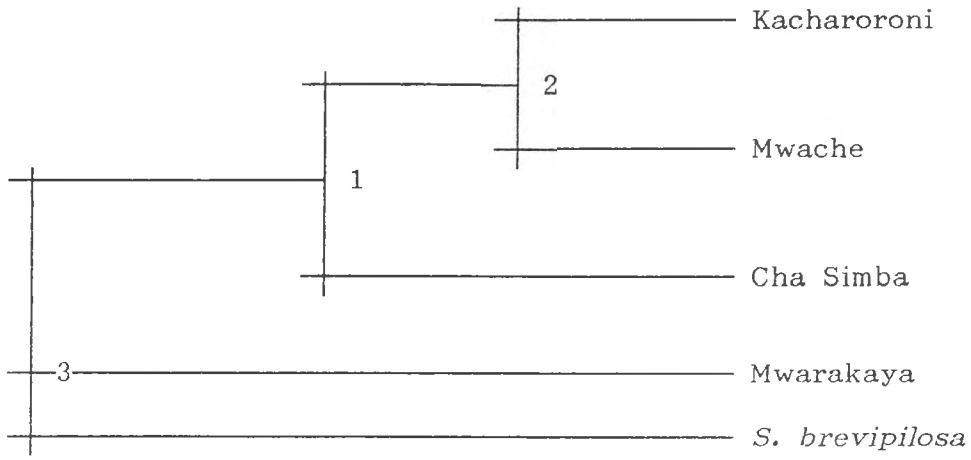
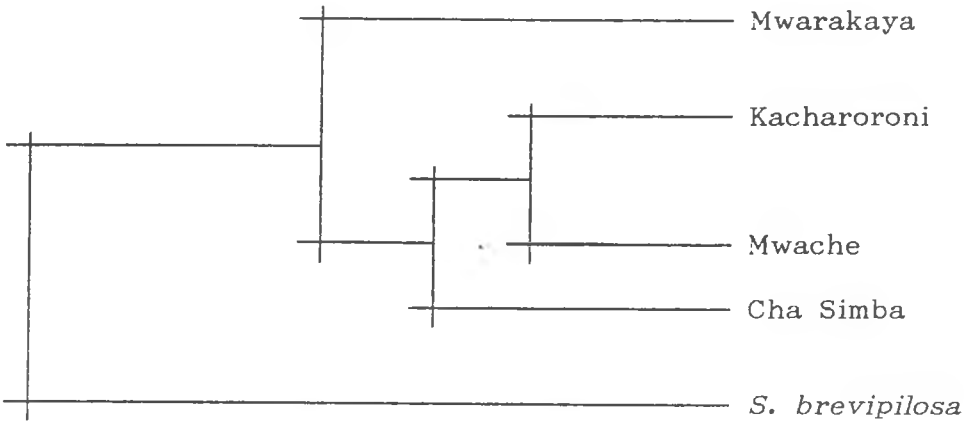


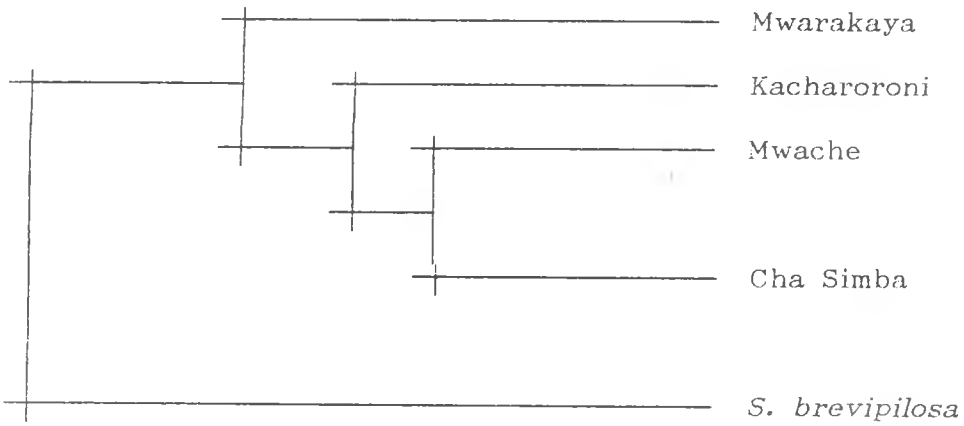
Figure 5: The phylogenetic tree (unrooted) inferred from the ITS sequences, generated by the nucleic acid sequence Maximum Likelihood method, version 3.5c. The bootstrap values and confidence limits are given in table 7.

Table 7: Bootstrap values and confidence limits for figure 5. *=Significantly positive, P<0.01

Between	And	Length	Approximate confidence limits
3	1	0.00159	0.00000-0.00471*
1	2	0.00156	0.00000-0.00456*
2	Kacharoroni	0.00003	0.00000-infinity
2	Mwache	0.00160	0.00000-0.00472*
1	Cha Simba	0.00163	0.00000-0.00479
3	Mwarakaya	0.00006	0.00000-infinity
3	*S. brevipilosa*	0.01452	0.00504-0.02403*



A



B

Figure 6: The two most parsimonious phylogenetic trees (unrooted) inferred from the ITS sequences, generated by the DNA Parsimony Algorithm, version 3.5c. Each tree required 23 steps for construction (table 9 and 10)

3.3.3 Sequence divergence

A total of 22 variable sites (3.4 %, including indels) were found, with only 3 (13.6 %) phylogenetically informative sites (Table 4). A phylogenetically informative site must have at least two character states shared by two or more taxa. Out of the 22 variable sites, 12 occur among the four populations of *S. rupicola*, with 8 of these being deletions. Thus, only 4 base substitutions were observed in *Saintpaulia rupicola* compared to 6 in *S. brevipilosa* only. This shows the conservative nature of the sequence in this region among these populations.

