PHYLOGEOGRAPHY, POPULATION STRUCTURE AND PHYTOCHEMICAL ANALYSIS OF THE AFROMONTANE ENDEMIC *Prunus africana* (Hook.f. Kalkman)

CAROLINE ANNE CHELEDI KADU (BSc, MSc)

I84/10257/2007

A THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF DOCTOR OF PHILOSOPHY (MOLECULAR GENETICS) IN THE SCHOOL OF PURE AND APPLIED SCIENCES OF KENYATTA UNIVERSITY

JUNE, 2021

DECLARATION

This thesis is my original work and has not been presented for a degree in any other University or any other award.

DECLARATION

This thesis is my original work and has not been presented for a degree in any other University or any other award.

CARLINE ANNE CHELEDI KADU

(I84/10257/2007)

Signature .... Date 23.06.2021

I/We confirm that the work reported in this thesis was carried out by the student under my/our supervision.

Prof. Geoffrey M. Muluvi
South Eastern Kenya University,
Nairobi, Kenya

Signature .... Date 14-6-2021

Prof. Thomas Geburek
Federal Research and Training Centre for Forests, Natural Hazards and Landscapes, Vienna, Austria

Signature .... Date 8/6/2021

Dr. Laura Snook
Alliance Bioversity - CIAT,
Rome, Italy

Signature .... Date 10 June 2021

Dr. Alice Muchugi
International Livestock Research Institute (ILRI)
Addis Ababa, Ethiopia

Signature .... Date 14/06/2021

DEDICATION

To my husband Geofrey Onyango Kadu, my late parents Lorna Musungu and Arthur Henry Kiliru Agufa and siblings Jacqueline, Violet, Clare, Doreen and the late James. Your loving support and encouragement were my constant inspiration.

ACKNOWLEDGEMENTS

Thanks be to God who leads us in triumphant victory no matter the odds. Sincere gratitude to my supervisors and colleagues in Federal Research Training Centre for Forests, Natural Hazards and Landscapes who made my stay in Vienna most enjoyable. Many thanks to Prof. Thomas Geburek who tirelessly gave his time and advice from the inception to the completion of this work. Extreme gratitude to Dr Heino Konrad who was my daily adviser from settling me in with immigration to proof reading the scripts. Sincere gratitude to Dr. Silvio who gave invaluable input in project development, data analysis and writing as well as his friendship. Most appreciated friendship of Mr. Thomas Thalmayer who provided technical assistance, and together with his dear wife Sonja opened up their home and friends to me. Many thanks to Monica Lotter, Renate Slunsky, Berthold Heinze, Michael Mengl, Frank Franner, Rudolf Litschauer, Karl Sieberer, Karen Robitschek, Gerald Golesch, Barbara Fussi, Ilse Strohschneider and all in Mariabrun who were so courteous and gracious to me. Vielen Dank.

Sincere gratitude to my supervisors and colleagues in Bioversity International. Deep appreciation to Dr. Laura Snook who supported me from the beginning to the completion of this work. Many thanks to Dr. Lex Thompson who left so soon after recruiting me, Elizabeth Goldberg, Elisabetta Rossetti, Lyn Hunt and all in the Rome campus who facilitated my settling in and all logistics throughout the programme. Many thanks to Dr. Oscar Eyog-Matig who facilitated the sample collections and all the other logistics involved and gave valuable input into the development of the project. Many thanks to Dr. Barbara Vincenti who was very instrumental in the inception of the project and in giving feedback in the manuscripts. Many thanks to Dr. Judy Loo who gave full support throughout the project. To Dr. Joseph Jojo Baidu-Forson, Caroline Genga, Josephine Osea, Patrick Litswa, John Tindi and all in the regional office in Nairobi who facilitated logistics for me and gave moral support many many thanks. To Jacinta Kimwaki, Humprey Keah, Caroline Mbogo and all on ICRAF campus who gave me unlimited support many thanks. To all the national partners who facilitated the supply of this magnitude of range wide samples many thanks.

Sincere gratitude to my supervisor and the staff in Kenyatta University. Deep appreciation and gratitude to Prof. Geoffrey Muluvi who tirelessly walked with and mentored me through this journey. Many thanks to Dr. Mathew P. Ngugi, Dr. Alice Muchugi and the colleagues in the Biochemistry, Microbiolgy and Biotechnology Department who made this achievement possible. Many thanks to the members of the Postgraduate board who facilitated all things necessary to make this achievement possible.

Sincere gratitude to Pastors Candace and Thomas Manning, Pastors Chapa and Shawn Ginder and all the fellowship at Vienna Christian Centre International who prayed with me and supported me all through my stay in Vienna. The good Lord Bless you.

Sincere gratitude to Pastors Nelly and Charles Mulema, Pastors Mercy and Simon Muasya, the diaspora and all in Nairobi Gospel Centre International who prayed with Geofrey and I and encouraged us throughout this journey. Our good Lord Almighty, Lord Jesus Christ and the Holy Spirit bless you all. Now to him who is able to establish this work and do immeasurably more than all we can ask or imagine, according to his power that is at work within us, to him be glory, power and honour now and forever. Amen

TABLE OF CONTENTS

TITLE i

DECLARATION ii

DEDICATION iii

ACKNOWLEDGEMENTS iv

TABLE OF CONTENTS v

LIST OF TABLES viii

LIST OF FIGURES x

ABBREVIATIONS, SYMBOLS AND ACRONYMS xii

ABSTRACT xiv

CHAPTER ONE 1

INTRODUCTION 1

1.1 Background information 1
1.2 Statement of the problem 5
1.3 Justification 7
1.4 Hypotheses 8
1.5 Objectives 8
1.5.1 Broad Objective 8
1.5.2 Specific Objectives 9

CHAPTER TWO 10

LITERATURE REVIEW 10

2.1 Taxonomy, distribution and ecology of *Prunus africana* 10
2.1.1 Taxonomy 10
2.1.2 Distribution 11
2.1.3 Ecology 14
2.2 Phylogeography 16
2.3 Vegetation history 17
2.4 Population Genetics 19
2.5 Intraspecific phytochemical variations 21
2.6 Molecular markers 22
2.6.1 Microsatellites (SSR) 23
2.6.2 Chloroplast microsatellite markers (cpSSR markers) 24

CHAPTER THREE 26

MATERIALS AND METHODS 26

3.1 Population sampling and DNA isolation 26

3.2 cpDNA Amplification and genotyping of the range wide study 28
3.3 nSSR genotyping and fragment analysis of range wide study 31
3.4 Intraspecific chemical diversity 33
3.4.1 Plant Material 33
3.4.2 Bark processing 34
3.4.3 Phytochemical analysis of the bark samples 34
3.5 Data Management and statistical analysis 37
3.5.1 Phylogeography and Population Structure revealed by cpSSR 37
3.5.1.1 cpDNA variation and haplotype construction 37
3.5.1.2 Population structure and phylogeography as revealed by cpSSR 38
3.5.1.3 Population history and coalescence as revealed by cpSSR 39
3.5.2 Population structure and history of *Prunus africana* as revealed by nuclear microsatellites (nSSR) 41
3.5.2.1 nSSR fragment analysis and diversity measures 41
3.5.2.2 Population differentiation 42
3.5.2.3 Individual based population assignment 43
3.5.2.4 Genetic barriers among populations 44
3.5.2.5 Estimation of recent migration rates among regions. 44
3.5.2.6 Likelihood estimates of the direction and magnitude of historic gene flow 46
3.5.3 Intraspecific Phytochemical diversity 47
3.5.3.1 Genetic data 47
3.5.3.2 Statistical analysis 48

CHAPTER FOUR 50

RESULTS 50

4.1 Phylogeography and population structure revealed by cpSSR 50
4.1.1 cpDNA variation and haplotype construction 50
4.1.2 Population structure and phylogeography 54
4.1.3 Population history and coalescence 57
4.2 Population structure and history as revealed by nuclear microsatellites (nSSR) 62
4.2.1 Nuclear microsatellite locus characterization and genetic diversity .. 62
4.2.2 Genetic differentiation and barriers to gene flow 65
4.2.3 Individual based population assignment 68
4.2.4 Recent Gene flow 70
4.2.5 Historic gene flow 71
4.3 Intraspecific chemical diversity 79

CHAPTER FIVE 88

DISCUSSION, CONCLUSION AND RECOMMENDATTIONS 88

5.1 Discussion 88
5.1.1 Introduction 88
5.1.2 Phylogeography and Population Structure revealed by cpSSR 88
5.1.2.1 Genetic diversity and population structure based on cpDNA 88
5.1.2.2 Phylogeography of *Prunus africana* – migration and dispersal within Africa 90

5.1.3 Population structure and history of *Prunus africana* as revealed by nuclear microsatellites (nSSR) 96
5.1.3.1 Genetic diversity and population differentiation 96
5.1.3.2 Recent patterns of migration 98
5.1.3.3 The Rift valley as major gene flow barrier – major shift of a gene flow barrier in East Africa 99
5.1.3.4 *Prunus africana* populations in Madagascar 104
5.1.3.5 Derived populations in southern and western Africa 105
5.1.4 Intraspecific chemical diversity 106
5.2 CONCLUSIONS 112
5.3 RECOMMENDATIONS 113

REFERENCES 116

LIST OF TABLES

Table 3.1: Description and location of the 32 populations of P. africana**.** 27

Table 3.2: cpSSR primer sequences used for the phylogeographucal study. 29

Table 3.3: nSSR primer sequences employed in the population structure analysis. 32

Table 4.1: Haplotype construction and frequency in 32 *Prunus africana* populations revealed by 4 cpDNA indels (Loci 1 – 4) and 3 cpDNA microsatellites (Loci 5 – 7).. 51

Table 4.2: Description, location and haplotype diversity measures of investigated populations. 53

Table 4.3: Genetic structure among *Prunus africana* populations analysed by SAMOVA.. 58

Table 4.4: Analysis of variance of population structure of 32 *Prunus africana* populations based on SAMOVA structure analysis using K = 2 to 20 groups. 59

Table 4.5: Results of coalescence modelling using a Bayesian analysis.. 60

Table 4.6: List of the 30 natural populations of *Prunus africana* analysed in this study... 64

Table 4.7: Allelic richness (*Rs*; after rarefaction to 13 individuals) and number of alleles (*a*) per locus and population in *Prunus africana*... 67

Table 4.8: Twenty-five natural populations of *Prunus africana* from the five population groups ranked by mean allelic richness. 70

Table 4.9: Null allele occurrence in 25 populations of *Prunus africana*......... 74

Table 4.10: Pairwise differentiation by F_{ST} among *Prunus africana* populations (below diagonal), and inter-population Nei's unbiased genetic distance (above diagonal). 75

Table 4.11: Results of analysis of molecular variance (AMOVA) from data of six nuclear microsatellite loci of 25 *Prunus africana* populations from seven countries. Groups correspond to the 5 biogeographic regions detected. 76

Table 4.12: Recent gene flow estimates (as rate of migration from population *i* into population *j*, average from five computations). 78

Table 4.13: Results of historic gene flow analysis among population groups derived from MIGRATE; θ is the estimated mutation-scaled population size,. 79

Table 4.14: Correlations between the concentrations of different bark constituents of *Prunus africana* 82

Table 4.15: Correlations between genetic parameters [Number of effective alleles (N_{e_c}), unbiased haplotypic diversity (h_c), number of effective nuclear alleles (N_{e_n}), ... 84

Table 4.16: Significance level of Mantel-tests between chemical and genetic distances based on chloroplast (Φ_{PT}) and nuclear microsatellites (R_{ST}), respectively. 85

Table 4.17: Description and location of 20 populations of *Prunus africana* sampled. 86

Table 4.18: Estimates of LOD/LOQ by means of the signal to noise ratios of the standards. 87

LIST OF FIGURES

Figure 1.1: *Prunus africana* tree....... 2

Figure 1.2: *Prunus africana* wildings showing shiny leaves and characteristic red stem 3

Figure 2.1: Distribution of *Prunus africana* adapted from Hall *et al.*, (2000). 13

Figure 2.2: *Prunus africana* from Madagascar 15

Figure 2.3: *Prunus africana* fruit from Lari, Kenya....... 15

Figure 2.4: Perfect flowers of *Prunus africana* from Lari, Kenya 17

Figure 3.1: Herbarium samples of *Prunus africana* from Kabale, Uganda 28

Figure 3.2: Ethidium Bromide-stained agarose gel (1% [w/v]) showing Qiagen DNA extraction of 24 *Prunus africana* samples....... 28

Figure 3.3: Microsatellite genotyping using the CEQ 8000 Genetic analysis system 9.0 (Beckman Coultier, USA) 31

Figure 3.4: Bark samples harvesting from *Prunus africana* in Kabale, Uganda....... 35

Figure 4.1: Approximate geographic distribution of *Prunus africana* (grey shaded area),. 52

Figure 4.2: Genealogical network revealing 22 haplotypes constructed from 19 alleles obtained from 4 cpDNA indels (Loci 1–4) and 3 cpDNA microsatellites (Loci 5–7) within 32 *Prunus africana* populations.. 56

Figure 4.3: Node support for various population splitting events identified by coalescence modelling using a Bayesian analysis.. 59

Figure 4.4: Time of population splitting events identified by coalescence modelling using a Bayesian analysis for two estimates of generation time (8 years – solid line; 15 years – broken line)....... 60

Figure 4.5: Hypothesised migration route of *Prunus africana* as deduced from coalescence modelling and the observed phylogeographic – genetic structure.. 61

Figure 4.6: Consensus tree based on 1,000 bootstrap replicates of Cavalli-Sforza chord distance matrices of nuclear SSR data for 25 *Prunus africana* populations. 65

Figure 4.7: Principal coordinate analysis showing the multivariate relationships of the *Prunus africana* populations along the first two axes which explain a total of 67% variation.......................... 68

Figure 4.8: Sampling locations, group membership and barriers to gene flow for 30 natural populations of *Prunus africana* (natural distribution range indicated by grey shaded area)........................ 72

Figure 4.9: Population assignment using structure reveals 5 clusters. 77

Figure 4.10: Mean chemical analyte concentration (mg/kg) of *Prunus africana* individuals from 20 populations within the natural range of the species ranked in ascending order. 81

Figure 4.11: Analysis of variance of chemical analytes in *Prunus africana*.. 82

Figure 4.12: Relative mean of the various chemical compounds................... 83

Figure 4.13: Clustering of populations based on chemical distance 85

Figure 4.14: *Prunus africana* bark sampled throughout the natural range...... 87

ABBREVIATIONS, SYMBOLS AND ACRONYMS

ANOVA	Analysis of variance
AMOVA	Analysis of molecular variance
BATWING	Bayesian Analysis of Trees with Internal Node Generation
CITES	Convention on International Trade in Endangered Species of Wild Fauna and Flora
cp	chloroplast
cpDNA	chloroplast deoxyribonucleic acid
δF	maximum change of inbreeding coefficient
δm	maximum change of migration rate
δp	maximum change of allele frequency
δ	average distance of pollen dispersal
Δ	delta
D	observed density
D_e	effective density
F_{IS}	fixation index
F_{SC}	differentiation among populations within groups
F_{ST}	a measure of population differentiation (intra-class kinship coefficient)
F_{CT}	differentiation among groups of populations
G_{ST}	differentiation based on allelic frequencies
H_o	observed level of heterozygosity
H_e	expected level of heterozygosity
h_c	haplotypic diversity
h_t	total gene diversity
h_s	average within population gene diversity
HWE	Hardy Weinberg Equilibrium
LD	linkage disequilibrium
LOD	method detection limits
LOQ	method quantification limits
MCMC	Markov chain Monte Carlo
ML	maximum likelihood
M	Migration rate
m	immigration rate
μ	Mutation rate
N	Population size
N_e	effective population size
N_{e_c}	effective haplotypic population size
N_{e_n}	effective nuclear population size
N_{ST}	differentiation among populations using ‘ordered alleles’ i.e. considering the phylogenetic distance between alleles (or haplotypes)
θ	scaled population size
PCoA	principal coordinate analysis
Φ_{PT}	population differentiation estimates for haploid loci
R_{ST}	F_{ST} analogue based on allele size
SAMOVA	Spatial analysis of molecular variance
SSR	simple sequence repeats (microsatellites)

cpSSR	chloroplast simple sequence repeats
nSSR	nuclear simple sequence repeats
SMT	southern migratory tract
θ	Estimates of scaled population size
UPGMA	Unweighted Pair Group Method with Arithmetic Mean

ABSTRACT

Prunus africana is a highly valued Afromontane species for its bark whose products are used for treatment of benign prostatic hypertrophy (BPH). Increasing demand of *P. africana* bark has led to overharvesting of the species natural populations thus threatening its survival in the wild. Consequently, cultivation of the species on farms offers an alternative to sustain the *P. africana* herbal market. Information on existing genetic and phytochemical variations within and among the species provenances is crucial in initiating *P. africana* farming and its conservation. This study proposed to generate such important information by assessing the genetic diversity using cpDNA and SSR markers and phytochemical variations from 32 populations sampled across the species distribution range. The colonisation history of the Afromontane tree *P. africana* was inferred using seven cpDNA loci of 582 individuals from 32 populations that revealed 22 haplotypes. Bayesian coalescence modelling suggested that 'east' and 'west' African types split early during southward migration of the species, while furthermore recent splitting events occurred among populations in the East of the continent. The high genetic similarity found between western Uganda and West African populations indicates that a former Afromontane migration corridor may have existed through Equatorial Africa. To gain new insights into the population structure, population history and historical gene flow among populations of this Afromontane species, six nuclear microsatellite markers were used on a total of 459 samples from 30 populations and revealed strong population divergence among Afromontane regions. Five main regions were identified: West Africa, East Africa west of the Eastern rift valley, East Africa east of the Eastern Rift Valley, southern Africa and Madagascar. The eastern Rift Valley posed as the second most important barrier to gene flow in *Prunus africana*. Discordance between recent and historic gene flow estimates and patterns obtained from the cpDNA study gives evidence for a major shift of gene flow barriers in East Africa. To determine the correlation between phytochemical diversity in active compounds of *Prunus africana* bark with geographic or genetic distribution, the content of the different antioxidants derived from bark extracts was assayed by means of GC_MSD in 20 of the 32 populations sampled. The average concentration [mg/kg w/w] of the bark constituents assayed in increasing order was: lauric acid (18), myristic acid (22), n-docosanol (25), ferulic acid (49), β-sitostenone (198), β-sitosterol (490), and ursolic acid (743). The concentrations of many bark constituents were significantly intercorrelated. Only haplotypic diversity was positively correlated with the variation of the content of β-sitosterol and the fixation index (F_{IS}) was negatively correlated with the content variation of ferulic acid. While genetic distances based on pairwise Φ_{PT} for chloroplast and R_{ST} for nuclear genetic differentiation and Euclidian chemical distances were subjected to Mantel's tests to associate concentration of bark constituents with the molecular phylogenetic pattern. For n-docosanol and ursolic acid, weak, but significant correlations were detected for both chloroplast as well as nuclear data, revealing that the molecular phylogenetic pattern is co-expressed by means of certain bark constituents. This correlation finding suggest that some populations exhibiting high amount of key phytochemical constituents thought to be important for BHP control could be a focus on *P. africana* domestication and improvement for on farm planting. The detailed implications of these results on the sustainable use and conservation of *Prunus africana* are further discussed.

CHAPTER ONE

INTRODUCTION

1.1 Background information

Prunus africana (Hook.f.) Kalkman (formerly *Pygeum africanum*; Hook.f.) (Rosaceae) is a long-lived monoecious evergreen tree that may grow to a height of more than 40 m and a diameter greater than 1 m (Figure 1.1). It is easily identified in the natural habitat by its shiny leaves and characterisitic red colour of young branches (Figure 1.2). It is a geographically widespread Afromontane tree reported from 22 African countries, restricted to African mountain forests "islands", which are generally above 1,000 m in altitude. These include the mountain highlands of mainland Africa (Angola, Cameroon, Ethiopia, Kenya, Malawi, Nigeria, Somalia, South Africa, Sudan, Swaziland, Tanzania, Uganda, Zaire and Zimbabwe) and outlying Islands (Bioko, Grand Comore, Madagascar and Sao Tome) (Cunningham and Mbenkum, 1993; Dawson and Powell, 1999).

Prunus africana leaves, roots and bark are used as traditional medicine in Africa to treat chest pain, malaria and fevers (Cunningham and Mbenkum, 1993) as well as a purgative for cattle and extract for arrow poison (Sunderland and Nkefor, 1997). The timber is hard and durable and is used for the manufacture of various household products such as axes, hoes and furniture (Sunderland and Nkefor, 1997). Extracts from the bark are also used in the treatment of benign prostatic hyperplasia, a disorder of the prostate gland

common in older men. These extracts significantly improve urologic symptoms (Ishani *et al*., 2000; Edgar *et al*., 2007) having anti-proliferative

Figure 1.1: *Prunus africana* tree

and apoptotic effects in the prostate (Quiles *et al*., 2010). The efficacy of bark extract is believed to be a synergistic effect of a cocktail of known and

unknown compounds (Simons *et al*., 1998). Among the known ones, three groups are of great importance: (1) phytosterols, especially β-sitosterol, having

Figure 1.2: *Prunus africana* wildings showing shiny leaves and characteristic red stem

anti-inflammatory properties inhibiting the swelling of the prostate gland (Carbin *et al*., 1990), (2) pentacyclic triterpenoids providing anti-edematous

activity (Bombardelli and Morazzani, 1997) and (3) ferulic acid esters or their chemical derivates possessing a powerful hypocholesterolemic activity in the prostate (Seetharamaiah and Chandrasekhara, 1993) and antitumor activity (Kampa *et al*., 2004). The medicinal property of *P. africana* bark in the treatment of benign prostatic hyperplasia (Simons *et al*., 1998) has led to extensive and commercially important international trade, with the species being one of the most valuable medicinal exports from Africa (Cunningham *et al*., 1997). From two initial brand-name products produced in France and Italy, there are now at least 40 brand-name products using *P. africana* bark extract. This has led to the over-the-counter retail value of the trade in *P. africana* herbal preparations of well over US$220 million/yr (Cunningham *et al*., 1997) and may be even higher today.

The discontinuous distribution of the species in montane "islands" over such a wide geographical range almost certainly implies extensive genetic variation among populations. This variation can be exploited in cultivation programs by selecting populations that produce the most bark and have the highest concentration of extract in bark. Molecular markers which have been employed in previous studies on the species have shown a great partitioning of genetic variation across geographical distance due to the wide scale but disjunct distribution of *P. africana* in Afromontane forest islands (Barker *et al*., 1994; Dawson and Powell, 1999; Muchugi *et al*., 2006). This suggests that a considerable number of individual sites may be required for effective conservation of genetic variation within the species. However, a more complete picture of the existing genetic structure within the species range

would facilitate development of national and international strategies for conservation and sustainable use.

An understanding of the phylogeography and range-wide spatial genetic structure of *P. africana* as revealed by molecular markers and the corresponding correlation to phytochemical components would provide essential data for the selection of gene sources for revegetation, planting at appropriate densities and development of management programs. The assessment of the spatial genetic structure within and between populations of *P. africana* is a prerequisite for developing optimum management strategies of this important genetic resource.

1.2 Statement of the problem

Prunus africana populations have experienced high commercial exploitation in the past decades in various parts of its natural distribution. Recent studies on the impacts of wild harvest on *P. africana* populations in Cameroon show that the current exploitation is unsustainable (Stewart, 2001; Stewart, 2003a: Stewart, 2003b) and debarking of *P. africana* often occurs within Afromontane forest habitat of global conservation significance (Stattersfield *et al*., 1998; Butynski and Koster, 1994). In addition to selective harvesting, trees are lost due to general forest clearance for agriculture, since the distribution of *P. africana* often coincides with regions of high human population density (Dawson *et al*., 2000), and through climate change that restricts regeneration at the edges of the species' natural range (Geldenhuys 1981 for South Africa). As

a result, *P. africana* is listed under Appendix II of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES), and various populations are considered as vulnerable according to International Union for the Conservation of Nature and Natural Resources criteria (IUCN, 2004).

African governments are thus faced with the challenge of preventing this decline continuing without negatively affecting the potential economic and social benefits of *P. africana*. Concerted efforts at developing national and international sustainable management interventions need to be implemented to conserve this genetic resource. This has not been achieved to date due to the limited knowledge of the species' ecology, biology, distribution and genetic diversity. In addition, limited knowledge on policies and practices necessary for adopting preservation, conservation, and management appropriate for sustainable exploitation has also played a key role.

A 'conservation by cultivation' approach has been recommended to forest-dependant rural communities to capture the great genetic variation across the species geographical range (Barker *et al.*, 1994; Dawson and Powell, 1999). However, this requires a proper supply of germplasm and, consequently, a sound understanding of the spatial distribution of genetic variation and its correlation with the content of important chemical compounds in the bark. Various studies have implicated the existence of genetic differentiation among populations (Dawson and Powell, 1999; Muchugi *et al.*, 2006) and have inferred the influence of gene flow on these genetic structures but to date there

has not been a study inferring the phylogeography and genetic differentiation among populations covering the geographic distribution range of the species simultaneously in correlation with phytochemical components. Knowledge of the phylogeography, extent of genetic variation and correlation to phytochemical components will allow for better selection of which populations to target for bark harvesting.

1.3 Justification

Prunus africana is a species of great commercial significance (Phillips and Meilleur, 1998) due to the proliferation of bark products used for an increasing demand for treatment of benign prostatic hypertrophy (BPH). Increasing demand of *P. africana* bark has led to rapid decline of the species natural populations. Thus, there is a need to develop alternative sources of *P. africana* bark for use by small farmers as a cash-crop. Although cultivation by local farmers is taking place in some countries (Franzel *et al*., 2009), all bark entering the international market is from wild harvest. To prevent this decline from continuing without inhibiting the potential economic and social benefits, it is crucial to develop management policies and practices appropriate for sustainable exploitation.

Molecular markers have been employed in previous studies and have shown a great partitioning of genetic variation across geographical distance due to the wide scale but disjunct distribution of *P. africana* in Afromontane forest islands (Barker *et al*., 1994; Dawson and Powell, 1999). This suggests that a

considerable number of individual sites may be required for effective conservation of genetic variation within the species. However, the overall picture of genetic variation through the species natural range is incomplete and current findings do not allow comparing populations with respect to phytochemical variations. This knowledge is important for conservation and sustainable use of *P. africana*, whereby populations of known phytochemical diversity can be used in planting programmes. This research employed the cpDNA SSRs and nuclear SSRs to carry out a range-wide characterization of the macro- and micro-spatial distribution of genetic diversity in the species and assesed phytochemical differentiation across the species range.

1.4 Hypotheses

i. Phylogeography exists within natural populations of *Prunus africana.*
ii. Distinct patterns of population genetic structure exist among natural populations of *P. africana* within Africa.
iii. *P. africana* exhibits intraspecific phytochemical diversity.

1.5 Objectives

1.5.1 Broad Objective

To elucidate phylogeography, population structure and phytochemical diversity of natural populations of *P. africana*.

1.5.2 Specific Objectives

i. To investigate the phylogeography of 32 *P. africana* natural populations within Africa using cpDNA SSR markers.

ii. To study patterns of population genetic structure in 30 natural populations of *P. africana* within Africa using nuclear microsatellites and Bayesian methods.

iii. To evaluate the chemical diversity of phytochemical extracts of 20 *P. africana* natural populations within Africa.

CHAPTER TWO

LITERATURE REVIEW

2.1 Taxonomy, distribution and ecology of *Prunus africana*

2.1.1 Taxonomy

Prunus africana belongs to the large Rosaceae family designated by Kalkman, (1988) as being a very distinct and isolated family. Twenty two groups of the genera have been considered and these groups are related to the four subfamilies (Spiraeoideae, Rosoideae, Maloideae and Prunoideae) of Schulze-Menz, (1964). Thorne (1992) recognizes an additional subfamily, the Quilajoideae, separating this from other genera placed in the Spiraeoideae and preferring the name Amygdaloideae to Prunoideae. Recent phylogenetic analysis based on sequence variation in the chloroplast encoded *rbc*L gene (Morgan *et al*., 1994), supports the general integrity of three of Schluze-Menz's subfamilies: the Rosoideae, Maloideae and Amygdaloideae. In a study restricted to the Amygdaloideae, Evans and Dickinson (1999) recognize Exochorda (5 species), Oemleria (1 species), Prinsepia (3-5 species), *Prunus* (ca 200 species) and probably Maddenia (3-5 species) as the genera in a monophyletic Amygdaloideae.

The species of *Prunus* are divided between two subgenera, Padus and Laurocerasus (Kalkman, 1988). Padus contains the deciduous species and Laurocerasus contains the evergreen species. *P. africana*, the only African and Madagascan species, is included in Laurocerasus as are all the Eurasian and tropical Asian species of the genus. Unfortunately, *P. africana* has not been

included in any study attempting to clarify relationships between different species of *Prunus* using the power of modern molecular/genetic tools. Inferences based only on morphology must be treated with reservations considering the uniformity in features across the genus. *P. africana* displays characteristics which suggest it may be rather isolated. It is the sole African species and Kalkman's (1988) suggested migration path would reduce the likelihood of recent ancestry shared with the species he found morphologically similar. Though various antioxidants have been identified in different members of the genus *Prunus* and found to be have anti-inflammatory, anticancer or antiviral effects (e.g., Donovan *et al*., 1998; Takeoka and Dao, 2003; Ferretti *et al*., 2010), it appears to be chemically distinctive – hence its pharmaceutical importance. It is the largest of all the species of *Prunus* (Cunningham *et al*., 1997) and its wood structure, too, is distinctive: “quite unlike any of the European woods belonging to the same natural order” (Stone, 1924).

2.1.2 Distribution

There have been various speculations about the processes leading to the current distribution of *P. africana*. Various migration paths have been inferred based on the present distribution (Aubréville, 1976; Kalkman, 1988). While Aubréville (1976) suggests a Laurasian origin of *Prunus* (only considering subgenus Laurocerasus, which includes *P. africana* in his work) with subsequent movement into Africa from the north-east, Kalkman (1988) proposes a Gondwanian origin of species of the tribe Pruneae (containing the genera *Maddenia*, *Prinsepia* and *Prunus*) with northwards movement along a

path starting in Australia, South America and Africa. Another open question is the phylogeographical history of the species within Africa, which seems to be strongly connected to past vegetation dynamics (White, 1983). Based on a genetic study using RAPD markers, Muchugi *et al*. (2006) suggested that long distant seed dispersal could have been responsible for the occurrence of *P. africana* in outlying islands. In cooler periods *P. africana* could have migrated along White's Southern Migratory Tract (SMT; White, 1983), which passes through southern DRC, Zambia, northern Angola and Gabon (White, 1983; 1993) (Figure 2.1). However, while the stepping stone model explained the migration through the SMT, it could not explain the large genetic differences observed among the west and eastern Kenya populations. Muchugi *et al*. (2006) thus concluded that the Aubréville path of distribution was more likely, supporting the fact that the rift valley could have acted as a barrier to gene flow in cooler times.

With the exception of some disjunct populations, *P. africana* is confined to 'High Africa', extending from latitude 33°40'S in South Africa to latitude 11°55' N near the Gulf of Aden. Within 'High Africa' the species is restricted primarily to mountains or volcanic regions, particularly the Great escarpment in south-eastern Africa and the Eastern and Western Rift Valley systems and Ethiopian Highlands of eastern Africa. The species can withstand mild or infrequent frost, but not severe or prolonged frost. The range in southern Africa, according to comments with voucher specimens, is between 600 m and 1,000 m but elevations as low as 60 m have been reported (Geldenhuys, 1981) in the Bloukrans River Gorge, South Africa. Both the range in elevation

(1,000–3,000 m) and the maximum elevation (3,500 m) are greatest in equatorial Africa. *P. africana* is restricted to

Figure 2.1: Distribution of *Prunus africana* adapted from Hall *et al*., (2000).

mountainous terrain within High Africa and occupies a distribution typified by a relatively temperate climate and mean annual rainfall from 500–700 mm (high latitudes) to over 3,000 mm (low latitudes).

2.1.3 Ecology

Prunus africana has often been recognized as a pioneer (e.g. White, 1983) or early secondary (e.g. Geldenhuys, 1981) species. Such a status is consistent with the view of Von Breitenbach (1965) that it is initially fairly fast growing: mean annual height increment 0.6–0.8 m, presumably for South Africa. Indications exist that the onset of flowering may be at an age below 10 years, although commonly it appears to be later (Mbonyimana, 1988; White, 1983). The early growth rate, early flowering and frequently reported association with disturbance (Eggeling, 1940; Von Breitenbach, 1965; Geldenhuys, 1981) are consistent with White's (1983) listing as a “nomad”, a term emphasizing adaptations for easily reaching (and persisting in) new forest gaps. A combination of nomadic behaviour and extended longevity explains why *P. africana* so often occurs as sparsely distributed large individuals in closed forest communities. The reproductive seasonality of *P. africana* differs within the three zones. Within the equatorial zones it is year-round while in the North and South it coincides with the cool dry conditions. The fruiting period seems to be the 2–3 months following flowering and is usually associated with rainfall. Figure 2.2 and 2.3 show slight morphological variation between fruits from Madagascar and fruit from Lari, Kenya respectively. Fruits are red drupes

Figure 2.2: *Prunus africana* from Madagascar

Figure 2.3: *Prunus africana* fruit from Lari, Kenya

with one seed. The flowers are always perfect and no unisexual functionality has been reported although the state is known in other species of *Prunus*

(Kalkman, 1965; Figure 2.4). The nature of the pollen grains, presence of nectar and scent of flower suggest cross-pollination by animal vectors possibly insects. However, the light weight of the pollen could result in some wind pollination. The effectiveness of mechanisms and floral structure prevent self-pollination (Hall *et al*., 2000). Despite its wide distribution in Africa and on islands off-shore, there is little published information on the seed dispersal process. Potential dispersal agents include animals and birds.

