Genetic marker resources for application in *Cyclopia* species

by

Marioné Niemandt

Thesis presented in fulfilment of the requirements for the degree of Master of Science in the Faculty of AgriSciences at Stellenbosch University

Supervisors: Prof Rouvay Roodt-Wilding and Dr Cecilia Bester

Co-supervisor: Mr Kenneth Tobutt

December 2016

Stellenbosch University https://scholar.sun.ac.za

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

December 2016

Copyright © 2016 Stellenbosch University

All rights reserved

Stellenbosch University https://scholar.sun.ac.za

Abstract

Cyclopia species are endemic to the Fynbos Biome of South Africa and have been utilised for many years as a health drink known as honeybush tea. Despite the commercial importance of *Cyclopia*, no molecular resources are available to characterise this genus. The polyploid nature furthermore limits the use of molecular markers as some species exhibit up to 14 sets of chromosomes (*Cyclopia intermedia* and *Cyclopia meyeriana*: $2n = 14x = 126$). This study optimised a DNA extraction protocol for various *Cyclopia* species in order to obtain high quality DNA as the first crucial step during molecular genetic studies. The use of young, fresh leaves as starting material for DNA extraction presents a challenge when sampling from distant locations; therefore, a CTAB/NaCl buffer was optimised to preserve the leaves for up to two weeks prior to DNA extraction under laboratory conditions. Microsatellite markers were developed in the commercially important *C. subternata* and transferred to six other *Cyclopia* species with a success rate of 81-88%. The Agricultural Research Council (ARC) maintain a field gene bank for several of the *Cyclopia* species and a set of six DNA fingerprinting markers were developed to characterise the accessions. This will facilitate the correct management of the gene bank, such as keeping track of clones for seed orchards or commercial release and the identification of duplicates in the gene bank. The genetic diversity of *C. subternata* wild populations was investigated and compared to the accessions to ensure that the gene bank accurately reflects the natural diversity. As such, the *C. subternata* accessions were representative of the wild samples, excluding the genetically distinct Haarlem population. The genetic resource tools developed in this study can be applied to detect the extent of cross-contamination of cultivated material to wild populations of *Cyclopia* as well as the characterisation of wild populations of all known species that could be included in the field gene bank. Further conservation strategies include the monitoring of wild-harvesting as well as the *in situ* conservation of genetically distinct populations.

Stellenbosch University https://scholar.sun.ac.za

Opsomming

Cyclopia spesies is endemies aan die Fynbos Bioom van Suid-Afrika en word al vir baie jare as ‘n gesondheidsdrankie, naamlik heuningbos tee, gebruik. Ten spyte van die kommersiële belangrikheid van *Cyclopia*, is daar geen molekulêre hulpbronne beskikbaar om die genus te karakteriseer nie. Die poliploïdie vlak van sommige spesies, wat tot en met 14 stelle chromosome kan insluit (*Cyclopia intermedia* en *Cyclopia meyeriana*: $2n = 14x = 126$) plaas verdere beperkings op die gebruik van molekulêre merkers. Hierdie studie het ‘n DNA ekstraksie protokol vir verskeie *Cyclopia* spesies geoptimiseer wat reeds tydens die eerste stap van molekulêre studies ‘n hoë kwaliteit DNA verseker. Die gebruik van jong, vars blare vir DNA ekstraksies is nie prakties wanneer monsterneming in afgeleë areas geskied nie, daarom is ‘n CTAB/NaCl buffer metode geoptimiseer om die blare tot en met twee weke te preserveer voordat DNA ekstraksie in ‘n laboratorium omgewing geskied. Mikrosatellietmerkers is vir die kommersieel belangrike spesie, *C. subternata*, ontwikkel en met ‘n sukseskoers van 81-88% na ses ander *Cyclopia* spesies oorgedra. Die Landbou Navorsingsraad (LNR) besit ’n veld genebank vir verskeie van die heuningbosspesies en ‘n stel van ses DNA vingerafdrukmerkers is ontwikkel om die plantaanwinste te karakteriseer. Dit sal die korrekte bestuur van die genebank verseker soos om tred te hou met klone vir saadboorde en kommersiële vrystelling asook identifisering van duplikate. Die genetiese diversiteit van wilde populasies van *C. subternata* is ondersoek en vergelyk met die verboude varieteite in die genebank om te verseker dat genoeg diversiteit vasgevang en behoue bly. Dit is gevind dat die *C. subternata* plantaanwinste verteenwoordigend is van die wilde populasies se genetiese diversiteit, behalwe vir die Haarlem populasie wat geneties van die ander twee populasies verskil. Die molekulêre hulpbronne wat tydens die studie ontwikkel is, kan gebruik word om genetiese kontaminasie tussen verboude plante en wilde populasies waar te neem, asook om wilde populasies van al die verskillende spesies te karakteriseer om later by die genebank in te sluit. Verdere bewaringstrategieë sluit in die kontrolering van veldoes, sowel as die *in situ* bewaring van unieke genetiese populasies in hulle natuurlike habitat.

Stellenbosch University https://scholar.sun.ac.za

Acknowledgements

I would like to extend my gratitude to the following institutions for financial support: Agricultural Research Council, Department of Science and Technology (DST), Monetary Treasury Economic Fund (MTEF), National Research Foundation (NRF), Stellenbosch University, Harry Crossley Foundation.

The following people need special mentioning for the part they played during my studies - Jessica Vervalle; not only were you there for technical assistance, but your friendship and advice kept me sane when things were at their worst. Johané Nienkemper-Swanepoel - you were always prepared to make the time to help me, even if you had to explain the statistics more than a few times. Ruhan Slabbert - your practical knowledge in the laboratory was of infinite value to me. Mlamuli Motsa - for always being prepared to take me out into the field for sampling, even if you had a thousand things to do with your own studies. Marlise Joubert - you are one of my mothers at the ARC and your vast knowledge on honeybush helped me with various questions. The ARC Honeybush Research Group - each of you contributed to my experience in your own way and everyday was a learning experience.

Thank you to my family: my mother, who is also my best friend and greatest calm-me-downer; my father, whose words of encouragement meant more to be than he realises and my brother, who often came to say hi in the lab. You are my rock and I would not have had the opportunity or courage to come this far in my studies, were it not for your support.

Thank you to my purple-couch friends: Kristin Oosthuizen, Kelly Breeds, Natasha Kitchin and Charné Rossouw. We made it from Honours to Masters, with lots of fun and tears of frustration. Thank you also to the rest of the MBB lab group, I could always count on you for advice regarding population genetic statistics and you made lab meetings fun.

Lastly, thank you to my supervisors. Dr Bester - I could walk into your office anytime and you were always prepared to lend a listening ear. Thank you for giving me the opportunity to work on honeybush. Prof Roodt-Wilding - your knowledge and encouragement through the polyploid problems kept me calm and I could always count on a perfect draft once it had been through your capable hands. Kenneth Tobutt - you always made me think outside the box and reminded me of the finer detail during my study.

Stellenbosch University https://scholar.sun.ac.za

Table of Contents

Declaration i
Abstract ii
Opsomming iii
Acknowledgements iv
Table of Contents v
List of Figures viii
List of Tables xi
List of Abbreviations xiii
Chapter 1: Literature Review 1
1.1 *Cyclopia* Species 1
1.1.1 Taxonomy and distribution 1
1.1.2 Botany 1
1.1.3 Commercial importance 2
1.2 DNA Marker Technologies 4
1.3 Applications of DNA Marker Technologies in Plant Genetic Resources 6
1.4 Polyploidy and Molecular Markers 8
1.5 Study Rationale, Aims and Objectives 10
1.5.1 Problem statement 10
1.5.2 Study aim and objectives 10
References 11
Chapter 2: *Cyclopia* Leaf Preservation and DNA Extraction 17
Abstract 17
2.1 Introduction 17
2.2 Optimisation of DNA Extraction Protocol and Testing of Different Tissue Types 19
2.2.1 Materials and methods 19
2.2.2 Results and discussion 21
2.3 Testing of Optimised DNA Extraction Protocol on Different *Cyclopia* Species 24
2.3.1 Materials and methods 24
2.3.2 Results and discussion 25
2.4 Testing of Preservation Buffers 27
2.4.1 Materials and methods 27
2.4.2 Results and discussion 28
2.5 Conclusion 30
References 30

Stellenbosch University https://scholar.sun.ac.za

Chapter 3: Microsatellite Marker Development, Cross-Species Amplification and DNA Fingerprinting 33
Abstract 33
3.1 Introduction 33
3.2 Materials and Methods 35
3.2.1 Sampling 35
3.2.2 DNA extractions 36
3.2.3 Marker development and optimisation 36
3.2.4 Cross-species amplification 38
3.2.5 DNA fingerprinting: genotyping and data analyses 39
3.3 Results and Discussion 42
3.3.1 Microsatellite marker optimisation and characteristics in *C. subternata* 42
3.3.2 Cross-species amplification 43
3.3.3 DNA fingerprinting and allelic diversity 46
3.4 Conclusion 51
References 52
Chapter 4: Population Genetic Diversity in the Polyploid *Cyclopia subternata* 56
Abstract 56
4.1 Introduction 56
4.2 Materials and Methods 59
4.2.1 Sampling and DNA extraction 59
4.2.2 PCR amplification 60
4.2.3 Genotyping 61
4.2.4 Data analysis 62
4.3 Results 64
4.3.1 Genetic diversity 64
4.3.2 Genetic differentiation and clustering among wild populations 64
4.3.3 Genetic differentiation and clustering between wild populations and accessions 67
4.4 Discussion 71
4.4.1 Comparison of various population genetic statistics and software for polyploids 71
4.4.2 Accessions and capturing of genetic variation 74
4.4.3 Genetic differentiation of wild populations and implications for conservation 75
4.5 Conclusion 77
References 78
Chapter 5: Study Conclusions 84
5.1 Background 84
5.2 Synopsis of Main Results and Discussion 84

Stellenbosch University https://scholar.sun.ac.za

5.3 Study Limitations 86
5.4 Future Work 88
5.5 Final Remarks 90
References 90
Appendix 95

Stellenbosch University https://scholar.sun.ac.za

List of Figures

Figure 1.1. The small, narrow, needle-like leaves of *C. genistoides* (A) in comparison to the larger, flattened leaves of *C. longifolia* (B) and *C. subternata* (C) 2

Figure 2.1. (A) Young leaves, (B) young stems and (C) old leaves in *C. subternata*. Young leaves are lighter in colour, while old leaves are dark green to grey 20

Figure 2.2. Leaves of the seven *Cyclopia* species included in this study showing variation in size and morphology. (A) *C. genistoides*, (B) *C. longifolia*, (C) *C. subternata*, (D) *C. sessiliflora*, (E) *C. intermedia*, (F) *C. pubescens* and (G) *C. maculata* 25

Figure 2.3. DNA concentration ranged from 205-727ng/µl for the seven *Cyclopia* species sampled from the ARC field gene bank for further molecular studies. A_{260}:A_{280} ratios ranged from 1.8 to 2.1 and the A_{260}:A_{230} ratios ranged from as low as 1.4, to a maximum of 2 26

Figure 2.4. DNA concentration for the ten accessions of *C. subternata* that were used to develop the microsatellite markers ranged from 200-650ng/µl. A_{260}:A_{280} ratios ranged from 1.88 to 2.0 and the A_{260}:A_{230} ratios was also well within the required range (1.76-2.14) 27

Figure 2.5. Visual comparison between the two buffers used to preserve the *Cyclopia* leaves before DNA was extracted. All figures denoted with A relate to the CTAB buffer and all figure denoted with B to the CTAB/NaCl buffer. A1) the buffer, (A2) leaves and (A3) DNA solution of the 2% CTAB buffer turned brown after 2 weeks. This indicates that phenolic compounds seeped from the leaves and contaminated the DNA during extraction. B1) the buffer, (B2) leaves and (B3) DNA solution of the CTAB/NaCl buffer stayed light green after 2 weeks, indicating less phenolic compounds seeped from the leaves and therefore better preservation 29

Figure 2.6. The average DNA quantity (in ng/µl) was plotted on the vertical axis and the quality ratios were plotted on a secondary vertical axis to compare how well *C. subternata* samples were preserved in two types of CTAB buffer. Data for day 1-14 was pooled and no significant differences were observed. Error bars indicates the standard deviation of DNA concentration 29

Figure 3.1. The average percentage transferability of 19 *C. subternata* microsatellite markers to six other *Cyclopia* species. Transferability was determined as the average percentage amplification across all accessions within a species. Error bars indicates the standard deviation from the average percentage amplification 43

Stellenbosch University https://scholar.sun.ac.za

Figure 3.2. The percentage transferability of 19 *C. subternata* microsatellite markers to six other *Cyclopia* species (represented by the various colours). Transferability was calculated as the percentage of accessions in which the marker amplified for the six species. The amplification of markers in *C. subternata* and the total percentage transferability per marker are also indicated 44

Figure 3.3. Phylogenetic relationships of *Cyclopia* species according to a study conducted by Boatwright *et al*. (2008). The maximum parsimony tree was constructed based on ITS and *rbcL* data. The numbers above the branches represent Fitch lengths, while values below the branches are bootstrap percentages above 50% (equal weights before the slash and successive weights after the slash). The seven species of interest in the current study are indicated by an * with (R) indicating a sprouter and (S) indicating a non-sprouter 45

Figure 3.4. Examples of microsatellite peaks from GeneMapper 4.0 for a *C. subternata* accession, SKB18 for locus (A) Cys5, (B) Cys10, (C) Cys21b, (D) Cys22, (E) Cys25 and (F) Cys88 46

Figure 3.5. Example of GeneMapper traces showing the complicated peak pattern of A) *C. genistoides*, a decaploid in comparison with the simpler pattern of B) *C. subternata*, a hexaploid, for the microsatellite locus Cys25. 47

Figure 3.6. Principal coordinate analysis (PCoA) of accessions from the ARC field gene bank of five *Cyclopia* species based on Bruvo's genetic distance measure 50

Figure 4.1. Natural distribution of six commercially important *Cyclopia* species in the Eastern and Western Cape of South Africa (Joubert *et al*., 2011). *Cyclopia subternata* is localised in the Tsitsikamma, Outeniqua and Langeberge mountains, stretching from Riversdal to Plettenberg Bay and Port Elizabeth (as denoted by a 6) 58

Figure 4.2. Maps showing the locations of the three wild populations sampled of *C. subternata*. Map A indicates the locations of the populations in South Africa, while map B focuses on the exact areas. Population 1 from Guava Juice (S 34.04595; E 024.34648) is in the Eastern Cape and populations 2 from Haarlem (S 33.777688; E 023.304393) and 3 from George (S 33.88580; E 022.34306) are in the Western Cape 60

Figure 4.3. Chart depicting the presence of alleles for six microsatellite loci (Cys5, Cys10, Cys21b, Cys22, Cys25 and Cys88) in 79 individuals of wild populations and 22 gene bank accessions of *C. subternata*. Note that this is not the distribution of alleles, since allele frequencies could not be calculated for this hexaploid species 66

Figure 4.4. Principal coordinate analysis plots based on three genetic distance matrices for three *C. subternata* wild populations and gene bank accessions. Population 1 was sampled from Guava

Stellenbosch University https://scholar.sun.ac.za

Juice in the Eastern Cape, population 2 is from Haarlem in the Western Cape and population 3 is from George in the Western Cape 70

Figure 4.5. The optimum K for populations of *C. subternata* was chosen as 2 according to Evanno's method implemented in Structure Harvester 71

Figure 4.6. STRUCTURE bar plots with (A) K = 2 chosen as the most likely number of clusters and (B) K = 3 (included for comparison) for cultivated accessions and three wild populations of *C. subternata*. The K = 4 minor cluster was not included as this did not depict suitable ancestry. Population 1 was sampled from Guava Juice in the Eastern Cape, population 2 is from Haarlem in the Western Cape and population 3 is from George in the Western Cape..... 71

Stellenbosch University https://scholar.sun.ac.za

List of Tables

Table 2.1. Distribution of residuals and homoscedasticity of *Cyclopia* DNA quality data before and after transformation with appropriate functions for three dependent variables ... 21

Table 2.2. Two-way factorial design (linear model) for three dependent variables of *Cyclopia* DNA quality data after transformation ... 21

Table 2.3. Bonferroni t-test groupings of the range and mean DNA quantity in ng/µl for the interaction between various tissue types as starting material for *C. genistoides* and *C. subternata* species ... 22

Table 2.4. Bonferroni t-test groupings of the range and mean A_{260}:A_{280} for the interaction between various tissue types as starting material for *C. genistoides* and *C. subternata* species ... 22

Table 2.5. Bonferroni t-test groupings of the range and mean A_{260}:A_{230} for the interaction between various tissue types as starting material for *C. genistoides* and *C. subternata* species ... 23

Table 2.6. The seven *Cyclopia* species and their number of accessions sampled from the ARC gene bank with different leaf sizes and structure ... 24

Table 3.1. Seven *Cyclopia* species and their accessions from the ARC field gene bank used for genetic marker development ... 35

Table 3.2. Screening of 27 microsatellite markers in 22 accessions of *C. subternata* according to five criteria ... 40

Table 3.3. Rules for genotyping with six microsatellite markers in the polyploid *Cyclopia* genus.. 41

Table 3.4. Information for the six primer pairs incorporated into *Cyclopia* microsatellite multiplex panels (HB-Panel 1 and 2) used for DNA fingerprinting ... 47

Table 3.5. Summary table for the performance of the six *C. subternata* markers for DNA fingerprinting purposes in each of the seven *Cyclopia* species ... 48

Table 3.6. Allelic diversity of accessions of five *Cyclopia* species from the ARC gene bank for six microsatellite markers (HB-Panel 1 and HB-Panel 2) ... 48

Table 4.1. Primer information for six *C. subternata* microsatellite markers in three wild populations and 22 accessions ... 65

Table 4.2. Allele diversity statistics for three wild populations of *C. subternata* as well as cultivated accessions ... 65

Stellenbosch University https://scholar.sun.ac.za

Table 4.3. Partitioning of the observed *C. subternata* microsatellite variation (A) within and among three wild populations and (B) within and among wild populations and accessions using three genetic distance measures appropriate for polyploid data. P-values indicated at the 5% significance level . 68

Table 4.4. Pairwise ϕ_{PT} and ρ_{ST} values between three wild populations and cultivated accessions of *C. subternata*. Significant P-values (5% level) are indicated above the diagonal 69

Table 4.5. Climatic conditions as measured by the ARC AgroMet weather stations closest to the wild populations of *C. subternata* (indicated in brackets) across three years (2012-2014)....................... 76

Table A1. The primer information for the 27 microsatellite markers developed by Genetic Marker Services for *C. subternata*. Only 22 primers were chosen for fluorescent labelling with the dyes indicated 95

Table A2. DNA fingerprints for the 81 accessions of 5 *Cyclopia* species in the ARC field gene bank as determined by six microsatellite markers. Missing genotypes are denoted by -9, as this is the symbol used in POLYSAT 97

Table A3. Locations from where the *Cyclopia* species accessions, currently in the ARC field gene bank situated at Nietvoorbij and Elsenburg in Stellenbosch, were originally obtained 110

Table A4. Characteristics of the three genetic distance matrices used throughout the study 113

Table A5. Visual representation of distribution of residuals and homoscedasticity of *Cyclopia* DNA quality data before and after transformation with appropriate functions for three dependent variables. Supplementary to the data given in Table 2.1 114

Stellenbosch University https://scholar.sun.ac.za

List of Abbreviations

%	Percentage
°C	Degrees Centigrade
~	Approximately
$>$	More than
$<$	Less than
®	Registered trademark
µl	Microliters
µg	Micrograms
µM	Micromolar
ϕ_{PT}	PhiPT value
ρ_{ST}	RhoST value
2*n*	Diploid chromosome number
3’	Three prime
5’	Five prime
6-FAM	Blue fluorescent dye
A	Adenine
ABI 3730	Applied Biosystems Analyzer
AFLP	Amplified Fragment Length Polymorphism
AMOVA	Analysis of Molecular Variance
ANOVA	Analysis of Variance
ARC	Agricultural Research Council
bp	Base pair
BSA	Bovine Serum Albumin
C	Cytosine
CAF	Central Analytical Facility
CBOL	Consortium for the Barcode of Life
CFR	Cape Floristic Region
Cj	Confusion probability
CTAB	Hexadecyltrimethylammonium bromide
ddH_2O	Double distilled water

Stellenbosch University https://scholar.sun.ac.za

df	Degrees of freedom
Dj	Discriminatory power
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleotide Triphosphate
DST	Department of Science and Technology
EC	Eastern Cape
EDTA	Ethylenediaminetetraacetic acid
EtBr	Ethidium bromide
EST	Expressed Sequence Tag
F	Forward primer
F_{ST}	Fixation Index
g	Centrifugal gravitational force
g	Grams
G	Guanine
GMS	Genetic Marker Services
gt	Genotypes
h	Hour
HB	Honeybush
HWE	Hardy-Weinberg Equilibrium
Hz	Hertz
IBD	Isolation by distance
ITS	Internal transcribed spacer
JM109	Competent cells
K	Ancestral cluster
LB	Luria-Bertani medium
m	Metres
MAC-PR	Microsatellite DNA allele counting - peak ratios
matK	Maturase K
Max	Maximum
MCMC	Markov chain Monte Carlo
mg	Milligrams

Stellenbosch University https://scholar.sun.ac.za

$MgCl_2$	Magnesium chloride
min(s)	Minute(s)
Min	Minimum
ml	Millilitres
mm	Millimetres
mM	Millimolar
MTEF	Monetary Treasury Economic Fund
m/v	Mass per volume
N_A	Number of alleles
NaCl	Sodium chloride
NED	Yellow fluorescent dye
ng	Nanograms
nm	Nanometres
NRF	National Research Foundation
p	Allele present
PBR	Plant Breeders Rights
PCoA	Principal Coordinate Analysis
PCR	Polymerase Chain Reaction
PET	Red fluorescent dye
pg	Picograms
PIC	Polymorphic Information Content
P_{ID}	Probability of Identity
pmol	Picomoles
Pop 1	Wild population 1, sampled from Guava Juice in the EC
Pop 2	Wild population 2, sampled from Haarlem in the WC
Pop 3	Wild population 3, sampled from George in the WC
pt	Phenotypes
P-value	Probability value at 5% significance level
PVP	Polyvinylpyrrolidone
q	Allele absent
R	Reverse primer

Stellenbosch University https://scholar.sun.ac.za

r^2	Squared correlation coefficient
RAD-seq	Restriction-site associated DNA sequencing
RAPD	Random Amplified Polymorphic DNA
rbcL	Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit
RCF	Relative centrifugal force
RFLP	Restriction Fragment Length Polymorphism
RFU	Relative Fluorescent Units
RNA	Ribonucleic Acid
RNase	Ribonuclease
*Rsa*I	Restriction enzyme isolated from *Rhodopseudomonas sphaeroides*
s	Seconds
SANBI	South African National Biodiversity Institute
SDRFs	Single Dose Restriction Fragment Markers
SMM	Stepwise Mutational Model
SNP	Single Nucleotide Polymorphism
SSC	Saline-sodium citrate
SSR	Simple Sequence Repeat
T	Thymine
TAE	Tris-acetate-EDTA buffer
Taq	*Thermus aquaticus* DNA polymerase
TBE	Tris/Borate/EDTA buffer
T_M	Annealing temperature
Tris-HCL	Tris(hydroxymethyl)aminomethane hydrochloride
VIC	Green fluorescent dye
v/v	Volume per volume
WC	Western Cape
w/v	Weight per volume
x	Basic chromosome number

Stellenbosch University https://scholar.sun.ac.za

Chapter 1: Literature Review

1.1 *Cyclopia* Species

1.1.1 Taxonomy and distribution

The Cape Floristic Region (CFR) is situated at the south-western tip of southern Africa. It comprises a large number of diverse species and is characterised by a Mediterranean climate and winter rainfall (Linder, 2003; Du Toit, 2005). Fabaceae is a leguminous plant family and the second largest in the CFR with approximately 760 species (Kamara *et al*., 2004; Du Toit, 2005; Joubert *et al*., 2011). The genus *Cyclopia* is placed in the tribe Podalyrieae (family Fabaceae) based on morphological characters such as the presence of trifoliate leaves and free stamens (Schutte, 1997). Tribal affinities were discerned based on phylogenetic data obtained using the ribosomal internal transcribed spacer (ITS) region. *Cyclopia* was found to be strongly monophyletic and sister to the rest of Podalyrieae, thereby forming a separate clade within this tribe. The other two clades comprised the Xiphothecinae (consisting of the species *Amphitalea*, *Coelidium* and *Xiphotheca*) and Podalyriinae (consisting of species *Calpurnia*, *Liparia*, *Stirtonanthus* and *Virgilia*) (Du Toit, 2005). Species of *Cyclopia* grow within the coastal and mountainous areas of the Cape. Currently, 23 species are recognised, of which six are utilised for commercial purposes (Marnewick, 2009; Joubert *et al*., 2011). Several of these species are, however, listed as endangered or critically endangered including *C*. *longifolia* and *C*. *pubescens*. Species listed as rare or near threatened include *C*. *genistoides* and *C*. *maculata*, while species such as *C*. *intermedia* are rapidly declining (SANBI, 2012).

1.1.2 Botany

Cyclopia species grow naturally as shrubs that reach a height of 1.5-3m (Schutte, 1997). Variation in size and shape exist between the trifoliate leaves of different species; from small, narrow, needle-like leaves (*C*. *genistoides*) to the larger, flattened leaves of *C*. *longifolia* and *C*. *subternata* (Figure 1.1) (Du Toit *et al.,* 1998). Bright yellow flowers appear during spring (September to October) and emit the characteristic sweet smell of honey which thereby lends its name to honeybush (Du Toit *et al.,* 1998).

Stellenbosch University https://scholar.sun.ac.za

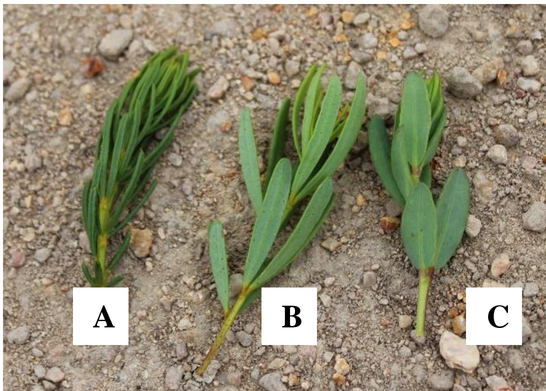


Figure 1.1. The small, narrow, needle-like leaves of *C. genistoides* (A) in comparison to the larger, flattened leaves of *C. longifolia* (B) and *C. subternata* (C).

Fynbos plants, including *Cyclopia*, require veld fires to ensure successful reproduction. In the absence of fire, fynbos becomes senescent and degenerates rapidly, thereby allowing invasive species to settle and take over (Du Toit, 2005). Fynbos plants can be classified into two categories on the basis of their fire survival strategies, namely sprouters and non-sprouters. Sprouters, such as *C. genistoides* and *C. intermedia*, have a woody rootstock that can produce new coppice shoots after the occurrence of a fire. Non-sprouters, including *C. maculata* and *C. subternata*, are dependent on seeds from the soil bank to germinate and grow (Schutte, 1997; Du Toit, 2005; Joubert *et al.*, 2011). In order for germination to take place, seeds need to be scarified. In nature, fire is responsible for the scarification process, but artificial methods such as chemicals or abrasion techniques can be applied under laboratory conditions (Joubert *et al.*, 2008; Luna *et al.*, 2009; Koen *et al.*, 2016).

Cyclopia species are polyploids with a chromosome basic number of $x = 9$. The only published data available on *Cyclopia* chromosome counts are those of *C. maculata* ($2n = 4x = 36$), *C. subternata* ($2n = 6x = 54$) and *C. intermedia* and *C. meyeriana* ($2n = 14x = 126$) (Goldblatt, 1981; Schutte, 1997). Recently it was established that *C. genistoides* is a decaploid ($2n = 10x = 90$) and *C. longifolia* is a hexaploid ($2n = 6x = 54$) (Motsa, 2016).

1.1.3 Commercial importance

The leaves, stems and flowers of *Cyclopia* are used for honeybush tea production. The brew is alternatively known as kustee, heuningtee, bergtee or vleitee, depending on the species from which it is produced (McKay and Blumberg, 2007; Joubert *et al.*, 2011). Owing to its potential health effects, honeybush tea has been utilised for years as a folk remedy to treat a number of health problems such as skin problems, respiratory irritations and digestive disorders as well as the stimulation of milk production in women (Kamara *et al.*, 2003; McKay and Blumberg, 2007; Marnewick, 2009). An

Stellenbosch University https://scholar.sun.ac.za

interest in these health benefits led to studies on the phytochemical composition of honeybush, specifically the phenolic compounds (Kamara *et al*., 2004; Kokotkiewicz *et al*., 2012). According to De Nysschen (1996), three major phenolic compounds are present in the leaves of *Cyclopia*, namely mangiferin (a xanthone) and two flavanones, hesperetin and isosakuranetin. Mangiferin has antidiabetic properties, while hesperetin is mainly found in citrus, exhibiting antioxidant as well as anti-inflammatory activity (Joubert *et al*., 2011; Parhiz *et al*., 2015).

Numerous *in vitro* studies have been performed to investigate the antimutagenic and antioxidant activities of honeybush tea. The brew contains low amounts of caffeine and only 0.45% tannins, thereby making it a healthier option than oriental teas (Du Toit, 2005; McKay and Blumberg, 2007). Several animal model studies corroborate the *in vitro* findings, but no clinical trials on humans have yet been conducted (McKay and Blumberg, 2007; Marnewick, 2009).

The main stages of honeybush tea processing are: harvesting, cutting, fermentation, drying, sieving and packaging. Traditionally, harvesting took place during the flowering season. The sweet honey flavour of the tea was first believed to be reliant on the flowers, but is now attributed to the spontaneous oxidation of phenolic compounds during fermentation (Du Toit *et al*., 1998; Joubert *et al*., 2008). The leaves, stems and flowers are cut into smaller pieces before fermentation in order to facilitate rapid oxidation of phenolic compounds (Du Toit *et al*., 1998; Joubert *et al*., 2008). During earlier practices, the leaves and flowers were simply sun-dried without any method of fermentation. The first fermentation method for honeybush was heap fermentation, but this lead to the growth of unwanted microorganisms. Batch rotary fermentation was an improved fermentation process that entailed the exposure of leaves to very high temperatures (80-85°C) for several hours, followed by drying under controlled conditions (Du Toit *et al*., 1998; Joubert *et al*., 2008). Lastly, the tea leaves are sieved to ensure a finer product. Honeybush is often mixed with other indigenous South African plants, such as buchu and rooibos, before being packaged either as loose tea leaves or tea bags (Joubert *et al*., 2008).

Even though honeybush tea has a long history of use as an herbal tea, it was at first not a successful commercial venture and processing was discontinued. Following the success of rooibos tea, the need for cultivation of honeybush was recognised by Dr J. H. de Lange of the South African National Biodiversity Institute (SANBI) (Joubert *et al*., 2008; Joubert *et al*., 2011). The ARC Honeybush Breeding and Selection Programme was initiated in 1992 as an improvement programme that aims to produce cultivars for commercial purposes (Joubert *et al*., 2011). A total of twelve different honeybush species were evaluated for their commercial potential from which only a few were chosen

Stellenbosch University https://scholar.sun.ac.za

to be developed commercially, including *C. genistoides*, C. *intermedia* and *C. subternata* (Bester *et al*., 2016). Breeding practices undertaken at the ARC have been mainly conventional and a honeybush field gene bank was established in 2010 that contains 15 accessions of *C. genistoides*, 30 accessions of *C. longifolia* and 25 accessions of *C. subternata* (Bester *et al*., 2013). The next phase in the breeding programme was the use of polycrosses, as well as controlled intra-species crosses (Joubert *et al*., 2011). Polycross designs allow for a group of clones in isolation to cross-pollinate freely, giving rise to half-sib progenies with unknown parental contributions (Poehlman, 1983; Acquaah, 2007). Eventually, an improved artificial variety in terms of seed yield is produced (Poehlman, 1983). Currently, genetic material with improved biomass is released to the honeybush industry as seeds obtained from clonal seed-orchards. Future aims also include the release of vegetatively propagated clones as registered cultivars after application for Plant Breeders Rights (PBR) (Bester *et al*., 2016).

Several factors can influence the cultivation potential of a plant. *Cyclopia* seeds often exhibit dormancy and low germination rates, even if environmental conditions are favourable (Koen *et al*., 2016). Species such as *C. intermedia* have slow growth rates and can only be harvested every second or third year (Joubert *et al*., 2011). Cultivation is therefore often not a fruitful venture and only 25% of the *Cyclopia* crop is obtained in this manner, while the majority is currently harvested from the wild to meet the increasing demand (Bester, 2012). Even if wild-harvesting is permitted, strict control should be implemented to ensure it does not lead to over-harvesting and a decline in the genetic diversity of wild populations (Schippmann *et al*., 2006). Additionally, it is recommended that honeybush plantations should not be situated in close proximity to wild populations, as uncontrolled gene flow can compromise the genetic integrity of endangered species (Campbell *et al*., 2016).

A review by Joubert *et al*. (2011) highlighted the need for more studies on the genetic composition of honeybush to infer phylogenetic relationships, chromosome numbers and ploidy levels. Despite the economic importance of *Cyclopia* species, no molecular markers have been developed for any of the species. This creates an ideal opportunity to exploit appropriate DNA marker technologies in order to generate DNA information for future applications in the breeding programme.

1.2 DNA Marker Technologies

Several types of molecular markers are available: each with its own advantages and disadvantages. Restriction fragment length polymorphisms (RFLPs) work on the principle of restriction enzymes that cut DNA at particular sites. Variations in restriction sites allow for different banding patterns (Lowe *et al*., 2004). RFLP markers are highly reproducible and their co-dominant nature allows for

Stellenbosch University https://scholar.sun.ac.za

distinguishing between homozygous and heterozygous genotypes (Agarwal *et al*., 2008). This method, however, usually requires large amounts of DNA (Lowe *et al*., 2004). Randomly amplified polymorphic DNA (RAPD) is a marker technique whereby a random primer, usually consisting of ten nucleotides, is amplified throughout the genome and bands are scored as present or absent. This method requires no sequence information and is cost-effective, but several studies have reported technical issues such as low reproducibility (Chalmers *et al*., 2001; Agarwal *et al*., 2008). Amplified fragment length polymorphisms (AFLPs) combine the use of restriction enzymes and PCR amplification to ensure reproducibility. This method utilises two pairs of restriction enzymes: a rare cutter and a frequent cutter. Fragments are amplified by PCR and scored on the basis of presence or absence of a band. Similar to RFLPs, large amounts of DNA are needed as starting material (Lowe *et al*., 2004).

Microsatellite markers, also referred to as simple sequence repeats (SSRs), are PCR-based single locus DNA markers that consist of short, tandemly repeated nucleotides (Selkoe and Toonen, 2006). Microsatellites display high levels of polymorphism and reproducibility, making them the preferred marker for DNA fingerprinting and population genetic diversity studies (Kalia *et al*., 2011; Nybom *et al*., 2014). Based on the nucleotide repeat motif, microsatellites can be classified as mono-, di-, tri-, tetra-, penta- or hexanucleotides. They are relatively abundant throughout the genome, occurring in both coding and non-coding regions (Wang *et al*., 2009; Kalia *et al*., 2011). Microsatellites are co-dominant markers; consequently, heterozygous and homozygous peak patterns can be distinguished in a diploid organism. In polyploid organisms, peak patterns are often complicated by the inability to determine allele dosage, where the maximum number of alleles present in one individual will be equal to the ploidy level (Ouborg *et al*., 1999; De Silva *et al*., 2005; Sampson and Byrne, 2012).

Microsatellite variation can be detected by PCR amplification as the number of microsatellite repeats that exist among individuals, which is manifested as size polymorphism (Chapuis and Estoup, 2007; Kalia *et al*., 2011). Several detection methods are available for separating the products and for allowing scoring and viewing of microsatellite data. Polyacrylamide or agarose gel electrophoresis are older methods that were frequently utilised before the advent of automated sequencers. With these methods, microsatellite alleles are visualised as bands by staining with ethidium bromide or silver. The major constraint of these gel methods is the loss of resolution that occurs, especially with agarose gel electrophoresis leading to inaccurate size calling. Automated systems, whether capillary or gel based, visualise alleles as peaks thereby enabling easier and more reliable data scoring and handling (Wang *et al*., 2009).

Stellenbosch University https://scholar.sun.ac.za

Various aspects, such as reproducibility and running costs, should be considered when choosing microsatellites above other molecular marker techniques. The level of polymorphism is much higher for microsatellites in comparison with RFLPs, RAPDs and AFLPs (Kalia *et al*., 2011). In a study by McGregor *et al*. (2000), RAPD markers had lower reproducibility and informativeness when compared to microsatellite and AFLP markers. Overall, microsatellites have been shown to display a large number of favourable characteristics in terms of running cost, automation, repeatability and even the level of training required in comparison with other methods (Rao and Hodgkin, 2002). A disadvantage to microsatellites is that development is an expensive and laborious effort, therefore species-specific microsatellite loci are only available for a limited number of important plant taxa. Cross-transferability of the available characterised markers to other closely related species would circumvent the need for species-specific markers (Kuleung *et al*., 2004). The success of cross-species transferability relies on conserved flanking regions across taxa that allow for primer development (Rossetto, 2001; Rai *et al*., 2013).

A considerable number of studies have shown that gene content and order are highly conserved among plant species that are closely related (Rai *et al*., 2013). Several related species display conserved primer sequences; for instance, up to 65% cross-amplification of microsatellites was found between species within the genus *Glycine* (Peakall *et al*., 1998). Similarly, wheat microsatellites were transferred to triticale with a success rate of 58% (Kuleung *et al*., 2004). Therefore, the utilisation of cross-species amplification of microsatellites for *Cyclopia* may be a feasible option. Challenges that arise during cross-species amplification should also, however, be considered. For example, null alleles, where changes in the primer flanking region cause the non-amplification of an allele, have been reported to be more prevalent with increasing phylogenetic distance (Chapuis and Estoup, 2007). The presence of null alleles can lead to an overestimation of homozygotes and have further implications for the interpretation of population statistics. Size homoplasy causes additional problems, since the presence of similar sized amplification products from related species might mask undetectable mutations such as reversions (Selkoe and Toonen, 2006). In addition, the duplication of primer sites can lead to multiple banding patterns, while false positives can occur when the microsatellite repeat is completely absent from the amplified product of the study species (Rossetto, 2001).

1.3 Applications of DNA Marker Technologies in Plant Genetic Resources

The development of improved and higher yielding plant cultivars through the selection of favourable characteristics by plant breeders has enabled crops to be grown at an increasingly larger scale. The

Stellenbosch University https://scholar.sun.ac.za

selection of genotypes from only a small number of elite lines and subsequent large scale production has resulted in a steady decline of genetic diversity within the cultivated material (Huang *et al*., 2002; Rao and Hodgkin, 2002). This, in turn, results in genetic vulnerability which can lead to greater susceptibility to disease and insect damage (Godwin, 2009). Conservation efforts are therefore crucial to ensure the accessibility of natural genetic variation and new genes for continuous enrichment of the existing gene pool (Ford-Lloyd, 2001).

There are two approaches to the conservation of plant genetic resources. *In situ* conservation is the maintenance of genetic resources in their natural habitat, while *ex situ* refers to the conservation of these resources outside their natural habitat in the form of field gene banks or seed stores (Rao, 2004). Ideally, plant genetic resources should capture all the genetic diversity of the cultivated species and their wild relatives (Rao and Hodgkin, 2002).