Of the mutations observed in the present study, 8 were transitions, 5 transversions and 9 deletions. More transitions reveal that the divergence of taxa in this study occured recently (Hsiao *et al*, 1993). Most of the deletions were observed in the Kacharoroni population. Most of the variation was found in ITS1 with 13 variable sites (2.04 %). The 5.8 rDNA had 6 (0.94 %) and ITS2 had 2 (0.31 %). These figures slightly vary from those quoted in section 1.1.1.3.1(having been based on analysis of interspecific variation). The sequences had an average C+Gcontent of 52.6 %, ranging from 52 to 52.9 %. The analysis of the length and CG content of ITS1, ITS2 and 5.8 rDNA is summarized in Table 8. The ITS region is thus conserved within the four populations of *S. rupicola*. Low divergence level is also consistent with recent origin of this group.

Table 8: Base composition of ITS1, 5.8S, and ITS2 regions in *Saintpaulia rupicola*, *Saintpaulia* sp. nov. and *S. brevipilosa*

Sample	ITS1 length	ITS 1 C+G (%)	5.8S length	5.8S C+G (%)	ITS2 length	ITS2 C+G (%)	Total ITS length	Total ITS C+G (%)
Kacharoroni	237	116(48.9)	179	98(54.7)	208	113(54.3)	624	327(54.2)
Mwaraka ya	241	120(49.8)	180	98(54.4)	208	114(54.8)	629	332(52.8)
Mwache Cha Simba	241	120(49.8)	180	99(55)	208	113(54.3)	629	332(52.8)
S.brevipilosa	241	121(50.2)	181	99(54.7)	208	113(54.3)	630	333(52.9)
	240	115(47.9)	181	98(54.1)	208	114(54.8)	629	327(52)

CHAPTER 4

GENERAL DISCUSSION

4.0 Extraction and quantification of genomic DNA

Despite the fact that the CTAB procedure was modified to improve DNA yield, the yield of genomic DNA was still too low (approximately 50 to 150 μg from 3 g leaf material). The size of the pellet obtained was quite large. However, the amount of DNA in this pellet was quite small. This shows that the DNA extracted by this procedure from the African Violets precipitates with a lot of impurities. This interferes with resuspension of DNA in TE, and in most cases an excess of TE was used to ensure that all the DNA was fully resuspended. The more succulent the leaves, the less the DNA yield, and the higher the ratio of impurity to DNA content. In most cases, this DNA had to be purified further to respond to restriction enzyme digestion and PCR amplification. The OD values obtained indicated that the samples were contaminated with impurities.

4.1 Sequencing PCR clones versus direct sequencing of the PCR product

Suh *et al* (1993) reported two types of ITS sequences in clones of one family (Winteraceae). This occurs if concerted evolution fails to homogenize ITS paralogues (e.g. those at different chromosomal loci) through a succession of speciation events (Baldwin *et al*, 1995). This presents the possibility of sampling sequences with different evolutionary histories, presenting a possible danger to phylogenetic analysis. A direct sequencing approach can sometimes aid in detection of intragenomic repeat-type variants, which can be tentatively diagnosed by two or

more nucleotide states at a site, implying superimposition of two or more sequence patterns. The small error rate of *Taq* polymerase dictates that the predominant signal in direct sequences of PCR products pooled from two or more reactions is likely to represent DNAs that were accurately replicated (Saiki *et al*, 1988; Gyllensten & Erlich, 1988; Bruns *et al*, 1990). In contrast, sequencing only one PCR clone risks sampling of misreplicated products and undersampling of any repeat-type variants.

However, it is still arguable whether sequencing of individual PCR clones or of pooled PCR products is the better approach to phylogenetic analysis of the ITS region (Baldwin *et al*, 1995). Sequencing a large number of PCR clones of the same individual, though ideal, is prohibitively expensive and labour intensive. In this study, only two clones per individual were sequenced.

4.2 Sequence analysis

Indels can be potentially phylogenetically informative (Baldwin *et al*, 1995). Thus, it may be desirable to take advantage of information provided by ambiguously placed indels rather than removing them entirely from phylogenetic analyses. Although indels can introduce complications to phylogenetic analysis, the effect of different indel treatments (including exclusion) on resultant tree topologies, where examined in angiosperms have been minimal or absent (Wojciechowski *et al*, 1993; Baldwin and Robichaux, 1995; Baum *et al*, 1994). In this study, the effect of using indels for phylogenetic analysis is to separate the Kacharoroni and Mwache populations, with Kacharoroni being revealed as the most evolutionary distant

(Figure 4A). This is in contrast with the evolutionary relationships observed when indels are not used in the analysis using UPGMA (Figure 4B), Maximum Likelihood (Figure 5) and DNA parsimony (Figure 6). In these analyses, gaps introduced in indel regions have been treated as missing data (Swofford and Olsen, 1990, Bruns *et al*, 1992, Wojciechowski *et al*, 1993). Inclusion of indels in this analysis therefore is not appropriate for this study.