2.2 Phylogeography

Phylogeography is a study that uses genetic data to understand the history of populations (Beheregaray, 2008). Most species display some degree of population structure that can be interpreted in geographic and chronological contexts. Deciphering spatial and temporal components of population structure and interpreting the evolutionary and ecological processes responsible are major goals of phylogeography. Afromontane forests are regions of high biodiversity and harbour important ecosystems in Africa (Taylor, 2015). Using *Prunus africana* as an example, the chain of events that have led to the current distribution of species within the Afromontane forests could be elucidated. Patterns of population genetic structure within plant species over a wide distribution range are shaped by the integration of many factors (Yan *et al*., 2009). Some of these are intrinsic to the species life history (breeding system,

Figure 2.4: Perfect flowers of *Prunus africana* from Lari, Kenya

modes of seed and pollen dispersal, life form, gregariousness), whereas others are perturbations induced by natural processes (ice ages, climatic stochasticity) or human impact (habitat fragmentation, global change) (Heuertz *et al*., 2004). Climatic fluctuations that could have happened in the past possibly as far back as the Quaternary could have produced important changes in the distribution ranges of many plant species and thereby strongly affected their large-scale genetic structure (Debout *et al*., 2011).

2.3 Vegetation history

Afromontane forests are regions of great biodiversity and centres of endemism (Hall *et al*., 2009). Long term environmental stability of these forests has been proposed as a mechanism for the accumulation and persistence of species during glacial periods, resulting in the diverse forests observed today (Finch *et al*., 2009). However, significant climate changes which occurred within Africa from the Last Glacial Maximum (LGM) to the present times caused

considerable modifications of species ranges (e.g. Livingstone, 1975; Castañeda *et al*., 2009). In comparison to latitudinal range shifts where a loss of genetic diversity due to repeated bottlenecks occurs (Hewitt, 1996), range shifts in mountain regions are less likely to affect genetic diversity. Here, colonization occurs over much shorter distances (Ehrich *et al*., 2007) and population sizes might be more stable (Ehrich *et al*., 2007). On the other hand, African mountain massifs are characterized by their isolation from each other, thus limiting gene flow and facilitating population subdivision and independent evolution of montane forest species (Hewitt, 2000).

Unfortunately, past vegetation patterns can only be reconstructed partly by pollen analysis due to lack of lake sediments, particularly in Central Africa. Existing data suggest that during the LGM around 20,000 years BP montane vegetation extended to wider areas of lower elevation not only in East Africa, but also in northern Angola and south of the Democratic Republic of Congo (Van Zinderen Bakker and Clark, 1962; Caratini and Giresse, 1979; Elenga, 1987;Maley, 1987; 1989). Later, during the early part of the Holocene, an increase in temperature caused an expansion of evergreen forests through tropical Africa, while at the same time montane vegetation disappeared from these areas and very likely the distribution of montane species retracted to higher elevations (Livingstone, 1975; Maley, 1991). The high similarities found between East and West Africa montane forests in their insect, bird and plant communities have long been suggested as evidence for a former connection via lowland bridges (Hall, 1973; Moreau, 1966). In addition to ecological indicators for past vegetation patterns, the genetic relatedness of

populations of typical representatives of the Afromontane flora could improve knowledge on the historical processes responsible for current distribution.

2.4 Population Genetics

Most of the theory of population genetics is concerned with the frequencies of alleles whose substitution has a discontinuous effect on the phenotype of an organism: the so called “major genes” which determine single locus Mendelian polymorphisms (Powell, 1992; Jones *et al*., 1997). However, the majority of phenotypic variation that is observed in populations is quantitative, determined by many genes of individual small effect, whose expression is modified by the environment (Mather and Jinks, 1982). Genes determining quantitative characters are inherited in the same manner as major genes, differing only in the magnitude of their individual effect on phenotype (Mather and Jinks, 1982; Powell, 1992).

Genetic structure refers to the distribution of genetic variation within and between populations. Population geneticists have long recognized that the genetic diversity present in a taxon is generally hierarchically structured. In addition to differences among individuals within any one population, variation may be structured among populations within a given geographical region and among populations from different geographical regions (Holsinger and Mason-Gamer, 1996). The distribution of variation among populations is the result of interactions among several evolutionary and biological factors such as selection to varying environmental pressures, effective population size,

breeding system and the ability of a taxon to disperse pollen and seed (Hamrick, 1989). In general, selection should increase population differentiation, as should genetic drift associated with small population sizes. Though generalisations must be made with caution, species without-breeding mechanisms and greater pollen or seed movement are likely to be less differentiated than selfing species or those with restricted gene flow (Bonnin *et al*., 1996). Levels and partitioning of genetic variation may differ significantly among species with different geographic ranges, life forms and taxonomic affinities (Loveless, 1992). Woody perennials generally possess more variation at marker loci than short-lived plants (Loveless, 1992).

Traditionally, genetic resources have been characterized by a combination of morphological and agronomic traits (Nei, 1975). The process is time consuming as many of the morphological descriptors can only be assessed at maturity (Wilde *et al*., 1992). Moreover these characteristics may be influenced by environmental factors and may therefore not reflect true genetic similarities or differences (Dawson *et al*., 1993). The effectiveness of this approach alone to estimate genetic diversity and genetic structure has therefore been questioned by several authors (Gottlieb, 1977; Brown, 1979; Russell *et al*., 1993). Molecular and biochemical markers that are not subject to environmental influences provide an opportunity to examine precisely the genetic relationship between accessions (Wilde *et al*., 1992).

2.5 Intraspecific phytochemical variations

Intraspecific variation of chemical compounds is common to many plant species, and often shows defined spatial patterns, reflecting environmental difference within the range of a species. Hence, chemical races among the plant kingdom are often described and provide taxonomists with a powerful tool (Bohm, 2009). One of the most probably cited examples of spatio-chemical diversity in secondary metabolites, including its genetic background, is the variation of cyanogenesis in *Trifolium repens* throughout Europe (e.g., Hayden and Parker, 2002 and references cited therein). Despite the vast body of literature, spatio-chemical differences among populations in plants providing natural products have only been occasionally described and studies linking a spatial pattern in compound concentrations with genetic differences among populations are even more limited.

Correlations between the metabolic contents and DNA fingerprints among geographically distinct plant populations were reported for volatile oil constituents in *Ocimum gratissimum* (Vieira *et al*., 2001), phenolic acids in *Fructus xanthii* (Han *et al*., 2008) or essential oils in *Primula ovaliformis* (Ma *et al*., 2005). However, Trindade *et al*., (2008) reported contrary results in *Thymus caespititius.* In their study, the variation of volatile oils was not associated with the genetic substructure of the populations; however, the study was based on a very small sample of only 31 individuals.

Disregarding whether the concentration of certain bark constituents derived from *P. africana* may be linked to genetic data or not, knowledge on regional variation of bark compounds is essential in order to optimize sustainable utilization and conservation of this endangered tree species (Heywood, 2002). A previous pilot study employing three regions of origin, i.e. Cameroon, Zaire and Madagascar has already shown that the composition of *P. africana* extracts depends on the origin of the bark (Martinelli *et al*., 1986). It is therefore probable that *P. africana* would also exhibit intraspecific plant chemical diversity, i.e., variation within and among populations.

2.6 Molecular markers

Morphological markers displaying Mendelian inheritance have been used for quite a long time to characterize genetic variation. They provide valuable insights into genetic relationships within and among natural populations of tree species (Marley and Parker, 1993; Harrier *et al*., 1997), but their utility is limited by environmental plasticity (Mather and Jinks, 1982). Molecular approaches can provide a more detailed insight into the underlying genetic diversity within and among populations, since analysis may be used to provide data that delineate complex genetic structures that cannot be determined from morphological studies alone, but are important for management of a taxon.

Currently, a variety of these molecular procedures provide unprecedented power to explore the genomes of diverse plant populations including forest tree species. These include biochemical markers such as isozyme markers (Jana

and Pietzrak, 1988; Jelinski and Cheliak, 1992), DNA markers such as the restriction fragment length polymorphisms (RFLP; Botstein *et al.*, 1980) and methods based on the polymerase chain reaction (PCR) (Vosberg, 1989) such as random amplified polymorphic DNA (RAPD; Williams *et al.*, 1990;Welsh and McClelland, 1990), simple sequence repeats polymorphism (SSR; Tautz, 1989), amplified fragment length polymorphism (AFLP; Vos *et al.*, 1995) and restriction endonuclease cleavage of sequence-tagged sites (RFLP-PCR; Tragoonrung *et al.*, 1992).

2.6.1 Microsatellites (SSR)

Microsatellites markers (SSRs, simple sequence repeats) are becoming the markers of choice for fingerprinting purposes due to their high polymorphism and reproducibility, codominance, amenability to automation and high transferability to other crosses and related species (Tautz, 1989; Bell and Ecker, 1994; Guilford *et al.*, 1997; Sosinski *et al.*, 2000). Thus, individual SSR genetic linkage maps can generally be used within a species and in related taxa. SSR markers have been used for genetic linkage mapping (e.g., Bell and Ecker, 1994; Xu *et al.*, 1999), comparative mapping (Diwan *et al.*, 2000), merging maps across populations (Cregan et al., 1999a), genotype identification and variety protection (Rongwen *et al.*, 1995), quantitative trait locus analysis, and MAS in breeding programs (Bradshaw *et al.*, 1998) and in enabling the study of contemporary gene flow as mediated by pollen and seeds (Dow and Ashley, 1996; 1998); Streiff *et al.*, 1999). SSRs have been developed for many plant species, including rice (Panaud *et al.*, 1996); Chen *et*

al., 1997), maize (Senior *et al*., 1998), potato (Milbourne *et al*., 1998), *Arabidopsis thaliana* L. (Bell and Ecker, 1994), barley (Ramsay *et al*., 2000), soybean (Cregan et al., 1999b), tomato (Broun and Tanksley, 1996), cherry (Cantini *et al*., 2001), peach (Cipriani *et al*., 1999; Sosinski *et al*., 2000), Citrus spp. (Sankar and Moore, 2001), and peanut (Hopkins *et al*., 1999). However, the high cost of initial isolation and characterization hinder their application in many plant species. Several microsatellite markers have already been reported in various *Prunus* species and found to be transferable within species, in other genera belonging to the Rosaceae and even in other families. A high number of nSSRs (approximately 185 in Dirlewanger *et al*., 2004) have been developed and mapped in different *Prunus* species. A high proportion of these markers are universal in *Prunus* species (e.g. Schueler *et al*., 2003) and some have already been tested in *P. africana* (Farwig *et al*., 2008).

2.6.2 Chloroplast microsatellite markers (cpSSR markers)

Organelle genomes (chloroplast (cp) DNA and mitochondrial (mt) DNA) have been intensively studied during the last decade to analyse haplotypic variation and phylogeography in a wide range of plants. Among the Rosaceae tree genera, *Sorbus* (Oddou-Muratorio *et al*., 2001), *Prunus* (Mohanty *et al*., 2001), and *Malus* (Vornam and Gebhardt, 2000), among others have been studied with organelle DNA. In all Rosaceae species, chloroplast genomes are maternally inherited (e.g. Ishikawa and Shimamoto, 1992; (Matsumoto *et al*., 1997); Oddou-Muratorio *et al*., 2001). The advantage of using cpDNA markers is that they provide information on the spatial distribution of genetic variation

as a consequence of seed mediated gene flow. Further, the smaller effective population size of cpDNA, compared with nuclear effective population size (Birky *et al*., 1989), makes cpDNA particularly useful for detecting the signature of founding events and other population bottlenecks associated with invasion and range expansion (McCauley *et al*., 2003).

In phylogeographical studies, cpDNA markers can be used to test between the effects of migration and mutation (Petit *et al*., 1993b). cpDNA marker polymorphisms are especially apt for investigating population history that occurred in the more distant past (Hewitt, 2000). Chloroplast markers, however, could not be fully utilized for the assessment of closely related chloroplast genomes until highly polymorphic genome regions were identified (McCauley, 1995). Chloroplast simple sequence repeats (cpSSRs), i.e. polymorphic mononucleotide repeats, were developed for such genetic analyses in the 1990s (McCauley, 1995; Vendramin and Ziegenhagen, 1997). This technology is based on highly polymorphic regions and has been used for phylogeographical analyses of many trees species, for example *Pinus* (Afzal-Rafii and Dodd, 2007), *Hagenia* (Ayele *et al*., 2009) and *Adansonia digitata* (Pock Tsy *et al*., 2009)

CHAPTER THREE

MATERIALS AND METHODS

3.1 Population sampling and DNA isolation

Leaf samples were obtained from 32 natural populations of *P. africana* covering the main block of the species' geographic range across sub-Saharan Africa (Figure 2.1; Table 3.1). This included the West Africa region: Cameroon (3 sites), Equatorial Guinea (Bioko) and Nigeria (1 site each); the East Africa region: Kenya (9 sites), Tanzania (6 sites), Uganda (4 sites); and the southern Africa region: Madagascar (3 sites), South Africa (2 sites) and Zimbabwe (3 sites). From each population, leaf samples from 6 to 36 trees were collected and dried with silica gel. Import, export and phytosanitary certificates were obtained according to national and CITES regulations.

Herbarium samples were collected for future reference (Figure 3.1). For each sampled tree, total genomic DNA was extracted from 40 to 60 mg of dry leaf material using the QIAGEN Plant mini kit (QIAGEN, Valencia, CA, USA) according to the manufacturer's instructions with slight modifications (Farwig *et al*., 2008). DNA quality was determined by Ethidium bromide stained 1% w/v agarose gel electrophoresis (Figure 3.2) and stored at 4° C. DNA quantity was determined with a Nanodrop ND 1000 spectrophotometer (NanoDrop Technologies).

Table 3.1: Description and location of the 32 populations of P. africana**.**

No.	Name of Population	Code	Country	Longitude	Latitude	n
1	Ngashie-Mt Oku	CA	Cameroon	10.5092	6.2048	19
2	Lower Mann's Spring, Mt Cameroon	CB	Cameroon	9.1192	4.1346	19
3	Mt Danoua	CC	Cameroon	11.9413	7.1871	20
4	Moka	GQ	Equatorial Guinea	8.6559	3.3625	18
5	Chuka, Central province	KA	Kenya	37.3645	-0.1798	20
6	Kinale, Central province	KB	Kenya	36.4151	-0.5278	19
7	Kapcherop, Cherangani Forest, Rift Valley	KC	Kenya	35.2161	1.0315	20
8	Kakamega Forest, Western Province	KD	Kenya	34.519	0.1415	20
9	Londiani, Rift Valley	KE	Kenya	35.1836	-0.2737	20
10	Ol Danyo Sambuk, Central Province	KF	Kenya	37.1501	-1.0744	19
11	Taita Hills, Coast Province	KG	Kenya	38.2088	-3.2462	20
12	Lari, Central Province	KT	Kenya	36.687	-1.0311	12
13	Kibiri forest, Western Province	KS	Kenya	34.8719	0.1502	16
14	Marovoay	MA	Madagascar	48.3069	-18.802	5
15	Lakato forest	MB	Madagascar	48.2779	-19.197	33
16	Antsahabiraoka	MC	Madagascar	49.2164	-14.402	18
17	Ngel Nyaki Forest Reserve, Nigeria	NG	Nigeria	11.0271	7.066	9
18	Mpumalanga	SA	South Africa	30.7306	-25.023	19
19	KwaZulu-Natal	SB	South Africa	30.2725	-29.287	17
20	Meru Catchment Forest	TA	Tanzania	36.8074	-3.2925	19
21	Kilimanjaro Catchment Forest Reserve	TB	Tanzania	37.5237	-3.0036	17
22	Kindoroko Catchment Reserve	TC	Tanzania	37.6267	-3.7389	14
23	Shume/Magamba Catchment Forest Reserve	TD	Tanzania	38.2521	-4.7544	20
24	Kidabaga	TE	Tanzania	35.9296	-8.1101	15
25	Udzungwa	TT	Tanzania	36.7791	-7.7668	16
26	Kibale Forest Natural Park	UA	Uganda	30.357	0.5644	20
27	Kalinzu Forest Reserve	UB	Uganda	30.1101	-0.3742	20
28	Bwindi Forest	UC	Uganda	29.7754	-1.0476	19
29	Mabira Forest	UD	Uganda	33.015	0.3806	20
30	Nyanga National Park	ZWA	Zimbabwe	32.7401	-18.287	20
31	Cashel Valley Chimanimani	ZWB	Zimbabwe	32.8004	-19.578	20
32	Chirinda forest Reserve Chipinge	ZWC	Zimbabwe	32.695	-20.409	19

Figure 3.1: Herbarium samples of *Prunus africana* from Kabale, Uganda

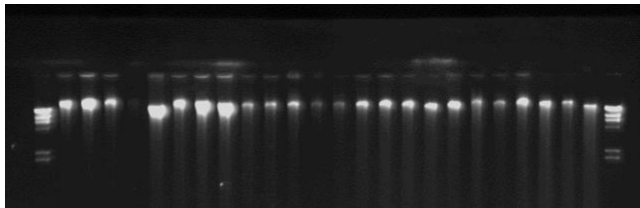

Figure 3.2: Ethidium Bromide-stained agarose gel (1% [w/v]) showing Qiagen DNA extraction of 24 *Prunus africana* samples.

3.2 cpDNA Amplification and genotyping of the range wide study

Chloroplast DNA variation was analysed with three microsatellite loci (cpSSR) originally characterised in Japanese plum (*Prunus salicinia,* primer pairs *TPSCP1*, *TPSCP5* and *TPSCP10*; S. Ohta *et al*. 2005) and two from *Sorbus* (primer pairs *rps16pm2* and *trnT-Lpm1*; Chester *et al*. 2007) (Table 3.2). PCR consisted of 1 µL (10–50 ng) total genomic DNA, 1 x PCR buffer, 0.2 mM

Table 3.2: cpSSR primer sequences used for the phylogeographucal study.

Marker	**Location**	**Repeat motif**	**Primer sequence**	**T_a (°C)**
TPSCP1	rpl16 intron	$(T)_9$	F: TTGAAAACGAATCCTAATG R: ATTTTCTTTTTCCTTTGTATTATC	
TPSCP5	atpB-rbcL intergenic	$(T)_8$	F TTTCTATCTCATTGGTCCTT R: ATTCGCTCTTGACAGTGAT	
TPSCP10	rps16 intron	$(T)_9$	F: GGTTTCTTTTGAGTTATTTGAG R: CTTTTTTCTTATTCTTCCCCAAC	
rps16pm2	rps16 intron	$(ATCAA)_3$	F: CAACTTGAGTTATGAGGATAC R: TCGGGATCGAACATCAATTGCAAC	48
trnT-Lpm1	trnT-L spacer	$(A)_4(TA)_2$	F: CATTACAAATGCGATGCTCT R: CGCTATATTAATAGGTATGTT	50

dNTPs, 1.5 mM $MgCl_2$, 0.06–0.5 µM of fluorescence-labelled forward primer, the same amount of unlabelled reverse primer, and 0.5 units of Platinum® Taq DNA Polymerase (Invitrogen, USA). Volumes were made up to 10 µL with autoclaved deionised water. Reactions were run in 0.2-mL microtitre plate wells (Greiner) overlaid with a drop of mineral oil (Sigma).

Amplifications were run on a PTC-200 thermocycler (MJ Research) using the following heating profile: 3 min at 95°C (initial denaturing step), followed by 35 cycles of 30–60 sec at 94°C (denaturing step), 50 sec at 50°C (for *TPSCP5*, *rps16pm2*, *trnT-Lpm1*) or 60 sec at 55°C (for *TPSCP1*, *TPSCP10*) (annealing step), 1–2 min at 72°C (extension step) and an additional 10 min at 72°C (final extension step). PCR products were analysed using a CEQ 8000 Beckmann Coulter Sequencer (Beckman-Coulter, USA; Figure 3.3) with an internal standard to measure the size of amplified bands. A volume of 1–2 µL of sample of each diluted PCR product was added to 0.4 µL of size standard in 35 µL sample loading solution. Differences in the sizes of products and the use of different fluorescent labels allowed the analysis of two to three microsatellites in a single run.

Figure 3.3: Microsatellite genotyping using the CEQ 8000 Genetic analysis system 9.0 (Beckman Coultier, USA)

3.3 nSSR genotyping and fragment analysis of range wide study

Nuclear variation was analysed with six microsatellite loci (nSSR) originally characterised in wild cherry (*Prunus avium* L., primer pairs EMPaS01, EMPaS06 and EMPaS10; Vaughan and Russell, 2004) and peach (*Prunus persica*, primer pairs U3 (UDP97–403) and U5 (UDP96–018); Cipriani *et al*., 1999) and P2 (PS12A02); Sosinski *et al*., 2000) (Table 3.3). PCR consisted of 1 µL (10–50 ng) total genomic DNA, 1x PCR buffer, 0.2 mM dNTPS, 1.5–2.5 mM $MgCl_2$, 0.04–0.2 µM of fluorescence-labelled forward primer and the same of unlabelled reverse primer and 0.5–1.5 units of Platinum® Taq DNA Polymerase (Invitrogen, USA). Volumes were made up to 10 µL with autoclaved deionised water. Reactions were run in 0.2-mL microtitre plate wells (Greiner) overlaid with a drop of mineral oil (Sigma). Amplifications

Table 3.3: nSSR primer sequences employed in the population structure analysis.

Marker	**Repeat motif**	**Primer sequence**	**T_a °C**
EMPaS01	$(GA)_9(GA)_{11}$	F: 5’ CAAAATCAACAAAATCTAAACC 3’ R: 5’ CAAGAATCTTCTAGCTCAAACC 3’	
EMPaS06	$(CT)_{12}$	F: 5’ AAGCGGAAAGCACAGGTAG 3’ R: 5’ TTGCTAGCATAGAAAAGAATTGTAG 3’	
EMPaS010	$(GA)_{28}$	F: 5’ GCTAATATCAAATCCCAGCTCTC 3’ R’ 5’ TGAAGAAGTATGGCTTCTGTGG 3’	
U3 (UDP97–403)	$(AG)_{22}$	F: 5’CTGGCTTACAACTCGCAAGC 3’ R: 5’CGTCGACCAACTGAGACTCA 3’	60
U5 (UDP96–018)	$(AC)_{21}$	F: 5’TTCTAATCTGGGCTATGGCG 3 R: 5’GAAGTTCACATTTACGACAGGG 3’	60
P2 (PS12A02)		F: 5’GCC ACC AAT GGT TCT TCC 3’ R: 5’AGC ACC AGA TGC ACC TGA 3’	62

were run on a PTC-200 thermocycler (MJ Research) using the following heating profile: 3 min at 95ºC (initial denaturation step), followed by 35 cycles of 30–45 sec 94ºC (denature step), 45–90 sec at 48ºC–64ºC (primer variable annealing step), 45 sec–2 min at 72ºC (extension step) with an additional 8 min at 72ºC (final extension step). PCR products were analysed as described in section 3.2.

For locus verification five individuals possessing homozygous alleles were selected from every primer and amplified in a total PCR reaction of 50 µL. The PCR products were cleaned using the Qiaquick PCR clean up kit (Qiagen, Germany). The purified PCR products were sequenced on both the forward and the reverse strand by the automated fluorescent cycle sequencing method, using the Big Dye Terminator Ready Reaction Kit according to the manufacturer’s instructions. (Applied Biosystems, Foster City CA, USA)

3.4 Intraspecific chemical diversity

3.4.1 Plant Material

Equal-sized bark disks were collected at 1.3 m above ground in 4 orientations (N, E, S, W) of each tree (Figure 3.4). Samples were air-dried, stored in calico bags and shipped to the Austrian laboratory. In total, 20 populations (20 trees each) were sampled (Table 4.17, Figure 4.11). Genetic data were available for all individually sampled trees (section 3.5.3.1).

3.4.2 Bark processing

All four samples per tree were crushed through a coarse mill and further processed by fine-milling. Five tree samples per population were randomly selected and 50 grams of fine-milled bark was aliquoted out. A sixth sample was prepared to form an aliquot mix of bark powder (5 gm each) originating from all 20 trees from each population. This sample was called population mix.

3.4.3 Phytochemical analysis of the bark samples

In total, eight antioxidants probably having the most significant pharmaceutical properties, i.e., ferulic acid, friedelin, n-docosanol, lauric acid, myristic acid, β-sitostenone, β-sitosterol and ursolic acid were analysed. One g of the bark powder was extracted with 20 ml of chloroform/methanol (80/20 v/v) in a 100 ml Erlenmeyer flask using an orbital shaker (180 min^{-1} for 70 min). The extracts were filtered through S&S 595$^1/_2$ folded filters (110 mm diameter, Schleicher & Schüll, Dassel, Germany) into 24 ml glass vials and stored at 4°C.

A 1 ml aliquot of the raw extract was transferred into a 2 ml glass vial, put on a heating plate (45°C) and evaporated to dryness under a gentle stream of nitrogen. Residues were first dissolved in 50 µl dimethylformamide (DMF) and shaken for 10 sec. Subsequently, 50 µl of the derivatisation reagent N-Methyl-N-(t-butyldimethylsilyl) trifluoroacetamide (MTBSTFA) was added, samples were shaken again for 10 sec and were left in an oven at 60°C for 60 min.

Figure 3.4: Bark samples harvesting from *Prunus africana* in Kabale, Uganda

After derivatisation, 0.45 ml isooctane and 1 ml KH_2PO_4-Buffer (pH=7.00; 50 ml 0.1 M KH_2PO_4 mixed with 29 ml 0.1 M NaOH) were added and vials were vortexed for at least 30 sec until the upper organic layer was clear. Approximately 300 µl of the organic layer containing the analytes or their derivates were transferred into a HPLC-microvial which was closed tightly and analysed by a GC-MS-system (GC 6890N and MSD 5975b inert XI/CI, Agilent, Santa Clara, USA).

Injection volume was 1 µl using splitless injection mode. Chromatographic separation was carried out by means of an Agilent HP-5-MS column (30 m x 0.25 mm x 0.25 µm), using Helium as carrier gas (1 ml/min) and a temperature gradient (starting with 80 °C for 0.5 min, heating up to 340 °C with 25 °C/ min and holding for 17 min). Analytes were detected using SCAN- and SIM-mode simultaneously. For each compound one target ion was chosen as quantifier ion namely m/z 257, 285, 365, 383, 412, 426, 471 and 627 used for lauric acid, myristic acid, ferulic acid, n-docosanol, β-sitostenone, friedelin, β-sitosterol and ursolic acid, respectively. The retention times and the area of the extracted ion chromatograms (EIC) of the target ions (SIM-mode) were used for quantification of the analytes, data of the SCAN-mode were additionally used for verification by means of the spectra. For calibration of the GC-MS instrument analyte mixtures containing isooctane between 2.5 µg/ml and 75 µg/ml per analyte were measured. Quantification of the target compounds in the bark extracts was carried out by evaluating EIC-peak areas.

3.5 Data Management and statistical analysis

3.5.1 Phylogeography and Population Structure revealed by cpSSR

3.5.1.1 cpDNA variation and haplotype construction

To confirm the single nucleotide repeat motif of cpSSR variation and to avoid misinterpretations caused through size homoplasy, various alleles from each locus were sequenced. Sequence comparisons revealed single nucleotide steps for loci *TPSCP1*, *TPSCP5* and *TPSCP10*, whereas both of the loci *rps16pm2* and *trnT-Lpm1* were found to contain two indels of 5 – 26 base pairs in the flanking region. These indels were treated as two additional loci, namely *rps16pm2a* and *rps16pm2b*, and *trnT-Lpm1a* and *trnT-Lpm1b*. Additional length variation at the locus *rps16pm2* was not considered because it could not be resolved to unique mononucleotide repeat variation. Sequences were deposited in GenBank under the accession numbers HM749746 – HM749785.

Multi-locus haplotypes were constructed by combining single loci. The number of haplotypes per population, the effective number of haplotypes, the number of private haplotypes and haplotype diversity were calculated using GENALEX 6.41 (Peakall and Smouse, 2006). To account for differences in sample size, rarefaction analysis of haplotype diversity was undertaken using CONTRIB (Petit *et al*., 1998), giving an improved estimate of haplotype diversity and its standard error. Since mononucleotide microsatellites are potentially prone to homoplasy (i.e., alleles can be identical in state but not identical by descent due to repeated forward and backward mutations), linkage disequilibrium (LD) among loci (including indels) was examined following

Hale *et al*. (2004) using the statistic r_d implemented in the software MULTILOCUS 1.2 (Agapow and Burt, 2001). If the multilocus LD is significantly high, we can expect that alleles identical in size are also identical in descent; phylogeographic inferences should then not be affected (Hale *et al*. 2004, Angioi *et al*. 2009).

Haplotype relationships were inferred by constructing a statistical parsimony network following a two-step strategy according to Bänfer *et al*., (2006). This procedure accounted for the (presumably) different mutation rates underlying indel and microsatellite variation. First, haplotype data from the indel variation were employed to construct a backbone network using the programme TCS 1.18 (Clement *et al*., 2000). Second, the network was enlarged by adding the variation at microsatellite loci manually at the respective positions of the backbone; mononucleotide microsatellites were coded according to a stepwise mutation model.

3.5.1.2 Population structure and phylogeography as revealed by cpSSR

Total gene diversity (h_t) and average within population gene diversity (h_s) were calculated according to Pons and Petit (1995; 1996) using the software PERMUT (http://www.pierroton.inra.fr/genetics/labo/Software). To test for the existence of phylogeographic patterns, we compared the two measures of differentiation N_{ST} and G_{ST}, which are based on ordered and unordered alleles, respectively. Significance was tested on the basis of 1,000 permutations.

To analyse relatedness among populations without *a priori* hypotheses, a spatial analysis of molecular variance (SAMOVA) was applied according to Dupanloup *et al*. (2002). This method is based on a simulated annealing procedure where populations are assigned to a defined number (*K*) of groups. Groups are maximally differentiated from each other but include populations which are geographically proximate. *F*-statistics (F_{SC}, F_{ST} and F_{CT}) were generated based on pairwise differences among haplotypes following an AMOVA approach (Excoffier *et al*., 1992). The behaviour of the indices F_{CT} and F_{SC} for different values of *K* were explored to identify the optimum number of groups for a set of populations (Dupanloup *et al*., 2002). One hundred simulated annealing processes were used for each value of *K*, from *K* = 2 to 30.

3.5.1.3 Population history and coalescence as revealed by cpSSR

In order to relate the genealogical network to colonisation history and the physical characteristics of Africa and its changes within the Holocene, Bayesian coalescence modelling of past population splitting events were used. The group structure defined by SAMOVA was used to identify the time and sequence of divergence among regional groups. The software BATWING was used to generate random samples from posterior distributions of a variety of population genetic parameters, e.g. mutation rate, effective population size and growth rate, based on coalescence theory (Wilson *et al*., 2003). Due to lack of detailed knowledge on population demographics, a simple model of constant population size, where the prior population size follows a uniform distribution

within an interval of 10 to 10,000 individuals was used. Although diffuse, these size limits cover the most plausible values for *N*, if neighbourhood sizes, gene flow distances and densities typical for tropical trees (Hardy *et al*., 2006) and other *Prunus* species (Schueler *et al*., 2006) are considered. Prior mutation rates were set according to previous reports of chloroplast loci (Provan *et al*., 1999). Two groups of loci, both following a gamma distribution were specified: the indels *rps16pm2a, rps16pm2b*, *trnT-Lpm1a* and *trnT-Lpm1b* were set to priors of μ = 8 x 10^{-5}, while the mononucleotide microsatellites *TPSCP1*, *TPSCP5* and *TPSCP10* were set to priors of μ = 3.2 x 10^{-5} (Afzal-Rafii and Dodd, 2007) because they showed higher variation. With these priors, 55,000 samples in total were generated, from which the first 5,000 were discarded as the 'burn-in'. Posterior parameter distributions for mutation rate, population size, splitting times and splitting sequence were estimated from the output of the remaining 50,000 samples.

For coalescence modelling of *P. africana*, populations were aggregated into *K* = 6 regional groups according to the spatial-genetic structure observed in SAMOVA; this grouping provided an unbiased clustering of populations with a clear geographic distribution. To calculate estimates for the time of splitting events, coalescent units given by BATWING were converted into years by multiplying the population size given by the program with the generation time (GT) of *P. africana*, which was assumed to be 8 years at the lower limit and 15 years at the upper limit (Hall *et al*., 2000).

3.5.2 Population structure and history of *Prunus africana* as revealed by nuclear microsatellites (nSSR)

3.5.2.1 nSSR fragment analysis and diversity measures

Individual diploid genotypes were manually identified using CEQ 8000 Genetic analysis system 9.0 (Beckman Coultier, USA; Figure 3.3). Individuals with missing data were omitted in order not to confound the observed genetic structure. Data set was checked using the MICROSATELLITE TOOLKIT (Park, 2001) for identical genotypes and only one individual per genotype per population was retained. The resulting number of individuals employed for data analysis was 459 with individuals per population ranging from 3–31. The microsatellite dataset was checked for the presence of null alleles and other genotyping errors using the software program MICRO-CHECKER (Van Oosterhout *et al*., 2004).

For each locus the observed allele frequencies and the observed (H_o) and expected (H_e) levels of heterozygosity per population were calculated using GENALEX version 6.41 (Peakall and Smouse, 2006). Deviations from HWE were assessed using GENEPOP (Raymond and Rousset, 1995) as significant heterozygote excess or deficiency using Markov Chain Monte Carlo simulations with the following parameters; 10,000 dememorizations, 100 batches and 10,000 iterations. Genotypic linkage disequilibrium between loci was also assessed in FSTAT version 2.9.3.2 (Goudet, 1995) for single populations and across all populations using a Bonferroni correction. Allelic richness with rarefaction to 10 individuals was also calculated using FSTAT.

The presence of recent changes in population size as evidenced by heterozygosity excess based on estimates of multilocus genotypes was tested using the software BOTTLENECK version 1.2.02 (Piry *et al*., 1999), implementing a Wilcoxon signed rank test and assuming a two-phased model of mutation (95% stepwise and 5% multistep mutations, 15% variance among multiple steps. Five sampling locations with less than 10 individuals (Table 4.6) were excluded from population diversity analysis, but were retained for the Bayesian assignment and historic gene flow analysis described below.