Traditional methods of identifying accessions in a field gene bank can be time consuming as they rely on phenotypic observations that are, in turn, influenced by environmental factors (Karp *et al*., 1997; Wünsch and Hormaza, 2002). DNA fingerprinting is an alternative method that allows for rapid identification of accessions on the DNA level at any stage of the plant development (Karp *et al*., 1997; Wünsch and Hormaza, 2002; Rao, 2004). Once the field gene bank is properly characterised, management of the accessions can include, for example, identifying incorrect labelling as well as duplicates (Fowler and Hodgkin, 2004; Rao, 2004). This creates space for wider variation in plants in the field gene bank and prevents the waste of land resources (Ford-Lloyd, 2001; Rao and Hodgkin, 2002; Rao, 2004). The identification of clones to be used in breeding applications or commercialisation is an important aspect that can be addressed using molecular markers. Furthermore, field gene banks often contain plants regenerated from seeds under different selection pressures than in their natural habitat. Molecular markers can be used to monitor the genetic integrity of these accessions (Spooner *et al*., 2005).

DNA profiles can also assist in the protection of plant breeder's rights in lending support to a plant breeder's claim that their variety is distinct. This is especially useful in cases where there are no apparent phenotypic differences between the different varieties (Lee and Henry, 2001; Rajora and Rahman, 2003). Furthermore, underrepresented wild populations can be identified and sampling strategies can include more wild samples that properly represent the genetic diversity of the species (Rao, 2004; Gepts, 2006).

Stellenbosch University https://scholar.sun.ac.za

1.4 Polyploidy and Molecular Markers

During meiosis in a diploid organism, four daughter cells, each with half the number of chromosomes, are formed from one parent cell (Zamariola *et al.*, 2014). This ensures that no chromosome doubling occurs, resulting in progeny with the same ploidy as the parents (Moore, 2013). During the first meiotic division, homologous chromosomes form bivalents and segregate from each other (Zamariola *et al.*, 2014). Polyploidy is the phenomenon whereby an organism has more than two sets of chromosomes and can originate when unbalanced or unreduced gametes formed during meiosis (Aversano *et al.*, 2012; Zamariola *et al.*, 2014).

Polyploidy is a common phenomenon among plants, especially in angiosperms where approximately 70% have experienced polyploidisation in their history (Masterson, 1994; Soltis and Soltis, 1999). Polyploids often exhibit higher genetic diversity than diploids as variation is introduced by the formation of multiple sets of chromosomes (Soltis and Soltis, 1993; Adams and Wendel, 2005; Eliášová *et al.*, 2014). An evolutionary advantage of polyploidy is that duplicated gene copies can be altered to diversify gene function in a process known as diploidization, thereby lending greater diversity to the organism (Comai, 2005).

Polyploid organisms can be classified as autopolyploids or allopolyploids, or even display mixed inheritance patterns. In autopolyploids, a doubling in the genome occurred within a single species; therefore, each chromosome can pair with more than one homologue. Chromosomes typically form multivalents during segregation in meiosis, leading to polysomic inheritance. In allopolyploids, two genomes from different species hybridised and homologous chromosomes pair, leading to disomic inheritance (Meirmans and Van Tienderen, 2013). Exceptions occur where the homoeologous parental chromosomes are similar enough and can pair during meiosis, thereby leading to mixed inheritance patterns in certain allopolyploids (Jeridi *et al.*, 2012; Dufresne *et al.*, 2014).

Allopolyploids are thought to be more widespread than autopolyploids due to several proposed evolutionary advantages (Parisod *et al.*, 2010). They exhibit fixed heterozygosity when parental chromosomes do not segregate during bivalent formation, thereby reducing inbreeding depression (Werth *et al.*, 1985; Städler *et al.*, 1993; Lowe *et al.*, 2004). Hybrid vigour in allopolyploids often leads to increased phenotypic characteristics such as larger flowers or seeds, an increase in growth rate, higher fertility and novel morphological characteristics (Lowe *et al.*, 2004; Sattler *et al.*, 2016).

Stellenbosch University https://scholar.sun.ac.za

Most population genetic statistical tools are based on diploid assumptions with a disomic mode of inheritance, and therefore cannot be used directly for analysis of polyploid genomes. The mode of inheritance will influence allele frequency calculations, which is a prerequisite during population genetic data analyses for further inferences such as the calculation of expected heterozygosities and F_{ST} values (Dufresne *et al*., 2014). Unfortunately, allele frequencies can often not be calculated for higher order polyploids, such as hexaploids and octaploids, as allele dosage is not discernible (De Silva *et al*., 2005; Pfeiffer *et al*., 2011). Further challenges arise during genotyping when true alleles cannot be distinguished from artifacts such as stutter peaks and pull-up peaks. The presence of null alleles also complicates the assessment of reliable allele and genotype frequencies (Dufresne *et al*., 2014). Several extensions of diploid population genetic statistics have, however, successfully been applied to polyploid studies and each of these will be discussed in more detail below (Dufresne *et al*., 2014; Wang and Scribner, 2014).

The most challenging population genetic statistic to apply to polyploids is the estimation of allele frequencies. Instead of full genotypes with a known number of allele copies, multilocus allelic phenotypes are determined that simply infer the presence and identity of the various alleles. The subsequent construction of binary arrays allows for the interpretation and analysis of dominant rather than co-dominant data (Sampson and Byrne, 2012). This creates a loss of genetic information, since full genotypes cannot be obtained to calculate allele frequencies. It is, however, still a preferred method for polyploids as their diverse microsatellite profiles (comprised of multiple alleles) often permit the use of fewer loci when compared to diploids (Pfeiffer *et al*., 2011; Zawedde *et al*., 2015). Several measures can be calculated from the binary arrays such as allelic diversity as well as the total number of alleles over all loci, the number of different alleles in each population and the number of alleles per individual (Sampson and Byrne, 2012).

Multivariate analysis is suitable for polyploid analyses, since it does not have any underlying assumptions about Hardy-Weinberg Equilibrium (HWE) (Dufresne *et al*., 2014). Genetic distances among individuals can be calculated from the binary arrays and visualised on a principal coordinate analysis (PCoA) plot (Sampson and Byrne, 2012; Dufresne *et al*., 2014). Bruvo distance is implemented in POLYSAT, an R based package especially developed for dealing with polyploid data, and calculates pairwise genetic distances without the need for allele dosage information (Bruvo *et al*., 2004; Clark and Jasieniuk, 2011; Dufresne *et al*., 2014). A disadvantage of using Bruvo distance is that it may falsely group individuals with the same ploidy level if there are mixed ploidies involved in the analyses (Dufresne *et al*., 2014).

Stellenbosch University https://scholar.sun.ac.za

1.5 Study Rationale, Aims and Objectives

1.5.1 Problem statement

Despite the commercial importance of *Cyclopia,* very little genetic information is available for this genus. The application of molecular markers, such as microsatellites, can provide valuable insight for breeding and conservation efforts of *Cyclopia* species.

1.5.2 Study aim and objectives

The aim of this project is to develop genetic marker resources, specifically microsatellites, for application in *Cyclopia* species.

No DNA extraction protocol that uses fresh leaves as starting material is available for this genus; therefore, the first experimental chapter (Chapter 2) will focus on the development and optimisation of a DNA extraction protocol for *Cyclopia* to minimise the interference of phenolic compounds. For this, different tissue types will be tested to evaluate which potentially yields the highest quality and quantity of DNA. Subsequently, to estimate whether the extraction protocol can successfully be applied across all species, DNA extraction will be performed on seven different *Cyclopia* species for which accessions are available in the ARC germplasm collection. Lastly, a preservation technique will be tested to examine the effect on DNA quality and quantity during sampling of wild populations from remote locations.

Chapter 3 will discuss the development of microsatellite markers for the commercially important species, *C. subternata*. These microsatellite markers will be applied and optimised to test their transferability to six other *Cyclopia* species that are currently represented in the ARC field gene bank (*C. genistoides*, *C. intermedia*, *C. longifolia*, *C. maculata*, *C. pubescens* and *C. sessiliflora*). DNA fingerprints will be generated for accessions of *C. longifolia*, *C. maculata*, *C. pubescens*, *C. sessiliflora* and *C. subternata* since the ARC needs to identify accessions in the gene bank with confidence.

Lastly, the microsatellites will be used to study the population genetic structure of three wild populations of *C. subternata* in Chapter 4. The high ploidy levels of *Cyclopia* species will not permit a full-scale population genetic study and focus will shift to using available statistical tools to combat problems that commonly arise when working with polyploid organisms.

Stellenbosch University https://scholar.sun.ac.za

References

Acquaah, G., 2007. Introduction to quantitative genetics, in: Principles of Plant Genetics and Breeding. Blackwell, Oxford, UK, 142-144.

Adams, K. & Wendel, J., 2005. Polyploidy and genome evolution in plants. Current Opinion in Genetics and Development 8, 135-141.

Agarwal, M., Shrivastava, N., Padh, H., 2008. Advances in molecular marker techniques and their applications in plant sciences. Plant Cell Reports 27, 617-631.

Aversano, R., Ercolano, M.R., Caruso, I., Fasano, C., Rosellini, D., Carputo, D., 2012. Molecular tools for exploring polyploid genomes in plants. International Journal of Molecular Sciences 13, 10316-10335.

Bester, C., 2012. A model for commercialisation of honeybush tea, an indigenous crop. Acta Horticulturae 1007, 889-894.

Bester, C., Tobutt, K., Mansvelt, E., Blomerus, L., Jolly, N., Pieterse, W., Smit, L., Van Schalkwyk, D., 2013. The value and impact of the ARC Infruitec-Nietvoorbij genebanks. Acta Horticulturae 1007, 975-980.

Bester, C., Joubert, M.E., Joubert, E., 2016. A breeding strategy for South African indigenous herbal teas. Acta Horticulturae, In Press.

Bruvo, R., Michiels, N., D'Souza, T., Schulenburg, H., 2004. A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. Molecular Ecology 13, 2101-2106.

Campbell, L.G., Lee, D., Shukla, K., Waite, T.A., Bartsch, D., 2016. An ecological approach to measuring the evolutionary consequences of gene flow from crops to wild or weedy relatives. Applications in Plant Sciences 4, 1-7.

Chalmers, K.J., Jefferies, S.P., Langridge, P., 2001. Comparison of RFLP and AFLP marker systems for assessing genetic diversity, in: Henry, R.J. (Ed.), Plant Genotyping: The DNA Fingerprinting of Plants. CABI, New York, USA, 161-178.

Chapuis, M.P. & Estoup, A., 2007. Microsatellite null alleles and estimation of population differentiation. Molecular Biology and Evolution 24, 621-631.

Clark, L.V. & Jasieniuk, M., 2011. POLYSAT: An R package for polyploid microsatellite analysis. Molecular Ecology Resources 11, 562-566.

Comai, L., 2005. The advantages and disadvantages of being polyploid. Nature 6, 836-846.

De Nysschen, A., Van Wyk, B., Van Heerden, F., Schutte, A.L., 1996. The major phenolic compounds in the leaves of *Cyclopia* species (honeybush tea). Biochemical Systematics and Ecology 24, 243-246.

Stellenbosch University https://scholar.sun.ac.za

De Silva, H., Hall, A., Rikkerink, E., McNeilage, M., Fraser, L., 2005. Estimation of allele frequencies in polyploids under certain patterns of inheritance. Heredity 95, 327-334.

Dufresne, F., Stift, M., Vergilino, R., Mable, B.K., 2014. Recent progress and challenges in population genetics of polyploid organisms: An overview of current state-of-the-art molecular and statistical tools. Molecular Ecology 23, 40-69.

Du Toit, J., Joubert, E., Britz, T.J., 1998. Honeybush tea - A rediscovered indigenous South African herbal tea. Journal of Sustainable Agriculture 12, 67-84.

Du Toit, N., 2005. Molecular phylogenetics of *Cyclopia* Vent. and its position within Podalyrieae (Fabaceae). Unpublished Ph.D. thesis, University of Johannesburg, Johannesburg, 1-9.

Eliášová, A., Trávníček, P., Mandák, B., Münzbergová, Z., 2014. Autotetraploids of *Vicia cracca* show a higher allelic richness in natural populations and a higher seed set after artificial selfing than diploids. Annals of Botany 113, 159-170.

Ford-Lloyd, B., 2001. Genotyping in plant genetic resources, in: Henry, R.J. (Ed.), Plant Genotyping: The DNA Fingerprinting of Plants. CABI, New York, USA, 59-81.

Fowler, C. & Hodgkin, T., 2004. Plant genetic resources for food and agriculture: Assessing global availability. Annual Review of Environment and Resources 29, 143-179.

Gepts, P., 2006. Plant genetic resources conservation and utilization: The accomplishments and future of a societal insurance policy. Crop Science 46, 2278-2292.

Godwin, I., 2009. Plant germplasm collections as sources of useful genes, in: Newbury, H.J. (Ed.), Plant Molecular Breeding. Wiley-Blackwell, Oxford, UK, 280-281.

Goldblatt, P., 1981. Chromosome numbers in legumes II. Annals of the Missouri Botanical Garden 68, 551-557.

Huang, Q., Börner, A., Röder, S., Ganal, W., 2002. Assessing genetic diversity of wheat (*Triticum aestivum* L.) germplasm using microsatellite markers. Theoretical and Applied Genetics 105, 699-707.

Jeridi, M., Perrier, X., Rodier-Goud, M., Ferchichi, A., D’Hont, A., Bakry, F., 2012. Cytogenetic evidence of mixed disomic and polysomic inheritance in an allotetraploid (AABB) *Musa* genotype. Annals of Botany 110, 1593-1606.

Joubert, E., Gelderblom, W.C., Louw, A., De Beer, D., 2008. South African herbal teas: *Aspalathus linearis*, *Cyclopia* spp. and *Athrixia phylicoides* - a review. Journal of Ethnopharmacology 119, 376-412.

Joubert, E., Joubert, M., Bester, C., De Beer, D., De Lange, J., 2011. Honeybush (*Cyclopia* spp.): From local cottage industry to global markets - The catalytic and supporting role of research. South African Journal of Botany 77, 887-907.

Stellenbosch University https://scholar.sun.ac.za

Kalia, R.K., Rai, M.K., Kalia, S., Singh, R., Dhawan, A.K., 2011. Microsatellite markers: An overview of the recent progress in plants. Euphytica 177, 309-334.

Kamara, B.I., Brandt, E.V., Ferreira, D., Joubert, E., 2003. Polyphenols from honeybush tea (*Cyclopia intermedia*). Journal of Agricultural and Food Chemistry 51, 3874-3879.

Kamara, B.I., Brand, D.J., Brandt, E.V., Joubert, E., 2004. Phenolic metabolites from honeybush tea (*Cyclopia subternata*). Journal of Agricultural and Food Chemistry 52, 5391-5395.

Karp, A., Kresovich, S., Bhat, K., Ayad, W., Hodgkin, T., 1997. Molecular tools in plant genetic resources conservation: A guide to the technologies. IPGRI Technical Bulletin No. 2. International Plant Genetic Resources Institute. Rome, Italy, 9-12.

Koen, J., Slabbert, M.M., Bester, C., Bierman, F., 2016. Germination characteristics of dimorphic honeybush (*Cyclopia* spp.) seed. South African Journal of Botany, In Press. Available at: http://dx.doi.org/10.1016/j.sajb.2016.03.006

Kokotkiewicz, A., Luczkiewicz, M., Sowinski, P., Glod, D., Gorynski, K., Bucinski, A., 2012. Isolation and structure elucidation of phenolic compounds from *Cyclopia subternata* Vogel (honeybush) intact plant and in vitro cultures. Food Chemistry 133, 1373-1382.

Kuleung, C., Baenziger, P.S., Dweikat, I., 2004. Transferability of SSR markers among wheat, rye, and triticale. Theoretical and Applied Genetics 108, 1147-1150.

Lee, L.S. & Henry, R.J., 2001. Commercial applications of plant genotyping, in: Henry, R.J. (Ed.), Plant Genotyping: The DNA Fingerprinting of Plants. CABI, New York, USA, 265-274.

Linder, H.P., 2003. The radiation of the Cape flora, southern Africa. Biological Reviews of the Cambridge Philosophical Society 78, 597-638.

Lowe, A., Harris, S., Ashton, P., 2004. Markers and sampling in ecological genetics, in: Ecological Genetics: Design, Analysis and Application. Blackwell, Oxford, UK, 36-44.

Luna, T., Wilkinson, K., Dumroese, R.K., 2009. Seed germination and sowing options, in: Dumroese, R.K., Luna, T., Landis, T.D. (Eds.), Nursery Manual for Native Plants: A Guide for Tribal Nurseries. Agriculture Handbook 730. Washington, D.C.: U.S. Department of Agriculture, Forest Service, 133-151.

Marnewick, J., 2009. Rooibos and honeybush: Recent advances in chemistry, biological activity and pharmacognosy. American Chemistry Society 1021, 277-294.

Masterson, J., 1994. Stomatal size in fossil plants: Evidence for polyploidy in majority of angiosperms. Science 264, 421-424.

McGregor, C.E., Lambert, C.A., Greyling, M.M., Louw, J.H., Warnich, L., 2000. A comparative assessment of DNA fingerprinting techniques (RAPD, ISSR, AFLP and SSR) in tetraploid potato (*Solanum tuberosum* L.) germplasm. Euphytica 113, 135-144.

Stellenbosch University https://scholar.sun.ac.za

McKay, D. & Blumberg, J., 2007. A review of the bioactivity of South African herbal teas: Rooibos (*Aspalathus linearis*) and honeybush (*Cyclopia intermedia*). Phytotherapy Research 16, 1-16.

Meirmans, P.G. & Van Tienderen, P.H., 2013. The effects of inheritance in tetraploids on genetic diversity and population divergence. Heredity 110, 131-137.

Moore, G., 2013. Meiosis in polyploids, in: Chen, Z.J., Birchler, J.A. (Eds.), Polyploid and Hybrid Genomics. Wiley-Blackwell, Oxford, UK, 187-197.

Motsa, M., 2016. Natural phenology, fecundity, genetic variation and seed dormancy of *Cyclopia* species (Honeybush). Unpublished D.Tech. thesis, Tshwane University of Technology, Pretoria, 159-163.

Nybom, H., Weising, K., Rotter, B., 2014. DNA fingerprinting in botany: Past, present, future. Investigative Genetics 5, 1-35.

Ouborg, N., Piquot, Y., Van Groenendael, J., 1999. Population genetics, molecular markers and the study of dispersal in plants. Journal of Ecology 87, 551-568.

Parhiz, H., Roohbakhsh, A., Soltani, F., Rezaee, R., Iranshahi, M., 2015. Antioxidant and anti-inflammatory properties of the citrus flavonoids hesperidin and hesperetin: An updated review of their molecular mechanisms and experimental models. Phytotherapy Research 29, 323-331.

Parisod, C., Holderegger, R., Brochmann, C., 2010. Evolutionary consequences of autopolyploidy. New Phytologist 186, 5-17.

Peakall, R., Gilmore, S., Keys, W., Morgante, M., Rafalski, A., 1998. Cross-species amplification of soybean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legume genera: Implications for the transferability of SSRs in plants. Molecular Biology and Evolution 15, 1275-1287.

Pfeiffer, T., Roschanski, A.M., Pannell, J.R., Korbecka, G., Schnittler, M., 2011. Characterization of microsatellite loci and reliable genotyping in a polyploid plant, *Mercurialis perennis* (Euphorbiaceae). Journal of Heredity 102, 479-488.

Poehlman, J., 1983. Methods of breeding: Cross-pollinated crops, asexually propagated crops, in: Breeding Field Crops. AVI, Connecticut, USA, 140-146.

Rai, M.K., Phulwaria, M., Shekhawat, N.S., 2013. Transferability of simple sequence repeat (SSR) markers developed in guava (*Psidium guajava* L.) to four Myrtaceae species. Molecular Biology Reports 40, 5067-5071.

Rajora, P. & Rahman, H., 2003. Microsatellite DNA and RAPD fingerprinting, identification and genetic relationships of hybrid poplar (*Populus x canadensis*) cultivars. Theoretical and Applied Genetics 106, 470-477.

Rao, N., 2004. Plant genetic resources: Advancing conservation and use through biotechnology. African Journal of Biotechnology 3, 136-145.

Stellenbosch University https://scholar.sun.ac.za

Rao, V.R. & Hodgkin, T., 2002. Genetic diversity and conservation and utilization of plant genetic resources. Plant Cell, Tissue and Organ Culture 68, 1-19.

Rossetto, M., 2001. Sourcing of SSR markers from related plant species, in: Henry, R.J. (Ed.), Plant Genotyping: The DNA Fingerprinting of Plants. CABI, New York, USA, 211-224.

Sampson, J. & Byrne, M., 2012. Genetic diversity and multiple origins of polyploid *Atriplex nummularia* Lindl. (Chenopodiaceae). Biological Journal of the Linnean Society 105, 218-230.

SANBI, 2012. Red List of South African Plants. [ONLINE] Available at: http://redlist.sanbi.org/. [Accessed 8 August 2016].

Sattler, M.C., Carvalho, C.R., Clarindo, W.R., 2016. The polyploidy and its key role in plant breeding. Planta 243, 281-296.

Schippmann, U., Leaman, D., Cunningham, A.B., 2006. A comparison of cultivation and wild collection of medicinal and aromatic plants under sustainability aspects, in: Bogers, R., Craker, L., Lange, D. (Eds.), Medicinal and Aromatic Plants. Springer, Netherlands, 75-95.

Schutte, A.L., 1997. Systematics of the genus *Cyclopia* Vent. (Fabaceae, Podalyrieae). Edinburgh Journal of Botany 54, 125-170.

Selkoe, K.A. &, Toonen, R.J., 2006. Microsatellites for ecologists: A practical guide to using and evaluating microsatellite markers. Ecology Letters 9, 615-629.

Soltis, D.E. & Soltis, P.S., 1993. Molecular data and the dynamic nature of polyploidy. Critical Reviews in Plant Sciences 12, 243-273.

Soltis, D.E. & Soltis, P.S., 1999. Polyploidy: Recurrent formation and genome evolution. Trends in Ecology and Evolution 14, 348-352.

Spooner, D., Treuren, R. Van, De Vicente, M.C., 2005. Molecular markers for genebank management. International Plant Genetic Resources Institute Technical Bulletin 10, 24-90.

Städler, A.T., Loew, M., Streit, B., 1993. Genetic evidence for low outcrossing rates in polyploid freshwater snails (*Ancylus fluviatilis*). Proceedings of the Royal Society of London B: Biological Sciences 251, 207-213.

Wang, J. & Scribner, K.T., 2014. Parentage and sibship inference from markers in polyploids. Molecular Ecology Resources 14, 541-553.

Wang, M., Barkley, N., Jenkins, T., 2009. Microsatellite markers in plants and insects. Part I: Applications of biotechnology. Genes, Genomes and Genomics 3, 1-10.

Werth, C.R., Guttman, S.I., Eshbaugh, W.H., 1985. Recurring origins of allopolyploid species in *Asplenium*. Science 228, 731-733.

Wünsch, A. & Hormaza, J., 2002. Cultivar identification and genetic fingerprinting of temperate fruit tree species using DNA markers. Euphytica 125, 59-67.

Stellenbosch University https://scholar.sun.ac.za

Zamariola, L., Tiang, C.L., De Storme, N., Pawlowski, W., Geelen, D., 2014. Chromosome segregation in plant meiosis. Frontiers in Plant Science 5, 1-19.

Zawedde, B.M., Ghislain, M., Magembe, E., Amaro, G.B., Grumet, R., Hancock, J., 2015. Characterization of the genetic diversity of Uganda's sweet potato (*Ipomoea batatas*) germplasm using microsatellites markers. Genetic Resources and Crop Evolution 62, 501-513.

Stellenbosch University https://scholar.sun.ac.za

Chapter 2: *Cyclopia* Leaf Preservation and DNA Extraction

Abstract

The extraction of high quality DNA is the first crucial step during molecular genetic studies. *Cyclopia* species are rich in phenolic compounds that can interfere with the DNA and render it useless for further downstream analyses, such as PCR amplification. The phenolic content of different species and tissue types vary greatly, therefore the optimisation of a DNA extraction protocol across *Cyclopia* species is of great importance for further molecular work. Choosing an optimal tissue type as starting material, as well as applying various strategies to remove phenolic compounds, should ensure the extraction of high quality DNA. Sampling from remote locations presented an additional problem, as the leaves would turn brown and wilt by the time DNA extraction can be performed. Therefore, different preservation techniques were also tested. Young leaves were found to yield the highest concentrations of DNA and was used throughout the rest of the study. A CTAB/NaCl buffer was used to preserve the leaves for up to two weeks in the field, before DNA was extracted. High quality DNA was obtained and successfully used in PCR amplification and microsatellite development for the commercially important C. *subternata* species.

2.1 Introduction

The South African genus *Cyclopia* consists of several leguminous species, more commonly known as honeybush. Teas brewed from plants of this genus exhibit various medicinal properties (Kamara *et al*., 2003; 2004). In particular, inhabitants from the Cape Fynbos region have recognised honeybush as a cure for ailments such as respiratory irritations and digestive disorders as well as the stimulation of milk production in women. This prompted an interest and deeper investigation as to which compounds are responsible for these health properties (Kamara *et al*., 2003; Marnewick, 2009). The leaves and roots of *Cyclopia* species are rich in phenolic compounds when the plant is subjected to external stresses such as drought, infection or extreme weather conditions. Recently, great emphasis has been placed on the health properties of these compounds when consumed by humans (Joubert *et al*., 2014). De Nysschen *et al*. (1996) found three main phenolic compounds in *Cyclopia,* namely mangiferin, hesperetin and isosakuranetin. These compounds do not occur in any of the other genera of the tribe Podalyrieae, thereby indicating the chemical uniqueness of *Cyclopia*.

Although phenolic compounds are beneficial in terms of human health, they cause problems during molecular studies of plants when DNA extraction procedures are performed (Khanuja *et al*., 1999).

Stellenbosch University https://scholar.sun.ac.za

Oxidation of phenolic compounds occurs spontaneously and the resulting products are known to covalently bind to DNA, resulting in a brown, highly viscous solution that renders the DNA inaccessible to certain restriction enzymes and *Taq* polymerase during polymerase chain reaction (PCR) amplification (Porebski *et al*., 1997; Weising *et al*., 2005; Calderón-Cortés *et al*., 2010; Borse *et al*., 2011).

Cyclopia species are reported to differ widely in terms of phenolic content. *Cyclopia genistoides* contains high levels of mangiferin, while the predominant compound in *C. intermedia* is hesperetin (Joubert *et al*., 2008). Since species vary greatly in their biochemical composition, the optimisation of species-specific DNA extraction protocols may be needed to ensure the extraction of good quality DNA as starting material. This will in turn benefit subsequent analyses such as PCR amplification during molecular marker studies (Weising *et al*., 2005) that can be applied in the honeybush breeding programme. Research also indicates that harvest time affects the phenolic content of *Cyclopia* leaves (Louw *et al*., 2013; Joubert *et al*., 2014). During winter time, phenolics are present in low concentrations therefore this is expected to be the optimal time for DNA extraction. When utilised for its medicinal properties, however, honeybush should be harvested during the summer months when the phenolic compounds are present in greater amounts (Louw *et al*., 2013; Joubert *et al*., 2014).

Various strategies can be applied during the DNA extraction process to counteract or remove phenolic compounds successfully. Strong reducing agents such as β-mercaptoethanol can decrease the browning of DNA preparations by preventing oxidation (Weising *et al*., 2005). Polyvinylpyrrolidone (PVP) forms complex hydrogen bonds with the phenolics and the complex co-precipitates with the cell debris, while high ratios of extraction buffer to plant material will dilute the phenolic compounds (Weising *et al*., 2005; Calderón-Cortés *et al*., 2010). Further precautions include the commencement of DNA extraction immediately after sampling to minimise the browning of the leaves that is caused by the oxidation of phenolic compounds (Joubert *et al*., 2008). Access to laboratory facilities during field work is, however, not always possible, especially when wild populations are sampled in remote areas (Štorchová *et al*., 2000). Several techniques of preservation prior to extraction have been described, including the drying of leaves with silica gel, preservation in CTAB/NaCl buffer solutions and freezing with liquid nitrogen (Bhattacharjee *et al*., 2009). Previous studies on *Cyclopia* extracted DNA from silica dried material using a CTAB extraction method developed by Doyle and Doyle (1987) (Du Toit, 2005; Boatwright *et al*., 2008). The current study will instead focus on DNA extraction from fresh leaves that are preserved in a CTAB/NaCl buffer, as the protocol was optimised specifically using fresh leaves.

Stellenbosch University https://scholar.sun.ac.za

The current study is aimed at several aspects of optimising a DNA extraction protocol for further molecular work such as microsatellite marker development and DNA fingerprinting in the genus *Cyclopia*. Firstly, different tissue types were tested to evaluate which tissue type yields the highest quality and quantity of DNA. Subsequently, to estimate whether the extraction protocol can successfully be applied across all species, DNA extraction was performed on seven different *Cyclopia* species for which accessions were available in the ARC germplasm collection. Lastly, a preservation technique was tested to examine the effect on DNA quality and quantity during sampling of wild populations from remote locations.

2.2 Optimisation of DNA Extraction Protocol and Testing of Different Tissue Types

2.2.1 Materials and methods

2.2.1.1 Sampling

Three different tissue types were tested namely, young leaves, young stems and old leaves (Figure 2.1). Two accessions were sampled from each of the two species growing in the ARC field gene bank, namely *C. genistoides* (accessions GG53 and GT1) and *C. subternata* (accessions SHL3 and SKB3), with four repetitions per tissue type, giving a total of 48 DNA extractions. Gloves were worn to pick the young and old leaves while the stems were cut using pruning shears and all tissues were kept on ice in 50ml Falcon tubes. Once back in the laboratory, processing of the stems entailed cutting them in half and slicing slivers from the phloem using a sterile scalpel and blade. The use of a fresh scalpel blade for each accession minimised any possible contamination and 100% ethanol was used to wash the blade between repetitions. A fixed amount of leaf and stem tissue (0.1g) were weighed in 2ml Eppendorf tubes.

2.2.1.2 DNA extraction

A DNA extraction protocol for medicinally important angiosperms described by Borse *et al*. (2011) was optimised for use in *Cyclopia*. A stainless steel bead was added to each Eppendorf tube and the leaf and stem material was pulsed for 1min at 30Hz using the TissueLyser (Qiagen). A high CTAB buffer to leaf ratio successfully diluted the phenolic compounds; therefore, 500µl of a pre-warmed (60°C) 2% (w/v) CTAB extraction buffer (100mM Tris-hydroxymethyl-aminomethane hydrochloride [Tris-HCl], pH 8.0; 20mM ethylenediaminetetraacetic acid [EDTA], pH 8.0; 1.4M sodium chloride [NaCl]; 2% (w/v) hexadecyltrimethylammonium bromide [CTAB]; 2% (m/v) polyvinylpyrrolidone [PVP]) was added to the ground tissue (Weising *et al*., 2005). This was thoroughly mixed and ground in the TissueLyser for another minute at 30Hz. After the addition of

Stellenbosch University https://scholar.sun.ac.za

2µl β-mercaptoethanol and 2µl protein kinase K (10mg/ml), the tubes were incubated at 60°C overnight. Chloroform:isoamylalcohol (24:1) was added in equal volumes and the solution was centrifuged at a relative centrifugal force of 16 100 for 10mins at 4°C. A constant volume of 100-200µl supernatant was removed to ensure that no cross-contamination with chloroform or other waste in the organic phase occurred, thereby ensuring a high DNA yield. Several protocols repeat the chloroform-isoamyl extraction step to ensure a clean aqueous phase, but it was often found that DNA recovery was decreased and thus only one step was performed. DNA was precipitated with 2/3 volume ice-cold isopropanol for 1h at -80°C and spun down at a relative centrifugal force (RCF) of 16 100 for 20mins at room temperature before the supernatant was discarded. The DNA pellet was cleaned using 200µl of 70% (v/v) ethanol after which the tubes were spun down at a RCF of 16 100 for 10mins at room temperature before the ethanol was decanted. This step was repeated and the pellet was air dried before re-suspension in 30µl ddH_2O. DNA concentrations were measured using the NanoDrop ND-1000 spectrophotometer (Thermo-Fisher).

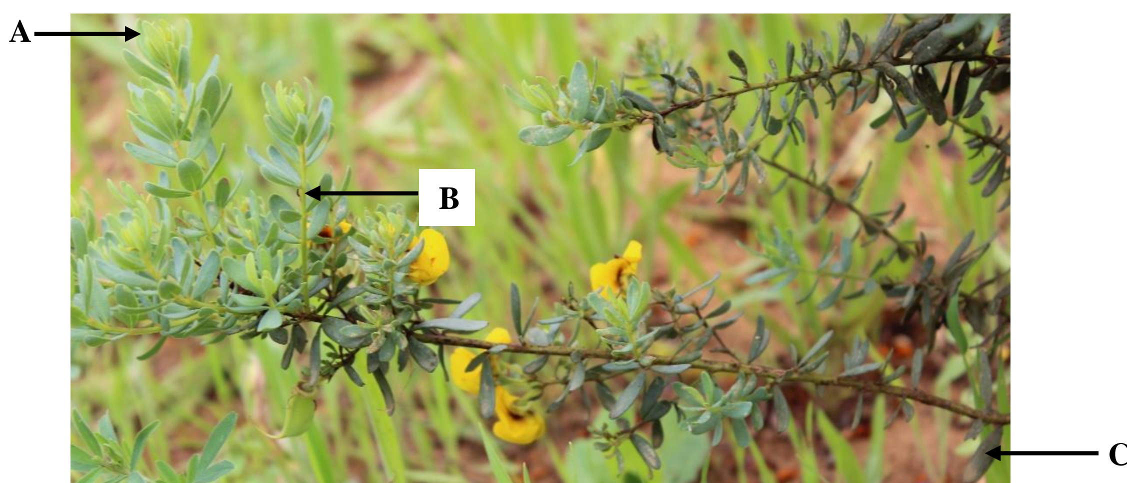


Figure 2.1. (A) Young leaves, (B) young stems and (C) old leaves in *C. subternata*. Young leaves are lighter in colour, while old leaves are dark green to grey.

2.2.1.3 Data analyses

Three variables were chosen to evaluate DNA quality and quantity: the DNA concentration (ng/µl) as well as the A_{260}:A_{280} and A_{260}:A_{230} absorbance ratios. To determine which plant tissue yielded optimal DNA, the data were first tested for homoscedasticity and normal distributions of residuals by means of the SAS® Enterprise Guide 7.1 software (SAS Institute, North Carolina, USA). Residuals were calculated using a linear model prediction; however, none of the data sets conformed to both a normal distribution and the null hypothesis that assumes equal variances.

For further analysis, the skewed data sets were transformed with appropriate functions to normalise the data (see Shapiro-Wilks P-values in Table 2.1) and to ensure homoscedasticity using Levene's

test for homogeneity (Table 2.1 and Table A5 in Appendix). The A_{260}:A_{230} values were transformed with a square root function, the A_{260}:A_{280} values with an x^4 function and the DNA concentration (ng/µl) using a log function. To test the interaction between the two different species and the type of tissue used (species by source interaction), a two-way factorial design (linear model) was selected. To investigate the significant interaction for each of the three dependent variables, one-way Analyses of Variance (ANOVAs) were conducted with Bonferroni t-tests in the SAS Enterprise Guide 7.1 software. Bonferroni mean values (x) for each of the variables were transformed back to their original values before data was interpreted: DNA concentration by using the function 10^x; A_{260}:A_{280} ratios by using the function $\sqrt[4]{x}$ and A_{260}:A_{230} by using the function x^2.

Table 2.1. Distribution of residuals and homoscedasticity of *Cyclopia* DNA quality data before and after transformation with appropriate functions for three dependent variables.

Before transformation						
	Distribution of residuals			Homoscedasticity		
Dependent variable	**P-value**	**Significant**	**Normal**	**P-value**	**Significant**	**Homoscedastic**
[DNA] in ng/µl	0.0005	No	No	0.0010	No	No
A_{260}:A_{280}	0.0673	Yes	Yes	< 0.0001	No	No
A_{260}:A_{230}	0.1700	Yes	Yes	0.0321	No	No
After transformation						
	Distribution of residuals			**Homoscedasticity**		
Transformation	**P-value**	**Significant**	**Normal**	**P-value**	**Significant**	**Homoscedastic**
Log (ng/µl)	0.9249	Yes	Yes	0.2067	Yes	Yes
$(A_{260}:A_{280})^4$	0.5679	Yes	Yes	0.0800	Yes	Yes
$\sqrt{}$ (A_{260}:A_{230})	0.1201	Yes	Yes	0.0502	Yes	Yes

Table 2.2. Two-way factorial design (linear model) for three dependent variables of *Cyclopia* DNA quality data after transformation.

Transformation	R-square coefficient	Linear model fit	Interaction (species and source)
Log (ng/µl)	0.907	Good	P < 0.0001
$(A_{260}:A_{280})^4$	0.808	Good	P < 0.0001
$\sqrt{}$ (A_{260}:A_{230})	0.836	Good	P < 0.0001

2.2.2 Results and discussion

For the testing of different tissue types, the two-way factorial design was an appropriate model for the three data sets (see relevant R-square values in Table 2.2) and significant interaction was found for each of the two species, *C. genistoides* and *C. subternata*, and the tissue types used as DNA source (P < 0.0001, Table 2.2).

Stellenbosch University https://scholar.sun.ac.za

Bonferroni t-tests revealed three groupings for the DNA concentration data set (Table 2.3). Within species, results indicated that the *C. subternata* young leaves yielded the highest DNA concentration, with an average of 1007.04ng/µl, while old leaves and stems yielded significantly lower values (60.14ng/µl and 216.86ng/µl, respectively). Young leaves of *C. genistoides* yielded significantly higher concentrations of DNA than either the old leaves or the stems with an average of 740.24ng/µl. Between species, the old leaves of *C. subternata* yielded significantly less DNA (60.14ng/µl) than the *C. genistoides* old leaves (130.15ng/µl), but no significant differences were found between the stems and young leaves (Table 2.3).

Table 2.3. Bonferroni t-test groupings of the range and mean DNA quantity in ng/µl for the interaction between various tissue types as starting material for *C. genistoides* and *C. subternata* species.

Bonferroni Grouping	Range of DNA quantity (ng/µl)	Mean DNA quantity (ng/µl)	N	Interaction
A	607.90-1647.50	1007.04	8	*C. subternata* young leaves
A	587.80-899.60	740.24	8	*C. genistoides* young leaves
B	103.00-529.00	216.86	8	*C. subternata* young stems
B	128.50-346.90	200.24	8	*C. genistoides* young stems
B	102.50-194.60	130.15	8	*C. genistoides* old leaves
C	35.00-95.10	60.14	8	*C. subternata* old leaves

Table 2.4. Bonferroni t-test groupings of the range and mean A_{260}:A_{280} for the interaction between various tissue types as starting material for *C. genistoides* and *C. subternata* species.