Four basic types of mutations occur at the DNA level. These are: substitution of a nucleotide with another nucleotide, deletion or insertion of a nucleotide with another nucleotide(s), and inversion of nucleotides. Nucleotide substitutions can be of transitional or transversional type. Transition is the substitution of a purine (adenine or guanine) for another purine, or a pyrimidine (thymine or cytosine) for another pyrimidine. Substitutions of a pyrimidine for a purine or vise-versa are transversions. In practice, transitions occur more commonly than transversions (Fitch, 1967; Vogel, 1972; Gojobori *et al* , 1982; Li *et al*, 1984) as was observed in this study.

4.3 Applicability of PCR sequencing

As observed, PCR-sequencing of the ITS region revealed some genetic differences among the samples analysed. However, the amount of genetic variation revealed by this method was too low, making this technique more appropriate for assessment of interspecific variation in *Saintpaulia*. This kind of analysis would be stronger if it was carried out with a larger selection of outgroup species from the genus *Saintpaulia*, or for comparison of intraspecific variation with that of interspecific variation. Thus, the ITS sequences did not provide enough phylogenetic information

to resolve stable relationships among this group of very closely related species.

Extending the analysis of DNA variation to other rapidly evolving regions such as the ***trnL*** intron sequences of the chloroplast genome may provide useful data that may broaden our understanding on the genetic variability that exists in these populations. This intron has been used successfully to reveal genetic variation within and between *Acacia tortilis, A. senegal and A. nilotica* (Obunga, 1995).

4.4 Genetic variation and delimitation of species boundaries

The genetic variation revealed by this analysis closely conforms with the classification based on morphological characters. The Kacharoroni and Mwache populations resemble each other as do the Cha Simba and Mwarakaya populations. Robertson and Luke (1993) lumped the Mwache and Kacharoroni populations together under the same taxa. In the present analysis, all the individuals have been grouped as separate genetic entities. The Mwarakaya and Cha Simba populations, separated geographically about 1 km, look morphologically more alike than the Mwache and Kacharoroni populations. The results of cladistic analysis discussed in chapter 3 conform with this observation. It is however not possible from this analysis to conclusively determine whether or not the Mwache and Kacharoroni populations differ from the Cha Simba and Mwarakaya populations enough to be classified as a different species. This may be possible if more outgroup species from the genus *Saintpaulia* were to be included in the analysis. However, ITS sequence analysis shows the four populations to be much more genetically closely related than to *S. brevipilosa.*

4.5 Evolutionary implications

From the observations made in section 3.4.2.it may be possible that the Mwarakaya population is the original *S. rupicola*. The Cha Simba population may have been dispersed from Mwarakaya. Alternatively, the two populations may have been one continuous population which was eventually separated into two by forest clearing and farming activities. It could be possible that the Mwache and Kacharoroni populations are evolutionary ancestors of Cha Simba.

4.6 Recommendations

A more thorough analysis is required before concrete recommendations regarding conservation strategies of this species can be made. A combination of more than one technique for assessment of genetic variation would provide a more conclusive relationship among these plants. However, despite the poor resolution and weak support of relationships, some recommendations can be made from the fact that genetic variation does exist among these populations.

The genetic variation observed among the populations of *S. rupicola* have considerable implications for their conservation. The demonstration of distinct genetic entities among these populations suggests that collection and preservation of germplasm (e.g. seeds, leaves for propagation) samples should be based on representation of all the populations of the species to ensure coverage of as much of the genetic diversity as possible. None of the populations should be treated as a representative of the others during conservation. The sustainability of genetic variation or the genetic improvement of any plant species depends upon the

existence of genetic variability between parental materials. The use of germplasm with a narrow genetic base can lead to limited genetic base or homozygosity.

4.7 Concluding remarks

The degree of ITS sequence divergence in *Saintpaulia* is more useful for inferring phylogenetic relationships at the species level than it is at the population level. This is consistent with findings of Hsiao *et al* (1995). This study provides a base for the assessment of genetic variation among other species and populations of the African Violets using the same technique or in combination with others. Although only one individual (*S. brevipilosa*) was used as an outgroup, the fact that it revealed more genetic diversity than that observed among the Kenyan populations demonstrates that valuable information on classification can be derived from a similar analysis of *Saintpaulia* species.

REFERENCES

Appels, R., and J. Dvorak. 1982. Relative rates of divergence of spacer and gene sequences within the rDNA region of species in the Triticeae: Implications for the maintenance of homogeneity of a repeated gene family. *Theor. Appl. Genet.*, **63**: 361-365

Arisumi, T. 1964. Interspecific hybridization in African Violets. *J. Heredity* **55:** 181-183

Arnheim, N. 1983. Concerted evolution of multigene families. *In* "Evolution of Genes and Proteins", M. Nei and R. Koehn, Eds, pp. 38-61, Sinaeur. Sunderland

Arnheim, N., M. Krystal, R. Schmickel, G. Wilson, O. Ryder and E. Zimmer. 1980. Molecular evidence for molecular exchanges among ribosomal genes on non-homologous chromosomes in man and apes. *Proc. Natl. Acad. Sci. USA*, **77**: 7323-7327

Avise, J. C. 1989. Gene trees and organismal histories: a phylogenetic approach to population biology. *Evolution,* **43**: 1192-1208

Avise, J. C., J. Arnold, R. M. Ball, E. Berminham, T. Lamb, J. E. Neigel, C. A. Reeb and N. C. Saunders. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Ann. Rev. Ecol. Syst.*, **18**: 489-522

Baldwin, B. G. 1992. Phylogenetic utility of the Internal Transcribed Spacers of Nuclear ribosomal DNA in Plants: An example from the Compositae. *Molecular phylogenetics and evolution*. **1**: 3-16

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

Baldwin, B. G. and R. H. Robichaux. 1995. Historical biogeography and ecology of the Hawaiian silversword alliance: New molecular phylogenetic perspectives. *In* **W. L. Wagner & V. A. Funk**. Hawaiian biogeography: evolution on a Hot Spot Archipelago. Smithsonian Information Press, Washington, D. C.