3.5.2.2 Population differentiation

Global F_{ST} and R_{ST} were calculated using ARLEQUIN 3.5 (Excoffier and Lischer, 2010). An unweighted pair group method arithmetic average (UPGMA) dendrogram based Cavalli-Sforza and Edwards (1967) chord distances was also produced, after creating 1,000 bootstrapped matrices in MSA (Dieringer and Schlötterer, 2003) and using the computer programs NEIGHBOUR and CONSENSE in the PHYLIP version 3.63 package (Felsenstein, 1989) for tree construction. Additionally, principal coordinate analysis (PCoA) was used to explore multivariate relationships among inter-individual genetic distances (Nei's standard genetic distances; Nei, 1972) and to identify a set of reduced dimension traits (e.g., PC eigen vectors). These analyses were conducted using GENALEX 6.41. Consequently, analysis of molecular variance (AMOVA) was used to investigate population differentiation at various levels of subdivision defined by the groups detected by PCoA, UPGMA clustering and individual based population assignment (see below), using both F_{ST} and R_{ST} in separate

analyses (10,000 permutations, significance level 0.05). These computations were performed by ARLEQUIN 3.5.

3.5.2.3 Individual based population assignment

Individual based population assignment methods were used to further analyse population structure. In these computations all available samples were included. The model-based Bayesian clustering method implemented in STRUCTURE version 2.3 (Pritchard *et al*., 2000) was used to infer clusters by assigning individual multilocus genotypes probabilistically to a user defined number *K* clusters or genepools (this is a simple approach for accounting for genotypic ambiguity in studies of population structure *i.e*. null alleles and errors in genotypic calling; Falush *et al*., 2007). If their genotypes indicated that they are admixed they were assigned to two or more genepools. This allowed for the detection of genotypes that are outliers in their sample of origin and in fact belong to another genepool.

Using *K* values ranging from 2 to 30, a burn-in period of 200,000 and run lengths of 500,000 iterations were chosen based on preliminary runs to determine run parameters. Ten runs per K were performed on the total data set. The average –ln probability (ln prop) of the data from multiple replicates of each *K* was plotted to evaluate the rate of probability changes with increasing values of *K*. The most likely number of clusters was inferred using Evanno *et al*. (2005) ΔK statistics. The top 10 highest likelihood runs were then analysed in CLUMPP 1.0 (Jakobsson and Rosenberg, 2007) to create averages between

different runs common to cluster analyses. The results of CLUMPP were then visualized in DISTRUCT 1.0 (Rosenberg, 2003).

3.5.2.4 Genetic barriers among populations

In order to examine genetic barriers between populations at continental and regional scale, the software BARRIER 2.2 (Manni *et al.*, 2004) was used. In this approach computational geometry and a Monmonier's maximum-difference algorithm were combined to identify edges associated with high rates of change in a genetic distance matrix. Mapped populations were linked into a network using Delaunay triangulation (Delaunay, 1934) and a set of Voronoi polygons (Voronoi, 1908) was superimposed on the network. One hundred bootstrapped Cavalli-Sforza and Edwards' pairwise relative genetic distances matrices created in MSA were used for the computations. Each edge of the Voronoi tessellation was associated with the estimated value of genetic distance between the two populations. Finally, (Monmonier, 1973) maximum difference algorithm was used to determine where breaks in gene flow occur. This was done by placing a barrier perpendicular to the edge of the triangulation that corresponds to the largest genetic distance. The barrier was then extended across the adjacent edge of the network until the boundary reached either the limits of the triangulation or a previously traced barrier.

3.5.2.5 Estimation of recent migration rates among regions.

To quantify average recent migration rates (last few generations) among populations, a Bayesian approach implemented in the software BAYESASS 1.3

(Wilson and Rannala, 2003) was applied. This method makes no assumptions of HWE within the sample, and is based on transient multilocus disequilibrium in multilocus genotypes of migrants relative to the host population. It calculates inbreeding coefficients for each population separately and the joint probabilities are used to estimate recent migration rates (Wilson and Rannala, 2003). Since the accuracy of estimating population assignment or migration rates with this approach depends on large population sample sizes and/or large number of loci, populations with less than 10 individuals were omitted. A run length of 5,000,000 generations, with 999,999 generations burn-in, sampling frequency of 2000, with δp (maximum change of allele frequency) at 0.30, a δm of 0.15 (maximum change of migration rate), and δF of 0.15 (maximum change in inbreeding coefficient).

The proportion of migrants (m) between geographic regions was calculated. The method allows for arbitrary genotype frequencies within and calculates separate inbreeding coefficients for each population, the joint probabilities of which are used to estimate recent migration rates (Wilson and Rannala, 2003). The first run used default settings with subsequent runs incorporating random seed and delta values, the latter chosen to ensure that proposed changes between chains at the end of the run were between 40 % and 60 %. This step is recommended to ensure that neither chain mixing gets stuck on a local maximum nor the chain slips out of the global maximum too easily (Austin *et al*., 2004). Each data set was run five times to ensure consistency of results and an average of the values for these runs were calculated.

3.5.2.6 Likelihood estimates of the direction and magnitude of historic gene flow

To estimate the direction and magnitude of gene flow between geographical groups determined in STRUCTURE, the maximum likelihood (ML) approach of Beerli and Felsenstein, (2001) implemented in MIGRATE version 3.2.7 (Beerli, 2010) was applied. The ML-model assumes that populations are at drift-migration equilibrium, had constant size and migration between population pairs occurred over the coalescent period (~4*Ne* generations). However, unlike *F*-statistics, this method allows for the possibility of non-symmetrical migration and differences in population sizes among populations, both biologically realistic scenarios.

MIGRATE jointly estimates the mutation-scaled effective population sizes ($\theta=4N_e\mu$; μ the mutation rates for diploid data) and the mutation-scaled effective migration rates ($M=m/\mu$; where m is the immigration rate) between geographical groups by estimating allele genealogies and then approximating the sum of probabilities across possible genealogies using Metropolis-Hastings Markov chain Monte Carlo (MCMC) sampling. MCMC concentrates the sampling in areas of the coalescent space that contribute most to the final likelihood and ignores genealogies that contribute little to the final likelihood (Beerli and Felsenstein, 2001). The Brownian motion model for microsatellites as an approximation of the ladder model (Ohta and Kimura, 1973) was used as mutation model and equal mutation rates among loci were assumed. As a search strategy 10 short chains with 5,000 recorded trees, followed by 3 long

chains with 50,000 recorded trees, a burn-in of 20,000 and a sampling increment of 100 was used. MIGRATE was run a minimum of four times per data set to verify that final chains were estimating the same ML values for Θ (i.e., population size) and 4*Nm* (as determined by overlapping 95% confidence intervals). For each run the random number seed, the starting values of Θ and 4*Nm* were changed. The first run estimated Θ and 4*Nm* from F_{ST} values, and subsequent runs incorporated the ML estimates of Θ and 4*Nm* from the previous run as the starting parameters. To avoid problems due to different sample size among groups, a reduced random subset (37 individuals) corresponding to the lowest group size of each population was used in a parallel computation.

3.5.3 Intraspecific Phytochemical diversity

3.5.3.1 Genetic data

Genetic data were based on seven chloroplast (*cp*) microsatellites and six nuclear (*n*) microsatellites as described in detail in section 3.2 & 3.3. Genetic parameters, effective number of alleles in the chloroplast (N_{e_c}) and nuclear genome (N_{e_n}), unbiased expected haplotypic diversity (h_c), unbiased expected heterozygosity (H_e) as well as fixation index (F_{IS}) were calculated using GENALEX v. 6.41 (Peakall and Smouse, 2006). For correlations of chemical distances to genetic distances, pairwise differentiation matrices were calculated both for chloroplast and nuclear microsatellites. For the chloroplast data, Φ_{PT}, the population differentiation for haploid loci was used, whereas R_{ST}, the population differentiation for codominant markers assuming stepwise

mutations was applied for the nuclear microsatellites. Both, Φ_{PT} and R_{ST} estimate the proportion of pairwise differentiation among populations, relative to the total differentiation using the analysis of molecular variance approach developed by Excoffier *et al*., (1992) and extended by Peakall *et al*., (1995) and Michalakis and Excoffier, (1996) to codominant data.

3.5.3.2 Statistical analysis

The measurements for ferulic acid, friedelin, n-docosanol, lauric acid, myristic acid, β-sitostenone, β-sitosterol and ursolic acid were scored for 20 populations along with those for the population mixes and the control samples. The method detection limits (LOD) and method quantification limits (LOQ), respectively, were calculated based on 20 total procedure blanks with LOD three times the standard deviation and LOQ ten times the standard deviation of mean blank concentrations (Table 4.18). The mean of the quantified chemical compounds of the five single trees per populations were calculated to establish an overview of relationships and then coefficients of variation within populations were calculated to have a normalized comparison of variation. Contents of different constituents were intercorrelated and correlated to environmental conditions and tree size. If needed, significance was Bonferoni corrected. ANOVA and a subsequent Duncan-test were used to test for differences among populations. The proportion of variance within and among populations were estimated by variance component analysis using restricted maximum likelihood technique.

In order to evaluate whether populations that show high variation also show higher genetic diversity, the various measures of genetic diversity for chloroplast and nuclear markers were correlated to the chemical coefficient of variation within population.

To test for a geographical structure in the bark chemical data and to infer the similarity of populations, matrices of pairwise chemical distances were calculated for each single compound (Euclidean distance) and for the complete spectra of substances using Ward's Jr., (1963) technique for joining substances. In addition, discriminant analysis was carried out to reveal a matrix of Mahabaloni-distances between populations as a basis for subsequent cluster analysis. The various chemical distance matrices (for single compounds and all compounds) were used to test for associations of geographical to genetic distances (Φ_{PT} and $R_{ST,}$ respectively) by Mantel tests. Significance was tested at 999 permutations using the Mantel procedure of GENALEX, v.6.41 (Peakall and Smouse, 2006). Other statistical calculations were done by means of STATISTICA (http://www.statsoft.de/)

CHAPTER FOUR

RESULTS

4.1 Phylogeography and population structure revealed by cpSSR

4.1.1 cpDNA variation and haplotype construction

Three cpDNA microsatellites and four cpDNA indels generated data for 582 individuals of *P. africana* from 32 populations. Data from all primers yielded 2 to 4 alleles per locus and 19 alleles in total. The most variable cpSSR loci were *TPSCP1* and *TPSCP10* with 4 alleles each, followed by *TPSCP5* with 3 alleles.

Combination of loci resulted in 22 haplotypes in total (Table 4.1). The predominant haplotype, HT1a (frequency 31%), occurred in 13 populations of East and southern Africa. HT1m, the second most frequent haplotype (frequency 14%), occurred only in populations of western Uganda and West Africa (Figure 4.1). The genetic characteristics of chloroplast haplotypes in populations are shown in Table 4.2. The number of haplotypes (N_a) per population ranged from one to four with an average of two, with populations from the East Africa region harbouring the highest number of haplotypes. Nine

Table 4.1: Haplotype construction and frequency in 32 *Prunus africana* populations revealed by 4 cpDNA indels (Loci 1 – 4) and 3 cpDNA microsatellites (Loci 5 – 7). Alleles are represented as 1 for presence of fragment and 2 for absence of fragment in the case of the indels (Loci 1 – 4) and repeat numbers in the case of microsatellites (Loci 5 – 7).

	Backbone Nt	1															2		3			4	5
	Nt subdivision	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	a	b	a	b	c	a	a
Indels†	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	2	2
	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1
	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	1	1	1	1	1
	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2
MS$	5	9	10	10	10	11	8	9	9	9	9	9	9	8	8	8	8	10	9	9	9	9	10
	6	9	9	9	9	9	9	10	10	9	8	8	8	8	8	8	8	8	9	10	9	9	9
	7	9	9	11	12	12	11	9	10	10	9	10	11	9	10	11	9	9	9	9	10	11	11
Population‡	CA														1	18							
	CB														19								
	CC													20									
	GQ											7	11										
	KA							20															
	KB	1			4		3		11														
	KC		3	12	5																		
	KD	20																					
	KE	14			6																		
	KF				7	12																	
	KG									20													
	KS	15									1												
	KT				12																		
	MA																		4		1		
	MB																		8	25			
	MC																				18		
	NG													9									
	SA	13						6															
	SB	13						4															
	TA						10																9
	TB	4		1					7													5	
	TC																					14	
	TD	16						4															
	TE	11						4															
	TT	12	1							3													
	UA													20									
	UB													20									
	UC													18	1								
	UD																13	7					
	ZWA	20																					
	ZWB	20																					
	ZWC	19																					
	Total	178	4	13	34	12	13	38	18	23	1	7	11	87	21	18	13	7	12	25	19	19	9

†Indel loci 1,2,3,4 are rps16pm2a, rps16pm2b, trnT-Lpm1a, and trnT-Lpm1b
$MS-Chloroplast microsatellite loci 5, 6, 7 are TPSCP1, TPSCP5, and TPSCP10
‡Populations are Ngashie Mt.Oku (CA), Lower Mann's Spring, Mt. Cam (CB), Mt. Danoua (CC), Equatorial Guinea (GQ), Chuka (KA), Kinale (KB), Kapcherop (KC), Kakamega (KD), Londiani (KE), Ol Danyo Sambuk (KF), Taita (KG), Lari (KT), Kibiri forest (KS), Marovoay (MA), Lakato forest (MB), Antsahabiraoka (MC), Nigeria (NG), Mpumalanga (SA), Kwazulu Natal (SB), Meru (TA), Kilimanjaro (TB), Kindoroko (TC), Shume Magamba (TD), Kidabaga (TE), Udzungwa (TT), Kibale (UA), Kalinzu (UB), Bwindi (UC), Mabira (UD), Nyanga (ZWA), Chimanimani (ZWB) and Chipinge (ZWC)

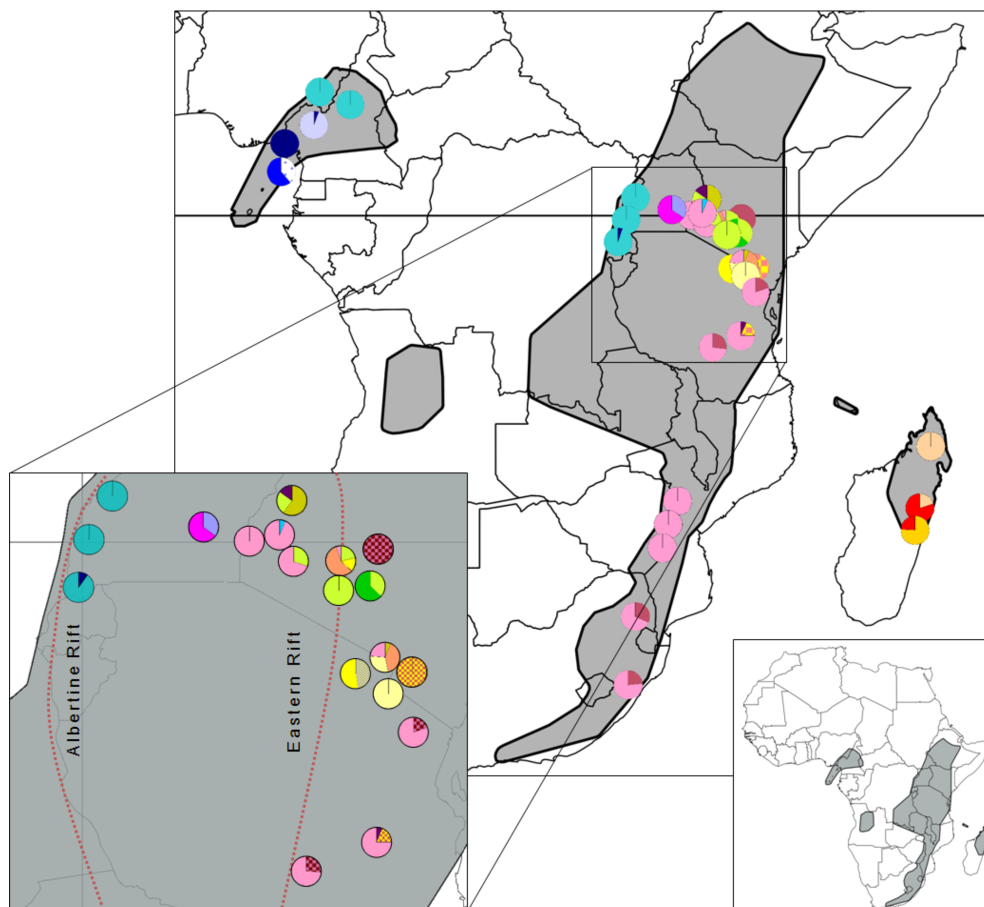

Figure 4.1: Approximate geographic distribution of *Prunus africana* (grey shaded area), the origins of the populations used in the current study and haplotype distributions. Colour codes demonstrate the proportions of haplotypes within each population. Country boundaries are also shown.

haplotypes were private to single populations, while 13 occurred in two or more stands. The island of Bioko (GQ) and one population from Uganda (UD) each had two private haplotypes (GQ: HT1k and HT1l; UD: 2a and 2b) and did not share haplotypes with other populations. Fourteen populations revealed only one haplotype. The highest haplotype diversity (H_e) of 0.728 was found in the Tanzanian population from Mt. Kilimanjaro, followed by the Kenyan population Kinale (H_e = 0.626).

Table 4.2: Description, location and haplotype diversity measures of investigated populations.

No.	Name of Population	Code	Country	Longitude	Latitude	n	N_a	N_e	N_p	H_e
1	Ngashie-Mt Oku	CA	Cameroon	10.5092	6.2048	19	2	1.1	1	0.105 (0.092-0.158)
2	Lower Mann's Spring, Mt Cameroon	CB	Cameroon	9.1192	4.1346	19	1	1	0	0.000 (0.000-0.000)
3	Mt Danoua	CC	Cameroon	11.9413	7.1871	20	1	1	0	0.000 (0.000-0.000)
4	Moka	GQ	Equatorial Guinea	8.6559	3.3625	18	2	1.9	2	0.503 (0.064-0.755)
5	Chuka, Central province	KA	Kenya	37.3645	-0.1798	20	1	1	0	0.000 (0.000-0.000)
6	Kinale, Central province	KB	Kenya	36.4151	-0.5278	19	4	2.5	0	0.626 (0.098-1.053)
7	Kapcherop, Cherangani Forest, Rift Valley	KC	Kenya	35.2161	1.0315	20	3	2.3	0	0.584 (0.091-0.955)
8	Kakamega Forest, Western Province	KD	Kenya	34.519	0.1415	20	1	1	0	0.000 (0.000-0.000)
9	Londiani, Rift Valley	KE	Kenya	35.1836	-0.2737	20	2	1.7	0	0.442 (0.087-0.663)
10	Ol Danyo Sambuk, Central Province	KF	Kenya	37.1501	-1.0744	19	2	1.9	1	0.491 (0.068-0.737)
11	Taita Hills, Coast Province	KG	Kenya	38.2088	-3.2462	20	1	1	0	0.000 (0.000-0.000)
12	Lari, Central Province	KT	Kenya	36.687	-1.0311	12	1	1	0	0.000 (0.000-0.000)
13	Kibiri forest, Western Province	KS	Kenya	34.8719	0.1502	16	2	1.1	1	0.125 (0.106-0.188)
14	Marovoay	MA	Madagascar	48.3069	-18.802	5	2	1.5	0	0.400 (0.237-0.600)
15	Lakato forest	MB	Madagascar	48.2779	-19.197	33	2	1.6	1	0.379 (0.079-0.568)
16	Antsahabiraoka	MC	Madagascar	49.2164	-14.402	18	1	1	0	0.000 (0.000-0.000)
17	Ngel Nyaki Forest Reserve, Nigeria	NG	Nigeria	11.0271	7.066	9	1	1	0	0.000 (0.000-0.000)
18	Mpumalanga	SA	South Africa	30.7306	-25.023	19	2	1.8	0	0.456 (0.085-0.684)
19	KwaZulu-Natal	SB	South Africa	30.2725	-29.287	17	2	1.6	0	0.382 (0.113-0.574)
20	Meru Catchment Forest	TA	Tanzania	36.8074	-3.2925	19	2	2	1	0.526 (0.040-0.789)
21	Kilimanjaro Catchment Forest Reserve	TB	Tanzania	37.5237	-3.0036	17	4	3.2	0	0.728 (0.060-1.256)
22	Kindoroko Catchment Reserve	TC	Tanzania	37.6267	-3.7389	14	1	1	0	0.000 (0.000-0.000)
23	Shume/Magamba Catchment Forest Reserve	TD	Tanzania	38.2521	-4.7544	20	2	1.5	0	0.337 (0.110-0.505)
24	Kidabaga	TE	Tanzania	35.9296	-8.1101	15	2	1.6	0	0.419 (0.113-0.629)
25	Udzungwa	TT	Tanzania	36.7791	-7.7668	16	3	1.7	0	0.425 (0.133-0.670)
26	Kibale Forest Natural Park	UA	Uganda	30.357	0.5644	20	1	1	0	0.000 (0.000-0.000)
27	Kalinzu Forest Reserve	UB	Uganda	30.1101	-0.3742	20	1	1	0	0.000 (0.000-0.000)
28	Bwindi Forest	UC	Uganda	29.7754	-1.0476	19	2	1.1	0	0.105 (0.092-0.158)
29	Mabira Forest	UD	Uganda	33.015	0.3806	20	2	1.8	2	0.479 (0.072-0.718)
30	Nyanga National Park	ZWA	Zimbabwe	32.7401	-18.287	20	1	1	0	0.000 (0.000-0.000)
31	Cashel Valley Chimanimani	ZWB	Zimbabwe	32.8004	-19.578	20	1	1	0	0.000 (0.000-0.000)
32	Chirinda forest Reserve Chipinge	ZWC	Zimbabwe	32.695	-20.409	19	1	1	0	0.000 (0.000-0.000)

The relationship between haplotypes is demonstrated by the genealogical network given in Figure 4.2. The backbone of the network consists of five main haplotypes (denoted HT1 – HT5), of which HT1 is the dominant type on mainland Africa. Haplotypes of the HT3 family (3a – 3c) form a geographically distinct group in Madagascar. The most common types HT1a and HT1m are separated only by two mutational steps but are geographically distributed either in the 'east' or 'west' African populations, respectively. HT1a and the haplotypes HT1b – HT1i can be found only in east (excluding Uganda) from here henceforth referred to as 'east' African populations and southern African populations, whereas populations from the eastern side of the Albertine Rift Valley in Uganda and the western part of Africa contain haplotypes related to HT1m (HT1j – HT1n) from here henceforth referred to as 'west' African populations.

4.1.2 Population structure and phylogeography

The total gene diversity (h_t) of 0.886 (SE = 0.0376) was more than three times higher than the mean within population diversity (h_s) of 0.234 (SE = 0.0432), suggesting high genetic differentiation among populations. This was also demonstrated by the high levels of differentiation of ordered (N_{ST}) and unordered (G_{ST}) alleles (0.840, SE = 0.0395 and 0.735, SE = 0.0481, respectively). N_{ST} was significantly higher than G_{ST} (P < 0.01), indicating a phylogeographic pattern for *P. africana*.

SAMOVA revealed a continuous increase of F_{CT} values from $K = 2$ until $K = 20$, again demonstrating the high differentiation among populations. F_{CT} values remained constant for $K > 20$ groups. Table 4.3 shows groupings for different values of *K*. Continuous bold lines depict barriers that remained constant from low to high numbers, while dotted bold lines depict barriers which occasionally disappear. Finer lines show boundaries which appear only at higher numbers of *K* (> 14). The clearest spatial-genetic group is formed by the three populations from Madagascar, because their grouping is stable from $K = 6$ upwards. Interestingly, three populations from western Uganda group with the West Africa stands rather than other geographically proximate populations from East Africa. The East African region shows the highest heterogeneity, where for $K = 2$ a barrier was already found. The mainland southern region is closely related to the East Africa region, but within the mainland southern region populations are much less differentiated. The distribution of molecular variance to populations, populations within groups and groups is shown in Table 4.4; with increasing *K*, variation among populations within groups decreased and variation among groups increased.

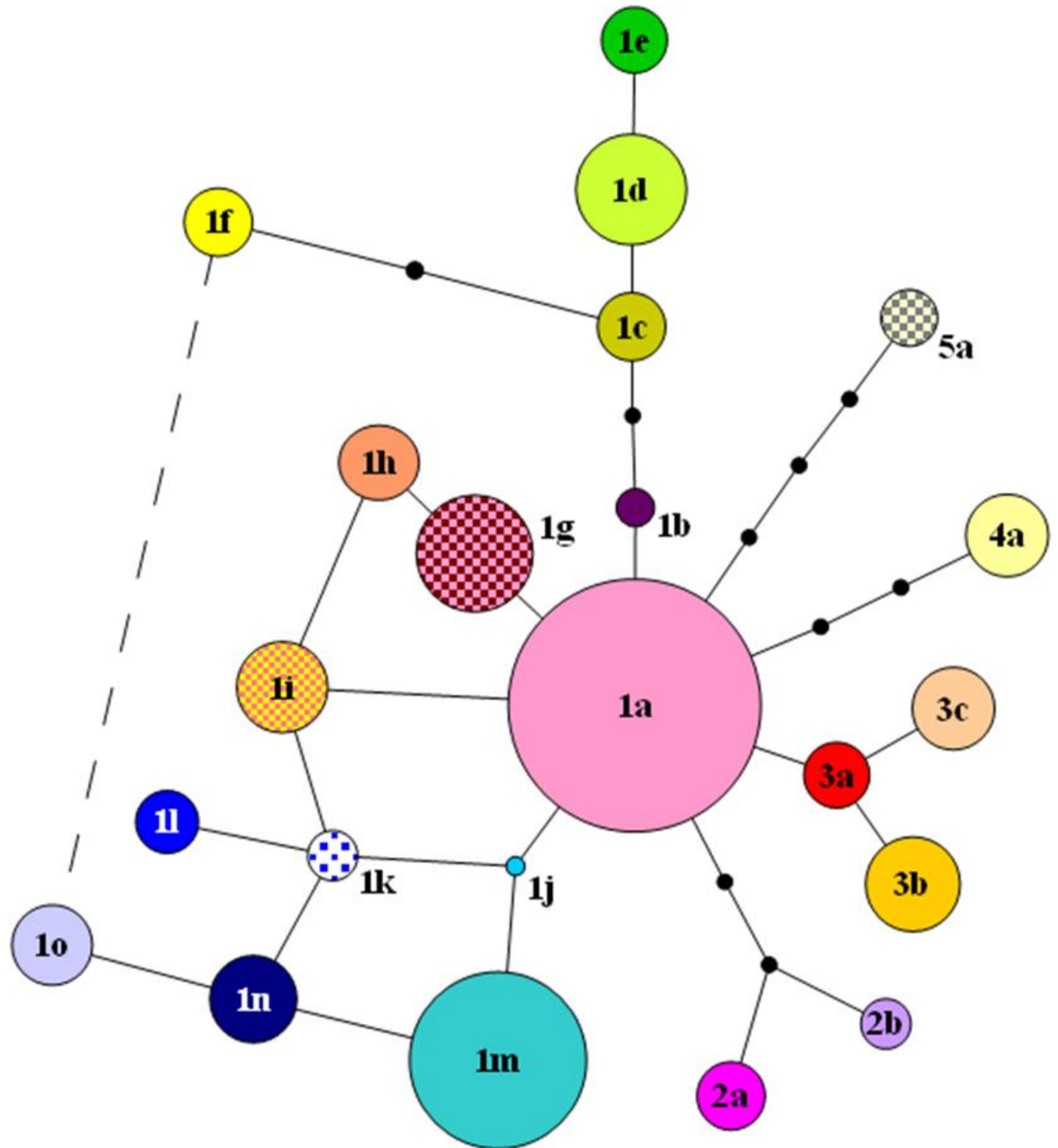

Figure 4.2: Genealogical network revealing 22 haplotypes constructed from 19 alleles obtained from 4 cpDNA indels (Loci 1–4) and 3 cpDNA microsatellites (Loci 5–7) within 32 *Prunus africana* populations. Circle size is representative of the number of individuals having each haplotype (see Table 4.1).

4.1.3 Population history and coalescence

The two most recent splitting events with a node support of more than 80% were found to have occurred among populations of eastern and southern Africa (Figure 4.3). The estimated confidence limit for the time of these population splitting events assuming a GT of 8 years ranged from 21,000 years before present (BP) to 1,500 years BP for the second youngest split and from 6,500 years BP until today for the youngest split respectively (see Table 4.5 and Figure 4.3). The splitting among ‘west’ African groups (including populations from Uganda) also received high node support (48%) and dates either to the second, third or fourth oldest splitting event to times between 1,400 years to 67,000 years BP (GT = 15: 2,700 to 125,000 years). The split between Madagascar and all East and South African groups dates to the second and third oldest splitting events (from 7,800 to 67,000 years BP using GT = 8 and from 14,600 to 125,000 years using GT = 15) and receives a node support of about 40%. Other potential splitting events at the second or third oldest splitting event receive a node support of less than 20%, suggesting branching among groups of ‘east’ and ‘west’ Africa occurred concurrently.

Table 4.3: Genetic structure among *Prunus africana* populations analysed by SAMOVA. Increasing values of F_{CT} from K = 2 to 20 groups are shown. The groupings for different values of K are indicated by line demarcations. Indicated groupings relate to geographic origin.

		West Africa Region and Uganda									Madagascar			Southern Region					East Africa Region excl. Uganda														
K	F_{CT}																																
2	0.529	CA	CB	GQ	CC	UA	UB	UC	NG	UD	MB	MA	MC	SA	SB	ZWA	ZWB	ZWC	TD	TE	TT	KA	KD	KS	KE	KG	TB	KB	TC	TA	KC	KF	KT
4	0.595	CA	CB	GQ	CC	UA	UB	UC	NG	UD	MB	MA	MC	SA	SB	ZWA	ZWB	ZWC	TD	TE	TT	KA	KD	KS	KE	KG	TB	KB	TC	TA	KC	KF	KT
6	0.667	CA	CB	GQ	CC	UA	UB	UC	NG	UD	MB	MA	MC	SA	SB	ZWA	ZWB	ZWC	TD	TE	TT	KA	KD	KS	KE	KG	TB	KB	TC	TA	KC	KF	KT
8	0.700	CA	CB	GQ	CC	UA	UB	UC	NG	UD	MB	MA	MC	SA	SB	ZWA	ZWB	ZWC	TD	TE	TT	KA	KD	KS	KE	KG	TB	KB	TC	TA	KC	KF	KT
10	0.743	CA	CB	GQ	CC	UA	UB	UC	NG	UD	MB	MA	MC	SA	SB	ZWA	ZWB	ZWC	TD	TE	TT	KA	KD	KS	KE	KG	TB	KB	TC	TA	KC	KF	KT
12	0.745	CA	CB	GQ	CC	UA	UB	UC	NG	UD	MB	MA	MC	SA	SB	ZWA	ZWB	ZWC	TD	TE	TT	KA	KD	KS	KE	KG	TB	KB	TC	TA	KC	KF	KT
14	0.776	CA	CB	GQ	CC	UA	UB	UC	NG	UD	MB	MA	MC	SA	SB	ZWA	ZWB	ZWC	TD	TE	TT	KA	KD	KS	KE	KG	TB	KB	TA	TC	KC	KF	KT
16	0.793	CA	CB	GQ	CC	UA	UB	UC	NG	UD	MB	MA	MC	SA	SB	ZWA	ZWB	ZWC	TD	TE	TT	KA	KD	KS	KE	KG	TB	KB	TA	TC	KC	KF	KT
18	0.795	CA	CB	GQ	CC	UA	UB	UC	NG	UD	MB	MA	MC	SA	SB	ZWA	ZWB	ZWC	TD	TE	TT	KA	KD	KS	KE	KG	TB	KB	TA	TC	KC	KF	KT
20	0.825	CA	CB	GQ	CC	UA	UB	UC	NG	UD	MB	MA	MC	SA	SB	ZWA	ZWB	ZWC	TD	TE	TT	KA	KD	KS	KE	KG	TB	KB	TA	TC	KC	KF	KT

Boundaries stable

Boundaries change occasionally

Boundaries change at $K > 14$

Table 4.4: Analysis of variance of population structure of 32 *Prunus africana* populations based on SAMOVA structure analysis using K = 2 to 20 groups.

K	vg	vpg	vp	F_{SC}	F_{ST}	F_{CT}
2	52.89	38.6	8.47	0.82	0.92	0.53
4	59.49	28.2	12.3	0.7	0.88	0.6
6	66.71	20.1	13.2	0.6	0.87	0.67
8	70.02	16.7	13.3	0.56	0.87	0.7
10	72.71	13	14.3	0.48	0.86	0.73
12	74.17	10.9	14.9	0.42	0.85	0.74
14	78.05	7.01	15	0.32	0.85	0.78
16	79.27	5.45	15.3	0.26	0.85	0.79
18	79.51	4.91	15.6	0.24	0.84	0.8
20	82.5	1.91	15.6	0.11	0.84	0.83

K, number of groups; vg, variation within groups; vpg, variation of populations within groups; vp, variation within populations; F_{SC}, differentiation among populations within groups; F_{ST}, differentiation among populations; F_{CT}, differentiation among groups of populations

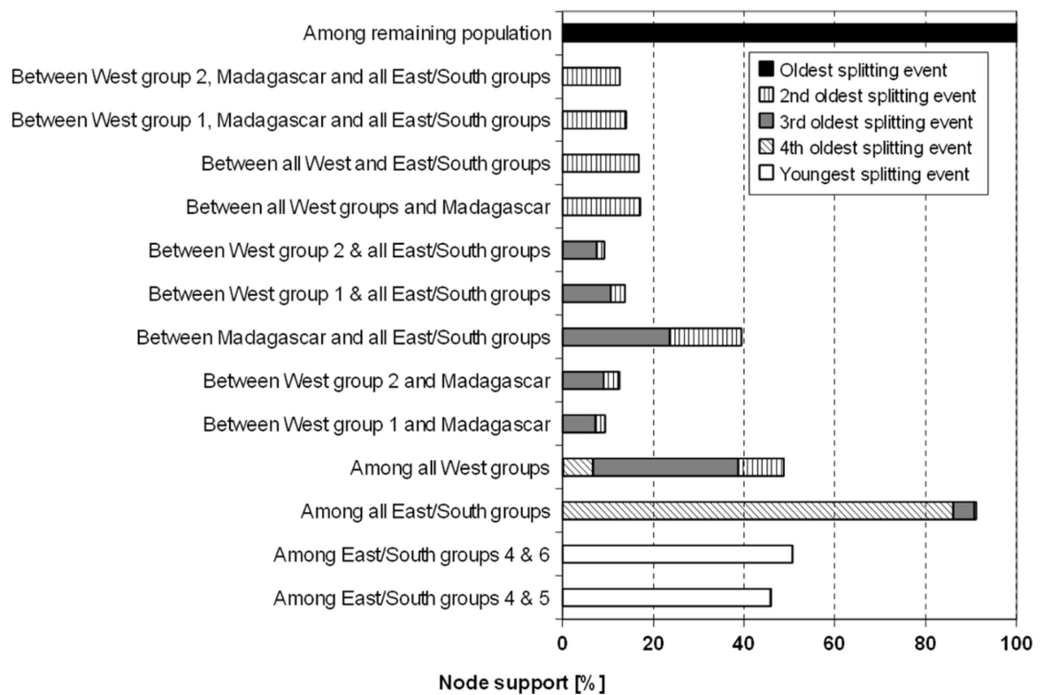

Figure 4.3: Node support for various population splitting events identified by coalescence modelling using a Bayesian analysis. Only splitting events with a node support of more than 5% are shown.