Bonferroni Grouping	Range of A_{260}:A_{280}	A_{260}:A_{280}	N	Interaction
A	1.99-2.06	2.02	8	*C. genistoides* young leaves
A	1.95-2.08	2.01	8	*C. subternata* young leaves
B	1.79-2.05	1.93	8	*C. subternata* young stems
B	1.81-1.97	1.89	8	*C. genistoides* old leaves
B	1.76-2.02	1.89	8	*C. genistoides* young stems
C	1.25-1.76	1.53	8	*C. subternata* old leaves

DNA absorbs light at 260nm, therefore the A_{260}:A_{280} absorbance ratio gives an indication of the DNA purity of a sample. The A_{260}:A_{280} values should ideally range from 1.8-2, indicating pure DNA (Borse *et al*., 2011; Moreira and Oliveira, 2011). An A_{260}:A_{280} value higher than 1.8 indicates the presence of RNA, while a value below 1.8 indicates protein contamination (Bhattacharjee *et al*., 2009). The young leaves and stems of both species yielded ratios above 1.8 which fall within the range for pure DNA (Table 2.4). Similarly, the old leaves from *C. genistoides* had an average ratio of 1.89, while the ratio for *C. subternata* old leaves was well below 1.8 (with a mean of 1.53). This indicates protein contamination and low DNA quality.

Stellenbosch University https://scholar.sun.ac.za

Young leaves from both species had significant ratios above 1.9, indicating no phenolic compound contamination (Table 2.5). Phenolic compounds and other contaminants absorb light at 230nm and an A_{260}:A_{230} ratio well below 2 is used to indicate contamination. The *C. genistoides* old leaves had no phenolic compound contamination (average ratio of 1.88) while the old leaves from *C. subternata* had significantly low A_{260}:A_{230} ratios. The quality of the stems averaged 1.31 for *C. genistoides* and 1.22 for *C. subternata*. A similar study by Moreira and Oliveira (2011) tested the effect of leaf age on the quality of DNA extracted from *Dimorphandra mollis* (Fabaceae). Isolation from old leaves was not successful and this was attributed to the presence of phenolic compounds. DNA from young leaves, however, yielded high qualities and quantities of DNA that successfully amplified during PCR amplification.

Table 2.5. Bonferroni t-test groupings of the range and mean A_{260}:A_{230} for the interaction between various tissue types as starting material for *C. genistoides* and *C. subternata* species.

Bonferroni Grouping	Range of A_{260}:A_{230}	A_{260}:A_{230}	N	Interaction
A	2.03-2.58	2.28	8	*C. genistoides* young leaves
A	1.74-2.22	1.92	8	*C. subternata* young leaves
A	1.50-2.42	1.88	8	*C. genistoides* old leaves
B	0.96-1.79	1.31	8	*C. genistoides* young stems
B	0.91-1.69	1.22	8	*C. subternata* young stems
C	0.51-1.06	0.70	8	*C. subternata* old leaves

Young leaves consistently yielded higher concentrations of good quality DNA in *Cyclopia*. Several factors affect the amount of DNA accessible during isolation procedures. Phenolic compounds covalently bind to DNA, lowering the yield and quality which will influence further downstream analyses such as PCR amplification (Calderón-Cortés *et al*., 2010). Old leaves accumulate large amounts of phenolics during maturation for chemical defence against herbivores while young leaves dedicate their nutrients to early growth and development (Khanuja *et al*., 1999; Moreira and Oliveira, 2011). Mature leaves of *Cyclopia* often produced less DNA during isolation procedures in addition to poor quality DNA with A_{260}:A_{230} ratios that fell well below the required range.

Old leaves become tougher as they mature, while young leaves have a softer texture and can easily be ground to release the DNA (Lowman and Box, 1983). A study conducted by Chen and Huang (2013) reported on the difference in cuticle thickness and epidermis of young and mature leaves in 76 woody species. Young leaves consistently displayed thinner cuticle layers, whereas mature leaves had a thicker cuticle that was difficult to disrupt.

Stellenbosch University https://scholar.sun.ac.za

Although *Cyclopia* stems yielded high quality and quantity DNA, they are extremely thin and only small amounts of phloem can be scraped per stem sampled. Acquiring an adequate amount of starting material is a laborious process, whereas a greater number of leaves can be picked in a short amount of time.

In the current study, young *Cyclopia* leaves consistently produced a combination of the highest concentration and quality of DNA as revealed by both the A_{260}:A_{280} and A_{260}:A_{230} ratios. Consequently, young *Cyclopia* leaves were chosen as starting material for this study when DNA extraction commenced.

2.3 Testing of Optimised DNA Extraction Protocol on Different *Cyclopia* Species

2.3.1 Materials and methods

2.3.1.1 Sampling and DNA extraction

All the accessions were sampled at the ARC Nietvoorbij and Elsenburg Research Farms near Stellenbosch, South Africa. Table 2.6 summarise the number of accessions that were sampled for each of the different species, and Table A3 (Appendix) provide more detail on the origin of the accessions. Seven species were obtained from the ARC germplasm collection namely *C. genistoides, C. intermedia, C. longifolia*, *C. maculata, C. pubescens*, *C. sessiliflora* and *C. subternata* (Figure 2.2).

Table 2.6. The seven *Cyclopia* species and their number of accessions sampled from the ARC gene bank with different leaf sizes and structure.

Species	No. accessions sampled	Leaf size	Leaf blade morphology
C. genistoides	15	Small	Needle-like
C. intermedia	4	Large	Flat
C. longifolia	30	Large	Flat
C. maculata	10	Small	Needle-like
C. pubescens	10	Small	Needle-like
C. sessiliflora	10	Large	Flat
C. subternata	22	Large	Flat

Stellenbosch University https://scholar.sun.ac.za

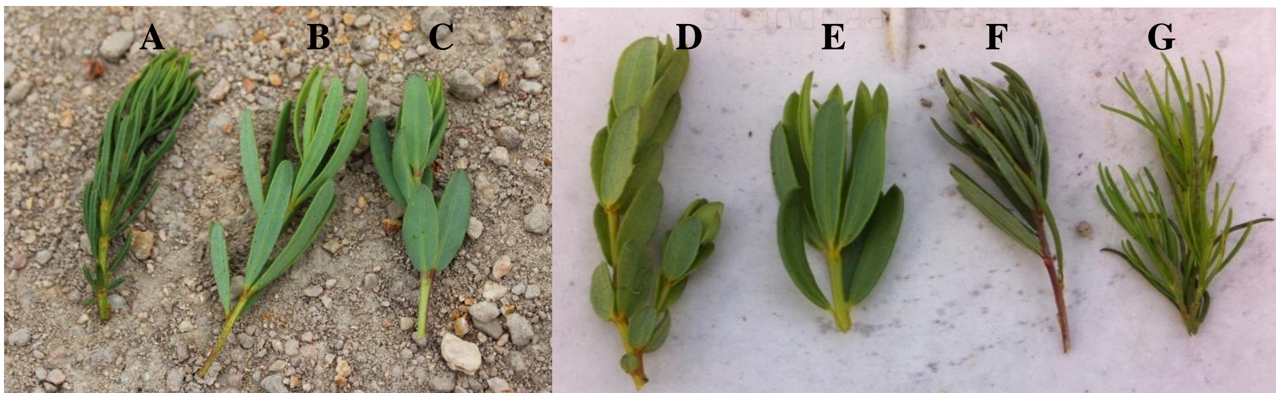


Figure 2.2. Leaves of the seven *Cyclopia* species included in this study showing variation in size and morphology. (A) *C. genistoides*, (B) *C. longifolia*, (C) *C. subternata*, (D) *C. sessiliflora*, (E) *C. intermedia*, (F) *C. pubescens* and (G) *C. maculata*.

Gloves were worn while picking young, healthy, green leaves by hand and leaves were subsequently deposited in 2ml Eppendorf tubes. These were continuously kept on ice during transportation to the laboratory facilities. Approximately 0.05-0.15g of leaf material per plant was used as starting material and DNA was extracted as described in section 2.2.1.2. The ten *C. subternata* accessions that had the highest quantity and quality of DNA were selected for microsatellite development (Chapter 3).

2.3.1.2 Data analysis

To investigate the effect of species on DNA concentration, three variables were chosen: the DNA concentration (ng/µl) as well as the A_{260}:A_{280} and A_{260}:A_{230} absorbance ratios. The data sets for the three variables were first tested for normality using SAS Enterprise Guide 7.1 software. A nonparametric one-way ANOVA was conducted in XLSTAT (Addinsoft, New York, USA) after data were found to be skewed (Shapiro-Wilk test for normality, $P < 0.0001$). Significant interaction between the different species, DNA yield and absorption values was tested using a Kruskal-Wallis test. The average DNA concentrations, A_{260}:A_{280} and A_{260}:A_{230} ratios were calculated in Microsoft Excel (Microsoft, Redmond, Washington) and a column chart was constructed to visualise the data.

2.3.2 Results and discussion

The average DNA quantity ranged from 205ng/µl (*C. genistoides*) to 727ng/µl (*C. sessiliflora*). Even though it was significant, these results could be influenced by the slightly different amount of tissue used as starting material and are therefore not directly comparable. The DNA quality is, however, independent of starting material and significant differences were found between the seven *Cyclopia* species for the A_{260}:A_{280} and A_{260}:A_{230} ratios tested using the Kruskal-Wallis test ($P < 0.0001$, data not shown).

Different species that belong to the same genus can exhibit variation in their biochemical composition, thereby influencing the DNA yield obtained (Khanuja *et al*., 1999). *Cyclopia* species are known to have different phenolic compositions which could, in turn, influence the A_{260}:A_{230} absorbance values (Joubert *et al*., 2014). All of the average values were above 1.8, except those of *C. maculata* and *C. pubescens* which were approximately 1.4. It should, however, be noted that DNA extraction was performed in different batches, thereby introducing experimental variations that could have influenced the results.

Leaf size and structure are variable among the seven *Cyclopia* species (Figure 2.2 and Table 2.4). Overall, higher concentrations of DNA were obtained from species that had large, flat, leaves (Figure 2.3) as opposed to species that had small, needle-like leaves, with the exception of the broad leaves of *C. subternata* that had similar DNA amounts as the small-leaved species. A study by Chung *et al*. (1998) found a positive correlation between leaf size and DNA content in *Glycine max*, thereby possibly explaining the general trend found in *Cyclopia* where a larger leaf morphology yielded more DNA.

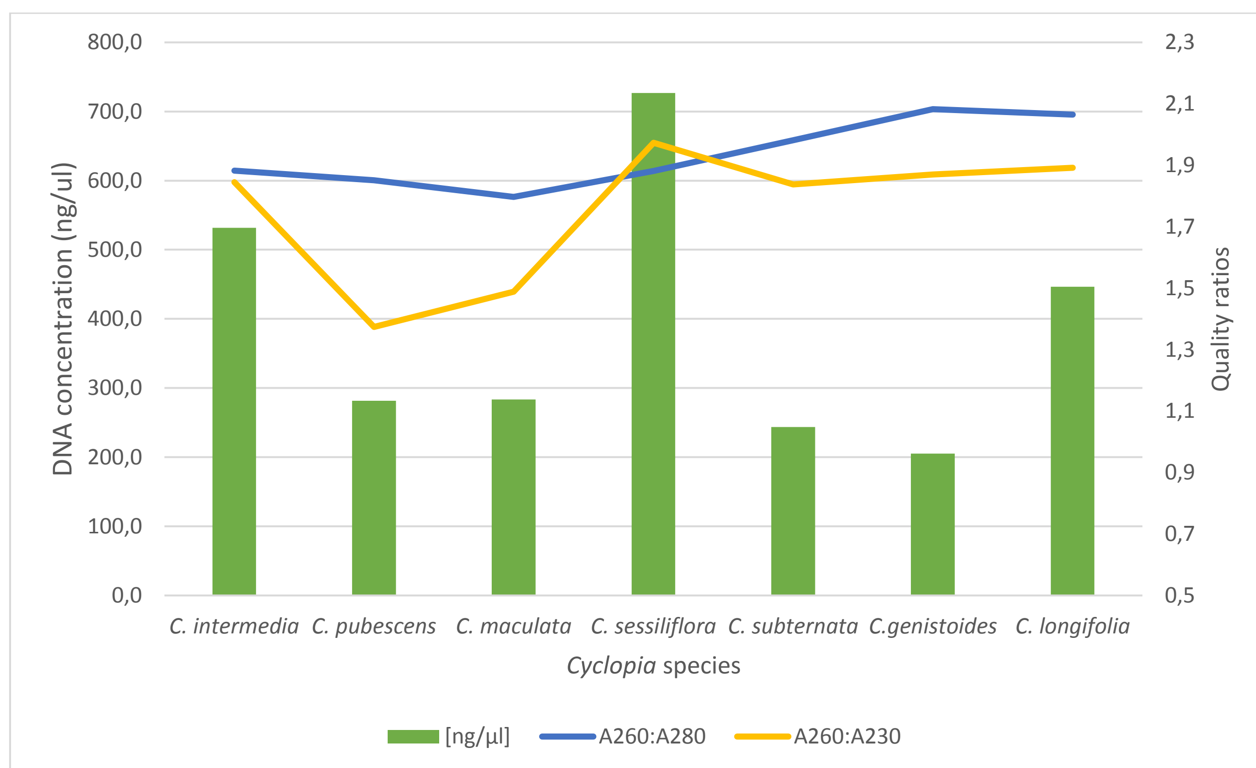


Figure 2.3. DNA concentration ranged from 205-727ng/µl for the seven *Cyclopia* species sampled from the ARC field gene bank for further molecular studies. A_{260}:A_{280} ratios ranged from 1.8 to 2.1 and the A_{260}:A_{230} ratios ranged from as low as 1.4, to a maximum of 2.

Stellenbosch University https://scholar.sun.ac.za

This application of one standard isolation protocol to different species within a genus is often not possible (Weising *et al*., 2005). Even though the DNA quantity and quality differed among the seven species, the DNA was still suitable for use in PCR amplification, indicating that the protocol optimised in *C. subternata* could successfully be applied for all seven *Cyclopia* species.

DNA was extracted from the young leaves of 10 *C. subternata* accessions for application in microsatellite marker development. Glenn and Schable (2005) advised that 2-3µg good quality DNA is sufficient for library preparation. The DNA concentration per sample ranged from 200-650ng/µl with 30µl total volume extracted, thereby giving a total DNA quantity per sample that ranged from 6-19µg. Likewise, the DNA quality ratios were within the required range as indicated in Figure 2.4.

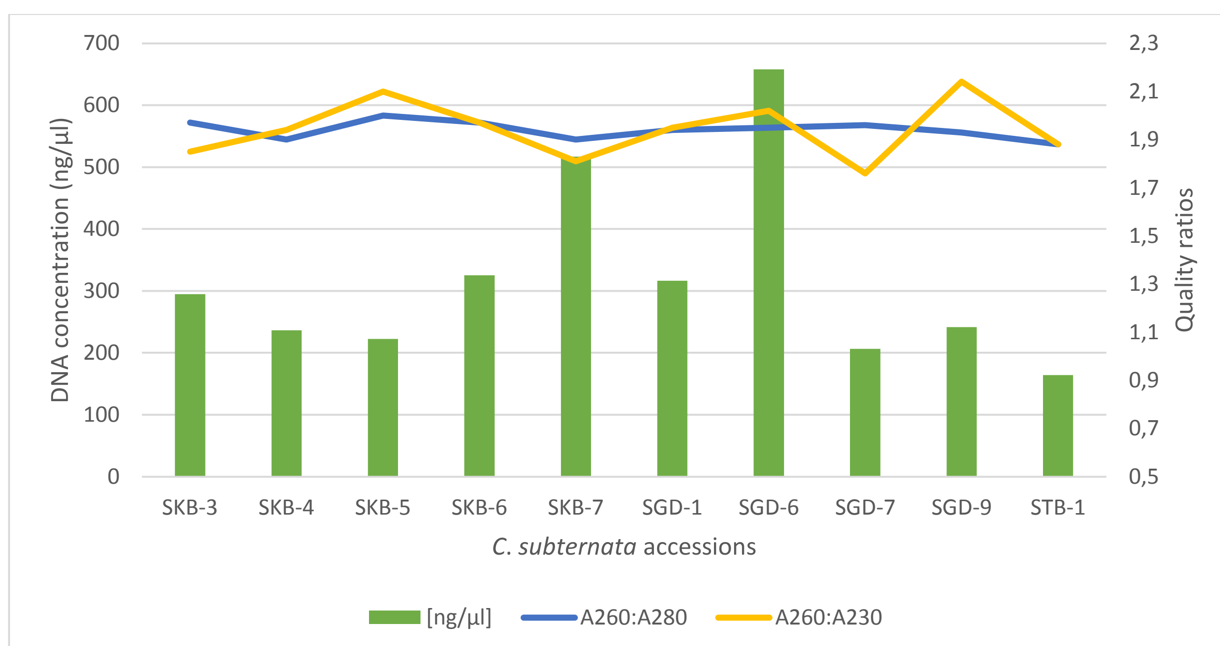


Figure 2.4. DNA concentration for the ten accessions of *C. subternata* that were used to develop the microsatellite markers ranged from 200-650ng/µl. A_{260}:A_{280} ratios ranged from 1.88 to 2.0 and the A_{260}:A_{230} ratios was also well within the required range (1.76-2.14).

2.4 Testing of Preservation Buffers

2.4.1 Materials and methods

2.4.1.1 Sampling and DNA extraction

Test extractions were first performed on two repetitions of the *C. subternata* accession SHL2 growing at the ARC Nietvoorbij Research Farm. The maximum duration of field sampling of wild populations was estimated to be two weeks and therefore storage was not tested beyond this point. Two types of

Stellenbosch University https://scholar.sun.ac.za

buffer were investigated: a 2% (w/v) CTAB extraction buffer (buffer 1) and a CTAB/NaCl buffer (buffer 2). Buffer 1 was prepared as described previously (section 2.2.1.2). Buffer 2 was prepared by dissolving 60g CTAB (hexadecyltrimethylammonium bromide) and 700g NaCl (sodium chloride) in 2000ml of distilled water according to a protocol by Rogstad (1992). All samples were stored at 4°C. DNA extraction, as previously described in Section 2.2.1.2, was performed on the test samples after storage for 2 days, 1 week and 2 weeks in the respective buffers.

2.4.1.2 Data analyses

The same three dependent variables were used to determine optimal DNA yield and quality: the DNA concentration (ng/µl) as well as the A_{260}:A_{280} and A_{260}:A_{230} absorbance ratios. Data were tested for homoscedasticity using Bartlett's test and a Shapiro-Wilk test for normality of residuals was performed with the SAS Enterprise Guide 7.1 software. The DNA concentration data set was normalised by appropriately transforming the data with a log function. To test for any interaction between the type of buffer and time of preservation, a two-way factorial design (linear model) was chosen.

2.4.2 Results and discussion

Visual observations indicated that both the 2% CTAB buffer and the leaves turned dark brown by week 2, whereas in the CTAB/NaCl buffer the leaves remained green and fresh (Figure 2.5). After tissue lysis, the extraction solution containing the 2% CTAB buffer leaves was dark brown, indicating contamination by phenolic compounds (Moreira and Oliveira, 2011) (Figure 2.5). In contrast, the solution resulting from extracted leaves stored in the CTAB/NaCl buffer remained light green.

There were no significant interactions observed between the two types of buffer and the time of preservation for any of the three dependent variables ($P > 0.05$) (Figure 2.6). It should be noted that the sample size was however small ($n = 4$) and this could influence the conclusion reached.

Several other studies reported successful use of the CTAB/NaCl preservation method on various species including *Taraxacum* sect. *Ruderalia* (Asteraceae), *Myosotis palustris* L. (Boraginaceae), *Rubus nessensis* (Rosaceae), and *Melampyrum sylvaticum* L. (Scrophulariaceae) (Štorchová *et al.*, 2000). In the current study, the CTAB/NaCl buffer (buffer 2) was preferred as the preservation medium since it consists of fewer reagents and visual inspection showed improved preservation of the leaves.

Stellenbosch University https://scholar.sun.ac.za

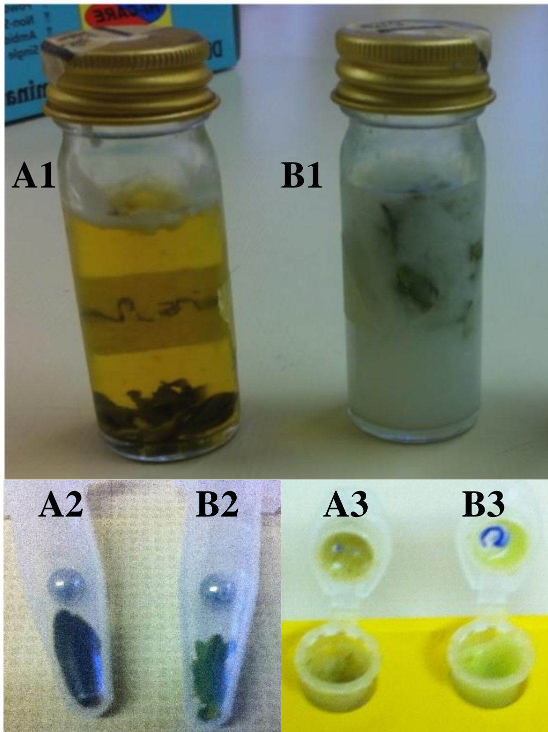


Figure 2.5. Visual comparison between the two buffers used to preserve the *Cyclopia* leaves before DNA was extracted. All figures denoted with A relate to the CTAB buffer and all figure denoted with B to the CTAB/NaCl buffer. A1) the buffer, (A2) leaves and (A3) DNA solution of the 2% CTAB buffer turned brown after 2 weeks. This indicates that phenolic compounds seeped from the leaves and contaminated the DNA during extraction. B1) the buffer, (B2) leaves and (B3) DNA solution of the CTAB/NaCl buffer stayed light green after 2 weeks, indicating less phenolic compounds seeped from the leaves and therefore better preservation.

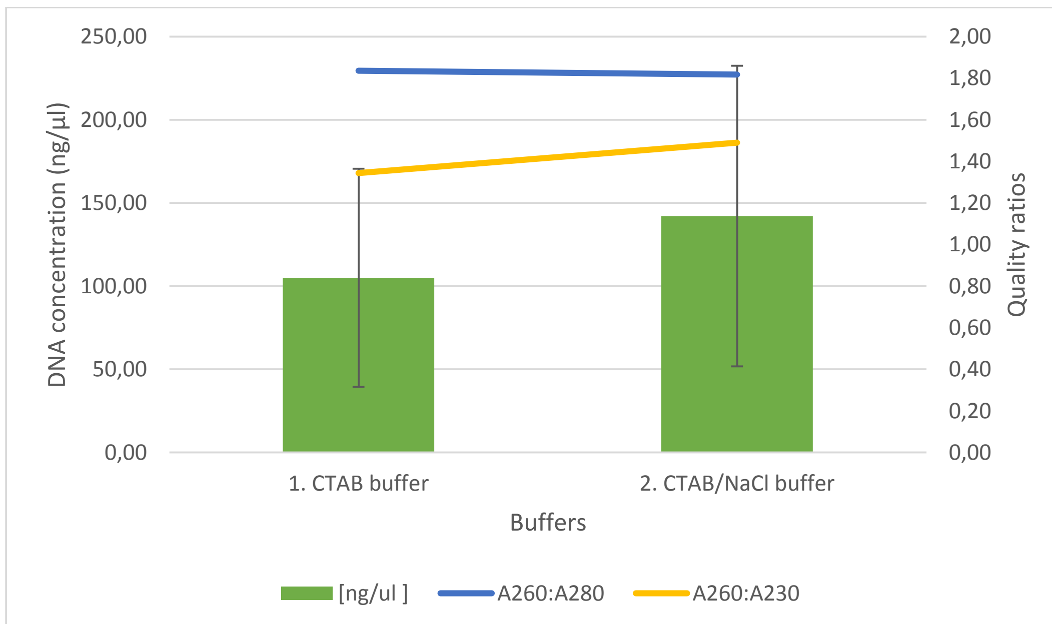


Figure 2.6. The average DNA quantity (in ng/µl) was plotted on the vertical axis and the quality ratios were plotted on a secondary vertical axis to compare how well *C. subternata* samples were preserved in two types of CTAB buffer. Data for day 1-14 was pooled and no significant differences were observed. Error bars indicates the standard deviation of DNA concentration.

Stellenbosch University https://scholar.sun.ac.za

2.5 Conclusion

This study successfully tested three tissue types of *Cyclopia* as potential starting material for DNA extraction. Young leaves were chosen as they produced high DNA yields and the tissue was easy to disrupt owing to its soft texture. The DNA extraction protocol was successfully optimised and applied across seven *Cyclopia* species to isolate high quality and quantity of DNA. During sampling in remote locations, a CTAB/NaCl preservation method was utilised where less browning of leaf material was observed than when the 2% CTAB buffer was used. The leaves were successfully stored for two weeks before DNA extraction was performed. Future work includes repeating the CTAB/NaCl buffer experiments with a larger sample size to investigate the interaction between the type of buffer and amount of storage time. Test extractions were performed after storage for only up to two weeks in this study, but further tests can include a maximum amount of time that the leaves can be preserved to allow for extended sampling periods during field trips. DNA obtained from this study will be applied in Chapter 3 and 4 for molecular studies on *Cyclopia* species.

References

Bhattacharjee, R., Ferguson, M., Gedil, M., Dumet, D., 2009. Field collection, preservation and large scale DNA extraction procedures for cassava (*Manihot esculenta* Crantz.). African Journal of Biotechnology 8, 3424-3430.

Boatwright, J., Savolainen, V., Van Wyk, B., Schutte-Vlok, A., Forest, F., Van Der Bank, M., 2008. Systematic position of the anomalous genus *Cadia* and the phylogeny of the tribe Podalyrieae (Fabaceae). Systematic Botany 33, 133-147.

Borse, T., Joshi, P., Chaphalkar, S., 2011. Biochemical role of ascorbic acid during the extraction of nucleic acids in polyphenol rich medicinal plant tissues. Journal of Plant Molecular Biology and Biotechnology 2, 1-7.

Calderón-Cortés, N., Quesada, M., Cano-Camacho, H., Zavala-Páramo, G., 2010. A simple and rapid method for DNA isolation from xylophagous insects. International Journal of Molecular Sciences 11, 5056-5064.

Chen, Y. & Huang, S., 2013. Red young leaves have less mechanical defence than green young leaves. Oikos 122, 1035-1041.

Chung, J., Lee, J.H., Arumuganathan, K., Graef, G.L., Specht, J.E., 1998. Relationships between nuclear DNA content and seed and leaf size in soybean. Theoretical and Applied Genetics 96, 1064-1068.

Stellenbosch University https://scholar.sun.ac.za

De Nysschen, A., Van Wyk, B., Van Heerden, F., Schutte, A.L., 1996. The major phenolic compounds in the leaves of *Cyclopia* species (honeybush tea). Biochemical Systematics and Ecology 24, 243-246.

Doyle J.J. & Doyle J.L., 1987. A rapid isolation procedure for small amounts of leaf tissue. Phytochemical Bulletin 19, 11-15.

Du Toit, N., 2005. Molecular phylogenetics of *Cyclopia* Vent. and its position within Podalyrieae (Fabaceae). Unpublished Ph.D. thesis, University of Johannesburg, Johannesburg, 1-9.

Glenn, T. & Schable, N., 2005. Isolating microsatellite DNA loci. Methods in Enzymology 395, 202-222.

Joubert, E., Gelderblom, W., Louw, A., De Beer, D., 2008. South African herbal teas: *Aspalathus linearis*, *Cyclopia* spp. and *Athrixia phylicoides* - a review. Journal of Ethnopharmacology 119, 376-412.

Joubert, E., De Beer, D., Hernández, I., Munné-Bosch, S., 2014. Accumulation of mangiferin, isomangiferin, iriflophenone-3-C-β-glucoside and hesperidin in honeybush leaves (*Cyclopia genistoides* Vent.) in response to harvest time, harvest interval and seed source. Industrial Crops and Products 56, 74-82.

Kamara, B.I., Brandt, E.V., Ferreira, D., Joubert, E., 2003. Polyphenols from honeybush tea (*Cyclopia intermedia*). Journal of Agricultural and Food Chemistry 51, 3874-3879.

Kamara, B.I., Brand, D.J., Brandt, E.V., Joubert, E., 2004. Phenolic metabolites from honeybush tea (*Cyclopia subternata*). Journal of Agricultural and Food Chemistry 52, 5391-5395.

Khanuja, S., Shasany, A., Darokar, M., Kumar, S., 1999. Rapid isolation of DNA from dry and fresh samples of plants producing large amounts of secondary metabolites and essential oils. Plant Molecular Biology Reporter 17, 1-7.

Louw, A., Joubert, E., Visser, K., 2013. Phytoestrogenic potential of *Cyclopia* extracts and polyphenols. Planta Medica 79, 580-590.

Lowman, M.D. & Box, J.D., 1983. Variation in leaf toughness and phenolic content among five species of Australian rain forest trees. Australian Journal of Ecology 8, 17-25.

Marnewick, J., 2009. Rooibos and honeybush: Recent advances in chemistry, biological activity and pharmacognosy. American Chemical Society (ACS) Symposium Series 1021, 277-294.

Moreira, P. & Oliveira, D., 2011. Leaf age affects the quality of DNA extracted from *Dimorphandra mollis* (Fabaceae), a tropical tree species from the Cerrado region of Brazil. Genetics and Molecular Research 10, 353-358.

Porebski, S., Bailey, L., Baum, B., 1997. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant Molecular Biology Reporter 15, 8-15.

Stellenbosch University https://scholar.sun.ac.za

Rogstad, S., 1992. Saturated NaCl-CTAB Solution as a means of field preservation of leaves for DNA analyses. Taxon 41, 701-708.

Štorchová, H., Hrdličková, R., Chrtek, J., Tetera, M., Fitze, D., Fehrer, J., 2000. An improved method of DNA isolation from plants collected in the field and conserved in saturated NaCl/CTAB solution. Taxon 49, 79-84.

Weising, K., Nybom, H., Wolff, K., Kahl, G., 2005. Methodology, in: DNA Fingerprinting in Plants: Principles, Methods and Applications. CRC, Florida, USA, 170-205.

Stellenbosch University https://scholar.sun.ac.za

Chapter 3: Microsatellite Marker Development, Cross-Species Amplification and DNA Fingerprinting

Abstract

Although *Cyclopia* species are recognised for their commercial importance, very few molecular studies have been conducted on them. This study successfully developed microsatellite markers for *Cyclopia subternata*, one of the commercially important species for the production of honeybush tea, thereby laying the foundation for future molecular studies. Cross-amplification to six other *Cyclopia* species was achieved with a success rate of 81-88%. A panel comprised of six microsatellite markers was developed for DNA fingerprinting of the accessions that are currently in the ARC field gene bank. This will aid in various stages of characterising and managing the *Cyclopia* plant genetic resources, such as keeping track of the accessions in the gene bank and clones chosen for commercialisation. A thorough understanding of genetic variation in the wild will, in turn, allow for the conservation of genetic material for future breeding programmes.

3.1 Introduction

Cyclopia species are utilised for the production of honeybush tea, a traditional South African herbal drink. Several species of this genus are grown at the Agricultural Research Council (ARC) where the Honeybush Breeding and Selection Programme aims to develop cultivars with improved biomass yield and tea quality (Joubert *et al*., 2011; Bester *et al*., 2013). Many of the species, such as *C. genistoides*, *C. intermedia*, *C. longifolia*, *C. maculata*, *C. sessiliflora* and *C. subternata*, are of commercial importance and accessions of each species are maintained in an ARC field gene bank (Bester *et al*., 2016). Field gene banks entail living collections of plant accessions, collected from populations in their natural habitat (Rao, 2004). *Cyclopia pubescens* is also grown at the ARC, not for commercial reasons, but rather to conserve the species as it is listed as critically endangered on the South African National Biodiversity Institute (SANBI) Red List (SANBI, 2012). Currently, there is a need for the development of marker technologies that will allow for fingerprinting of the accessions in the field gene bank. DNA fingerprinting entails the genetic identification of accessions in the form of a unique banding pattern. The comparison of fingerprints is useful for two purposes: to evaluate the genetic diversity in the field gene bank and to estimate the usefulness of the markers in distinguishing individual accessions (Laurentin, 2009). Once the gene bank is fingerprinted, the genetic data can be used in the management of the accessions to detect incorrect labelling and identify duplicates (Fowler and Hodgkin, 2004; Rao, 2004). Sampling strategies can be adapted to include a

Stellenbosch University https://scholar.sun.ac.za

broader range of wild populations that properly represent the genetic diversity of the species (Rao, 2004; Gepts, 2006).

Microsatellite markers (also known as simple sequence repeats or SSRs) consist of short tandem repeat motifs, usually 1-6 base pairs in length, and occur throughout the eukaryotic genome (Holton, 2001). Stepwise mutations, caused by proof-reading errors during DNA replication, change the number of repeat units. The differences in repeat units manifest as length polymorphisms that vary among individuals and can be detected using high-resolution gel electrophoresis (Selkoe and Toonen, 2006). Microsatellites have high mutation rates when compared to other molecular markers, making them especially informative for detecting recent genetic changes, allowing them to be used in intra-population studies (Putman and Carbone, 2014).

Despite the numerous advantages of microsatellites, they have to be developed *de novo* in plants that have not been previously studied. The development of microsatellites is an expensive venture that necessitates investment in specialised laboratory equipment (Zane *et al*., 2002; Kalia *et al*., 2011). Fortunately, there is the option of enlisting commercial services that can develop customised microsatellite markers. The cost and time of development depends on the extent of services required: basic development or additional quality testing and optimisation of PCR amplification protocols (Selkoe and Toonen, 2006).

An alternative method to *de novo* microsatellite marker development is the transfer of microsatellites available in one species to detect polymorphism in closely related species of interest. The success of cross-species amplification relies on the evolutionary distance between the source and target species, as the conserved regions that flank the microsatellite locus are used for primer design to amplify the locus. The larger the evolutionary distance, the less likely it is that the target species contains the same conserved flanking regions as the source species (Rossetto, 2001).

Microsatellites have been most widely used in diploid species as the co-dominant inheritance nature of these markers allows for the distinction between heterozygous and homozygous genotypes. Polyploid organisms, however, have the added complexity of allele dosage ambiguity as the maximum number of alleles present in one individual will be equal to the ploidy level (Sampson and Byrne, 2012). *Cyclopia* species are highly polyploid organisms with a chromosome basic number of $x = 9$ and ploidy levels that range from $4x$ to $14x$ (Goldblatt, 1981; Schutte, 1997; Motsa, 2016).

Stellenbosch University https://scholar.sun.ac.za

Due to challenges associated with allele dosage in polyploids, their microsatellite profiles are often interpreted rather as allele phenotypes, consequently disregarding dosage, with the alleles scored as a matrix of present versus absent (Pfeiffer *et al*., 2011; Dufresne *et al*., 2014). Subsequent analyses, such as the calculation of pairwise genetic distances are based on the allele phenotypes matrices rather than on genotypes (Sampson and Byrne, 2012).

The aim of this study is to develop and optimise microsatellite markers for use in *Cyclopia* species for further applications in field gene bank management such as the DNA fingerprinting of accessions. Markers will be developed in one species, namely *C. subternata*, and transferability will be tested to six other *Cyclopia* species. Ideally, a universal marker panel for DNA fingerprinting of *Cyclopia* accessions will be developed to allow for quick and efficient identification of accessions in the field gene bank. Although polyploid data analyses presents a challenge, appropriate methods will be applied throughout the study and strict rules for genotyping will be implemented.

3.2 Materials and Methods

3.2.1 Sampling

Seven species currently growing in the ARC field gene bank which are situated at the Nietvoorbij and Elsenburg Research Farms, near Stellenbosch, were chosen for this study. Table 3.1 summarises the number of accessions that were sampled for each of the different species and Table A3 in the Appendix provides further detail on the origins of the accessions. Results from Chapter 2 indicated that young leaves yielded the highest concentration and quality of DNA, therefore leaves were chosen from the tips of the shoots and healthy, green, young leaves were picked with gloved hands and placed in 1.5ml Eppendorf tubes. The tubes were kept on ice during transportation to the laboratory facilities. Approximately 0.05-0.15g of leaf material per plant was used as starting material for DNA extraction.

Table 3.1. Seven *Cyclopia* species and their accessions from the ARC field gene bank used for genetic marker development.

Species	No. accessions sampled
C. longifolia	30
C. subternata	22
C. genistoides	15
C. sessiliflora	10
C. maculata	10
C. pubescens	10
C. intermedia	4

Stellenbosch University https://scholar.sun.ac.za

3.2.2 DNA extractions

The optimised protocol developed in Chapter 2 was used for DNA extraction. A stainless steel bead was added to each tube and, using the TissueLyser (Qiagen), the leaf material was pulsed for 1min at 30Hz. A prewarmed (60°C) 2% (w/v) CTAB buffer (100mM Tris-hydroxymethyl-aminomethane hydrochloride [Tris-HCL], pH 8.0; 20mM ethylenediaminetetraacetic acid [EDTA], pH 8.0; 1.4M sodium chloride [NaCl]; 2% (w/v) hexadecyltrimethylammonium bromide [CTAB]; 2% (m/v) polyvinylpyrrolidone [PVP]) was added to the lysed tissue and thoroughly mixed and ground in the TissueLyser for 1min at 30Hz. After the addition of 2µl β-mercaptoethanol and 2µl protein kinase K (10mg/ml), the tubes were incubated at 60°C overnight. Chloroform:isoamylalcohol (24:1) was added in equal volumes after which the solution was centrifuged at a relative centrifugal force (RCF) of 16 100 for 10mins at 4°C. The aqueous phase was removed and the DNA was precipitated with 2/3 volume ice-cold isopropanol for 1h at -80°C. The precipitated DNA was spun down at a RCF of 16 100 for 20mins at room temperature and the supernatant discarded. The DNA pellet was cleaned using 200µl of 70% (v/v) ethanol after which the tubes were spun down at a RCF of 16 100 for 10mins at room temperature before the ethanol was decanted. This step was repeated and the pellet was dried completely before being re-suspended in 30µl ddH_2O. DNA concentrations were measured using the NanoDrop ND-1000 spectrophotometer (Thermo-Fisher).

3.2.3 Marker development and optimisation

Approximately 1-3µg of DNA for each of ten *C. subternata* accessions were sent to the laboratory of Genetic Marker Services (GMS) (Brighton, UK; www.geneticmarkerservices.com) for marker development using an enriched library method. The genomic DNA was digested with the *Rsa*I restriction enzyme and fragment sizes suitable for cloning were produced. DNA adapters were ligated to the cut ends of DNA and the fragments were amplified by PCR to increase the number of products and eliminate the fragments without adapters. The fragments were then enriched with known microsatellite sequence repeats bound to a filter: namely AG^{17}, AC^{17}, AAC^{10}, CCG^{10}, CTG^{10}, and AAT^{10}. The filters were washed at different temperatures ranging from 50-60°C with variations in saline-sodium citrate (SSC) stringencies (0.5X-2X). Thereby, positive clones were bound to the filter and unbound fragments could be eliminated. Fragments were PCR amplified, cloned into an appropriate plasmid, transformed into the *Escherichia coli* competent cell line JM109 and plated onto LB-agar plates that contain ampicillin. The colonies were screened for vectors that contain the microsatellite insertion and a total of 86 positive colonies were isolated, cultured and sequenced from three libraries. The sequences were edited to remove plasmid sequences and Tandem Repeat Finder

Stellenbosch University https://scholar.sun.ac.za

was used to locate repeat motifs (Benson, 1999). A total of 27 sequences were chosen for further development based on the presence of continuous repeat units with minimum interspersed interruptions, as well as sufficient flanking regions to allow for primer development using the online Primer3 software (Rozen and Skaletsky, 2000).