Baldwin, B. G., M. J. Sanderson, J. M. Porter, M. F. Wojciechowski, C. S. Campbell and M. J. Donoghue. 1995. The ITS region of the nuclear ribosomal DNA: A valuable source of evidence on angiosperm phylogeny. *Ann. Missouri Bot. Gard,* **82:** 247-277

Baum, D. A., K. J. Systma and P. C. Hoch. 1994. A phylogenetic analysis of *Epilobium* (Onagraceae) based on nuclear ribosomal DNA sequences. *Syst. Bot,* **19:** 363-388

Birnboim, H. C. and J. Doly. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic acids Res.*, 7: 1513

Bruns, T. D., R. Fogel and J. W. Taylor. 1990. Amplification and sequencing of DNA from fungal herbarium specimens. *Mycologia*. **82:** 175-184

Burtt, B. L. 1947. Species of *Saintpaulia, Gard. Chron 3* **ser cxxii.** 22-23

Burtt, B. L. 1956. A confusion in African Violets. *Baileya,* **4:** 163-164

Burtt, B. L. 1958. Studies in the Gesneriaceae of the Old World XV: the Genus *Saintpaulia*. *Notes Bot. Gard. Edinb.* **22:** 547-568

Burtt, B. L. 1960. *Saintpaulia teitensis*. *Notes Roy. Bot. Gard. Edinb*, **13**: 559

Burtt, B. L. 1964. *Saintpaulia rupicola*. *Notes Roy. Bot. Gard. Edinb*, **25**: 193

Burtt, B. L. 1976. *Saintpaulia orbicularis var. purpurea. Curt. Bot. Mag.* **181.** t. 722

Chalmers, K. J., Waugh, R., Sprent, J. I., Simons, A. J., and Powell, W. 1992. Detection of genetic variation between and within populations of *Gliricidia sepium* and *G. maculata* using RAPD markers *Heredity:*, **69:** 465-472

Crisci, J. V., E. A. Zimmer, P. C. Hoch, G. B. Johnson, C. Mudd and N. S. Pan, 1990. Phylogenetic implications of ribosomal DNA restriction site variation in the family Onagraceae. *Ann. Mo. Bot. Gard*, **77**: 523-538

Downie, R. S. and J. D. Palmer. 1992. Use of chloroplast DNA rearrangements in reconstructing plant phylogeny. *In* Soltis, P. S., D. E. Soltis and J. J. Doyle (Eds). Molecular systematics of plants. Chapman and Hall, London. pp 14-35

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

Doyle, J. J., J. L. Doyle and A. H. D. Brown. 1990. Chloroplast DNA polymorphism and phylogeny in the B genome of Glycine subgenus Glycine (Leguminosae). *Amer. J. Bot.*, **77:** 772-782

Ehrlich, H. 1958. Cytological studies in *Saintpaulia*. *Am J. Bot.* **45:** 177-182

Fitch, W. M. 1967. evidence suggesting a non-random character to nucleotide replacements in naturally occuring mutations. *J. Mol. Biol.*, **26:** 499-507

Fussell, C. 1958. Chromosome numbers in the Gesneriaceae. *Baileya* **6:** 117-125

Gielly, L. and P. Taberlet. 1994. The use of chloroplast DNA to resolve plant phylogenies: Non-coding versus *rbcL* sequences. *Mol. Biol. Evol.*, **11**(5): 769-777

Gojobori, T., Li, W-H and Graur, D. 1982. Patterns of nucleotide substitutions in pseudogenes and functional genes. *J. Mol. Evol,* **18:** 360-369

Gottlieb, L. D. 1977. Electrophoresis evidence and plant systematics. *Ann. Missouri Bot. Gard*, **64**: 161-180

Gyllesten, U. B. and H. A. Erlich. 1988. Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA DQ locus. *Proc. Natl. Acad. Sci. USA*, **85**: 7652-7658

Hamby, R. K. and E. A. Zimmer. 1992. Ribosomal RNA as a phylogenetic tool in plant systematics. *In* Soltis, P. S., D. E. Soltis and J. J. Doyle (*Eds)*. Molecular systematics of plants pp 50-91. Chapman and Hall, New York. London

Harrison, R. G. 1989. Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. *Trends Ecol. Evol*, **4**: 6-11

Harvey, P. M. & A. Knap. 1979. A technique for *in-vitro* propagation of African Violets using petioles. *Can. J. Plant. Sci.* **59:** 263-266

Hillis, D. M. and M. T. Dickson. 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. *Quarterly Review of Biology,* **66:** 411-453

Hillis, D. M., C. Moritz, C. A. Porter and R. J. Baker. 1991. Evidence for biased gene conversion in concerted evolution of ribosomal DNA. *Science*, **251**: 308-310

Hsiao, C., N. J. Chatterton, K. H. Asay and K. B. Jensen. 1993. Phylogenetic relationships of 10 grass species: an assessment of the phylogenetic utility of the internal transcribed spacer region in nuclear ribosomal DNA in monocots. *Genome*. **37:** 112-120

Hsiao, C., N. J. Chatterton, K. H. Asay and K. B. Jensen. 1994. Phylogenetic relationships of the monogenomic species of the wheat tribe, Triticeae (Poaceae), inferred from nuclear rDNA (internal transcribed spacer) sequences. *Genome*, **38**: 211-223

