Landscape Genetics and Ecological Niche Modeling of

***Afzelia quanzensis* (Pod Mahogany)**

BY

PERCY JINGA
B.S., Bindura University of Science Education, 2005
M.S., Bindura University of Science Education, 2009

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Biological Sciences
in the Graduate College of the
University of Illinois at Chicago, 2017

Chicago, Illinois

Defense Committee:

Mary V. Ashley, Chair and Advisor
Roberta J. Mason-Gamer
Henry F. Howe
Andrea T. Kramer, Chicago Botanic Garden
Andrew L. Hipp, The Morton Arboretum

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Mary V. Ashley, for giving me the opportunity to pursue graduate studies in the Ashley Lab, and the accompanying tremendous effort in guiding my learning and research. On one hand, I am so happy for the knowledge I gained from her throughout my time under her guidance, and on the other, I am so sad to be leaving the Ashley Lab. I would like to thank all members of the Ashley Lab, particularly Jason Palagi, Dr. Janet Backs, and Eun Sun Kim, who assisted me with developing lab and data analysis skills. Appreciation goes to members of my committee, Dr. Roberta J. Mason-Gamer, Dr. Henry, F. Howe, Dr. Andrea T. Kramer, and Dr. Andrew L. Hipp, for their guidance throughout my graduate school work.

Acknowledgements go to Kevin Feldheim of the Pritzker Lab at The Field Museum of Natural History for facilities and equipment. I would like to thank Karolis Ramanauskas of the Igic Lab at the University of Illinois at Chicago for providing computing platforms. Appreciation goes to staff at the Forestry Commission of Zimbabwe for helping with plant species identification and sampling.

Funding for this work was provided by Fulbright, International Institute of Education, the Office of International Affairs at the University of Illinois at Chicago, the Chicago Consular Corps Scholarship, and the Biological Sciences Department at the University of Illinois at Chicago.

Lastly, I would like to thank my family, particularly Patience Marange, my fiancée, for enduring my absence, and supporting me all the way.

CONTRIBUTION OF AUTHORS

Jinga, P., Palagi, J. and Ashley V. M. 2016. Development of microsatellite loci of pod mahogany, *Afzelia quanzensis* (Fabaceae), by Illumina shotgun sequencing, and cross-amplification in *A. africana*. *Applications in Plant Sciences*, **4**: doi:10.3732/apps.1600010.

Percy Jinga

I collected *Afzelia quanzensis* samples, and extracted DNA for preparation of a library for next generation sequencing. I set up the sequencing run, and searched for microsatellite loci from output raw reads. I tested amplification and polymorphisms of potential loci, and cross-amplified polymorphic ones in a congeneric species, *A. africana*. I wrote the manuscript.

Jason Palagi

JP helped with DNA extraction.

Mary V. Ashley

MVA helped with data analysis, and proofreading the manuscript.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (continued)

LIST OF ABBREVIATIONS

ABC	Approximate Bayesian Computation
AUC	Area Under the Receiver Operating Characteristic Curve
C	Congolia rain forest
DNA	Deoxyribonucleic acid
ENM	Ecological Niche Modeling
GBIF	Global Biodiversity Information Facility
GCM	Global Climate Model
IBD	Isolation by Distance
KZ axis	Kalahari-Zimbabwe axis
LG	Lower Guinea rain forest
NGS	Next Generation Sequencing
PCR	Polymerase chain reaction
RCP	Representative Concentration Pathway
ROC	Receiver Operating Characteristic
SDM	Species Distribution Model
UG	Upper Guinea rain forest

SUMMARY

The distribution and genetic structure of plant populations have important consequences for survival. Past climate changes may have genetic signatures that affect current structure and distributions of species. The current climate change may result in range shifts, extinctions, and adaptations. The impacts of the current climate change on the distribution of species are important for designing mitigatory measures. Genetic structure and distributions are also affected by biogeographic barriers. My work aimed at establishing how Pleistocene climate changes might have affected genetic structure of tropical African trees, how climate change may affect the distribution of *Afzelia quanzensis* in its native sub-Saharan Africa range, and how a geographic barrier, the Kalahari-Zimbabwe (KZ) axis, might be impacting gene flow between *A. quanzensis* populations located on different sides of the mountain range.

In chapter one, genetic studies of African trees were reviewed to determine genetic signatures of past climate changes. Several hypotheses have been proposed to explain the occurrence and distribution of biodiversity in the tropics. The Forest refuge hypothesis was among the earliest hypothesis to be proposed. The Forest refuge hypothesis proposes that during climate changes of the Pleistocene, species survived in climatically stable and isolated refuges during cold and dry periods, and expanded from these refuges during warm and mesic conditions. When species were restricted in the isolated refugia, they underwent differentiation due to genetic drift, and eventually, speciation, due to a combination of factors that include adaptation to microenvironments. The refuges may be identified today as areas of high species

richness and endemism. Data from tropical South America have largely disputed the existence of Pleistocene refugia. In Africa, areas of high species richness and endemism have been identified in rain forests as putative Pleistocene refugia.

Sixteen genetic studies covering fourteen tree species were reviewed. In all the studies, at least a refugial population, that is a population with high allelic richness, and unique haplotypes or alleles and located in areas of high species richness and endemism, was identified. Refugial populations were identified in all the three domains of the African rain forest, that is the Upper Guinea, the Lower Guinea, and the Congolia phytogeographic regions. Populations were generally highly differentiated. These findings suggest that tropical African trees were once isolated, and expansion from isolation can be determined from their genetic imprint. African tropical biodiversity may be explained by the Forest refuge hypothesis, unlike in the neotropics. African trees may be clustered into populations of high allelic richness with unique alleles or haplotypes. Identifying populations with high genetic diversity is important for management, especially of commercially harvested species.

The current distribution of *A. quanzensis*, a species that occurs in sub-Saharan Africa, is largely unknown despite its economic importance. In chapter two, my objectives were to establish the current distribution of *A. quanzensis*, and to determine the consequences of climate change. *Afzelia quanzensis* is a medium to large tree species that has a variety of uses in its native range. I used Maxent modeling, incorporating bioclimatic variables at known occurrence

sites, to estimate the current distribution. Occurrence sites were obtained from the Global Biodiversity Information Facility, and the Tropicos database maintained by the Missouri Botanical Garden. Bioclimatic variables were obtained from the Worldclim-Global Climate Data database. The estimated current distribution shows the species occurring south of the Sahara, particularly in eastern and southern African countries, such as Kenya, Tanzania, Malawi, Zimbabwe, Zambia, Mozambique, and South Africa.

Projections of the current distribution were made to 2050 and 2070 under four Representative Concentration Pathways (RCPs), and two Global Climate Models (GCMs) to determine the impact of climate change. The four RCPs were RCP2.6, RCP4.5, RCP6.0, and RCP8.5, and the two GCMs were the Community Climate System Model version 4.0 (CCSM4.0), and the Institut Pierre Simon Laplace Climate Model version 5A (IPSL-CM5A-LR). When projected to 2050 and 2070, the distribution increased for both GCMs and all RCPs, by 24-307 %. The distribution increased the most under RCP8.5, where no mitigation to emission of greenhouse gases is assumed. Therefore, a warming environment may not be a threat to the existence of *A. quanzensis*. It is, however, important to note that indirect impacts of climate change were not considered in the modeling. Hindcasting to mid-Holocene showed a decline in the distribution. The distribution declined by 41 % and 20 % under CCSM4.0 and IPSL-CM5A-LR, respectively. *Afzelia quanzensis* might have been restricted into isolated populations during the mid-Holocene.

Microsatellites are genetic markers of choice because of their codominance, putative neutrality, and ease of genotyping. In chapter three, I developed microsatellite loci of *A. quanzensis* that will be used in genetic studies of the species. Next generation sequencing was used to generate raw reads from where microsatellite loci were directly identified. Seventy potential microsatellite loci were tested, and 39 consistently amplified. The loci that amplified were checked for polymorphisms in 40 individuals randomly collected from Zimbabwe. A total of 12 loci were polymorphic, comprised of four di-, three tri-, one tetra-, one penta-, one hexa-, and two compound motifs. The number of alleles for the 12 loci varied from three to 10 with a mean of 5.583. Observed heterozygosity ranged from 0.138 to 0.737, while expected heterozygosity ranged from 0.313 to 0.832.

The twelve polymorphic loci were cross-amplified in a congeneric species, *A. africana*. Eight loci were transferable. Next generation sequencing produced an enormous number of potential microsatellite loci for testing. This is in contrast to old enrichment methods, which are highly laborious but yielding a small number of potential loci. With declining costs of sequencing, next generation sequencing should be the method of choice for marker development. Dinucleotide repeat motifs were highly polymorphic, and they should be prioritized when testing for polymorphisms in future. The loci developed in this chapter will be used for genetic studies in *A. quanzensis*. Some of the loci may also be used in congeners, because of the high transferability in *A. africana*.

Barriers to gene flow have important consequences for genetic structure and the persistence of populations. If populations are isolated by barriers, their genetic diversity may decrease due to genetic drift. Low genetic diversity may reduce adaptability of populations to future changes. In chapter four, the objective was to examine the impact of a geographic barrier, the KZ axis, a mountain range that divides *A. quanzensis* populations into northern and southern distributions. A total of 192 samples were collected across the mountain range, and genotyped at 10 microsatellite loci developed in chapter three. Genetic diversity was relatively low in both northern and southern individuals. Overall genetic diversity estimates were 2.917, 2.208, 0.466, and 0.452 for allelic richness, effective number of alleles, observed heterozygosity, and expected heterozygosity, respectively. There was no significant difference between most of the genetic diversity estimates between northern and southern individuals. STRUCTURE analysis and principal components analysis delineated southern from northern individuals. A genetic barrier was detected that coincided with the KZ axis, and gene flow across the mountain was very low. The low genetic diversity may affect the long-term persistence of the species. The differentiated populations should be considered for germplasm management in order to capture all the extant genetic diversity.

1. GENETIC STUDIES OF TROPICAL TREES IDENTIFY PUTATIVE PLEISTOCENE REFUGIA IN AFRICAN RAIN FORESTS

1.1 Introduction

The tropics hold an enormous amount of biodiversity. It is estimated that the tropics hold over seventy percent of the species on earth, but comprise only six percent of the earth's surface (Wilson, 1988). The origin and sustainability of tropical biodiversity have been subjects of intense research. Several hypotheses have been suggested to explain the high species richness and evenness in the Amazon and other tropical forests. The Forest refuge hypothesis (Haffer, 1969) was among the earliest to be proposed, and it gained a lot of traction. The Forest refuge hypothesis states that during the cold and dry climates of the Quaternary, when ice sheets encroached the lower latitudes, the Amazon suffered species attrition, resulting in huge swathes of savanna and isolated tropical forested areas that remained climatically stable. Tropical species retreated into these refuges, and expanded when the climate became favorable. Speciation and genetic differentiation occurred in the small and isolated populations. The regression to and expansion from refugia happened repeatedly, therefore, these refuges should be areas of high species diversity and endemism at present. Historic climate changes have been associated with changes in global distribution of species (Walther et al., 2002; Pearson and Dawson, 2003; Permesan and Yohe, 2003; Chen et al., 2011). Using Amazonian bird diversity, Haffer (1969) identified nine putative forest refugia in central and northern South America. The Forest refuge hypothesis seemed to hold for a time (Prance, 1982; Haffer, 1985; Haffer, 1997; Haffer and

Prance, 2001), until it started to be unraveled by a spectrum of evidence collected in South America.

The inadequacy of the Forest refuge hypothesis in tropical South America has been demonstrated from phylogenetic, mineralogy, geology, and paleoecology studies. Phylogenetic studies have shown that a number of Amazonian species, such as butterflies (Brower and Egan, 1997), birds (Moritz et al., 2000), and a neotropical tree genus *Inga* (Richardson et al., 2001), have a Miocene and Pliocene origin, well before the Quaternary. The high species richness and endemism in the neotropics, thus, predates the Pleistocene and cannot be explained by the Forest refuge hypothesis. Pollen analyses have shown that the Amazon was largely forested during the Quaternary, and evidence of huge swathes of savannas are not supported (Bush et al., 1990; Liu and Colinvaux, 1985; Colinvaux et al., 2000; Bush and de Oliveira, 2006). Analyses of lignin, phenols, cutin acids and stable carbon isotopes in organic matter from deep sea sediments show a constancy of tropical vegetation cover over central and north South America (Kastner and Goni, 2003).

With such robust evidence against the Forest refuge hypothesis, alternative hypotheses have been proposed, and these include the Paleogeography (Emsely, 1965), River (Hershkovitz, 1977), River-refuge (Aryes and Clutton-Brock, 1992), Disturbance-vicariance (Colinvaux, 1993), and Gradient hypotheses (Endler, 1982).

The Paleogeography hypothesis states that extant species and their distributions arose as a result of periodic changes in sea level, and tectonic movements. These sea level fluctuations and

uplifts resulted in species isolation that was sufficient to drive differentiation (Emsely, 1965). Under the River hypothesis, wide rivers that developed in the Amazon during the late Tertiary and early Quaternary acted as boundaries that isolated species and resulted in their differentiation (Hershkovitz, 1977). The River-refuge hypothesis combines isolation by rivers (River hypothesis), and climate vegetation changes (Haffer, 1969), while the Disturbance-vicariance hypothesis explains high biodiversity by competitive interactions among species (Colinvaux, 1993). The Gradient hypothesis states that contiguous populations differentiated due to presence of steep environmental changes, and thus species assemblages change from one environment to the next, in the absence of physical separation (Endler, 1982).

While the existence of forest refuges has largely been disputed in the neotropics, could the theory be revived in other centers of high species richness, such as in tropical Africa? Palynological studies in tropical Africa have largely detected vegetation changes consistent with that suggested under the Forest refuge hypothesis. Analyses of pollen from three marine cores off the Gulf of Guinea have indicated fluctuations in vegetation cover during the Quaternary over the central African watershed (Hooghiemstra and Agwu 1988; Bengo and Maley, 1991). Vegetation changes in tropical Africa were also deduced from pollen and carbon isotopes of organic matter collected inland, at Lake Bosumtwi in Ghana (Maley, 1991) and Lake Barombi Mbo in Cameroon (Maley, 1991; Giresse et al., 1994). The inland analyses showed the persistence of rain forest refugia during the Quaternary in the midst of periodic tropical forest regression. The presence of a large swathe of savanna vegetation, the Dahomey Gap, separating the west from the central Africa rain forests, is further evidence of the vegetation changes in tropical Africa. Palynological studies have shown that the Dahomey Gap oscillated between

savanna and rain forest vegetation during the last 150 000 years (Talbot et al., 1984; Dupont and Weinelt, 1996; Russel et al., 2003). Sedimentary geochemistry studies collected at Lac Sélé in southern Benin showed the rapid deterioration of tropical vegetation in the Dahomey Gap during the cold and dry conditions of the Quaternary, and an expansion of Sudano-Guinean savannas (Salzmann and Hoelzmann, 2005), and the trend was reversed during warm and mesic conditions. These studies show that vegetation oscillated between rain forest and savanna types, opening the possibility of the presence of Pleistocene forest refuges in tropical Africa.

The identification of areas of high species richness and endemism in tropical Africa has been ongoing as these are thought to be Pleistocene refugia. Early studies identified two putative refugia, one in western Gabon, and the other in central Democratic Republic of Congo (DRC) (Kingdon, 1971; Diamond and Hamilton, 1980; Grubb, 1978; Prigogine, 1988). The presence of contact zones between suggested refugia provide strong evidence for their existence (Mayr and O'Hara, 1986). More areas of high species richness and endemism were identified in central DRC (Colyn et al., 1991) and southwest Ghana (Hall and Swine, 1981). As speciation is believed to have occurred in the isolated refugia, detailed maps of the location of putative refugia in tropical Africa have been attempted (eg. Moreau, 1969; Hamilton, 1976; Sosef, 1991) based on high species richness and endemism. An aggregation of these maps, in addition to palaeoecological data, led to a comprehensive description and location of putative Pleistocene refugia in the Guineo-Congolian rain forest by Maley (1996) (Figure 1.1).