A subset of seven of the ten *C. subternata* accessions was used to optimise PCR amplification and identify polymorphic loci, also at the Genetic Marker Services laboratory. PCR reactions were prepared in 25µl volumes with approximately 20-50ng of genomic DNA, 1X PCR buffer, 0.5U *Taq* polymerase, 1.5mM of $MgCl_2$, 7pmol of each primer, 0.2mM of each dNTP and 0.8µg/µl bovine serum albumin (BSA). A touchdown PCR protocol was used with the following cycling conditions: 60s at 95°C, annealing for 60s with 2 cycles each at 64-59°C, 10 cycles at 58°C and 10 cycles at 57°C. An elongation step was performed at 72°C for 60s followed by a final 5mins at 72°C. The PCR products were run on 4% (m/v) MetaPhor high-resolution agarose gels (Lonza) in Tris-acetate-EDTA buffer (TAE) at 10°C.

Primer sequence information was received from the Genetic Marker Services laboratory for 27 developed markers. Markers were screened for usefulness in *C. subternata* accessions according to a specific set of five criteria (see Table 3.2). Failure of adherence to a single criterion excluded the marker from further investigation.

Firstly, polymorphism in the ten accessions was determined from the high-resolution agarose gel test pictures that were received from GMS. A total of 22 microsatellite markers were found to be polymorphic. These were divided into six multiplexes, after which forward primers fluorescently labelled with either PET, 6-FAM, NED or VIC (Applied Biosystems) were synthesised (see Appendix, Table A1).

Successful PCR amplification in the 22 *C. subternata* accessions was the second criterion. The original touchdown PCR cycling conditions were adapted and optimised for use with the Qiagen multiplex kit. PCR reactions were prepared in a total volume of 25µl with 50ng of genomic DNA, 1X QIAGEN Multiplex PCR Master Mix (HotStarTaq® DNA polymerase, Multiplex PCR Buffer with 6mM $MgCl_2$, dNTP Mix), 0.2µM of the forward and reverse primer (Applied Biosystems) and RNase free water. To minimise ‘pull-up’ peaks that were often found for the green- and blue-labelled primers (Cys10 and Cys22) only half the amount of those primers were added (0.1µM). Pull-up peaks arise because of intense signal from a certain fluorescence that affects the height of other markers in the same area (Hamilton, 2009). The relative intensities, measured as relative fluorescent units (RFU),

Stellenbosch University https://scholar.sun.ac.za

of the various dyes differ. VIC and 6-FAM emit stronger signals (100 RFU) than NED (40 RFU) and PET (25 RFU). Primers labelled with blue and green dye are more prone to a higher signal and therefore 'pull-up' peaks (Applied Biosystems, 2014). The cycling conditions were as follows: 95°C for 15mins followed by a denaturing step for 30s at 94°C, annealing for 90s with two cycles each at 64-59°C, 10 cycles at 58°C and 10 cycles at 57°C. An elongation step was performed at 72°C for 90s followed by a final extension step of 10mins at 72°C.

PCR products were run on a 2% (m/v) agarose gel with 1X Tris/Borate/EDTA buffer (TBE) and 7.5µl of ethidium bromide (EtBr) (10mg/ml) for initial screening. Only products that yielded visible bands were further prepared for sizing by capillary electrophoresis to correct for off-scale peaks using the following procedure: the PCR product was cleaned at the Central Analytical Facility (CAF) using the Nucleofast® 96 PCR clean-up kit (Macherey-Nagel) after which the DNA concentration was measured using a NanoDrop ND-1000 spectrophotometer (Thermo-Fisher). DNA was diluted to an end concentration of 12-15ng/µl and, for each sample, amplification products were sized using an ABI 3730 capillary sequencer (Applied Biosystems). GeneMapper 4.0 (Applied Biosystems) was used to visualise the microsatellite peaks and amplification was regarded as successful if a visible peak was observed in the correct range as specified by GMS (Table A1, Appendix). A marker was regarded as unsuccessful if it amplified in fewer than 80% of the 22 *C. subternata* individuals, thereby excluding a further three markers from the original 22 polymorphic markers, leaving a total of 19 for the transferability study (see Table 3.2).

A second round of polymorphism testing was performed during product sizing (criterion 3). At this stage of the primer optimisation, allele scoring was not yet regarded as a criterion, since transferability simply required the successful amplification of the marker. Therefore, all 19 markers were used for the transferability study in the seven *Cyclopia* species.

Unambiguous allele scoring was added as the fourth criterion (Table 3.2). Finally, reproducibility was tested (criterion 5) on a subset of 15-20% of the samples, for each species. Six markers were chosen for DNA fingerprinting (HB-Panel 1 with Cys21b, Cys22 and Cys25 and HB-Panel 2 with Cys5, Cys10 and Cys88).

3.2.4 Cross-species amplification

The 19 polymorphic markers that satisfied criteria 1 and 2 (Table 3.2) in *C. subternata*, the focal species, were tested for transferability to six other *Cyclopia* species. Amplification was regarded as

Stellenbosch University https://scholar.sun.ac.za

successful (denoted with a 'yes') if peaks were distinct and visible in the correct range on GeneMapper. The absence of a peak, or peaks below a certain threshold (500), were regarded as unsuccessful and denoted with a 'no'. For each species, the percentage transferability was calculated across all individuals, per marker:

$$Amplification\ (\%) = \frac{Number\ of\ amplified\ individuals}{Total\ number\ of\ individuals\ within\ species}\ x\ 100$$

An average was also determined per species across all markers. Standard deviations were calculated and the data are represented as a bar chart (see Section 3.3.2, Figure 3.1).

3.2.5 DNA fingerprinting: genotyping and data analyses

Two species, *C. genistoides* and *C. intermedia*, were excluded from the study because of complicated peak patterns associated with their highly polyploid nature (*C. genistoides* with $2n = 10x = 90$ and *C. intermedia* with $2n = 14x = 126$). Furthermore, for the remaining five species, any individuals for which genotypes were missing for more than one out of the six markers (more than ~10% missing data) were excluded from the study. The data set of 82 accessions for *C. longifolia*, *C. maculata*, *C. pubescens*, *C. sessiliflora* and *C. subternata* was reduced to 81 accessions, LMD30 (*C. longifolia*) being excluded because of amplification problems.

GeneMapper 4.0 software was used for peak detection and product size estimation for the six markers in HB-Panel 1 (Cys21b, Cys22 and Cys25) and HB-Panel 2 (Cys5, Cys10 and Cys88) for all seven species (refer to Table 3.4 for primer details). The expected product sizes in *C. subternata* were used as a guideline to identify allele sizes and create bins for each marker in the other six species. Automated scoring was performed and data were verified manually.

Only a few samples were fully heterozygous, with the same number of alleles as the ploidy level; therefore, full allelic configuration was not obtained for most samples. This is, however, not necessary for DNA fingerprinting purposes, since the main objective is to achieve a reproducible banding pattern that can distinguish between accessions within a species. General methods accounting for genotyping errors in diploids include the use of Micro-Checker to identify null alleles, large allele dropout and stutter peaks (Van Oosterhout *et al.*, 2004). These methods are often not applicable to polyploid genotyping and were not used. Strict rules were followed for allele calling to ensure reliability (see Table 3.3).

Table 3.2. Screening of 27 microsatellite markers in 22 accessions of *C. subternata* according to five criteria.

Marker	Criteria				
	1	**2**	**3**	**4**	**5**
	1st polymorphism testing	**Successful PCR amplification***	**2nd polymorphism testing**	**Unambiguous scoring**	**Reproducibility**
Cys2	✗	-	-	-	-
Cys3	✓	✓	✓	✗	-
Cys4	✓	✓	✓	✗	-
Cys5	✓	✓	✓	✓	✓
Cys6	✓	✓	✓	✗	-
Cys9	✓	✓	✓	✗	-
Cys10	✓	✓	✓	✓	✓
Cys11	✗	-	-	-	-
Cys12	✓	✓	✓	✗	-
Cys13	✓	✗	-	-	-
Cys14	✗	-	-	-	-
Cys16	✓	✓	✓	✗	-
Cys17	✓	✓	✓	✗	-
Cys18	✗	-	-	-	-
Cys20	✓	✓	✓	✗	-
Cys21	✗	-	-	-	-
Cys21b	✓	✓	✓	✓	✓
Cys22	✓	✓	✓	✓	✓
Cys23	✓	✓	✓	✗	-
Cys24	✓	✓	✓	✗	-
Cys25	✓	✓	✓	✓	✓
Cys26	✓	✓	✓	✗	-
Cys27	✓	✓	✓	✗	-
Cys36	✓	✗	-	-	-
Cys37	✓	✗	-	-	-
Cys88	✓	✓	✓	✓	✓
Cys90	✓	✓	✗	-	-
No. markers	22	19	18	6	6

***Amplification was regarded as successful if the marker amplified in at least 80% of the 22 *C. subternata* individuals.**

Stellenbosch University https://scholar.sun.ac.za

Table 3.3. Rules for genotyping with six microsatellite markers in the polyploid *Cyclopia* genus.

Marker	Peak characteristics	Rules for genotyping
Cys5	Quite easy to score, no stutter	/
Cys10	Sometimes produced a double peak	Only call right-handed peak (usually the highest)
Cys21b	No stutter, very polymorphic	Call all peaks, usually not more than 6 Watch out for pull-up from Cys10
Cys22	Some stutter	Only call highest peaks
Cys25	Sometimes produced a double peak	Only call right-handed peak Do not call small peak directly before main peak Small peak completely separate from main peak can be called
Cys88	No stutter, very polymorphic	/

Binary matrices of presence or absence of the particular alleles were constructed in POLYSAT 1.4 for each of the studied species (Clark and Jasieniuk, 2011). A binary matrix represents allele phenotypes that can be interpreted in a way to dominant data. This allows for the calculation of statistics such as the probability that two individuals in a population will have the same genotype at multiple loci, denoted as the probability of identity (P_{ID}) (Silva *et al*., 2012). This was calculated for 'dominant data' as:

$$P_{(ID)} = p^2 + (2pq)^2 + (q^2)^2$$

For each locus, each allele was denoted as present (1) or absent (0) across all individuals. A frequency was obtained by dividing by the total number of individuals to determine the presence (p) or absence (q) of each allele within the species. Each allele was examined separately before combining all the alleles per locus for overall probability using the product rule (Waits *et al*., 2001). Total probability of identity was calculated by combining the multilocus probabilities across all accessions within a species.

Allelic diversity statistics were calculated as follows. The number of alleles per locus, averaged over loci (N_A), and the maximum number of alleles was determined from the binary matrices in Microsoft Excel (Microsoft, Redmond, Washington). The number of private alleles per population, where a population represents a group of genebank accessions from the same species, was calculated in GenAlEx 6.501 using the frequency option for binary data (Peakall and Smouse, 2006). The polymorphic information content (PIC) is a measure by which the usefulness of each marker is determined and this was calculated per marker for 'dominant data':

$PIC = 1 - \sum pi^2$, where the squared frequency of allele i is represented as pi (Powell *et al*., 1996).

Stellenbosch University https://scholar.sun.ac.za

Pairwise genetic distances were calculated between all accessions based on the Bruvo distance measure in POLYSAT 1.4 (Clark and Jasieniuk, 2011). Principal coordinate analysis (PCoA) plots were constructed from the genetic distance matrix, also in GenAlEx 6.501, to visualise the relationships among individuals (Peakall and Smouse, 2006).

3.3 Results and Discussion

3.3.1 Microsatellite marker optimisation and characteristics in *C. subternata*

The Genetic Marker Services (GMS) laboratory reported 86 positive clones that were obtained from three plated libraries, yielding 27 microsatellite loci. The majority of the microsatellites had dinucleotide repeats, only two out of 27 (7.4%) having trinucleotide repeats (see Appendix, Table A1). Twenty-five of the markers exhibited perfect repeats, where the repeat is present as a single, uninterrupted motif (Weising *et al*., 2005). The other two markers, Cys25 and Cys90, fall in the compound repeat class, in which two different types of repeat follow one after the other (Weising *et al*., 2005).

Mono- and dinucleotide repeats are generally more frequent in the plant genome than tri-, tetra- and pentanucleotides (Weising *et al*., 2005). Dinucleotide repeats, however, suffer from the occurrence of stutter peaks caused by DNA polymerase slippage and are often more challenging to score (Holton, 2001). It is not known whether the *Cyclopia* microsatellite repeats are located in coding or non-coding areas of the genome. Trinucleotides, which were only found in two of the *C. subternata* markers, are reported to be the most prevalent repeat in expressed sequence tags (ESTs), since insertions or deletions will not affect the open reading frame and result in altered amino acids (Cardle *et al*., 2000; Weising *et al*., 2005). As microsatellites obtained from ESTs occur in regions of the genome that are readily transcribed, these are often more conserved across taxa. However, non-coding regions, in turn, display a higher level of polymorphism, since mutations accumulate more readily (Testolin *et al*., 2000; Gutierrez *et al*., 2005). Overall, AG/CT microsatellite repeats generally seem to be more prevalent in plants, as opposed to AC/GT repeats (Morgante and Olivieri, 1993; Cipriani *et al*., 1999; Yasodha *et al*., 2005). A similar trend was observed for the current study, with the majority of *Cyclopia* markers obtained from an AG-only library.

Based on the first round of polymorphism testing, a final marker set of 22 was chosen for further optimisation in the focal species, *C. subternata*. For the second criterion, three markers (Cys13, Cys36

and Cys37) were excluded, due to unsuccessful PCR amplification. The second round of polymorphism testing (criterion 3) found Cys90 to be monomorphic, however the marker was still included in the transferability study since it amplified in 100% of individuals. A total of 19 markers were therefore included for the transferability study (section 3.3.2). The difficulty of identifying alleles unambiguously in polyploids, especially in the presence of stutter peaks, led to strict implementation of the fourth criterion: ease of scoring. Only six markers (Cys5, Cys10, Cys21b, Cys22, Cys25 and Cys88) adhered to this criterion in *C. subternata* with limited or no stutter bands.

3.3.2 Cross-species amplification

The nineteen markers were tested for amplification in the focal species, *C. subternata*, followed by transferability testing in six other *Cyclopia* species. The markers amplified best in *C. subternata* (97%, Figure 3.1). Across species, the transferability ranged from 81% (*C. pubescens*) to 88% (*C. intermedia*). Two markers (Cys20 and Cys27) had reduced performance, even in *C. subternata* where Cys27 amplified in 82% and Cys20 amplified in 86% of individuals (Figure 3.2). These two markers amplified in only 26% (Cys27) and 35% (Cys20) of the total accessions (Figure 3.2). The other 17 markers amplified in 75% or more individuals across all accessions. Interestingly, some of the markers (such as Cys5, Cys9, Cys12, Cys17, Cys24 and Cys88) had higher success rates in some of the non-focal species, than in *C. subternata* (see Figure 3.2). This could however be biased by the small sample size of the *C. intermedia*, *C. maculata*, *C. pubescens* and *C. sessiliflora* species.

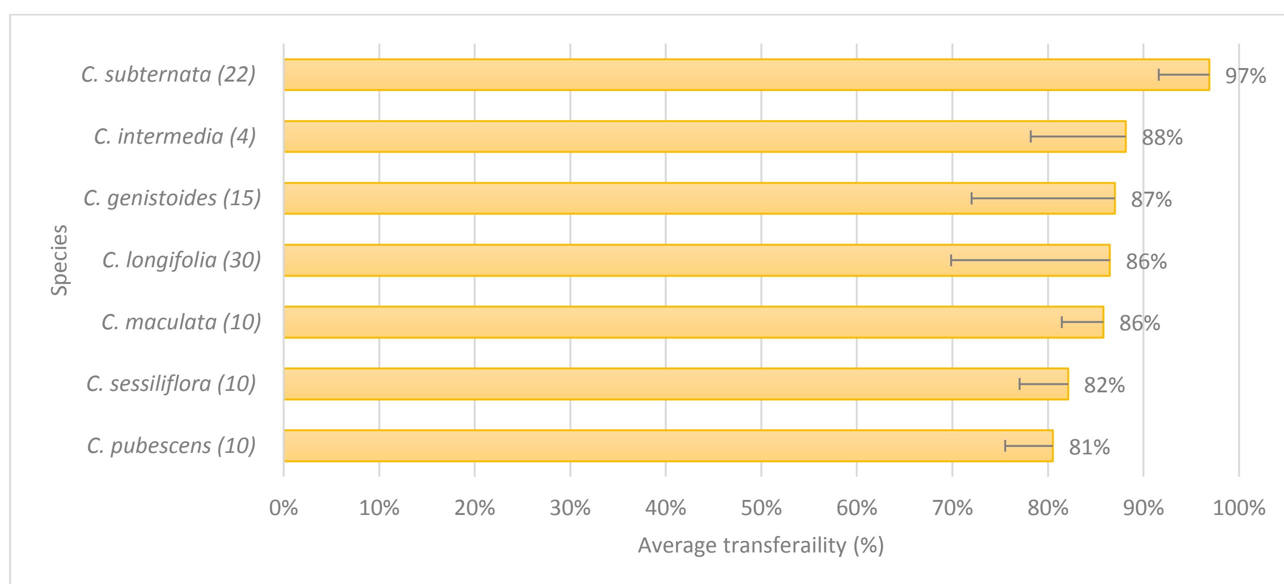


Figure 3.1. The average percentage transferability of 19 *C. subternata* microsatellite markers to six other *Cyclopia* species. Transferability was determined as the average percentage amplification across all accessions within a species. Error bars indicates the standard deviation from the average percentage amplification.

Stellenbosch University https://scholar.sun.ac.za

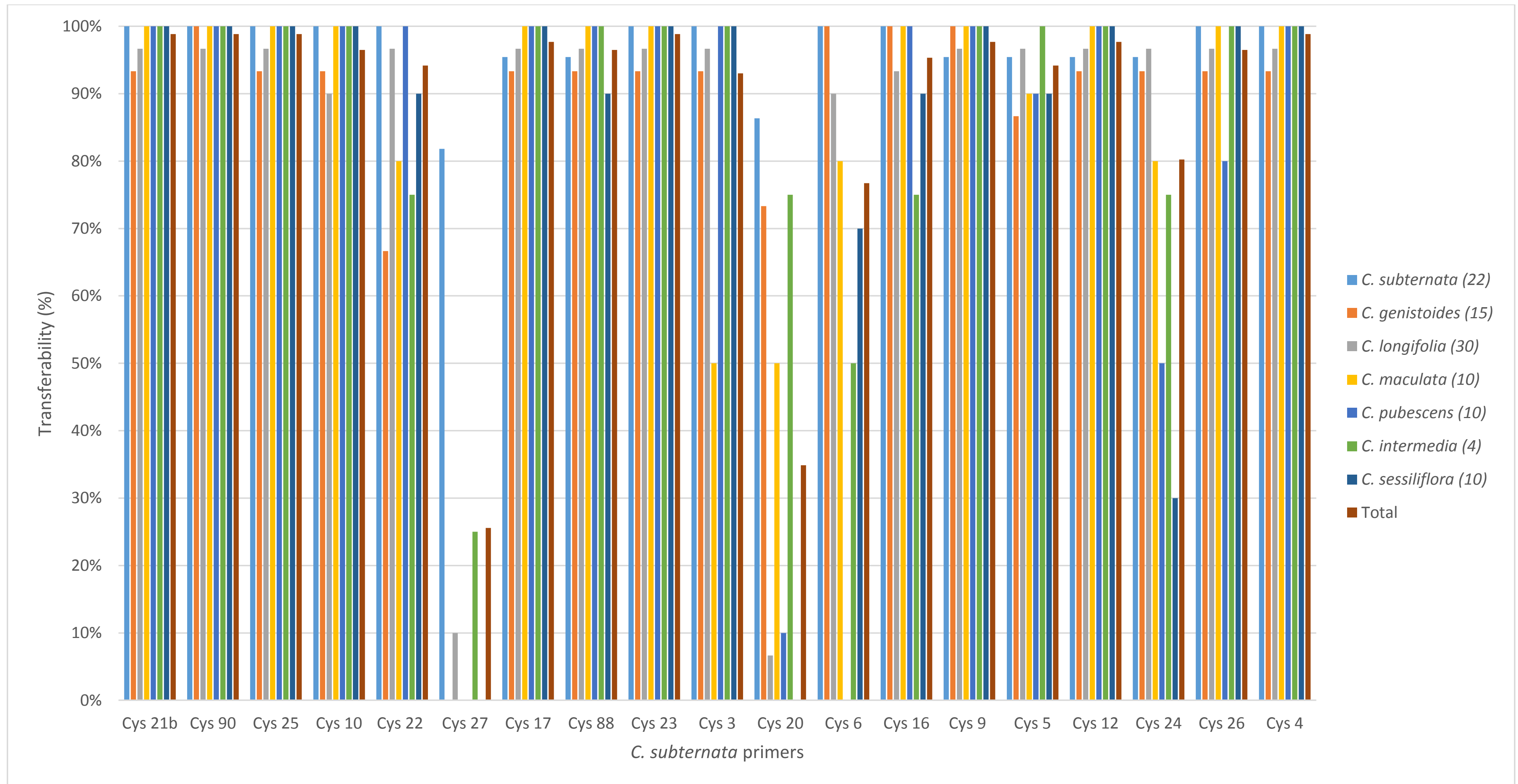


Figure 3.2. The percentage transferability of 19 *C. subternata* microsatellite markers to six other *Cyclopia* species (represented by the various colours). Transferability was calculated as the percentage of accessions in which the marker amplified for the six species. The amplification of markers in *C. subternata* and the total percentage transferability per marker are also indicated.

The success of cross-species amplification within the genus *Cyclopia* was demonstrated by the high rate of transferability, suggesting that flanking regions are generally well conserved between species. The phylogenetic position of the genus *Cyclopia* relative to the rest of the Podalyrieae has been studied using the internal transcribed spacer (ITS) region of ribosomal DNA as well as the ribulose-1, 5-bisphosphate carboxylase/oxygenase large subunit (*rbcL*) gene region (Boatwright *et al*., 2008). The phylogenetic relationships among the seven species of interest in the current study have, however, not yet been resolved on a fine scale (Figure 3.3). The transferability success of the species did not correlate with the topology of the tree, as, for example, *C. subternata* and *C. pubescens* cluster together but the transferability of the *C. subternata* primers had the lowest success (81%) in *C. pubescens* when compared to the other species. Nevertheless, the high amplification success of 81-88% in *Cyclopia* species correlate with a review by Rossetto *et al.* (2001) in which transferability within genera in plants was generally at least 76%. These numbers decreased significantly as evolutionary distance increased; within family transferability being only 35%. The *Cyclopia* markers could potentially be tested on other genera to determine inter-genus transferability. An example would include the commercially important *Aspalathus* (family Fabaceae; tribe Crotalarieae) better known as rooibos (Joubert *et al*., 2008).

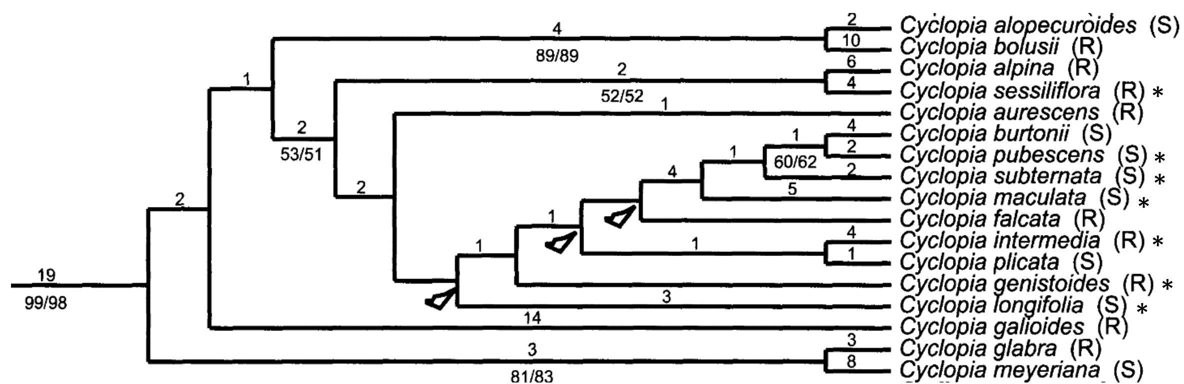


Figure 3.3. Phylogenetic relationships of *Cyclopia* species according to a study conducted by Boatwright *et al.* (2008). The maximum parsimony tree was constructed based on ITS and *rbcL* data. The numbers above the branches represent Fitch lengths, while values below the branches are bootstrap percentages above 50% (equal weights before the slash and successive weights after the slash). The seven species of interest in the current study are indicated by an * with (R) indicating a sprouter and (S) indicating a non-sprouter.

Sized-based comparisons of microsatellite alleles are often applied with the assumption that distinct alleles will vary in length (Selkoe and Toonen, 2006). Comparisons across different species should be accompanied with caution, since alleles identical in length might not indicate an identical descent (Hoshino *et al*., 2012). This phenomenon is known as size homoplasy and can disguise underlying sequence variation such as rearrangements and duplications (Rossetto *et al*., 2001). Homoplasy associated with sequence variation can be detected by sequencing the alleles. Another type of

Stellenbosch University https://scholar.sun.ac.za

homoplasy that is often undetected, however, occurs when two alleles are identical in sequence and therefore length, but not identical by descent (Selkoe and Toonen, 2006). In the current study none of the amplified PCR products in *Cyclopia* were sequenced to investigate potential homoplasy.

3.3.3 DNA fingerprinting and allelic diversity

The six markers that adhered to criterion 4 gave reproducible patterns. They were combined into two multiplexes (see Table 3.4) namely HB-Panel 1 (Cys21b, Cys22 and Cys25) and HB-Panel 2 (Cys5, Cys10 and Cys88). Table A2 in the appendix indicates the DNA fingerprint profiles that were generated for each of the accessions for five *Cyclopia* species in the ARC field gene bank using the six markers. See Figure 3.4 for GeneMapper output showing example peaks for each locus.

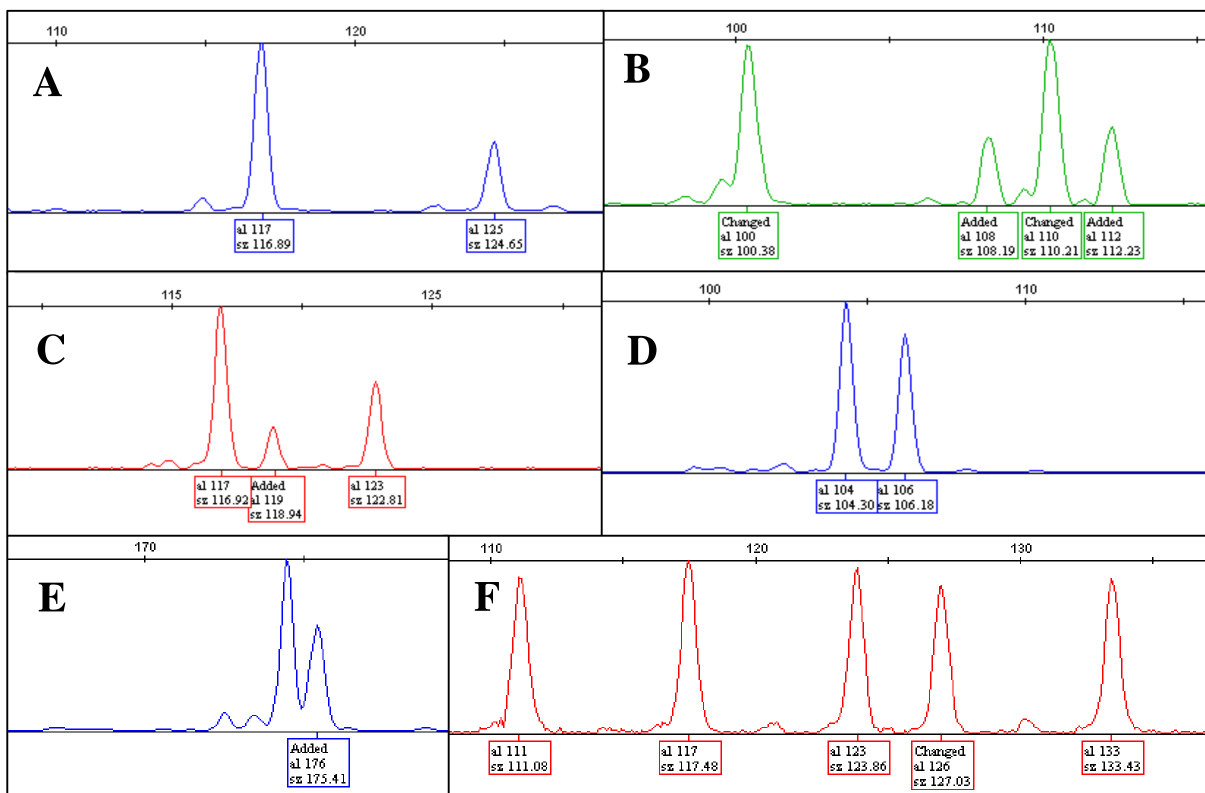


Figure 3.4. Examples of microsatellite peaks from GeneMapper 4.0 for a *C. subternata* accession, SKB18 for locus (A) Cys5, (B) Cys10, (C) Cys21b, (D) Cys22, (E) Cys25 and (F) Cys88.

Cyclopia genistoides (see Figure 3.5) and *C. intermedia* showed complicated peak patterns, possibly due to their high ploidy levels (10*x* and 14*x*). These two species were therefore excluded from the DNA fingerprinting study, and subsequently only five out of seven species in the field gene bank were fingerprinted. The number of alleles per locus across the five species ranged from 13 (Cys22) to 21 (Cys25) (Table 3.4). A total of 75 alleles were detected across all six loci in *C. subternata* (data not shown) and 100 alleles were detected across all species (Table 3.4). Table 3.5 summarises the

Stellenbosch University https://scholar.sun.ac.za

performance of the six microsatellite markers for DNA fingerprinting purposes in each of the seven species.

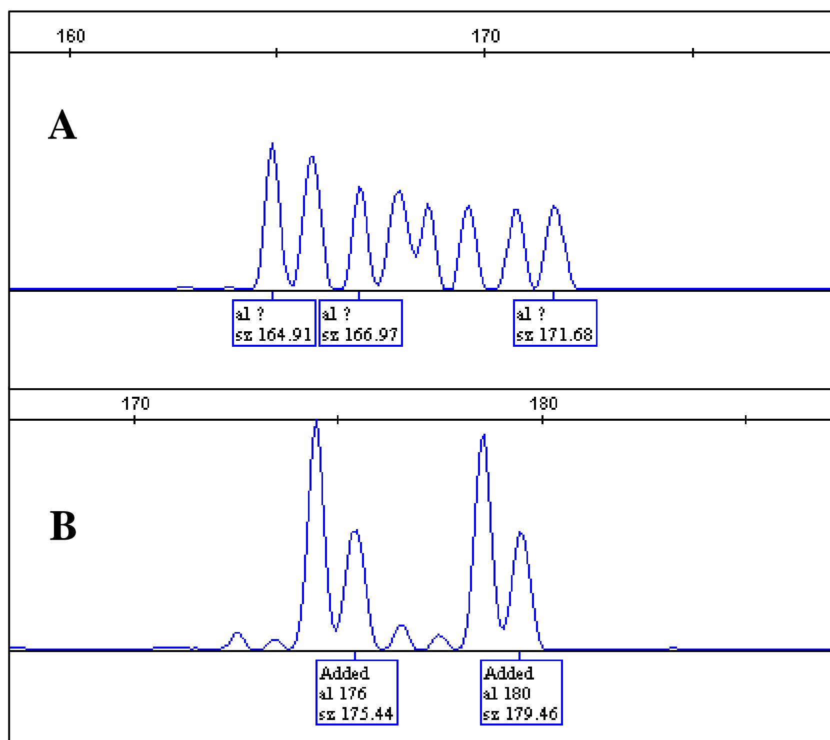


Figure 3.5. Example of GeneMapper traces showing the complicated peak pattern of A) *C. genistoides*, a decaploid in comparison with the simpler pattern of B) *C. subternata*, a hexaploid, for the microsatellite locus Cys25.

Table 3.4. Information for the six primer pairs incorporated into *Cyclopia* microsatellite multiplex panels (HB-Panel 1 and 2) used for DNA fingerprinting.

MP	Locus	F/R	Primer sequence (5'-3')	T_M	Repeat	Range (bp)	Label	N_A
HB-Panel 1	Cys21b	F	AAAGTGTTCGTAGTTTGCTTC	54	CT^{17}	90-160	PET	18
	Cys21b	R	GTACTAGAATAGCTGCAAGACA	52				
	Cys22	F	GCGTGAGGACAGAGACACAA	60	AG^{12}	76-140	FAM	13
	Cys22	R	GACGGCGGAAAAGAAAATAA	60				
	Cys25	F	TGAATGGATGAGTGGCAAGA	60	$GT^{10}C^{1}AG^{6}$	153-233	FAM	21
	Cys25	R	TCACCAGCAGATTCAGTTGAG	60				
HB-Panel 2	Cys5	F	CGTGTGTGTGTGTGTGAGAAA	60	AG^{14}	100-160	FAM	16
	Cys5	R	GCGTTAGCGAAGGACTAGCT	60				
	Cys10	F	TTGTCATGTTCTCACGCTGTT	60	TC^{8}	90-130	VIC	17
	Cys10	R	TCTCTGCACACAAAAATCACG	60				
	Cys88	F	TTTGCCTGCTCAGAAATCCC	60	GTT^{7}	100-160	PET	15
	Cys88	R	TGTTTTCTTCTTGTTTTCAATCCA	60				

MP = multiplex; F/R = forward or reverse primer; T_M = melting temperature; N_A = number of alleles per locus, averaged over loci was calculated across all accessions.

Table 3.5. Summary table for the performance of the six *C. subternata* markers for DNA fingerprinting purposes in each of the seven *Cyclopia* species.

	Cys5	Cys10	Cys21b	Cys22	Cys25	Cys88
C. genistoides	High ploidy – complex peak patterns					
C. intermedia	High ploidy – complex peak patterns					
C. longifolia	✓	✓	✓	✓	✓	✓
C. maculata	Monomorphic	✓	✓	Monomorphic	Monomorphic	✓
C. pubescens	✓	✓	✓	✓	✓	✓
C. sessiliflora	✓	✓	✓	Monomorphic	✓	✓
C. subternata	✓	✓	✓	✓	✓	✓

Cyclopia maculata, a known tetraploid, had a maximum of 4 alleles per marker per individual. In *C. longifolia* and *C. subternata*, which are hexaploid, the maximum number of alleles per individual ranged from 3 to 6. For the two species which have unknown chromosome numbers, ploidy can be estimated based on the maximum number of alleles per individual. *Cyclopia pubescens* had a maximum of 4 alleles, indicating that this species could be tetraploid, while *C. sessiliflora* had a maximum of 5 alleles, indicating possible hexaploidy. The six markers were generally polymorphic in all species, except in *C. maculata*, in which only some were polymorphic. Polymorphic information content (PIC) ranged from 0.44 in *C. maculata* to 0.87 in *C. subternata*. The small sample sizes of *C. maculata*, *C. pubescens* and *C. sessiliflora* (10 samples per species) could have resulted in the underrepresentation of some alleles, thereby affecting the discriminatory power of the utilised markers in these species.

Table 3.6. Allelic diversity of accessions of five *Cyclopia* species from the ARC gene bank for six microsatellite markers (HB-Panel 1 and HB-Panel 2).

	N_A	Total alleles	Private alleles	Max bands	Mean bands	PIC	P_{ID}
C. longifolia	11	83	12	6	3	0.82	3.5×10^{-12}
C. maculata	3	14	3	3	1	0.44	5.6×10^{-4}
C. pubescens	3	22	5	4	2	0.74	9.0×10^{-9}
C. sessiliflora	3	24	3	5	2	0.68	1.2×10^{-9}
C. subternata	13	62	14	6	3	0.87	4.9×10^{-14}

N_A = number of alleles per locus, averaged over loci; Maximum number of alleles calculated per individual; Mean number of alleles calculated per individual; PIC = polymorphic information content; P_{ID} = probability of identity for dominant data

Three markers exhibited limited polymorphism in some of the *Cyclopia* species. The cross-amplification of Cys22 in *C. maculata* gave a 76bp product in 80% of individuals. This falls outside the expected range of 87-137bp. Interestingly, the flanking regions of this marker in *C. subternata*

Stellenbosch University https://scholar.sun.ac.za

without a microsatellite repeat is 88bp in length, indicating that this marker contains no microsatellite repeat between the primer flanking regions in *C. maculata*. A polymorphic locus in the focal species could be monomorphic in another species, because the microsatellite repeat is either absent or not as variable between the primer binding sites (Weising *et al*., 2005). The possibility of a false positive, where the microsatellite repeat is completely absent and the product is monomorphic, cannot be excluded for Cys22 in *C. maculata*. Furthermore, this marker had limited polymorphism in *C. sessiliflora* with a single allele size of 104bp in 90% of individuals. Similarly, Cys25 and Cys5 yielded only one and three alleles respectively in *C. maculata*. Cys25 had a consistent allele size of 170bp in all *C. maculata* individuals, indicating there is little variation for this locus across individuals within this species. Although Cys5 yielded three alleles in *C. maculata*, it displayed less allelic variation than was exhibited in *C. subternata*, which had up to 14 different alleles (data not shown). For DNA fingerprinting purposes, additional microsatellites would be required to replace these monomorphic markers for fingerprinting of *C. maculata* and *C. sessiliflora*.

A method called MAC-PR (microsatellite DNA allele counting-peak ratios) is often utilised in lower-order polyploids to infer allele dosage by using the peak area ratios provided by the software program GeneMapper 4.0 (Esselink *et al*., 2004). This method is, however, very dependent on high quality data and known genetic relationships of the plants (Brandrud, 2014). It was therefore not applied during the current DNA fingerprinting study, since the study species had a large number of alleles (maximum of six) and genetic relationships are, as yet, largely unknown, making this particular method inappropriate.

The binary matrices that represent allele phenotypes were therefore used to calculate the probability of identity for all five species and applied to estimate whether DNA fingerprints could successfully distinguish among accessions (Table 3.6). Even though the markers had the lowest P_{ID} in *C. maculata*, it was still sufficient as one in every 1774 accessions may have the same genotype. This is attributed to the high genetic diversity of polyploid microsatellite profiles (Pfeiffer *et al*., 2011). One in every $2x10^{13}$ *C. subternata* accessions may have the same genotype according to the P_{ID} value of the six markers in this species. Similarly, a study by Silva *et al*. (2012) used only three markers in the octoploid sugarcane to distinguish one in every million accessions.

The analyses of genetic variation among polyploid individuals of different species are further complicated when species exhibit different ploidy levels. Principal coordinate analysis (PCoA) is a graphical representation of a distance matrix that is represented on a smaller scale while still accurately reflecting the original distances (Zuur *et al*. 2007). The PCoA method has been applied

Stellenbosch University https://scholar.sun.ac.za

successfully to *Ononis* species, which are also legumes, to investigate genetic variation across diploids and tetraploids (Kloda *et al*., 2008). Furthermore, the Bruvo genetic distance measure (used in the current study to construct PCoA plots – see Appendix, Table A4) can accommodate mixed ploidy levels (Bruvo *et al*., 2004; Dufresne *et al*., 2014). It might, however, cluster individuals according to ploidy level, leading to an overestimation of genetic distance between individuals with unrelated ploidy levels (Clark and Jasieniuk, 2011). In the current study this does not seem to be a limitation as species did not cluster according to their ploidy level in the PCoA plot (Figure 3.6). For example, *C. pubescens*, most likely a tetraploid, clustered with *C. subternata*, a known hexaploid rather than with *C. maculata*, a known tetraploid.