Hsiao, C., N. J. Chatterson, K. H. Asay and K. B. Jensen. 1995. Phylogenetic relationships of the monogenomic species of the wheat tribe, Triticeae (Poaceae), inferred form nuclear rDNA (internal transcribed spacer) sequences. *Genome* , **38:** 211-223

Intelligenetics, Inc. 1991. Geneworks. Reference manual. 700 East El Camino Real. Mountain View, CA 94040

Isc-Horowicz, D. and J. F. Burke. 1981. Rapid and efficient cosmid cloning. *Nucl. Acid Res.*, **9**: 2989

Ivaniva, L. 1966. Application of the carpological method to the taxonomy of the Gesneriaceae. *Notes Bot. Gard. Edin.*, **26:** 383-403

Johansen, D. R. 1978. *Saintpaulia* in their natural environment with notes on their present status in Kenya and Tanzania. *Biol. Conserv.* **14:** 45-62

Kim, K. J., R. K. Jansen and B. L. Turner. 1992. Evolutionary implications of chloroplast DNA variation in dwarf dandelions (*Krigia*; Asteraceae). *Amer. J. Bot.*, **79:** 708-715

Kim, K. J. and T. J. Mabry. 1991. Phylogenetic and evolutionary implications of nuclear ribosomal DNA variation in dwarf dandelions (*Krigia,* Lactuceae, Asteraceae). *Plant Syst. Evol.*, **177**: 53-69

Kimura, M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.*, **16:** 111-120

Kocher, T. D., W. K. Thomas, A. Meyer, S. V. Edwards, S. Paabo, F. X. Villblanca and A. C. Wilson. 1989. Dynamics of mitochondrial evolution in animals: amplification and sequencing with conserved primers. *Proc. Natl. Acad. Sci.*, USA, **86:** 6196-6200

Learn, G. H. and B. A. Schaal. 1986. Population sub-division for ribosomal DNA repeat variants in *Clematis fremontii*. *Evolution*, **41**(2): 433-438

Lee, R. 1962. Chromosome numbers in the Gesneriaceae. *Baileya* **10:** 33-45

Li, W-H, Wu, C-I, and Luo, C-C. 1984. Non-randomness of point mutation as reflected in nucleotide substitution in pseudogenes and its evolutionary implications. *J. Mol. Evol.*, **21:** 58-71

Milne, C. 1975. Chromosome numbers in the Gesneriaceae, 5,*Notes Bot. Gard. Edinb.*, **36:** 523-525

Nei, M. 1975. Molecular population genetics and evolution. Amsterdam, North Holland

Nei, M. 1987. Molecular Evolutionary Genetics. Columbia University Press. New York. N. Y. pp293-98

Obunga, E. O. 1995. A study of genetic systems of four African species of *Acacia*. Ph.D thesis submitted to the university of Sussex, U. K.

Olmstead , R. G. and J. D. Palmer. 1994. Chloroplast DNA systematics: a review of methods and data analysis. *Amer. J. Bot.*, **81**(9): 1205-1224

Palmer, J. D. 1987. Chloroplast DNA evolution and Biosystematic uses of chloroplast DNA variation. *Amer. Nat.* , **130**: S6-S29

Palmer, J. D. 1988. Intraspecific variation and multicircularity in *Brassica* mitochondrial DNAs. *Genetics*, **118**: 341-351

Palmer, J. D. 1992. Mitochondrial DNA in plant systematics: Application and limitations. *In*: Soltis, P. S., D. E. Soltis and J. J. Doyle (Eds), Molecular systematics of plants. Chapman and Hall, London. pp 36-49

Ratter, J. 1975. A survey of the chromosome numbers in the Gesneriaceae of the Old World. *Notes Bot. Gard. Edinb.* **33:** 527-543

Rieserberg, L. H., D. E. Soltis and J. D. Palmer. 1988. A reexamination of introgression between *Helianthus annuus* and *H. bolnderi* (Compositae). *Evolution*, **42**: 227-238

Rieserberg, L. H., S. M. Beckstrom-Sternberg, A. Liston and D. M. Arias. 1991. Phylogeny and systematic inferences from chloroplast DNA and isozyme variation in *Helianthus* sect. *Helianthus* (Asterceae). *Systematic Botany*, **16:** 50-76

Robertson, S. A. and W. R. Q. Luke. 1993. KENYA COASTAL FORESTS. The report of the NMK/WWF Coast Forest Survey, WWF Project 3256: Kenya, Coast Forest Status, Conservation and Management

Rogers S. O., and A. J. Bendich. 1987. Ribosomal RNA genes in plants: Variability in copy number and in the intergenic spacer. *Plant Mol. Biol.*, **9**: 509-520

Russel, J. R., F. Hosein, E. Johnson, R. Waugh and W. Powell. 1993. genetic differentiation of cocoa ((*Theobroma cacao* L.) populations revealed by RAPD

analysis. *Mol. Ecol.*, **2:** 89-97

Saghai-Maroof, M. A., K. M. Soliman, R. A. Jorgensen and R. W. Allard. 1984. Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl. Acad. Sci. USA,* **81:** 8014-8018

Saiki, R. K. 1989. The design and optimization of the PCR *In* PCR Technology: Principles and applications for DNA amplification. pp 7-16. Erlich, H. A. (Eds) Stockton Press, USA

Saiki, R. K., S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Erlich and N. Arnheim. 1985. Enzymatic amplification of Ë-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anaemia. *Science*, **230**: 1350

Saiki, R. K., D. H. Gelfand. S. Stofel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis and H. A Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, **239**: 487-491