Table 4.5: Results of coalescence modelling using a Bayesian analysis. Upper and lower 5% quantiles, mean and standard deviation of a posteriori estimates of ancestral population size (N) and mutation rates (μ) are given.

	Quantile 5%	Mean	Quantile 95%	Std.dev.
Ancestral population size				
N	3807	6419	9290	1663
Posterior mutation rates				
μ (loci 1-4)	1.94×10^{-5}	3.57×10^{-5}	5.74×10^{-5}	1.18×10^{-5}
μ (loci 5-7)	8.10×10^{-5}	12.06×10^{-5}	16.93×10^{-5}	2.72×10^{-5}

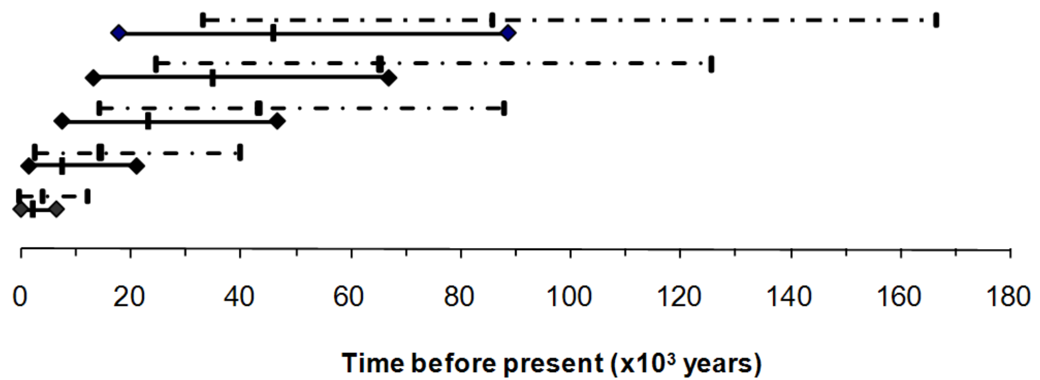

Figure 4.4: Time of population splitting events identified by coalescence modelling using a Bayesian analysis for two estimates of generation time (8 years – solid line; 15 years – broken line). The graph gives the upper and lower 5% quantiles and the mean of the time estimates. In descending order, the upper line identifies the oldest event and the lowest one the youngest event.

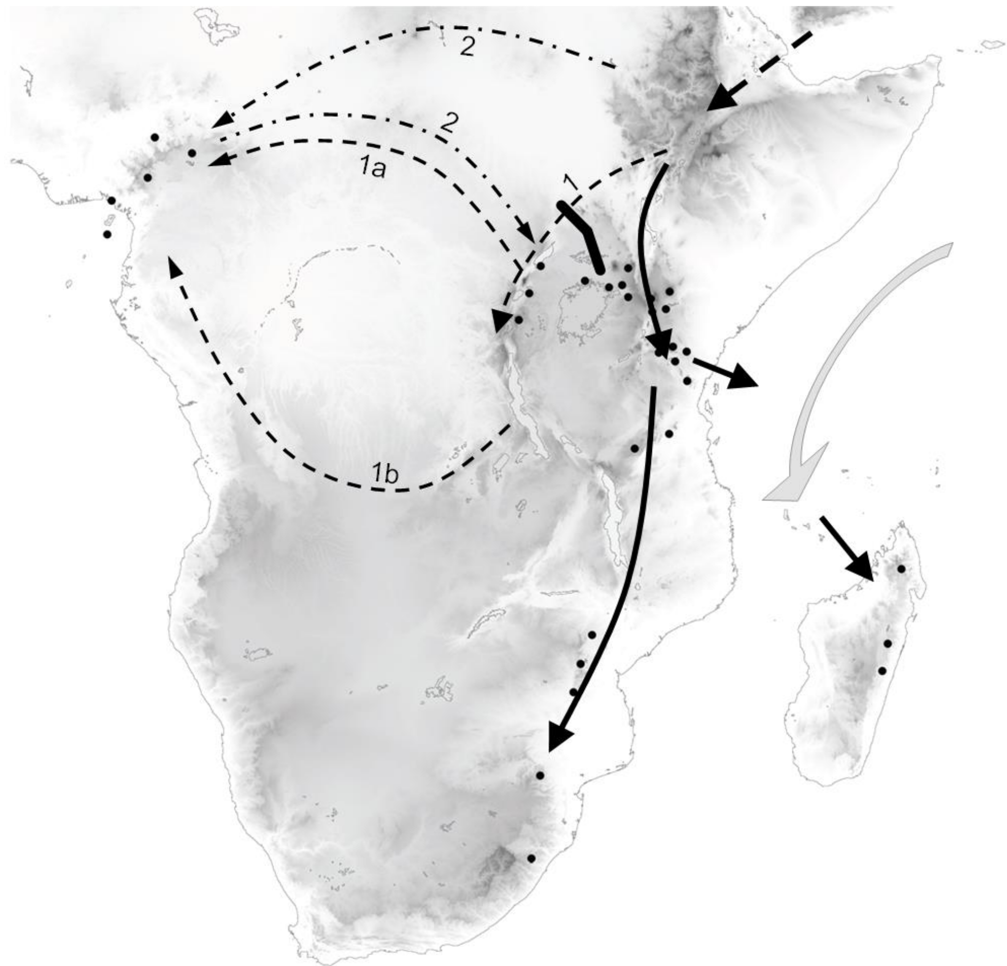

Figure 4.5: Hypothesised migration route of *Prunus africana* as deduced from coalescence modelling and the observed phylogeographic – genetic structure. Black solid arrows indicate the hypothetical route according to the present study, the black broken arrow indicates Aubréville's path of immigration of the genus *Prunus* into Africa (Aubréville, 1976). The putative migration paths to West Africa are shown by thin broken arrows: 1 – split of eastern and western populations during southward migration at the southern fringe of the Ethiopian highlands, subsequently: 1a – migration from the Albertine Rift to West Africa via a northern connection; 1b – southern migratory tract (SMT) following White (1993); 2 – independent colonisation of West Africa via a northern migration corridor and subsequent colonisation of the Albertine Rift from West Africa. The grey arrow indicates ocean currents involved in the possible dispersal to Madagascar. Black dots represent locations where population sampling was carried out in this study. The bold bar indicates the putative barrier between eastern and western lineage created by the upper river Nile and Lake Victoria basins.

4.2 Population structure and history as revealed by nuclear microsatellites (nSSR)

4.2.1 Nuclear microsatellite locus characterization and genetic diversity

Nuclear microsatellite markers exhibited a high level of polymorphism in *P. africana* with 10–45 alleles per locus summing up to 142 alleles in total. Nevertheless, a high level of divergence among populations was evident as some populations were fixed for single alleles: populations SA (SA1), SB (SA2) and MA (MG1) did not show variation at locus U3, likewise SA (SA1) and CA (CM2) were not polymorphic at locus U5, and population ZWC (ZW3) was fixed at locus EMPaS01. Number of alleles, number of effective alleles, observed and expected heterozygosity and fixation Index are shown in Table 4.8.

Observed heterozygosity ranged from 0.352 in SA (SA1) to 0.808 in UA (UG3) with an average of 0.622. Lower values were mostly found in the southern population of Zimbabwe and South Africa, and in Madagascar. The lower levels of genetic diversity are also reflected in measures of average allelic richness, which ranged from 3.398 in SB (SA2) to 8.436 in KS (KE1). The highest average values of allelic richness were found in East African populations west of the Eastern Rift valley (Table 4.7 and Table 4.9). In many populations, private alleles were detected, highlighting the strong regional divergence among the populations of *P. africana* in Africa (Table 4.8). On the regional level, populations from Madagascar and populations from Eastern Africa (east of the Eastern Rift Valley), harboured the highest number of

private alleles (nine). Also, East African populations west of the Eastern Rift had many private alleles (eight), whereas other regions harboured less (southern African populations: five; western African populations: three). The frequencies of these private alleles were generally below 0.05, however, in Madagascar three alleles at locus U5 had frequencies of 0.14, 0.11 and 0.23, respectively; similarly, one allele in locus P2 occurred in a frequency of 0.31 in southern African populations.

Although several populations showed heterozygote deficiency or excess, only few highly significant deviations from HWE ($p < 0.001$) were detected (populations UD, KC and TT; Table 4.9). Deviations were probably caused by the occurrence of null alleles, which affected all loci except locus P2 to a varying degree (Table 4.9); however, these probably did not affect the further analysis. No evidence for linkage disequilibrium between loci within populations was found after sequential Bonferroni correction. Assuming a two-phased mutation model, the Wilcoxon signed rank test implemented in BOTTLENECK ($p < 0.01$) detected recent decrease in population size in southern Africa (SA (SA1), SB (SA2), ZWC (ZW3)), Kenya (KA (KE5)), Uganda (UD (UG4)) and Madagascar (MA (MG1)) (Table 4.7)

Table 4.6: List of the 30 natural populations of *Prunus africana* analysed in this study

Code		Population name	Country	Population	Longitude	Latitude	Elevation	*N*
GQ	GQ1	Moka	Equ. Guinea	WG	8.6559	3.3625	1700	13
CB	CM1	Lower Mann's Spring, Mt Cameroon	Cameroon	WG	9.1192	4.1346	1600	20
CA	CM2	Ngashie-Mt Oku	Cameroon	WG	10.5092	6.2048	2450	16
CC	CM3	Mt Danoua	Cameroon	WG	11.9413	7.1871	1400	19
NG	NG1	Ngel Nyaki Forest Reserve	Nigeria	WG	11.0271	7.066	1100	7
UC	UG1	Bwindi Forest	Uganda	EG1	29.7754	-1.0476	1900	20
UB	UG2	Kalinzu Forest Reserve	Uganda	EG1	30.1101	-0.3742	1500	20
UA	UG3	Kibale Forest Natural Park	Uganda	EG1	30.3570	0.5644	1350	20
UD	UG4	Mabira Forest	Uganda	EG1	33.0150	0.3806	1250	20
KS	KE1	Kibiri forest, Western Province	Kenya	EG1	34.8719	0.1502	1400	12
KD	KE2	Kakamega Forest, Western Province	Kenya	EG1	34.5190	0.1415	1400	17
KC	KE3	Kapcherop, Cherangani Forest, Rift Valley	Kenya	EG1	35.2161	1.0315	1900	18
KE	KE4	Londiani, Rift Valley	Kenya	EG1	35.1836	-0.2737	1750	20
KA	KE5	Chuka, Central province	Kenya	EG2	37.3645	-0.1798	3550	19
KT	KE6	Lari, Central Province	Kenya	EG2	36.687	-1.0311	2250	5
KF	KE7	Ol Danyo Sambuk, Central Province	Kenya	EG2	37.1501	-1.0744	1450	15
KG	KE8	Taita Hills, Eastern Arc	Kenya	EG2	38.2088	-3.2462	900	18
TB	TZ1	Kilimanjaro Catchment Forest Reserve	Tanzania	EG2	37.5237	-3.0036	2000	11
TC	TZ2	Kindoroko Catchment Reserve, Eastern Arc	Tanzania	EG2	37.6267	-3.7389	1750	6
TD	TZ3	Shume Magamba Forest Reserve, Eastern Arc	Tanzania	EG2	38.2521	-4.7544	1700	3
TA	TZ4	Meru Catchment Forest	Tanzania	EG2	36.8074	-3.2925	2000	16
TT	TZ5	Udzungwa, Eastern Arc	Tanzania	EG2	36.7791	-7.7668	1150	10
TE	TZ6	Kidabaga, Eastern Arc	Tanzania	EG2	35.9296	-8.1101	2000	15
ZWA	ZW1	Nyanga National Park	Zimbabwe	SG	32.7401	-18.2873	1800	19
ZWB	ZW2	Cashel Valley Chimanimani	Zimbabwe	SG	32.8004	-19.5781	1350	13
ZWC	ZW3	Chirinda forest Reserve Chipinge	Zimbabwe	SG	32.6950	-20.4090	1200	20
SA	SA1	Mpumalanga	South Africa	SG	30.7306	-25.0225	1300	18
SB	SA2	KwaZulu-Natal	South Africa	SG	30.2725	-29.2872	1500	12
MA	MG1	Marovoay	Madagascar	MG	48.2825	-19.1183	1100	31
MC	MG2	Antsahabiraoka	Madagascar	MG	49.2164	-14.4015	1350	6

N, number of samples.

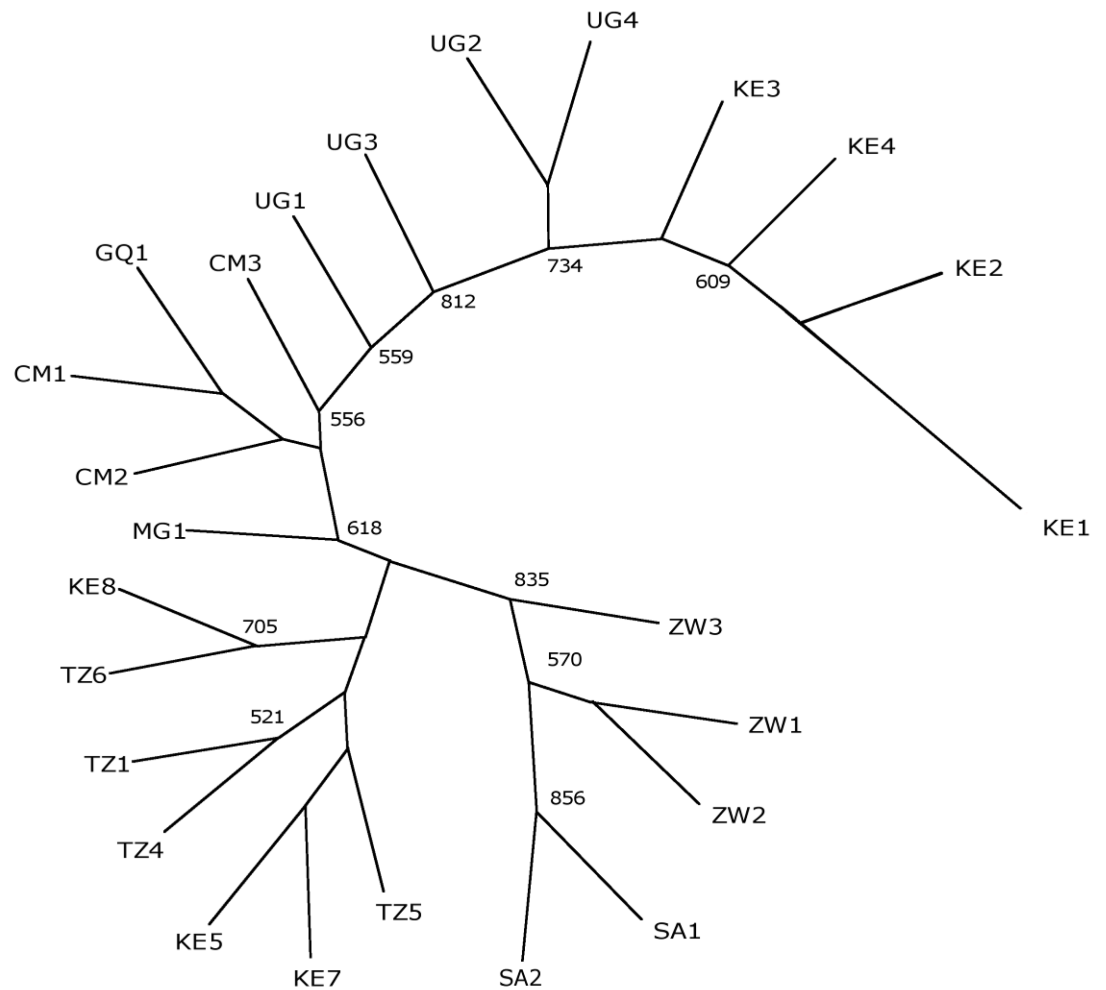

Figure 4.6: Consensus tree based on 1,000 bootstrap replicates of Cavalli-Sforza chord distance matrices of nuclear SSR data for 25 *Prunus africana* populations. Numbers indicate bootstrap support for nodes.

4.2.2 Genetic differentiation and barriers to gene flow

Over all populations a high global F_{ST} value of 0.27 was obtained. Population differentiation taking into account allele size and a stepwise mutation model was significantly higher with R_{ST} value of 0.70; indicating a strong phylogeographic pattern within the sample. Relationships among populations based on a bootstrapped Cavalli-Sforza and Edwards (1967) chord distance

matrix are shown in the UPGMA tree (Figure 4.6): the population from Madagascar appeared central in the unrooted tree; South African and populations east of the Eastern Rift Valley appeared close to each other, as did populations from west Africa and west of the Eastern Rift Valley (with population CC (CM3) on the same branch with these east African populations). Despite geographic proximity, populations east and west of the Eastern Rift were the most diverged on the tree. A similar result was obtained in the PCA (Figure. 4.7); in this analysis population CA (CM2) from Cameroon clustered with populations from East Africa west of the Eastern Rift, also these two groups were not as clearly differentiated as the other groups.

Main barriers to gene flow as detected with BARRIER 2.2 are indicated in Figure 4.8. The first barrier (91.8 % support) was found between Madagascar and mainland Africa. The second barrier (54.8 % support) was found between populations of east and west of the Eastern Rift Valley. The third barrier between populations of West and East Africa was only weakly supported (34 %), similarly to all further barriers. Because of the strong regional structuring of the dataset, a separate analysis of gene flow barriers was conducted within the four groups that held more than two populations (Figure 4.8). In West Africa, the most important gene flow barrier among populations (61 % support) was detected between CB (CM1) and CA (CM2); in populations west of the Eastern Rift the main barrier (55 % support) separated population UC (UG1) from neighbouring populations; in populations east of the Eastern Rift the sharpest break in allele frequencies was detected between population KG (KE8) and all other populations (76 % support); in the Southern African group

the strongest barrier (72 %) support was found between populations of South Africa and Zimbabwe.

Table 4.7: Allelic richness (*Rs*; after rarefaction to 13 individuals) and number of alleles (*a*) per locus and population in *Prunus africana*.

Pop. code	Pop. group	N_a	N_e	H_o	H_e	F		Ap	Rs	Wilcoxon signed rank test (TPM)	
GQ (GQ1)	WG	5.833	4.027	0.744	0.735	-0.074		2	5.388	0.2188	
CB (CM1)	WG	7.500	4.866	0.683	0.719	0.158		0	6.090	0.0156	
CA (CM2)	WG	6.167	3.730	0.531	0.647	0.174	*	0	5.287	0.0391	
CC (CM3)	WG	5.667	3.066	0.658	0.645	-0.051		2	4.681	0.0156	
UC (UG1)	EG1	8.667	5.334	0.733	0.771	0.015	*	1	6.789	0.0156	
UB (UG2)	EG1	7.833	4.530	0.692	0.726	0.042		0	6.094	0.0234	
UA (UG3)	EG1	8.167	4.807	0.808	0.722	-0.146	**	0	6.588	0.2188	
UD (UG4)	EG1	7.833	5.521	0.658	0.775	0.158	***	0	6.606	0.0078	a
KS (KE1)	EG1	9.167	6.936	0.736	0.827	0.104	*	1	8.436	0.0156	
KD (KE2)	EG1	8.500	5.533	0.725	0.761	0.036		1	7.081	0.0234	
KC (KE3)	EG1	8.167	5.160	0.620	0.737	0.119	***	1	6.696	0.0156	
KE (KE4)	EG1	8.000	5.441	0.692	0.763	0.073	**	1	6.596	0.2813	a
KA (KE5)	EG2	6.667	4.199	0.781	0.716	-0.130		0	5.662	0.0078	a
KF (KE7)	EG2	4.667	3.200	0.711	0.616	-0.188	**	1	4.247	0.0078	
KG (KE8)	EG2	4.667	2.794	0.491	0.547	0.098	*	1	4.079	0.0156	
TB (TZ1)	EG2	6.000	3.770	0.530	0.657	0.133	*	0	5.797	0.0156	
TA (TZ4)	EG2	5.500	3.423	0.604	0.629	0.019		1	4.928	0.0156	
TT (TZ5)	EG2	7.333	5.339	0.517	0.753	0.364	***	1	7.333	0.0156	
TE (TZ6)	EG2	6.167	3.880	0.578	0.613	-0.011		0	5.499	0.0156	
ZWA (ZW1)	SG	5.167	3.284	0.632	0.612	-0.072		1	4.495	0.0156	
ZWB (ZW2)	SG	6.000	3.355	0.590	0.574	-0.067		3	5.389	0.0781	a
ZWC (ZW3)	SG	4.000	2.688	0.575	0.540	-0.088		0	3.548	0.0078	a
SA (SA1)	SG	4.500	2.937	0.352	0.430	0.182	**	0	3.836	0.0078	a
SB (SA2)	SG	3.500	2.597	0.417	0.467	0.048	*	0	3.398	0.0078	a
MA (MG1)	MG	6.500	3.577	0.500	0.566	0.212		9	4.734	0.0078	

$N_{a,}$ no. of observed alleles; N_e, no of effective alleles; H_o, observed heterozygosity; H_e, unbiased expected heterozygosity; F, fixation index; *, **, *** deviation from HWE significant at the 0.05, 0.01 and 0.001 level, respectively; A_p, number of private alleles; R_s, average allelic richness; test for population bottlenecks based on the Wilcoxon signed rank test under the two-phased model of mutation (95% stepwise, 5% multistep), [a], significant at the p < 0.001 level.

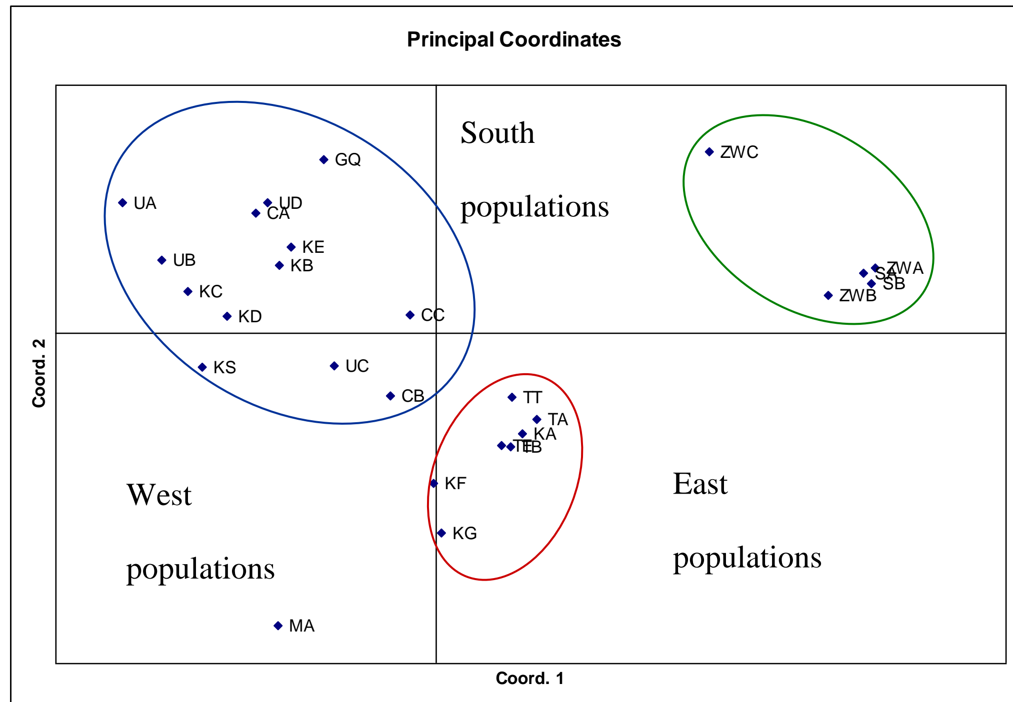

Figure 4.7: Principal coordinate analysis showing the multivariate relationships of the *Prunus africana* populations along the first two axes which explain a total of 67% variation. The first and second axes explain 50% and 17% of the variance, respectively. Population codes are as in Table 4.6).

4.2.3 Individual based population assignment

The results from the individual based assignment with admixture using STRUCTURE are shown in Figure 4.9. Assignment coefficients to a specific cluster were in general very high, bespeaking strong population differentiation. Using the method of Evanno *et al*., (2005) Δln P (D) reached a maximum at five groups, so that the grouping into five clusters was supported as the most likely grouping: West Africa (WG), East Africa west of the Eastern Rift Valley (EG1), East Africa east of the Eastern Rift Valley (EG2), Southern Africa (SG) and Madagascar (MG). When *K* was set to 6 and 7, further clusters

differentiating populations from West Africa and populations from East Africa were detected, respectively (Figure 4.9); but these groupings showed less clear ancestry coefficients. At the *K* value of 7, UC (UG1) showed heavy admixture from the West African Cluster. Above a *K* value of 10, no biologically meaningful splitting in the data set occurred. These five groups were also supported by the results of the distance tree and PCA (Figures 4.6 & 4.7) and thus were taken as the basis for further analyses.

In the AMOVA with F_{ST} genetic differences were partitioned to 14.4 % among groups, 12.6 % within populations among groups and 73.0 % within populations; in the R_{ST} based analysis inflated effects of mutation on regional differentiation were evident as 62.5 % of variation was found among groups, and 7.5 % within populations among groups, and only 30.0 % within populations (Table 4.11)

Table 4.8: Twenty-five natural populations of *Prunus africana* from the five population groups ranked by mean allelic richness.

Pop.	Region	EMPaS01		EMPaSO6		EMPaS10		U3		U5		P2		*Rs*
Code		*Rs*	*a*	*Rs*	*a*	*Rs*	*a*			*Rs*	*a*	*Rs*	*a*	mean
SB (SA2)	SG	3	3	6.6	7	6	6	1	1	1.8	2	2	2	3.398
ZWC (ZW3)	SG	1	1	4.6	5	6.7	9	3	3	2	2	3.9	4	3.548
SA (SA1)	SG	2	2	6.3	8	9.3	11	1	1	1	1	3.4	4	3.836
KG (KE8)	EG2	2.9	3	4.4	5	6.1	7	1.6	2	4.2	5	5.4	6	4.079
KE (KE4)	EG1	2	2	5	5	8.5	10	2.8	3	2.7	3	4.6	5	4.247
ZWA (ZW1)	SG	2	2	8	9	8.1	10	2	2	2.5	3	4.4	5	4.495
CC (CM3)	WG	5.7	7	6.2	7	6.1	7	3.6	5	2	2	4.5	6	4.681
MG (MG1)	MG	2.7	3	6	8	2	2	1.5	2	7.7	11	8.6	13	4.734
TA (TZ4)	EG2	2.6	3	7	8	8.8	10	3.5	4	4	4	3.6	4	4.928
CA (CM2)	WG	6.1	7	8.2	10	6.8	8	3	3	1	1	6.6	8	5.287
GQ (GQ1)	WG	4.7	5	7.3	8	7.2	8	2	2	4.5	5	6.5	7	5.388
ZWB (ZW2)	SG	3.5	4	8.8	10	9.7	11	3.7	4	3	3	3.7	4	5.389
TE (TZ6)	EG2	2.7	3	8.5	10	8.3	9	1.9	2	4	4	7.7	9	5.499
KA (KE5)	EG2	4.6	5	5.4	6	10	13	2.8	3	4.5	5	6.6	8	5.662
TB (TZ1)	EG2	3	3	7.8	8	7.6	8	1.9	2	4.9	5	9.5	10	5.797
CB (CM1)	WG	7.8	10	8.1	10	8.4	11	2.5	3	3	3	6.8	8	6.090
UB (UG2)	EG1	4.6	5	7.2	10	9.3	14	2	2	3.8	4	9.5	12	6.094
UA (UG3)	EG1	7.3	9	6.7	9	9.1	12	1.9	2	6.3	7	8.3	10	6.588
KE (KE4)	EG1	7.2	8	5	6	11	14	2	2	5.5	6	9.1	12	6.596
UD (UG4)	EG1	6.6	8	7.4	9	9.2	11	2	2	5.3	6	9.3	11	6.606
KC (KE3)	EG1	7.9	9	8	10	9.1	12	2	2	2.9	3	10	13	6.696
UC (UG1)	EG1	7.4	10	7.7	10	10	13	2.5	3	4	4	8.9	12	6.789
KD (KE2)	EG1	7.2	9	8.1	9	13	16	2	2	4.6	5	8.1	10	7.081
TT (TZ5)	EG2	3	3	8	8	15	15	2	2	4	4	12	12	7.333
KS (KS1)	EG1	7.6	8	8.5	9	13	14	2	2	6.5	7	13	15	8.436
All		10	19	10	23	15	45	4.2	10	6.6	17	13	30	9.928

R_s, allelic richness per locus and population (after rarefaction to 10 individuals); *a*, number of alleles.

4.2.4 Recent Gene flow

Estimates of recent migration between populations obtained with BAYESASS were relatively low, but differed (significantly) among the five regions (Table 4.12). In general, recent migration occurred mainly within population groups. Only few estimates >1% suggested inter-group migration and this only applied to migration between WG and EG1, and EG2 and SG, respectively. No

evidence for recent migration between spatially close populations east and west of the Eastern Rift Valley was found. While populations from West Africa, Southern Africa and Madagascar (one population) showed evidence of very little current gene flow among them, with average self-recruitment of 0.90, 0.83, 0.99, respectively, values were higher in populations of East Africa. Within this region populations belonging to the EG1 on average showed the highest level of recent migration among populations (self-recruitment average: 0.79), while populations bordering the eastern side of the Rift (KA (KE5) and KF (KE7)) were mostly self-recruiting (0.88). Higher migration estimates (0.78) were found in the populations from Tanzania (populations around Mount Kilimanjaro), which also appeared to have a minor migration rate to populations of the Southern range. The population from Madagascar had the lowest migration rate of all populations; sharing of migrants within WG was inferred only between CB (CM1) and the island population GQ (GQ1).

4.2.5 Historic gene flow

Historic rates of gene flow among groups derived from the coalescent computations implemented in MIGRATE were similarly low and divergent among regions (Table 4.13). Estimates of migration rates (given as scaled migration rate *M*) were similar between computations involving the whole dataset and a random subset (not shown). Estimates of *M* between Madagascar

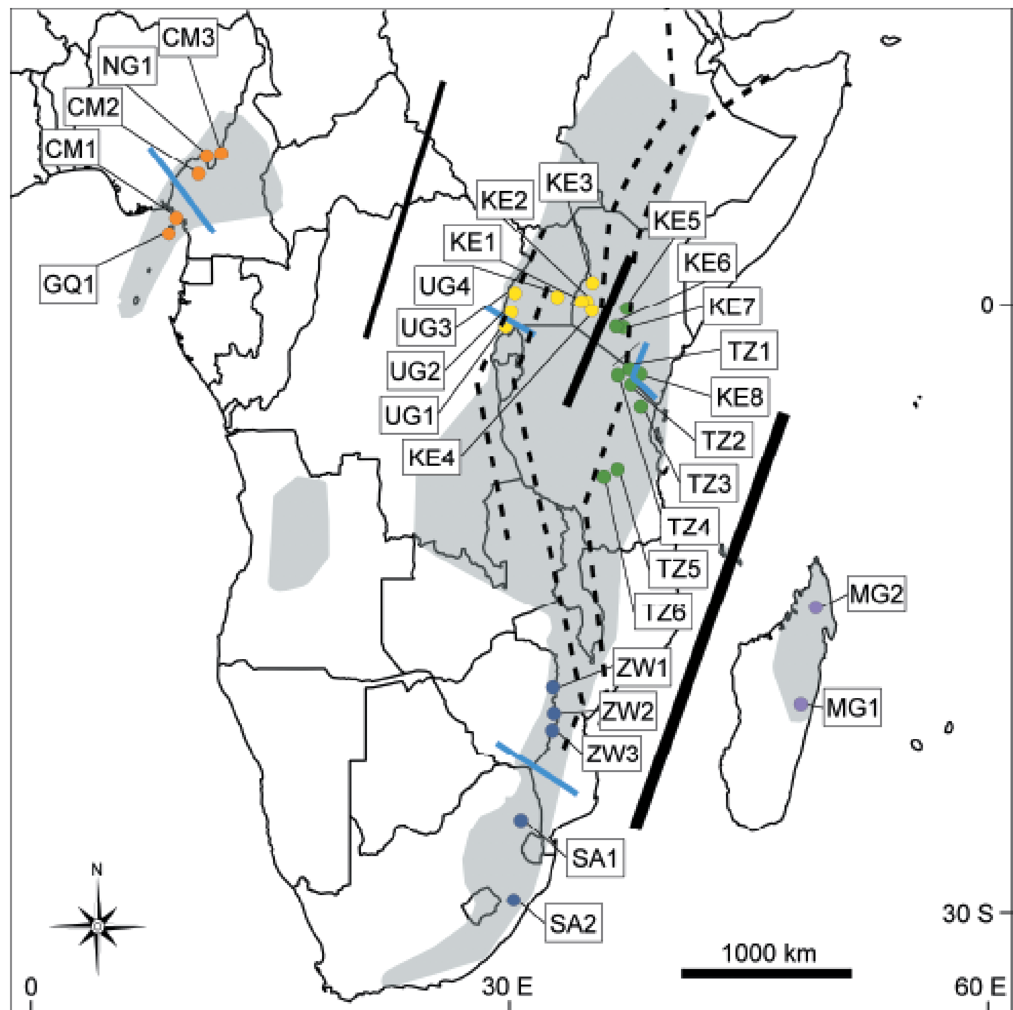

Figure 4.8: Sampling locations, group membership and barriers to gene flow for 30 natural populations of *Prunus africana* (natural distribution range indicated by grey shaded area). Broken lines indicate the Rift Valley System. Colours indicate membership to genetic population groups. The main gene flow barriers among regions as recovered in barrier are indicated by black bars; barriers within regions are indicated by blue bars. The confidence of the barrier is indicated in line thickness with strong lines indicating the best-supported barriers

and the mainland populations were low. Historic gene flow appeared exclusive between SG populations and EG2 populations. There was, opposed to the results obtained for recent migration rates, evidence for moderate bidirectional historic gene flow between populations east (EG2) and west of the Eastern Rift Valley (EG1). The highest historic gene flow estimates were derived for

populations west of the Rift valley (EG1) and West Africa (WG). The derived direction of gene flow was primarily from EG1 to WG. Estimates of scaled population size (θ) were similar in EG1, EG2 and highest for MG, but comparatively low for WG and SG. Estimates of θ were similar for the different computations, although in the random subset (smaller sample size) θ for EG1 was about 20 times larger than estimated with the full data set. However, the θ estimates need to be interpreted cautiously, as sub-group structure and observed deviations from model assumptions can influence estimates (cf. Beerli and Felsenstein, 2001).