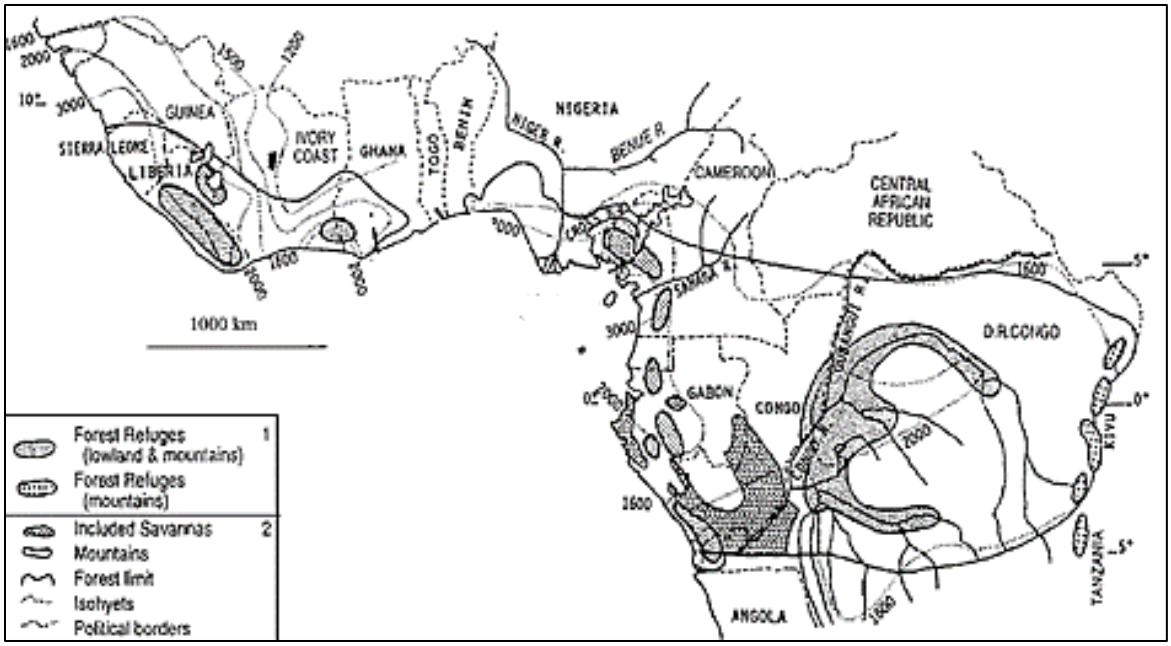


Figure 1.1. Location of putative Pleistocene refugia, areas of high species richness and endemism in tropical Africa (modified from Maley, 1996)

The Guineo-Congolian rain forest holds most of the tropical species in Africa, and has been divided into three domains, which are Upper Guinea (UG), Lower Guinea (LG) and Congolia (C) phytogeographic regions (White, 1979). The UG is made up of the west African rain forest, in Ivory Coast, Sierra Leonne, Guinea and Liberia. Three putative refugia have been identified in the UG. The LG domain is centered around Cameroon and Gabon, and harbors seven putative refugia. The C domain is located in the DRC, and is made up of one large putative refugium in the central Congo Basin (Maley, 1996; Budde et al., 2013).

The location of putative refugia using species richness and endemism, however, is strongly challenged because habitat heterogeneity may promote speciation without historical isolation (Moritz et al., 2000). Species richness data and endemism are also problematic because

some sister taxa diverged before the Pleistocene (Brower and Egan, 1997; Moritz et al., 2000; Plana, 2004; Plana et al, 2004). Phylogenetic studies and population genetics of tree species should provide valuable data to test the Forest refuge hypothesis. Trees generally have long longevity, and low speciation rates (Petit and Hampe, 2006), and therefore they are expected to harbor endemic alleles within species (Dainou et al., 2010). The presence of endemic haplotypes and alleles in putative refugia, evidence of demographic expansion of putative refugial populations, and admixture at putative refugia contact zones are obtained from population genetics data, and may help to support or refute the Forest refuge hypothesis.

The aim of this study was to identify congruences between putative Pleistocene refugia delineated by studies of species richness and endemism and those identified in genetic studies of tropical African tree species. I conducted a literature search for relevant genetic studies of trees in order to evaluate how well they supported the refuge hypothesis for explaining diversification of African forests. Genetic studies compliment work on species richness and endemism, and strengthen the conclusion of the existence of Pleistocene refugia in tropical Africa. Identification of refugia is important in delineating areas of high priority for conservation. Knowledge of the historical distribution of species in response to climate change is also valuable in predicting their distribution under the current climate change (Aitken et al., 2008). Measures that assist adaptation, and migration under global warming will be devised once historical responses under similar conditions are known.

1.2 Study Selection and Results

The Web of Science database was searched for genetic studies in tree species of the African rain forests using the titles “Genetic diversity of tropical African tree species”, “Population structure of tropical African tree species”, “Phylogeography of tropical African tree species”, “Pleistocene refugia in the Guineo-Congolian rain forest”, “Upper Guinea tree species genetic diversity and population structure”, “Lower Guinea tree species genetic diversity and population structure”, and “Congolian tree species genetic diversity and population structure”. An initial total of 112 published articles was recovered. The articles were searched for further retrieval of relevant articles.

The first filter was to check whether retrieved studies used genetic methods and involved tropical Africa tree species. The second filter was checking for the presence of an objective to identify Pleistocene refugia suggested in Maley (1996) in any of the three domains of the Guineo-Congolian rain forest. The numbers of studies recovered, included, and filtered out were tracked using the Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) method (Moher et al., 2009) (Figure 1.2). Sixteen studies covering 14 tropical tree species were finally used in this study (Table I). Sample size of populations was very variable among the selected articles, ranging from 43 to 864. A variety of genetic markers were used, including SSRs, cpDNA, RAPDs, AFLPs and RFLPs. The presence of rare haplotypes and alleles, and high allelic richness was used to assign a population to a putative refugium in the 16 studies.

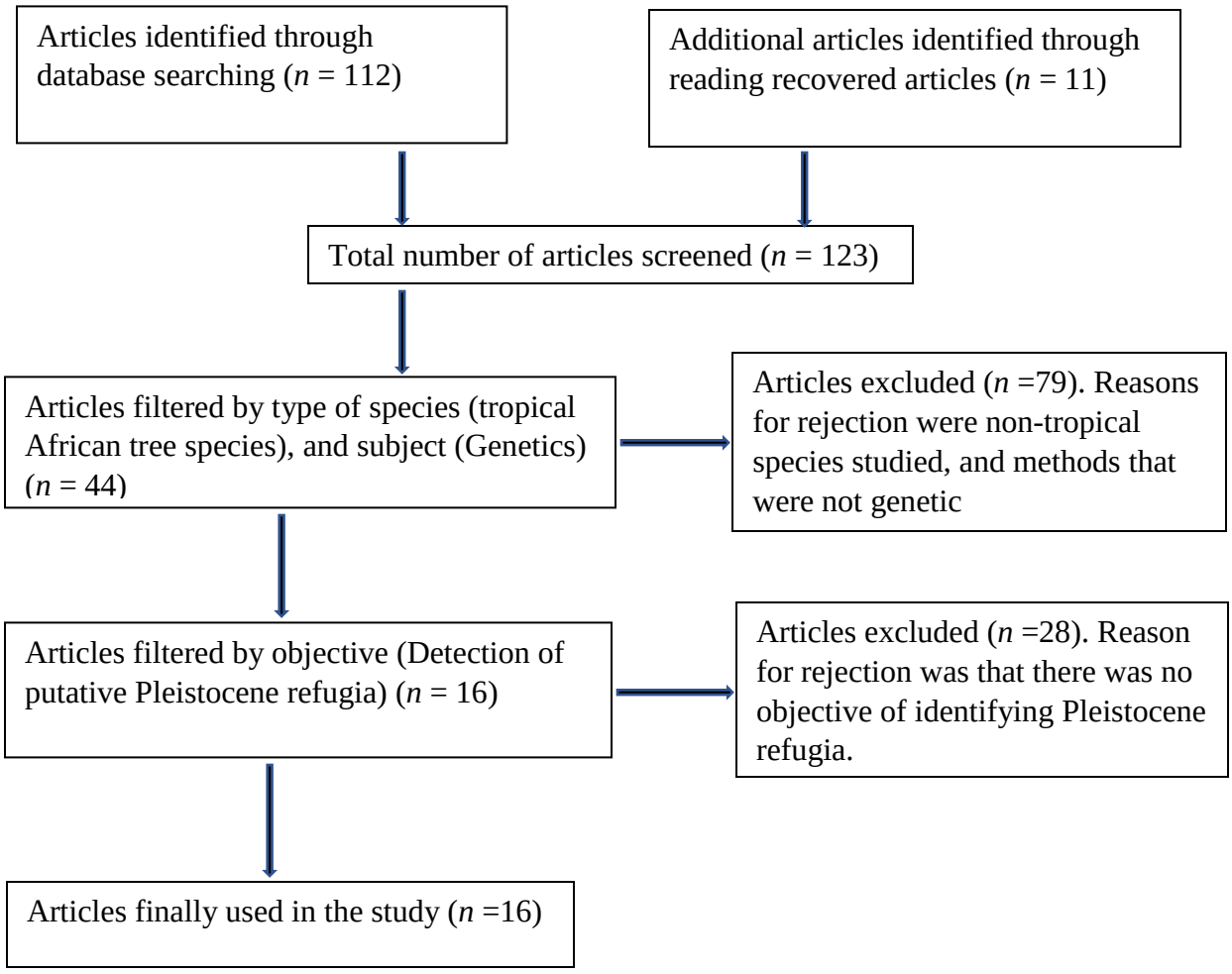


Figure 1.2. Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) flow diagram used to track studies of detection of putative Pleistocene refugia in the Guineo-Congolian rain forest

Table I. Summary of studies of tropical African tree species used to identify putative Pleistocene refugia in the Guineo-Congolian rain forest

Species	**Family**	***N***	**Domain**	**Genetic marker**	**Putative Refugia**	**Reference**
Irvingia gabonensis *Irvingia wombolu*	Irvingiaceae	87, 43	LG	RAPDs	Southern Cameroon, Southeast Nigeria, Southcentral Gabon	Lowe et al. (2000)
Aucoumea klaineana	Burseraceae	79	LG	cpDNA	Northern Gabon (Cristal Mountains), Southcentral Gabon (Chaillu Massif Mountains), Western Gabon (Doudou Mountains)	Muloko-Ntoutoume et al. (2000)
Vitellaria paradoxa	Sapotaceae	179	UG, LG	RAPDs, cpDNA	Southwest Ghana, Southwestern Cameroon	Fontaine et al. (2004)
Coffea canephora	Rubiaceae	61	UG, LG, C	nSSRs, RFLPs	Southwest Cameroon, Congo Basin	Gomez et al. (2009)
Greenwayodendron suaveolens	Annonaceae	186	LG	cpDNA	Southwest Cameroon, Western Gabon	Dauby et al. (2010)
Milicia excelsa	Moraceae	550	UG, LG	nSSRs, cpDNA	Southwest Cameroon, Southwest Ghana, Congo basin[a], Gabon[a]	Dainou et al. (2010)
Irvingia gabonensis	Irvingiaceae	85	UG, LG, C	RFLPs, cpDNA	Southern Cameroon, Southeastern Nigeria, Southwestern Cameroon	Lowe et al. (2010)
Distemonanthus benthamianus	Fabaceae	295	LG	nSSRs	Southern Cameroon, Western Gabon, Southcentral Gabon	Debout et al. (2011)
Aucoumea klaineana	Burseraceae	864	LG	nSSRs	Northern Gabon (Cristal Mountains), Western Gabon (Doudou Mountains), Southcentral Gabon (Chaillu Massif Mountains)	Born et al. (2011)

Species	**Family**	***N***	**Domain**	**Genetic marker**	**Putative Refugia**	**Reference**
Santiria trimera	Burseraceae	594	UG, LG	cpDNA	Southwestern Cameroon, Southern Cameroon	Koffi et al. (2011)
Symphonia globulifera	Clusiaceae	251	LG, C	nSSRs, cpDNA	Southwest Cameroon, Sao Tome[b]	Budde et al. (2013)
Erythrophleum spp.	Fabaceae	648	UG, LG	nSSRs, cpDNA	Northern LG[a], Southern LG[a]	Duminil et al. (2013)
Khaya senegalensis	Maliaceae	507	LG	nSSRs, cpDNA	Southwest Cameroon/Southeast Nigeria, West African river systems[a]	Sexton et al. (2015)
Distemonanthus benthamianus	Fabaceae	429	UG, LG	nSSRs	Dahomey Gap[b]	Demenou et al. (2016)
Chasmanthera dependens	Minispermaceae	139	UG, LG	AFLPs, cpDNA	Dahomey Gap[b]	Iloh et al. (2017)

[a] Refugium not specified within the Upper Guinea, Lower Guinea, or Congolia domains.

[b] Refugium outside the Upper Guinea, Lower Guinea, and Congolia domains.

The LG was included as a sampling location in all studies, nine included the UG, while three included the C. The disproportionate sampling effort is probably caused by the general inaccessibility of the C, and the anthropogenic degradation of the UG such that some species are now difficult to find (Hardy et al., 2013). It is not surprising, therefore, that most populations with high allelic richness, and unique haplotypes and/or alleles (hereafter referred to as refugial populations) were detected in the LG. All studies identified at least a refugial population in the study area (Table 1). There was some congruence in results of studies of the same species, particularly for refugium located in the LG, such as southwest Cameroon. Genetic studies on *A. klaineana* (Muloko-Ntoutoume et al., 2000; Born et al., 2011), show refugial populations in the LG despite differences in genetic markers. The two studies show that the refugial populations for *A. klaineana* were located in the Cristal Mountains in northern Gabon, the Doudou Mountains in western Gabon, and the Chaillu Massif Mountains in south-central Gabon. Likewise, *I. gabonensis* refugial populations were identified in southern Cameroon and southeastern Nigeria (Lowe et al., 2000; Lowe et al., 2010).

Commonly, refugial populations of different species were located in the same refugia. For example, refugial populations of *Distemonanthus benthamianus* (Debout et al., 2011), *Santiria trimera* (Koffi et al., 2011), and *Irvingia gabonensis* (Lowe et al., 2000; Lowe 2010) were located in the southern Cameroon refugium, while the southwest Cameroon putative refugium had refugial populations of *Symphonia globulifera* (Budde et al., 2013), *Coffea canephora* (Gomez et al., 2009), *Greenwayodendron suaveolens* (Dauby et al., 2010), *Vitellaria paradoxa* (Fontaine et al., 2004), and *Milicia excelsa* (Dainou et al., 2010).

When sampling was expanded outside the three domains, refugial populations were detected in Sao Tome (Budde et al., 2013), an island some 200 km off the cost of Gabon, the Dahomey Gap (Demenou et al., 2016; Iloh et al., 2017), the Cape Floristic Province, and east

African coast (Linder, 2001; Kuper et al., 2004). Refugial populations were also identified along African river systems (eg. Sexton et al., 2015), outside the putative Pleistocene refugia delineated by Maley (1996). Significant genetic differentiation was generally detected, and there were distinct gene pools in the domains. For example, three distinct gene pools were detected for *Erythrophylum* spp. corresponding to the three domains (Duminil et al., 2013), while three of five gene pools observed for *C. canephora* were in the LG (Gomez et al., 2009).

1.3 Discussion

Genetic studies of tropical African tree species show that refugial populations were located in postulated Pleistocene refugia. Combined with palynological studies showing oscillations in vegetation type between rain forest and savanna species (Hooghiemstra and Agwu 1988; Bengo and Maley, 1991; Maley, 1991; Giresse et al., 1994), the congruence of genetic data and studies of species richness strengthens the forest refuge origin of high biodiversity in tropical Africa. The presence of distinct gene pools between tree populations, together with a concentration of unique haplotypes and alleles in particular areas, suggest populations expanded fairly recently. These results substantially weaken a pre-Pleistocene differentiation of these African tree species. An expectation under a pre-Pleistocene differentiation scenario would be a fairly uniform distribution of alleles and haplotypes across the entire landscape, unless there are genetic barriers. Although some genetic barriers were detected, such as the Adamawa highlands (Allal et al., 2011), and the Dahomey Gap (Fontaine et al., 2004; Gomez et al., 2009), they could not explain much of the phylogeographic pattern obtained for all species.

The identification of refugial populations of different species in a single putative Pleistocene refugium, such as in the southwest Cameroon refugium (Fontaine et al., 2004;

Gomez et al., 2009; Dauby et al., 2010; Dainou et al., 2010; Budde et al., 2013), and the southern Cameroon refugium (Lowe et al., 2000; Lowe 2010; Debout et al., 2011 Koffi et al., 2011), suggests that these putative refugia harbored many species during cold and dry spells. The refugia possibly sustained the high species richness in tropical Africa.