Cyclopia longifolia clustered mainly in quadrant 4 of the PCoA plot, thereby distinctly separating it from the other species (Figure 3.6), while *C. maculata* clustered in quadrant 3. *Cyclopia pubescens*, *C. sessiliflora* and *C. subternata* could, however, not be distinguished from each other and were dispersed across quadrants 1 and 2. Analysis using the six microsatellite markers seems to indicate that these three species share a similar genepool. This emphasises the need for additional microsatellite markers to help resolve the abovementioned cluster.

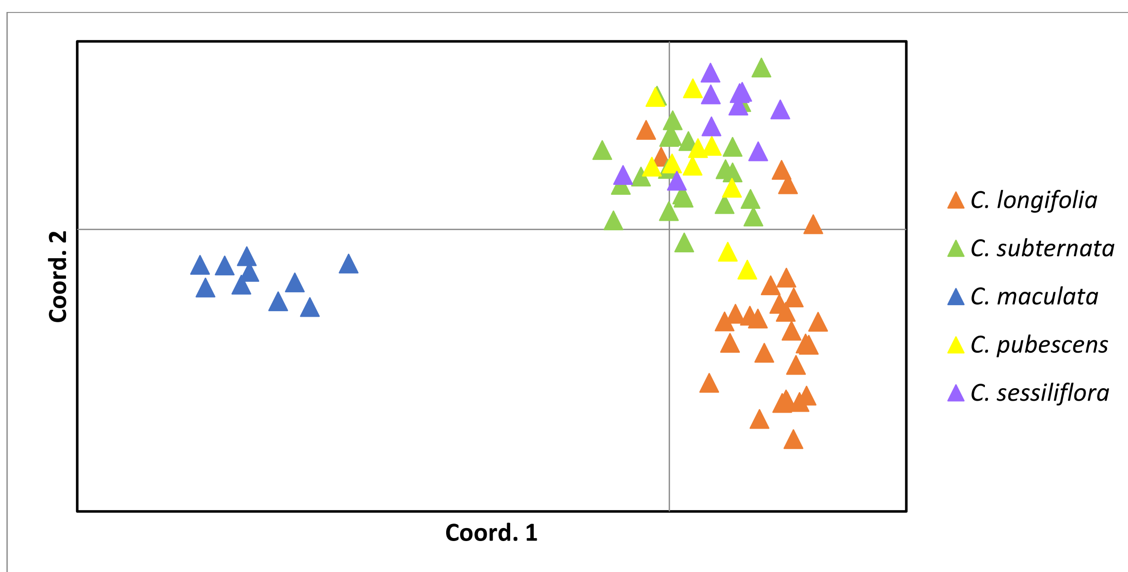


Figure 3.6. Principal coordinate analysis (PCoA) of accessions from the ARC field gene bank of five *Cyclopia* species based on Bruvo's genetic distance measure.

Reproductive isolation often ensues between species of different ploidy level (Husband and Schemske, 2000). This can be attributed to prezygotic reproductive barriers such as differing flowering times, geographical isolation and pollinators that are specific to a certain plant (Husband and Sabara, 2011). A study by Motsa *et al*. (2016) concluded that the flowering times of *C. genistoides*

Stellenbosch University https://scholar.sun.ac.za

(a decaploid) and *C. subternata* (a hexaploid) occurred mainly between September and October, but that minor changes was to be expected between the various species. As all *Cyclopia* species are pollinated by large xylocopid bees (van Wyk, 1993), a more likely barrier would be post-zygotic where a cross between a polyploid plant and one of lower ploidy will often result in the formation of a sterile hybrid with an uneven number of chromosomes (Köhler *et al*., 2010). Kloda *et al*. (2008) found separate clustering using principal coordinate analysis (PCoA) between *Ononis* species of different ploidy levels. Additionally, a study based on six microsatellite markers as well as single nucleotide polymorphism (SNPs) derived from restriction-site associated DNA sequencing (RAD-seq) investigated the genetic variation across related species with varying ploidy levels in the genus *Cochlearia* (Brandrud, 2014). The PCoA plots based on the six microsatellite markers indicated an overlap between species, but PCoA plots based on high resolution RAD-seq SNP data separated the groups quite clearly. In future applications, SNPs may therefore be a useful addition to the microsatellite markers developed in the current study to distinguish the species more accurately. It should be kept in mind that, as with microsatellites, the application of SNPs in polyploid studies present challenges in the determination of dosage. Whereas the presence and absence of a nucleotide would indicate a heterozygous or homozygous individual in diploids, polyploids have multiple allele copies (Cordeiro *et al*., 2006). Several studies have successfully characterised SNPs in polyploid plants such as in the tetraploid potato using pyrosequencing and in the highly polyploid sugarcane using ESTs as well as 454 sequencing (Rickert *et al*. 2002; Cordeiro *et al*., 2006; Bundock *et al*., 2009). Additional marker technologies will therefore be of great use in future studies of *Cyclopia*.

3.4 Conclusion

This study successfully developed 19 *C. subternata* microsatellite markers and tested the transferability of these to six other *Cyclopia* species. The markers amplified in 97% of *C. subternata* individuals. The average cross-species amplification was high, ranging from 81% (*C. pubescens*) to 88% (*C. intermedia*). Two multiplexes consisting of three microsatellites each were sufficient for distinguishing among various accessions within a species, which will be useful in keeping track of accessions in the ARC field gene bank. Although the aim was to develop a universal panel for the five C*yclopia* species to facilitate comparisons, future studies should include species-specific panels. The development of additional markers will therefore create scope for more informative, species-specific markers, as markers monomorphic in one species could be polymorphic in another. The limited knowledge of the *Cyclopia* genome and unknown chromosome inheritance patterns would first need to be investigated before further applications, such as marker-assisted breeding, can be attempted.

Stellenbosch University https://scholar.sun.ac.za

References

Applied Biosystems, 2014. Experimental design, in: DNA Fragment Analysis by Capillary Electrophoresis. Thermo Fisher Scientific Inc, 36-40.

Benson, G., 1999. Tandem Repeats Finder: A program to analyse DNA sequences. Nucleic Acids Research 27, 573-578.

Bester, C., Tobutt, K., Mansvelt, E., Blomerus, L., Jolly, N., Pieterse, W., Smit, L., Van Schalkwyk, D., 2013. The value and impact of the ARC Infruitec-Nietvoorbij genebanks. Acta Horticulturae 1007, 975-980.

Bester, C., Joubert, M.E., Joubert, E., 2016. A breeding strategy for South African indigenous herbal teas. Acta Horticulturae, In Press.

Boatwright, J., Savolainen, V., Van Wyk, B., Schutte-Vlok, A., Forest, F., Van Der Bank, M., 2008. Systematic position of the anomalous genus *Cadia* and the phylogeny of the tribe Podalyrieae (Fabaceae). Systematic Botany 33, 133-147.

Brandrud, M.K., 2014. Polyploidy and ecotype variation in *Cochlearia officinalis* L. and related species. Unpublished M.Sc. thesis, University of Oslo, Norway, 34-42.

Bruvo, R., Michiels, N., D'Souza, T., Schulenburg, H., 2004. A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. Molecular Ecology 13, 2101-2106.

Bundock, P.C., Eliott, F.G., Ablett, G., Benson, A.D., Casu, R.E., Aitken, K.S., Henry, R.J., 2009. Targeted single nucleotide polymorphism (SNP) discovery in a highly polyploid plant species using 454 sequencing. Plant Biotechnology Journal 7, 347-354.

Cardle, L., Ramsay, L., Milbourne, D., Macaulay, M., Marshall, D., Waugh, R., 2000. Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 156, 847-854.

Cipriani, G., Lot, G., Huang, W., Marrazzo, M., Peterlunger, E., Testolin, R., 1999. AC/GT and AG/CT microsatellite repeats in peach: Isolation, characterisation and cross-species amplification in *Prunus*. Theoretical and Applied Genetics 99, 65-72.

Clark, L.V. & Jasieniuk, M., 2011. POLYSAT: An R package for polyploid microsatellite analysis. Molecular Ecology Resources 11, 562-566.

Cordeiro, G.M., Eliott, F., McIntyre, C.L., Casu, R.E., Henry, R.J., 2006. Characterisation of single nucleotide polymorphisms in sugarcane ESTs. Theoretical and Applied Genetics 113, 331-343.

Dufresne, F., Stift, M., Vergilino, R., Mable, B.K., 2014. Recent progress and challenges in population genetics of polyploid organisms: An overview of current state-of-the-art molecular and statistical tools. Molecular Ecology 23, 40-69.

Stellenbosch University https://scholar.sun.ac.za

Esselink, G.D., Nybom, H., Vosman, B., 2004. Assignment of allelic configuration in polyploids using the MAC-PR (microsatellite DNA allele counting-peak ratios) method. Theoretical and Applied Genetics 109, 402-408.

Fowler, C. & Hodgkin, T., 2004. Plant genetic resources for food and agriculture: Assessing global availability. Annual Review of Environment and Resources 29, 143-179.

Gepts, P., 2006. Plant genetic resources conservation and utilization: The accomplishments and future of a societal insurance policy. Crop Science 46, 2278-2292.

Goldblatt, P., 1981. Chromosome numbers in legumes II. Annals of the Missouri Botanical Garden 68, 551-557.

Gutierrez, M.V, Vaz Patto, M.C., Huguet, T., Cubero, J.I., Moreno, M.T., Torres, A.M., 2005. Cross-species amplification of *Medicago truncatula* microsatellites across three major pulse crops. Theoretical and Applied Genetics 110, 1210-1217.

Hamilton, M., 2009. Genotype frequencies, in: Hamilton, M. (Ed.), Population Genetics. Wiley-Blackwell, Oxford, UK 23-24.

Holton, T.A., 2001. Plant genotyping by analysis of microsatellites, in: Henry, R.J. (Ed.), Plant Genotyping: The DNA Fingerprinting of Plants. CABI, New York, USA, 15-27.

Hoshino, A.A., Bravo, J.P., Nobile, P.M., Morelli, K.A., 2012. Microsatellites as tools for genetic diversity analysis, in: Caliskan, M. (Ed.), Genetic Diversity in Microorganisms. InTech, Croatia, EU, 149-170.

Husband, B.C. & Sabara, H., 2011. Reproductive isolation between autotetraploids and their diploid progenitors in fireweed, *Chamerion angustifolium* (Onagraceae). New Phytologist 161, 703-713.

Husband, B.C. & Schemske, D.W., 2000. Ecological mechanisms of reproductive isolation between diploid and tetraploid *Chamerion angustifolium*. Journal of Ecology 88, 689-701.

Joubert, E., Gelderblom, W.C.A., Louw, A., De Beer, D., 2008. South African herbal teas: *Aspalathus linearis*, *Cyclopia* spp. and *Athrixia phylicoides* - a review. Journal of Ethnopharmacology 119, 376-412.

Joubert, E., Joubert, M., Bester, C., De Beer, D., De Lange, J., 2011. Honeybush (*Cyclopia* spp.): From local cottage industry to global markets - the catalytic and supporting role of research. South African Journal of Botany 77, 887-907.

Kalia, R.K., Rai, M.K., Kalia, S., Singh, R., Dhawan, A.K., 2011. Microsatellite markers: An overview of the recent progress in plants. Euphytica 177, 309-334.

Kloda, J.M., Dean, P., Maddren, C., MacDonald, D.W., Mayes, S., 2008. Using principle component analysis to compare genetic diversity across polyploidy levels within plant complexes: An example from British Restharrows (*Ononis spinosa* and *Ononis repens*). Heredity 100, 253-260.

Stellenbosch University https://scholar.sun.ac.za

Köhler, C., Mittelsten Scheid, O., Erilova, A., 2010. The impact of the triploid block on the origin and evolution of polyploid plants. Trends in Genetics 26, 142-148.

Laurentin, H., 2009. Data analysis for molecular characterization of plant genetic resources. Genetic Resources and Crop Evolution 56, 277-292.

Morgante, M. & Olivieri, A., 1993. PCR-amplified microsatellites as markers in plant genetics. The Plant Journal 3, 175-182.

Motsa, M., 2016. Natural phenology, fecundity, genetic variation and seed dormancy of *Cyclopia* species (Honeybush). Unpublished D.Tech. thesis, Tshwane University of Technology, Pretoria, 159-163.

Peakall, R. & Smouse, P.E., 2006. GenAlEx 6: Genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6, 288-295.

Pfeiffer, T., Roschanski, A.M., Pannell, J.R., Korbecka, G., Schnittler, M., 2011. Characterization of microsatellite loci and reliable genotyping in a polyploid plant, *Mercurialis perennis* (Euphorbiaceae). Journal of Heredity 102, 479-488.

Powell, W., Morgante, C., Andre, M., Hanafey, M., Vogel, J., Tingey, S., Rafalski, A., 1996. The comparison of RFLP, RAPD, AFLP, and SSR (microsatellite) markers for germplasm analysis. Molecular Breeding 2, 225-238.

Putman, A.I. & Carbone, I., 2014. Challenges in analysis and interpretation of microsatellite data for population genetic studies. Ecology and Evolution 4, 4399-4428.

Rao, N., 2004. Plant genetic resources: Advancing conservation and use through biotechnology. African Journal of Biotechnology 3, 136-145.

Rickert, A.M., Premstaller, A., Gebhardt, C., Oefner, P.J., 2002. Genotyping of SNPs in a polyploid genome by pyrosequencing. BioTechniques 32, 592-603.

Rossetto, M., 2001. Sourcing of SSR markers from related plant species, in: Henry, R.J. (Ed.), Plant Genotyping: The DNA Fingerprinting of Plants. CABI, New York, USA, 211-224.

Rozen, S. & Skaletsky, H., 2000. Primer3 on the WWW for general users and for biologist programmers, in: Misener, S., Krawets, S. (Eds.), Methods in Molecular Biology. Humana, 365-386.

Sampson, J. & Byrne, M., 2012. Genetic diversity and multiple origins of polyploid *Atriplex nummularia* Lindl. (Chenopodiaceae). Biological Journal of the Linnean Society 105, 218-230.

SANBI, 2012. Red List of South African Plants. [ONLINE] Available at: http://redlist.sanbi.org/. [Accessed 8 August 2016].

Schutte, A.L., 1997. Systematics of the genus *Cyclopia* Vent. (Fabaceae, Podalyrieae). Edinburgh Journal of Botany 54, 125-170.

Stellenbosch University https://scholar.sun.ac.za

Selkoe, K. & Toonen, R.J., 2006. Microsatellites for ecologists: A practical guide to using and evaluating microsatellite markers. Ecology Letters 9, 615-629.

Silva, D.C., Sérgio, L., Duarte, C., Messias, J., 2012. DNA fingerprinting based on simple sequence repeat (SSR) markers in sugarcane clones from the breeding program RIDESA. African Journal of Biotechnology 11, 4722-4728.

Testolin, R., Marrazzo, T., Cipriani, G., Quarta, R., Verde, I., Dettori, M.T., Pancaldi, M., Sansavini, S., 2000. Microsatellite DNA in peach (*Prunus persica* L. Batsch) and its use in fingerprinting and testing the genetic origin of cultivars. Genome 43, 512-520.

Van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M., Shipley, P., 2004. MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes 4, 535-538.

Van Wyk, B.E., 1993. Nectar sugar composition in Southern African *Papilionoideae* (Fabaceae). Biochemical Systematics and Ecology 21, 271-277.

Waits, L.P., Luikart, G., Taberlet, P., 2001. Estimating the probability of identity among genotypes in natural populations: Cautions and guidelines. Molecular Ecology 10, 249-256.

Weising, K., Nybom, H., Wolff, K., Kahl, G., 2005. Methodology, in: DNA Fingerprinting in Plants: Principles, Methods and Applications. CRC, Florida, USA, 170-205.

Yasodha, R., Ghosh, M., Sumathi, R., Gurumurthi, K., 2005. Cross-species amplification of eucalyptus SSR markers in Casuarinaceae. Acta Botanica Croatica 64, 115-120.

Zane, L., Bargelloni, L., Patarnello, T., 2002. Strategies for microsatellite isolation: A review. Molecular Ecology 11, 1-16.

Zuur, A.F., Ieno, E.N., Smith, G.M., 2007. Principal coordinate analysis and non-metric multidimensional scaling, in: Gail, M., Krickeberg, K., Sarnet, J., Tsiatis, A., Wong, W. (Eds.), Analysing Ecological Data. Springer, New York, USA, 259–261.

Stellenbosch University https://scholar.sun.ac.za

Chapter 4: Population Genetic Diversity in the Polyploid *Cyclopia subternata*

Abstract

Cyclopia species are commercially utilised for honeybush tea production and currently the majority of plant material is harvested from the wild. Over-harvesting is of great concern and can lead to a loss of genetic diversity in wild populations. Conservation strategies, such as sustainable harvesting, can be implemented to protect *Cyclopia* species. Additionally, the cultivation of *Cyclopia* is already underway to relieve pressure placed on wild populations and an ARC field gene bank was established to secure as wide a genetic basis as possible for future applications. Knowledge on the population structure of wild *Cyclopia* is necessary to facilitate management strategies such as *in situ* conservation in the natural habitat, but thus far no such studies have been performed. This study successfully applied microsatellite markers and compared various population genetic statistics, without violating the assumption of disomic inheritance, in a polyploid species. The population structure of three wild populations of *C. subternata* was inferred and compared to the structure of the accessions that are currently in the ARC field gene bank. The field gene bank represented the majority of genetic diversity present in these wild populations, with the exception of the population sampled from Haarlem. The unique genetic structure of this population warrants further investigation and it is recommended that genetic material be captured for the field gene bank. The other two populations (Guava Juice and George) proved to be panmictic with a high amount of gene flow.

4.1 Introduction

Over-harvesting of wild plant populations for medicinal use is a common phenomenon that can lead to a loss of genetic diversity, thereby affecting the evolutionary potential of a species to adapt successfully to its environment (Cruse-Sanders and Hamrick, 2004). The cultivation of commercially important plants offers a solution to meet high demands and simultaneously relieve the pressure placed on wild populations. Not all species react to harvesting pressure in an identical manner. Species that have a narrow geographical range, small population sizes and are specifically adapted to their habitat will be highly susceptible to over-harvesting. Additional factors such as reproductive biology, slow growth rate and low germination rates will further determine how species are affected by over-harvesting (Schippmann *et al.*, 2006).

Stellenbosch University https://scholar.sun.ac.za

Cyclopia subternata is one of the commercially most important honeybush tea species and genetic improvement is underway at the Agricultural Research Council (ARC) where 25 elite selections, that have been evaluated and selected based on increased performance, are conserved in the field gene bank. The aim of the gene bank is to conserve advanced selections as well as plants from the wild and to provide the genetic resources for the breeding programme (Bester *et al*., 2013). Currently, 75% of the *Cyclopia* crop is wild-harvested to cater for demand, since not all species can be cultivated economically (Bester, 2012). *Cyclopia intermedia*, for instance, grows slowly and can only be harvested once every other year (Joubert *et al*., 2011). Additionally, as it is a resprouter, frequent harvesting can cause die-back of the plant as the nutrients are not given adequate time to be replenished (Barnardo thesis). Furthermore, *Cyclopia* seeds exhibit dormancy and low germination rates, even if environmental conditions are favourable (Koen *et al*., 2016). Such factors place some *Cyclopia* species at great risk and 9 of the 23 species are listed as vulnerable or endangered, including *C. longifolia* and *C. pubescens*. *C. subternata* is listed as declining and is therefore still of conservation concern (SANBI, 2012).

Cyclopia species have an endemic distribution range that spans the coastal and mountainous regions of the Eastern and Western Cape (Joubert *et al*., 2008; Joubert *et al*., 2011). Specifically, wild populations of *C. subternata* are localised in the Tsitsikamma, Outeniqua and Langeberge mountains, stretching from Riversdal to Plettenberg Bay and Port Elizabeth (Figure 4.1). The natural distribution of *C. subternata* overlap with that of *C. intermedia* and *C. longifolia*, as seen in Figure 4.1. Wild populations of *Cyclopia* are not well protected and an increase in demand has led to unsustainable harvesting practices (Bester, 2012). Sustainable harvesting can be adapted as a conservation method to prevent further decline and eventual extinction of the species (Schippmann *et al*., 2006; Mills, 2007).

To conserve the genetic diversity present in the wild populations for future breeding programmes, a thorough understanding of population structure and distribution is required for comparisons between cultivated and wild plants (Ford-Lloyd, 2001). One such aspect is the flow of genetic material from one population to another, whereby genotype frequencies will eventually change over time (Hamilton, 2009). The amount of gene flow will affect the potential for genetic differentiation among populations as well as their capacity for local adaptation (Keyghobadi *et al*., 2005). The ability of an organism to adapt successfully to changing environmental conditions to avoid extinction is important (Valladares *et al*., 2014). Abundant gene flow can lead to panmixia, where populations are genetically undifferentiated and are deemed a single entity (Allendorf and Luikart, 2007; Hamilton, 2009). Limited gene flow will often lead to genetic differentiation and loss of heterozygosity, thereby

inflating the occurrence of inbreeding (Mills, 2007). The exchange of genes is dependent on the dispersal ability of the species as well as the geographical distance and land barriers that isolate populations from each other (Frankham *et al*., 2002; Freeland, 2005).

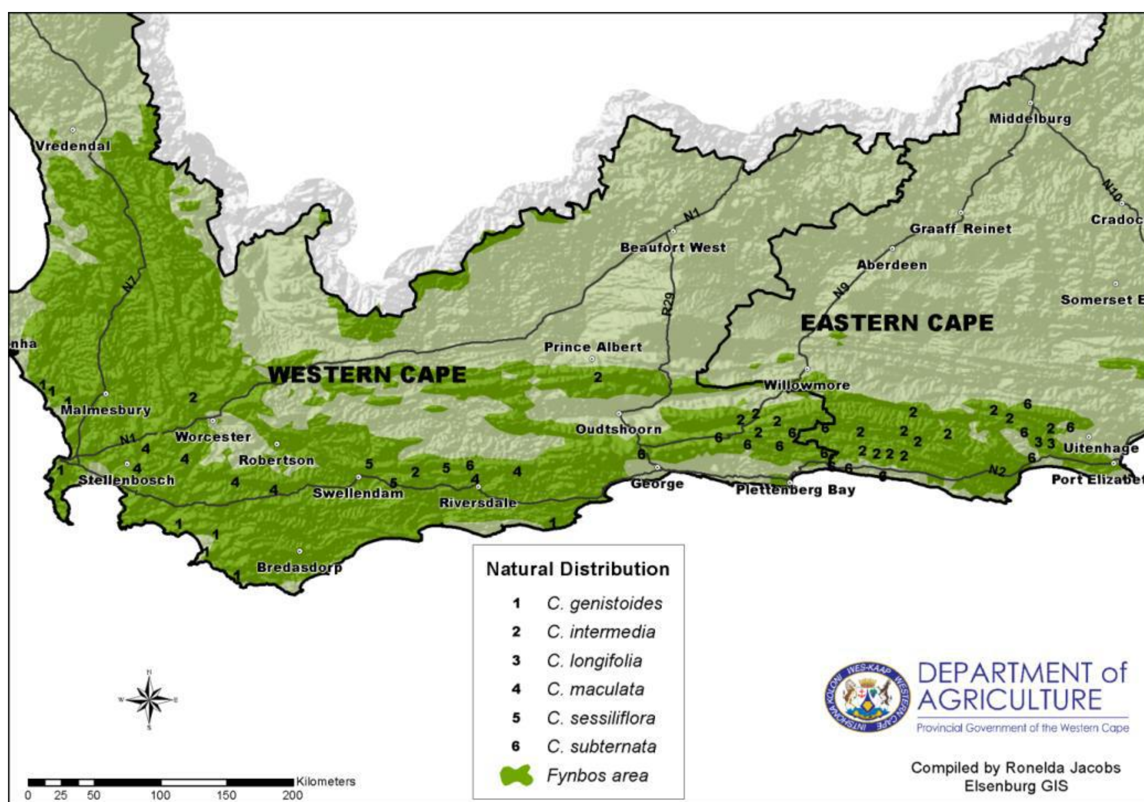


Figure 4.1. Natural distribution of six commercially important *Cyclopia* species in the Eastern and Western Cape of South Africa (Joubert *et al*., 2011). *Cyclopia subternata* is localised in the Tsitsikamma, Outeniqua and Langeberge mountains, stretching from Riversdal to Plettenberg Bay and Port Elizabeth (as denoted by a 6).

Previous studies indicate that *C. subternata* is a hexaploid with a chromosome base number $x = 9$ ($2n = 6x = 54$) (Goldblatt, 1981; Schutte, 1997) and this was confirmed by Motsa (2016). The mode of reproduction (outcrossing or self-fertilising) and inheritance patterns (polysomic or disomic) is still unknown and very little genetic information is available for this species. Furthermore, population genetic statistics for polyploids are limited as allele copy number is often ambiguous in polyploids (Sampson and Byrne, 2012). Dufresne *et al*. (2014) thoroughly reviewed several extensions of diploid statistics to polyploid organisms such as pairwise genetic distance measures and *Rho* as an alternative to F-statistics.

The aim of this study is to investigate the population genetic structure of three wild populations of *C. subternata*. Two populations are from the Western Cape (Haarlem and George) while the other population is from the Eastern Cape (Guava Juice). Several population statistics are applied and

Stellenbosch University https://scholar.sun.ac.za

compared for a small-scale population genetic study of this polyploid species. The conclusions may have implications for conservation of the species, dependant on the amount of genetic differentiation present (panmixia versus discrete populations). In addition, it compares the genetic diversity of field gene bank of *C. subternata* to wild populations to determine how much genetic diversity has been captured in the ARC field gene bank.

4.2 Materials and Methods

4.2.1 Sampling and DNA extraction

For the cultivated samples, 22 *C. subternata* field gene bank accessions were sampled from the ARC Nietvoorbij and Elsenburg Research Farms near Stellenbosch. The origin of where the samples were evaluated includes regions such as Kanetberg, Tolbos and Haarlem from the Western Cape and Groendal from the Eastern Cape, which can be seen in the Appendix, Table A3. Three wild *C. subternata* populations were chosen due to their accessibility and the availability of the plants with the advice of local farmers who harvest the populations for vleitee production. Sites included one area from the Eastern Cape - Guava Juice (population 1 with 30 individuals sampled), and two locations from the Western Cape - Haarlem (population 2 with 30 individuals sampled) and George (population 3 with only 24 individuals in the population to sample) (Figure 4.2).

Samples collected in the wild had to be preserved in a buffer (Chapter 2) since sampling took place in remote locations. A CTAB/NaCl preservation buffer was prepared in advance in the laboratory by dissolving 60g CTAB and 700g NaCl in 2000ml of distilled water according to a protocol by Rogstad (1992). One hundred 15ml tubes were filled halfway with the buffer for subsequent wild sampling. Sampling was conducted during March 2015 and shoots containing adequate young leaves were picked by a gloved hand and placed in the tubes containing CTAB. Upon completion of the sampling at each location, the tubes were filled up with the remaining buffer using a 10ml pipette. Shoots were kept in the CTAB buffer for up to two weeks before the leaves were washed with distilled water in the laboratory and the DNA extraction protocol was performed as described in Chapter 3 (section 3.2.2).

Stellenbosch University https://scholar.sun.ac.za

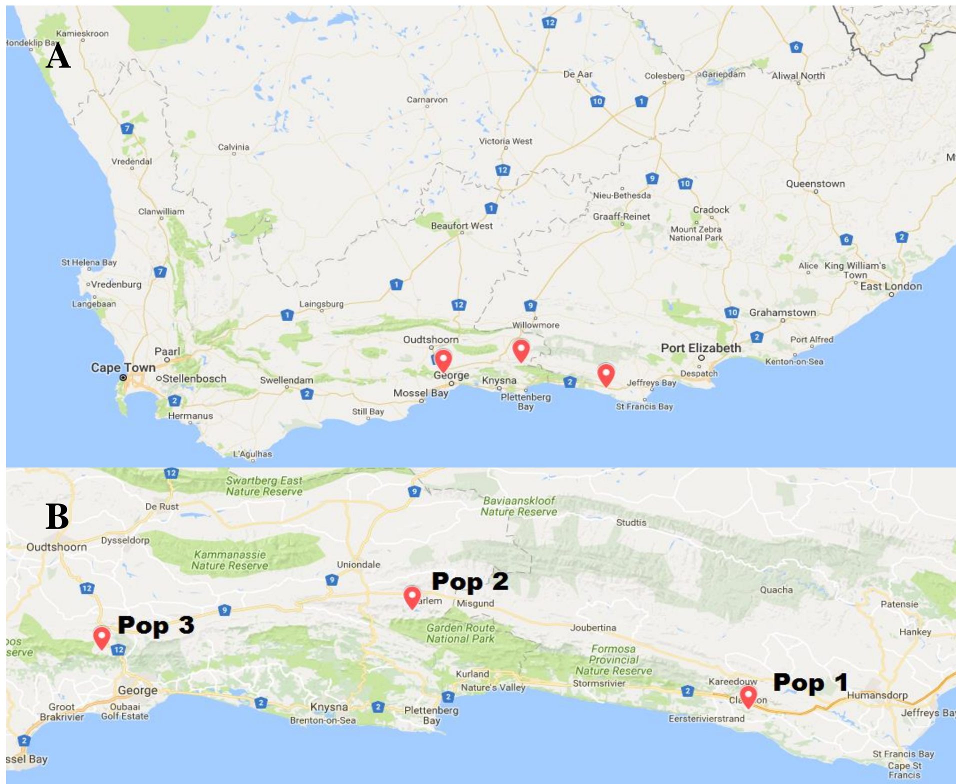


Figure 4.2. Maps showing the locations of the three wild populations sampled of *C. subternata*. Map A indicates the locations of the populations in South Africa, while map B focuses on the exact areas. Population 1 from Guava Juice (S 34.04595; E 024.34648) is in the Eastern Cape and populations 2 from Haarlem (S 33.777688; E 023.304393) and 3 from George (S 33.88580; E 022.34306) are in the Western Cape.

4.2.2 PCR amplification

DNA was amplified using the primer pairs from two honeybush-specific multiplex panels for six microsatellite loci (see Table 4.1 in section 4.3.1 for primer detail). A touchdown PCR cycling protocol was adapted and optimised for use with the Qiagen multiplex kit. PCR reactions were prepared in a total volume of 25µl with 50ng of genomic DNA, 1X QIAGEN Multiplex PCR Master Mix (HotStarTaq® DNA polymerase, Multiplex PCR Buffer with 6mM MgCl2 and dNTP Mix), 0.2µM of the forward and reverse primer (excluding primers Cys10 and Cys22, for which only 0.1µM was added) (Applied Biosystems) and RNase free water. The cycling conditions were as follows: 95°C for 15mins followed by a denaturing step for 30s at 94°C, annealing for 90s with 2 cycles each at 64-59°C, 10 cycles at 58°C and 10 cycles at 57°C. An elongation step was performed at 72°C for 90s followed by a final extension step of 10mins at 72°C.

PCR products were run on a 2% (m/v) agarose gel with 1X TBE and 7.5µl of ethidium bromide (10mg/ml) for initial screening and only products that yielded visible bands were further prepared for

Stellenbosch University https://scholar.sun.ac.za

sizing by capillary electrophoresis to correct for off-scale peaks using the following procedure. The PCR product was cleaned at CAF using the Nucleofast® 96 PCR clean-up kit (Macherey-Nagel) after which the DNA quantity and quality was measured using a NanoDrop ND-1000 spectrophotometer (Thermo-Fischer). The DNA was diluted to an end concentration of 12-15ng/µl and the products sized using an ABI 3730 capillary sequencer (Applied Biosystems).

4.2.3 Genotyping

GeneMapper 4.0 software (Applied Biosystems) was used to detect peaks and estimate product sizes. To minimise genotyping errors, several strategies were applied in accordance with the recommendation of Bonin *et al*. (2004). A negative control was included during PCR amplification to monitor contamination. Any individuals that had missing genotypes, caused by problems with amplification and repeatability, for more than one out of the six markers (more than ~10% missing data) were excluded from the study. The *C. subternata* wild population data set of 84 individuals was thus reduced to 79 individuals; where population 2 had two individuals with missing genotypes (S44 and S45) and population 3 had three (S72, S78, S84). A previously typed sample (either SKB3, SKB4 or SGD7) was included as a positive control during amplification for calibration of data scoring.

During data analysis in GeneMapper, automated scoring was performed using a maximum of six alleles and data were checked manually. The six markers chosen for this study yielded polymorphic products showing limited or no stutter bands in any of the *C. subternata* samples. This is especially important for polyploids to make the process of allele scoring easier (Pfeiffer *et al*., 2011).

No more than six peaks are expected in any of the individuals, as the ploidy level of *C. subternata* was recently confirmed to be hexaploid (Motsa, 2016). A data coding method previously employed for polyploids was applied (Jørgensen *et al*., 2008; Kloda *et al*., 2008; Sampson and Byrne, 2012; Teixeira *et al*., 2014). Multilocus allelic profiles were generated by combining all alleles seen at each locus. As explained, these allelic phenotypes are not genotypes, since allele dosage could not be determined, but indicate the absence or presence of the allele. The multilocus data were transformed into binary arrays where an allele was either recorded as absent (0), present (1) or missing (-9) using POLYSAT 1.4 (Clark and Jasieniuk, 2011). Allele dosage ambiguity also meant that testing for Hardy-Weinberg equilibrium was not possible, since allele frequencies, which cannot be calculated for the polyploid species, is a prerequisite (Dufresne *et al*., 2014).

Stellenbosch University https://scholar.sun.ac.za

4.2.4 Data analysis

Data were treated as dominant and several calculations were performed from the binary matrices. Allelic diversity was measured as the number of alleles per locus, averaged over loci (N_A). Private alleles per population were calculated in GenAlEx 6.501 using the frequency option for binary data (Peakall and Smouse, 2006). The maximum number of alleles was determined and correlated with the ploidy data.

The polymorphic information content (PIC) is a measure where the usefulness of each marker is determined. This was calculated per marker:

$PIC = 1 - \sum pi^2$, where the squared frequency of allele *i* is represented as pi (Powell *et al*., 1996).

A second measure, the discriminatory power (D_j), was used for comparison of the usefulness of each marker.

$$D_j = 1 - C_j$$

where C_j is the confusion probability for each marker:

$$C_j = \sum_{i=1}^{I} pi \frac{(N.pi-1)}{N-1}$$

The same pi (frequency of allele *i*) was applied in the formula as for calculation of PIC where *N* equals the number of individuals.

The data for all the *C. subternata* wild individuals as well as gene bank accessions were combined and pairwise genetic distances between individuals were calculated according to three methods for comparison (see Appendix, Table A4 for detail):

1. A Euclidian distance measure implemented in GenAlEx based on the binary data.
2. Lynch distance in POLYSAT based on the original genotypic data matrix.
3. Bruvo distance in POLYSAT based on the original genotypic data matrix.

An Analysis of Molecular Variance (AMOVA) was used to examine molecular variance within and between the populations. The pairwise genetic distance matrices were utilised as input data and ϕ-

Stellenbosch University https://scholar.sun.ac.za

statistics, analogous to the population differentiation measure F_{ST}, were computed to calculate the proportion of variance among populations relative to total variance. This measure allows for comparison between binary and co-dominant data (Teixeira *et al*., 2014). The statistical significance of the variance components was assessed using a permutation method in GenAlEx. To test for isolation by distance (IBD), a Mantel test (Mantel, 1967) was performed with 999 permutations in GenAlEx using the correlation between [ϕ_{PT} / (1 − ϕ_{PT})] and the natural logarithm of geographic distance between populations.

RhoST (ρ_{ST}) is another analogue to F_{ST} that can be applied specifically to polyploids since it is independent of selfing rate, double reduction, ploidy level and mode of inheritance (Ronfort *et al*., 1998; Meirmans and Van Tienderen, 2013; Brandrud, 2014). Pairwise ρ_{ST} was calculated in SPAGeDi 1.5 and statistical significance was tested using 10 000 permutations (Hardy and Vekemans, 2002). The input file for SPAGeDi was created in POLYSAT.

To visualise the relationships among individuals, principal coordinate analysis (PCoA) plots based on the genetic distance matrices were constructed in GenAlEx. PCoA is a multivariate method, based on any chosen distance or dissimilarity measure that can reduce complex data sets to a lower dimension and plot them on a set of axes for visualisation (Anderson and Willis, 2003; Zuur *et al*., 2007). This method was previously employed by Kloda *et al*. (2008) to visualise relationships among diploid and tetraploid *Ononis* species (Fabaceae) and the lack of assumptions regarding population genetics make it ideal for polyploids.

Population structure was investigated using a Bayesian clustering method in the software STRUCTURE 2.3.4 (Pritchard *et al*., 2000). The software can handle allele copy number ambiguities, which are often found in polyploid data analyses. The STRUCTURE input file for the three wild populations and the accessions was generated using POLYSAT software (Clark and Jasieniuk, 2011). STRUCTURE analysis was run without prior assumptions of population structure using the admixture ancestry model and the recessive allele (=1) model to account for ambiguities. The optimum K, or ancestral genetic clusters, was determined from 10 replicate runs for each value of K (chosen as 1-4) with the length of burn-in period set to 200 000 and the Markov chain Monte Carlo (MCMC) repeats to 2 000 000. The STRUCTURE results were used as an input file for the Structure Harvester 0.6.94 website (http://taylor0.biology.ucla.edu/structureHarvester/) to find the optimum K using the method of Evanno *et al*. (2005) that generates likelihood values across different K values (Earl and vonHoldt, 2012). The highest delta K value was regarded as the true number of clusters and was visualised on a plot. The STRUCTURE results file was also subjected to analysis by CLUMPAK,

Stellenbosch University https://scholar.sun.ac.za

which is a post-processing program that can cluster the multiple STRUCTURE runs to visualise a consensus bar plot of the results (Kopelman *et al*., 2015).

4.3 Results

4.3.1 Genetic diversity

The number of alleles per locus across the three populations and the set of accessions ranged from 12 (Cys10) to 22 (Cys5) (Table 4.1). The maximum number of alleles for each locus per individual ranged from 4 to 6, consistent with *C. subternata* being hexaploid (Table 4.1 and 4.2). A total of 98 alleles were detected across all 6 loci with 23 private alleles found in the wild populations (23.5%) and only 7 found in the accessions (7.1%). All six markers were polymorphic with high PIC values ranging from 0.80 (population 2) to 0.87 (accessions) (Table 4.2). The discriminatory power of the markers (D_j) followed a similar trend with values ranging from 0.83 (population 2) to 0.91 (accessions). The distribution of alleles can be seen in Figure 4.3 where the wild populations and cultivated accessions share most of the alleles. For markers Cys5, Cys21b and Cys22 the wild populations often had private alleles, whereas for Cys88 the private alleles were seen in the accessions.