Sambrook, J., E. F. Fritsch and T. Maniatis. 1987. Molecular cloning, a laboratory manual. pp. 150-161. Cold Spring Harbour laboratory, N. Y., USA

Sambrook, J., E. F. Fritsch and T. Maniatis. 1989. Molecular cloning, a laboratory manual. Volume 1 pp. 1.25-1.28. Cold Spring Harbour laboratory, N. Y., USA

Sanderson, M. J., and J. J. Doyle. 1992. Reconstruction of organismal and gene phylogenies from data on multigene families: concerted evolution, homoplasy and confidence. *Syst. Biol.*, **41:** 4-17

Sanger, F., A. R. Nicklen and R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci., USA*. **74**: 5463-5467

Saghai-Maroof, M. A., K. M. Soliman, R. A. Jorgensen and R. W. Allard. 1984. Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl. Acad. Sci. USA*. **81**: 8014-8018

Schilling, E. E. and R. K. Jansen. 1989. Restriction fragment analysis of chloroplast DNA and the systematics of *Viguiera* and related genera (Asteraceae: Helianthae). *Amer. J. Bot.*, **76:** 1769-1778

Sheil, D. 1993. What future for the African Violets? *In* Swara vol. 16 No. 5. East African Wildlife Society, EAWS

Sera, T. & K. Karasawa. 1984. Karyomorphological studies on 22 taxa of *Saintpaulia* (Gesneriaceae). *Bulletin of the Hiroshima Bot. Gard.*, **7:** 1-30

Sheil, D. 1993. What future for African Violets? *Swara magazine* **16(5),** 18-21

Smith, R. L., and K. J. Systsma. 1990. Evolution of *Populus nigra* (sect. *Aïgeiros*): Introgressive hybridization and the chloroplast contribution of *Populus alba* (Sect. *Populus*). *Amer. J. Bot.* **77**: 1176-1187

Soltis, D. E., M. S. Mayer, P. S. Soltis and M. Edgerton. 1991a. Chloroplast DNA variation in *Tellima grandiflora* (Saxifgagaceae). *Amer. J. Bot.*, **78**: 1379-1390

Soltis, D. E., P. S. Soltis, T. G. Collier and M. L. Edgerton. 1991b. Chloroplast DNA variation within and among genera of the *Heuchera* group [Saxifragaceae]: evidence for chloroplast transfer and paraphyly. *Amer. J. Bot.*, **78**(8): 1091-1112

Soltis, D. E., P. S. Soltis and B. G. Milligan. 1992. Intraspecific chloroplast DNA variation: systematic and phylogenetic implications. *In* Soltis P. S., D. E. Soltis J. J. Doyle (Eds). Molecular systematics of plants. Chapman and Hall, New York. pp 117-150

Sperling, F. A. 1993. Mitochondrial DNA variation and Haldane's rule in the *Papilio glaucus* and *P. troilus* species groups. *Heredity,* **71**: 227-233

Suh, Y., L. B. Thien, H. E. Reeve and E. A. Zimmer. 1993. Molecular evolution and phylogenetic implications internal transcribed spacer sequences of ribosomal DNA in Winteraceae. *Amer. J. Bot.*, **80**(9): 1042-1055

Swensen, S. M., G. J. Allan, M. Howe, W. J. Elisens, S. A. Junak and L. H. Rieseberg. 1995. Genetic analysis of the endangered island endemic *Malacothamnus fasciculatus* (Nutt) Greene var. *nesioticus* (Rob.) Kearn. (Malvaceae). *Conservation Biology*, **9**: 404-415

Swofford, D. L. and G. L. Olsen. 1990. Phylogeny reconstruction. Pp 411-501 *In* D. M. Hillis and C. Moritz (Eds), Molecular Systematics. Sinauer, Sunderland, Massachusetts

Sytsma, K. J. and B. A. Schaal. 1985. Phylogenetics of the *Lisianthus skinneri* (Gentianaceae) species complex in Panama utilizing DNA restriction fragment analysis. *Evolution*, **39**: 594-608

Taberlet, P., L. Gielly, G. Pautou and J. Bouvet. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Pl. Molec. Biol.*, **27**(1): 69-78

Takaiwa, F., K. Oono and M. Sugiura. 1985. Nucleotide sequence of the 17S-25S spacer region from rice rDNA. *Pl. Molec. Biol.* **4:** 355-364

Vogel, F. 1972. Non-randomness of base replacement in point mutation. *J. Mol. Evol.,* **1:** 344-367

Wallace, R. S. and R. K. Jansen. 1990. Systematic implications of chloroplast DNA variation in the genus *Microseris* (Asteraceae, Lactuceae). *Syst. Bot.,* **15**(4): 606-616

Welsh, J. and M. McCleland. 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucl. Acids Res.*, **18**: 7213-7218

White, T. J., T. Bruns, S. Lee and J. Taylor. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *In* M. Innis, D. Gelfand, J. Sninsky, and T. White ''PCR protocols: A guide to methods and applications'', (Eds.), pp. 315-322, Academic Press, San Diego

Williams, J. G. K., A. R. Kubelik, K. J. Livak, J. A. Rafalski and S. V. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucl. Acid Res.*, **18**: 6531-6535

Wojciechowski, M. F., M. J. Sanderson, B. G. Baldwin and M. J. Donoghue. 1993. Monophyly of aneuploid astragalus (Fabaceae): Evidence from ribosomal DNA Internal Transcribed Spacer sequences. *Amer. J. Bot.* **80**(6): 711-722