In some putative refugia, such as the southwest Ghana refugium, however, only a few refugial populations were identified. In the southwest Ghana refugium, refugial populations of only two species were identified (Fontaine et al., 2004; Dainou et al., 2010). While such a result may be caused by inadequate sampling, it is also plausible that species dispersed at different speeds when climate became warmer and mesic, resulting in different genetic signatures. The African mahogany, *K. senegalensis*, showed very low genetic differentiation, and unique haplotypes and alleles were extensively distributed along West African river systems (Sexton et al., 2015). Long distance gene flow in *K. senegalensis* is enabled by wind pollination and wind dispersal of seeds. The northeasterly trade winds coincide with the flowering and fruiting of the species, and the small, flat, and winged seed enables wind propulsion throughout the range of the species (Sexton et al., 2015). The fruit is also buoyant and can be carried over long distances by the labyrinth river system in tropical Africa. Such rapid and extensive dispersal mechanisms may produce a different genetic signature than that produced by a species with restricted dispersion. The rate of gene flow from refugia might have been species-specific, resulting in different genetic signatures.

The Forest refuge hypothesis asserts that small and isolated populations in refugia often experienced genetic drift and inbreeding (Haffer, 1997). Genetic drift is a change in allelic frequency in a population because sexual reproduction only allows a sample of alleles from parents to be passed to offspring (Ellstrand and Elam, 1993). In a large population, the sample of alleles passed to offspring will be a close representative of those of the parents, and

therefore changes in allele frequency due to drift are small. Inbreeding is mating of related individuals, and in plants, it increases when gene flow through pollination and seed dispersal is restricted, as is often the case in isolated populations. Genetic drift and inbreeding in small populations often result in differentiation, a precursor to speciation. An expectation under the Forest refuge hypothesis is high population differentiation because of genetic drift and inbreeding in the isolated refugia (Haffer, 1969). The high population differentiation and admixture in contact zones observed in most tropical African tree populations is likely a result of repeated historical isolation in and expansion from refugia.

The presence of refugial populations along African river systems suggests that rivers also remained climatically stable for riverine species during the climate oscillations of the Pleistocene. Fluvial refugia were suggested by Colyn et al. (1991), and refugial populations of *K. senegalensis* were detected along river systems. This observation suggests that *K. senegalensis* persisted along rivers when the climate was colder and drier. The identification of refugial populations of *C. dependens* and *D. benthamianus* in the Dahomey Gap is particularly notable because this is a stretch of savanna vegetation separating the UP and the LG. The reduced ecological requirements of *C. dependens* possibly allowed it to survive in the Dahomey Gap (Iloh et al., 2017), in contrast to other tropical tree species. The low genetic diversity in *D. benthamianus* in the Dahomey Gap suggests the individuals survived in small and isolated populations susceptible to genetic drift (Demenou et al., 2016), since the Dahomey Gap does not adequately support tropical tree species. The presence of refugial populations outside the Guineo-Congolia rain forest, such as in the Cape Floristic Province, east African coast (Linder, 2001; Kuper et al., 2004), and Sao Tome (Budde et al., 2013), suggests that refugia were not restricted to African rain forests, but must have been widespread to include other biomes, such as the Afromontane of east Africa.

A complimentary method to detect the restriction of populations into refugia is to hindcast their current distributions to the time of interest using Ecological Niche Models (ENMs). ENMs use environmental variables at species occurrence points to predict their distribution (Guisan and Zimmermann, 2000; Soberon and Nakamura, 2009). A recent study of hindcasting the distribution of *Afzelia quanzensis* to the Pleistocene showed a niche decline (Jinga et al., in preparation). The decline in the distribution of the species when hindcasted to the Pleistocene suggests the species might have been restricted to climatically stable areas, since these areas seem to have been widespread in sub-Saharan Africa. Populations in climatically stable areas may be targeted for conservation since they will have high genetic diversity.

1.4 Conclusions

The Forest refuge hypothesis was proposed to explain biodiversity in neotropics, but has largely been abandoned. However, my review of studies of African tree species shows that biodiversity in the continent may be explained by Haffer's hypothesis. Genetic studies of tree species in tropical Africa have largely shown that populations with high allelic richness, and unique alleles or haplotypes are located in areas of high species richness and endemism which have been delineated as putative Pleistocene refugia. Together with palynological studies, the occurrence of refugial populations in areas of high endemism suggests that tree species survived in isolated refugia that remained climatically stable during the Pleistocene, and expanded when the climate became warmer and mesic. Repeated expansion and contraction of long-lived tree species may be revealed by their genetic signature. The Forest refuge hypothesis seem to be strongly validated by genetic signature of tropical African tree species, unlike in tropical South America. Putative Pleistocene refugia have also been suggested in other African biomes, apart from the rain forests. Refugia might have been

widespread on the African continent. Genetic drift and inbreeding acting in refugia might have resulted in high population differentiation observed in most tropical African tree populations. The high species richness and endemism observed in refugia is likely a result of speciation in isolated populations. When populations are isolated, mutations and adaptation in the absence of gene flow accelerates speciation. Putative Pleistocene refugia in Africa should be priority areas for conservation since they have high genetic diversity, high species richness, and are centers of endemism. In order to capture all the extant genetic diversity, individuals in putative refugia should be strongly considered as candidates for germplasm conservation, such as in the creation of seed banks, and *ex situ* preservation sites.

1.5 References

Aitken, S. N., Yeaman, S., Holliday, J. A., Wang, T. and Curtis-McLane, S. 2008. Adaptation, migration or exrtipation: climange change outcomes for tree populations. *Evolutionary Applications*, **1**: 95-111.

Allal, F., Sanou, H., Millet, L. et al. 2011. Past climate changes to explain the phylogeography of *Vitellara paradoxa* over Africa. *Heredity*, **107**: 174-186.

Ayres, J.M.C. and Clutton-Brock, T.H. 1992. River boundaries and species range size in Amazonian primates. *American Naturalist*, **140**: 531-537.

Bengo, M. D. and Maley, J. 1991. Analysis of pollen flux on the southern margin of the Gulf of Guinea in the last 135 000 years. *Comptes Rendus de I'Academie des Sciences, Paris, II*, **313**: 843-849.

Born C., Alvarez, N., McKey, D. et al. 2011. Insights into the biogeographical history of the Lower Guinea Forest Domain: evidence for the role of refugia in the intraspecific differentiation of *Aucoumea klaineana*. *Molecular Ecology*, **20**: 131-142.

Brower, A. V. Z and Egan, M. G. 1997. Cladistic analysis of Heliconius butterflies and relatives (Nymphalidae: Heliconiiti): a revised phylogenetic position for Eueides based on sequences from mtDNA and a nuclear gene. *Proceedings of the Royal Society of London B*, **264**: doi:10.1098/rspb.1997.0134.

Budde, K. B., Gonzalez-Martinez, S. C., Hardy, O. J. and Heuertz, M. 2013. The ancient tropical rainforest tree Symphonia globulifera L. F. (Clusiaceae) was not restricted to postulated Pleistocene refugia in Atlantic Equatorial Africa. *Heredity*, **111**: 66-76.

Bush, M. B. and de Oliveira, P. E. 2006. The rise and fall of the Refugial hypothesis of Amazonian speciation: a paleoecological perspective. *Biota Neotropica*, **6**: doi.org/10.1590/S1676-06032006000100002.

Bush, M. B., Weimann, M., Piperno, D. R., Liu, K. and Colinvaux, P. A. 1990. Pleistocene temperature depression and vegetation change in Ecuadorian Amazonia. *Quaternary Research*, **34**: 330-345.

Chen, I-C., Hill, J. K., Ohlemuller, R., Roy, D. B. and Thomas, C. D. 2011. Rapid range shifts of species associated with high levels of climate warming. *Science*, **333**: 1024-1026.

Colinvaux, P. 1993. Pleistocene biogeography and diversity in tropical forests of South America. In: *Biological Relationships between Africa and South America* (Goldblatt, P. ed.), pp. 473-99. Yale University Press, New Haven.

Colinvaux, P. A., de Oliveira, P. E. and 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.

Colyn, M., Gautier-Hion, A. and Verheyen, W. 1991. A re-appraisal of palaeoenvironmental history in Central Africa: evidence for a major fluvial refuge in the Zaire Basin. *Journal of Biogeography*, **18**: 403-407.

Dainou, K., Bizoux, J-P., Doucet, J-L., Mahy, G., Hardy, O. J. and 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**: 4462-4477.

Dauby, G., Duminil, J., Heuertz, M. and Hardy, O. J. 2010. Chloroplast DNA polymorphism and phylogeography of a Central African tree species widespread in mature rainforests: *Greenwayodendron suaveolens* (Annonaceae). *Tropical Plant Biology*, **3**: 4-13.

Debout, G. D. G., Doucet, J-L. and 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**: 88-99.

Demenou, B. B., Pineiro, R. and Hardy, O. J. 2016. Origin and history of the Dahomey Gap separating West and Central African rain forests: insights from the phylogeography of the legume tree *Distemonanthus benthamianus*. *Journal of Biogeography*, **43**: 1020-1031.

Diamond, A. W. and Hamilton, A. C. 1980. The distribution of forest passerine birds and Quaternary climatic change in tropical Africa. *Journal of Zoology*, **191**: 379-402.

Duminil, J., Brown, R. P., Ewedje, E-E., Mardulyn, P., Doucet, J-L. and Hardy, O. J. 2013. Large-scale pattern of genetic differentiation within African rainforest trees: insights on the roles of ecological gradients and past climate changes on the evolution of *Erythrophleum* spp. (Fabaceae). *BMC Evolutionary Biology*, **13**: 195.

Dupont, L. M. and Weinelt, M. 1996. Vegetation history of the savanna corridor between the Guinean and the Congolian rain forest during the last 150000 years. *Vegetation History and Archaeobotany*, **5**: 273-92.

Ellstrand, N. C. and Elam, D. R. 1993. Population genetic consequences of small population size: Implications for plant conservation. *Annual Review of Ecology and Systematics*, **24**: 217-242.

Emsley, M.G. 1965. Speciation in *Heliconius* (Lep., Nymphalidae): Morphology and geographic distribution. *Zoologica*, **50**: 191-254.

Endler, J. 1982. Pleistocene forest refuges: fact or fancy? In: *Biological diversification in the tropics* (Prance, G. T. ed.), pp. 179-200. Columbia University Press, New York.

Fontaine, C., Lovett, P. N., Sanou, H., Maley, J. and Bouvet, J-M. 2004. Genetic diversity of the shea tree (*Vitellaria paradoxa* C. F. Gaerth), detected by RAPD and chloroplast microsatellite markers. *Heredity*, **93**: 639-648.

Giresse, P., Maley, J. and Brenac, P. 1994. Late Quaternary palaeoenvironments in the Lake Barombi Mbo (West Cameroon) deduced from pollen and carbon isotopes of organic matter. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **107**: 65-78.

Gomez. C., Dussert, S., Hamon, P., Hamon, S., Kochko, A. and Poncet, V. 2009. Current genetic differentiation of *Coffea canephora* Pierre ex A. Froehn in the Guineo-Congolian African zone: cumulative impact of ancient climatic changes and recent human activities. *BMC Evolutionary Biology*, **9**: 167.

Grubb, P. 1978. Patterns of speciation in African mammals. Bulletin of Carnegie Museum of Natural History, **6**: 152-167.

Guisan, A. and Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**:147–186.

Haffer, J. 1969. Speciation in Amazonian forest birds. *Science*, **165**: 131-137.

Haffer, J. 1985. Avian zoogeography of the Neotropical lowlands. *Ornithological Monographs*, **36**: 113-146.

Haffer, J. 1997. Alternative models of vertebrate speciation in Amazonia: an overview. *Biodiversity and Conservation*, **6**: 451-476.

Haffer, J. and Prance, G. T. 2001. Climatic forcing of evolution in Amazonia during the Cenozoic: On the refuge theory of biotic differentiation. *Amazoniana*, **16**: 579-607.

Hall, J. B. and Swaine, M. D. 1981. *Distribution and ecology of vascular plants in a tropical rain forest: Forest vegetation in Ghana*, W. Junk, The Hague.

Hamilton, A. C. 1976. The significance of patterns of distribution shown by forest plants and animals in tropical Africa for the reconstruction of upper Pleistocene palaeoenvironments: a review. *Palaeoecology of Africa*, **9**: 63-97.

Hardy, O J., Born, C., Budde, K. et al. 2013. Comparative phylogeography of African rain forest trees: A review of genetic signatures of vegetation history in the Guineo-Congolian region. *Comptes Rendus Geoscience*, **345**: 284-296.

Hershkovitz, P. 1977. *Living New World Monkeys* (*Platyrrhini*). University of Chicago Press, Chicago.

Hooghiemstra, H. and Agwu, C. O. 1988. Changes in the vegetation and trade winds in equatorial northwest Africa 140 000-70 000 yr B.P. as deduced from tow marine pollen records. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **66**: 173-213.

Iloh, A. C., Schmidt, M., Muellner-Riehl, A., Ogundipe, O. T. and Paule, J. 2017. Pleistocene refugia and genetic diversity patterns in West Africa: Insights from the liana *Chasmanthera dependens* (Menispermaceae). *PLOS ONE*, **12**: e0170511.

Kastner, T. P. and Goni, M. A. 2003. Constancy in the vegetation of the Amazon Basin during the late Pleistocene: Evidence from the organic matter composition of Amazon deep sea fan sediment. *Geology*, **31**: 291-294.

Kingdon, J.S. 1971. *East African mammals: an atlas of evolution in Africa*. Academic Press, London.

Koffi, K. G., Hardy, O. J., Doumenge, C., Cruaud, C. and Heuertz, M. 2011. Diversity gradients and phylogeographic patterns in *Santiria trimera* (Burseraceae), a widespread African tree typical of mature rainforests. *American Journal of Botany*, **98**: 254-264.

Kuper, W., Sommer, J. H., Mutke, J. et al. 2004. Africa's hotspots of biodiversity redefined. *Annals of Missouri Botanical Garden*, **91**: 525-535.

Linder, H. P. 2001. Plant diversity and endemism in sub-Saharan tropical Africa. *Journal of Biogeography*, **28**: 169-182.

Liu, K. and Colinvaux, P. A. 1985. Forest changes in the Amazon Basin during the last glacial maximum. *Nature*, **318**: 556-557.

Lowe, A. J., Gillies, C. M., Wilson, J. and Dawson, I. K. 2000. Conservation genetics of bush mango from central/west Africa: implications from random amplified polymorphic DNA analysis. *Molecular Ecology*, **9**: 831-841.

Lowe, A. J., Harris, D., Dormontt, E. and Dawson, I. K. 2010. Testing putative African tropical forest refugia using chloroplast and nuclear DNA phylogeography. *Tropical Plant Biology*, **3**: 50-58.

Maley, J. 1991. The African rain forest vegetation and palaeoenvironments during the late Quaternary. *Climatic Change,* **19**: 79-98.

Maley, J. 1996. The African rain forest: main characteristics of the changes in vegetation and climate from the upper Cretaceous to the Quaternary. *Proceedings of the Royal Society of Edinburgh Section B,* **104**: 31-73.

Mayr, E. and O'Hara, R. J. 1986. The biogeographic evidence supporting the Pleistocene forest refuge hypothesis. *Evolution,* **40**: 55-67.

Moher, D., Liberati, A., Tetzlaff, J. and Altman, D.G. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *The British Medical Journal*, **339**: b2535 doi: 10.1136/bmj.b2535.

Moreau, R. E. 1969. Climatic changes and the distribution of forest vertebrates in West Africa. *Journal of Zoology*, **158**: 39-61.

Moritz, C., Patton, J. l., Scheinder, C. J. and Smith, T. B. 2000. Diversification of rainforest faunas: an integrated molecular approach. *Annual Review of Ecology and Systematics*, **31**: 533-563.

Muloko-Ntoutoume, N., Petit, R. J., White, L. and Abernethy, K. 2000. Chloroplast DNA variation in a rainforest tree (*Aucoumea klaineana*, Burseraceae) in Gabon. *Molecular Ecology*, **9**: 359-363.

Pearson, R. G. and Dawson, T. P. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography,* **12**: 361-371.

Permesan, C. and Yohe, G. 2003. A globally coherent fingerprint of climate change impacts across natural systems. *Nature,* **421**: 37-42.

Petit, J. R. and Hampe, A. 2006. Some evolutionary consequences of being a tree. *Annual Review of Ecology, Evolution, and Systematics,* **37**: 187-214.

Plana, V. 2004. Mechanisms and tempo of evolution in the African Guineo-Congolian rainforest. *Philosophical Transactions of the Royal Society of London B,* **359**: 1585-1594.

Plana, V., Gascoigne, A., Forest, L. L., Harris, D. and Pennington, R. T. 2004. Pleistocene and pre-Pleistocene *Begonia* speciation in Africa. *Molecular Phylogenetics and Evolution,* **31**: 449-461.