Table 4.9: Null allele occurrence in 25 populations of *Prunus africana*.

Name of Population	Pop.code	Grou	EMPaS	EMPaS	EMPaS	U3	U5	P
1. Moka	GQ (GQ1)	WG	-	-	-	-	-	-
2.Lower Mann's Spring, Mt	CB (CM1)	WG	-	-	-	*** 0.265	-	-
3. Ngashie-Mt Oku	CA (CM2)	WG	-	-	-	*** 0.326	-	-
4. Mt Danua	CC (CM3)	WG	-	-	-	-	-	-
5. Bwindi Forest	UC (UG1)	EG2	*** 0.1788	-	-	-	-	-
6. Kalinzu Forest Reserve	UB (UG2)	EG2	*** 0.1988	-	-	-	-	-
7. Mabira Forest	UD (UG4)	EG2	*** 0.1646	-	-	-	*** 0.140	-
8. Kibale Forest Natural Park	UA (UG3)	EG2	-	-	-	-	-	-
9. Kibiri forest, Western Province	KS (KE1)	EG2	-	-	-	-	-	-
10.Kapcherop, Cherangani Forest,	KC (KE3)	EG2	*** 0.3242	-	-	-	*** 0.251	-
11.Kakamega Forest, Western	KD (KE2)	EG2	-	-	-	-	-	-
12. Londiani, Rift Valley	KE (KE4)	EG2	*** 0.1399	-	-	-	-	-
13.Ol Danyo Sambuk, Central	KF (KE7)	EG2	-	-	-	-	-	-
14. Chuka, Central province	KA (KE5)	EG2	-	-	-	-	-	-
15. Taita Hills, Coast Province	KG (KE8)	EG2	-	-	-	-	*** 0.255	-
16. Kidabaga	TE (TZ6)	EG2	-	-	*** 0.1112	-	-	-
17. Udzungwa	TT (TZ5)	EG2	*** 0.2509	-	-	*** 0.407	*** 0.304	-
18. Meru Catchment Forest	TA (TZ4)	EG2	-	-	-	-	*** 0.178	-
19.Kilimanjaro Catchment Forest	TB (TZ1)	EG2	-	-	-	-	*** 0.247	-
20. Nyanga National Park	ZWA	SG	-	-	-	-	-	-
21. Cashel Valley Chimanimani	ZWB	SG	-	-	-	-	-	-
22. Chirinda forest Reserve Chipinge	ZWC	SG	-	-	-	-	-	-
23. Mpumalanga	SA (SA1)	SG	-	-	-	-	-	-
24. KwaZulu-Natal	SB (SA2)	SG	-	-	-	-	-	-
25. Marovoay	MA	MG	-	*** 0.1274	-	*** 0.163	-	-

*** significant null alleles detected, estimated frequency of null alleles after

Table 4.10: Pairwise differentiation by F_{ST} among *Prunus africana* populations (below diagonal), and inter-population Nei's unbiased genetic distance (above diagonal).

	GQ	CB	CA	CC	UC	UB	UD	UA	KS	KC	KD	KE	KF	KA	KB	KG	TE	TT	TA	TB	ZW	ZW	ZW	SA	SB	MA
GQ	-	0.47	0.67	0.69	0.63	0.66	0.51	0.50	0.41	0.42	0.46	0.44	1.30	1.20	0.68	1.47	1.04	1.05	1.25	1.14	1.85	1.61	1.44	1.98	2.04	1.92
CB	0.10	-	0.73	0.40	0.37	1.11	0.79	1.14	0.31	0.79	0.51	0.66	0.61	0.66	0.60	0.66	0.45	0.48	0.71	0.41	1.59	1.45	1.58	1.60	1.65	0.65
CA	0.13	0.14	-	0.66	0.98	0.87	0.89	1.18	0.85	1.32	1.01	0.94	1.78	1.43	0.37	1.83	1.10	0.90	1.35	1.61	1.92	2.54	1.14	3.72	3.47	1.53
CC	0.12	0.08	0.15	-	0.43	0.94	0.74	1.00	0.61	0.84	0.79	0.56	0.95	1.03	0.30	0.97	0.61	0.86	0.94	0.65	1.45	1.45	1.38	1.31	1.35	0.72
UC	0.09	0.06	0.14	0.07	-	0.50	0.54	0.62	0.27	0.47	0.31	0.44	0.63	0.84	0.57	0.52	0.59	0.65	0.71	0.52	1.69	1.66	1.72	1.69	1.93	0.82
UB	0.09	0.14	0.15	0.14	0.08	-	0.23	0.40	0.29	0.42	0.38	0.31	1.42	1.29	0.62	0.87	1.16	1.08	1.02	1.16	2.53	2.97	1.71	3.48	3.30	1.52
UD	0.07	0.12	0.13	0.11	0.07	0.04	-	0.25	0.23	0.30	0.34	0.26	1.19	1.13	0.51	1.30	1.17	0.99	1.28	1.08	2.06	2.05	1.62	2.25	2.03	1.44
UA	0.08	0.16	0.16	0.15	0.10	0.07	0.05	-	0.27	0.28	0.30	0.29	1.15	1.44	0.75	1.58	1.68	1.34	1.63	1.41	4.15	3.08	2.23	2.68	3.21	2.05
KS	0.06	0.07	0.13	0.09	0.04	0.05	0.04	0.06	-	0.08	0.06	0.09	0.57	0.60	0.54	0.65	0.56	0.44	0.68	0.52	2.82	2.42	1.73	2.64	2.51	0.85
KC	0.07	0.13	0.17	0.13	0.07	0.07	0.05	0.05	0.03	-	0.13	0.16	0.97	1.10	0.95	1.01	1.21	0.80	1.10	1.00	3.13	2.58	2.05	2.83	2.59	1.51
KD	0.07	0.09	0.15	0.12	0.05	0.06	0.05	0.06	0.02	0.03	-	0.19	0.75	0.82	0.77	0.90	1.00	0.68	0.92	0.79	2.83	2.40	1.76	2.20	2.34	1.13
KE	0.06	0.10	0.14	0.10	0.06	0.05	0.04	0.06	0.03	0.03	0.03	-	0.91	0.91	0.71	1.04	0.94	0.78	0.96	0.80	2.08	1.88	1.63	1.79	1.94	1.40
KF	0.18	0.10	0.23	0.16	0.11	0.18	0.16	0.18	0.10	0.16	0.13	0.14	-	0.17	1.12	0.62	0.82	0.38	0.42	0.31	1.82	1.71	1.25	1.69	1.91	0.87
KA	0.14	0.09	0.17	0.14	0.10	0.14	0.12	0.15	0.08	0.13	0.10	0.11	0.05	-	1.07	0.59	0.60	0.15	0.30	0.40	0.95	1.08	0.69	1.25	1.37	0.89
KB	0.11	0.11	0.10	0.08	0.09	0.10	0.08	0.12	0.08	0.13	0.11	0.10	0.17	0.14	-	1.12	0.92	0.90	1.42	1.09	1.72	2.00	1.35	2.40	3.26	1.13
KG	0.22	0.12	0.25	0.18	0.12	0.18	0.20	0.23	0.13	0.19	0.17	0.18	0.15	0.13	0.20	-	0.25	0.42	0.65	0.52	1.35	1.51	2.38	1.72	1.71	0.85
TE	0.17	0.09	0.19	0.13	0.11	0.18	0.17	0.21	0.11	0.18	0.16	0.15	0.15	0.12	0.16	0.07	-	0.34	0.73	0.54	0.94	1.07	1.56	1.31	1.22	0.83
TT	0.13	0.09	0.15	0.13	0.09	0.13	0.11	0.15	0.07	0.11	0.10	0.10	0.10	0.04	0.13	0.12	0.10	-	0.42	0.37	0.70	0.78	0.62	1.31	1.27	0.93
TA	0.17	0.11	0.20	0.15	0.10	0.15	0.16	0.19	0.10	0.15	0.13	0.14	0.10	0.07	0.18	0.15	0.15	0.09	-	0.23	1.05	1.11	0.76	1.07	1.17	1.09
TB	0.17	0.07	0.21	0.12	0.09	0.17	0.16	0.20	0.10	0.16	0.13	0.14	0.07	0.08	0.17	0.12	0.11	0.09	0.06	-	1.25	1.03	0.89	1.23	1.02	0.79
ZW	0.19	0.19	0.24	0.19	0.17	0.21	0.18	0.22	0.18	0.21	0.20	0.18	0.24	0.15	0.21	0.24	0.19	0.13	0.19	0.20	-	0.11	0.39	0.34	0.42	3.29
ZW	0.20	0.20	0.27	0.21	0.19	0.23	0.20	0.24	0.19	0.22	0.21	0.20	0.25	0.17	0.24	0.26	0.22	0.15	0.21	0.21	0.04	-	0.53	0.38	0.53	2.56
ZW	0.20	0.21	0.22	0.21	0.20	0.21	0.19	0.23	0.19	0.22	0.20	0.20	0.24	0.15	0.21	0.32	0.27	0.15	0.17	0.20	0.13	0.16	-	0.69	0.52	2.83
SA	0.28	0.28	0.36	0.27	0.25	0.30	0.27	0.30	0.26	0.30	0.27	0.26	0.32	0.24	0.31	0.35	0.31	0.24	0.26	0.29	0.16	0.16	0.25	-	0.10	3.29
SB	0.27	0.27	0.34	0.26	0.25	0.28	0.25	0.30	0.24	0.28	0.26	0.25	0.31	0.23	0.31	0.34	0.28	0.23	0.26	0.26	0.16	0.19	0.21	0.06	-	3.21
MA	0.22	0.11	0.23	0.15	0.14	0.21	0.19	0.24	0.14	0.21	0.17	0.19	0.16	0.15	0.19	0.17	0.16	0.16	0.18	0.14	0.28	0.29	0.30	0.37	0.36	-

Table 4.11: Results of analysis of molecular variance (AMOVA) from data of six nuclear microsatellite loci of 25 *Prunus africana* populations from seven countries. Groups correspond to the 5 biogeographic regions detected.

	F_{ST}			R_{ST}		
Source of variation	Sum of squares	Variance components	Percentage variation	Sum of squares	Variance components	Percentage variation
Among groups	320.96	0.39	14.44	508857.65	750.14	62.49
Among populations within groups	266.36	0.34	12.57	66982.60	89.95	7.49
Within populations	1662.02	1.98	72.99	302357.47	360.38	30.02
Total	2249.33	2.71		878197.73	1200.47	

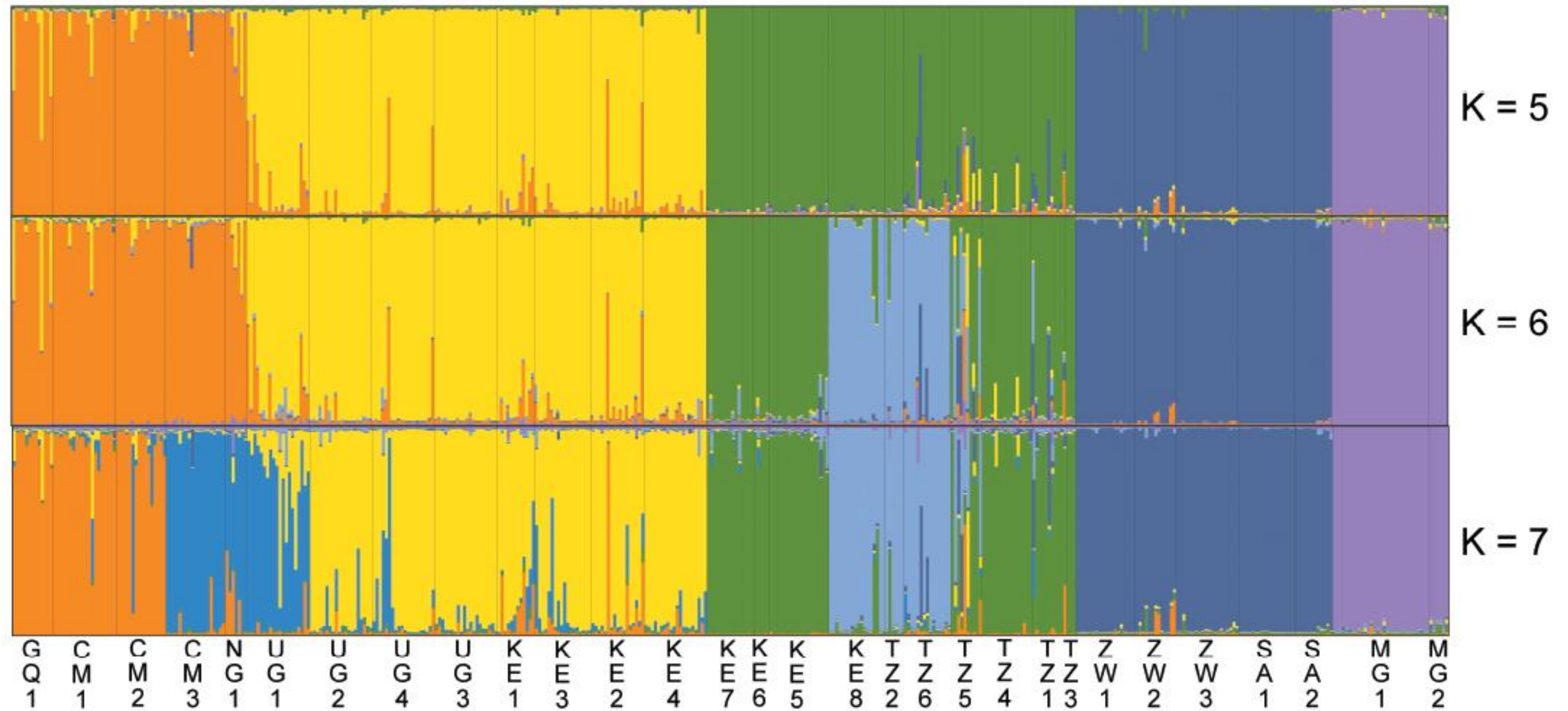

Figure 4.9: Population assignment using structure reveals 5 clusters. At 6 Clusters we see a division in the Tanzanian group and at 7 clusters Bwindi (UG1) clusters with Mt Danoua (CM3) and Nigeria (NG1)

Table 4.12: Recent gene flow estimates (as rate of migration from population *i* into population *j*, average from five computations) among 25 populations of *Prunus africana*; standard deviation in brackets. Rate of self-recruitment is shown in bold. Lines indicate the five regional groups. Values below 0.01 not depicted.

	Population *j*																								
	GQ	CB	CA	CC	UC	UB	UA	UD	KS	KD	KC	KE	KA	KF	KG	TB	TA	TT	TE	ZWA	ZWB	ZWC	SA	SB	MA
Population *i*	WG				EG1								EG2							SG					MG
GQ	**0.8598**	0.0892		0.0141																					
	(0.1403)	(0.1093)		(0.0283)																					
CB	0.0652	**0.8627**																							
	(0.0799)	**(0.1443)**																							
CA			**0.9799**																						
			0.0023																						
CC				**0.8933**																					
				(0.1198)																					
UC					**0.8281**																				
					(0.0952)																				
UB					0.0440	**0.8626**	0.0115	0.0861	0.0395	0.0540	0.0579	0.0454													
					(0.0836)	**(0.1474)**	(0.0230)	(0.1055)	(0.0455)	(0.0702)	(0.0826)	(0.0907)													
UA							**0.8306**																		
							(0.1260)																		
UD							0.0123	**0.8543**																	
							(0.0162)	**(0.1408)**																	
KS					0.0173	0.0327	0.0310		**0.7450**	0.0398	0.0425	0.0452													
					(0.0283)	(0.0654)	(0.0621)		**(0.1089)**	(0.0795)	(0.0850)	(0.0903)													
KD				0.0242	0.0215				0.0516	**0.8008**	0.0838	0.0872													
				(0.0485)	(0.0293)				(0.0636)	**0.1427**	(0.1026)	(0.1069)													
KC					0.0160	0.0137	0.0494				**0.6836**														
					(0.0240)	(0.0273)	(0.0724)				**(0.0013)**														
KE				0.0212		0.0404			0.0138	0.0194	0.0252	**0.7423**													
				(0.0423)		(0.0808)			(0.0276)	(0.0389)	(0.0504)	**(0.1188)**													
KA													**0.8592**	0.0478	0.0429	0.0267	0.0490	0.0484	0.0171						
													(0.1449)	(0.0726)	(0.0859)	(0.0343)	(0.0773)	(0.0422)	(0.0342)						
KF													0.0897	**0.9024**		0.0260		0.0357							
													(0.1098)	**(0.1144)**		(0.0343)		(0.0445)							
KG															**0.8629**				0.0037						
															(0.1466)				(0.0073)						
TB																**0.6931**									
																(0.0004)									
TA															0.0434	0.0743	**0.8988**	0.0277	0.0735						
															(0.0868)	(0.0458)	**(0.1126)**	(0.0340)	(0.0901)						
TT																	0.0126	**0.7134**	0.0371						
																	(0.0252)	**(0.0382)**	(0.0742)						
TE																			**0.7598**						
																			(0.1066)						
ZWA																				**0.8633**	0.0949		0.0438	0.0304	
																				(0.1468)	(0.0775)		(0.0876)	(0.0609)	
ZWB																					**0.6887**				
																					(0.0005)				
ZWC																0.0182		0.0235		0.0445	0.0330	**0.9845**			
																(0.0063)		(0.0040)		(0.0889)	(0.0660)	**(0.0008)**			
SA																				0.0443	0.0351		**0.9233**	0.1218	
																				(0.0887)	(0.0633)		**(0.1202)**	(0.0609)	
SB																								**0.6907**	
																								(0.0005)	
MA																									**0.9900**
																									(0.0003)

Table 4.13: Results of historic gene flow analysis among population groups derived from MIGRATE; θ is the estimated mutation-scaled population size, M is the mean estimated mutation-scaled migration rate (receiving populations in rows); upper and lower 99.5% profile likelihood percentiles given in brackets. Values for M >1 in bold.

		M (m/μ)				
		WG	EG1	EG2	SG	MG
	θ ($4N_e\mu$)	1->	2->	3->	4->	5->
WG	2.4562 (2.2561-2.6808)	x	2.6182 (2.3228-2.9376)	0.5895 (0.4542-0.7487)	0.1465 (0.0645-0.2315)	0.2266 (01480-0.3290)
EG1	4.3005 (4.0447-4.5784)	1.388 (1.2220-1.5681)	x	1.0171 (0.8767-1.1718)	0.1007 (0.0610-0.1547)	0.3796 (0.2966-0.4769)
EG2	3.803 (3.5489-4.0820)	0.0898 (0.0491-0.1486)	1.3086 (1.1290-1.5063)	x	1.5507 (1.3543-1.7653)	0.3543 (0.2647-0.4622)
SG	2.7108 (2.4981-2.9483)	0.5901 (0.4586-0.7447)	0.2525 (0.1701-0.3579)	1.4215 (1.2102-1.6556)	x	0.0001 (0.0000-0.0171)
MG	7.1183 (6.3558-8.0076)	0.2157 (0.1635-0.2778)	0.0655 (0.0389-0.1021)	0.1321 (0.0923-0.1819)	0.0023 (0.0000-0.0139)	x

4.3 Intraspecific chemical diversity

The average concentration [mg/kg w/w] based on 20 bulked population samples (population mix) in increasing order was: lauric acid (18), myristic acid (22), n-docosanol (25), ferulic acid (49), β-sitostenone (198), β-sitosterol (490), and ursolic acid (743). Friedelin was not detected in any of the samples above the limit of detection (7 mg/kg). Concentration of the other bark constituents varied considerably among and within populations. However, both for lauric and myristic acid the content in individual samples within certain populations partly varied erratically. Due to the high variation of individual

concentrations in these two analytes, they were not considered for further analysis.

The variation in concentration among different populations of the other key compounds studied is shown in Figure 4.9. Docosanol was found to be highest in Ngashie Mt Oku (CA-Cameroon) followed by Kapcherop (KC-Kenya) and lowest in Kidabaga (TE-Tanzania) followed by Moka (GQ-Equatorial Guinea) with concentrations of 10 to 28 mg/kg. Ferulic acid was found to be highest in Kinale (KB-Kenya) followed by Mpumalanga (SA-South Africa) and lowest in Moka (GQ-Equatorial Guinea) followed by Kidabaga (TE-Tanzania) with concentrations of 30 to 90 mg/kg. Content of ferulic acid in the samples was significantly correlated to tree diameter (r = 0.23, $p \leq 0.05$) and annual precipitation (r = -0.50, $p \leq 0.025$). Concentration of β-sitostenone was found to be highest in Mt. Danoua (CC-Cameroon) followed by Lakato forest (MB-Madagascar) and lowest in Chimanimani (ZW-Zimbabwe) followed by Bwindi (UC-Uganda) with concentrations of 116 to 260 mg/kg. Concentration of Beta-sitosterol was found to be highest in Chimanimani (ZW-Zimbabwe) followed by Ngashie Mt Oku (CA-Cameroon) and lowest in Chuka (KA-Kenya) followed by Kinale (KB-Kenya) and ranged from 349 to 583 mg/kg. Among all compounds, the highest concentration was found for ursolic acid varying between 317 and 2000 mg/kg. Concentration of ursolic acid was highest in the Marovoay (MA-Madagascar) followed by Lakato forest (MB-Madagascar) and lowest in Chimanimani (ZW-Zimbabwe) followed by Kidabaga (TE-Tanzania). The concentration of ursolic acid was negatively correlated with the diameter of the trees (r = -0.33, $p \leq 0.01$).

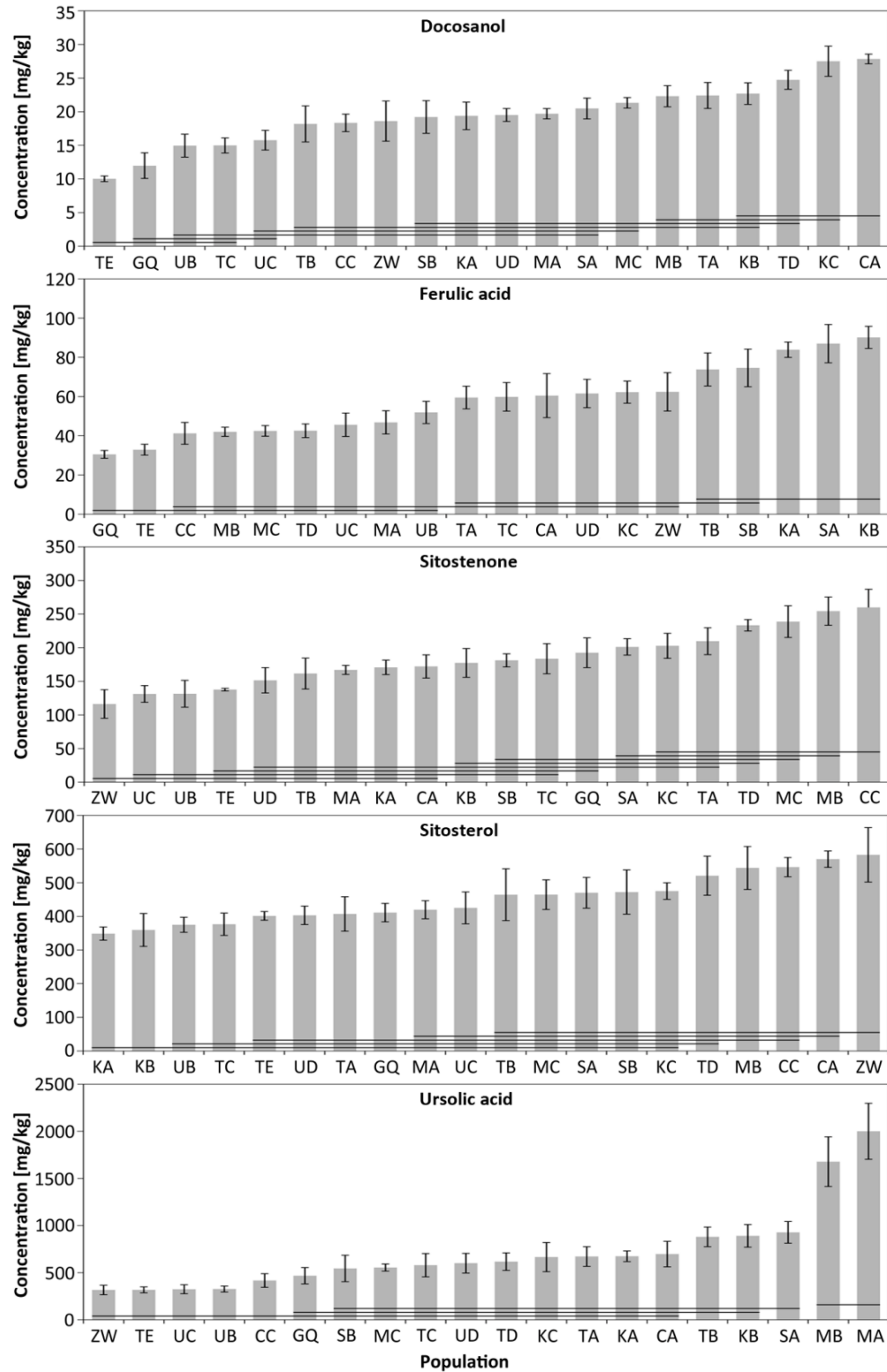

Figure 4.10: Mean chemical analyte concentration (mg/kg) of *Prunus africana* individuals from 20 populations within the natural range of the species ranked in ascending order. Populations that are connected with a horizontal line are statistically not different. Respective vertical lines indicate standard errors

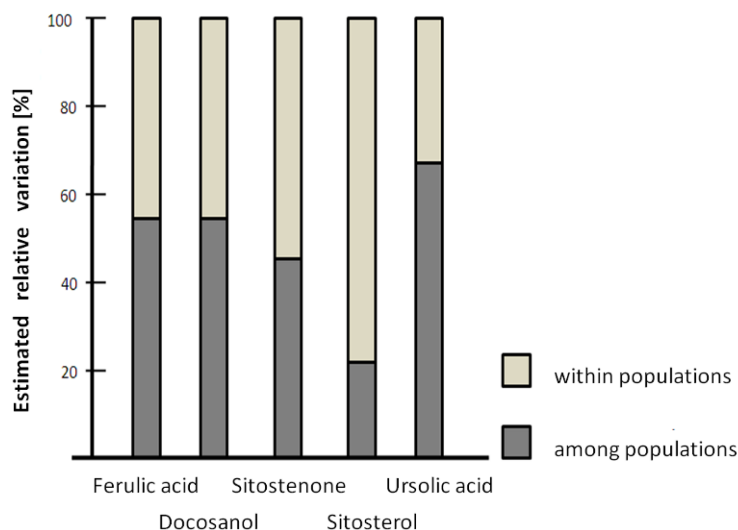

Figure 4.11: Analysis of variance of chemical analytes in *Prunus africana*. Estimates of relative variance components of each chemical analyte are shown.

From the analysis of variance of chemical analytes in *P. africana* carried out (Figure 4.11), three constituents, ferulic acid, n-docosanol and β-sitostenone, showed equivalent variation within and among population. The variation with highest among population component, with roughly two thirds, was attributed to ursolic acid, whilst the among population component for β-sitosterol amounted to approximately 20 %.

Table 4.14: Correlations between the concentrations of different bark constituents of *Prunus africana*. Asterisks indicate significance level (*** $p \leq .001$, ** $p \leq .01$, * $p \leq .05$)

	Docosanol	Lauric acid	Myristic acid	Sitostenone	Sitosterol	Ursolic acid
Ferulic acid	.266 **	.118	.175	-.153	-.041	.114
Docosanol		-.420***	-.315***	.508***	.557***	.339***
Lauric acid			.365***	-.294**	-.177	-.038
Myristic acid				-.281***	-.198*	-.053
Sitostenone					.391***	.263**
Sitosterol						.153

The relative mean of various chemical compounds was estimated in order to enable significant comparisons between populations (Figure 4.12). Here, the maximum mean was used as the denominator and the relative fraction of each analyte per population was obtained. Ursolic acid was found in very high concentrations in two out of three Madagascan populations, while in most other populations the concentration amounted to approximately 30% of the maximum found in Marovoay (MA-Madagascar); this is the only sample whose contents exceeded 50% of the respective maximum in all constituents. On the other hand, in the Tanzanian sample Kidabaga (TE) very low relative contents for all constituents were found.

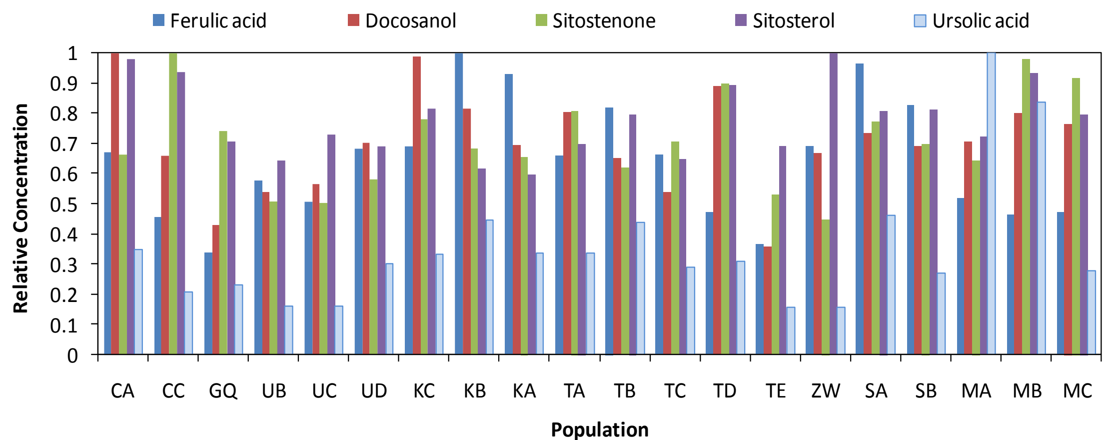

Figure 4.12: Relative mean of the various chemical compounds.

The correlation of haplotype and nuclear genetic diversity showed that content variation of β-sitosterol of the populations was significantly correlated to the effective number of haplotypes ($r = 0.63$, $p \leq 0.05$) and haplotype diversity ($r = 0.66$, $p \leq 0.05$). Variation of ferulic acid was positively correlated to the

fixation index F_{IS} (r = 0.68, $p \leq 0.05$) (Table 4.15). Hence, association between genetic diversity and chemical diversity was very small.

Table 4.15: Correlations between genetic parameters [Number of effective alleles (N_{e_c}), unbiased haplotypic diversity (h_c), number of effective nuclear alleles (N_{e_n}), unbiased expected nuclear heterozygosity (H_e), Fixation Index (F_{IS}) of the populations] and variation coefficient of five bark constituents of *Prunus africana.* Asterisk indicates significance $p \leq 0.05$.

Genetic Parameter	Ferulic acid	Docosanol	Sitostenone	Sitosterol	Ursolic acid
N_{e_c}	-.20	.24	.20	.63*	-.05
h_c	-.19	.25	.17	.65	-.02
N_{e_n}	.11	.24	.32	.04	.03
H_e	.10	.29	.32	-.08	.11
F_{IS}	.48*	-.05	-.10	-.04	.41

The population cluster based on the content differences of bark constituents was very weakly, but in certain cases significantly correlated with the molecular phylogenetic structure (Figure 4.13). Significant correlations were detected for chloroplast as well as nuclear data both for n-docosanol and ursolic acid. When all constituents were considered for a single Mantel test, both pairwise Φ_{PT} for chloroplast and R_{ST} for nuclear genetic differentiation used as genetic distance resulted in weak, but significant correlations (Table 4.16).

Table 4.16: Significance level of Mantel-tests between chemical and genetic distances based on chloroplast (Φ_{PT}) and nuclear microsatellites (R_{ST}), respectively.