4.3.2 Genetic differentiation and clustering among wild populations

Analyses of Molecular Variance (AMOVA) based on the three distance measures yielded similar results. The greatest amount of variance (~90%, P = 0.001) was found within the three wild populations, as opposed to ~10% between population variance (Table 4.3A). Pairwise ϕ_{PT} values were found to be significant between all populations for all three distance measures ($P < 0.05$, Table 4.4). The binary and Lynch measures yielded similar results. Higher ϕ_{PT} values were observed when population 2 was compared to population 1 (Lynch $\phi_{PT} = 0.125$, P = 0.001 and binary $\phi_{PT} = 0.123$, P = 0.001) and population 3 (Lynch $\phi_{PT} = 0.103$, P = 0.001 and binary $\phi_{PT} = 0.101$, P = 0.001) than when population 1 and 3 were compared (Lynch $\phi_{PT} = 0.058$, P = 0.001 and binary $\phi_{PT} = 0.059$, P = 0.001). The Bruvo distance measure yielded slightly different pairwise ϕ_{PT} values than the Lynch and binary methods. Values were similar when population 2 was compared to both populations 1 and 3 (population 1, $\phi_{PT} = 0.106$, P = 0.001 and population 3, $\phi_{PT} = 0.108$, P = 0.001). The ϕ_{PT} values correlate with the ρ_{ST} values obtained where higher values could be observed between populations 1 and 2 ($\rho_{ST} = 0.193$, $P < 0.001$) and between populations 2 and 3 ($\rho_{ST} = 0.145$, $P < 0.001$) than between populations 1 and 3 ($\rho_{ST} = 0.064$, $P < 0.001$).

Stellenbosch University https://scholar.sun.ac.za

Table 4.1. Primer information for six *C. subternata* microsatellite markers in three wild populations and 22 accessions.

MP	Locus	F/R	Primer sequence (5'-3')	T_M	Repeat	Range	Label	N_A	Max
HB-Panel 1	Cys21b	F	AAAGTGTTCGTAGTTTGCTTC	54	CT^{17}	90-160	PET	18	6
	Cys21b	R	GTACTAGAATAGCTGCAAGACA	52					
	Cys22	F	GCGTGAGGACAGAGACACAA	60	AG^{12}	76-140	FAM	16	4
	Cys22	R	GACGGCGGAAAAGAAAATAA	60					
	Cys25	F	TGAATGGATGAGTGGCAAGA	60	$GT^{10}C^{1}AG^{6}$	153-233	FAM	13	4
	Cys25	R	TCACCAGCAGATTCAGTTGAG	60					
HB-Panel 2	Cys5	F	CGTGTGTGTGTGTGTGAGAAA	60	AG^{14}	100-160	FAM	22	4
	Cys5	R	GCGTTAGCGAAGGACTAGCT	60					
	Cys10	F	TTGTCATGTTCTCACGCTGTT	60	TC^{8}	90-130	VIC	12	5
	Cys10	R	TCTCTGCACACAAAAATCACG	60					
	Cys88	F	TTTGCCTGCTCAGAAATCCC	60	GTT^{7}	100-160	PET	17	6
	Cys88	R	TGTTTTCTTCTTGTTTTCAATCCA	60					

MP = multiplex; F/R = forward or reverse primer; T_M = melting temperature; N_A = number of alleles per locus, averaged over loci; Max = maximum alleles per individual per locus.

Table 4.2. Allele diversity statistics for three wild populations of *C. subternata* as well as cultivated accessions.

C. subternata	N_A	Total alleles	Private alleles	Max bands	Mean bands	PIC	D_j
Accessions	13	62	7	6	3	0.87	0.91
Population 1	12	83	6	5	3	0.86	0.89
Population 2	10	61	5	5	2	0.80	0.83
Population 3	10	61	6	6	3	0.85	0.89

N_A = number of alleles per locus, averaged over loci; Maximum number of alleles per individual per locus; Mean number of alleles per individual per locus; PIC = polymorphic information content; D_j = Discriminatory power.

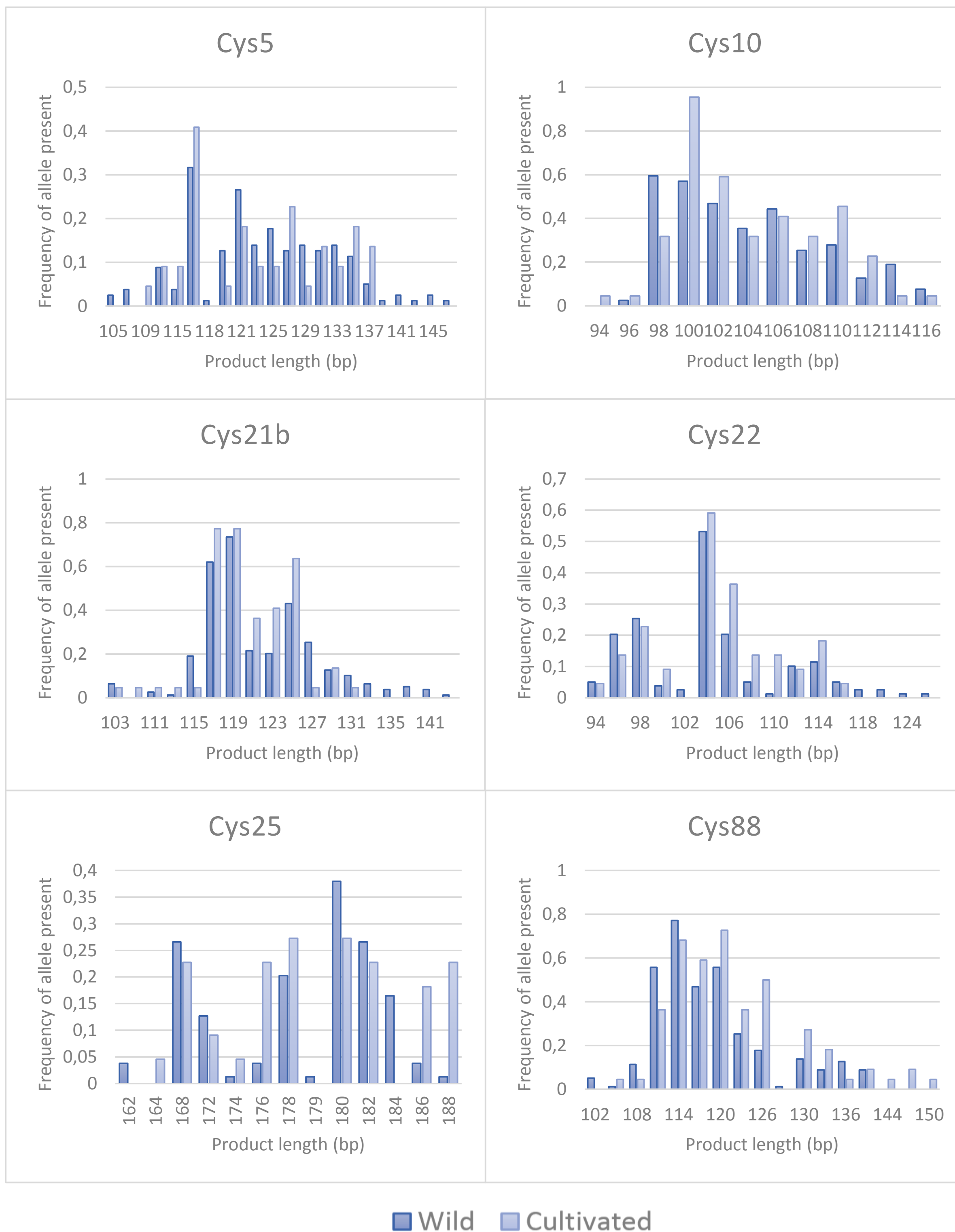


Figure 4.3. Chart depicting the presence of alleles for six microsatellite loci (Cys5, Cys10, Cys21b, Cys22, Cys25 and Cys88) in 79 individuals of wild populations and 22 gene bank accessions of *C. subternata*. Note that this is not the distribution of alleles, since allele frequencies could not be calculated for this hexaploid species.

Stellenbosch University https://scholar.sun.ac.za

Clustering patterns revealed similar results to the pairwise ϕ_{PT} and ρ_{ST} comparisons. The Guava Juice and Haarlem populations clustered separately from each other in all three distance measures (Figure 4.4). In the Lynch and binary distance PCoA plots, population 3 was dispersed among populations 1 and 2. The Bruvo distance PCoA plot clustered population 3 with population 1, but quite separately from population 2. Furthermore, there was a weak but significant correlation between genetic and geographic distance ($r^2 = 0.0685$, $P = 0.001$, data not shown).

4.3.3 Genetic differentiation and clustering between wild populations and accessions

Analysis of Molecular Variance (AMOVA) based on the three distance measures yielded identical results. The greatest amount of variance (~92%, $P = 0.001$) was found within the populations and accessions, compared to ~8% between populations (Table 4.3B).

Pairwise ϕ_{PT} values were found to be significant between all populations and accessions for all three distance measures ($P < 0.05$, Table 4.4). The Bruvo and Lynch measures yielded similar results. Higher ϕ_{PT} values were observed when population 2 was compared to the accessions (Lynch $\phi_{PT} = 0.097$, $P = 0.001$ and Bruvo $\phi_{PT} = 0.081$, $P = 0.001$) than when populations 1 and 3 were compared with the accessions (population 1 Lynch $\phi_{PT} = 0.022$, $P = 0.009$ and Bruvo $\phi_{PT} = 0.018$, $P = 0.008$; and population 3 Lynch $\phi_{PT} = 0.033$, $P = 0.007$ and Bruvo $\phi_{PT} = 0.033$, $P = 0.003$). The binary distance measure yielded slightly different pairwise ϕ_{PT} values. The values were lower when comparing the accessions to population 1 ($\phi_{PT} = 0.020$, $P = 0.011$) and population 3 ($\phi_{PT} = 0.032$, $P = 0.005$) than when comparing the accessions with population 2 ($\phi_{PT} = 0.117$, $P = 0.001$).

The ϕ_{PT} values correlate with the ρ_{ST} values. Lower values were observed between population 1 and the accessions ($\rho_{ST} = 0.029$, $P = 0.006$) and between population 3 and the accessions ($\rho_{ST} = 0.036$, $P = 0.012$) than between population 2 and the accessions ($\rho_{ST} = 0.122$, $P < 0.001$).

PCoA clustering patterns revealed results similar to those from the pairwise ϕ_{PT} and ρ_{ST} comparisons (Figure 4.4). The accessions (yellow diamonds) clustered mostly with the Guava Juice (red squares) and George (blue triangles) population in quadrant 2 and 4 of the PCoA plot. The Haarlem population (green circles) grouped separately in quadrant 1 and 3 using all three distance measures. The binary and Lynch PCoA plots had some of the blue individuals dispersed across quadrant 1 and 3 with the Haarlem population, whereas in Bruvo's method, the individuals showed tighter clustering.

Stellenbosch University https://scholar.sun.ac.za

STRUCTURE analysis partitioned all individuals into two clusters (K = 2) (Figure 4.5). The majority of individuals had admixed ancestry drawn from both clusters (Figure 4.6). Population 2 was assigned mostly to cluster 1 (orange) while populations 1 and 3 and the accessions were assigned mostly to cluster 2 (blue). This corroborates the PCoA results where population 2 grouped separately from the other populations and accessions. The STRUCTURE plot of K = 3 was also included for comparison, while the minor cluster K = 4 was excluded as it did not reflect the true ancestry of the various sites. Again, population 2 had a distinct composition with low levels of admixture.

Table 4.3. Partitioning of the observed *C. subternata* microsatellite variation (A) within and among three wild populations and (B) within and among wild populations and accessions using three genetic distance measures appropriate for polyploid data. P-values indicated at the 5% significance level.

	Distance measure	Source	df	Est. Var.	%	ϕ	P-value
A)		Among Pops	2	0.029	9%	0.091	0.001
	Bruvo	Within Pops	76	0.286	91%		0.001
		Total	78	0.314	100%		
		Among Pops	2	0.035	10%	0.099	0.001
	Lynch	Within Pops	76	0.315	90%		0.001
		Total	78	0.349	100%		
		Among Pops	2	0.996	10%	0.097	0.001
	Binary (Euclidian)	Within Pops	76	9.306	90%		0.001
		Total	78	10.302	100%		
B)		Among Pops	3	0.022	7%	0.070	0.001
	Bruvo	Within Pops	97	0.289	93%		0.001
		Total	100	0.311	100%		
		Among Pops	3	0.027	8%	0.078	0.001
	Lynch	Within Pops	97	0.318	92%		0.001
		Total	100	0.345	100%		
		Among Pops	3	0.811	8%	0.079	0.001
	Binary (Euclidian)	Within Pops	97	9.483	92%		0.001
		Total	100	10.294	100%		

Stellenbosch University https://scholar.sun.ac.za

Table 4.4. Pairwise ϕ_{PT} and ρ_{ST} values between three wild populations and cultivated accessions of *C. subternata*. Significant P-values (5% level) are indicated above the diagonal.

ϕ_{PT} (GenAlEx): Bruvo				
Pop1	Pop2	Pop3	Accessions	
	0.001	0.001	0.008	Pop1
0.106		0.001	0.001	Pop2
0.053	0.108		0.003	Pop3
0.018	0.081	0.033		Accessions
ϕ_{PT} (GenAlEx): Lynch				
Pop1	Pop2	Pop3	Accessions	
	0.001	0.001	0.009	Pop1
0.125		0.001	0.001	Pop2
0.058	0.103		0.007	Pop3
0.022	0.097	0.033		Accessions
ϕ_{PT} (GenAlEx): Binary (Euclidian)				
Pop1	Pop2	Pop3	Accessions	
	0.001	0.001	0.011	Pop1
0.123		0.001	0.001	Pop2
0.059	0.101		0.005	Pop3
0.020	0.117	0.032		Accessions
ρ_{ST} (SPAGeDi)				
Pop1	Pop2	Pop3	Accessions	
	< 0.001	< 0.001	0.006	Pop1
0.193		< 0.001	< 0.001	Pop2
0.064	0.145		0.012	Pop3
0.029	0.122	0.036		Accessions

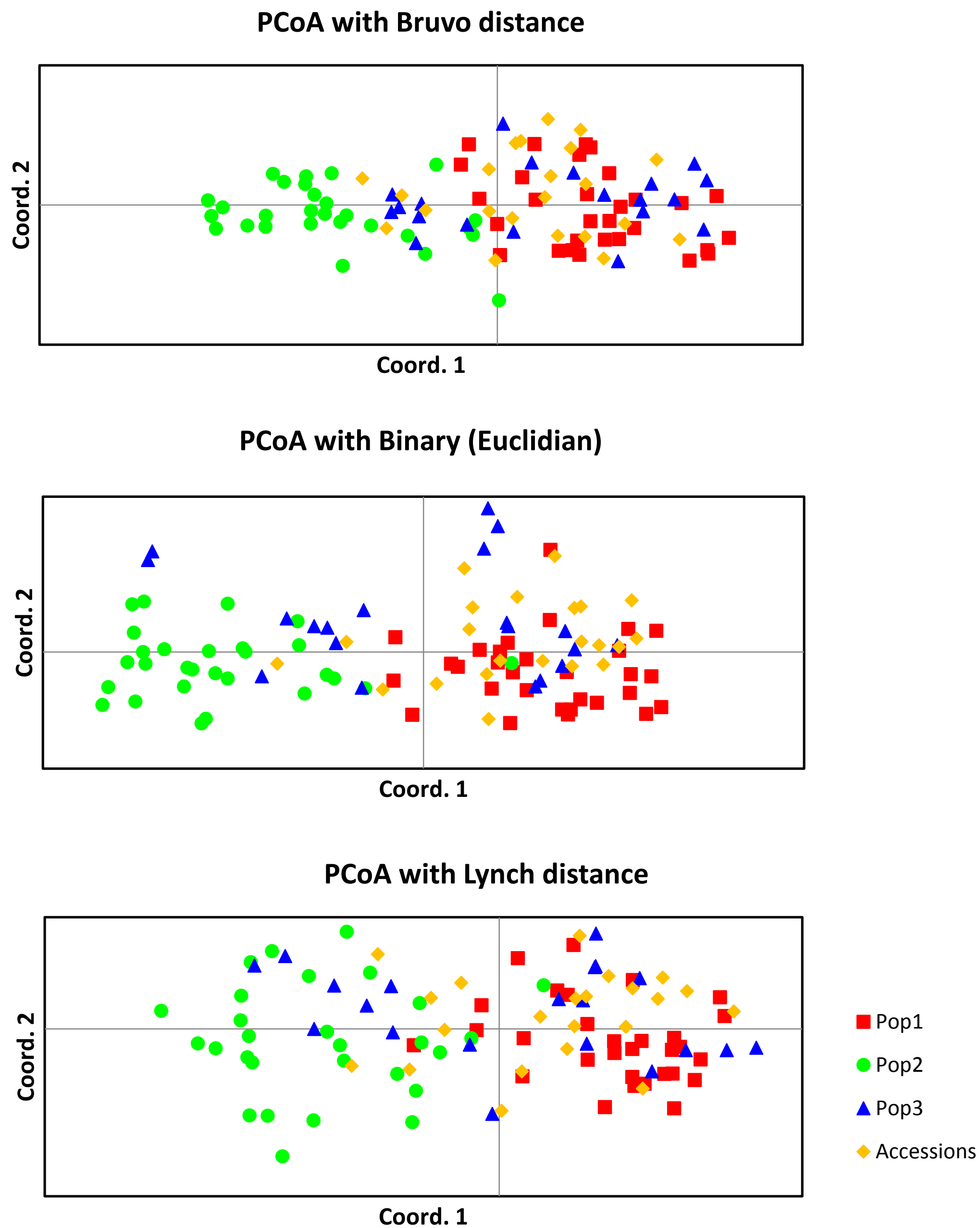


Figure 4.4. Principal coordinate analysis plots based on three genetic distance matrices for three *C. subternata* wild populations and gene bank accessions. Population 1 was sampled from Guava Juice in the Eastern Cape, population 2 is from Haarlem in the Western Cape and population 3 is from George in the Western Cape.

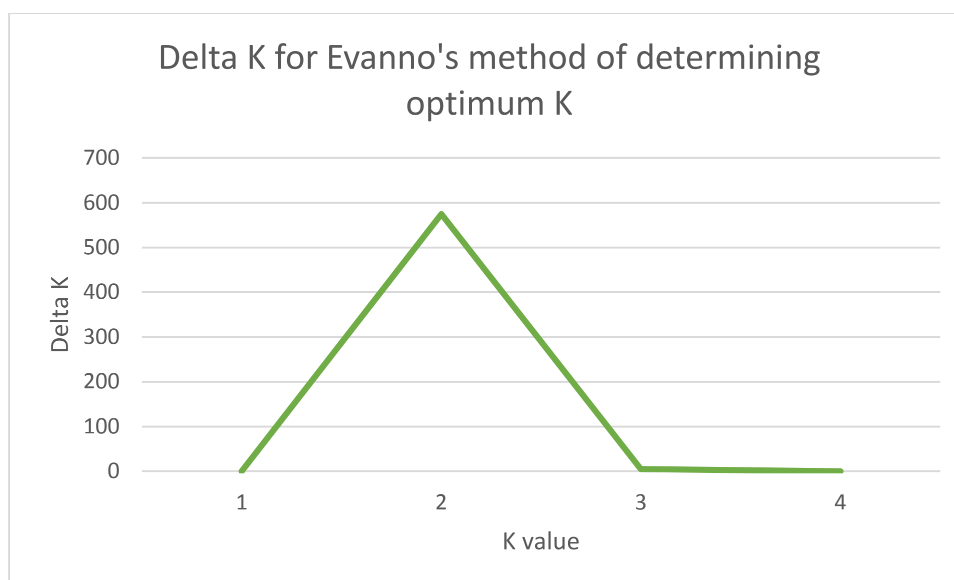


Figure 4.5. The optimum K for populations of *C. subternata* was chosen as 2 according to Evanno's method implemented in Structure Harvester.

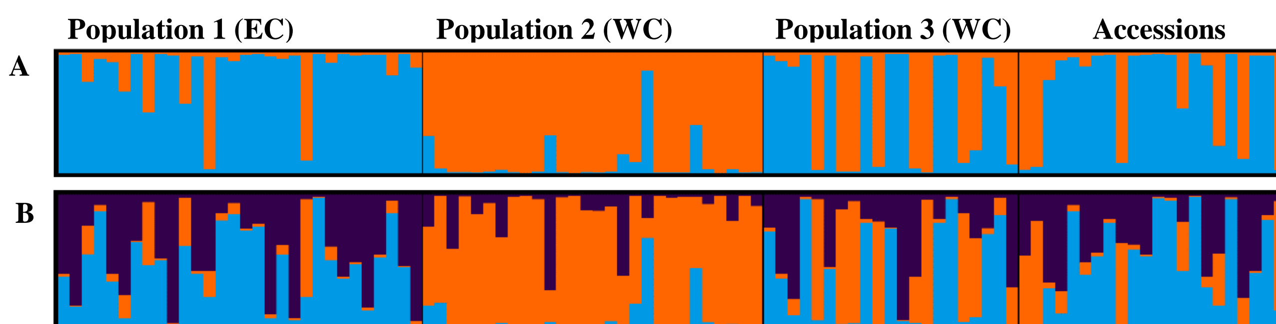


Figure 4.6. STRUCTURE bar plots with (A) K = 2 chosen as the most likely number of clusters and (B) K = 3 (included for comparison) for cultivated accessions and three wild populations of *C. subternata*. The K = 4 minor cluster was not included as this did not depict suitable ancestry. Population 1 was sampled from Guava Juice in the Eastern Cape, population 2 is from Haarlem in the Western Cape and population 3 is from George in the Western Cape.

4.4 Discussion

4.4.1 Comparison of various population genetic statistics and software for polyploids

This is the first population genetic study of the polyploid *C. subternata* using microsatellites. The problem of allele dosage in a polyploid, such as the focal species, could perhaps have been addressed by using the MAC-PR (microsatellite DNA allele counting - peak ratios) method (Esselink *et al.*, 2004). This has been applied successfully to several low order polyploid species, such as tetraploids to resolve allele dosage (Pfeiffer *et al.*, 2011). However, in higher order polyploids (≥ 6*x*), such as

Stellenbosch University https://scholar.sun.ac.za

C. subternata, this method is often not feasible since microsatellite peak ratios are analysed in pairwise comparisons and at least one fully heterozygous individual per locus is necessary for a baseline comparison to determine the copy number in each individual (Esselink *et al*., 2004; Caruso *et al*., 2010). Allele dosage determination was therefore not attempted in this study and microsatellite output was alternatively interpreted as allele phenotypes. Although a loss of genetic information is expected, polyploids exhibit higher allelic diversity than diploids, since variation is introduced by the formation of multiple sets of chromosomes (Soltis and Soltis, 1993; Adams and Wendel, 2005; Eliášová *et al*., 2014). This is demonstrated by the following: the maximum number of possible genotypes (gt) increases proportionally to the ploidy level and can be determined by $\frac{(2n-1)!}{n!\,(n-1)!}$, where n denotes the ploidy level. The number of allele phenotypes can similarly be described as pt = $2^n - 1$. Where a diploid individual has gt = pt = 3, a hexaploid individual can display 462 possible genotypes and 63 possible allelic phenotypes (Pfeiffer *et al*., 2011).

In the present study, similar trends were observed for both ϕ_{PT} and ρ_{ST} values, with ϕ_{PT} ranging from 0.018-0.125, while ρ_{ST} ranged from 0.029-0.193. These measures are analogues of F_{ST} and both have been successfully used in previous studies to test for population differentiation in polyploids (Clarke *et al*., 2012; Brandrud, 2014; Dennhardt *et al*., 2016). Meirmans and Van Tienderen (2013) recommend the use of *Rho* (ρ) in polyploid population genetic studies where the mode of inheritance is not known. However, ρ_{ST} values should be interpreted differently from F_{ST}, since it yields consistently higher values that range from zero (no population differentiation) to one (complete population differentiation) (Ronfort *et al*., 1998; Wiesner *et al*., 2011). A similar trend for ϕ_{PT} values have been observed in previous studies. Teixeira *et al*. (2014) reported ϕ_{PT} values that ranged from 0.004-0.127 and concluded that there was genetic differentiation between Moroccan and Spanish *Juniperus thurifera* populations. Similarly, Clarke *et al*. (2012) reported values that ranged from 0.009-0.214 for an *Atriplex* species. A study by Dennhardt (2016) reported moderate genetic differentiation between wild populations of grass species with ρ_{ST} values ranging from 0.023-0.224.

Genetic distance quantifies the degree of similarity between groups of individuals and was calculated by means of several different distance measures (Lowe *et al*., 2004). Euclidian distance is based on the shared presence and absence of alleles expressed as binary data; therefore, it can be applied independently of ploidy level.

Lynch distances disregard the shared absence of a band as a similarity between individuals and use allele identity alone to calculate distances. This method is preferred for loci that do not conform to a

Stellenbosch University https://scholar.sun.ac.za

stepwise mutational model (Clark and Jasieniuk, 2011; Dufresne *et al*., 2014). Bruvo distances are applicable to autopolyploids, where alleles are expected to be closely related by mutation, as they are derived from the same genome (Clark and Jasieniuk, 2011; Clarke *et al*., 2012). This measure can, however, be applied to allopolyploids if no alleles at the two isoloci are within five mutational steps of each other (Clark and Jasieniuk, 2011). Although it is not known whether *C. subternata* is an allopolyploid or autopolyploid, the Bruvo distance was included for comparison to the other two distance measures. Of the three measures used, Lynch and Euclidian are both similarity indices, whereas Bruvo is a distance index. The relationship between similarity and distance can be expressed with the equation: similarity = 1 - distance (Dufresne *et al*., 2014).

Similar clustering patterns were revealed for the various genetic distance measures, in which the Haarlem population individuals (population 2) mostly clustered separately from the other two wild populations. Bruvo distance takes mutational relationships between alleles into account and a tighter clustering pattern than the other two distance measures was observed. This signifies that individuals might not necessarily share identical alleles, but that alleles may be closely related by mutational events (Clark and Jasieniuk, 2011). Very few studies have compared various genetic distance measures in polyploid organisms. Clarke *et al*. (2012) compared Bruvo and Lynch for an *Atriplex* species (Amaranthaceae) where the individuals were more interspersed in the Bruvo PCoA plot. This was attributed to the possibility that the loci did not conform to a stepwise mutational model as explicitly assumed by this distance measure. Assoumane *et al*. (2013) reported that, even though the tetraploid *Acacia senegal* (Fabaceae) is thought to be allopolyploid, the Bruvo distance successfully separated the various populations.

Bayesian clustering has been applied successfully in several polyploid studies where the mode of inheritance is not certain (Sampson and Byrne, 2012; Teixeira *et al*., 2014; Zawedde *et al*., 2015). Structuring is, however, based on Hardy-Weinberg assumptions and violations could lead to inconsistent population clustering. The extent of the effect that violations have on spurious clustering is still unknown (Dufresne *et al*., 2014). Since the STRUCTURE results corroborate the findings of the multivariate clustering and AMOVA methods, it was deemed suitable for use in the current study.

Software for polyploid studies is extremely limited and only a handful, such as SPAGeDi, STRUCTURE, POPDIST and GENODIVE, can accommodate polyploid data. GENODIVE (Meirmans and Van Tienderen, 2004) can also calculate Bruvo distance and additional features including the testing of deviations from Hardy-Weinberg Equilibrium (HWE) (Dufresne *et al*., 2014). This software is, however, restricted to a Mac OS which unfortunately limits its use. POLYSAT is

Stellenbosch University https://scholar.sun.ac.za

an extremely useful software program for conversion and analysis of polyploid data (Clark and Jasieniuk, 2011). Various previous studies successfully utilised this software to calculate pairwise genetic distances (Lynch or Bruvo), construct PCoA plots, calculate allele frequencies and estimate the ploidy level based on the maximum number of alleles per individual (Clarke *et al*., 2012; Sampson and Byrne, 2012; Brandrud, 2014; Dennhardt *et al*., 2016). Furthermore, POLYSAT can import GeneMapper files and allows for the conversion of data sets to several different formats for easy import into software such as STRUCTURE, SPAGeDi and POPDIST. A useful application in POLYSAT was the conversion of genotype data to a binary array and subsequently writing a table that can be edited in Microsoft Excel (Microsoft, Redmond, Washington).

4.4.2 Accessions and capturing of genetic variation

The natural distribution of *C. subternata* spans across the coastal mountain ranges of the Eastern and Western Cape. The *C. subternata* accessions currently located in the ARC gene bank originate from wild populations where seeds were collected by J. H. de Lange and planted for evaluation and selection in various regions that included Kanetberg, Tolbos and Haarlem from the Western Cape and Groendal from the Eastern Cape (see Table A3 in Appendix). STRUCTURE results revealed high levels of admixture among the accessions and the wild populations from George and Guava Juice. The clustering of accessions with these two wild populations indicates that the collection of plant material for the germplasm collection is well representative of the genetic diversity present in these two wild populations. Some natural populations are situated in hard-to-reach areas, including mountain slopes that can only be reached by off-road vehicles and experienced hikers. Additionally, at the time the wild seeds were collected by J. H de Lange, several populations were still unknown and it is therefore possible that some *C. subternata* wild populations are underrepresented in the gene bank (M. Joubert, pers. comm., 2016).

Overall, a high level of within population variation (>90%) was maintained in the *C. subternata* wild populations (Table 4.3). However, the AMOVA results indicated the accessions had a slight decrease in genetic variation (7%) relative to the three wild populations that were compared to each other (10%) (Table 4.3). The similar, albeit small, decrease in levels of genetic diversity could indicate that the accessions represent the genetic diversity present in most of the wild populations but is not adequately representative of population 2. Although the accessions denoted as SHL were selected from Haarlem, they came from a plantation and not a wild population. The Haarlem wild population is genetically distinct and should be sampled for addition to the ARC field gene bank.

Stellenbosch University https://scholar.sun.ac.za

4.4.3 Genetic differentiation of wild populations and implications for conservation

4.4.3.1 Populations 1 (Guava Juice, Eastern Cape) and 3 (George, Western Cape)

Weak genetic differentiation was observed between these wild populations and no evidence of isolation by distance was detectable. In contrast, these two populations were geographically the furthest apart and seemed to be a single panmictic population with low estimates of subdivision, indicating possible gene flow.

The great distance (185 km) between these populations could potentially act as a barrier for gene flow in the form of pollen dispersal; however, studies are still underway to determine how far pollen can disperse in *Cyclopia* species (J. Koen, pers. comm., 2016). Dispersal of pollen over large distances ensures high levels of gene flow, often causing populations to be genetically undifferentiated (Kamm *et al*., 2009). This is, however, not likely scenario in this case and it is more feasible that populations 1 and 3 may have resulted from similar source populations, possibly caused by human interference such as the translocation of plant species. A high within population genetic diversity, as was revealed by the AMOVA results, often points to the exchange of genetic material between localities (Zawedde *et al*., 2015; Dennhardt *et al*., 2016). Another explanation is the spontaneous spread of plants from Guava Juice (population 1) to George (population 3), or *vice versa* since *C. subternata* is known to colonise new environments rapidly and has a high establishment rate compared to other *Cyclopia* species (Motsa, 2016).

Population structure may also be affected by the polyploid nature of *Cyclopia*. The presence of multiple gene copies, as opposed to only two in a diploid individual, will lead to the homogenisation of alleles among populations and therefore a lower degree of genetic differentiation than what can be expected in a diploid individual (Meirmans and Van Tienderen, 2013).

4.4.3.2 Population 2 (Haarlem, Western Cape)

A unique leaf morphology was detected in the Haarlem population (thin and narrow) compared to that of the two other populations (large, flat leaves). Local farmers harvest this population for vleitee production and young leaves were in abundance during sampling in March, indicating recent wild harvesting of the plants and consequent regrowth. The unique leaf morphology, together with its distinct genetic structure, indicates clear genetic differentiation and possibly local adaptation of the Haarlem population. The population in Haarlem, against a slope of the Langkloof surrounded by mountains, is completely isolated from the Guava Juice and George wild populations. This was

Stellenbosch University https://scholar.sun.ac.za

supported by the separate PCoA clustering pattern and higher pairwise ϕ_{PT} and ρ_{ST} values compared to the other wild populations.

The phenotypic differences across various *C. subternata* wild populations could potentially be attributed to varying climatic conditions (Table 4.5). However, similar rainfall patterns and maximum temperatures of 20-23°C, are observed for all three populations. Haarlem does exhibit a slightly lower minimum temperature, but not extreme enough to explain such phenotypic changes. As the weather data was obtained from Joubertina weather station, which is the closest to the sampled Haarlem population but still 60km away, the weather could differ slightly to what is reported.

Table 4.5. Climatic conditions as measured by the ARC AgroMet weather stations closest to the wild populations of *C. subternata* (indicated in brackets) across three years (2012-2014).

Year	Grasslands (Guava Juice, EC) Altitude = 210m			Joubertina (Haarlem, WC) Altitude = 746m			Mosselbay (George, WC) Altitude = 203m		
	Max (°C)	Min (°C)	Rain (mm)	Max (°C)	Min (°C)	Rain (mm)	Max (°C)	Min (°C)	Rain (mm)
2012	21.7	9.7	100.4	20.3	6.6	79.3	21.9	11.2	77.6
2013	22.1	8.9	72.1	23.2	7.8	53.2	22.6	11.1	52.0
2014	23.6	10.2	75.0	23.2	7.9	50.6	22.3	11.7	62.5

Cyclopia subternata is known to occur within a wide range of altitudes (160-1000m) (Schutte, 1997). The Haarlem population was sampled at an altitude of ~746m whereas the other two populations had lower altitudes of 200-210m (Table 4.5). Plants exposed to increased altitudes adapt to conditions such as increased irradiation, a reduction in atmospheric temperature and a decrease in atmospheric pressure (including oxygen) (Körner, 2007). Therefore, populations from the same species located at various altitudes can be quite distinct (Valladares *et al.*, 2014). Leaf sizes tend to decrease with higher altitude to reduce their exposure to direct irradiation (McDonald *et al.*, 2003). Altitudinal differences could therefore potentially explain the observed phenotypic leaf differences of the Haarlem population, but this would need to be studied in more detail before final conclusions can be reached.

4.4.3.3 Conservation

It is recommended that conservation efforts be implemented to ensure genetic distinctness and preservation of the Haarlem population, as suggested by the preliminary population genetic results of the current study. Once further research on the Haarlem population confirms its genetic distinctness, management plans should capture the genetic material from this area for preservation in the field gene bank. Furthermore, it should be a high priority for *in situ* conservation. The monitoring of wild harvesting in this area, as well as the prevention of artificial movement of plants from and to this

Stellenbosch University https://scholar.sun.ac.za

distinct population is important. The mixing of populations that are genetically distinct can have implications such as outbreeding depression, where the progeny have a decreased fitness due to the disruption of gene complexes that confer fitness when they are inherited collectively (Kramer and Havens, 2009; Frankham *et al*., 2011; Clarke *et al*., 2012). Also, the introduction of maladaptive alleles can cause previously adapted populations to have decreased fitness (Allendorf *et al*., 2001). However, when populations exhibit low levels of genetic differentiation, minimal negative genetic effects are expected when individuals are exchanged such as between the panmictic populations from Guava Juice and George (Swarts *et al*., 2009).

Concern arises for wild populations when uncontrolled gene flow from cultivated material could cause the locally adapted genotypes to be replaced by hybrids. This process is known as genetic swamping (Campbell *et al*., 2016; Todesco *et al*., 2016). In some circumstances, this can lead to a reduction of genetic diversity in the wild relatives and therefore decreased fitness, placing rare species at risk of extinction (Todesco *et al*., 2016). Commercial plantations that could potentially contaminate the Haarlem population should be avoided in this area and further field studies should investigate the minimum isolation distance that will limit gene flow. The molecular tools developed in this study can be applied to study the gene flow of wild populations for which concern exists due to cultivated plantations that occur in close proximity.

4.5 Conclusion

This is the first study to use microsatellite markers for inferring population genetic structure in the polyploid species *C. subternata*. Several methods, each verifying the same findings, were successfully employed for a small-scale population genetic diversity study.

The polyploid nature of *Cyclopia* introduced several limitations to the study. These limitations were taken into account throughout the study and discussed in detail. As *Cyclopia* research is still at its initial phase, information on the mode of inheritance, reproduction and pollen dispersal abilities are needed for a more thorough population genetic study. Additional *C. subternata* wild populations should also be investigated and more individuals per population should be included.

Even though microsatellite markers have remained the most frequently used marker in population genetic studies, single nucleotide polymorphism (SNPs) are also a viable option. The use of SNPs in polyploids, however, present similar problems as the use of microsatellite markers in terms of allele dosage, but have been applied successfully in several studies (Rickert *et al*., 2002; Cordeiro *et al*.,

Stellenbosch University https://scholar.sun.ac.za

2006; Bundock *et al*., 2009). Amplified fragment length polymorphisms (AFLPs) are an attractive alternative for polyploids, as their dominant nature ensures that dosage problems are avoided (Dufresne *et al*., 2014).

It is recommended that *C. subternata* wild populations should be evaluated as single entities before conservation strategies are implemented. In this study, the Guava Juice and George populations had some degree of admixture and exhibited low levels of genetic differentiation. Minimal negative effects, such as outbreeding depression, are expected if individuals are exchanged between these populations. The Haarlem population was, however, found to be genetically distinct and efforts should be made to include material from this population in the germplasm collection to ensure that the full scope of genetic diversity of this species is captured. Several *C. subternata* wild populations that were not sampled in this study should also be included to ensure that the field gene bank capture their genetic diversity. These span across areas from Riversdal and Albertinia, the Garden Route (Plettenberg Bay, Knysna) up to Karreedouw.

References

Adams, K. & Wendel, J., 2005. Polyploidy and genome evolution in plants. Current Opinion in Genetics and Development 8, 135-141.

Allendorf, F.W. & Luikart, G., 2007. Conservation and the Genetics of Populations. Blackwell, Oxford, UK, 205-227.

Allendorf, F.W., Leary, R.F., Spruell, P., Wenburg, J.K., 2001. The problems with hybrids: Setting conservation guidelines. Trends in Ecology and Evolution 16, 613-622.

Anderson, M.J., Willis, T.J., 2003. Canonical analysis of principal coordinates: A useful method of constrained ordination for ecology. Ecology 84, 511–525.

Assoumane, A., Zoubeiro, A., Rodier-Goud, M., Favreau, B., Bezancon, G., Verhaegen, D., 2013. Highlighting the occurrence of tetraploidy in *Acacia senegal* (L.) Willd. and genetic variation patterns in its natural range revealed by DNA microsatellite markers. Tree Genetics and Genomes 9, 93-106.

Barnardo, N., 2013. Sustainable harvesting of wild populations of *Cyclopia intermedia* in Kouga, Eastern Cape, South Africa. Unpublished M.Sc. thesis, Nelson Mandela Metropolitan University, Port Elizabeth, 7-8.

Bester, C., 2012. A model for commercialisation of honeybush tea, an indigenous crop. Acta Horticulturae 1007, 889-894.

Stellenbosch University https://scholar.sun.ac.za

Bester, C., Tobutt, K., Mansvelt, E., Blomerus, L., Jolly, N., Pieterse, W., Smit, L., Van Schalkwyk, D., 2013. The value and impact of the ARC Infruitec-Nietvoorbij genebanks. Acta Horticulturae 1007, 975-980.

Bonin, A., Bellemain, E., Eidesen, P.B., Pompanon, F., Brochmann, C., Taberlet, P., 2004. How to track and assess genotyping errors in population genetics studies. Molecular Ecology 13, 3261-3273.

Brandrud, M.K., 2014. Polyploidy and ecotype variation in *Cochlearia officinalis* L. and related species. Unpublished M.Sc. thesis, University of Oslo, Norway, 22-23.

Bundock, P.C., Eliott, F.G., Ablett, G., Benson, A.D., Casu, R.E., Aitken, K.S., Henry, R.J., 2009. Targeted single nucleotide polymorphism (SNP) discovery in a highly polyploid plant species using 454 sequencing. Plant Biotechnology Journal 7, 347-354.