Prance, G. T. 1982. Forest refuges: evidence from woody angiosperms. In: *Biological diversification in the tropics* (Prance, G. T. ed.), pp. 137-157. Columbia University Press, New York.

Prigogine, A. 1988. Speciation pattern of birds in the Central African forest refugia and their relationship with other refugia. *Acta XIX Congressus Internationalis Ornithologici,* **2**: 144-157.

Richardson, J. E., Pennington, R. T., Pennington, T. D. and Hollingsworth, P. M. 2001 Rapid Diversification of a species-rich genus of Neotropical rain forest trees. *Science*, **293**: 2242-2245.

Russel, J., Talbot, M. R. and Haskell, B. J. 2003. Mid-Holocene climatic change in Lake Bowumtwi, Ghana. *Quaternary Research*, **60**: 133-141.

Salzmann, U. and Hoelzmann, P. 2005. The Dahomey Gap: an abrupt climatically induced rain forest fragmentation in West Africa during the late Holocene. *The Holocene*, **15**: 190-199.

Sexton, G. J., Frere, C. H., Kalinganire, A. et al. 2015. Influence of putative refugia and biogeographic barriers on the level and distribution of genetic variation in an African savannah tree, *Khaya senegalensis* (Desr.) A. Juss. *Tree Genetics and Genomes*, **11**:103.

Soberon, J. M. and Nakamura, N. 2009. Niches and distributional areas: concepts, methods, and assumptions. *Proceedings of the National Academy of Sciences, USA*, **106**: 19644–19650.

Sosef, M. S. M. 1991. New species of Begonia in Africa and their relevance to the study of glacial rain forest refuges. *Wageningen Agricultural University Press*, **91**: 120-151.

Talbot, M. R., Livingstone, D. A., Palmer, P. G., Maley, J., Melack, J. M., Delibrias, G. and Guliksen, S. 1984. Preliminary results from sediment cores from Lake Bosumtwi, Ghana. *Palaeoecology of Africa*, **16**: 173-192.

Walther, G-R., Post, E., Convey, P. et al. 2002. Ecological responses to recent climate change. *Nature,* **416**: 389-395.

White, F. 1979. The Guineo-Congolian region and its relationships to other phytochoria. *Bulletin du Jardin Botanique National de Belgique*, **49**: 11-55.

Wilson, O. 1988. *Biodiversity*. The National Academies Press, Washington D. C.

2. MAXENT MODELING OF PAST, PRESENT, AND FUTURE DISTRIBUTIONS OF *AFZELIA QUANZENSIS* (POD MAHOGANY) IN ITS NATIVE SUB-SAHARA AFRICAN RANGE

2.1 Introduction

There is strong evidence that increased concentration of atmospheric greenhouse gases, particularly carbon dioxide and methane, due to human activities, has started to markedly modify global climate (IPCC, 2007). On average, global surface temperatures increased by 0.61 (± 0.16 ºC) between 1861 and 2000 (Folland et al., 2001). The correlation between atmospheric concentration of greenhouse gases and temperature is well established (Petit et al., 1999). The recently measured atmospheric concentration of carbon dioxide of nearly 404.49 ppm (Scripps Institution of Oceanography, 2017, https://scripps.ucsd.edu/programs/keelingcurve/) is unparalled in history, and is likely to accelerate the rate of climate change. The surface temperature of the earth is projected to increase by between 1 ºC to 3.7 ºC by 2100, depending on the choice of emission scenario or Representative Concentration Pathway (RCP) (IPCC, 2014). Climate change affects the global distribution of organisms, particularly plants because their distribution is directly affected by temperature and water availability.

Evidence from palynology and fossil records have shown that, in the past, plants have responded to climate change by remaining within the changing climate through adaptation or tolerance, and migrating to track suitable climate (Bush et al., 2004; Jump and Penuelas, 2005; Petit et al., 2008). Recent studies have shown shifts in the distribution of plant species in response to climate change (e.g Kelly and Guolden, 2008; Parolo and Rossi, 2008; Chen et

al., 2011; Crimmins et al., 2011; Feeley, 2012; Bitencourt et al., 2016). However, the magnitude of the current rate of climate change may require plants to adapt rapidly or migrate faster, which they might be unable to do. Therefore, the current rate of climate change may threaten biodiversity and may result in extinctions.

Xerophytes or desert plants may expand their habitat in a warming environment because of their adaptations to withstand elevated temperatures and dry environments. Conversely, the suitable range of temperate, alpine and hydrophytic plants may decline because these plants lack the evolutionary adaptations to conserve water, and may be unable to develop such adaptations in time to catch up with the rapidity of the current climate change. To determine the impact of climate change, it is important to model species-specific distributions because variability in traits results in unique estimations of niche availability.

An established method to predict past and future distributions of plants is the use of species distribution models (SDMs) or ecological niche models (ENMs) (Guisan and Zimmermann, 2000; Crimmins et al., 2013). These models correlate environmental variables with species occurrence data using statistical or machine learning procedures. Despite their limitations, such as dispersal assumptions (Thuiller et al., 2004), assumption of constancy of species' niches over time, and the exclusion of species interactions in modeling (Roberts and Hamann, 2012), SDMs or ENMs have been shown to be useful tools for species distribution modeling based on occurrence records and climate variables (Soberon and Nakamura, 2009).

Among the commonly used ENMs is the maximum entropy method (Maxent) (Phillips et al., 2006). The Maxent method uses presence-only records and correlates them to topographic, climatic, edaphic, biogeographic, or remotely sensed variables observed at the

occurrence coordinates, to predict species distributions. A niche-based model represents an approximation of a species' ecological niche in the examined environmental dimensions. A species' fundamental niche consists of the set of all conditions that allow for its long-term survival, whereas its realized niche is a subset of the fundamental niche that it actually occupies (Hutchinson, 1957). A general assumption of ENMs is that environmental conditions at the occurrence sites constitute samples from the realized niche (Phillips et al., 2006), particularly for non-vagile organisms. A niche-based model, therefore, represents an estimation of the realized niche in the study area and environmental variables under consideration. In that regard, the Maxent method has been shown to have high predictive power (Phillips et al., 2006; Pearson et al., 2007; Ramírez-Amezcua et al., 2016) compared to other models that use presence-only data.

Afzelia quanzensis (Fabaceae) is a low-altitude (up to 1 300 m), medium to large tree species (15-35 m in height) which naturally occurs in sub-Saharan Africa, and is part of Miombo woodlands (Munyanziza and Oldeman, 1996). Miombo woodlands are predominantly made up of tree species in the subfamily Caesalpinioideae, and apart from *Afzelia*, other genera of the woodlands include *Isoberlinia*, *Julbernardia* and *Brachystegia* (Frost, 1996). The tree species thrive in relatively warm temperatures of 18.0-23.1°C, and dry environments with annual precipitation of 710-1365 mm (Frost, 1996). Most of the rainfall falls during the summer, and trees shed off leaves during the dry winter months. The soils of the woodlands are generally well-drained sandy, with poor nutrient composition.

Afzelia quanzensis has a variety of uses, which include production of household furniture, roofing planks, fencing posts, railway sleepers, and woodcarvings (Germishuizen *et al*., 2005). Extracts from leaves, roots, and bark of *A. quanzensis* have been used as

traditional medicine to treat a myriad of health problems (Germishuizen *et al*., 2005), although the pharmacological efficacy of the extracts still needs to be scientifically tested. The species has been heavily logged in its native range (Gerhardt and Todd, 2009).

The aims of this study were to construct a niche-based model for *A. quanzensis* from occurrence records, and provide projections of the past and future distributions under different RCPs and Global Circulation Models (GCMs). I sought to answer the following specific questions: (i) What is the current distribution of *A. quanzensis*? (ii) What was the distribution of *A. quanzensis* during the mid-Holocene? (iii) What is the projected distribution of *A. quanzensis* in 2050 and 2070? (iv) Are there management implications that arise from the forecasted distributions? The current distribution of *A. quanzensis* is only estimated from archived sampling geographic coordinates. Distribution maps are important for locating, delineating and designing conservation areas. Distribution maps help in identifying suitable areas for reintroductions, establishment of corridors, and botanical exploration. Gaps in the distribution of species are identified and, possibly, their causes determined. The identification of environmental variables that significantly affects the distribution of species helps to manage unsuitable or degraded sites (Manel et al., 2001). Hindcasting and forecasting SDMs may allow planning of mitigation against the impacts of climate change.

2.2 Materials and Methods

2.2.1. Occurrence records and environmental variables

Occurrence records of *A. quanzensis* were obtained from the Global Biodiversity Information Facility (GBIF) (http://www.gbif.org) and Tropicos (http://www.tropicos.org)

(Missouri Botanical Garden) online data portals. Automatic filters on the GBIF portal were used to remove records with georeferencing errors. Occurrence records were also obtained from the Zimbabwe National Herbarium (SRGH). All records were manually checked to remove duplicates. To construct models, precipitation, maximum and minimum temperatures, and 19 bioclimatic variables (Appendix A) derived from precipitation and temperature, were correlated to the occurrence records. All variables were obtained from the WorldClim dataset (www.worldclim.org) version 1.4 (Hijmans et al., 2005), interpolated to 30 seconds resolution. Bioclimatic variables are important abiotic indicators of the distribution of plants, and have been used widely in ENMs (eg. Adeyemi et al., 2012; Ramirez-Villegas et al., 2014; Ramírez-Amezcua et al., 2016).

2.2.2 Ecological niche modeling and model evaluation

A maximum entropy algorithm, Maxent version 3.3.3 (Phillips et al., 2006), was used for ecological niche modeling and projections. Model evaluation statistics were produced from 10 replicate model evaluation runs using the 19 bioclimatic variables as well as precipitation, and maximum and minimum temperatures. For these runs, 75% of the total records were used to build the model and the remaining 25% were used for testing. The random seed option was enabled, and number of iterations were increased to 1000 to allow adequate time for model convergence. The number of background points was set at 10 000, where model performance has been shown to be reliable (Phillips and Dudik, 2008).

The full set of presence records was used to build the final model to obtain the best estimate of the current distribution and for projections. Receiver Operating Characteristic

(ROC) analysis from the 10 runs was used to measure model performance. The area under the ROC curve (AUC) is a measure of model performance in terms of sensitivity (correctly predicted presences) versus specificity (correctly predicted absences) (Phillips et al., 2006). Despite having shortcomings (Lobo et al., 2008; Warren et al., 2013), the AUC is still a useful measure of model performance (Graham et al., 2008; Elith and Graham, 2009; VanDerWal et al., 2009; Braunisch et al., 2013), and has been used widely. The value of the AUC varies from 0.5 to 1.0. An AUC value of 0.5 denotes random prediction, and the closer the value is to 1.0, the better the model performance (Phillips and Dudik, 2008; Adeyemi et al., 2012). The final model with the full set of presence records was jack-knifed to check variable importance.

2.2.3 Past and future projections

For future distribution projections, I used all four RCPs proposed by the IPCC to years 2050 and 2070. The RCPs predict climate variables, particularly temperature, under different greenhouse gas emission scenarios. Between 2046 to 2065, global mean surface temperatures are predicted to rise by 1.0 ℃, 1.4 ℃, 1.3 ℃, 2.0 ℃ under RCP2.6, RCP4.5, RCP6.0, RCP8.5, respectively. Between 2081 to 2100, the global mean surface temperatures are predicted to rise by 1.0 ℃, 1.8 ℃, 2.2 ℃, and 3.7 ℃ under RCP2.6, RCP4.5, RCP6.0 and RCP8.5, respectively (IPCC, 2013). For past distributions, the current model was projected to mid-Holocene, about 6 000 years ago.

Climate variables used in all projections were downloaded from two Global Climate Models (GCMs), and these are the Community Climate System Model version 4.0 (CCSM4.0) (Lawrence et al., 2012), and the Institut Pierre Simon Laplace Climate Model

version 5A (IPSL-CM5A-LR) (Mignot and Bony, 2013). Niche loss or gain was calculated as the difference between the size of the current predicted distribution and the size of the projected distribution. The species range change was calculated as the percentage of the niche gain or loss compared to the current range. A threshold on the probability of niche suitability used for calculations was determined by reading off the shortest distance to the top-left corner of the ROC plot. The point on the top-left corner of the ROC plot is where specificity and sensitivity are 100 % (Cantor et al., 1999). Determining the threshold from the ROC plot was shown to be more accurate compared to other methods (Liu et al., 2005). Niche gain and loss were calculated in ArcMap version 10.4.1 (ESRI, 2017) using the Conversion, Data Management and Spatial Analyst tools.

2.3 Results

A total of 170 occurrence records of *A. quanzensis* were retrieved from the GBIF and Tropicos online data portals, after removing duplicates. Herbaria that contributed occurrence records to GIBF include PRECIS, South Africa, Royal Botanic Gardens, UK, The Field Museum of Natural History, USA, Geneva Herbarium, Switzerland, Real Jardin Botanico, Spain, IICT Herbario, Portugal, and Herbarium Senckenbergianum, France. Nine records were obtained from the Zimbabwe National Herbarium. Overall, 179 occurrence records were used for model development, evaluation, and projection.

The 10 models for evaluation showed a high level of performance when compared to random prediction. Mean test AUC value was (0.977 ± 0.002) (Figure 2.1). The evaluation models correctly predicted most of the test locations.

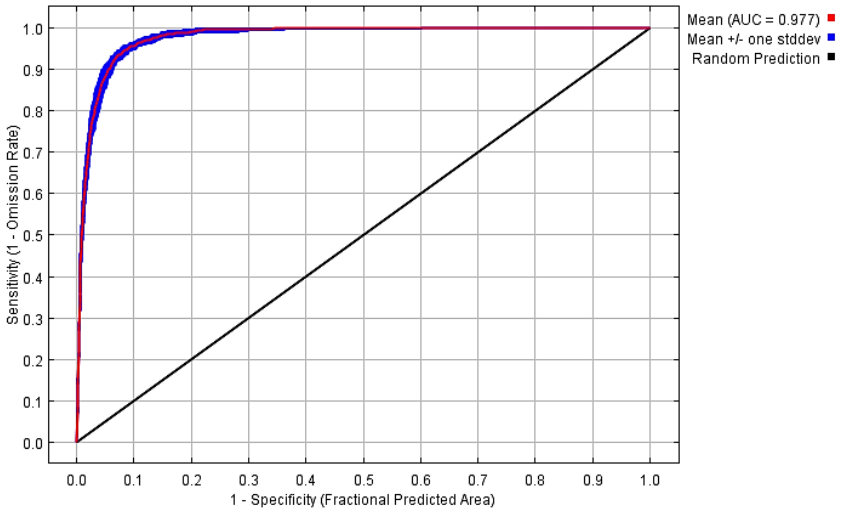


Figure 2.1**.** Mean training AUC of 10 evaluation models for ecological niche modeling of *Afzelia quanzensis*

The modeled current distribution of *A. quanzensis* shows occurrence in south of the Sahara, including Kenya, Malawi, Mozambique, Uganda, Botswana, Zimbabwe, Angola, Namibia, and South Africa (Figure 2.2 and Figure 2.3). Precipitation of the twelfth (36.1%), eleventh (6.1%), tenth (5.2%) and ninth (8.3%) months contributed significantly to the model. Temperature seasonality (bio4) (10.1%) and precipitation of the driest month (bio14) (6.4%) also contributed significantly to the model (Table II, Appendix B). The jack-knife test showed that the environmental variable with the highest gain when used in isolation is precipitation of the twelfth month, and therefore it has the most useful information by itself. The environmental variable that decreases the gain the most when it is excluded is temperature seasonality, which therefore appears to have the most information that is absent in the other variables.