Wolfe, K. H., W-H Li and P. M. Sharp. 1987. Rates of nucleotide substitution vary among plant mitochondrial, chloroplast and nuclear DNAs. *Proc. Natl. Acad. Sci. USA*, **84**: 9054-9058

Zimmer, E. A., S. L. Martin, S. M. Beverley, Y. W. Kan and A. C. Wilson. 1990. Rapid duplication and loss of genetic coding for the alpha chains of hemoglobin. *Proc. Natl. Acad. Sci. USA*, **77**: 2158-2162

Zurawski, G., M. T. legg and A. D. H. Brown. 1984. The nature of nucleotide sequence divergence between barley and maize chloroplast DNA. *Genetics,* **106**: 735-749

APPENDICES

Appendix 1: Bacterial media and plates

LB/amp/IPTG/X-gal plates

10 g Bacto-tryptone

5 g Bacto-yeast extract

5 g NaCl

15 g agar

This mixture was made up to 1 liter, autoclaved for 20 min, allowed to cool to 50°C. Ampicillin was then added to a final concentration 100 μl/ml, IPTG to 0.5 mM and X-gal to 80 μg/ml. The agar in the plates was allowed to harden and the plates stored at 4°C for up to 1 month.

SOC MEDIUM

2 % (w/v) Bacto Tryptone

0.5 % (w/v) Bacto Yeast extract

10 mM NaCl

2.5 mM KCl

10 mM MgCl

10 mM $MgSO_4$

20 mM glucose

Appendix 2: Buffers and solutions

3x CTAB isolation buffer

100 mM Tris-HCL pH 8.0

1.4 M NaCl

20 mM EDTA

3 % CTAB

DNA loading buffer

7 % sucrose

0.25 % bromophenol blue

0.25 % xylene blue

15 % Ficoll

Ethanol wash solution

50 % Ethanol

0.1 M NaCl

10 mM Tris-HCl pH 7.5

This solution was sterilized by autoclaving for 20 min. It was stored at 4°C.

NaI Solution

NaI	90.8 g
Na_2SO_4	1.5 g

The volume was made up to 100 ml and filtered through whatman. This solution was then saturated with 0.5 g Na_2SO_3. It was stored at 4°C.

Solution 1 (RF extraction)

50 mM glucose

10 mM EDTA pH 8.0

25 mM Tris-HCl pH 8.0

4 mg/ml lysozyme

Solution II (Lysis solution for RF Extraction)

0.2 N NaOH

1 % SDS

Both solution 1 and II for RF extraction were prepared fresh for every extraction

Solution III (RF extraction)

KOAc pH 4.8

This solution is stored at 4°c

Formamide stop solution

95 % Formamide

20 mM EDTA

0.05 % Bromophenol Blue

0.05 % Xylene Cyanol FF

10 X *Taq* polymerase buffer

50 mM KCl

10 mM Tris-HCl pH 9.0

1.5 mM $MgCl_2$

0.001 % (w/v) gelatin

1 % Triton X-100

50X TAE buffer

Tris base

Glacial acetic acid

EDTA, pH 8.0

TE Solution

10 mM Tris pH 7.5

1 mM EDTA

10X TBE buffer

900 mM Tris

880 mM boric acid

25 mM EDTA

TE RNAse

10 mM Tris pH 7.5

1 mM EDTA

100 μg/ml RNAse A

RNAse was boiled at 90°C for 10 min to destroy any DNAse present

Wash buffer (for genomic DNA)

75 % Ethanol

10 mM Ammonium Acetate

X-gal

20 µg/ml X-gal in dimethylformamide

Table 9: Steps used in constructing evolutinary tree (figure 6A).

steps in each site:

	0	1	2	3	4	5	6	7	8	9
0!		0	0	0	0	0	0	0	0	0
10!	0	0	0	0	0	0	0	0	0	0
20!	0	0	0	0	0	0	0	0	0	0
30!	0	0	0	0	0	0	0	0	0	0
40!	0	0	0	0	0	1	0	0	0	0
50!	0	0	0	0	0	0	0	0	0	0
60!	0	0	0	1	0	0	0	0	0	0
70!	0	0	0	0	0	0	0	0	0	0
80!	0	0	0	0	0	0	1	0	0	0
90!	0	0	0	0	0	0	0	0	0	0
100!	0	1	0	0	0	0	0	0	0	0
110!	0	0	0	0	1	0	0	0	0	0
120!	0	0	0	0	0	0	0	0	0	0
130!	1	1	0	0	0	0	0	0	0	0
140!	0	0	0	0	0	0	0	0	0	0
150!	0	0	0	0	0	0	0	0	0	0
160!	0	0	0	0	0	0	0	0	0	0
170!	0	0	0	0	0	0	0	0	0	0
180!	0	0	0	0	0	0	0	0	0	0
190!	0	1	0	1	0	0	0	0	0	1
200!	0	0	0	0	0	0	0	0	0	0
210!	0	0	0	0	0	0	0	0	0	0
220!	0	0	0	1	0	0	1	0	0	0
230!	0	1	0	0	0	0	0	0	0	0
240!	0	0	0	0	1	0	0	0	0	0
250!	0	0	0	0	0	0	0	0	0	0
260!	0	0	0	0	0	0	0	0	0	0
270!	0	0	0	0	0	0	0	0	0	0
280!	0	0	0	0	0	0	0	0	0	0
290!	0	0	0	0	0	0	0	0	0	0
300!	0	0	0	0	0	0	0	0	0	0
310!	0	0	0	0	0	0	0	0	0	0
320!	0	0	0	0	0	0	0	0	0	0
330!	0	0	0	0	0	0	0	0	0	0
340!	0	0	0	0	0	0	0	0	0	0
350!	0	0	0	0	0	0	0	0	0	0
360!	1	0	0	0	0	0	0	0	0	0
370!	0	0	0	0	0	0	0	0	1	0
380!	1	0	0	0	2	0	0	0	0	0
390!	0	0	0	0	0	0	0	0	0	0
400!	0	0	0	0	0	0	0	0	0	0
410!	0	0	0	0	1	0	0	0	0	0
420!	0	0	0	0	0	0	0	0	1	0
430!	0	0	0	0	0	0	0	0	0	0
440!	0	0	0	0	0	0	0	0	0	0
450!	0	0	0	0	0	0	0	0	0	0
460!	0	0	0	1	0	0	0	0	0	0
470!	0	0	0	0	0	0	0	0	0	0
480!	0	0	0	0	0	0	0	0	0	0
490!	0	0	0	0	0	0	0	0	0	0
500!	0	0	0	0	0	0	0	0	0	0
510!	0	0	0	0	0	0	0	0	0	0
520!	0	0	0	0	0	0	0	0	0	0
530!	0	0	0	0	0	0	0	0	0	0
540!	0	0	0	0	0	0	0	0	0	0
550!	0	0	0	0	0	0	0	0	0	0
560!	0	0	0	0	0	0	0	0	0	0
570!	0	0	0	0	0	0	0	0	0	0
580!	0	0	0	0	0	0	0	0	0	0
590!	0	0	0	0	0	0	0	0	0	0
600!	0	0	0	0	0	0	0	0	0	0
610!	0	0	0	0	0	0	0	0	0	0
620!	0	0	1	0	0	0	0	0	0	0
630!	0	0								