Campbell, L.G., Lee, D., Shukla, K., Waite, T.A., Bartsch, D., 2016. An ecological approach to measuring the evolutionary consequences of gene flow from crops to wild or weedy relatives. Applications in Plant Sciences 4, 1-7.

Caruso, M., Curro, S., Las Casas, G., La Malfa, S., Gentile, A., 2010. Microsatellite markers help to assess genetic diversity among *Opuntia ficus indica* cultivated genotypes and their relation with related species. Plant Systematics and Evolution 290, 85-97.

Clark, L.V. & Jasieniuk, M., 2011. POLYSAT: An R package for polyploid microsatellite analysis. Molecular Ecology Resources 11, 562-566.

Clarke, L., Jardine, D., Byrne, M., Shepherd, M., Lowe, A., 2012. Significant population genetic structure detected for a new and highly restricted species of *Atriplex* (Chenopodiaceae) from Western Australia, and implications for conservation management. Australian Journal of Botany 60, 32-41.

Cordeiro, G.M., Eliott, F., McIntyre, C.L., Casu, R.E., Henry, R.J., 2006. Characterisation of single nucleotide polymorphisms in sugarcane ESTs. Theoretical and Applied Genetics 113, 331-343.

Cruse-Sanders, J.M. & Hamrick, J.L., 2004. Genetic diversity in harvested and protected populations of wild American ginseng, *Panax quinquefolius* L. (Araliaceae). American Journal of Botany 91, 540-548.

Dennhardt, L., Dekeyser, E., Tennefos, S., Travers, S., 2016. There is no evidence of geographical patterning among invasive *Poa pratensis* L. populations in the northern Great Plains. Unpublished Ph.D. thesis, North Dakota State University, Grand Forks, 7-12.

Dufresne, F., Stift, M., Vergilino, R., Mable, B.K., 2014. Recent progress and challenges in population genetics of polyploid organisms: An overview of current state-of-the-art molecular and statistical tools. Molecular Ecology 23, 40-69.

Stellenbosch University https://scholar.sun.ac.za

Earl, D.A. & VonHoldt, B.M., 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4, 359-361.

Eliášová, A., Trávníček, P., Mandák, B., Münzbergová, Z., 2014. Autotetraploids of *Vicia cracca* show a higher allelic richness in natural populations and a higher seed set after artificial selfing than diploids. Annals of Botany 113, 159-170.

Esselink, G.D., Nybom, H., Vosman, B., 2004. Assignment of allelic configuration in polyploids using the MAC-PR (microsatellite DNA allele counting-peak ratios) method. Theoretical and Applied Genetics 109, 402-408.

Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Molecular Ecology 14, 2611-2620.

Ford-Lloyd, B., 2001. Genotyping in plant genetic resources, in: Henry, R.J. (Ed.), Plant Genotyping: The DNA Fingerprinting of Plants. CABI, New York, USA, 59-81.

Frankham, R., Ballou, J., Briscoe, D., 2002. Introduction to Conservation Genetics. Cambridge University, Cambridge, UK, 330-441.

Frankham, R., Ballou, J.D., Eldridge, M.D.B., Lacy, R.C., Ralls, K., Dudash, M.R., Fenster, C.B., 2011. Predicting the probability of outbreeding depression. Conservation Biology 25, 465-475.

Freeland, J., 2005. Molecular Ecology. John Wiley and Sons, England, 127-129.

Goldblatt, P., 1981. Chromosome numbers in legumes II. Annals of the Missouri Botanical Garden 68, 551-557.

Hamilton, M., 2009. Population structure and gene flow, in: Population Genetics. Wiley-Blackwell, England, 88-153.

Hardy, O.J. & Vekemans, X., 2002. SPAGeDi: A versatile computer program to analyse spatial genetic structure at the individual or population levels. Molecular Ecology Notes 2, 618-620.

Jørgensen, M., Carlsen, T., Skrede, I., Elven, R., 2008. Microsatellites resolve the taxonomy of the polyploid *Cardamine digita* aggregate (Brassicaceae) 57, 882-892.

Joubert, E., Gelderblom, W.C.A, Louw, A., De Beer, D., 2008. South African herbal teas: *Aspalathus linearis*, *Cyclopia* spp. and *Athrixia phylicoides* - a review. Journal of Ethnopharmacology 119, 376-412.

Joubert, E., Joubert, M., Bester, C., De Beer, D., De Lange, J., 2011. Honeybush (*Cyclopia* spp.): From local cottage industry to global markets - the catalytic and supporting role of research. South African Journal of Botany 77, 887-907.

Kamm, U., Rotach, P., Gugerli, F., Siroky, M., Edwards, P., Holderegger, R., 2009. Frequent long-distance gene flow in a rare temperate forest tree (*Sorbus domestica*) at the landscape scale. Heredity 103, 476-482.

Stellenbosch University https://scholar.sun.ac.za

Keyghobadi, N., Roland, J., Strobeck, C., 2005. Genetic differentiation and gene flow among populations of the alpine butterfly, *Parnassius smintheus*, vary with landscape connectivity. Molecular Ecology 14, 1897-1909.

Kloda, J.M., Dean, P., Maddren, C., MacDonald, D.W., Mayes, S., 2008. Using principle component analysis to compare genetic diversity across polyploidy levels within plant complexes: An example from British Restharrows (*Ononis spinosa* and *Ononis repens*). Heredity 100, 253-260.

Koen, J., Slabbert, M.M., Bester, C., Bierman, F., 2016. Germination characteristics of dimorphic honeybush (*Cyclopia* spp.) seed. South African Journal of Botany, In Press. Available at: http://dx.doi.org/10.1016/j.sajb.2016.03.006

Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A., Mayrose, I., 2015. CLUMPAK: A program for identifying clustering modes and packaging population structure inferences across K. Molecular Ecology Resources 15, 1179-1191.

Körner, C., 2007. The use of “altitude” in ecological research. Trends in Ecology and Evolution 22, 569-574.

Kramer, A.T. & Havens, K., 2009. Plant conservation genetics in a changing world. Trends in Plant Science 14, 599-607.

Lowe, A., Harris, S., Ashton, P., 2004. Genetic diversity and differentiation, in: Ecological Genetics: Design, Analysis and Application. Blackwell, Oxford, UK, 52-100.

Mantel, N., 1967. The detection of disease clustering and a generalized regression approach. Cancer Research 27, 209-220.

McDonald, P.G., Fonseca, C.R., Overton, J.M., Westoby, M., 2003. Leaf-size divergence along rainfall and soil-nutrient gradients: Is the method of size reduction common among clades? Functional Ecology 17, 50-57.

Meirmans, P.G. & Van Tienderen, P.H., 2004. GENOTYPE and GENODIVE: Two programs for the analysis of genetic diversity of asexual organisms. Molecular Ecology Notes 4, 792-794.

Meirmans, P.G. & Van Tienderen, P.H., 2013. The effects of inheritance in tetraploids on genetic diversity and population divergence. Heredity 110, 131-137.

Mills, L., 2007. Conservation of Wildlife Populations: Demography, Genetics and Management. Blackwell, Oxford, UK, 176-188.

Motsa, M., 2016. Natural phenology, fecundity, genetic variation and seed dormancy of *Cyclopia* species (Honeybush). Unpublished D.Tech. thesis, Tshwane University of Technology, Pretoria, 159-163.

Peakall, R. & Smouse, P.E., 2006. GenAlEx 6: Genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6, 288-295.

Stellenbosch University https://scholar.sun.ac.za

Pfeiffer, T., Roschanski, A.M., Pannell, J.R., Korbecka, G., Schnittler, M., 2011. Characterization of microsatellite loci and reliable genotyping in a polyploid plant, *Mercurialis perennis* (Euphorbiaceae). Journal of Heredity 102, 479-488.

Powell, W., Morgante, C., Andre, M., Hanafey, M., Vogel, J., Tingey, S., Rafalski, A., 1996. The comparison of RFLP, RAPD, AFLP, and SSR (microsatellite) markers for germplasm analysis. Molecular Breeding 2, 225-238.

Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics Society of America 155, 945-959.

Rickert, A.M., Premstaller, A., Gebhardt, C., Oefner, P.J., 2002. Genotyping of SNPs in a polyploid genome by pyrosequencing. BioTechniques 32, 592-603.

Rogstad, S., 1992. Saturated NaCl-CTAB solution as a means of field preservation of leaves for DNA analyses. Taxon 41, 701-708.

Ronfort, J., Jenczewski, E., Bataillon, T., Rousset, F., 1998. Analysis of population structure in autotetraploid species. Genetics 150, 921-930.

SANBI, 2012. Red List of South African Plants. [ONLINE] Available at: http://redlist.sanbi.org/. [Accessed 8 August 2016].

Sampson, J. & Byrne, M., 2012. Genetic diversity and multiple origins of polyploid *Atriplex nummularia* Lindl. (Chenopodiaceae). Biological Journal of the Linnean Society 105, 218-230.

Schippmann, U., Leaman, D., Cunningham, A.B., 2006. A comparison of cultivation and wild collection of medicinal and aromatic plants under sustainability aspects, in: Bogers, R., Craker, L., Lange, D. (Eds.), Medicinal and Aromatic Plants. Springer, Netherlands, 75-95.

Schutte, A.L., 1997. Systematics of the genus *Cyclopia* Vent. (Fabaceae, Podalyrieae). Edinburgh Journal of Botany 54, 125-170.

Soltis, D.E. & Soltis, P.S., 1993. Molecular data and the dynamic nature of polyploidy. Critical Reviews in Plant Sciences 12, 243-273.

Swarts, N.D., Sinclair, E.A., Krauss, S.L., Dixon, K.W., 2009. Genetic diversity in fragmented populations of the critically endangered spider orchid *Caladenia huegelii*: Implications for conservation. Conservation Genetics 10, 1199-1208.

Teixeira, H., Rodríguez-Echeverría, S., Nabais, C., 2014. Genetic diversity and differentiation of *Juniperus thurifera* in Spain and Morocco as determined by SSR. PLoS ONE 9, 1-7.

Todesco, M., Pascual, M.A., Owens, G.L., Ostevik, K.L., Moyers, B.T., Hübner, S., Heredia, S.M., Hahn, M.A., Caseys, C., Bock, D.G., Rieseberg, L.H., 2016. Hybridization and extinction. Evolutionary Applications 9, 892-908.

Valladares, F., Matesanz, S., Guilhaumon, F., Araujo, M.B., Balaguer, L., Benito-Garzon, M., Cornwell, W., Gianoli, E., Van Kleunen, M., Naya, D.E., Nicotra, A.B., Poorter, H., Zavala,

Stellenbosch University https://scholar.sun.ac.za

M.A., 2014. The effects of phenotypic plasticity and local adaptation on forecasts of species range shifts under climate change. Ecology Letters 17, 1351-1364.

Wiesner, K.R., Loxdale, H.D., Schneider, A.R.R., Tiedemann, R., Weisser, W.W., 2011. Patterns of local and regional genetic structuring in the meadow grasshopper, *Chorthippus parallelus* (Orthoptera: Acrididae), in Central Germany revealed using microsatellite markers. Biological Journal of the Linnean Society 103, 875-890.

Zawedde, B.M., Ghislain, M., Magembe, E., Amaro, G.B., Grumet, R., Hancock, J., 2015. Characterization of the genetic diversity of Uganda's sweet potato (*Ipomoea batatas*) germplasm using microsatellites markers. Genetic Resources and Crop Evolution 62, 501-513.

Zuur, A.F., Ieno, E.N., Smith, G.M., 2007. Principal coordinate analysis and non-metric multidimensional scaling, in: Gail, M., Krickeberg, K., Sarnet, J., Tsiatis, A., Wong, W. (Eds.), Analysing Ecological Data. Springer, New York, USA, 259–261.

Stellenbosch University https://scholar.sun.ac.za

Chapter 5: Study Conclusions

5.1 Background

The use of molecular markers to characterise plant genetic resources at the DNA level can provide valuable information on genetic diversity, not only among accessions, but also among wild populations and accessions (Laurentin, 2009). This can aid in conservation strategies and the effective management of plant genetic resources, which is crucial for the preservation of genetic diversity (Fowler and Hodgkin, 2004). *Cyclopia* is an indigenous plant of commercial importance and, as part of its improvement programme, the Agricultural Research Council (ARC) maintain a field gene bank for several of the species (Bester *et al*., 2013). Some accessions originate from wild populations as a form of *ex situ* conservation, but there has been little systematic sampling. Very little genetic information is available for *Cyclopia* and no molecular markers are available to characterise the gene bank of this genus, nor to explore the genetic structure of wild populations. The primary aim of this study was to develop genetic marker resources and apply them to the DNA fingerprinting of the accessions currently maintained in the ARC gene bank as well as the determination of the genetic diversity among accessions and wild populations of *C. subternata*.

5.2 Synopsis of Main Results and Discussion

A DNA extraction protocol was optimised that yielded high concentrations of good quality DNA across seven *Cyclopia* species, eliminating the need for species-specific protocols. Results showed that young leaves as starting material were a significantly better option than old leaves or young stems. Young leaves are softer, easier to break down and contain smaller quantities of phenolic compounds that could potentially interfere with the DNA during PCR amplification (Lowman and Box, 1983; Khanuja *et al*., 1999; Moreira and Oliveira, 2011). Sampling expeditions to remote locations present a problem for the preservation of plant tissue pending DNA extraction under laboratory conditions (Štorchová *et al*., 2000). The problem was addressed in the current study by the optimisation of a CTAB/NaCl buffer method to preserve *C. subternata* leaves successfully for up to two weeks.

Microsatellite markers were developed for *C. subternata* and 19 of these microsatellites were successfully cross-amplified to six other *Cyclopia* species. This reduced the excessive costs that often accompany the process of species-specific marker development (Kalia *et al*., 2011). The cross-species amplification success ranged from 81-88% within the genus *Cyclopia*.

Stellenbosch University https://scholar.sun.ac.za

The two multiplexes comprised of six microsatellite markers were developed for DNA fingerprinting of the accessions that are currently in the ARC field gene bank. This will be useful for gene bank curation, particularly as new accessions are incorporated as well as for the proper identification of accessions (Acquaah, 2007). A general prerequisite for microsatellite profiles is their reproducibility. This is especially important during analysis of polyploids, where genotyping ambiguities are frequently magnified as the ploidy level increases and stutter peaks could be misidentified as true peaks (Weising *et al*., 2005; Pfeiffer *et al*., 2011). In the current study, *Cyclopia* species that exhibited chromosome numbers up to hexaploid level were genotyped effectively using the abovementioned multiplexes, since peaks were scored in a straightforward manner, thereby generating reproducible profiles. Species such as *C. genistoides* and *C. intermedia* with high ploidies up to 10*x* and 14*x* (Goldblatt, 1981; Schutte, 1997; Motsa, 2016) had a large number of peaks that could not be scored unambiguously, thereby reducing the utility of the currently developed markers in these species. The probability of identity (P_{ID}) was established for all fingerprints and the multiplex marker panels were sufficient as one in every $2x10^{13}$ *C. subternata* accessions may have the same genotype, while the lowest P_{ID} was still efficient as one in every 1774 *C. maculata* accessions may have the same genotype.

This study addressed the limitations posed by the polyploid nature of *Cyclopia* species by comparing suitable methods that did not violate the underlying assumptions of disomic inheritance (Dufresne *et al*., 2014). Various methods have been applied in polyploid population genetic studies. Principal coordinate analysis (PCoA) clustering lacks any underlying population genetic assumptions and is independent of ploidy level, as long as the pairwise genetic distance measure was chosen correctly (Kloda *et al*., 2008; Dufresne *et al*., 2014). Analysis of Molecular Variance (AMOVA) partitioning is used to explain the distribution of genetic variation among populations using ϕ-statistics as an analogue to F-statistics (Lowe *et al*., 2004; Meirmans, 2006). STRUCTURE software can handle genotype ambiguities, which are often found in polyploid data analyses (Clark and Jasieniuk, 2011). The results from these various methods led to similar findings regarding the population structure and genetic diversity of *Cyclopia* wild populations and accessions.

Wild populations exhibited low levels of genetic differentiation, except for the Haarlem population that was genetically distinct. This indicates that, in the future, *C. subternata* wild populations should be evaluated as separate entities before general conservation efforts can be applied, as the movement of plant material between genetically distinct populations could have severe genetic implications for some populations, such as outbreeding depression (Swarts *et al*., 2009; Frankham *et al*., 2011). The

Stellenbosch University https://scholar.sun.ac.za

ARC accessions seem to represent the natural genetic diversity of the sampled populations sufficiently, excluding the genetic diversity represented by the Haarlem population. Efforts should therefore be made to include genetically distinct samples in order to ensure that a wider range of genetic diversity is captured in the accessions for genetic preservation (Rao and Hodgkin, 2002). In addition, several *C. subternata* wild populations have not yet been sampled, stretching from areas such as from Riversdal and Albertinia, and the Garden Route (Plettenberg Bay, Knysna) up to Karreedouw. The genetic diversity in the gene bank might consequently be unevenly representative across all wild populations (Rao and Hodgkin, 2002).

5.3 Study Limitations

Sample sizes in this study were often limited which could have biased the results. For the CTAB/NaCl buffer preservation tests, various species with more accessions per species could be included and tested for an extended period of time to see how long *Cyclopia* material can be stored in this manner. This will allow for prolonged sampling expeditions to distant and hard-to-reach sampling sites such as in the Langkloof. The number of accessions available for *C. intermedia* was especially limited, since cultivation of this species is challenging. For the population study, more *C. subternata* populations can be included across a wider geographical range in South Africa, as recommended above. This will ensure a more thorough analysis of the population genetic status of this species and will be important for future conservation efforts to give an indication of unique populations that might harbour *C. subternata* genetic diversity which is currently not represented in the gene bank.

Taxonomic classification based on morphology is a challenging venture in *Cyclopia* species. This is reflected in the periodic reclassification of the genus since it was established in 1808. Schutte (1997) eventually recognised 23 *Cyclopia* species based on morphological characteristics but still, conflicting classifications persist. *Cyclopia subternata* (non-sprouter) has often been mistaken for *C. intermedia* (sprouter) based on similar leaf morphology, but they can be distinguished by their different growth forms. Furthermore, *C. laxiflora* was found to be closely related to *C. subternata* based on morphology with several shared traits such as small bracts and broad upper sutures of pods and, although it was retained as a separate species, according to Schutte (1997) it could possibly be included in *C. subternata*. Another example involves *C. dregeana* that was subsequently reclassified as *C. buxifolia*. It was concluded that there were no distinguishing characteristics to separate these two species and their similar distribution ranges indicate a single species (Schutte, 1997). These discrepancies in classification warrant deeper investigation of the genetically distinct Haarlem population to investigate the possibility of local adaptation or even ecological speciation. Local

Stellenbosch University https://scholar.sun.ac.za

adaptation is often the first step in speciation, whereby an organism experiences microevolutionary changes to adapt to a given environment (Lenormand, 2011; Montesinos-Navarro *et al*., 2011). Once barriers to gene flow evolve as a result of ecologically-based divergent selection, ecological speciation can occur (Rundle and Nosil, 2005). The study of adaptation requires that the phenotypic variation is associated with the environment in which the related study groups have survived (Montesinos-Navarro *et al*., 2011).

Investigation into the mode of reproduction (outcrossing or self-fertilising) as well as inheritance patterns (disomic or polysomic) of *Cyclopia* species by using microsatellite tools will aid in future studies where population statistics are to be applied. The microsatellite genotyping of progenies from crosses of polyploids will aid in determining the inheritance pattern for these species and have been applied in *Rorippa amphibia*, a tetraploid confirmed to be of autotetraploid origin (Stift *et al*., 2008). It will also provide useful information for the breeding programme when raising progenies.

Studies into the dispersal of seed and pollen will provide valuable insights as to the population structuring of *Cyclopia*. Seed dispersal is often spatially limited, while pollen dispersal in outcrossing plants can occur over larger distances by insect pollinators (Scheepens *et al*., 2012) such as the carpenter bee in *Cyclopia* (Van Wyk, 1993). Self-fertilising plants have limited pollen-mediated gene flow as they lack a mechanism for pollen dispersal (Freeland, 2005). Pollen dispersal studies can observe various pollinators during field experiments that contribute to gene flow (Ellstrand, 1992). Long-distance gene flow will also be influenced by pollen viability which can be tested by colorimetric tests that stain only viable pollen (Gaaliche *et al*., 2013). Additionally, studies on dispersal would help address the concerns of levels of gene flow from cultivated material to the wild populations (Campbell *et al*., 2016). The molecular tools developed in this study can be applied to study the gene flow of wild populations for which concern exists due to nearby cultivated plantations. Di Vecchi-Staraz *et al*. (2009) studied the pollen-mediated gene flow from cultivated to wild plants in a *Vitis vinifera* subspecies. They studied two wild populations and concluded that a higher frequency of cross-pollination occurred at the wild population located close to a commercial vineyard as opposed to the wild population that occurred 10km from a commercial vineyard.

Finally, the polyploid nature of *Cyclopia* presented a major limitation for a full-scale population genetic study. Alternative methods were employed throughout the study, such as the use of ϕ_{PT} and ρ_{ST} as an alternative to F_{ST}, to ensure that no assumptions based on disomic inheritance were violated (Brandrud, 2014). Principal coordinate analysis (PCoA) present an attractive option for visualising

Stellenbosch University https://scholar.sun.ac.za

pairwise genetic distances between individuals, assuming that the underlying genetic distance matrix is chosen correctly (Kloda *et al*., 2008).

5.4 Future Work

The development of microsatellite markers is an important step for further molecular characterisation of *Cyclopia* species. The microsatellites developed in this study can be applied for population genetic studies in other commercially important species of *Cyclopia*, such as *C. longifolia*. For this species, similar strategies as used in this study can be applied to investigate the genetic diversity of the accessions and wild populations as *C. longifolia* is also hexaploid (Motsa, 2016).

For the purpose of marker assisted breeding in future studies, the construction of linkage maps that display the position and genetic distance of markers is a prerequisite (Weising *et al*., 2005). Markers which are linked to important agronomic traits can be identified from the linkage map. To ensure full genome coverage, a large number of segregating markers dispersed throughout the genome is often needed for linkage studies (Collard *et al*., 2005), therefore many additional microsatellites will be required for *Cyclopia*. Linkage maps have proven especially useful for the genetic improvement of other tea plants such as *Camellia sinensis*, where traits associated with the leaf biomass yield and tea quality could be mapped (Hackett *et al*., 2000). A similar linkage map would greatly benefit *Cyclopia* breeding, however, linkage mapping in polyploids is more complex. Gene copies are unknown and large segregating population sizes are required to reliably estimate the genetic distances between the markers (Albertini *et al*., 2003; Baker *et al*., 2010). Linkage mapping of polyploids becomes more straightforward when diploid relatives are available for comparison, but polyploids such as sugarcane (Wu *et al*., 1992) and even *Cyclopia* do not have diploid counterparts. Single dose restriction fragment markers (SDRFs) that segregate in a 1:1 (presence:absence) ratio have been applied successfully to construct linkage maps in sugarcane as well as to study chromosome behaviour in order to determine the allopolyploid or autopolyploid state of this species (Wu *et al*., 1992; Albertini *et al*., 2003). For allopolyploids, SDRFs are the segregation equivalent of an allele at a heterozygous locus in a diploid and for autopolyploids it represents a single dose allele (Fregene *et al*., 1997).

Mitochondrial and chloroplast markers have previously been used in *Cyclopia* for phylogenetic inference, but these markers could be used to provide insight into the phylogeography of *Cyclopia* species across their distribution range (Boatwright *et al*., 2008). A study by Malgas *et al*. (2010) utilised the trnL–trnF region in *Aspalathus linearis*, commonly known as rooibos, to identify four unique haplotypes that strongly correlated with the various growth forms. A similar study on *Cyclopia*

Stellenbosch University https://scholar.sun.ac.za

might help investigate the underlying genetics of sprouters, such as *C. genistoides*, *C. intermedia* and *C. sessiliflora* versus non-sprouters, such as *C. maculata* and *C. subternata*. This could aid in clarifying the unknown growth forms of species such as *C. alopecuroidis* and *C. buxifolia* (Joubert *et al*., 2011). Although growth forms can be used as a taxonomic character for identification of species that have similar morphological features, other tools such as DNA barcoding could be worth exploring. DNA barcoding utilises variations in short DNA segments from a universal part of the genome to distinguish species. For plants, two plastid coding regions, *rbcL* and *matK*, were recommended by the Consortium for the Barcode of Life (CBOL) Plant Working group (Hollingsworth *et al*., 2011). These, however, do not always show sufficient variation and therefore additional nuclear markers are necessary (Spooner, 2009). Since polyploids have multiple copies of single nuclear genes, the application of nuclear DNA barcodes will be problematic as sequencing requires the PCR amplification of a single amplicon (Griffin *et al*., 2011). The haploid nature of plastid markers would therefore be a feasible alternative for polyploids (Hamilton, 2009).

Single nucleotide polymorphism (SNPs) present an attractive alternative to microsatellites, since they are amenable to high-throughput genotyping. There is a wide array of techniques that have been applied to discover and genotype SNPs in polyploid organisms such as the TaqMan assay (De Jong *et al*., 2003; Giancola *et al*., 2006), expressed sequence tags (ESTs) (Mochida *et al*., 2003; Cordeiro *et al*., 2006) and 454 sequencing (Bundock *et al*., 2009; Oliver *et al*., 2011). Each presents its own limitations in a polyploid and largely unexplored genome such as *Cyclopia*. TaqMan requires gene specific primers and currently no ESTs are available for this genus. SNP discovery by sequencing would most likely be the most viable future option for *Cyclopia*, as long as it is kept in mind that multiple genomes are involved (Akhunov *et al*., 2009).

Although genetic research on *Cyclopia* is at its initial phase, future collaborative efforts will require the calibration of microsatellite data to ensure the easy comparison between various laboratories (Jones *et al*., 1997). Allelic binning is a crucial step during genotyping whereby raw allele sizes are assigned a fixed integer (Guichoux *et al*., 2011). It is, however, difficult to harmonise allele sizes from different laboratories, since allele calling and binning methods are often not standardised (Urrestarazu *et al*., 2016). The inclusion of reference genotypes as controls should ensure the standardisation of allele calling across various research efforts (Guichoux *et al*., 2011).

Stellenbosch University https://scholar.sun.ac.za

5.5 Final Remarks

This was the first study to develop molecular markers, specifically microsatellites, for the polyploid genus, *Cyclopia*. Markers were successfully applied to generate unique profiles for a subset of the accessions that are currently maintained in the ARC field gene bank. In addition, a pilot population genetic study gave important insights into the genetic diversity distribution in wild populations of the species *C. subternata* and investigated the degree to which this genetic diversity is represented in the field gene bank. Knowledge obtained during this study will aid the ARC in proper management of its accessions in terms of correct labelling and future sampling efforts to include additional wild germplasm for breeding purposes, as well as keeping track of clones for seed orchards or commercial release. In conclusion, the present study provides valuable information about the genetic diversity of *Cyclopia* and lays the foundation for future, more in-depth studies.

References

Acquaah, G., 2007. Plant genetic resources for plant breeding, in: Principles of Plant Genetics and Breeding. Blackwell, Oxford, UK, 110-114.

Akhunov, E., Nicolet, C., Dvorak, J., 2009. Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. Theoretical and Applied Genetics 119, 507-517.

Albertini, E., Porceddu, A., Marconi, G., Barcaccia, G., Pallottini, L., Falcinelli, M., 2003. Microsatellite-AFLP for genetic mapping of complex polyploids. Genome research 46, 824-832.

Baker, P., Jackson, P., Aitken, K., 2010. Bayesian estimation of marker dosage in sugarcane and other autopolyploids. Theoretical and Applied Genetics 120, 1653-1672.

Bester, C., Tobutt, K., Mansvelt, E., Blomerus, L., Jolly, N., Pieterse, W., Smit, L., Van Schalkwyk, D., 2013. The value and impact of the ARC Infruitec-Nietvoorbij genebanks. Acta Horticulturae 1007, 975-980.

Boatwright, J., Savolainen, V., Van Wyk, B., Schutte-Vlok, A., Forest, F., Van Der Bank, M., 2008. Systematic position of the anomalous genus *Cadia* and the phylogeny of the tribe Podalyrieae (Fabaceae). Systematic Botany 33, 133-147.

Brandrud, M.K., 2014. Polyploidy and ecotype variation in *Cochlearia officinalis* L. and related species. Unpublished M.Sc. thesis, University of Oslo, Norway, 34-42.

Bundock, P.C., Eliott, F.G., Ablett, G., Benson, A.D., Casu, R.E., Aitken, K.S., Henry, R.J., 2009. Targeted single nucleotide polymorphism (SNP) discovery in a highly polyploid plant species using 454 sequencing. Plant Biotechnology Journal 7, 347-354.

Stellenbosch University https://scholar.sun.ac.za

Campbell, L.G., Lee, D., Shukla, K., Waite, T.A., Bartsch, D., 2016. An ecological approach to measuring the evolutionary consequences of gene flow from crops to wild or weedy relatives. Applications in Plant Sciences 4, 1-7.

Clark, L.V. & Jasieniuk, M., 2011. POLYSAT: An R package for polyploid microsatellite analysis. Molecular Ecology Resources 11, 562-566.

Collard, B.C.Y., Jahufer, M.Z.Z., Brouwer, J.B., Pang, E.C.K., 2005. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 142, 169-196.

Cordeiro, G.M., Eliott, F., McIntyre, C.L., Casu, R.E., Henry, R.J., 2006. Characterisation of single nucleotide polymorphisms in sugarcane ESTs. Theoretical and Applied Genetics 113, 331-343.

De Jong, W.S., De Jong, D.M., Bodis, M., 2003. A fluorogenic 5' nuclease (TaqMan) assay to assess dosage of a marker tightly linked to red skin colour in autotetraploid potato. Theoretical and Applied Genetics 107, 1384-1390.

Di Vecchi-Staraz, M., Laucou, V., Bruno, G., Lacombe, T., Gerber, S., Bourse, T., Boselli, M., This, P., 2009. Low level of pollen-mediated gene flow from cultivated to wild grapevine: Consequences for the evolution of the endangered subspecies *Vitis vinifera* L. subsp. *silvestris*. Journal of Heredity 100, 66-75.

Dufresne, F., Stift, M., Vergilino, R., Mable, B.K., 2014. Recent progress and challenges in population genetics of polyploid organisms: An overview of current state-of-the-art molecular and statistical tools. Molecular Ecology 23, 40-69.

Ellstrand, N.C., 1992. Gene flow by pollen: Implications for plant conservation genetics. Oikos 63, 77-86.

Fowler, C. & Hodgkin, T., 2004. Plant genetic resources for food and agriculture: Assessing global availability. Annual Review of Environment and Resources 29, 143-179.

Frankham, R., Ballou, J.D., Eldridge, M.D.B., Lacy, R.C., Ralls, K., Dudash, M.R., Fenster, C.B., 2011. Predicting the probability of outbreeding depression. Conservation Biology 25, 465-475.

Freeland, J., 2005. Molecular Ecology. Wiley, England, 129-133.

Fregene, M., Angel, F., Gomez, R., Rodriguez, F., Chavarriaga, P., Roca, W., Tohme, J., Bonierbale, M., 1997. A molecular genetic map of cassava (*Manihot esculenta* Crantz). Theoretical and Applied Genetics 95, 431-441.

Gaaliche, B., Majdoub, A., Trad, M., Mars, M., 2013. Assessment of pollen viability, germination, and tube growth in eight Tunisian caprifig (*Ficus carica* L.) cultivars. ISRN Agronomy 2013, 1-4.

Giancola, S., McKhann, H.I., Bérard, A., Camilleri, C., Durand, S., Libeau, P., Roux, F., Reboud, X., Gut, I.G., Brunel, D., 2006. Utilization of the three high-throughput SNP genotyping methods,

Stellenbosch University https://scholar.sun.ac.za

the GOOD assay, Amplifluor and TaqMan, in diploid and polyploid plants. Theoretical and Applied Genetics 112, 1115-1124.

Goldblatt, P., 1981. Chromosome numbers in legumes II. Annals of the Missouri Botanical Garden 68, 551-557.

Griffin, P.C., Robin, C., Hoffmann, A.A., 2011. A next-generation sequencing method for overcoming the multiple gene copy problem in polyploid phylogenetics, applied to *Poa* grasses. BMC Biology 9, 1-19.

Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Leger, P., Lepais, O., Lepoittevin, C., Malausa, T., Revardel, E., Salin, F., Petit, R.J., 2011. Current trends in microsatellite genotyping. Molecular Ecology Resources 11, 591-611.

Hackett, C.A., Wachira, F.N., Paul, S., Powell, W., Waugh, R., 2000. Construction of a genetic linkage map for *Camellia sinensis* (tea). Heredity 85, 346-355.

Hamilton, M., 2009. Genetic drift and effective population size, in: Hamilton, M. (Ed.), Population Genetics. Wiley-Blackwell, Oxford, UK 87-88.

Hollingsworth, P.M., Graham, S.W., Little, D.P., 2011. Choosing and using a plant DNA barcode. PLoS ONE 6, 1-14.

Jones, C., Edwards, K., Castaglione, S., 1997. Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. Molecular Breeding 3, 381-390.

Joubert, E., Joubert, M., Bester, C., De Beer, D., De Lange, J., 2011. Honeybush (*Cyclopia* spp.): From local cottage industry to global markets - The catalytic and supporting role of research. South African Journal of Botany 77, 887-907.

Kalia, R.K., Rai, M.K., Kalia, S., Singh, R., Dhawan, A.K., 2011. Microsatellite markers: An overview of the recent progress in plants. Euphytica 177, 309-334.

Khanuja, S., Shasany, A., Darokar, M., Kumar, S., 1999. Rapid isolation of DNA from dry and fresh samples of plants producing large amounts of secondary metabolites and essential oils. Plant Molecular Biology Reporter 17, 1-7.

Kloda, J.M., Dean, P., Maddren, C., MacDonald, D.W., Mayes, S., 2008. Using principle component analysis to compare genetic diversity across polyploidy levels within plant complexes: An example from British Restharrows (*Ononis spinosa* and *Ononis repens*). Heredity 100, 253-260.

Laurentin, H., 2009. Data analysis for molecular characterization of plant genetic resources. Genetic Resources and Crop Evolution 56, 277-292.

Lenormand, T., 2011. From local adaptation to speciation: Specialization and reinforcement. International Journal of Ecology 2012, 1-11.

Lowe, A., Harris, S., Ashton, P., 2004. Genetic diversity and differentiation, in: Ecological Genetics: Design, Analysis and Application. Blackwell, Oxford, UK, 52-100.

Stellenbosch University https://scholar.sun.ac.za

Lowman, M.D. & Box, J.D., 1983. Variation in leaf toughness and phenolic content among five species of Australian rain forest trees. Australian Journal of Ecology 8, 17-25.

Malgas, R.R., Potts, A.J., Oettle, N.M., Koelle, B., Todd, S.W., Verboom, G.A., Hoffman, M.T., 2010. Distribution, quantitative morphological variation and preliminary molecular analysis of different growth forms of wild rooibos (*Aspalathus linearis*) in the northern Cederberg and on the Bokkeveld Plateau. South African Journal of Botany 76, 72-81.

Meirmans, P.G., 2006. Using the AMOVA framework to estimate a standardized genetic differentiation measure. Evolution 60, 2399-2402.

Mochida, K., Yamazaki, Y., Ogihara, Y., 2003. Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags. Molecular Genetics and Genomics 270, 371-377.

Montesinos-Navarro, A., Wig, J., Pico, F.X., Tonsor, S.J., 2011. *Arabidopsis thaliana* populations show clinal variation in a climatic gradient associated with altitude. New Phytologist 189, 282-294.

Moreira, P. & Oliveira, D., 2011. Leaf age affects the quality of DNA extracted from *Dimorphandra mollis* (Fabaceae), a tropical tree species from the Cerrado region of Brazil. Genetics and Molecular Research 10, 353-358.

Motsa, M., 2016. Natural phenology, fecundity, genetic variation and seed dormancy of *Cyclopia* species (Honeybush). Unpublished D.Tech. thesis, Tshwane University of Technology, Pretoria, 159-163.

Oliver, R.E., Lazo, G.R., Lutz, J.D., Rubenfield, M.J., Tinker, N.A., Anderson, J.M., Wisniewski Morehead, N.H., Adhikary, D., Jellen, E.N., Maughan, P.J., Brown Guedira, G.L., Chao, S., Beattie, A.D., Carson, M.L., Rines, H.W., Obert, D.E., Bonman, J.M., Jackson, E.W., 2011. Model SNP development for complex genomes based on hexaploid oat using high-throughput 454 sequencing technology. BMC Genomics 12, 1-15.

Pfeiffer, T., Roschanski, A.M., Pannell, J.R., Korbecka, G., Schnittler, M., 2011. Characterization of microsatellite loci and reliable genotyping in a polyploid plant, *Mercurialis perennis* (Euphorbiaceae). Journal of Heredity 102, 479-488.

Rao, V.R. & Hodgkin, T., 2002. Genetic diversity and conservation and utilization of plant genetic resources. Plant Cell, Tissue and Organ Culture 68, 1-19.

Rundle, H.D. & Nosil, P., 2005. Ecological speciation. Ecology Letters 8, 336-352.

Scheepens, J.F., Frei, E.S., Armbruster, G.F.J., Stöcklin, J., 2012. Pollen dispersal and gene flow within and into a population of the alpine monocarpic plant *Campanula thyrsoides*. Annals of Botany 110, 1479-1488.

Stellenbosch University https://scholar.sun.ac.za

Schutte, A.L., 1997. Systematics of the genus *Cyclopia* Vent. (Fabaceae, Podalyrieae). Edinburgh Journal of Botany 54, 125-170.

Spooner, D.M., 2009. DNA barcoding will frequently fail in complicated groups: An example in wild potatoes. American Journal of Botany 96, 1177-1189.

Stift, M., Berenos, C., Kuperus, P., Van Tienderen, P.H., 2008. Segregation models for disomic, tetrasomic and intermediate inheritance in tetraploids: A general procedure applied to *Rorippa* (yellow cress) microsatellite data. Genetics 179, 2113-2123.

Štorchová, H., Hrdličková, R., Chrtek, J., Tetera, M., Fitze, D., Fehrer, J., 2000. An improved method of DNA isolation from plants collected in the field and conserved in saturated NaCl/CTAB solution. Taxon 49, 79-84.

Swarts, N.D., Sinclair, E.A., Krauss, S.L., Dixon, K.W., 2009. Genetic diversity in fragmented populations of the critically endangered spider orchid *Caladenia huegelii*: Implications for conservation. Conservation Genetics 10, 1199-1208.

Urrestarazu, J., Denancé, C., Ravon, E., Guyader, A., Guisnel, R., Feugey, L., Poncet, C., Lateur, M., Houben, P., Ordidge, M., Fernandez-Fernandez, F., Evans, K.M., Paprstein, F., Sedlak, J., Nybom, H., Garkava-Gustavsson, L., Miranda, C., Gassmann, J., Kellerhals, M., Suprun, I., Pikunova, A. V, Krasova, N.G., Torutaeva, E., 2016. Analysis of the genetic diversity and structure across a wide range of germplasm reveals prominent gene flow in apple at the European level. BMC Plant Biology 16, 1-20.

Van Wyk, B.E., 1993. Nectar sugar composition in Southern African *Papilionoideae* (Fabaceae). Biochemical Systematics and Ecology 21, 271-277.