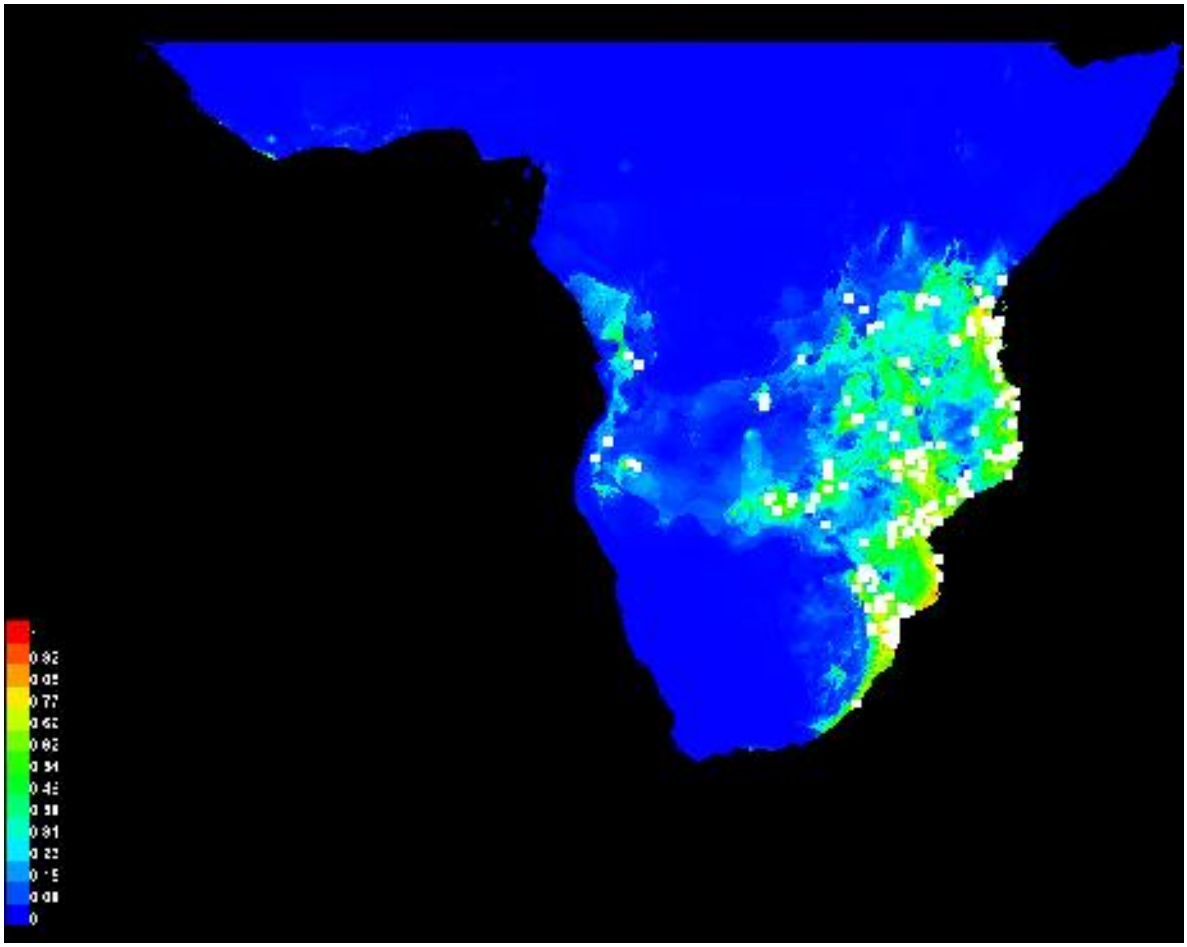


Figure 2.2. Map of predicted current distribution of *Afzelia quanzensis* from Maxent modeling

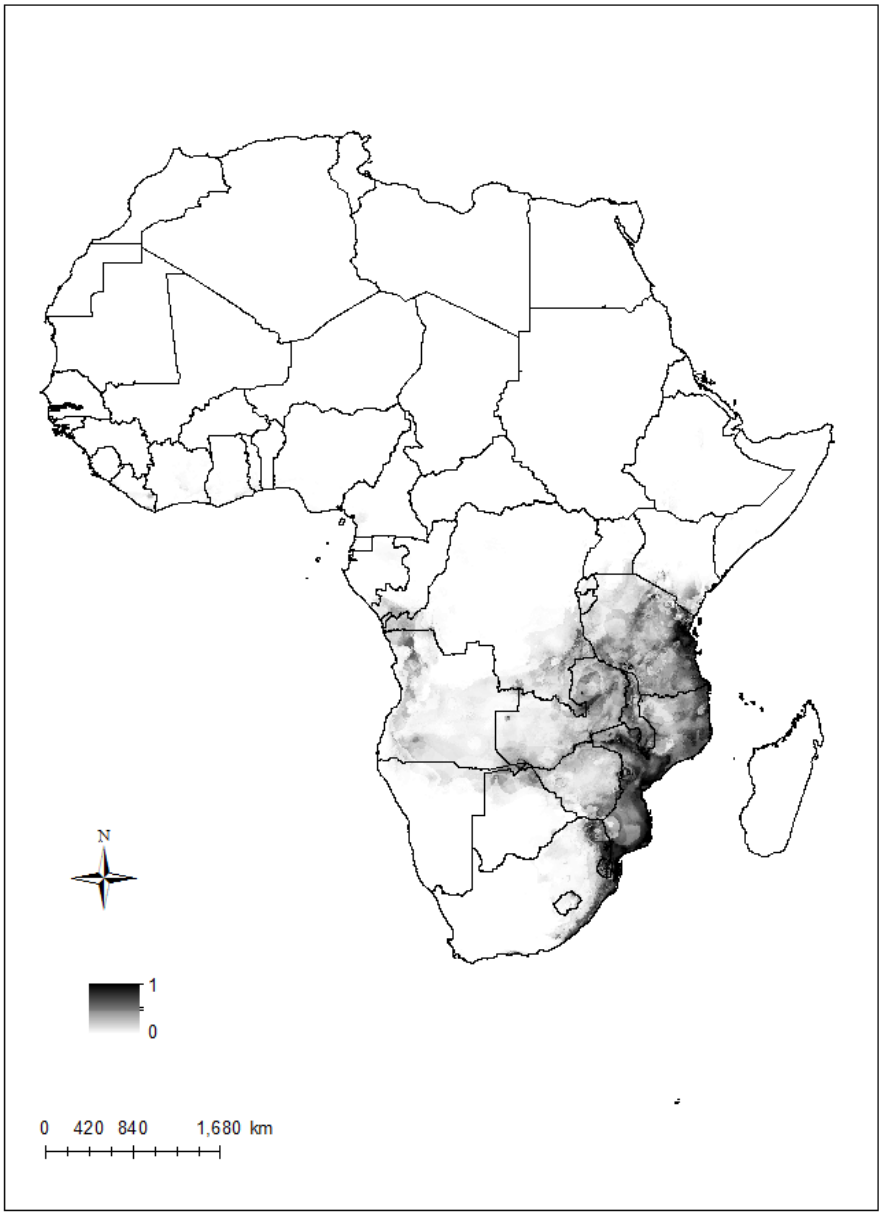
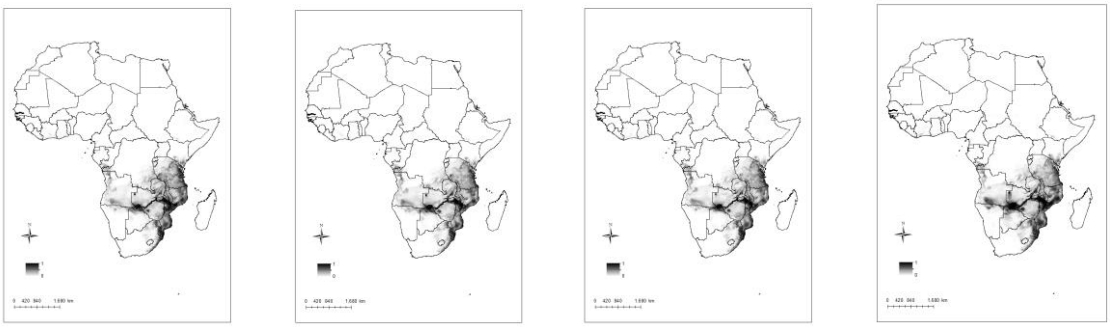


Figure 2.3. Map of predicted current distribution of *Afzelia quanzensis* with political boundaries

Forecasting the current distribution resulted in niche increase while hindcasting resulted in niche decrease (Figure 2.4). The distribution of *A. quanzensis* decreased when the model was projected to the mid-Holocene for both GCMs. The distribution decreased by 41% and 20 % for the CCSM4.0 and IPSL-CM5A-LR models, respectively (Figure 2.5). Future projections showed an increase in the distribution of *A. quanzensis* for both GCMs. Generally, the IPSL-CM5A-LR future niche expansions were greater than for CCSM4.0, but both GCMs produced a similar trend. The highest increase was observed under IPSL-CM5A-LR for year 2070 under RCP8.5, where the suitable range of the species increased by 307 % (Figure 2.5). The least niche expansion was observed in 2070, under IPSL-CM5A-LR, RCP2.6, where the niche increased by 24 %.

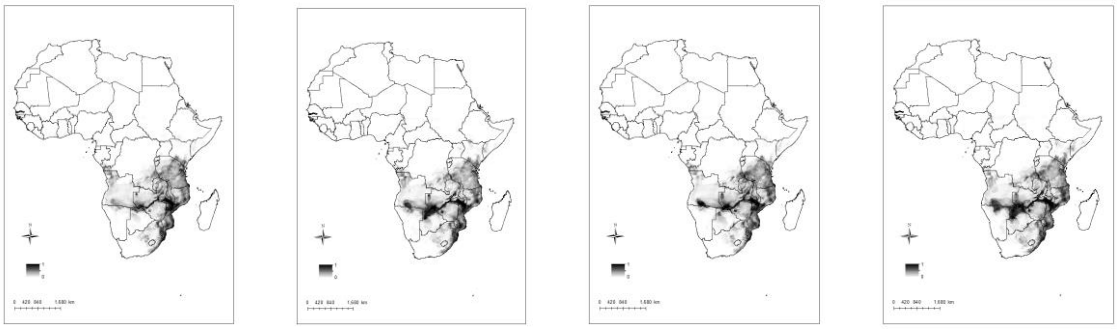
A. 2050 and CCMS4.0

RCP2.6 (48%) RCP4.5 (32%) RCP6.0 (36%) RCP8.5 (92%)



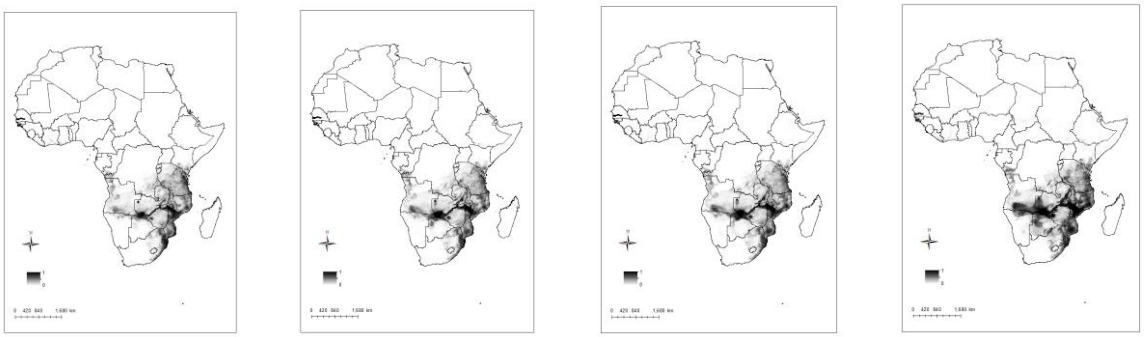
B. 2050 and IPSL-CM5A-LR

RCP2.6 (34%) RCP4.5 (77%) RCP6.0 (86%) RCP8.5 (124%)



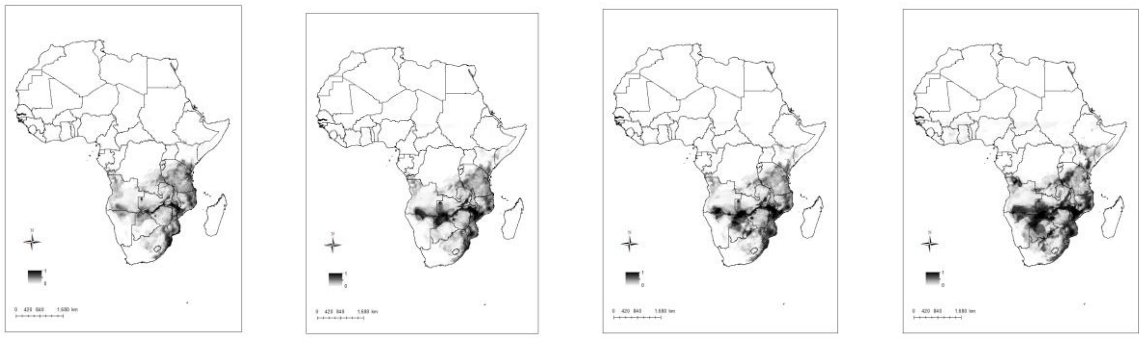
C. 2070 and CCSM4.0

RCP2.6 (27%) RCP4.5 (58%) RCP6.0 (71%) RCP8.5 (213%)



D. 2070 and IPSL-CM5A-LR

RCP2.6 (24%) RCP4.5 (126%) RCP6.0 (150%) RCP8.5 (307%)



E. Mid-Holocene

CCSM4.0 (-41%) IPSL-CM5A-LR (-20%)

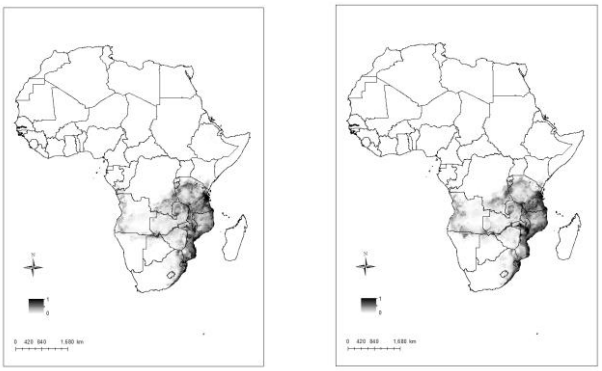


Figure 2.4. Projected distribution maps of *Afzelia quanzensis* showing range changes under two Global Climate Models, CCSM4.0 and IPSL-CM5A-LR, when forecasted to 2050 and 2070, and hindcasted to the mid-Holocene

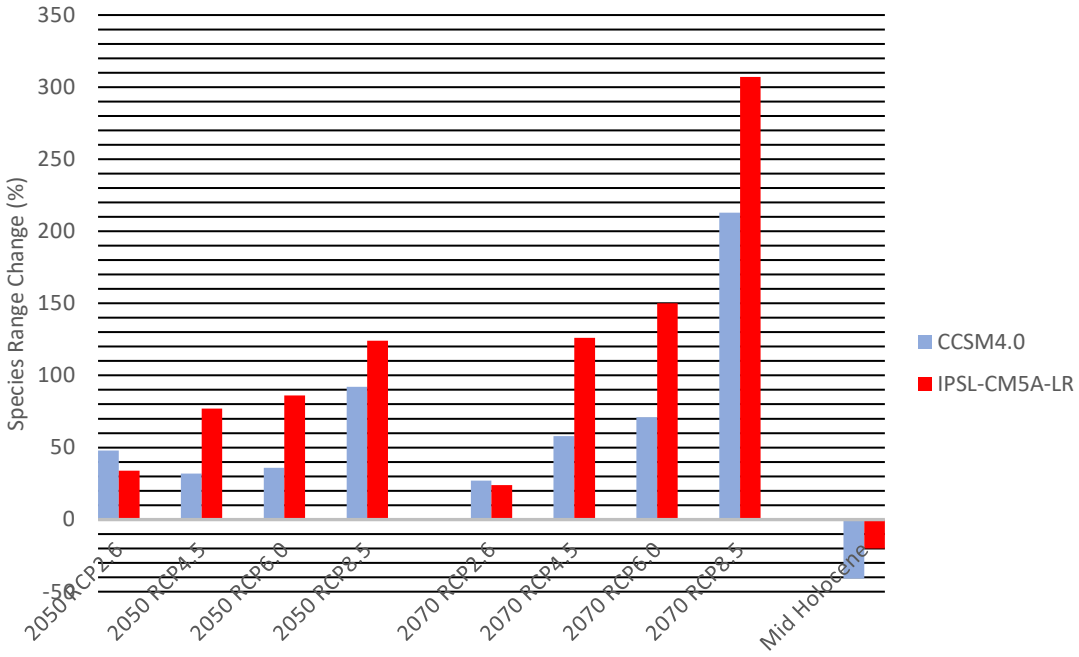


Figure 2.5. *Afzelia quanzensis* niche expansion and contraction under two Global Climate Models and different Representative Concentration Pathways

2.4 Discussion

The modeled current distribution of *A. quanzensis* includes most countries where the species has been reported to occur in several studies (eg. Gathua, 2000; Braedt and Standa-Gunda, 2000; Germishuizen, et al., 2005; Gerhardt and Nemarundwe, 2006; Gerhardt and Todd, 2009). *Afzelia quanzensis* is distributed largely in miombo woodlands of central, eastern and southern Africa. Miombo woodlands thrive in seasonally hot and dry environments (Gerhardt and Nemarundwe 2006). In Zimbabwe, the current distribution shows that *A. quanzensis* is largely found in agro-ecoregion 4, along the Zambezi and Save valleys, characterized by mean annual precipitation of 450-600 mm, and average temperatures of 18-24 ℃ (Mugandani et al., 2012). The distribution is also along the

relatively hot and dry Limpopo Province of South Africa, and large swathes of Tanzania, Mozambique and Zambia. Because the species grows well in relatively hot and dry environments, it is not surprising that the current distribution largely avoids the equatorial region of the DRC, the extremely hot and dry Namib desert of Namibia, and the Mediterranean climate of South Africa.