	Φ_{PT}	R_{ST}
Docosanol	.032	.048
Ferulic acid	.213	.202
Lauric acid	.321	.094
Myristic acid	.143	.094
Sitostenone	.174	.224
Sitosterol	.293	.412
Ursolic acid	.026	.019
All constituents (Euclidean)	.027	.010
All constituents (Mahaloboni)	.046	.080

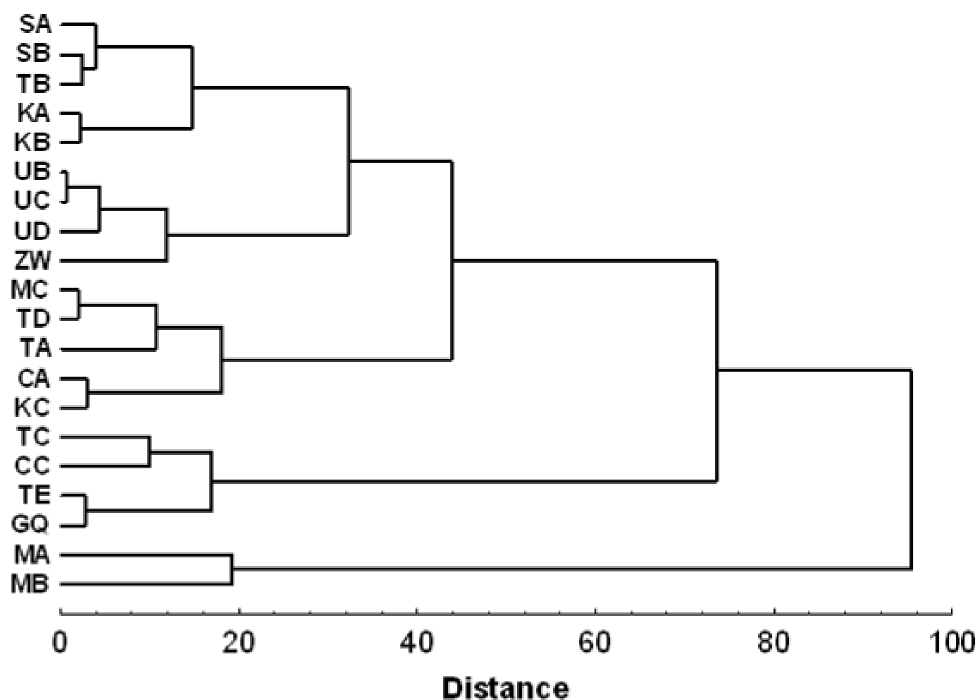

Figure 4.13: Clustering of populations based on chemical distance

Table 4.17: Description and location of 20 populations of *Prunus africana* sampled.

Name of Population	Code	Country	Longitude	Latitude	dbh (cm)	Elevation asl (m)	Annual MeanTemperature ^{O}C (Worldclim)	Annual Precipitation mm (Worldclim)
1. Ngashie Mt. Oku	CA	Cameroon	10,50916	6,20489	58,36	2729,0	13,20	2088,0
2. Mt. Danoua	CC	Cameroon	11,94175	7,18728	35,20	1731,8	19,14	1687,0
3. Moka	GQ	Equatorial Guinea	8,65631	3,36217	0,00	1510,4	20,56	2364,4
4. Chuka	KA	Kenya	37,74682	-0,53182	69,50	1174,6	20,84	1021,0
5. Kinale	KB	Kenya	36,72654	-0,98154	50,10	2165,4	15,48	1176,8
6. Kapcherop	KC	Kenya	35,52572	1,06878	91,34	2720,6	13,46	1111,4
7. Marovoay	MA	Madagascar	48,30586	-18,80238	18,40	1062,2	18,76	1487,0
8. Lakato Forest	MB	Madagascar	48,27377	-19,21113	14,40	936,6	19,26	1705,4
9. Antsahabiraoka	MC	Madagascar	49,21566	-14,40023	28,60	1271,6	19,00	1430,0
10. Mpumalanga	SA	South Africa	30,77360	-24,92951	72,80	1474,0	16,30	926,2
11. KwaZulu Natal	SB	South Africa	30,27138	-29,28646	0,00	1566,2	13,30	888,6
12. Meru	TA	Tanzania	36,70766	-3,29235	75,80	1884,0	16,50	971,0
13. Kilimanjaro	TB	Tanzania	37,52374	-3,00349	77,66	1976,0	16,30	999,0
14. Kindoroko	TC	Tanzania	37,64340	-3,73895	66,83	1875,0	17,60	1005,0
15. Shume Magamba	TD	Tanzania	38,28784	-4,72057	84,72	1738,4	15,44	984,6
16. Kidabaga	TE	Tanzania	35,93016	-8,10947	64,20	1988,8	16,98	1052,0
17. Kalinzu	UB	Uganda	30,00184	-0,37422	59,88	1100,0	22,10	1015,0
18. Bwindi	UC	Uganda	29,77540	-1,04760	54,00	2264,4	15,26	1267,8
19. Mabira	UD	Uganda	33,01496	0,38061	53,58	1266,0	21,20	1319,0
20. Chimanimani	ZW	Zimbabwe	32,80107	-19,57720	63,20	1343,2	17,06	1159,4

Table 4.18: Estimates of LOD/LOQ by means of the signal to noise ratios of the standards.

Bark Constituent	LOD mg/kg	LOQ mg/kg
Ferulic acid	1	4
Friedelin	7	24
Docosanol	0.3	1
Lauric acid	0.1	0.3
Myristic acid	0.1	0.3
Sitostenone	23	76
Sitosterol	0.8	3
Ursolic acid	45	150

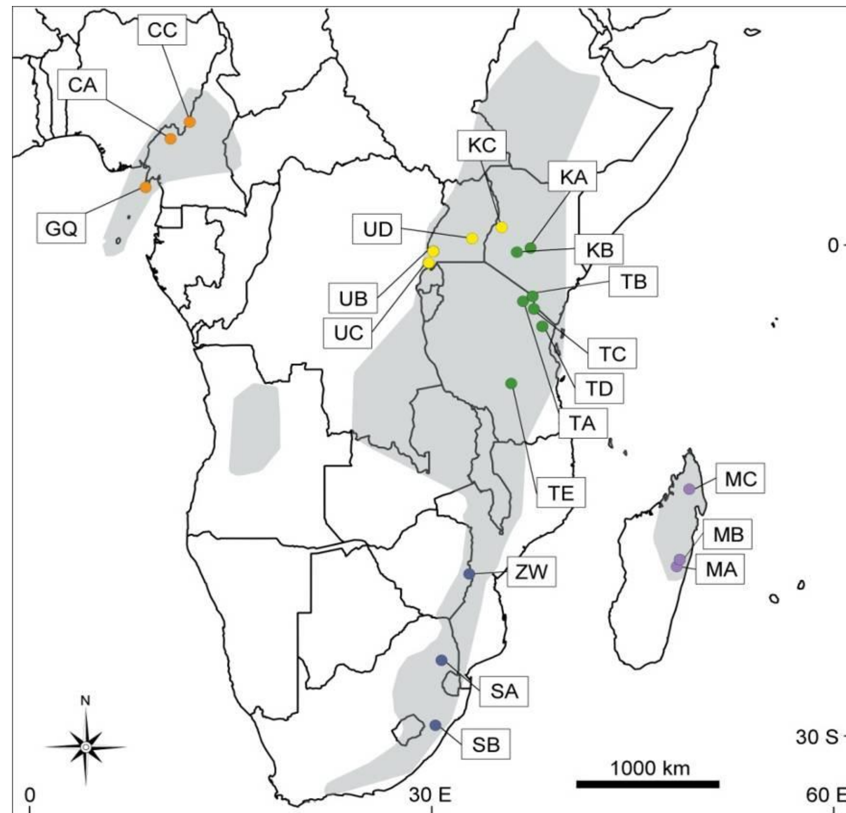

Figure 4.14: *Prunus africana* bark sampled throughout the natural range. Different colours of the sampling sites indicate phylogeographic groups (orange = West Africa, yellow = East Africa west of the Eastern Rift Valley, green = East Africa east of the Eastern Rift Valley, purple = Madagascar, blue =Southern Africa).

CHAPTER FIVE

DISCUSSION, CONCLUSION AND RECOMMENDATTIONS

5.1 Discussion

5.1.1 Introduction

In this study the phylogeography of *P. africana* as a model species to derive vegetation history of Afromontane forests was investigated, population structure and history was elucidated and intraspecific chemical diversity revealed. High population differentiation and significant phylogeographic structure within the species were detected. This finding provides strong evidence for a former migration corridor between 'east' and 'west' African populations. However, evidence for the dislocation of a historic immigration barrier to a more recent barrier to gene flow over several hundred kilometres was found. Similarly, a pronounced variation in the concentration of selected bark constituents was found among populations suggesting spatial genetic effects are likely to be present.

5.1.2 Phylogeography and Population Structure revealed by cpSSR

5.1.2.1 Genetic diversity and population structure based on cpDNA

Compared to other tree species from Africa the observed differentiation among populations (G_{ST} = 0.735) is relatively high and comparable to that found in *Adansonia digitata* (G_{ST} = 0.97; Pock Tsy *et al.*, 2009) and *Hagenia abyssinica* (G_{ST} = 0.899; Ayele *et al.*, 2009), although the observed within population diversity is higher in *P. africana* (h_s = 0.237) than in either *Adansonia* (h_s =

0.017) or *Hagenia* (h_s = 0.079) maybe due to the lower sample size in these studies. Overall multilocus linkage disequilibrium was highly significant (r_d = 0.074, $p < 0.002$), indicating that homoplasy plays a minor role in explaining the observed phylogeographic pattern in comparison to migration. Also, tropical species from South America show similar high patterns of differentiation, e.g. *Cedrela odorata* (G_{ST} = 0.96; Cavers *et al*., 2004) and *Vouacapoua americana* (G_{ST} = 0.89; Dutech *et al*., 2000).

In contrast, other range wide cpDNA studies on Rosaceous trees (e.g., from Europe) show markedly lower genetic differentiation among populations, but higher within population diversity, e.g. *P. avium (*G_{ST} = 0.29; h_s = 0.33; Mohanty *et al*., 2001*), P. spinosa (*G_{ST} = 0.32; h_s = 0.49; Mohanty *et al*., 2002) and *Sorbus torminalis (*G_{ST} = 0.33, h_s = 0.39; Oddou-Muratorio *et al*., 2001). Differences may be explained either by different seed dispersal mechanisms or different population history. Seed dispersal mechanisms play a crucial role in shaping distributions of haplotype diversity and genetic differentiation (Dumolin-Lapègue *et al*., 1997; Demesure *et al*., 1996; Petit, Kremer, and Wagner, 1993a; 2003). Farwig *et al*. (2006) listed 36 frugivorous birds and mammals with body sizes of 10 to 80 cm as seed dispersers for *P. africana*. These dispersers should facilitate short- and medium-distance seed dispersal at least as efficiently as seed dispersers of *Prunus* species in Europe. Therefore, it is probable that differences in the population history of *P. africana* and European *Prunus* species are responsible for the stronger differentiation in the former. European Rosaceous tree populations have been affected by ice ages and recent (within the last 10,000 years) postglacial recolonization of central

and northern European areas, whereas within the same period *P. africana* is likely to have experienced only smaller altitudinal range shifts. In comparison to latitudinal range shifts where a loss of genetic diversity due to repeated bottlenecks can occur (Hewitt, 1996), range shifts in mountain regions occur over much shorter geographic distances and population sizes may be more stable, thereby maintaining diversity (Ehrich *et al*., 2007). The isolation of mountain massifs from each other might limit gene flow and facilitate differentiation of local populations (Hewitt, 2000). Thus, the history of single populations in Africa is doubtless much older than that of European Rosaceous tree populations. Similar strong differentiation patterns were found for Afroalpine populations of *Arabis alpina* in a comparison to European and Arctic populations (Ehrich *et al*., 2007).

5.1.2.2 Phylogeography of *Prunus africana* – migration and dispersal within Africa

Mountain ranges which provide habitats for Afromontane species today were formed within the late Pliocene and throughout the Pleistocene (Hedberg, 1994). During the Pleistocene, several Afroalpine species colonised Africa from Asia through the coastal mountain ranges of the Arabian Peninsula (Koch *et al*., 2006). According to Aubréville (1976), *P. africana* could have followed a similar path based on the putative Laurasian origin of the genus *Prunus*. The current study was not able to provide data on the original immigration path of *P. africana* into the African continent as populations are not extant in the Arabian Peninsula and palaeoecological evidence is not available. However,

current data provides the first significant insights into the population history of *P. africana* within mainland Africa and its neighbouring islands, as a significant signal for phylogeographic pattern was found.

Supported by Bayesian analysis where splitting events between Madagascar and other African populations were estimated to be among the oldest, populations from Madagascar were found to be genetically most distinct, suggesting an early split from the main lineage. Since the separation of Madagascar from Africa was completed 100 million years ago (long before *P. africana* reached Africa), the colonisation of Madagascar may reflect early and possibly unique dispersal events facilitated by birds or the southern Monsoon drift or the Mozambique current (Figure 4.5).

On mainland Africa, a clear division between western populations (including Ugandan populations east of the Albertine Rift valley) and stands in eastern and southern Africa was detected. The haplotype sharing between West African and Ugandan populations suggests the former existence of a migration corridor from east to west. Although the exact migration routes to West Africa cannot to be determined at present, the data supports two scenarios: (i) a split during southward migration of *P. africana* at the southern fringe of the Ethiopian highlands (Figure 4.5, route 1) with migration of Albertine Rift valley populations to West Africa (route 1a, 1b); or (ii) the independent colonisation of West Africa via a north-western migration corridor and the subsequent colonisation of western Uganda from West Africa (route 2). Both scenarios imply that a strong barrier for gene flow has existed within East

Africa and that the distribution range of *P. africana* included large areas of Central Africa at one time, where presently no populations exist (Hall *et al*., 2000).

The first scenario of colonisation from the Albertine Rift to West Africa is supported by the relatively small genetic distances (< 4 mutations) among the haplotypes found in West Africa, which suggests that all six observed haplotypes originate from the dominant type HT1m. In contrast, the various haplotypes found in East Africa have a deeper root with up to 5 mutations among them and include other main backbone lineages. Thus, the high haplotype diversity found in West Africa could be a result of younger differentiation events coupled with isolated mountain habitats that facilitated the fixation of newly mutated types. In contrast, further differentiation at the Albertine Rift may have been prevented by regular seed exchange. The colonisation from East to West could have been through either the SMT as defined according to White (1981) or through a more northern direct connection. The data suggest a more direct connection since the SMT, if it functioned as a bridge, must have originated south of Lake Tanganyika. Here, the Albertine Rift and the Eastern Rift meet and this should have provided chance for lineage mixing: i.e., populations from West Africa would then also be expected to contain haplotypes from the eastern lineage. On the other hand, the absence of such haplotypes in the west could be explained by the later splitting events of the east and southern populations compared to the splitting events between the 'east' and 'west' African populations. More data from

populations sampled south of the Albertine Rift and following the SMT would be needed to resolve this definitively.

Under the second scenario, *P. africana* could have migrated directly from the Ethiopian highlands to West Africa, possibly using the Marra Mountains (Sudan) as stepping stones (cf. Wickens, 1976). Thereafter, subsequent colonisation of Uganda from West Africa could have occurred. This route would provide an alternative explanation for the lower diversity shown in the Albertine Rift valley populations. As there is no extant Afromontane forest or *P. africana* in the central African regions which must have been crossed, only an extensive phylogeographic study of populations from Ethiopian highlands could shed further light on this second hypothesis.

As key barriers to dispersal in the early population history of the species the upper river Nile basin and the Lake Victoria basin is suggested, because these areas experienced severe drought conditions before 12,000 BP (Livingstone, 1975) and probably also earlier in the Pleistocene. According to pollen analysis and radiocarbon data of lake sediments, the water level of Lake Victoria at 14,000 BP was between 26 and 75 m lower than today and the surrounding landscape was likely too dry for *P. africana* and characterised mainly by open vegetation (Livingstone, 1975). About 10,000 years BP evergreen rainforest expanded into the region, which again likely did not provide suitable habitat. Today, the river Nile basin with Lake Kyoga and its surrounding extensive marshy areas are still not suitable habitat for Afromontane forest, although the climate has been somewhat dryer since 6,000 years BP, facilitating more semi-

deciduous forest (Livingstone, 1975). Another reason for past absence of *P. africana* in this part of Uganda may be the relatively flat terrain, which did not allow for altitudinal range shifts during past climatic fluctuations. Within the adjacent rift valleys and mountain ranges, such range shifts might have supported the survival of populations during the unstable climatic history. The data does not support the suggestion of Muchugi *et al*. (2006) who from a study based on RAPD markers of Cameroonian, Kenyan, and Ugandan populations of *P. africana* concluded that the Eastern Rift Valley acted as a barrier to gene flow, since in our case on either side of both branches of the Rift similar chloroplast haplotypes occurred: HT1a east and west of the Eastern Rift Valley, and HT1m east and west of the Albertine Rift.

The data on the close relationship between Albertine Rift and West African populations of *P. africana* correspond with ecological comparisons between east and West African ornithological and entomological records. Already in the 1960s, Moreau (1966) suggested that montane vegetation belts were depressed to lower elevations connecting east and west African mountain regions during the last glacial period, although Livingstone (1975) doubted that such a belt existed. Based on data from palynological records from the last 15,000 years, Livingstone (1975) rather argued that forest trees were extremely rare in mountain regions and at lower altitudes during the LGM. Instead, he explained ecological affinities by the high migration capacity of birds and butterflies and suggested smaller isolated habitats to have served as stepping stones. Although *P. africana* is a typical species of Afromontane forest, it has also been reported to inhabit elevations below the montane zone (Hall *et al*.,

2000). It is difficult to conclude whether such site conditions would have allowed spread across a distance of around 2,000 km, but future genetic studies of additional wide-spread Afromontane species should help to shed more light on past montane vegetation belts.

The proposed migration paths and splits were well supported by both the genealogical network and the Bayesian simulations of coalescence. In general, the Bayesian simulations provided a broad estimate of splitting times for events. There is for example considerable overlap between the three oldest splitting events (Figure 4.3). Although the uncertainties in estimating generation time, population size and mutation rate limit the accuracy of the obtained splitting times, they elucidate the sequence of events. For example, the two most recent splits are sharply different from the older ones. Using the range of generation times given in the literature (Hall *et al.*, 2000), the time of the population splits vary by a factor of 2 to 4. The overall time frame of the simulations shows that the complete colonisation of the investigated distribution of *P. africana* has most likely occurred within the last 100,000 to 180,000 years. On a global scale this time period is known as the last glaciation period suggesting that the colonisation of eastern and western Africa by *P. africana* occurred mainly during colder climatic episodes.

5.1.3 Population structure and history of *Prunus africana* as revealed by nuclear microsatellites (nSSR)

While current patterns of genetic diversity in temperate northern regions have been determined mainly by recolonization after the LGM, patterns are more complicated in tropical areas that have not suffered total species disappearance during glaciations. Rather, the patterns existing today are the result of several cycles of climatic change (Fjeldsa and Lovett, 1997). The current results indicate that – as suggested by White (1983) – the population history of *P. africana* and the Afromontane forests in general is long and complicated and probably started in the early Pleistocene (2.5 - 0.012 million yr BP). The biogeography of the species thus is multifaceted and has been determined by rare long-distance dispersal events coupled with constant migration at intermediate geographic ranges, and strong gene flow barriers. The observed patterns corroborate the main result of the cpDNA study (section 5.1.2 & 5.1.3; Kadu *et al*., 2011). Comparing nuclear data with the cpDNA study, evidence for the dislocation of a historic immigration barrier to a more recent barrier to gene flow over several hundred kilometres is found, exemplifying the highly dynamic environmental history of African highlands.

5.1.3.1 Genetic diversity and population differentiation

Within population genetic diversity (H_e) in *P. africana* was comparable to other hardwood species, e.g. *Fraxinus excelsior* (Heuertz *et al*., 2004), or other tropical species, e.g. *Milicia excelsa* (Daïnou *et al*., 2010). Genetic diversity differed among regions. Allelic richness was highest in populations of Kenya

and Uganda west of the Eastern Rift. In these montane forests *P. africana* appears to be in its optimal habitat as it forms a main component of Afromontane forests there (White, 1983). Levels of genetic diversity were lowest in populations of southern Africa and Madagascar. Population bottlenecks were detected in populations of southern Africa, Kenya, Uganda, and Madagascar; these were probably anthropogenic as severe deforestation and changes in land use starting ca. 1000-2000 years BP have had dramatic impact on Afromontane ecosystems (Gade, 1996; Finch *et al*., 2009).

Population differentiation with a global F_{ST} of 0.27 and R_{ST} of 0.70 was high, reflecting the large and disjunct distribution area of the species and the strong divergence among populations even within regions. This was highlighted by the findings of the AMOVA (based on F_{ST}), as 14.4 % of the variation was partitioned among regions and 12.6 % among populations within regions. A very high value compared to other tree species (Petit *et al*., 2005). Stepwise-mutations had as strong effect on population differentiation as shown by R_{ST} >> F_{ST}, indicating long-term isolation of groups. Also, the frequent occurrence of private alleles and null alleles within regional samples underline the strong divergence of populations. Null alleles might have affected the consecutive analysis; however, the inference made is believed to be based on strong evidence, and is corroborated by previous findings based on cpDNA markers (section 5.1.2; Kadu *et al*., 2011).

Five main groups of populations were resolved in the analysis corresponding to main biogeographic regions: Afromontane forests of Western Africa,

Eastern Africa west of the Eastern Rift Valley, Eastern Africa east of the Eastern Rift Valley, Southern Africa and Madagascar. This grouping was supported by the UPGMA distance tree, PCA and the results of individual population assignment. A similar pattern had been obtained for *P. africana* from RAPD data by Dawson and Powell (1999), but the relationship between WG and EG1 derived in that study was even closer; thus gene flow between WG and EG1 is shown to have been disrupted for a relatively short period, but has allowed populations to diverge.

5.1.3.2 Recent patterns of migration

Estimates of recent migration rates between populations as detected in the BAYESASS analysis showed that gene flow between populations has recently been largely restricted in the Western and Madagascan groups. Given the island nature of Madagascar the result is not surprising there, while in West Africa gene flow was restricted to the island of Bioko (GQ; GQ1) and CB (CM1), the nearest population from the mainland, and appeared disrupted among the other mainland populations. A different pattern was observed in East Africa, and in particular in EG1: here gene flow was estimated to contribute to genetic diversity in most populations to a higher degree. On the other hand, Kenyan populations east and close to the Eastern Rift Valley appeared to be less connected (0.88; actually, the Eastern Rift was identified as the second most important barrier for gene flow throughout the range of *P. africana*), though some populations especially from the Eastern Arc appeared to be better connected and also clearly were the original populations for the

populations sampled in southern Africa, which were similar in self-recruitment. These findings are in general agreement with the estimates of population divergence obtained from cpDNA (section 5.1.2; Kadu *et al*., 2011): western and eastern lineages diverged early, and the population migration from EG1 to West Africa appears to be a relatively recent event, while east of the Rift effects of fragmentation are more evident and appear to be older as evidenced by the fixation of unrelated cp haplotypes in spatially close populations (section 5.1.2; Kadu *et al*., 2011). The better connectivity among Eastern Arc populations is probably associated to the more stable climate in that region (Hamilton and Taylor, 1991; Finch *et al*., 2009).

5.1.3.3 The Rift valley as major gene flow barrier – major shift of a gene flow barrier in East Africa

The situation in East Africa is remarkable in particular. While only one group was supported in the other main geographic regions, East Africa holds two major groups; also, the most evident genetic barrier on the mainland were detected in this region. This result is in line with previous findings in the species of Dawson and Powell (1999) and Muchugi *et al*., (2006). Historic patterns of direction and magnitude of migration among regions and effective population size over the coalescent period ($4N_e$ generations) as investigated using MIGRATE (Beerli and Felsenstein, 2001), corroborated some findings from the other methods and provided important new insight into population history. In contrast to estimates of recent gene flow, historic estimates indicate low levels of gene flow between EG1 and EG2, giving evidence that during an

(putatively colder) earlier period of the Pleistocene both groups exchanged migrants. At the same time, all analyses (chord distance tree, assignment, migration) support the strong divergence between western and eastern genetic groups. Ancestral population size appears to have also been large in populations of East Africa (both sides of the Eastern Rift), providing strong evidence that this region was central for population differentiation and spread.

The divergent climatic conditions between East and West Africa probably best explain the observed pattern: during the Pleistocene in West Africa the montane vegetation descended to lowland areas - in particular as riverine gallery forests - several times and thus probably allowed ample gene flow between populations (Hamilton and Taylor, 1991). Accumulated evidence also suggests that montane forests in the Albertine Rift had more constant climatic conditions; e.g. Marchant *et al*., (1997), Marchant and Taylor, (1998), Jolly *et al*., (1997) report that Afromontane forests were more or less stable in south-west Uganda from at least 42,000 years until ca. 1000 years BP (start of human influence). In contrast, the climate east of the Eastern Rift Valley was generally dryer and more variable at least during the last 10,000 years, and montane vegetation became sparse several times as exemplified by pollen records (reviewed in White, 1983; Street-Perrott and Perrott, 1993; Elenga *et al*., 2000), thus probably survived in smaller and more isolated populations which led to higher population differentiation due to drift (UPGMA tree, Figure 4.5; cpDNA, Kadu *et al*., 2011).

In the Eastern Arc Mountains of Tanzania and Kenya (populations KE8, TZ2, TZ5 and TZ6) the situation is different again, as that region is buffered against extreme drought by orographic rainfall from the Indian Ocean (Hamilton, 1982; Lovett, 1993), which may also explain the better connectivity among these populations inferred from recent gene flow estimates and the domination of few chloroplast haplotypes in the region (Kadu *et al.*, 2011). In addition to climatic factors, also the geology of the region has most likely impacted the population demography: in the Albertine Rift, volcanic activity has occurred until ca. 40,000 years BP, very likely heavily impacting the forests occurring on these volcanic mountains (Tolley *et al.*, 2011). This disturbance factor may further explain the observed pattern of relative genetic homogeneity in EG1, due to repeated re-colonization from surrounding populations.

Interestingly, in the cpDNA analysis (section 5.1.2; Kadu *et al.*, 2011), populations on the west side of the Eastern Rift shared cp haplotypes with eastern populations and not with western populations close to the Albertine Rift (and other western clade populations). In this analysis, these populations (UG4, KE1, KE2, KE3 and KE4) are shown to clearly cluster with populations east of the Albertine Rift in nuclear microsatellites. At the same time estimates of historic gene flow indicate low levels of gene flow between EG1 and EG2, which currently appears fully disrupted. The results on nuclear data therefore support a shift in gene flow barriers probably during the early to mid Pleistocene: apparently the Eastern Rift Valley has replaced the Lake Victoria basin as the main barrier to gene flow in Eastern Africa for the species since at least the mid Holocene (cf. patterns of recent gene flow). Shifts in gene flow

barriers have only been rarely reported, and this finding highlights the dynamic environmental history of the continent and the evolutionary impact it has had on shaping species' genetic diversity.

In respect to population history, there are two possible scenarios to explain the observed pattern: (i) population divergence between EG1 and EG2 took place relatively recently, and these populations acted as "genetic bridges" for the movement from east to west; consecutively gene flow was fully disrupted east and west of the Eastern Rift Valley and the populations evolved separately; (ii) divergence could be older and the estimated migration between groups could reflect a more recent gene flow event, i.e. the observed pattern could support a scenario that during a colder period, when Afromontane forests moved down to lower altitudes (Jolly *et al*., 1997), gene flow by pollen was restricted to occur only between populations on the western side of the Eastern Rift that were actually descendants from populations (cp haplotype) east of the Eastern Rift, and consecutively were becoming isolated from EG2 gene flow due to the increasing aridity that affected the Eastern Rift during the Pleistocene (Hamilton and Taylor, 1991). The current distribution of genetic diversity and the strong divergence between western and eastern groups supports the latter hypothesis.

On the other hand, evidence from cpDNA analysis rather supports the first hypothesis, since the pattern of cp diversity in the western groups indicates a relatively recent spread of the species with a single haplotype dominant both in WG and EG1 indicating low founding population size and high impact of drift

effects; though cp haplotypes currently dominating in WG and EG1 are not closely related to haplotypes found in western Kenya/ eastern Uganda. Further intense sampling is strongly suggested in the Rift Valley region and in particular from Ethiopia - putatively the original region from where the species migrated into Africa (section 5.1.2 and 5.1.3; Kadu *et al*., 2011) - to finally resolve the group relationships.

The importance of the Rift Valley system as gene flow and dispersal barrier in the species deserves special attention: apparently relatively recent gene flow was possible across distances of more than 2000 km, but not across the 100 km of savannah separating the mountainous areas east and west of the Eastern Rift. The ecological reasons for this are intriguing, but are still mostly in the dark. The Eastern Rift valley was found to be a major gene flow barrier in several other Afromontane species, e.g. *Arabis alpina* (Assefa *et al*., 2007), *Lobelia gibberoa* (Kebede *et al*., 2007), *Hagenia abyssinica* (Ayele *et al*., 2009), *Acacia senegal* (Omondi *et al*., 2010), as well as a number of other organisms (e.g., Nicolas *et al*., 2008; and references therein). Strong population divergence between populations associated with the Eastern Rift Valley was also observed in a recent study on *Senegalia mellifera*, a species of the savannah (Ruiz Guajardo *et al*., 2010). There the striking breaks in allelic patterns were attributed to selection due to site (soil) conditions. Trauth *et al*., (2005) report the existence of deep lakes in the Eastern Rift between 1.0 to 0.9 million years BP, which could also have acted as gene flow barriers for *P. africana* affecting both pollen and seed dispersal by animals.

The most likely explanation, however, are adaptations to climate (precipitation) that may be more important in the moisture demanding *P. africana*. The observed discordance between chloroplast and nuclear variation further strengthens the indication of a selective component determining group distribution, as nuclear variation reflects site adaptation better (cf. Petit *et al*., 2001; Tsitrone *et al*., 2003), and obviously was influenced by gene flow to a lesser extent than cytoplasmic variation.

5.1.3.4 *Prunus africana* populations in Madagascar

Madagascar holds the most diverged populations and as suggested by the cpDNA analysis (section 5.1.2; Kadu *et al*., 2011) probably the colonization of that island was the result of a single dispersal event that dates back to the oldest splitting events within the species. This finding is in agreement with previous studies showing very low rates of propagules exchange between Madagascar and mainland Africa (Meve and Liede, 2002). The high estimated ancestral population size also indicates long time isolation and independent evolution for the population. The long-time climatic stability of this large island has allowed undisturbed evolution and accumulation of a rich flora (Vences *et al*., 2009). Madagascar must have been colonized very early by the species and gene flow with mainland populations has most likely been very low to non-existent. Further, the colonisation of Madagascar must have been independent from the colonization of southern Africa, as indicated by the lowest *M* estimate (Table 4.13) between these two groups. Sampling from the Comoros Islands would be important to evaluate possible gene flow routes

between mainland Africa and Madagascar. The evidence for recent population decline due to recent fragmentation (as exemplified by the detected bottleneck), but also morphological divergence (unpublished results) strengthens the demand for the protection of these populations.

5.1.3.5 Derived populations in southern and western Africa

The results clearly show that the Southern African populations are derived from the East African group (EG2) and that the founding population size for these populations has been low compared to the other areas. The divergence appears recent and at present probably mainly driven by drift (monomorphic loci, private alleles). The fact that samples from that region were collected over a relatively large geographic range but clearly clustered in the STRUCTURE computations indicates that populations in this region have only recently become fragmented and are probably derived from a single population expansion event. A significant barrier to gene flow within the region was only detected between populations from South Africa and Zimbabwe.

The MIGRATE analysis also sheds some light on the relationship between the East and West African populations of the species. Historic migration estimates between WG and EG1 populations west of the Eastern Rift (Uganda) were strong especially from EG1 to WG; considering both these results and patterns of diversity as well as the results from cpDNA analysis (section 5.1.2; Kadu *et al.*, 2011), colonization from east to west appears to be the most plausible hypothesis. In fact, the only other study (to my best knowledge) scrutinizing

gene flow patterns between West and East Africa is the study of Ackermann and Bishop, (2010) on gorillas, where an opposite direction of migration was inferred. Estimates of recent migration still indicate low levels of gene flow between these regions in *P. africana*. Interestingly, the most similar populations in WG were Bioko Island (GQ; GQ1) and CB (CM1), indicating that the island population is a direct descendent of CB (CM1); while other populations from that region recently appear in much less contact. Given the long-term absence of the species north of the Congolian plateau and the existence of satellite populations extant in the Congo, Zambia and Angola, the southern migration corridor (*sensu* White/ Moreau) – south of the central African rain forest region – may have served as migration corridor for this movement. In addition, aridification of Africa during the Pleistocene has driven forests south (Hamilton and Taylor, 1991), making this route of migration even more plausible. Sampling from these areas (western Tanzania, Zambia and Angola) would be needed to solve this question finally. Also, future collections of *P. africana* from riverine populations in Southern Africa and other regions appear promising.

5.1.4 Intraspecific chemical diversity

The average concentration [mg/kg w/w] of chemical constituents analysed based on 20 bulked population samples (population mix) varied considerably among and within populations with lauric acid (18) being the least and ursolic acid (743) the most. Friedelin was not detected in any of the samples above the limit of detection (7 mg/kg). This finding was quite surprising because this

triterpenoid which is known for its anti-inflammatory effects (Antonisamy *et al*., 2011) has been previously identified in *P. africana* bark extracts (Catalano *et al*., 1984) as well as in *P. lusitanica* (Biessels *et al*., 1974) and *P. turfosa* (Sainsbury, 1970). However, it is absent in *P. serotina* (Biessels *et al*., 1974). The concentrations of the other seven bark constituents varied considerably among and within populations. For both lauric and myristic acid, the content in individual samples within certain populations partly varied erratically. Both fatty saturated acids were previously reported in *P. africana*, but each with very low relative compositions compared to other free fatty acids (Ganzera *et al*., 1999). In *P. amygdalus* (Munshi and Sukhija, 1984) and *Artocarpus heterophyllus* (Chowdhury *et al*., 1997) a similar proportion of lauric and myristic acid was reported. However, Munshi and Sukhija (1984) found that the concentration of both analytes decreased rapidly and eventually only traces were found in developing kernels of almonds. An open question is, whether genetic, environmental or even methodological factors are causing this pattern.

Content of Docosanol among populations strongly varied and differed significantly. However, it was not affected by the environmental factors studied. Data on natural variation of this long-chain alcohol are virtually not existent. In crops, highly significant genotypic differences were observed (Irmak *et al*., 2008) and Dunford and Edwards (2010) found content to be significantly affected by the environment. Content of Ferulic acid was found to be highest in Kinale (KB-Kenya) and lowest in Moka (GQ-Equatorial Guinea) with concentrations of 30 to 90 mg/kg. Content of ferulic acid was significantly correlated to tree diameter (r = 0.23, $p \leq 0.05$) and annual

precipitation (r = -0.50, $p \leq 0.025$). Mpofu et al., (2006) showed that ferulic acid concentrations in different wheat cultivars varied at sampling sites differing in environmental conditions. This hydroxycinnamic acid is found in many plants, has already been detected in *P. africana* bark (Fourneau *et al*., 1996) and is also present in leaves of different *Prunus* species (Liakopoulos, 2001). It may effectively scavenge deleterious radicals and suppress radiation-induced oxidative reactions (Kikuzaki *et al*., 2002). Although ferulic acid as a component of lignocelluloses is found in many plants, it is unexpected that its variation in natural plant population has not been studied yet. In wheat (*Triticum aestivum*) cultivars, content variation showed significant difference (Irmak *et al*., 2008), while environmental effects were considerably larger than genotypic ones (Mpofu *et al*., 2006).