Weising, K., Nybom, H., Wolff, K., Kahl, G., 2005. Evaluation of Molecular Marker Data, in: DNA Fingerprinting in Plants: Principles, Methods and Applications. CRC, Florida, USA, 207-292.

Wu, K.K., Burnquist, W., Sorrells, M.E., Tew, T.L., Moore, P.H., Tanksley, S.D., 1992. The detection and estimation of linkage in polyploids using single-dose restriction fragments. Theoretical and Applied Genetics 83, 294-300.

Stellenbosch University https://scholar.sun.ac.za

Appendix

Table A1. The primer information for the 27 microsatellite markers developed by Genetic Marker Services for *C. subternata*. Only 22 primers were chosen for fluorescent labelling with the dyes indicated.

Locus	F/R	Primer sequence (5'-3')	T_M	Repeat	Product size (bp)	Dye
Cys2	F	GATCGAGAAATTGAGCTTCCA	60	CT^{16}	149	N/A
Cys2	R	TTTATTTCTCGCGCGATTCT	60		149	
Cys3	F	CCATCTGCATCCTTCATTCC	60	AG^{15}	170	VIC
Cys3	R	CAAACCAAATCGAACACTGC	60		170	
Cys4	F	CTGGCCAGAGTTGTGGAAGT	60	AG^{11}	197	FAM
Cys4	R	CAGCAAGAGTCGGTTATGTTTG	60		197	
Cys5	F	CGTGTGTGTGTGTGTGAGAAA	60	AG^{14}	134	FAM
Cys5	R	GCGTTAGCGAAGGACTAGCT	60		134	
Cys6	F	TCTGTTGTAAGGCAACGAAA	60	CT^{14}	154	NED
Cys6	R	TGACATCAGATAAGGAAATTAAAAA	60		154	
Cys9	F	GCCCATGCCTATCTCTTCCT	60	TC^{7}	153	PET
Cys9	R	TGAGGGTGATTAAGGTTGAAGAA	60		153	
Cys10	F	TTGTCATGTTCTCACGCTGTT	60	TC^{8}	100	VIC
Cys10	R	TCTCTGCACACAAAAATCACG	60		100	
Cys11	F	AATTGAAAACAGAGGAACAGACA	60	AG^{6}	100	N/A
Cys11	R	CTGCCTCTCCATCTTGTAAGC	60		100	
Cys12	F	TCATTTTCATTGTTCGCTGA	60	AG^{7}	130	NED
Cys12	R	TGCAATATATCACAAATTCAGTTTT	60		130	
Cys13	F	ACACCAGAACCACCATTGAT	60	AG^{8}	148	NED
Cys13	R	GCAACTGAAAAACGGTTGCT	60		148	
Cys14	F	TCAAAGGATTAGGGGTGTGG	60	AG^{10}	124	N/A
Cys14	R	CCATGTCACACTCGAAACCA	60		124	
Cys16	F	CCCAAGTCCGATTTTGAAGA	60	AG^{9}	179	PET
Cys16	R	ACTTCGTCTACACCGCCAAG	60		179	
Cys17	F	GTACTTGAAACAAAGGAAGGTGTG	60	AG^{22}	107	FAM
Cys17	R	TGTGAACTGAACTGAACTGAACTG	60		107	
Cys18	F	ACCTTCAACCACACCCAATC	60	AG^{6}	149	N/A
Cys18	R	CCCCTCTTTCCCTCTAGGTG	60		149	
Cys20	F	AAGGGGTGTGGATTGTATGG	60	CT^{12}	141	VIC
Cys20	R	GGGTTTATCACCACGGAAAA	60		141	
Cys21	F	AACCTCTTGTCAGCTGTTG	54	AC^{23}	114	N/A
Cys21	R	GAGAGAGAGAGAGGGAGAG	50		114	
Cys21b	F	AAAGTGTTCGTAGTTTGCTTC	54	CT^{17}	131	PET
Cys21b	R	GTACTAGAATAGCTGCAAGACA	52		131	
Cys22	F	GCGTGAGGACAGAGACACAA	60	AG^{12}	112	FAM
Cys22	R	GACGGCGGAAAAGAAAATAA	60		112	
Cys23	F	AGCGGAAAACCGATCAGAAT	60	GA^{18}	151	NED

Stellenbosch University https://scholar.sun.ac.za

Locus	F/R	Primer sequence (5'-3')	T_M	Repeat	Product size (bp)	Dye
Cys23	R	GTACAAACTCGGCCACAAGA	60		151	
Cys24	F	TCACGTGTTGCTATGCAATG	60	AG^{6}	100	VIC
Cys24	R	GCCAACCAGGAAGAAAACAA	60		100	
Cys25	F	TGAATGGATGAGTGGCAAGA	60	$GT^{10}C^{1}AG^{6}$	178	FAM
Cys25	R	TCACCAGCAGATTCAGTTGAG	60		178	
Cys26	F	GGGGGCATGTTAGAGAAGTTT	60	CT^{11}	220	VIC
Cys26	R	GGCTTTGATTCATCTCAATTCAT	60		220	
Cys27	F	GCATTGCCCTTTGAAAAATC	60	CT^{15}	181	VIC
Cys27	R	GGAACAACATCCGGGTAGAA	60		181	
Cys36	F	CACCTCCCCTGCCTTACACA	60	AAC^{6}	155	PET
Cys36	R	CGGAGCTATTGTGAGGATCA	60		155	
Cys37	F	TCATGTTGGTGTTGGGAGTT	60	GT^{6}	103	FAM
Cys37	R	AAAGGTTATGTTGTTCGTTTGTAGA	60		103	
Cys88	F	TTTGCCTGCTCAGAAATCCC	60	GTT^{7}	122	PET
Cys88	R	TGTTTTCTTCTTGTTTTCAATCCA	60		122	
Cys90	F	AGCAAGAAAGTGATGGAAAA	55	$GT^{6}GA^{4}$	151	VIC
Cys90	R	GCCTAGCTTCCAGAAACTCT	55		151	

Stellenbosch University https://scholar.sun.ac.za

Table A2. DNA fingerprints for the 81 accessions of 5 *Cyclopia* species in the ARC field gene bank as determined by six microsatellite markers. Missing genotypes are denoted by -9, as this is the symbol used in POLYSAT.

Accession	Alleles	Microsatellite marker					
		Cys5	Cys10	Cys21b	Cys22	Cys25	Cys88
L04	1	125	98	119	98	180	111
	2		102	121	108	186	114
	3		106	127			136
	4		113	131			
	5						
	6						
LBP05	1	133	98	123	102	180	105
	2		102	127	108	184	111
	3		106			204	114
	4		108				117
	5						
	6						
LBP22	1	115	98	115	102	166	117
	2	117	100	117	106	178	123
	3		102	119			126
	4		106	125			133
	5		110	131			136
	6						
LBP24	1	121	98	121	98	182	111
	2	125	100	123	108	184	114
	3	129	102			186	117
	4		106			212	123
	5		108				
	6		110				
LBP26	1	117	100	115	102	166	114
	2		102	117		172	117
	3		108	119		178	123
	4			121			133
	5						
	6						
LBP41	1	121	98	115	104	184	108
	2	129	102	119		218	111
	3		104	122			114
	4		106	127			117
	5		110				123
	6						

Stellenbosch University https://scholar.sun.ac.za

Accession	Alleles	Microsatellite marker					
		Cys5	Cys10	Cys21b	Cys22	Cys25	Cys88
LGR01	1	123	98	115	98	180	111
	2	125	102	119	102	184	114
	3		104	123	108		117
	4		106				
	5						
	6						
LGR02	1	125	98	117	96	180	114
	2		104	121	108	184	117
	3		106	127			
	4						
	5						
	6						
LHK02	1	127	98	119	100	180	111
	2		102	121		184	117
	3		106	122		206	
	4		110				
	5		114				
	6						
LHK08	1	123	102	115	98	180	111
	2	125	106	119	108	184	117
	3	129		121		194	123
	4			123		208	133
	5			127			
	6			129			
LHK16	1	127	98	115	108	182	111
	2		106	129		184	114
	3					206	117
	4						120
	5						123
	6						
LHK19	1	125	98	115	108	180	114
	2		102	122		184	117
	3		104	127		206	126
	4		106	129		208	
	5						
	6						
LHK23	1	125	98	119	98	180	114
	2		102	121	108	182	117
	3		106	127		186	123
	4		110				

Accession	Alleles	Microsatellite marker					
		Cys5	Cys10	Cys21b	Cys22	Cys25	Cys88
	5						
	6						
LHK31	1	123	98	119	98	180	114
	2	127	102	122	104	182	117
	3		110	127	108	208	120
	4						
	5						
	6						
LHK32	1	125	102	119	98	180	105
	2		104	122	108	184	111
	3		106	127		206	114
	4			129		208	117
	5						
	6						
LHK35	1	121	102	115	104	180	111
	2		106	119	108	184	114
	3			121		206	117
	4			129			126
	5						130
	6						
LHK36	1	121	104	115	106	180	114
	2	127	106	117	108	182	120
	3		110	121	114		123
	4		114	123			126
	5						
	6						
LHK41	1	115	98	117	104	180	105
	2	125	102	119	106	182	111
	3		104	123	108	194	114
	4		106	127			117
	5						
	6						
LHK44	1	125	102	117	108	180	105
	2		106	119		184	114
	3		110	127			117
	4		112				
	5						
	6						

Accession	Alleles	Microsatellite marker					
		Cys5	Cys10	Cys21b	Cys22	Cys25	Cys88
LHK45	1	117	98	119	98	180	114
	2	125	102	127		184	117
	3	127	104	129		194	123
	4		106				126
	5		112				130
	6						
LHK47	1	125	102	117	98	182	114
	2		104	119	108	184	117
	3		110	131		206	150
	4		112				
	5						
	6						
LHK51	1	127	98	119	108	180	111
	2		106	123		184	114
	3			127		200	117
	4						126
	5						
	6						
LMD09	1	121	102	119	108	180	111
	2	125	110	123	116	184	117
	3		112	127		218	126
	4						
	5						
	6						
LMD11	1	123	98	117	104	180	111
	2	127	104	119		182	120
	3	129	106	121		184	123
	4						
	5						
	6						
LMD13	1	125	104	117	100	180	111
	2	129	106	119	108	182	114
	3		112	121		184	117
	4			125		218	
	5			127			
	6						
LMD14	1	123	102	115	104	180	111
	2	125	104	117	108	184	114
	3	127	106	127		218	117
	4		113				123

Accession	Alleles	Microsatellite marker					
		Cys5	Cys10	Cys21b	Cys22	Cys25	Cys88
	5						
	6						
LMD23	1	125	98	115	100	182	111
	2		106	117	104	184	114
	3			127	108	218	
	4						
	5						
	6						
LMD34	1	123	98	117	98	182	105
	2	125	104	119	104	186	111
	3		106	123			114
	4		110	127			117
	5						
	6						
LMD37	1	129	102	117	108	180	111
	2		106	129		184	114
	3		108	133			117
	4		111				126
	5						136
	6						
MBV1	1	115	98	117	76	170	105
	2		102	127			
	3						
	4						
	5						
	6						
MBV2	1	115	96	125	76	170	105
	2		98				114
	3						
	4						
	5						
	6						
MBV3	1	-9	96	117	76	170	105
	2		98	125			114
	3						
	4						
	5						
	6						

Stellenbosch University https://scholar.sun.ac.za

Accession	Alleles	Microsatellite marker					
		Cys5	Cys10	Cys21b	Cys22	Cys25	Cys88
MBV4	1	115	98	119	76	170	105
	2		102	127			111
	3						114
	4						
	5						
	6						
MBV5	1	117	96	125	76	170	105
	2		98	129			114
	3		102				
	4						
	5						
	6						
MBV6	1	115	96	117	76	170	105
	2		98	125			
	3		102				
	4						
	5						
	6						
MBV7	1	115	98	125	76	170	114
	2		102				
	3						
	4						
	5						
	6						
MBV8	1	104	96	125	-9	170	105
	2		98	127			114
	3						
	4						
	5						
	6						
MBV9	1	104	96	119	76	170	111
	2		98	129			114
	3		102				
	4						
	5						
	6						
MBV10	1	115	96	117	-9	170	111
	2	117		129			114
	3						
	4						

Stellenbosch University https://scholar.sun.ac.za

Accession	Alleles	Microsatellite marker					
		Cys5	Cys10	Cys21b	Cys22	Cys25	Cys88
	5						
	6						
Pu01	1	121	98	119	104	176	111
	2		114	121	106	180	114
	3			135	108	184	120
	4						126
	5						
	6						
Pu02	1	-9	98	127	104	182	114
	2		116				117
	3						120
	4						
	5						
	6						
Pu03	1	123	98	123	104	174	111
	2		102	127	106	180	117
	3		106		108	186	120
	4		110				
	5						
	6						
Pu04	1	119	98	117	104	182	111
	2	139		127	108		114
	3			131			117
	4			139			120
	5						
	6						
Pu05	1	119	98	123	104	182	111
	2		112	137	108		117
	3		124				123
	4						
	5						
	6						
Pu06	1	123	98	123	104	182	111
	2	131	112	139			114
	3						126
	4						
	5						
	6						

Stellenbosch University https://scholar.sun.ac.za

Accession	Alleles	Microsatellite marker					
		Cys5	Cys10	Cys21b	Cys22	Cys25	Cys88
Pu07	1	131	98	117	104	182	111
	2		124	123	108		114
	3						120
	4						
	5						
	6						
Pu08	1	119	98	123	104	182	114
	2	139	100	131	108		117
	3			139			
	4						
	5						
	6						
Pu09	1	139	98	117	104	182	108
	2		100	123	108		111
	3		112	129			120
	4			131			126
	5						
	6						
Pu10	1	123	98	123	104	182	111
	2	137	106	131	108		117
	3		112	139			
	4						
	5						
	6						
Se01	1	127	102	119	104	172	111
	2	129	110	123		190	117
	3		112	127			120
	4		120				123
	5						
	6						
Se02	1	127	102	123	104	172	-9
	2	129	106	125		178	
	3		112	127			
	4			131			
	5						
	6						
Se03	1	127	98	113	-9	168	111
	2	129	102	123		172	114
	3					176	117
	4						120

Accession	Alleles	Microsatellite marker					
		Cys5	Cys10	Cys21b	Cys22	Cys25	Cys88
	5						147
	6						
Se04	1	127	102	113	104	172	117
	2	129	106	121			120
	3		108	123			133
	4		110	125			
	5						
	6						
Se05	1	125	106	113	104	186	111
	2	129	108	121			114
	3		110	123			120
	4		112	125			130
	5		114				133
	6						
Se06	1	117	98	121	104	172	117
	2	127	104	123		182	120
	3		106	127			
	4		108	131			
	5		122				
	6						
Se07	1	115	98	113	104	180	114
	2	121	106	121			117
	3	129	110	123			123
	4						133
	5						136
	6						
Se08	1	127	102	117	104	178	111
	2		106	133			114
	3		122				
	4						
	5						
	6						
Se09	1	129	98	113	104	172	111
	2		102	123		178	114
	3		106	127			117
	4			131			120
	5						
	6						

Accession	Alleles	Microsatellite marker					
		Cys5	Cys10	Cys21b	Cys22	Cys25	Cys88
Se10	1	127	106	123	104	174	111
	2		110			180	120
	3					186	133
	4						
	5						
	6						
SGD01	1	127	98	121	114	176	111
	2	129	100	125	116	180	117
	3	131	102	129			120
	4						147
	5						
	6						
SGD02	1	109	100	117	104	178	111
	2	121	102	119	108	180	114
	3		104	125			120
	4		106				126
	5						130
	6						136
SGD03	1	123	100	111	106	172	114
	2		110	119	112	178	123
	3			121			147
	4			127			
	5						
	6						
SGD06	1	135	100	115	104	178	105
	2		102	117			117
	3			119			126
	4			123			
	5			125			
	6						
SGD07	1	117	98	117	98	176	108
	2	131	100	125	104	182	117
	3		106		108	186	120
	4		108				150
	5		110				
	6						
SGD09	1	115	94	117	94	168	114
	2	121	98	123		178	117
	3	137	110			182	120
	4		116				123

Stellenbosch University https://scholar.sun.ac.za

Accession	Alleles	Microsatellite marker					
		Cys5	Cys10	Cys21b	Cys22	Cys25	Cys88
	5						126
	6						139
SHL02	1	115	100	119	96	168	114
	2	137	102	123	104	176	120
	3		104		106	186	123
	4		108		110	188	
	5						
	6						
SHL03	1	117	100	119	104	168	114
	2	133	106	121	106	172	117
	3		110	125			120
	4			129			
	5						
	6						
SKB03	1	117	100	117	98	180	111
	2	127	102	119	106		114
	3	135	106	121	110		120
	4		110	125			123
	5						
	6						
SKB04	1	117	100	117	106	188	111
	2		106	119			114
	3		108	123			120
	4						130
	5						133
	6						
SKB05	1	117	98	103	96	168	111
	2	125	100	117	98	178	114
	3	127	102	125		180	117
	4		106				120
	5		108				123
	6						130
SKB06	1	121	98	119	98	164	111
	2	127	100	123	104	182	120
	3	135	102	125	112		126
	4		104	129	114		130
	5		106				133
	6						

Accession	Alleles	Microsatellite marker					
		Cys5	Cys10	Cys21b	Cys22	Cys25	Cys88
SKB07	1	117	100	117	96	188	114
	2		102	119	104		117
	3		104	123	106		120
	4		114				
	5						
	6						
SKB11	1	131	100	117	104	174	114
	2		106	119		180	117
	3		112	125			120
	4						126
	5						
	6						
SKB13	1	113	98	117	104	186	114
	2		100	119	108		117
	3		112		110		
	4						
	5						
	6						
SKB14	1	113	100	109	104	168	114
	2		102	117	114		117
	3		104	121			120
	4		108				139
	5		110				
	6						
SKB15	1	117	96	117	106	176	111
	2	119	100	119		186	117
	3		110	125		188	120
	4			131			126
	5						133
	6						
SKB18	1	117	100	117	104	176	111
	2	125	108	119	106		117
	3		110	123			123
	4		112				126
	5						133
	6						
STB01	1	127	98	117	114	178	114
	2	137	100	119		180	123
	3		102	121			126
	4		106	125			130

Stellenbosch University https://scholar.sun.ac.za

Accession	Alleles	Microsatellite marker					
		Cys5	Cys10	Cys21b	Cys22	Cys25	Cys88
	5		112				144
	6						
STB101	1	133	100	117	98	182	114
	2		102	119	100		120
	3		104	121			126
	4		112	123			
	5			125			
	6						
STB102	1	117	100	113	100	188	114
	2	123	102	117	104		126
	3	135	110	119			130
	4			121			
	5			125			
	6						
STB103	1	121	100	117	104	182	117
	2		102	119			120
	3		104	123			123
	4		108	125			126
	5		110				
	6						

Accessions starting with: S = *C. subternata*; L = *C. longifolia*; M = *C. maculata*; Pu = *C. pubescens*; Se = *C. sessiliflora*.

Stellenbosch University https://scholar.sun.ac.za

Table A3. Locations from where the *Cyclopia* species accessions, currently in the ARC field gene bank situated at Nietvoorbij and Elsenburg in Stellenbosch, were originally obtained.

Accession	Species	Farm selected	Closest Town	GPS	Plantation/ Wild
GG03	*C. genistoides*	Gouriqua	Gouritsmond	S 34.35955 E 021.7932	P
GG09	*C. genistoides*	Gouriqua	Gouritsmond	S 34.35955 E 021.7932	P
GG31	*C. genistoides*	Gouriqua	Gouritsmond	S 34.35955 E 021.7932	P
GG34	*C. genistoides*	Gouriqua	Gouritsmond	S 34.35955 E 021.7932	P
GG53	*C. genistoides*	Gouriqua	Gouritsmond	S 34.35955 E 021.7932	P
GK01	*C. genistoides*	Koksrivier	Pearly beach	S 34.69463 E 019.6094	P
GK02	*C. genistoides*	Koksrivier	Pearly beach	S 34.69463 E 019.6094	P
GK03	*C. genistoides*	Koksrivier	Pearly beach	S 34.69463 E 019.6094	P
GK04	*C. genistoides*	Koksrivier	Pearly beach	S 34.69463 E 019.6094	P
GK05	*C. genistoides*	Koksrivier	Pearly beach	S 34.69463 E 019.6094	P
GK06	*C. genistoides*	Koksrivier	Pearly beach	S 34.69463 E 019.6094	P
GK07	*C. genistoides*	Koksrivier	Pearly beach	S 34.69463 E 019.6094	P
GK08	*C. genistoides*	Koksrivier	Pearly beach	S 34.69463 E 019.6094	P
GT01	*C. genistoides*	Toekomst	Bredasdorp	S 34.54340 E 019.8798	P
GT02	*C. genistoides*	Toekomst	Bredasdorp	S 34.54340 E 019.8798	P
N/A	*C. intermedia*	Haarlem	Haarlem	S 33.75265 E 023.3598	P
L04	*C. longifolia*	Moffetdam	Longmore Thornhill	S 33.48500 E 025.9014	W
LBP05	*C. longifolia*	Barendspas	Longmore Thornhill	S 33.86943 E 025.1416	W
LBP22	*C. longifolia*	Barendspas	Longmore Thornhill	S 33.86943 E 025.1416	W
LBP24	*C. longifolia*	Barendspas	Longmore Thornhill	S 33.86943 E 025.1416	W
LBP26	*C. longifolia*	Barendspas	Longmore Thornhill	S 33.86943 E 025.1416	W
LBP41	*C. longifolia*	Barendspas	Longmore Thornhill	S 33.86943 E 025.1416	W
LGR01	*C. longifolia*	Geelhoutboomrivier	Longmore Thornhill	S 33.83339 E 025.1503	W
LGR02	*C. longifolia*	Geelhoutboomrivier	Longmore Thornhill	S 33.83339 E 025.1503	W
LHK02	*C. longifolia*	Helsekloof	Longmore Thornhill	S 33.50831 E 25.63200	W
LHK08	*C. longifolia*	Helsekloof	Longmore Thornhill	S 33.50831 E 25.63200	W
LHK16	*C. longifolia*	Helsekloof	Longmore Thornhill	S 33.50831 E 25.63200	W
LHK19	*C. longifolia*	Helsekloof	Longmore Thornhill	S 33.50831 E 25.63200	W

Stellenbosch University https://scholar.sun.ac.za

Accession	Species	Farm selected	Closest Town	GPS	Plantation/ Wild
LHK23	*C. longifolia*	Helsekloof	Longmore Thornhill	S 33.50831 E 25.63200	W
LHK31	*C. longifolia*	Helsekloof	Longmore Thornhill	S 33.50831 E 25.6320	W
LHK32	*C. longifolia*	Helsekloof	Longmore Thornhill	S 33.50831 E 25.63200	W
LHK35	*C. longifolia*	Helsekloof	Longmore Thornhill	S 33.50831 E 25.63200	W
LHK36	*C. longifolia*	Helsekloof	Longmore Thornhill	S 33.50831 E 25.63200	W
LHK41	*C. longifolia*	Helsekloof	Longmore Thornhill	S 33.50831 E 25.63200	W
LHK44	*C. longifolia*	Helsekloof	Longmore Thornhill	S 33.50831 E 25.63200	W
LHK45	*C. longifolia*	Helsekloof	Longmore Thornhill	S 33.50831 E 25.63200	W
LHK47	*C. longifolia*	Helsekloof	Longmore Thornhill	S 33.50831 E 25.63200	W
LHK51	*C. longifolia*	Helsekloof	Longmore Thornhill	S 33.50831 E 25.63200	W
LMD09	*C. longifolia*	Moffetdam	Longmore Thornhill	S 33.48500 E 25.90140	W
LMD11	*C. longifolia*	Moffetdam	Longmore Thornhill	S 33.48500 E 25.90140	W
LMD13	*C. longifolia*	Moffetdam	Longmore Thornhill	S 33.48500 E 25.90140	W
LMD14	*C. longifolia*	Moffetdam	Longmore Thornhill	S 33.48500 E 25.90140	W
LMD23	*C. longifolia*	Moffetdam	Longmore Thornhill	S 33.48500 E 25.90140	W
LMD30	*C. longifolia*	Moffetdam	Longmore Thornhill	S 33.48500 E 25.90140	W
LMD34	*C. longifolia*	Moffetdam	Longmore Thornhill	S 33.48500 E 25.90140	W
LMD37	*C. longifolia*	Moffetdam	Longmore Thornhill	S 33.48500 E 25.90140	W
MBV01	*C. maculata*	Bereaville	Genadendal	S 34.04450 E 019.5122	P
MBV02	*C. maculata*	Bereaville	Genadendal	S 34.04450 E 019.5122	P
MBV03	*C. maculata*	Bereaville	Genadendal	S 34.04450 E 019.5122	P
MBV04	*C. maculata*	Bereaville	Genadendal	S 34.04450 E 019.5122	P
MBV05	*C. maculata*	Bereaville	Genadendal	S 34.04450 E 019.5122	P
MBV06	*C. maculata*	Bereaville	Genadendal	S 34.04450 E 019.5122	P
MBV07	*C. maculata*	Bereaville	Genadendal	S 34.04450 E 019.5122	P
MBV08	*C. maculata*	Bereaville	Genadendal	S 34.04450 E 019.5122	P
MBV09	*C. maculata*	Bereaville	Genadendal	S 34.04450 E 019.5122	P
MBV10	*C. maculata*	Bereaville	Genadendal	S 34.04450 E 019.5122	P
N/A	*C. pubescens*	Baywest road reserve	Port Elizabeth	S 33.94658 E 025.4644	W

Stellenbosch University https://scholar.sun.ac.za

Accession	Species	Farm selected	Closest Town	GPS	Plantation/ Wild
N/A	*C. sessiliflora*	Farm of Mr van Niekerk	Heidelberg	S 34.01729 E 020.9186	W
SGD01	*C. subternata*	Groendal	Louterwater	S 33.75399 E 023.5830	P
SGD02	*C. subternata*	Groendal	Louterwater	S 33.75399 E 023.5830	P
SGD03	*C. subternata*	Groendal	Louterwater	S 33.75399 E 023.5830	P
SGD06	*C. subternata*	Groendal	Louterwater	S 33.75399 E 023.5830	P
SGD07	*C. subternata*	Groendal	Louterwater	S 33.75399 E 023.5830	P
SGD09	*C. subternata*	Groendal	Louterwater	S 33.75399 E 023.5830	P
SHL02	*C. subternata*	Haarlem	Haarlem	S 33.75265 E 023.3598	P
SHL03	*C. subternata*	Haarlem	Haarlem	S 33.75265 E 023.3598	P
SKB03	*C. subternata*	Kanetberg	Barrydale	S 33.93763 E 021.0656	P
SKB04	*C. subternata*	Kanetberg	Barrydale	S 33.93763 E 021.0656	P
SKB05	*C. subternata*	Kanetberg	Barrydale	S 33.93763 E 021.0656	P
SKB06	*C. subternata*	Kanetberg	Barrydale	S 33.93763 E 021.0656	P
SKB07	*C. subternata*	Kanetberg	Barrydale	S 33.93763 E 021.0656	P
SKB11	*C. subternata*	Kanetberg	Barrydale	S 33.93763 E 021.0656	P
SKB13	*C. subternata*	Kanetberg	Barrydale	S 33.93763 E 021.0656	P
SKB14	*C. subternata*	Kanetberg	Barrydale	S 33.93763 E 021.0656	P
SKB15	*C. subternata*	Kanetberg	Barrydale	S 33.93763 E 021.0656	P
SKB18	*C. subternata*	Kanetberg	Barrydale	S 33.93763 E 021.0656	P
STB01	*C. subternata*	Tolbos	Napier	S 34.47784 E 019.7571	P
STB101	*C. subternata*	Tolbos	Napier	S 34.47784 E 019.7571	P
STB102	*C. subternata*	Tolbos	Napier	S 34.47784 E 019.7571	P
STB103	*C. subternata*	Tolbos	Napier	S 34.47784 E 019.7571	P

Note: Accession numbers indicate plants cloned from the original selected plant in plantation or the wild. If no accession number is present, it indicates seedlings from seed collected in plantations or the wild.

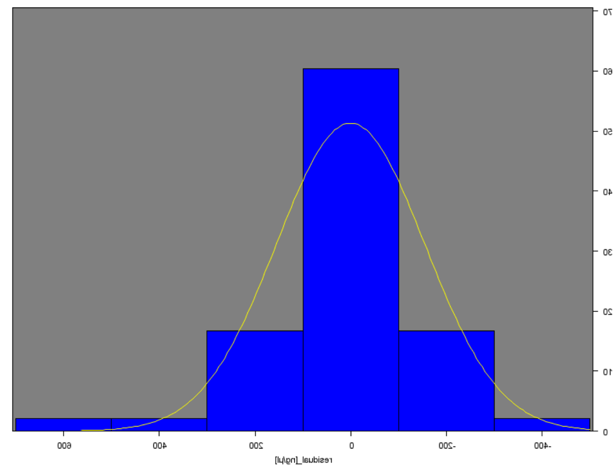
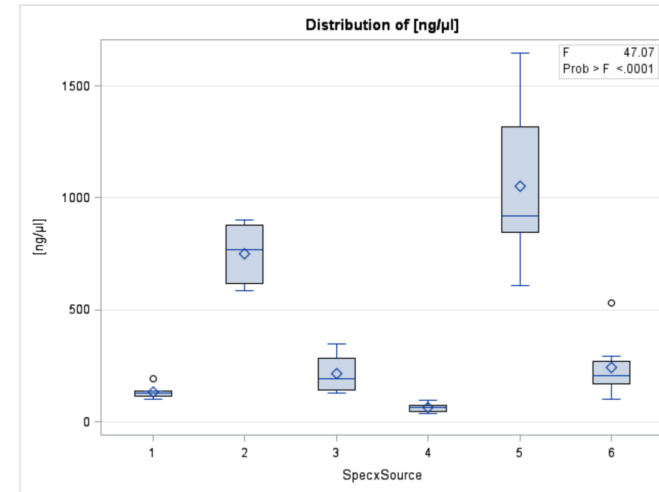
Stellenbosch University https://scholar.sun.ac.za

Table A4. Characteristics of the three genetic distance matrices used throughout the study.

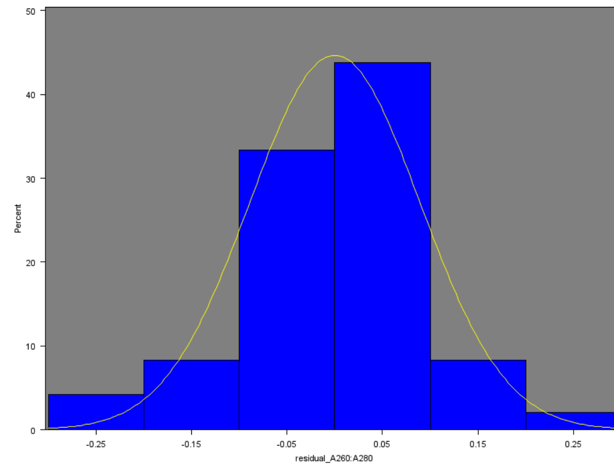
	Genetic distance matrix		
Characteristics	**Euclidian**	**Lynch**	**Bruvo**
Description	Calculate similarity based on the presence/absence of binary data	Band-sharing dissimilarity index	Calculate relative distances between microsatellite genotypes while taking stepwise mutational events into account
Advantage	Can be applied independent of ploidy level and inheritance	Shared absence does not contribute to similarity – can be used independent of level of expected relatedness	Can deal with dosage uncertainty
Disadvantage	Shared absence contribute to similarity Loss of genetic information as allele dosage is not taken into account	Very dependent on mode of inheritance	Can lead to overestimation of genetic distance between individuals
Autopolyploid/ allopolyploid	N/A	Autopolyploid	Mainly autopolyploid, but can be applied to allopolyploids if no alleles at the two isoloci are within five mutational steps of each other
Similarity/ distance indices	Similarity	Similarity	Distance
Software used	GenAlEx	POLYSAT	POLYSAT
Type of data	Binary	Genotypic	Genotypic

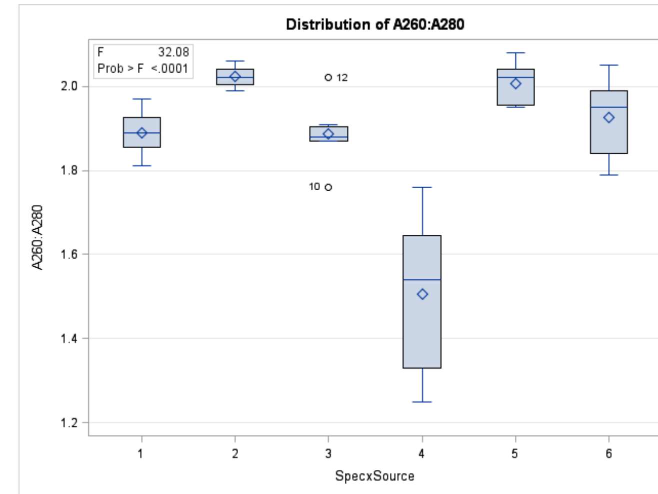
Stellenbosch University https://scholar.sun.ac.za

Table A5. Visual representation of distribution of residuals and homoscedasticity of *Cyclopia* DNA quality data before and after transformation with appropriate functions for three dependent variables. Supplementary to the data given in Table 2.1.

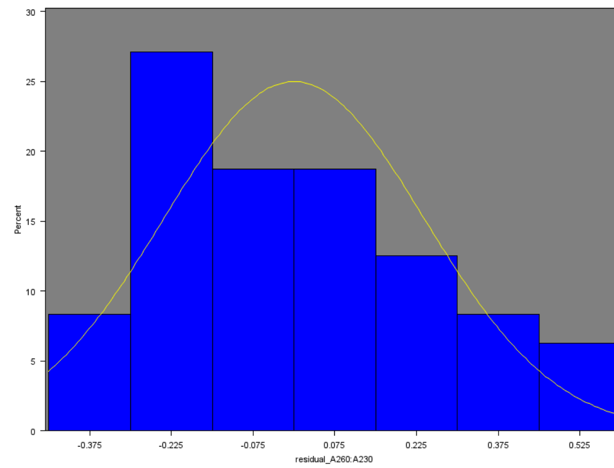
Before transformation		
	Normality	**Homoscedasticity**
[DNA] in ng/µl		

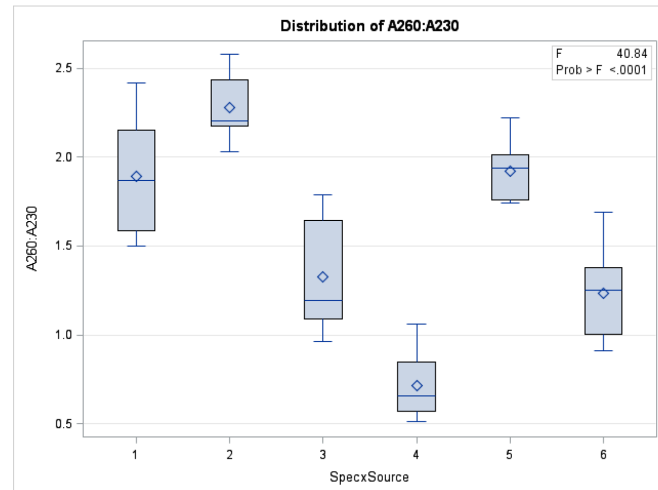
A260:A280



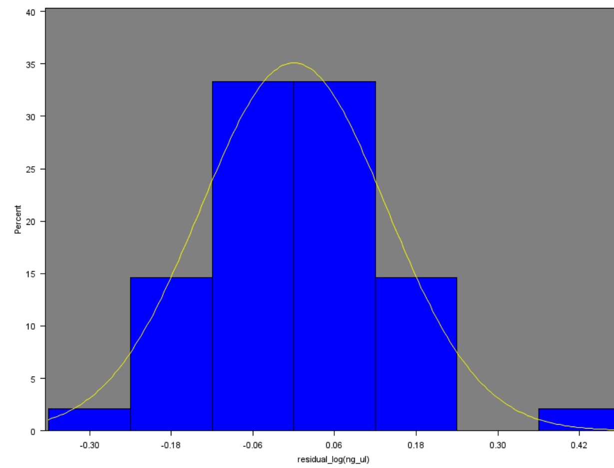
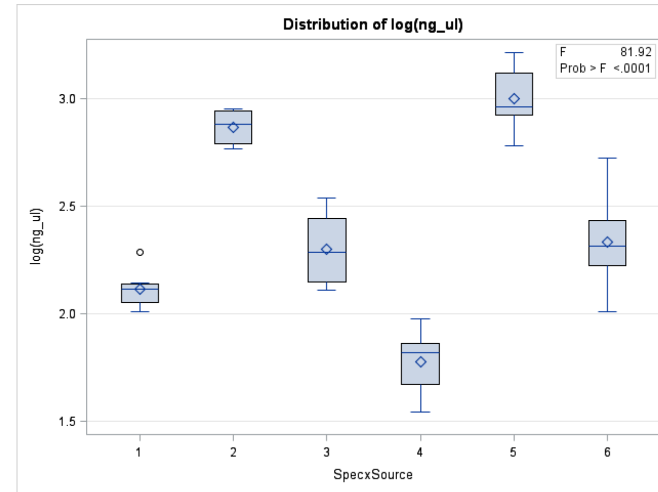
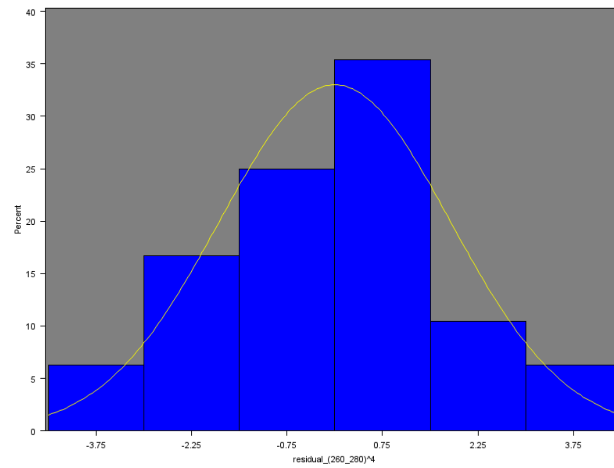
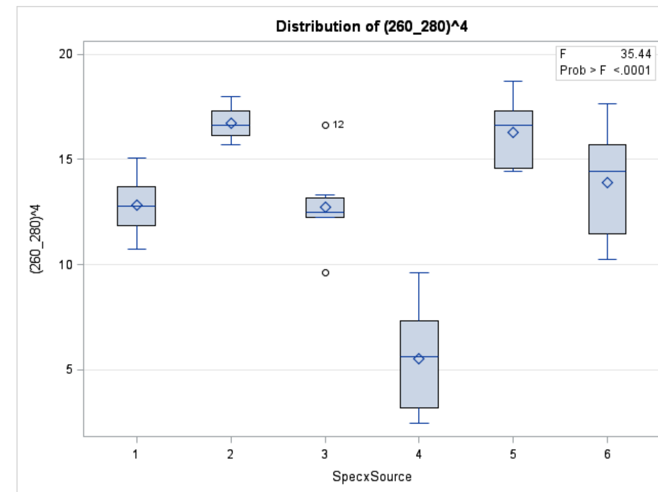


A_{260}:A_{230}





Stellenbosch University https://scholar.sun.ac.za

After transformation		
	Normality	**Homoscedasticity**
Log (ng/µl)		
$(A_{260}:A_{280})^4$		

Stellenbosch University https://scholar.sun.ac.za

$\sqrt{(A_{260}:A_{230})}$