Projections of the distribution of plant species to future climate scenarios have largely demonstrated range declines, and altitudinal and latitudinal shifts. Future niche declines have been reported in Mexican alpine (Ramírez-Amezcua et al., 2016) and Andean (Feeley and Silman, 2010; Tovar et al., 2013; Ramírez-Villegas et al., 2014) plant species. Suitable habitat completely disappeared for all Mexican alpine plant species by 2070 under RCP4.5 and GCM ACCESS 1.0 (Ramírez-Amezcua et al., 2016). Range contraction of alpine vegetation in a progressively warming climate is not surprising since these plants are currently restricted and adapted to relatively low temperature habitats (3-6 ℃ mean annual temperature) (Ramírez-Amezcua et al., 2016). Similarly, by 2050, it was predicted that more than 50 % of species studied in the Andes will experience reductions of at least 45 % in their niche size, while 10 % of the species may be extinct (Ramírez-Villegas et al., 2014). Globally, 57 % of plant species and 34 % of animal species were predicted to lose at least 50 % of their current climatic range by the 2080s (Warren et al., 2013). Even with mitigation, such range contractions are only expected to be reduced by 60 % (Warren et al., 2013).

Results of this study, however, show a different trend; an increase in the climatically suitable habitat of *A. quanzensis* in a warming environment. The increase is largest when there is no climate change mitigation, that is under RCP8.5 for both GCMs. Species' responses to changes in environmental conditions are affected by species-specific tolerance to

environmental variability. Amazon angiosperm species with a poor tolerance to water shortages and high temperatures markedly declined in climatic range compared to those well adapted to moisture deficits (Miles et al., 2004). The species which tolerate water deficits have morphological and physiological adaptations, which include presence of thick, waxy cuticles, sunken stomata, small leaves with low density stomata, and some photosynthesize through the C4 and CAM pathways. The increase in the future distribution of *A. quanzensis* suggests the species is adapted to a warm and dry environment, and climate change may not be a threat to its long-term persistence. Species in the genera *Brachystegia*, *Julbernardia*, and *Isoberlinia*, which also thrive in hot and dry environments, and constitute miombo woodlands, may similarly expand their ranges in a warming environment.

Modeling in this study assumes that the species can disperse to the projected niche. That assumption may not hold given the heterogeneity of landscapes and anthropogenic activities. Barriers to seed dispersal in the form of urban areas, mountain ranges, huge water bodies, and farming areas (Funk et al., 2005; Antolin et al., 2006; Riley et al., 2006; Storfer et al., 2007) may disallow a fundamental niche to be turned into a realized niche. For conservation purposes, in the presence of barriers to dispersal, assisted migration to projected suitable niches may be implemented.

Climate change may have indirect impacts on species distributions, for example, through the action of pests and diseases. There are biotic interactions that are important for species survival that are not taken into consideration when using ENMs. Competition with co-occurring species may reduce expansion. Moreover, the relatively long generation times of long-lived tree species may decrease establishment in suitable habitats within a short period of time (Ramirez-Villegas et al., 2014). Taken collectively, these factors may alter the

projected future distributions of *A. quanzensis*, and thus conclusions should be drawn with caution.

Hindcasting current distributions to the mid-Holocene has largely been done for northern hemisphere plants. In the northern hemisphere, the mid-Holocene experienced a warmer and higher precipitation environment compared to the Last Glacial Maximum (LGM) (Garzón et al., 2007; Alba-Sánchez et al., 2010). When hindcasted to the mid-Holocene, there was an increase in the distribution of *Castanea sativa*, *Fagus sylvatica* (Garzón et al., 2007), *Abies alba*, *A. pinsapo* (Alba-Sánchez et al., 2010), *Picea abies* and *Juniperus communis* (Pearman et al., 2008). Generally, in Europe, trees expanded their ranges during the mid-Holocene from their LGM refugia in response to a warming and wetter environment, with a tendency for species to rise in altitude and latitude.

On the contrary, results in this study show that the distribution of *A. quanzensis* during the mid-Holocene was more restricted than present, potentially supporting the Forest refuge hypothesis proposed by Haffer (1969). Studies have shown a complex mosaic of drier and mesic conditions in the sub region (Holmgren et al., 2003; Wanner et al., 2008; Burrough and Thomas, 2013), as a result of the impacts of the tropical Atlantic and Indian Ocean moisture sources. Palynological records show the emergence of the Fabaceae (Nash et al., 2006) and the miombo woodlands (Scott et al., 2012) during the early Holocene in regions with seasonal rainfall patterns. There was a gradual displacement of drought-intolerant tropical seasonal forest by grassland and Zambezian woodlands in south-central Africa during the early Holocene (Ivory et al., 2012). The reduced distribution of *A. quanzensis* during the Mid-Holocene in this study could be the emergence of the Fabaceae in response to

seasonality and the emergence of drought-tolerant miombo woodlands in drier areas of the sub region.

Although climate change may not negatively affect the future distribution of *A. quanzensis*, anthropogenic activities, particularly logging, pose a huge threat (Gerhardt and Todd, 2009). The species is logged in its entire native range for many uses, including for woodcarving. Woodcarvings derived from *A. quanzensis* dominate at market stalls (Braedt and Standa-Gunda, 2000; Braedt and Schroeder, 2003), and such dominance increases pressure on the resource in the wild. The failure of formal and informal institutions to regulate harvesting of natural resources in sub-Saharan Africa has rendered them open-access (Braedt and Schroeder, 2003), often resulting in unsustainable use. For conservation, it is recommended that anthropogenic threats of *A. quanzensis* be mitigated since they pose a threat to the long-term persistence of the species. This may involve controlling logging in unprotected forests, and reforestation in overharvested areas. Connectivity between forests is recommended for the species to successfully disperse to the projected niches. In the presence of barriers, such as highways and buildings, the seeds are unlikely to be dispersed.

2.5 Conclusions

The distribution of *A. quanzensis* in sub Saharan Africa expanded when modeled in a warming environment. The expansion in largest when there is no mitigation of climate change, under RCP8.5. When hindcasted to the mid-Holocene, the distribution decreased. Climate change may not be a direct threat to the occurrence of *A. quanzensis* because the species seems to be drought-tolerant. The distribution is prevalent in low-altitude and relatively hot regions of eastern and southern African countries. The distributions of other

miombo woodland trees that co-occur with *A. quanzensis* may similarly expand their ranges in a warming environment. The decline in the distribution of *A. quanzensis* during the mid-Holocene suggests that the climate was relatively colder than present. Although the distribution of *A. quanzensis* increased in a warming environment, other indirect impacts of climate change, such as occurrence of pests and diseases, were not considered in the modeling, and thus the overall impact of climate change remains unknown. Caution is advised because successful dispersal is assumed under the projected niche expansion. Dispersal may not always be successful because of barriers to gene flow. Connectivity should be maintained or established where the species has been shown to most likely occur. With uncertainties on the indirect impacts of climate change and dispersal to suitable niches, *A. quanzensis* should be conserved since it is unsustainably harvested for a myriad of uses in unprotected forests in the native range.

2.6 References

Adeyemi, T. O., Ogundipe, O. T. and Olowokudejo, J. D. 2012. Species distribution modelling of family Sapindaceae in West Africa. *International Journal of Botany*, **8**: 45-49.

Alba-Sanchez, F., Lopez-Saez, A., Pando, B. B., Linares, J. C., Nieto-Lugilde, D. and Lopez-Merino, L. 2010. Past and present potential distribution of the Iberian *Abies* species: a phytogeographic approach using pollen data and species distribution models. *Diversity and Distributions*, **16**: 214-228.

Antolin, M., Savage, L. and Eisen, R. 2006. Landscape features influence genetic structure of black-tailed prairie dogs (*Cynomys ludovicianus*). *Landscape Ecology*, **21**: 867–875.

Bitencourt, C., Rapinia, A., Damascena, L. S. and Junior, P. D. M. 2016. The worrying future of the endemic flora of a tropical mountain range under climate change. *Flora*, **218**: 1–10.

Braedt, O. and Schroeder, J. M. 2003. Institutions governing the use of forest products: woodcraft commercialization in Southern Zimbabwe. *Forest, Trees and Livelihoods*, **13**: 59-78.

Braedt O. and Standa-Gunda, W. 2000. Woodcraft markets in Zimbabwe. *International Tree Crops Journal*, **10**: 367-384.

Braunisch, V., Coppes, J., Arlettaz, R., Suchant, R., Schmid, H. and Bollmann, K. 2013. Selecting from correlated climate variables: A major source of uncertainty for predicting species distributions under climate change. *Ecography*, **36**: 971-983.

Burrough, S. L. and Thomas, D. S. G. 2013. Central southern Africa at the time of the African Humid Period: a new analysis of Holocene palaeoenvironmental and paleoclimate data. *Quaternary Science Reviews*, **80**: 29-46.

Bush, M. B., Silman, M. R. and Urrego, D. H. 2004. 48,000 years of climate and forest change in a biodiversity hot spot. *Science*, **303**: 827–829.

Cantor, B. S., Sun, C. C., Tortolero-Luna, G., Richards-Kortum, R. and Follen, M. 1999. A comparison of C/B ratios from studies using Receiver Operating Characteristic curve analysis. *Journal of Clinical Epidemiology*, **52**: 885-892.

Chen, I. C., Hill, J. K., Ohlemuller, R., Roy, D. B. and Thomas, C. D. 2011. Rapid range shifts of species associated with high levels of climate warming. *Science*, **333**: 1024–1026.

Crimmins, S. M., Dobrowski, S. Z., Greenberg, J. A., Abatzoglou, J. T. and Mynsberge, A. R. 2011. Changes in climatic water balance drive downhill shifts in plant species' optimum elevations. *Science*, **331**: 324–327.

Crimmins, S. M., Dobrowski, S. Z. and Mynsberge, A. R. 2013. Evaluating ensemble forecasts of plant species distributions under climate change. *Ecological Modelling*, **266**: 126– 130.

Elith, J. and Graham, C. H. 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography*, **32**: 66-77.

ESRI (Environmental Systems Resource Institute). 2017. ArcMap 10.4.1, ESRI, Redlands.

Feeley, K. J., Malhi, Y., Zelazowski, P. and Silman, M. R. 2012. The relative importance of deforestation, precipitation change, and temperature sensitivity in determining the future distributions and diversity of Amazonian plant species. *Global Change Biology*, **18**: 2636–2647.

Feeley, K. J., and Silman, M. R. 2010. Land-use and climate change effects on population size and extinction risk of Andean plants. *Global Change Biology*, **16**: 3215–3222.

Folland, C. K., Rayner, N. A., Brown, S. J., Smith, T. M., Shen, S. S. P., Parker, D. E., Macadam, I., Jones, P. D., Jones, R. N., Nicholls, N. and Sexto, D. M. H. 2001. Global temperature change and its uncertainties since 1861. *Geophysical Research Letters*, **28**: 2621-2624.

Frost, P. 1996. The ecology of Miombo woodlands. In: *The Miombo in transition: woodlands and welfare in Africa* (Campbell, B. ed.), pp. 11-57. Center for International Forestry Research, Bogor.

Funk, W. C., Blouin, M. S., Corn, P. S., Maxell, B. A., Pilliod, D. S., Amish, S. and Allendorf, F. W. 2005. Population structure of Columbia spotted frogs (*Rana luteiventris*) is strongly affected by the landscape. *Molecular Ecology*, **14**: 483–496.

Garzon, M. B., de Dios, R. S. and Ollero, H. S. 2007. Predictive modelling of tree species distributions on the Iberian Peninsula during the Last Glacial Maximum and the Mid-Holocene. *Ecography*, **30**: 120-134.

Gathua, M. 2000. The effects of primates and squirrels on seed survival of a canopy tree, *Afzelia quanzensis*, in Arabuko-Sokoke forest, Kenya. *Biotropica*, **32**: 127-132.

Gerhardt, K. and Nemarundwe, N. 2006. Participatory planting and management of indigenous trees: Lessons from Chivi District, Zimbabwe. *Agriculture and Human Values*, **23**: 231-243.

Gerhardt, K. and Todd, C. 2009. Natural regeneration and population dynamics of the tree *Afzelia quanzensis* in woodlands in Southern Africa. *African Journal of Ecology*, **47**: 583-591.

Germishuizen, G., Crouch, N. R. and Condy, G. 2005. *Afzelia quanzensis*. *Flowering Plants of Africa*, **59**: 74–83.

Graham, C. H., Elith, J., Hijmans, R. J., Guisan, A., Peterson, A. T., Loiselle, B. A. and The Nceas Predicting Species Distributions Working Group. 2008. The influence of spatial errors in species occurrence data used in distribution models. *Journal of Applied Ecology*, **45**: 239-247.

Guisan, A., and Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**: 147-186.

Hijmans, R. J., Cameron, S.E., Parra, J. L., Jones, P. G. and Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**: 1965-1978.

Haffer, J. 1969. Speciation in Amazonian forest birds. *Science*, **165**: 131-137.

Holmgren, K., Lee-Thorp, J. A., Cooper, G. R. J., Lundblad, K., Partridge, T. C., Scott, L., Sithaldeen, R., Talma, A. S. and Tyson, P. D. 2003. Persistent millennial-scale climatic variability over the past 25,000 years in southern Africa. *Quaternary Science Reviews*, **22**: 2311-2326.

Hutchinson, G.E. 1957. Concluding remarks. *Cold Spring Harbor Symposium on Quantitative Biology*, **22**: 415–457.

IPCC. 2007. *Synthesis report. Climate change 2007: contribution of working group I to the fourth assessment report of the inter-governmental panel on climate change*. IPCC, Geneva.

IPCC. 2013. *Summary for Policymakers. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK.

IPCC. 2014. *Summary for Policymakers. Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK.

Ivory, S. J., Lézine, A.-M., Vincens, A. and Cohen, A. S. 2012. Effect of aridity and rainfall seasonality on vegetation in the southern tropics of East Africa during the Pleistocene/Holocene transition. *Quaternary Research*, **77**: 77-86.

Jump, A. S. and Penuelas, J. 2005. Running to stand still: adaptation and the response of plants to rapid climate change. *Ecology Letters*, **8**: 1010–1020.

Keeling, F. R. 2008. Recording earth's vital signs. *Science*, **319**: 1771-1772.

Kelly, A. E. and Goulden, M. L. 2008. Rapid shifts in plant distribution with recent climate change. *PNAS,* **105**: 11823–11826.

Lawrence, D. M., Oleson, K. W., Flanner, M. G., Fletcher, C. G., Lawrence, P. J., Levis, S., Swenson, S. C. and Bonan, G. B. 2012. The CCSM4 land simulation, 1850-2005: Assessment of surface climate and new capabilities. *Journal of Climate*, **25**, 2240-2260.

Liu, C., Berry, M. P., Dawson, P. T. and Pearson, G. R. 2005. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, **28**: 385-393.

Lobo, J. M., Jiménez-Valverde, A. and Real, R. 2008. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, **17**: 145–151.

Manel, S., Williams, H. C. and Ormerod, S. J. 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology*, **38**: 921-931.

Mignot, J. and Bony, S. 2013. Presentation and analysis of the IPSL and CNRM climate models used in CMIP5. *Climate Dynamics*, **40**: 2089: doi: 10.1007/s00382-013-1720-1.

Miles, L., Grainger, A. and Phillips, O. 2004. The impact of global climate change on tropical forest biodiversity in Amazonia. *Global Ecology and Biogeography*, **13**: 553–565.

Mugandani, R., Wuta, M., Makarau, A. and Chipindu, B. 2012. Re-classification of agro-ecological regions of Zimbabwe in conformity with climate variability and change. *African Crop Science Journal*, **20**: 361-369.

Munyanziza, E. and Oldeman, R. A. 1996. Miombo trees: ecological strategies, silviculture and management. *Ambio*, **25**: 454–458.