Content of β-sitostenone was found to be highest in Mt. Danoua (CC-Cameroon) and lowest in Chimanimani (ZW-Zimbabwe) with concentrations of 116 to 260 mg/kg. These concentrations are much lower than in β-sitosterol (see below). Catalano *et al*., (1984) also found several fold lower concentrations of sitostenone compared to sitosterol in *P. africana* bark extracts. Beta-sitosterol was found in concentrations of 349 to 583 mg/kg. Concentration was highest in Chimanimani (ZW-Zimbabwe) and lowest in Chuka (KA-Kenya). These concentrations exceed those found in fruits and are at least similar to avocado which is known as a very rich source of β-sitosterol (Duester, 2001). Seeds of the soybean (*Glycine max*) (Yamaya *et al*., 2007), also known as an important source contain less, but seeds of rapeseed (*Brassica napus*) (Amar *et al*., 2009) contain 5-fold as much of this compound

as the *P. africana* samples. This phytosterol mostly known for its cholesterol-lowering activity (Satou, 2003) is believed also to have anti-cancer effects (Awad and Fink, 2000). *P. spinosa* (Wolbiš *et al*., 2001) as well as *Moringa oleifera* (Anwar *et al*., 2007) having a high concentration of β-sitosterol are known to be used for their diuretic properties in traditional medicine to increase urine flow.

To the best of my knowledge data on genetic variation of β-sitosterol content in natural plants is not at hand, however, both in *Glycine max* (Yamaya *et al*., 2007) and *Brassica napus* (Amar *et al*., 2009) genotypic variation was high. Content of β-sitosterol was independent of environmental conditions. This is contrasted by findings in *Glycine max*. Varieties grown in warmer areas produced seeds with higher sitosterol content than at colder sites (Yamaya *et al*., 2007), however, with the *P. africana* samples no environmental effect on the concentration was found.

Among all compounds, the content of ursolic acid was found to be the highest with concentrations varying between 317 and 2000 mg/kg. This natural pentacyclic triterpenoid carboxylic acid has been identified as the major component in many traditional medicinal plants (for a recent review see Ikeda *et al*., 2008) and serves as a basic material for synthesis of more potent bioactive derivatives, such as anti-tumor agents (Ma *et al*., 2005). It is found in many plants. For instance, it was reported in the peels of the loquat (*Eriobytrya japonica*) fruits in comparable concentration as in the present study. The concentration of ursolic acid was negatively correlated with the

diameter of the trees (*r* = -0.33, *p*≤ 0.01). This finding is to be expected because tree diameter is a good proxy for age and many triterpenoids, including ursolic acid act as plant defences (Arnason and Bernards, 2010) whose concentrations are likely to decrease with age as a result of reduced pressure from herbivores. In *Betula pendula*, the concentrations of triterpenoids declined roughly one hundred-fold from 1-year to 20-year old trees demonstrating the antibrowsing effect of triterpenoids (Laitinen, 2003).

In order to facilitate efficient provenance selection or breeding it is essential to know how much of the variation is found among and within populations. For three constituents, ferulic acid, n-docosanol and β-sitostenone, the variation could be equally attributed to the within and among components (Figure 4.11), whilst the among population component for β-sitosterol amounted to approximately 20 %. The variation with highest among population component, with roughly two thirds, was attributed to ursolic acid. The high variation among populations and the lacking effects of environment on concentration of ursolic acid bespeak population genetic effects. Similarly to what was detected in this study, the content of this component was found to vary considerably among cultivars of *Eriobotrya japonica* (Zhou *et al*., 2011).

Concentration of different bark extracts were significantly intercorrelated (Table 4.14). For certain constituent pairs a close correlation was to be expected. Beta-sitostenone is an oxidized product of β-sitosterol and both saturated fatty acids share a very similar chemical structure. Interestingly, the content of n-docosanol was significantly correlated with all other tested agents.

Thus, the concentration of n-docosanol in bark extracts can be used as a good predictor of all agents studied. It is noteworthy that the content of n-docosanol was negatively correlated with lauric and myristic acids, although the data of both fatty acids must be considered with care (see above).

None of the molecular markers used in this study encode any of studied bark constituents. However, haplotypic just as nuclear genetic diversity are a good indicator of genetic drift. This evolutionary force affects the whole genome equally and is only dependent on the gene frequencies and population size. Genes associated with the direct or indirect syntheses of the antioxidants studied are affected by this evolutionary force as are the molecular markers used in this study. Thus, one may assume that the content variation is correlated to genetic diversity (in the chloroplast and nuclear genome) and/or to outbreeding, quantified as fixation index. Content variation of β-sitosterol of the populations was significantly correlated to haplotype diversity and ferulic acid was correlated to the fixation index F_{IS} (Table 4.15). This indicates a higher variation in ferulic acid concentration in populations with an increasing excess of homozygotes compared to HWE. This could be the case if several inbred lines, which differ in ferulic acid concentrations merged in the populations or the findings were due to a Wahlund-effect. Regardless of the true biological reasons, the association between measures of genetic diversity and the variation in bark concentrations must be interpreted as weak.

Nuclear and cp-DNA microsatellites have shown a strong population divergence among the Afromontane region and five main regions were

identified as demonstrated in Figure 4.14. (1) West Africa, (2) East Africa west of the Eastern Rift Valley, (3) East Africa east of the Eastern Rift Valley, (4) Southern Africa and (5) Madagascar (section 5.1.2 & 5.1.3; Kadu *et al*., 2011, Kadu *et al*., 2013). Genetic divergence between mainland African and Madagascan populations, as well as between populations east and west of the Rift Valley, was very pronounced. The population cluster based on the content differences of bark constituents was very weakly, but in certain cases significantly correlated with the molecular phylogenetic structure (Table 4.16; Figure 4.10). It is especially noteworthy that significant correlations were detected for chloroplast as well as nuclear data both for n-docosanol and ursolic acid. Moreover, also when all constituents were considered for a single Mantel test, both pairwise Φ_{PT} for chloroplast and R_{ST} for nuclear genetic differentiation used as genetic distance resulted in weak, but significant correlations.

5.2 CONCLUSIONS

In this study the phylogeography of *P. africana* was investigated as a model species to derive vegetation history of Afromontane forests. Phylogeography was found to exist within the 32 *P. africana* natural populations. Further, populations were found to be highly differentiated and colonisation dynamics revealed evidence of a former northern migration corridor.

Secondly, patterns of population genetic structure in 30 natural populations of *P. africana* within Africa were studied and distinct divergent patterns were

found to exist. Moreover, comparison of nuclear data with plastid data provided evidence for the dislocation of a historical immigration barrier to a more recent barrier to gene flow over several hundreds of kilometres, exemplifying the highly dynamic environmental history of African highlands. There were two strong barriers found be

Thirdly, the chemical diversity of phytochemical extracts of 20 *P. africana* natural populations were evaluated and evidence of intraspecific chemical diversity was demonstrated though the geographical pattern presented was not very distinct. It is concluded that the molecular phylogeographic patterns is reflected in the spatial patterns of certain bark constituents, notably in ursolic acid.

Compare genetic and phytochemical date the correlations There were weak correlations

5.3 RECOMMENDATIONS

The following recommendations are suggested:

i. Investigation of further aspects in colonisation dynamics requiring sampling in the Ethiopian highlands and in the south-western/ central African range. More data from populations sampled south of the Albertine Rift and following the SMT would be needed to resolve the possibility of migration route via SMT and verify lack of lineage mixing. Similarly, sampling of *P. africana* riverine populations appear promising. Sampling

from the Comoros Island would be important for evaluating possible gene-flow routes from mainland Africa to Madagascar.

ii. Further intense sampling is strongly suggested in the Rift Valley region and in particular from Ethiopia - putatively the original region from where the species migrated into Africa, to finally resolve the EG1 and EG2 group relationships.

iii. To connect extant populations, planting of stepping stone populations is suggested, which could be achieved within the framework of a scheme to support the livelihood of the rural population in these countries, as the demand for bark is rising and sustainable use needs to be established.

iv. For *in situ* conservation, populations with high estimates of allelic richness and population size should be given priority. The divergence among the five detected groups putatively has led to site adaptation which needs to be conserved. *Ex-situ* efforts need to take care not to mix populations east and west of the Eastern Rift Valley, to avoid possible negative effects of genomic divergence, e.g. due to outbreeding depression or disruption of local selection-adaptation balance (Verhoeven *et al*., 2011). Data on spatio-chemical diversity are expedient for a conservation strategy for African cherry.

v. A careful analysis using GIS based methods is needed to identify regions of urgent need for protection, which have not been covered by current conservation efforts.

vi. Investigations on this and other species with disjunct distribution in West and East Africa (and the Rift Valley in particular) as well as other parts of the continent are strongly encouraged to further elucidate the underlying

processes. Promising candidates are suggested by (White, 1981), and include *Myrica salicifolia*, *Nuxia congesta*, and *Podocarpus latifolius*.

vii. If a conservation strategy focuses exclusively on molecular diversity, populations in Kenya, Uganda and Tanzania will be prime regions since Madagascan populations are not quite as diverse as the mainland populations from East Africa. However, populations from Madagascar feature a relatively high concentration of all bark constituents (Figure 4.12) and, although it is unknown which role bark constituents may play for adaptation, it is highly recommended to include Madagascan populations into a pan-African conservation strategy. These island populations appear also to be highly vulnerable due to bark overexploitation.

viii. Further investigations are needed to cast more light on the environmental factors affecting bark quality. To fully understand opportunities for genetic gains through selection of populations and individuals, genetic field tests are necessary.

REFERENCES

Ackermann, R. R., & Bishop, J. M. (2010). Morphological and molecular evidence reveals recent hybridization between gorilla taxa. *Evolution: International Journal of Organic Evolution*, ***64***(1), 271–290. http://www.ncbi.nlm.nih.gov/pubmed/19804402

Afzal-Rafii, Z., & Dodd, R. S. (2007). Chloroplast DNA supports a hypothesis of glacial refugia over postglacial recolonization in disjunct populations of black pine (*Pinus nigra*) in Western Europe. *Molecular Ecology*, ***16***(4), 723–736. https://doi.org/10.1111/j.1365-294X.2006.03183.x

Agapow, P.-M., & Burt, A. (2001). Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes*, ***1***(1–2), 101–102. https://doi.org/10.1046/j.1471-8278.2000.00014.x

Amar, S., Becker, H. C., & Möllers, C. (2009). Genetic variation in phytosterol content of winter rapeseed (*Brassica napus* L.) and development of NIRS calibration equations. *Plant Breeding*, ***128***(1), 78–83. https://doi.org/10.1111/j.1439-0523.2008.01531.x

Angioi, S. A., Desiderio, F., Rau, D., Bitocchi, E., Attene, G., & Papa, R. (2009). Development and use of chloroplast microsatellites in *Phaseolus* spp. and other legumes. *Plant Biology*, ***11***(4), 598–612. http://dx.doi.org/10.1111/j.1438-8677.2008.00143.x

Antonisamy, P., Duraipandiyan, V., & Ignacimuthu, S. (2011). Anti-inflammatory, analgesic and antipyretic effects of friedelin isolated from *Azima tetracantha* Lam. in mouse and rat models. *The Journal of Pharmacy and Pharmacology*, ***63***(8), 1070–1077. https://doi.org/10.1111/j.2042-7158.2011.01300.x

Anwar, F., Latif, S., Ashraf, M., & Gilani, A. H. (2007). *Moringa oleifera*: A Food Plant with Multiple Medicinal Uses. *Phytotherapy Research*, ***21***(1), 17–25. https://doi.org/10.1002/ptr

Arnason, J. T., & Bernards, M. A. (2010). Impact of constitutive plant natural products on herbivores and pathogens. *Canadian Journal of Zoology*, ***88***(7), 615–627. https://doi.org/10.1139/Z10-038

Assefa, A., Ehrich, D., Taberlet, P., Nemomissa, S., & Brochmann, C. (2007). Pleistocene colonization of afro-alpine “sky islands” by the arctic-alpine *Arabis alpina*. *Heredity*, ***99***(2), 133–142. https://doi.org/10.1038/sj.hdy.6800974

Aubréville, A. (1976). Centres tertiaires d’origine, radiations et migrations des flores angiospermiques tropicales. *Adansonia, Serie 2.*, ***16***, 297–354.

Austin, J. D., Lougheed, S. C., & Boag, P. T. (2004). Controlling for the Effects of History and Nonequilibrium Conditions in Gene Flow Estimates in Northern

Bullfrog *(Rana catesbeiana) Populations*. *Genetics* **168**, 1491–1506. https://doi.org/10.1534/genetics.104.027987

Awad, A. B., & Fink, C. S. (2000). Phytosterols as Anticancer Dietary Components: Evidence and Mechanism of Action. *Journal of Nutrition*, ***130***, 2127–2130.

Ayele, T. B., Gailing, O., Umer, M., & Finkeldey, R. (2009). Chloroplast DNA haplotype diversity and postglacial recolonization of *Hagenia abyssinica* (Bruce) J. F. Gmel. in Ethiopia. *Plant Systematics and Evolution*. **280**, 175 – 185 https://doi.org/10.1007/s00606-009-0177-5

Bänfer, G., Moog, U. T. E., Fiala, B., Mohamed, M., Weising, K., & Blattner, F. R. (2006). A chloroplast genealogy of myrmecophytic *Macaranga* species (Euphorbiaceae) in Southeast Asia reveals hybridization, vicariance and long-distance dispersals. *Molecular Ecology*, ***15***(14), 4409–4424. http://dx.doi.org/10.1111/j.1365-294X.2006.03064.x

Barker, N. P., Cunningham, A. B., Morrow, C., Harley, E. H., & Huntley, B. J. (1994). A preliminary investigation into the use of RAPD to assess the genetic diversity of a threatened African tree species: *Prunus africana*. *Strelitzia*, ***1***, 221–230.

Beerli, P. (2010). *Migrate version 3.2 documentation* (Available from http://popgen.sc.fsu.edu/Migrate/Migrate-n.html). http://popgen.sc.fsu.edu/Migrate/Migrate-n.html

Beerli, P., & Felsenstein, J. (2001). Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences of the United States of America*, ***98***(8), 4563–4568. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=31874&tool=pmcentrez&rendertype=abstract

Beheregaray, L. B. (2008). Twenty years of phylogeography: the state of the field and the challenges for the Southern Hemisphere. *Molecular Ecology*, ***17***(17), 3754–3774. http://www.ncbi.nlm.nih.gov/pubmed/18627447

Bell, C. J., & Ecker, J. R. (1994). Assignment of 30 Microsatellite Loci to the Linkage Map of *Arabidopsis*. *Genomics*, ***19***(1), 137–144.

Biessels, H. W. A., van der Kerk-van Hoof, A. C., Kettenes-van den Bosch, J. J., & Salemink, C. A. (1974). Triterpenes of *Prunus serotina* and *Prunus lusitanica*. *Phytochemistry*, ***13***, 203–207.

Birky, C. W., Fuerst, P., & Maruyamat, T. (1989). Organelle Gene Diversity Under Migration, Mutation, and Drift: Equilibrium Expectations, Approach. *Genetics*, ***121***, 613–627.

Bohm, B. A. (2009). The Geography of Phytochemical Races. In *The Geography of Phytochemical Races.: Vol. Bohm, B.A.* (pp. XVIII, 350 p). Springer, Heidelberg.

Bombardelli, E., & Morazzani, P. (1997). *Prunus africana* (Hook. f.) Kalkm. *Fitoterapia*, ***68***, 205–218.

Bonnin, I., Huguet, T., Gherardi, M., Prosperi, J. M., & Olivieri, I. (1996). High level of polymorphism and spatial structure in a selfing plant species, *Medicago truncatula* (Leguminosae), shown using RAPD markers. *American Journal of Botany*, ***83***(7), 843–855.

Botstein, D., Whitew, R. L., Skolnick, M., & Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphism. *American Journal of Human Genetics*, ***32***, 314–331.

Bradshaw, J. E., Hackett, C. A., Meyer, R. C., Milbourne, D., McNicol, J. W., Phillips, M. S., & Waugh, R. (1998). Identification of AFLP and SSR markers associated with quantitative resistance to Globodera pallida (Stone) in tetraploid potato (*Solanum tuberosum subsp. tuberosum*) with a view to marker-assisted selection. *Theoretical and Applied Genetics*, ***97***(1), 202–210. https://doi.org/10.1007/s001220050886

Broun, P., & Tanksley, S. (1996). Characterization and genetic mapping of simple repeat sequences in the tomato genome. *Molecular and General Genetics*, ***250***(1), 39–49. https://doi.org/10.1007/BF02191823

Brown, A. D. H. (1979). Enzymes polymorphisms in plant populations. *Theoretical Population Biology*, ***15***, 1–42.

Butynski, T. M., & Koster, S. H. (1994). Distribution and Conservation Status of Primates in Bioko Island, Equatorial-Guinea. *Biodiversity & Conservation*, ***3***(9), 893–909.

Cantini, C., Legnose, S., Ponte, V., Iezzoni, A. F., & Lamboy, W. F. (2001). DNA Fingerprinting of Tetraploid Cherry Germplasm Using Simple Sequence Repeats. *Journal of the American Society for Horticultural Science*, ***126***(2), 205–209.

Caratini, C., & Giresse, P. (1979). Contribution palynologique a la connaissance d'environnements continentaux et marins du Congo a la fin du Quaternaire. *CR Acad Sc Paris*, ***288***, 379–382.

Carbin, B. E., Larsson, B., & Lindahl, O. (1990). Treatment of benign prostatic hyperplasia with phytosterols. *British Journal of Urology*, ***66***(6), 639–641. http://www.ncbi.nlm.nih.gov/pubmed/1702340

Castañeda, I. S., Werne, J. P., Johnson, T. C., & Filley, T. R. (2009). Late Quaternary vegetation history of southeast Africa: The molecular isotopic record from Lake Malawi. *Palaeogeography, Palaeoclimatology,*

Palaeoecology, ***275***(1–4), 100–112. https://doi.org/10.1016/j.palaeo.2009.02.008

Catalano, S., Ferretti, M., Marsili, A., & Morelli, I. (1984). New Constituents of *Prunus africana* Bark Extract. *Journal of Natural Products*, ***47***(5), 910.

Cavalli-Sforza, L. L., & Edwards, A. W. F. (1967). Phylogenetic analysis. Models and estimation procedures. *The American Journal of Human Genetics*, ***19***(3 Pt 1), 233–257. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1706274&tool=pmcentrez&rendertype=abstract

Cavers, S., Navarro, C. M., & Lowe, A. J. (2004). Targeting genetic resource conservation in widespread species: a case study of *Cedrela odorata* L. *Forest Ecology and Management*, ***197***(1–3), 285–294. https://doi.org/10.1016/j.foreco.2004.05.019

Chen, X., Temnykh, S., Xu, Y., Cho, Y. G., & McCouch, S. R. (1997). Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics*, ***95***(4), 553–567. https://doi.org/10.1007/s001220050596

Chester, M., Cowan, R. S., Fls, M. F. F., & Rich, T. C. G. (2007). Parentage of endemic *Sorbus* L. (Rosaceae) species in the British Isles: evidence from plastid DNA. *Botanical Journal of the Linnean Society*, ***154***, 291–304.

Chowdhury, F. A., Raman, A. Md., & Mian, A. J. (1997). Distribution of free sugars and fatty acids in jackfruit (*Artocarpus heterophyllus*). *Food Chemistry*, ***60***(1), 25–28. https://doi.org/10.1016/S0308-8146(96)00294-4

Cipriani, G., Lot, G., Huang, W. G., Marrazzo, M. T., Peterlunger, E., & Testolin, R. (1999). AC/GT and AG/CT microsatellite repeats in peach [*Prunus persica* (L) Batsch]: isolation, characterisation and cross-species amplification in *Prunus*. *Theoretical and Applied Genetics*, ***99***(1–2), 65–72. https://doi.org/10.1007/s001220051209

Clement, M., Posada, D., & Crandall, K. A. (2000). TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, ***9***(10), 1657–1659. http://dx.doi.org/10.1046/j.1365-294x.2000.01020.x

Colinvaux, P. A., Oliveira, P. E. de, & Bush, M. B. (2000). Amazonian and neotropical plant communities on glacial time-scales: The failure of the aridity and refuge hypotheses. Quaternary Science Reviews, ***19***, 141 - 169.

Cregan, P. B., Jarvik, T., Bush, A. L., Shoemaker, R. C., Lark, K. G., Kahler, A. L., Kaya, N., VanToai, T. T., Lohnes, D. G., Chung, J., & Specht, J. E. (1999a). An integrated genetic linkage map of the soybean genome. *Crop Science*, ***39***(5), 1464–1490.

Cregan, P. B., Mudge, J., Fickus, E. W., Marek, L. F., Danesh, D., Denny, R., Shoemaker, R. C., Matthews, B. F., Jarvik, T., & Young, N. D. (1999b). Targeted isolation of simple sequence repeat markers through the use of bacterial artificial chromosomes. *Theoretical and Applied Genetics*, ***98***(6–7), 919–928.

Cunningham, A. B., & Mbenkum, F. T. (1993). Sustainability of harvesting *Prunus africana* bark in Cameroon: a medicinal plant in international trade. In *People and plants working paper 2*. UNESCO, Paris.

Cunningham, M., Cunningham, A. B., & Schippmann, U. (1997). *Trade in Prunus africana and the implementation of CITES*. Bundesamt fur Naturschutz, Bonn, Germany.

Daïnou, K., Bizoux, J.-P., Doucet, J.-L., Mahy, G., Hardy, O. J., & Heuertz, M. (2010). Forest refugia revisited: nSSRs and cpDNA sequences support historical isolation in a wide-spread African tree with high colonization capacity, *Milicia excelsa* (Moraceae). *Molecular Ecology*, ***19***(20), 4462–4477. https://doi.org/10.1111/j.1365-294X.2010.04831.x

Dawson, I. K., Chalmers, K. J., Waugh, R., & Powell, W. (1993). Detection and analysis of genetic variation in *Hordeum spontaneum* populations from Israel using RAPD markers. *Molecular Ecology*, ***2***(3), 151–159. www.ncbi.nlm.nih.gov

Dawson, I. K., & Powell, W. (1999). Genetic variation in the Afromontane tree *Prunus africana*, an endangered medicinal species. *Molecular Ecology*, ***8***(1), 151–156. https://doi.org/10.1046/j.1365-294X.1999.00518.x

Dawson, I. K., Were, J., & Lengkeek, A. G. (2000). Conservation of *Prunus africana*, an over-exploited African medicinal tree. *Forest Genetic Resources*, ***28***(28), 27–33.

Debout, G. D. G., Doucet, J.-L., & Hardy, O. J. (2011). Population history and gene dispersal inferred from spatial genetic structure of a Central African timber tree, *Distemonanthus benthamianus* (Caesalpinioideae). *Heredity*, ***106***(1), 88–99. http://www.ncbi.nlm.nih.gov/pubmed/20389306

Delaunay, B. (1934). Sur la sphère vide. Bulletin of the Academy of Sciences of the U.S.S.R. In *Classe des Sciences Mathématiques et Naturelle* (pp. 793–800).

Demesure, B., Comps, B., & Petit, R. J. (1996). Chloroplast DNA phylogeography of the common beech (*Fagus sylvatica* L.) in Europe. *Evolution*, ***50***, 2515–2510.

Dieringer, D., & Schlötterer, C. (2003). MICROSATELLITE ANALYSER (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes*, 2002–2004. https://doi.org/10.1046/j.1471-8286

Dirlewanger, E., Cosson, P., Howad, W., Capdeville, G., Bosselut, N., Claverie, M., Voisin, R., Poizat, C., Lafargue, B., Baron, O., Laigret, F., Kleinhentz, M., Arús, P., & Esmenjaud, D. (2004). Microsatellite genetic linkage maps of myrobalan plum and an almond-peach hybrid—location of root-knot nematode resistance genes. *Theoretical and Applied Genetics*, ***109***(4), 827–838. https://doi.org/10.1007/s00122-004-1694-9

Diwan, N., Bouton, J. H., Kochert, G., & Cregan, P. B. (2000). Mapping of simple sequence repeat (SSR) DNA markers in diploid and tetraploid *alfalfa*. *Theoretical and Applied Genetics*, ***101***(1), 165–172.

Donovan, J. L., Meyer, A. S., & Waterhouse, A. L. (1998). Phenolic Composition and Antioxidant Activity of Prunes and Prune Juice (*Prunus domestica*). *Journal of Agricultural and Food Chemistry*, ***46***(4), 1247–1252.

Dow, B. D. (1998). High levels of gene flow in bur oak revealed by paternity analysis using microsatellites. *Journal of Heredity*, ***89***(1), 62–70. https://doi.org/10.1093/jhered/89.1.62

Dow, B. D., & Ashley, M. v. (1996). Microsatellite analysis of seed dispersal and parentage of saplings in bur oak, *Quercus macrocarpa*. *Molecular Ecology*, ***5***(5), 615–627.

Duester, K. C. (2001). Avocado fruit is rich in Source of beta-sitosterol. *Journal of the American Dietetic Association*, ***101***(4), 404–405.

Dumolin-Lapègue, S., Demesure, B., Fineschi, S., le Come, V., & Petit, R. J. (1997). Phylogeographic Structure of White Oaks throughout the European Continent. *Genetics*, ***146***, 1475–1487.

Dunford, N. T., & Edwards, J. (2010). Nutritional bioactive components of wheat straw as affected by genotype and environment. *Bioresource Technology*, ***101***(1), 422–425. https://doi.org/10.1016/j.biortech.2009.08.009

Dupanloup, I., S, S., & Excoffier, L. (2002). A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology*, ***11***, 2571–2581.

Dutech, C., Maggia, L., & Joly, H. I. (2000). Chloroplast diversity in *Vouacapoua americana* (Caesalpiniaceae), a neotropical forest tree. *Molecular Ecology*, ***9***(9), 1427–1432. http://www.ncbi.nlm.nih.gov/pubmed/10972782

Edgar, A. D., Levin, R., Constantinou, C. E., & Denis, L. (2007). A Critical Review of the Pharmacology of the Plant Extract of *Pygeum africanum* in the Treatment of LUTS. *Neurology and Urodynamics*, ***26***, 458–463. https://doi.org/10.1002/nau

Eggeling, W. J. (1940). Budongo – an East African mahogany forest. *Empire Forestry Journal*, ***19***, 179–196.

Ehrich, D., Gaudeul, M., Assefa, A., Koch, M. A., Mummenhoff, K., Nemomissa, S., & Brochmann, C. (2007). Genetic consequences of Pleistocene range shifts: contrast between the Arctic, the Alps and the East African mountains. *Molecular Ecology*, ***16***(12), 2542–2559. http://www.ncbi.nlm.nih.gov/pubmed/17561912

Elenga, H. (1987). *Les Plateaux Batéké (Congo). Paléoenvironnements Quaternaries d'après l'étude du sondage du bois de Bilanko*. Université d'Aix-Marseille II.

Elenga, H., Peyron, O., Bonnefille, R., Jolly, D., Cheddadi, R., Guiot, J., Andrieu, V., Bottema, S., Buchet, G., de Beaulieu, J. L., Hamilton, A. C., Maley, J., Marchant, R., Perez-Obiol, R., Reille, M., Riollet, G., Scott, L., Straka, H., Taylor, D., … Jonson, H. (2000). Pollen-based biome reconstruction for southern Europe and Africa 18,000 yr bp. *Journal of Biogeography*, ***27***(3), 621–634. https://doi.org/10.1046/j.1365-2699.2000.00430.x

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, ***14***(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Evans, R. C., & Dickinson, T. a. (1999). Floral ontogeny and morphology in sub-family Amygdaloideae T. & G. (Rosaceae). *International Journal of Plant Sciences*, ***160***, 955–979.

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, ***10***(3), 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of Molecular Variance Inferred mfrom Metric Distances Among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data. *Genetics*, ***131***(2), 479–491.

Falush, D., Stephens, M., & Pritchard, J. K. (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes*, ***7***(4), 574–578. https://doi.org/10.1111/j.1471-8286.2007.01758.x

Farwig, N., Böhning-Gaese, K., & Bleher, B. (2006). Enhanced seed dispersal of *Prunus africana* in fragmented and disturbed forests? *Oecologia*, ***147***(2), 238–252. https://doi.org/10.1007/s00442-005-0288-9

Farwig, N., Braun, C., & Böhning-Gaese, K. (2008). Human disturbance reduces genetic diversity of an endangered tropical tree, *Prunus africana* (Rosaceae). *Conservation Genetics*, ***9***(2), 317–326. https://doi.org/10.1007/s10592-007-9343-x

Felsenstein, J. (1989). PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics*, ***5***(5), 164–166. http://www.citeulike.org/group/894/article/2344765

Ferretti, G., Bacchetti, T., Belleggia, A., & Neri, D. (2010). Cherry antioxidants: from farm to table. *Molecules*, ***15***(10), 6993–7005. https://doi.org/10.3390/molecules15106993

Finch, J., Leng, M. J., & Marchant, R. (2009). Late Quaternary vegetation dynamics in a biodiversity hotspot, the Uluguru Mountains of Tanzania. *Quaternary Research*, ***72***(1), 111–122. https://doi.org/10.1016/j.yqres.2009.02.005

Fjeldsa, J., & Lovett, J. C. (1997). Geographical patterns of old and young species in African forest biota: the significance of specific montane areas as evolutionary centres. *Biodiversity and Conservation*, ***6***(3), 325–346. https://doi.org/10.1023/A:1018356506390

Fourneau, C., Hocquemiller, R., & Cave, A. (1996). Triterpenes from *Prunus africana* bark. *Phytochemistry*, ***42***(5), 1387–1389. https://doi.org/10.1016/0031-9422(96)00038-6

Franzel, S., Ayuk, E., Cunningham, A., Asanga, C., & Duguma, B. (2009). Bark for sale: the adoption potential of *Prunus africana* as an agroforestry tree for small-scale farmers in Cameroon. In A. B. Cunningham, B. M. Campbell, & M. Luckert (Eds.), *Bark: use, management and commerce in Africa Winter 2009, vol. 17, Advances in Economic Botany.* (pp. 408–417). New York Botanical Garden.

Gade, D. (1996). Deforestation and its effects in Highland Madagascar. *Mountain Research and Development*, ***16***(2), 101–116. http://www.jstor.org/stable/10.2307/3674005

Ganzera, M., Croom Jr., E. M., & Khan, I. A. (1999). Determination of the Fatty Acid Content of Pumpkin Seed, Pygeum, and Saw Palmetto. *Journal of Medicinal Food*, ***2***(1), 21–27.

Geldenhuys, C. J. (1981). *Prunus africana* in the Bloukrans River George Southern Cape. *South African Forestry Journal*, ***118***, 61 – 66.

Gottlieb, L. D. (1977). Electrophoretic evidence and plant systematics. *Annals of Missouri Botanical Gardens*, ***64***, 161–180.

Goudet, J. (1995). FSTAT version 2.9.3.2. A computer software to calculate F-statistics. *Journal of Heredity*, ***86***, 485–486.

Guilford, P., Prakash, S., Zhu, J. M., Rikkerink, E., Gardiner, S., Bassett, H., & Forster, R. (1997). Microsatellites in Malus × domestica (apple): abundance, polymorphism and cultivar identification. *Theoretical and Applied Genetics*, ***94***(2), 249–254. https://doi.org/10.1007/s001220050407

Hale, M., Borland, A. M., Gustafsson, M. G., & Wolff, K. (2004). Causes of Size Homoplasy among Chloroplast Microsatellites in Closely Related Clusia Species. *Journal of Molecular Evolution*, ***58***(2), 182–190. https://doi.org/10.1007/s00239-003-2540-4

Hall, J. B. (1973). Vegetational zones on the southern slopes of Mount Cameroon. *Vegetatio*, ***27***, 49–69.

Hall, J. B., O'Brien, E. M., & Sinclair, F. L. (2000). *Prunus africana - a monograph*. School of Agricultural and Forest Sciences Publication Number 18, University of Wales, Bangor.

Hall, J., Burgess, N. D., Lovett, J. C., Mbilinyi, B., & Gereau, R. E. (2009). Conservation implications of deforestation across an elevational gradient in the Eastern Arc Mountains, Tanzania. *Biological Conservation*, ***142***(11), 2510–2521. https://doi.org/10.1016/j.biocon.2009.05.028

Hamilton, A. C. (1982). *Environmental history of East Africa: a study of the Quaternary*. Academic Press, London.

Hamilton, A. C., & Taylor, D. (1991). History of climate and forests in tropical Africa during the last 8 million years. *Climatic Change*, ***19***(1–2), 65–78. https://doi.org/10.1007/BF00142215

Hamrick, J. L. (1989). Isozymes and the analysis of genetic structure in plant populations. In D. E. Soltis & P. S. Soltis (Eds.), *Isozymes in Plant Biology* (pp. 87–105). Dioscorides Press.

Han, T., Hu, Y., Zhou, S., Li, H., Zhang, Q., Zhang, H., Huang, B., Rahman, K., Zheng, H., & Qin, L. (2008). Correlation between the genetic diversity and variation of total phenolic acids contents in *Fructus xanthii* from different populations in China. *Biomedical Chromatography*, ***22***, 478–486. https://doi.org/10.1002/bmc

Hardy, O. J., Maggia, L., Bandou, E., Breyne, P., Caron, H., Chevallier, M.-H., Doligez, A., Dutech, C., Kremer, A., Latouche-Hallé, C., Troispoux, V., Veron, V., & Degen, B. (2006). Fine-scale genetic structure and gene dispersal inferences in 10 neotropical tree species. *Molecular Ecology*, ***15***(2), 559–571. https://doi.org/10.1111/j.1365-294X.2005.02785.x

Harrier, L. A., Whitty, P. W., Sutherland, J. M., & Sprent, J. I. (1997). Phenetic investigation of non-nodulating African species of *Acacia* (Leguminosae) using morphological and molecular markers. *Plant Systematics and Evolution*, ***205***(1–2), 27–51.