Table 10: Steps used in constructing evolutionary tree (figure 6B).

steps in each site:

	0	1	2	3	4	5	6	7	8	9
0!		0	0	0	0	0	0	0	0	0
10!	0	0	0	0	0	0	0	0	0	0
20!	0	0	0	0	0	0	0	0	0	0
30!	0	0	0	0	0	0	0	0	0	0
40!	0	0	0	0	0	1	0	0	0	0
50!	0	0	0	0	0	0	0	0	0	0
60!	0	0	0	1	0	0	0	0	0	0
70!	0	0	0	0	0	0	0	0	0	0
80!	0	0	0	0	0	0	1	0	0	0
90!	0	0	0	0	0	0	0	0	0	0
100!	0	1	0	0	0	0	0	0	0	0
110!	0	0	0	0	1	0	0	0	0	0
120!	0	0	0	0	0	0	0	0	0	0
130!	1	1	0	0	0	0	0	0	0	0
140!	0	0	0	0	0	0	0	0	0	0
150!	0	0	0	0	0	0	0	0	0	0
160!	0	0	0	0	0	0	0	0	0	0
170!	0	0	0	0	0	0	0	0	0	0
180!	0	0	0	0	0	0	0	0	0	0
190!	0	1	0	1	0	0	0	0	0	1
200!	0	0	0	0	0	0	0	0	0	0
210!	0	0	0	0	0	0	0	0	0	0
220!	0	0	0	1	0	0	1	0	0	0
230!	0	1	0	0	0	0	0	0	0	0
240!	0	0	0	0	1	0	0	0	0	0
250!	0	0	0	0	0	0	0	0	0	0
260!	0	0	0	0	0	0	0	0	0	0
270!	0	0	0	0	0	0	0	0	0	0
280!	0	0	0	0	0	0	0	0	0	0
290!	0	0	0	0	0	0	0	0	0	0
300!	0	0	0	0	0	0	0	0	0	0
310!	0	0	0	0	0	0	0	0	0	0
320!	0	0	0	0	0	0	0	0	0	0
330!	0	0	0	0	0	0	0	0	0	0
340!	0	0	0	0	0	0	0	0	0	0
350!	0	0	0	0	0	0	0	0	0	0
360!	2	0	0	0	0	0	0	0	0	0
370!	0	0	0	0	0	0	0	0	1	0
380!	1	0	0	0	1	0	0	0	0	0
390!	0	0	0	0	0	0	0	0	0	0
400!	0	0	0	0	0	0	0	0	0	0
410!	0	0	0	0	1	0	0	0	0	0
420!	0	0	0	0	0	0	0	0	1	0
430!	0	0	0	0	0	0	0	0	0	0
440!	0	0	0	0	0	0	0	0	0	0
450!	0	0	0	0	0	0	0	0	0	0
460!	0	0	0	1	0	0	0	0	0	0
470!	0	0	0	0	0	0	0	0	0	0
480!	0	0	0	0	0	0	0	0	0	0
490!	0	0	0	0	0	0	0	0	0	0
500!	0	0	0	0	0	0	0	0	0	0
510!	0	0	0	0	0	0	0	0	0	0
520!	0	0	0	0	0	0	0	0	0	0
530!	0	0	0	0	0	0	0	0	0	0
540!	0	0	0	0	0	0	0	0	0	0
550!	0	0	0	0	0	0	0	0	0	0
560!	0	0	0	0	0	0	0	0	0	0
570!	0	0	0	0	0	0	0	0	0	0
580!	0	0	0	0	0	0	0	0	0	0
590!	0	0	0	0	0	0	0	0	0	0
600!	0	0	0	0	0	0	0	0	0	0
610!	0	0	0	0	0	0	0	0	0	0
620!	0	0	1	0	0	0	0	0	0	0
630!	0	0								