Nash, D. J., Meadows, M. E. and Gulliver, V. L. 2006. Holocene environmental change in the Okavango Panhandle, northwest Botswana. *Quaternary Science Reviews*, **25**: 1302-1322.

Parolo, G. and Rossi, G. 2008. Upward migration of vascular plants following a climate warming trend in the Alps. *Basic and Applied Ecology*, **9**: 100-107.

Pearman, P. B., Randin, C. F., Broennimann, O., Vittoz, P., van der Knaap, W. O., Engler, R., Lay, G. L., Zimmermann, N. E. and Guisan, A. 2008. Prediction of plant species distributions across six millennia. *Ecology Letters*, **11**: 357-369.

Pearson, R. G., Raxworthy, C. J., Nakamura, M. and Peterson, A. T. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography*, **34**: 102–117.

Petit, J. R., Jouzel, J., Raynaud, D. et al. 1999. Climate and atmospheric history of the past 420 000 years from the Vostok ice core, Antarctica. *Nature*, **399**: 429-436.

Petit, R. J., Hu, F. S. and Dick, C. W. 2008. Forests of the past: a window to future changes. *Science*, **320**: 1450–1452.

Phillips, S.J., Anderson, R. P. and Schapire, R. E. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**: 231–259.

Phillips, S. and Dudik, M. 2008. Modeling of species distributions with MaxEnt: new extensions and a comprehensive evaluation. *Ecography*, **31**: 161-175.

Ramirez-Amezcua, Y., Steinmann, V. W., Ruiz-Sanchez, E. and Rojas-Soto, O. R. 2016. Mexican alpine plants in the face of global warming: potential extinction within a specialized assemblage of narrow endemics. *Biodiversity and Conservation*, **25**: 865-885.

Ramirez-Villegas, J., Cuesta, F., Devenish, C., Peralvo, M., Jarvis, A. and Arnillas, C. A. 2014. Using species distribution models for designing conservation strategies of Tropical Andean biodiversity under climate change. *Journal of Nature Conservation*, **22**: 391-404.

Riley, S. P. D., Pollinger, J. P., Sauvajot, R. M., York, E.C., Bromley, C., Fuller, T. K. and Wayne, R. K. 2006. FAST-TRACK: A southern California freeway is a physical and social barrier to gene flow in carnivores. *Molecular Ecology*, **15**: 1733–1741.

Roberts, D. R. and Hamann, A. 2012. Predicting potential climate change impacts with bioclimate envelope models: a palaeoecological perspective. *Global Ecology and Biogeography*, **21**: 121–133.

Scott, L., Neumann, F. H., Brook, G. A., Bousman, C. B., Norström, E.and Metwally, A. A. 2012. Terrestrial fossil-pollen evidence of climate change during the last 26 thousand years in Southern Africa. *Quaternary Science Reviews*, **32**: 100-118.

Scripps Institution of Oceonagraphy, 2017 (https://scripps.ucsd.edu/programs/keelingcurve/) Accessed October 21st, 2017.

Soberon, J. and Nakamura, M. 2009. Niches and distributional areas: Concepts, methods, and assumptions. *PNAS*, **106**: 19644–19650.

Storfer, A., Murphy, M. A., Evans, J. S., Goldberg, C. S., Robinson, S., Spear, S. F., Dezzani, R., Delmelle, E., Vierling, L. and Waits, L. P. 2007. Putting the 'landscape' in landscape genetics. *Heredity*, **98**: 128–142.

Thuiller, W., Arau´ jo, M. B., Pearson, R. G., Whittaker, R. J., Brotons, L. and Lavorel, S. 2004. Biodiversity conservation: uncertainty in predictions of extinction risk. *Nature*, **430**: 33.

Tovar, C., Arnillas, C. A., Cuesta, F. and Buytaert, W. 2013. Diverging responses of Tropical Andean biomes under future climate conditions. *PLoS ONE*, **8**: e63634.

Van der Wal, J., Shoo, L. P. and Williams, S. E. 2009. New approaches to understanding late Quaternary climate fluctuations and refugial dynamics in Australian wet tropical rain forests. *Journal of Biogeography*, **36**: 291–301.

Wanner, H., Beer, J., Bütikofer, J. et al. 2008. Mid- to Late Holocene climate change: an overview. *Quaternary Science Reviews*, **27**: 1791-1828.

Warren, R., Van Der Wal, J., Price, J., Welbergen, J. A., Atkinson, I., Ramirez-Villegas,J., Osborn, T. J., Jarvis, A., Shoo, L. P., Williams, S. E. and Lowe, J. 2013. Quantifying the benefit of early climate change mitigation in avoiding biodiversity loss. *Nature Climate Change*, **3**: doi: 10.1038.

3. DEVELOPMENT AND EVALUATION OF MICROSATELLITE LOCI IN *AFZELIA QUANZENSIS* (FABACEAE)

Acknowledgements are in order for the contribution of the following paper in chapter three.

Jinga, P., Palagi, J. and Ashley V. M. 2016. Development of microsatellite loci of pod mahogany, *Afzelia quanzensis* (Fabaceae), by Illumina shotgun sequencing, and cross-amplification in *A. africana*. *Applications in Plant Sciences*, **4**: doi:10.3732/apps.1600010.

3.1 Introduction

Microsatellites, also known as simple sequence repeats (SSRs, Jacob et al., 1991) and short tandem repeats (STRs, Craig et al., 1988), are short lengths of DNA made up of tandemly repeated mono-, di-, tri-, tetra-, penta- or hexa-nucleotide motifs (Ashley, 2010). Microsatellites are classified into simple perfect, simple imperfect, compound perfect and compound imperfect (Wang et al., 2009). Simple perfect microsatellites are made up of one motif (eg, $[ATT]_n$), while simple imperfect ones are made up of one motif with different lengths (eg. $[ATT]_n[AAA][ATT]_{n+1}$). Compound perfect microsatellites are made up of two or more motifs of the same length (eg. $[ATT]_n[AGG]_n$). Compound imperfect microsatellites have different motifs of different lengths (eg. $[ATT]_n[AAGG]_{n+1}$) (Wang et al., 2009). The existence of microsatellites was first reported in 1982 (Hamada et al., 1982), and they were subsequently found to be widely distributed in eukaryotic genomes (Tautz and Renz, 1984). Since then, the development of microsatellites has been rapid in many organisms, and applications have been in an array of subjects, including forensics, population genetics, disease diagnosis, conservation genetics, phylogeography and genome mapping.

Microsatellites have become one of the markers of choice because of their hypervariability, codominant inheritance, requirement of minute amounts of template DNA, extensive genome coverage, including in mitochondrial and chloroplast DNAs, high reproducibility, and that they can be shared among laboratories as primer sequences (Powell et al., 1996; Ashley, 2010; Kalia et al., 2011). They are relatively inexpensive compared to next generation sequencing (NGS) techniques to generate enough data to differentiate closely related individuals or populations. The relatively small number of loci required to produce sufficient data means each locus can be carefully genotyped separately, reducing errors in the output. Because microsatellites are PCR-based, they can be amplified from low quantity and quality DNA (Hodel et al., 2016), thus becoming useful in ancient DNA studies and those involving archived specimens.

The hypervariability of microsatellites in brought about by mutations affecting the number of repeat units (Tautz, 1989; Weber and May, 1989; Dow et al., 1995). Change in motif number may be caused by single-stranded DNA slippage, double-stranded DNA recombination, and retrotransposition (Wang et al., 2009; Kalia et al., 2011). Slippage of DNA polymerase III on the template strand has been identified as the common mechanism of microsatellite evolution (Ellengren et al., 2004). Slippage during DNA replication may cause a loss or gain of a motif if the error is not repaired (Wang et al., 2009). Slippage errors are corrected by a mismatch repair (MMR) mechanism, and thus microsatellite evolution depends on the balance between DNA slippage and the fidelity of the MMR mechanism (Kalia et al., 2011), and since the balance is unique to individuals, the result is an array of alleles at a single locus in a population.

Microsatellites have their drawbacks, and chief among them is homoplasy (Viard et al., 1998). Slippage during DNA replication is susceptible to back mutations, which result in homology in state among individuals lacking identity by decent. Homoplasy may inflate homology, which lowers confidence in microsatellite analyses. Homoplasy, however, may be modeled, such as by using the stepwise mutation model, to lower its influence (Hodel et al., 2016). Another concern of microsatellites is that they inflate *F*-statistic estimates because of their large number of alleles. A remedy may be to use other statistics, such as G'_{ST} (Hedrick, 2005) and *D* (Jost, 2008), in place of the traditional *F*-statistics.

Microsatellites have been traditionally developed by ligating a known linker to DNA fragments digested by specific restriction enzymes. Probes containing repeat sequences are hybridized to the fragments, and the probes are bound to a nylon membrane or biotinylated. The probes are recovered by streptavidin-coated beads (Numone et al., 2006). The recovered fragments are amplified, cloned, and sequenced. This enrichment protocol is laborious and requires expertise, and the yield is typically very low. A less laborious method has been to develop microsatellites by interspecific and intergeneric transferability or cross-amplification. Microsatellite primers may amplify in related taxa because some sections of DNA are highly conserved. Specifically, genic microsatellites are expected to have a high rate of transferability because transcribed regions are highly conserved in related organisms (Ellis and Burke, 2007).

The advent of NGS has simplified microsatellite marker development. A large number of microsatellites can be quickly developed in non-model organisms from NGS platforms. The declining costs of sequencing is making development of microsatellites by NGS more attractive. Different NGS platforms are in use, and these include the Roche 454

(454 Life Sciences, Braford, CA, USA), Illumina (Illumina, San Diego, CA, USA), and Ion Torrent (Thermo Fisher Scientific, Waltham, MA, USA) platforms. These platforms have different capabilities, and the most important in microsatellite development is read length. Longer reads are ideal because they are more likely to include the flanking region required for primer development (Schoebel et al., 2013; Elliot et al., 2014). In that regard, the Illumina platform has been shown to be the most cost effective (Glenn, 2011; Hodel et al., 2016) although the technology is rapidly evolving.

In tandem with the evolution of NGS platforms is the development of software for microsatellite recovery from NGS data. Software for microsatellite recovery include Geneious (Kearse et al., 2012), GMATo (Wang et al., 2013), MISA (Thiel et al., 2003), MSATCOMMANDER (Faircloth, 2008), PAL_FINDER (Castoe et al., 2012), QDD3 (Meglecz et al., 2014), and SSR Locator (da Maia et al., 2008). These softwares have different functionalities, such as post sequencing processing requirements, ability for inherent primer designing, and whether they are executed by command line or graphic user interphase. Thus, the choice of software is user specific, and depends on the operating system and speed of the computer, type of NGS output data, and computer literacy of the user, among several factors.

The aims of the study were to (1) develop and evaluate microsatellite loci in *Afzelia quanzensis* from NGS data, and (2) check transferability of the markers in *A. africana*. *Afzelia quanzensis* is a deciduous tree that naturally occurs in eastern and southern Africa. The microsatellites will be used in genetic studies necessary for genetic resources management.

3.2 Materials and Methods

Genomic DNA was extracted from a leaf of one *A. quanzensis* individual (Sample coordinates: 19º36.056´S, 32º30.084´E) using the DNeasy Plant Mini Kit (Qiagen, Valencia, California, USA) following the manufacturer's instructions. The DNA was used to prepare a sequencing library using the KAPA DNA Library Preparation Kit for Illumina Sequencing (Kapa Biosystems, Wilmington, Massachusetts, USA) following the manufacturer's instructions. The final library was quantified using the KAPA Library Quantification Kit for Illumina. The DNA library was sequenced by an Illumina Miseq Benchtop Sequencer (Illumina, San Diego, California, USA).

The resulting raw Illumina paired-end sequencing reads were analysed with a perl script, PAL_FINDER_v0.02.04 (available at http://sourceforge.net/projects/palfinder), which identifies microsatellite loci without the need for prior sequence trimming and assembly (Castoe et al., 2012). The perl script was run with Primer3 version 2.0.0 (Rozen and Skaletsky, 1999) (available at http://primer3.sourceforge.net/releases/php) for simultaneous primer designing. Default settings on Primer3 were used, with the following adjustments: primer minimum annealing temperature 50 ℃, primer maximum annealing temperature 60 ℃, and primer optimum annealing temperature 55 ℃. Seventy potential amplifiable loci (PALs) with amplifiable primer pairs that occurred only once were tested for amplification.

Of the 70 loci tested, 39 amplified successfully and these were checked for polymorphisms in 40 individuals randomly collected from a population near Chaseyama, south-eastern Zimbabwe. Forward primers were tagged with a labelled M13 primer tail (TGTAAAACGACGGCCAGT). All PCR reactions were performed in a total volume of 10

µL, with 10 ng of template DNA, 0.6 µM of the reverse primer, 0.15 µM of the forward primer, 0.25 mM each dNTP, 0.6 µL BSA (10% w/v), 1 µL 10x reaction buffer with 15 mM $MgCl_2$, and 0.25 units of *Taq* DNA polymerase (Bulldog Bio, Rochester, New York, USA). Loci Afq45, Afq51, Afq62, Afq68 and Afq69 had an additional 0.1 mM $MgCl_2$.

The thermocycling profile consisted of an initial denaturation at 94 °C for 5 minutes, then 35 cycles of 94 °C for 30 seconds, 55.0 °C or 59.4 °C (Table 2) for 30 seconds, 72 °C for 30 seconds, and a final extension at 72 °C for 7 minutes. The PCR amplicons were electrophoresed on an ABI 3730 DNA analyser with GeneScan 500 LIZ (Applied Biosystems, Foster City, California, USA) as the size standard. The genotypes were scored using GeneMapper version 3.7 (Applied Biosystems). Polymorphic loci were cross-amplified in 24 individuals of a congeneric species, *A. africana*, using conditions specified above, to check transferability.

For the polymorphic loci, number of alleles per locus (*A*), observed heterozygosity (H_O), and expected heterozygosity (H_E) were calculated using GenAlEx version 6.5 (Peakall and Smouse, 2012). The program Arlequin version 3.5 (Excoffier and Lischer, 2010) was used to perform an exact test (Guo and Thomson, 1992) with a Markov chain for testing Hardy-Weinberg equilibrium, while the program Micro-Checker version 2.2.3 (van Oosterhout et al., 2004) was used to estimate null allele frequencies (F_{NULL}) with Bonferroni correction.

3.3 Results

The perl script identified a total of 961,804 potential loci, and 7,789 had primer pairs. Table III shows 39 loci that amplified, their repeat motifs, number of alleles per locus, allele size range, and annealing temperature. Twenty-seven loci were monomorphic while 12 were polymorphic. For the 12 polymorphic loci, *A*, H_O, and H_E are shown in Table IV. The number of alleles for polymorphic loci ranged from 3–10, with an average of 5.583. Observed heterozygosity ranged from 0.138 to 0.737, while H_E ranged from 0.313 to 0.832. Loci Afq35, Afq51, and Afq67 showed departure from Hardy-Weinberg equilibrium (Table IV). Of the 12 loci that were polymorphic in *A. quanzensis*, eight amplified in *A. africana* (Table V).