Hayden, K. J., & Parker, I. M. (2002). Plasticity in cyanogenesis of Trifolium repens L.: inducibility, fitness costs and variable expression. *Evolutionary Ecology Research*, ***4***, 155–168.

Hedberg, O. (1994). *Afroalpine region. East and northeast tropical Africa. In: Centres of plant diversity: A guide and strategy for their conservation* (S. Davis, V. Heywood, & A. Hamilton, Eds.). Information Press, Oxford, UK.

Heuertz, M., Hausman, J.-F., Hardy, O. J., Vendramin, G. G., Frascaria-Lacoste, N., & Vekemans, X. (2004). Nuclear microsatellites reveal contrasting patterns of genetic structure between western and southeastern European populations of the common ash (*Fraxinus excelsior* L.). *Evolution: International Journal of Organic Evolution*, ***58***(5), 976–988. http://www.ncbi.nlm.nih.gov/pubmed/15212379

Hewitt, G. (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, ***58***(3), 247–276. https://doi.org/10.1006/bijl.1996.0035

Hewitt, G. (2000). The genetic legacy of the quaternary ice ages. *Nature*, ***405***(6783), 220–227. http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Global+patterns+in+biodiversity#0

Heywood, V. H. (2002). The conservation of genetic and chemical diversity in medicinal and aromatic plants. In B. Şener (Ed.), *Biodiversity: Biomolecular Aspects of Biodiversity and Innovative Utilization* (pp. 13–22.).

Holsinger, K. E., & Mason-Gamer, R. J. (1996). Hierarchical analysis of nucleotide diversity in geographically structured populations. *Genetics*, ***142***(2), 629–639.

Hopkins, M. S., Casa, A. M., Wang, T., Mitchell, S. E., Dean, R. E., Kochert, G. D., & Kresovich, S. (1999). Discovery and Characterization of Polymorphic Simple Sequence Repeats (SSRs) in Peanut. *Crop Science*, ***39***, 1243–1247.

Ikeda, Y., Murakami, A., & Ohigashi, H. (2008). Ursolic acid: an anti- and pro-inflammatory triterpenoid. *Molecular Nutrition & Food Research*, ***52***(1), 26–42. https://doi.org/10.1002/mnfr.200700389

Irmak, S., Jonnala, R. S., & MacRitchie, F. (2008). Effect of genetic variation on phenolic acid and policosanol contents of Pegaso wheat lines. *Journal of Cereal Science*, ***48***(1), 20–26. https://doi.org/10.1016/j.jcs.2007.07.007

Ishani, A., MacDonald, R., Nelson, D., Rutks, I., & Wilt, T. J. (2000). Pygeum africanum for the treatment of patients with benign prostatic hyperplasia: a systematic review and quantitative meta-analysis. *The American Journal of Medicine*, ***109***(8), 654–664. http://www.ncbi.nlm.nih.gov/pubmed/11099686

Ishikawa, S., Kato, S., Imakawa, S., Mikami, T., & Shimamoto, Y. (1992). Organelle DNA polymorphism in apple cultivars and rootstocks. *Theoretical and Applied Genetics (TAG)*, ***83***(8), 963–967.

IUCN. (2004). *2004 IUCN Red List of Threatened Species*. www.redlist.org

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, ***23***(14), 1801–1806. https://doi.org/10.1093/bioinformatics/btm233

Jana, S., & Pietzrak, L. N. (1988). Comparison assessment of genetic diversity in wild and primitive cultivated barley in a center of diversity. *Genetics*, ***119***, 981–990.

Jelinski, D. E., & Cheliak, W. M. (1992). Genetic diversity and spatial subdivision of Populus tremuloides (Salicaceae) in a heterogeneous landscape. *American Journal of Botany*, ***79***(7), 728–736.

Jolly, D., Taylor, D., Marchant, R., Hamilton, A. C., Bonnefille, R., Buchet, G., & Riollet, G. (1997). Vegetation dynamics in central Africa since 18, 000 yr BP: pollen records from the interlacustrine highlands of Burundi, Rwanda and western Uganda. *Journal of Biogeography*, ***24***, 495–512.

Jones, N., Ougham, H., & Thomas, H. (1997). Markers and mapping: we are all geneticists now. *New Phytologist*, ***137***, 165–177.

Kadu, C. A. C., Konrad, H., Schueler, S., Muluvi, G. M., Eyog-Matig, O., Muchugi, A., Williams, V. L., Ramamonjisoa, L., Kapinga, C., Foahom, B., Katsvanga, C., Hafashimana, D., Obama, C., & Geburek, T. (2013). Divergent pattern of nuclear genetic diversity across the range of the Afromontane *Prunus africana* mirrors variable Pleistocene climate of African highlands. *Annals of Botany*, ***111***(1), 47–60.

Kadu, C. A. C., Schueler, S., Konrad, H., Muluvi, G. M., Eyog-Matig, O., Muchugi, A., Williams, V. L., Ramamonjisoa, L., Kapinga, C., Foahom, B., Katsvanga, C., Hafashimana, D., Obama, C., & Geburek, T. (2011). Phylogeography of the Afromontane *Prunus africana* reveals a former migration corridor between East and West African highlands. *Molecular Ecology*, ***20***(1), 165–178. https://doi.org/10.1111/j.1365-294X.2010.04931.x

Kalkman, C. (1988). The phylogeny of the Rosaceae. *Botanical Journal of the Linnean Society*, ***98***(1), 37–59. https://doi.org/10.1111/j.1095-8339.1988.tb01693.x

Kampa, M., Alexaki, V.-I., Notas, G., Nifli, A.-P., Nistikaki, A., Hatzoglou, A., Bakogeorgou, E., Kouimtzoglou, E., Blekas, G., Boskou, D., Gravanis, A., & Castanas, E. (2004). Antiproliferative and apoptotic effects of selective phenolic acids on T47D human breast cancer cells: potential mechanisms of action. *Breast Cancer Research*, ***6***(2), 63–74. https://doi.org/10.1186/bcr752

Kebede, M., Ehrich, D., Taberlet, P., Nemomissa, S., & Brochmann, C. (2007). Phylogeography and conservation genetics of a giant lobelia (*Lobelia giberroa*) in Ethiopian and Tropical East African mountains. *Molecular Ecology*, ***16***(6), 1233–1243. https://doi.org/10.1111/j.1365-294X.2007.03232.x

Kikuzaki, H., Hisamoto, M., Hirose, K., Akiyama, K., & Taniguchi, H. (2002). Antioxidant properties of ferulic acid and its related compounds. *Journal of Agricultural and Food Chemistry*, ***50***(7), 2161–2168. http://www.ncbi.nlm.nih.gov/pubmed/11902973

Koch, M. A., Kiefer, C., Ehrich, D., Vogel, J., Brochmann, C., & Mummenhoff, K. (2006). Three times out of Asia Minor: the phylogeography of *Arabis alpina* L. (Brassicaceae). *Molecular Ecology*, ***15***(3), 825–839. http://www.ncbi.nlm.nih.gov/pubmed/16499705

Laitinen, M.-L. (2003). Variation in secondary chemistry within a natural population of birch: Effects of genotype, environment and ontogeny. *University of Joensuu, PhD Dissertations in Biology*, ***20***, 98.

Liakopoulos, G. (2001). Analysis of Epicuticular Phenolics of Prunus persica and Olea europaea Leaves: Evidence for the Chemical Origin of the UV-induced Blue Fluorescence of Stomata. *Annals of Botany*, ***87***(5), 641–648. https://doi.org/10.1006/anbo.2001.1387

Livingstone, D. A. (1975). Late Quaternary Climatic Change in Africa. *Annual Review of Ecology and Systematics*, ***6***(1), 249–280. https://doi.org/10.1146/annurev.es.06.110175.001341

Loveless, M. D. (1992). Isozyme variation in tropical trees: patterns of genetic organization. *New Forests*, ***6***, 67–94.

Lovett, J. C. (1993). *Climatic history and forest distribution in eastern Africa. Biogeography and ecology of the rain forests of eastern Africa* (J. C. Lovett & S. K. Wasser, Eds.). Cambridge University Press, Cambridge.

Ma, C.-M., Cai, S.-Q., Cui, J.-R., Wang, R.-Q., Tu, P.-F., Hattori, M., & Daneshtalab, M. (2005). The cytotoxic activity of ursolic acid derivatives. *European Journal of Medicinal Chemistry*, ***40***(6), 582–589. http://www.ncbi.nlm.nih.gov/pubmed/15922841

Maley, J. (1987). Fragmentation de la forêt dense humide africaine et extension des biotopes montagnardes au Quaternaire récent: nouvelles données polliniques et chronologiques. Implications paleoclimatiques et biogéographiques. *Palaeoecology of Africa*, ***18***, 307–334.

Maley, J. (1989). Late Quaternary Climatic Changes in the African Rain-Forest - Forest Refugia and the Major Role of Sea-Surface Temperature-Variations. In M. Leinen & M. Sarnthein (Eds.), *Paleoclimatology and paleometeorology Modern and past patterns of global atmospheric transport* (Issue 282, pp. 585–616). Kluwer Academic Publishers.

Maley, J. (1991). The African rain forest vegetation and palaeoenvironments during late quaternary. *Climatic Change*, ***19***(1), 79–98. https://doi.org/10.1007/BF00142216

Manni, F., Guérard, E., & Heyer, E. (2004). Geographic Patterns of (Genetic, Morphologic, Linguistic) Variation: How Barriers Can Be Detected by Using Monmonier's Algorithm. *Human Biology*, ***76***(2), 173–190.

Marchant, R., & Taylor, D. (1998). Dynamics of montane forest in central Africa during the late Holocene: a pollen-based record from western Uganda. *The Holocene*, ***8***(4), 375–381. https://doi.org/10.1191/095968398672993971

Marchant, R., Taylor, D., & Hamilton, A. C. (1997). Late Pleistocene and Holocene History at Mubwindi Swamp, Southwest Uganda. *Quaternary Research*, ***47***(3), 316–328. https://doi.org/10.1006/qres.1997.1887

Marley, M. L., & Parker, W. H. (1993). Phenotypic variation in cone and needle characters of Pinus banksiana (jack pine) in North-Western Ontario. *Canadian Journal of Botany*, ***71***, 43–51.

Martinelli, E. M., Seraglia, R., & Pifferi, G. (1986). Characterization of Pygeum africanum bark extracts by HRGC with computer assistance. *Journal of High Resolution Chromatography & Chromatography Communications*, ***9***(2), 106–110.

Mather, K., & Jinks, J. L. (1982). *Biometrical Genetics*. Chapman and Hall.

Matsumoto, S., Wakita, H., & Fukui, H. (1997). Molecular classification of wild roses using organelle DNA probes. *Scientia Horticulturae*, ***68***(191 – 196).

Mbonyimana, F. (1988). Plantation d'essences autochtones à Rangiro. In V. Pleines (Ed.), *Compte-rendu du premier séminaire national sur la sylviculture des plantations forestières au Rwanda* (pp. 245–252). Départment de Foresterie de l'Institut des Sciences Agronomiques du Rwanda.

McCauley, D. E. (1995). The use of chloroplast DNA polymorphisam in studies of gene flow in plants. *Trends in Ecology and Evolution*, ***10***, 198–202.

McCauley, D. E., Smith, R. A., Lisenby, J. D., & Hsieh, C. (2003). The hierarchical spatial distribution of chloroplast DNA polymorphism across the introduced range of *Silene vulgaris*. *Molecular Ecology*, ***12***(12), 3227–3235. https://doi.org/10.1046/j.1365-294X.2003.01992.x

Meve, U., & Liede, S. (2002). Floristic exchange between mainland Africa and Madagascar: Case studies in Apocynaceae-Asclepiadoideae. *Journal of Biogeography*, ***29***, 865–873.

Michalakis, Y., & Excoffier, L. (1996). A Generic Estimation of Population Subdivision Using Distances between Alleles with Special Reference for Microsatellite Loci. *Genetics*, ***142***(3), 1061–1064. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1207006/

Milbourne, D., Meyer, R. C., Collins, A. J., Ramsay, L. D., Gebhardt, C., & Waugh, R. (1998). Isolation, characterisation and mapping of simple sequence repeat

loci in potato. *Molecular and General Genetics*, ***259***(3), 233–245. https://doi.org/10.1007/s004380050809

Mohanty, A., & Marti, J. P. (2001). A population genetic analysis of chloroplast DNA in wild populations of *Prunus avium L.* in Europe. *Heredity,* ***87***(May), 421–427.

Mohanty, A., Martin, J. P., & Aguinagalde, I. (2002). Population genetic analysis of European *Prunus spinosa* (Rosaceae) using chloroplast DNA markers. *American Journal of Botany*, ***89***(8), 1223–1228.

Monmonier, M. S. (1973). Maximum-Difference Barriers: An Alternative Numerical Regionalization Method. *Geographical Analysis*, ***5***(3), 245–261.

Moreau, R. E. (1966). *The Birds Faunas of Africa and its Islands*. Academic Press, New York.

Morgan, D. R., Soltis, D. E., & Robertson, K. R. (1994). Systematic and Evolutionary Implications of rbcL Sequence Variation in Rosaceae. *American Journal of Botany*, ***81***(7), 890–903.

Mpofu, A., Sapirstein, H. D., & Beta, T. (2006). Genotype and environmental variation in phenolic content, phenolic acid composition, and antioxidant activity of hard spring wheat. *Journal of Agricultural and Food Chemistry*, ***54***(4), 1265–1270. https://doi.org/10.1021/jf052683d

Muchugi, A., Lengkeek, A. G., Kadu, C. A. C., Muluvi, G. M., Njagi, E., & Dawson, I. K. (2006). Genetic variation in the threatened medicinal tree *Prunus africana* in Cameroon and Kenya: Implications for current management and evolutionary history. *South African Journal of Botany*, ***72***(4), 498–506. https://doi.org/10.1016/j.sajb.2006.01.003

Munshi, S. K., & Sukhija, P. (1984). Compositional changes and biosynthesis of lipids in developing kernels of almond (*Prunus amydalus* Batsch). *Journal of the Science of Food and Agriculture*, ***35***(6), 689–697.

Nei, M. (1972). Genetic distance between populations. *The American Naturalist*, ***106***(949), 283–292. https://doi.org/10.1086/282771

Nei, M. (1975). *Molecular population genetics and evolution*. North-Holland Publishing Company.

Nicolas, V., Mboumba, J.-F., Verheyen, E., Denys, C., Lecompte, E., Olayemi, A., Missoup, A. D., Katuala, P., & Colyn, M. (2008). Phylogeographic structure and regional history of *Lemniscomys striatus* (Rodentia: Muridae) in tropical Africa. *Journal of Biogeography*, ***35***(11), 2074–2089. https://doi.org/10.1111/j.1365-2699.2008.01950.x

Oddou-Muratorio, S., Petit, R. J., le Guerroue, B., Guesnet, D., & Demesure, B. (2001). Pollen- versus seed-mediated gene flow in a scattered forest tree

species. *Evolution*, ***55***(6), 1123–1135. http://dx.doi.org/10.1111/j.0014-3820.2001.tb00632.x

Ohta, S., Nishitani, C., & Yamamoto, T. (2005). Chloroplast microsatellites in *Prunus*, Rosaceae. *Molecular Ecology Notes*, ***5***(4), 837–840. https://doi.org/10.1111/j.1471-8286.2005.01080.x

Ohta, T., & Kimura, M. (1973). A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetical Research*, ***22***(2), 201–204.

Omondi, S. F., Kireger, E., Dangasuk, O. G., Chikamai, B., Odee, D. W., Cavers, S., & Khasa, D. P. (2010). Genetic Diversity and Population Structure of *Acacia senegal* (L) Willd. in Kenya. *Tropical Plant Biology*, ***3***(1), 59–70. https://doi.org/10.1007/s12042-009-9037-2

Panaud, O., Chen, X., & McCouch, S. R. (1996). Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). *Molecular and General Genetics*, ***252***, 597–607.

Park, S. D. E. (2001). *Trypanotolerance in West African cattle and the population genetic effects of selection*.

Peakall, R., & Smouse, P. E. (2006). genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, ***6***(1), 288–295. https://doi.org/10.1111/j.1471-8286.2005.01155.x

Peakall, R., Smouse, P. E., & Huff, D. R. (1995). Evolutionary implications of allozyme and RAPD variation in diploid populations of dioecious buffalograss Buchloe dactyloides. *Molecular Ecology*, ***4***(2), 135–147.

Petit, R. J., Bodénès, C., Ducousso, A., Roussel, G., & Kremer, A. (2003). Hybridization as a mechanism of invasion in oaks. *New Phytologist*, ***161***(1), 151–164. https://doi.org/10.1046/j.1469-8137.2003.00944.x

Petit, R. J., Duminil, J., Fineschi, S., Hampe, A., Salvini, D., & Vendramin, G. G. (2005). Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology*, ***14***(3), 689–701. https://doi.org/10.1111/j.1365-294X.2004.02410.x

Petit, R. J., el Mousadik, A., & Pons, O. (1998). Identifying Populations for Conservation on the Basis of Genetic Markers. *Conservation Biology*, ***12***(4), 844–855. https://doi.org/10.1046/j.1523-1739.1998.96489.x

Petit, R. J., Kremer, A., & Wagner, D. B. (1993a). Geographic structure of chloroplast DNA polymorphisms in European oaks. *Theoretical and Applied Genetics*, ***87***(1), 122–128. https://doi.org/10.1007/BF00223755

Petit, R. J., Kremer, A., & Wagner, D. B. (1993b). Finite island model for organelle and nuclear genes in plants. *Heredity*, ***71***(6), 630–641. http://dx.doi.org/10.1038/hdy.1993.188

Petit, R. J., Pineau, E., Bialozyt, R., Brewer, S., Cheddadi, R., & Combs, B. (2001). *From spatial patterns of genetic diversity to post glacial migration processes in forest trees. Integrating ecology and evolution in a spatial context* (J. Silvertown & J. Antonovics, Eds.). Blackwell Science, Oxford, U.K.

Phillips, O. L., & Meilleur, B. A. (1998). Usefulness and Economic Potential of the Rare Plants of the United States: A Statistical Survey. *Economic Botany*, ***52***(1), 57–67.

Piry, S., Luikart, G., & Cornuet, J. M. (1999). BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity*, ***90***(4), 502–503. https://doi.org/10.1093/jhered/90.4.489

Pock Tsy, J.-M. L., Lumaret, R., Mayne, D., Vall, A. O. M., Abutaba, Y. I. M., Sagna, M., Rakotondralambo Raoseta, S. O., & Danthu, P. (2009). Chloroplast DNA phylogeography suggests a West African centre of origin for the baobab, *Adansonia digitata* L. (Bombacoideae, Malvaceae). *Molecular Ecology*, ***18***(8), 1707–1715. https://doi.org/10.1111/j.1365-294X.2009.04144.x

Pons, O., & Petit, R. J. (1995). Estimation, variance and optimal sampling of gene diversity I. Haploid locus. *Theoretical and Applied Genetics*, ***90***, 462–470.

Pons, O., & Petit, R. J. (1996). Measuring and Testing Genetic Differentiation with Ordered Versus Unordered Alleles. *Genetics*, ***144***(3), 1237–1245.

Powell, W. (1992). Plant genomes, gene markers and linkage maps. In J. P. Moss (Ed.), *Biotechnology and Crop Improvement in Asia* (pp. 297–322). International Crops Research Institute for the Semi-Arid Tropics.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, ***155***(2), 945–959. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1461096&tool=pmcentrez&rendertype=abstract

Provan, J., Soranzo, N., Wilson, N. J., Goldstein, D. B., & Powell, W. (1999). A Low Mutation Rate for Chloroplast Microsatellites. *Genetics*, ***153***(2), 943–947. http://www.genetics.org/content/153/2/943.abstract

Quiles, M. T., Arbós, M. A., Fraga, A., de Torres, I. M., Reventós, J., & Morote, J. (2010). Antiproliferative and apoptotic effects of the herbal agent *Pygeum africanum* on cultured prostate stromal cells from patients with benign prostatic hyperplasia (BPH). *The Prostate*, ***70***(10), 1044–1053.

Ramsay, L., Macaulay, M., degli Ivanissevich, S., MacLean, K., Cardle, L., Fuller, J., Edwards, K. J., Tuvesson, S., Morgante, M., Massari, A., Maestri, E.,

Marmiroli, N., Sjakste, T., Ganal, M., Powell, W., & Waugh, R. (2000). A Simple Sequence Repeat-Based Linkage Map of Barley. *Genetics*, ***156***(4), 1997–2005.

Raymond, M., & Rousset, F. (1995). GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, ***86***(3), 248–249.

Rongwen, J., Akkaya, M. S., Bhagwat, A. A., Lavi, U., & Cregan, P. B. (1995). The use of microsatellite DNA markers for soybean genotype identification. *Theoretical and Applied Genetics*, ***90***, 43–48.

Rosenberg, N. A. (2003). Distruct: A Program for the Graphical Display of Population Structure. *Molecular Ecology Notes*, ***4***(1), 137–138. https://doi.org/10.1046/j.1471-8286.2003.00566.x

Ruiz Guajardo, J. C., Schnabel, A., Ennos, R., Preuss, S., Otero-Arnaiz, A., & Stone, G. (2010). Landscape genetics of the key African acacia species *Senegalia mellifera* (Vahl)-the importance of the Kenyan Rift Valley. *Molecular Ecology*, ***19***(23), 5126–5139. https://doi.org/10.1111/j.1365-294X.2010.04833.x

Russell, J. R., Hussein, F., Johnson, E., Waugh, R., & Powell, W. (1993). Genetic differentiation of cocoa (*Theobroma cacao* L.) populations revealed by RAPD analysis. *Molecular Ecology*, ***2***, 89–97.

Sainsbury, M. (1970). Friedelin and Epifriedelinol from the Bark of Prunus turfosa and A Review of their Natural Distribution. *Phytochemistry*, ***9***, 2209–2215.

Sankar, A. A., & Moore, G. A. (2001). Evaluation of inter-simple sequence repeat analysis for mapping in Citrus and extension of the genetic linkage map. *Theoretical and Applied Genetics*, ***102***(2), 206–214. https://doi.org/10.1007/s001220051637

Satou, H. (2003). Physiological effects and safety of plant sterols. *Oleoscience*, *3*, 395–401.

Schueler, S., Tusch, A., & Scholz, F. (2006). Comparative analysis of the within-population genetic structure in wild cherry (*Prunus avium* L.) at the self-incompatibility locus and nuclear microsatellites. *Molecular Ecology*, ***15***(11), 3231–3243. https://doi.org/10.1111/j.1365-294X.2006.03029.x

Schueler, S., Tusch, A., Schuster, M., & Ziegenhagen, B. (2003). Characterization of microsatellites in wild and sweet cherry (*Prunus avium* L.)-markers for individual identification and reproductive processes. *Genome National Research Council Canada Genome Conseil National de Recherches Canada*, ***46***(1), 95–102. http://www.ncbi.nlm.nih.gov/pubmed/12669801

Schulze-Menz, G. K. (1964). Rosaceae. In H. Melchior (Ed.), *Engler's syllabus der Pflanzenfamilien, II* (pp. 209 – 218). Gebüder Borntraeger, Berlin.

Seetharamaiah, G. S., & Chandrasekhara, N. (1993). Comparative hypocholesterolemic activities of oryzanol, curcumin and ferulic acid in rats. *Journal of Food Science and Technology*, ***30***, 249–252.

Senior, M. L., Murphy, J. P., Goodman, M. M., & Stuber, C. W. (1998). Utility of SSRs for determining Genetic Similarities and Relationships in Maize Using an Agarose Gel System. *Crop Science*, ***38***(4), 1088–1098.

Simons, A. J., Dawson, I. K., Duguma, B., & Tchonndjeu, Z. (1998). Passing problems: prostrate and Prunus. *Herbalgram*, ***43***, 49–53.

Sosinski, B., Gannavarapu, M., Hager, L. D., Beck, L. E., King, G. J., Ryder, C. D., Rajapakse, S., Baird, W. v, Ballard, R. E., & Abbott, A. G. (2000). Characterization of microsatellite markers in peach *Prunus persica* (L.) Batsch. *Theoretical and Applied Genetics*, ***101***(3), 421–428.

Stattersfield, A. J., Crosby, M. J., Long, A. J., & Wege, D. C. (1998). Endemic Bird Areas of the World: Priorities for Biodiversity Conservation. In A. J. Stattersfield, M. J. Crosby, A. J. Long, & D. C. Wege (Eds.), *BirdLife Conservation Series No 7* (Vol. 7, Issue 7). Birdlife International. http://en.wikipedia.org/wiki/Endemic_Bird_Areas_of_the_World:_Priorities_for_Biodiversity_Conservation

Stewart, K. M. (2001). *The commercial bark harvest of the African cherry (Prunus africana) on Mount Oku, Cameroon: effects of traditional uses and population dynamics*. PhD Dissertation. Florida International University.

Stewart, K. M. (2003a). The African cherry (*Prunus africana*): can lessons be learned from an over-exploited medicinal tree? *Journal of Ethnopharmacology*, *89*(1), 3–13. http://linkinghub.elsevier.com/retrieve/pii/S037887410300299X

Stewart, K. M. (2003b). The African cherry (*Prunus africana*): from hoe-handles to the international herb market. *Economic Botany*, ***57***(4), 559–569. https://doi.org/10.1663/0013-0001(2003)057[0559: TACPAF]2.0.CO;2

Stone, H. (1924). *The timbers of commerce and their identification* (2nd ed.). Rider.

Street-Perrott, F. A., & Perrott, R. A. (1993). Holocene vegetation, lake levels, and climate of Africa. In H. E. Wright & Et Al. (Eds.), *Global climates since the last glacial maximum* (pp. 318–356). University of Minnesota Press.

Streiff, R., Ducousso, A., Lexer, C., Steinkellner, H., Gloessl, J., & Kremer, A. (1999). Pollen dispersal inferred from paternity analysis in a mixed oak stand of *Quercus robur* L. and *Q. petraea* (Matt.) Liebl. *Molecular Ecology*, ***8***(5), 831–841. http://dx.doi.org/10.1046/j.1365-294X.1999.00637.x

Sunderland, T., & Nkefor, J. (1997). *Conservation through cultivation. A case study: The propagation of pygeum - Prunus africana*.

Takeoka, G. R., & Dao, L. T. (2003). Antioxidant constituents of almond [*Prunus dulcis* (Mill.) D.A. Webb] hulls. *Journal of Agricultural and Food Chemistry*, ***51***(2), 496–501. https://doi.org/10.1021/jf020660i

Tautz, D. (1989). Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research*, ***17***(16), 6463–6471.

Taylor, S. J. (2015). *African biodiversity hotspots and other important conservation designations for unique areas which may include mountains*.

Thorne, R. F. (1992). Classification and geography of the flowering plants. *Botanical Review*, ***58***, 225 – 348.

Tolley, K. A., Tilbury, C. R., Measey, J. G., Menegon, M., Branch, W. R., & Matthee, C. A. (2011). Ancient forest fragmentation or recent radiation? Testing refugial speciation models in chameleons within an African biodiversity hotspot. *Journal of Biogeography*, ***38***(9), 1748–1760. https://doi.org/10.1111/j.1365-2699.2011.02529.x

Tragoonrung, S., Kanazin, V., Hayes, P. M., & Blake, T. K. (1992). Sequence tagged sites facilitated PCR for barley genome mapping. *Theoretical and Applied Genetics*, *b*, 1002–1008.

Trauth, M. H., Maslin, M. a, Deino, A., & Strecker, M. R. (2005). Late Cenozoic moisture history of East Africa. *Science*, ***309***(5743), 2051–2053. https://doi.org/10.1126/science.1112964

Trindade, H., Costa, M. M., Sofia, A., Lima, B., Pedro, L. G., Figueiredo, A. C., & Barroso, J. G. (2008). Genetic diversity and chemical polymorphism of *Thymus caespititius* from Pico, São Jorge and Terceira islands (Azores). *Biochemical Systematics and Ecology*, ***36***(10), 790–797. https://doi.org/10.1016/j.bse.2008.09.001

Tsitrone, A., Kirkpatrick, M., & Levin, D. A. (2003). A model for chloroplast capture. *Evolution: International Journal of Organic Evolution*, ***57***(8), 1776–1782. http://www.ncbi.nlm.nih.gov/pubmed/14503619

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***(3), 535–538. https://doi.org/10.1111/j.1471-8286.2004.00684.x

van Zinderen Bakker, E. M., & Clark, J. D. (1962). Pleistocene climates and cultures in northeastern Angola. *Nature*, ***196***(4855), 639–642.

Vaughan, S. P., & Russell, K. (2004). Characterization of novel microsatellites and development of multiplex PCR for large-scale population studies in wild cherry, *Prunus avium*. *Molecular Ecology Notes*, ***4***(3), 429–431. https://doi.org/10.1111/j.1471-8286.2004.00673.x

Vences, M., Wollenberg, K. C., Vieites, D. R., & Lees, D. C. (2009). Madagascar as a model region of species diversification. *Trends in Ecology & Evolution*, ***24***(8), 456–465. http://www.ncbi.nlm.nih.gov/pubmed/19500874

Vendramin, G. G., & Ziegenhagen, B. (1997). Characterisation and inheritance of polymorphic plastid microsatellites in *Abies*. *Genome*, ***40***(6), 857–864.

Verhoeven, K. J. F., Macel, M., Wolfe, L. M., & Biere, A. (2011). Population admixture, biological invasions and the balance between local adaptation and inbreeding depression. *Proceedings of the Royal Society B: Biological Sciences*, ***278***(1702), 2–8. http://rspb.royalsocietypublishing.org/cgi/doi/10.1098/rspb.2010.1272

Vieira, R. F., Grayer, R. J., Paton, A., & Simon, J. E. (2001). Genetic diversity of *Ocimum gratissimum* L. based on volatile oil constituents, flavonoids and RAPD markers. *Biochemical Systematics and Ecology*, ***29***(3), 287–304. http://www.ncbi.nlm.nih.gov/pubmed/11152947

von Breitenbach, F. (1965). *The indigenous trees of Southern Africa. 4 vols*. Government Printer.

Vornam, B., & Gebhardt, K. (2000). PCR-based markers reveal genetic identity and diversity in subset collections of wild and cultivated apple. *Acta Horticulturae*, ***530***, 463–467.

Voronoi, G. (1908). Nouvelles applications des parameters continus à la théorie des formes quadratiques, deuxième memoire, recherche sur les parallelloèdres primitifs. *Journal Fur Die Reine Und Angewandte Mathematik*, ***134***, 198–287.

Vos, P., Hogers, R., Bleecker, M., Reijans, van de L. T., Hornes, M., Freijters, A., Pot, J., Peleman, J., & Zabeau, M. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, ***23***, 4407–4414.

Vosberg, H.-P. (1989). The polymerase chain reaction: an improved method for the analysis of nucleic acids. *Human Genetics*, ***83***, 1–15.

Ward Jr., J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, ***58***(301), 236–244. https://doi.org/10.2307/2282967

Welsh, J., & McClelland, M. (1990). Fingerprinting genomes using PCR with arbitrary primers*. *Nucleic Acids Research*, ***18***(24), 7213–7218.

White, F. (1981). The history of the Afromontane archipelago and the scientific need for its conservation. *African Journal of Ecology*, ***19***, 33–54.

White, F. (1983). Long-distance Dispersal. *Sonderbd. Naturwiss. Ver. Hamburg*, ***7***, 87–116.

White, F. (1993). Refuge theory, ice-age aridity and the history of tropical biotas: an essay in plant geography. *Fragmenta Floristica et Geobotanica Supplementum*, ***2***(2), 385–409.

Wickens, G. E. (1976). Speculations on long distance dispersal and the flora of Jebel Marra, Sudan Republic. *Kew Bulletin*, ***31***, 105–150.

Wilde, J., Waugh, R., & Powell, W. (1992). Genetic fingerprinting of *Theobroma* clones using RAPD markers. *Theoretical and Applied Genetics*, ***88***, 871–877.

Williams, J. G., Kubelik, A R., Livak, K. J., Rafalski, J. A, & Tingey, S. V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*, ***18***(22), 6531–6535. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=332606&tool=pmcentrez&rendertype=abstract

Wilson, G. A., & Rannala, B. (2003). Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, ***163***(3), 1177–1191. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1462502&tool=pmcentrez&rendertype=abstract

Wilson, I. J., Weale, M. E., & Balding, D. J. (2003). Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, ***166***(2), 155–188. http://dx.doi.org/10.1111/1467-985X.00264

Wolbiš, M., Olszewska, M., & Wesolowski, W. J. (2001). Triterpenes and Sterols in the Flowers and Leaves of *Prunus apinosa* L. (Rosaceae). *Acta Poloniae Pharmaceutica*, ***58***(6), 459–462.

Xu, M. L., Melchinger, A. E., Xia, X. C., & Lübberstedt, T. (1999). High-resolution mapping of loci conferring resistance to sugarcane mosaic virus in maize using RFLP, SSR, and AFLP markers. *Molecular and General Genetics*, ***261***(3), 574–581. http://dx.doi.org/10.1007/s004380051003

Yamaya, A., Endo, Y., Fujimoto, K., & Kitamura, K. (2007). Effects of genetic variability and planting location on the phytosterol content and composition in soybean seeds. *Food Chemistry*, ***102***(4), 1071–1075. https://doi.org/10.1016/j.foodchem.2006.07.001

Yan, J., Chu, H.-J., Wang, H.-C., Li, J.-Q., & Sang, T. (2009). Population genetic structure of two *Medicago* species shaped by distinct life form, mating system and seed dispersal. *Annals of Botany*, ***103***(6), 825–834. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2707894&tool=pmcentrez&rendertype=abstract

Zhou, C.-H., Li, X., Zhang, W.-S., Sun, C.-D., & Chen, K.-S. (2011). Oleanolic and ursolic acid in the fruit of *Eriobotrya japonica* Lindl. *Journal of Medicinal Plants Research*, ***5***(9), 1735–1740.