Table III. Characteristics of 39 microsatellite loci of *Afzelia quanzensis* that successfully amplified

Locus	**Primer sequences (5′−3′)**	**Repeat motif**	***A***	**Allele size range (bp)**	**T_a (°C)**	**NCBI Probe Database accession number**
Afq1	F:CCTATACCAGAAATTGATAAATTAGAGAGC R: GCTTAGCCAAGGGACATTGC	$(ATT)_{78}$	1	417	55.0	Pr032805619
Afq5	F:GACTCACAAGTGGCAAGTGAGG R:GTCCAACGATTGAAAGATTTAAGG	$(AAAT)_{20}$	1	341	55.0	Pr032805635
Afq6	F:CATGACCCAAGCATGACTCC R:CCTACAGTTGTTGAGAAGTCCGC	$(ACCCC)_{25}$	1	252	55.0	Pr032805642
Afq8	F:TTAATAATGCAAAGATGATTGGC R:GGGGCAATAAGTCAAAATGG	$(AAAT)_{24}$	1	410	55.0	Pr032805644
Afq9	F:CATTGACAAAGATGCATGATAGC R:TCATTGTAGTTTTCATTCACAACCC	$(AAAG)_{20}$	1	205	55.0	Pr032805645
Afq10	F:CAGGCAAGGGGTAAAATTGG R:CTGCTCCAAATTCCAAAGCC	$(TTC)_{33}$	1	154	55.0	Pr032805620
Afq12	F:CTCCTCTGCGCCACTATTCC R:CACTCCTCTCTCAGGCAGGG	$(AAC)_{15}$	3	265−271	55.0	Pr032754338
Afq13	F:AAATATTTTCGAGACCACAAACG R:AACTCGATTTCTTCATGTACGG	$(ATT)_{18}$	1	170	55.0	Pr032805621
Afq15	F:AGAAAACCAGCGGTACGAGC R:CATTATCGCCGGTAAGCTGC	$(CGG)_{18}$	1	212	55.0	Pr032805622
Afq20	F:AGAAAACCAGCGGTACGAGC R:CATTATCGCCGGTAAGCTGC	$(CGG)_{18}$	1	326	55.0	Pr032805623
Afq24	F:GGAAAGACTCCAGATCACTTCCC R:ACAAAACTGACCTGAACAAGGC	$(ATT)_{15}$	1	350	55.0	Pr032805624
Afq31	F:TGCACGAATGCAAATAGTGG R:ATTCTAAGGCATTAACATGGAGC	$(AAAT)_{20}$	1	249	55.0	Pr032805625
Afq33	F:GGATTCCATTCTAACCAGAGACC R:AAAGTTAGCTTTGCACCCTCC	$(ATTTT)_{20}$	3	210−220	55.0	Pr032754339

Locus	Primer sequences (5′−3′)	Repeat motif	*A*	Allele size range (bp)	T_a (ºC)	NCBI Probe Database accession number
Afq34	F:AAACTGATGCAAATAAGATGGG R:TCCTAGTTTGATACCAATTAATGTAACG	$(ATTTT)_{20}$	1	349	55.0	Pr032805626
Afq35	F:TGATATCGGTTATGTGCAGGG R:TGCTGGGTCATATTTACTAGTGCC	$(ATT)_{21}$	6	364−382	55.0	Pr032754340
Afq38	F:ACACCATGGGTGAACTTGAGG R:CCCAGAGATTCAGCTTAGGCG	$(TC)_{32}$	1	150	55.0	Pr032805627
Afq39	F:AGGTGGTCATCCACAGTCCC R:GCTCACATTTAGACGGTGACG	$(AT)_{30}$	1	152	55.0	Pr032805628
Afq40	F:CATGCATATATGACGATTTTGTCC R:TGACTGTTCATTTATATACACACATTCACC	$(AT)_{32}$	1	265	55.0	Pr032805629
Afq41	F:TGCATAACCACCCAAAAGGG R:TCCTAATGGTTGATAGGTCCCC	$(ATT)_{21}$	1	290	55.0	Pr032805630
Afq42	F:AATGGCATGTTGCGTACACC R:AAAGCATTTGAAGATTTGGTAGGG	$(ATT)_{33}$	1	343	55.0	Pr032805631
Afq43	F:GAAGAAGGAAGCTTGTCGGC R:ATCACATTACCCGCATTGGG	$(TCC)_{21}$	3	227−239	55.0	Pr032754341
Afq44	F:AATTTACATTTGCTTCAACAGGG R:AAACACTCTTATTAGTTTATTCACCTGG	$(ATCT)_{20}$	4	147−163	55.0	Pr032754342
Afq45	F:CAAAACTAAACGACATCTCCTGC R:TTCCCTTCTTGCTTAGGGAGC	$(TTGGGC)_{24}$	4	297−315	55.0	Pr032754343
Afq46	F:CCATGTGTGAATATATCCCTTTGC R:GGAGGATGTTGTTCCTGTCG	$(AAAAC)_{20}$	1	230	55.0	Pr032805632
Afq47	F:TGACATCAGTTTCCTTGTGCC R:TTTTGCCTAAAGAAAATAGGTTTGG	$(AAAAG)_{20}$	1	195	55.0	Pr032805633
Afq48	F:TTGACCCACGTTCCTTCC R:TCACATGACTTCACAATATTTCCG	$(AAATT)_{20}$	1	152	55.0	Pr032805634
Afq49	F:ATCCTTTTGCCCATTCCTGC R:ATGGCACCCAAAGAAGAAGC	$(TC)_{26}$	7	273−291	55.0	Pr032754344

Locus	Primer sequences (5′−3′)	Repeat motif	*A*	Allele size range (bp)	T_a (ºC)	NCBI Probe Database accession number
Afq50	F:CCAAAGGAATAGTTGGGTTTGC R:TATCGCCTTGTTCAACTGCC	$(AC)_{26}$	1	223	55.0	Pr032805636
Afq51	F:CATGGCTTCAACCTATCCTGG R:CCTTTCTCTGGTCCTTCCCC	$(ATT)_{39}(ATT)_{24}$	7	247–264	55.0	Pr032754345
Afq52	F:GGCAGGATTCATAGTTTACTTTCG R:ACAGGTGACATCGGAGTTGC	$(ATT)_{18}(ATT)_{15}$	1	319	55.0	Pr032805637
Afq54	F:CAAAGAGTAACAAAATCCCTGCG R:CATCGCTGGTTAGATGTTTTAGC	$(AAAT)_{20}(AAAT)_{20}$	1	352	55.0	Pr032805638
Afq56	F:TGCGAACAAGGTTCCTAACG R:TTTGGCATATGACAGTTGATGG	$(ATT)_{36}(ATT)_{15}$	1	408	55.0	Pr032805639
Afq57	F:CCTATTTGAAAGGTAATTTCTAAGACCC R:TCCCACACTTCATAAAACGGG	$(ATT)_{18}(ATT)_{18}$	1	271	55.0	Pr032805640
Afq58	F:TGTTAGCAGCATTGTTGAGGG R:CACTAATGGATTGCCTTTTCCC	$(ATT)_{30}(ATT)_{30}$	1	362	55.0	Pr032805641
Afq62	F:TGTATACAAAACGATTTGACGGC R:TTTCCAATCAAGCAAATCTCG	$(ATT)_{21}(ATT)_{18}$	4	219−234	59.4	Pr032754346
Afq66	F:TGAACAGATCAATCAAAGTGCG R:CCATATTCATCCCACTCCCG	$(TC)_{18}$	1	276	55.0	Pr032805643
Afq67	F:CTTCATCATATAGCATAAGATAATCGG R:TTTAAGATAGGCTCAAGGACGG	$(AC)_{26}$	8	330–354	55.0	Pr032754347
Afq68	F:AGGCACACGAGCACACTAGG R:CAGGACCCTCCAGTGTTTCC	$(TC)_{20}$	8	215–235	55.0	Pr032754348
Afq69	F:TGACCGTTTTAAGAAAAGTCAAGC R:TCGATGATCCAGGAAAGTTGG	$(TC)_{16}$	10	294–318	55.0	Pr032754349

A= number of alleles per locus.
T_a = annealing temperature.

Table IV. Polymorphic microsatellite locus-specific measures of genetic diversity of a population of 40 individuals of *Afzelia quanzensis*[a]

Locus	***A***	H_O	H_E	F_{NULL}
Afq12	3	0.487	0.605	0.1075[c]
Afq33	3	0.410	0.347	-0.0833
Afq35[b]	6	0.730	0.728	-0.0010
Afq43	3	0.550	0.546	-0.0037
Afq44	4	0.350	0.343	-0.0095
Afq45	4	0.556	0.564	0.0079
Afq49	7	0.641	0.612	-0.0231
Afq51[b]	7	0.138	0.832	0.8119[c]
Afq62	4	0.359	0.313	-0.0690
Afq67[b]	8	0.325	0.423	0.1304[c]
Afq68	8	0.583	0.606	0.0195
Afq69	10	0.737	0.748	0.0077

A = number of alleles per locus.
H_O = observed heterozygosity.
H_E = expected heterozygosity.
F_{NULL}= null allele estimates.
[a] Geographic coordinates for the population are 19°36.056´S, 32°30.084´E. A voucher is deposited at the National Herbarium and Botanic Garden, Harare, Zimbabwe (SRGH), with voucher number 1. The specimen was collected by Mr. Percy Jinga.
[b] loci not in Hardy–Weinberg equilibrium.
[c] loci showing evidence of null alleles.

Table V. Genetic properties of eight *Afzelia quanzensis* microsatellite loci that amplified in *Afzelia africana*[a]

Locus	***A***	**Allele size range (bp)**	H_O	H_E
Afq12	2	265−268	0.105	0.100
Afq33	2	210−215	0.348	0.386
Afq35	2	379−382	0.333	0.420
Afq43	3	230−236	0.087	0.084
Afq44	2	159−163	1.000	0.500
Afq49	5	283−291	0.750	0.622
Afq67	1	348	0.000	0.000
Afq69	1	289	0.000	0.000

A = number of alleles per locus.
H_O = observed heterozygosity.
H_E = expected heterozygosity.
[a] Geographic coordinates for the population are 1.55586°N, 9.26674°E. Vouchers are deposited at the Université Libre de Bruxelles, Belgium (BRLU), with voucher numbers AD61-AD85. The population is located in Bassila, central Benin. Specimens were collected by Dr. Olivier Hardy.

The percentage of loci that successfully amplified, and the percentage that were polymorphic, from the total loci tested for amplification are shown on Figure 3.1. Loci with pentanucleotide motifs had the highest success rate of amplification (75.00%), while loci with trinucleotide motifs were the least successful (40.74%). Half of the loci with hexanucleotide motifs were polymorphic, which was the highest success rate, while the least successful were loci with trinucleotide motifs (11.11%). Figure 3.2 shows the percentage of polymorphic loci out of those that amplified. All loci with hexanucleotide motifs that amplified were polymorphic, followed by loci with dinucleotide motifs (44.44 %). Loci with pentanucleotide motifs were the least polymorphic out of those that amplified (16.66 %) (Figure 3.2).

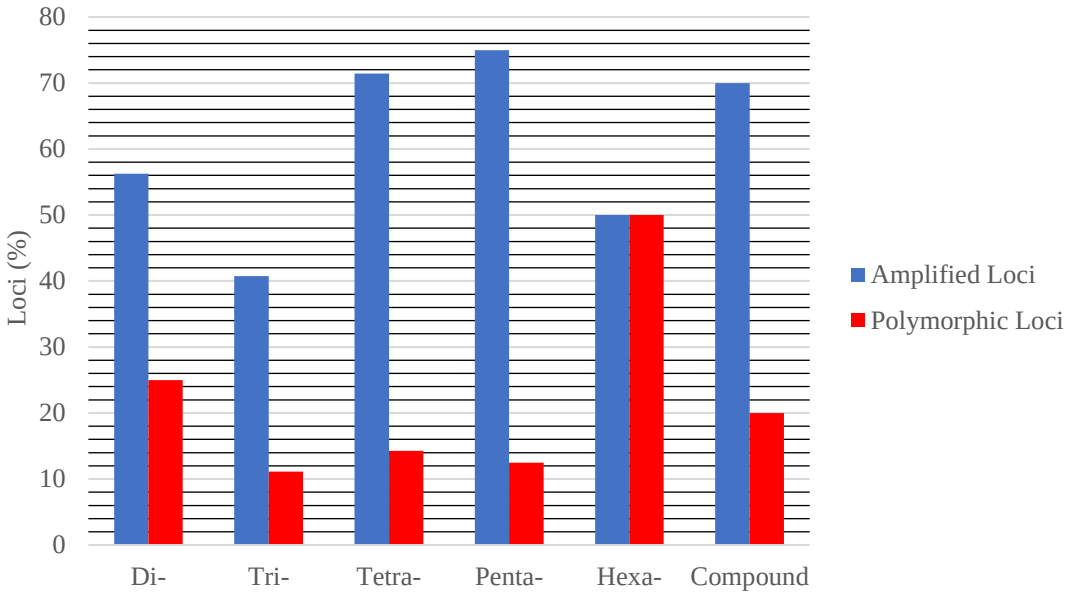


Figure 3.1. Percentage of loci that amplified, and were polymorphic, from total number of loci tested in *Afzelia quanzensis*

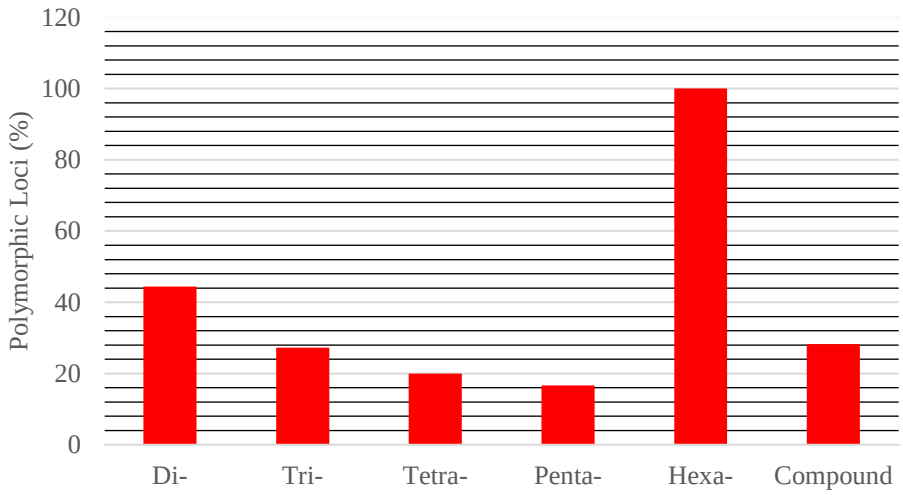


Figure 3.2. Percentage of polymorphic loci from number of amplified loci in *Afzelia quanzensis*

3.4 Discussion

NGS produced a phenomenal number of potential microsatellite loci. This is in contrast to the enrichment protocol where the attrition rate is very high (Ramsay et al., 2000; Parida et al., 2009). The enrichment protocol is time-consuming, and the DNA for enrichment should be of very high quality and quantity (Hodel et al., 2016). Several steps of the enrichment protocol are prone to failure. For example, the clone insertion sites generated may be too close to the repeat motifs such that there will be no space for primer design (Parida et al., 2009). The yield of the enrichment protocol is typically less than 20 polymorphic loci for 30-60 primer pairs developed (Zalapa et al., 2012). With declining costs, NGS should, therefore, become the method of choice for microsatellite marker development given the sheer number of potential loci that are recovered for evaluation.

Dinucleotide repeats have been observed to be most variable compared to tri-, tetra, penta-, and hexanucleotide repeats (Grist et al., 1993; Ellengren 2000). A review of 6,782 published genomic microsatellite markers revealed that dinucleotide motifs showed highest number of alleles per locus (Merrit et al., 2015). Results of this study also show that dinucleotide motifs have highest variability; 44.44% of amplified loci were polymorphic. Although all hexanucleotide markers that amplified were polymorphic (100%), their small sample size in this study prohibits making conclusions with high confidence. The high polymorphism of dinucleotide motifs may be due to the relative ease of mutation through DNA slippage during replication (Grist et al., 1993; Tautz and Schlotterer, 1994; Ellengren, 2000, 2004) compared on the other motifs.

All polymorphic loci amplified at the specified annealing temperature range of 50 to 60℃. All but one polymorphic loci amplified at the specified optimum annealing temperature of 55℃. The one loci, Afq62, amplified at 59.4℃. These results demonstrate the utility of Primer3 software in designing primers. A small sample size might have caused three loci to depart from HWE. It is recommended that these loci should be further tested when the sample size is large.

Cross-amplification of primers in *A. africana* was 75 % successful. Cross-amplification has been demonstrated in a number of plant species (Saha et al., 2006; Gimenes et al., 2007; Nazareno et al., 2009; Ince et al., 2010; Liu et al., 2017; Riser et al., 2017; Shi et al., 2017; Tosso et al., 2017), and it is now almost a standard requirement for microsatellite marker publication. Transferability has been shown to work not only in species within a genus, but in different genera within a family (Wang et al., 2004; Ellis and Burke, 2007; Stagel et al., 2008). High

transferability of above 60 % have been commonly reported, for example, in microsatellites of finger millet transferred to pearl millet (Arya et al., 2009), and of sugarcane transferred to other *Saccharum* species (Parida et al., 2009). These results are not surprising considering that DNA is largely conserved in related organisms. The markers developed in this study may not only be transferable in *A. africana*, but in other congeners, such as *A. xylocarpa, A. bipidensis, A. bella*, and *A. pachyloba*. The markers may also be transferable in genera of the legume tribe Detarieae, subfamily Caesalpinioideae, such as *Hymenostegia, Isoberlinia, Microberlinia*, and *Tessmannia*.

3.5 Conclusions

Next generation sequencing produces thousands of potential microsatellite loci. The NGS method is cost effective in microsatellite marker development compared to enrichment protocols. Dinucleotide motifs seem to be the most variable. The twelve microsatellite loci developed for *A. quanzensis* will be used in genetic studies of the species. The loci may also be used in other congeneric species since transferability in *A. africana* was successful for most of them.

3.6 References

Arya, L., Verma, M., Gupta, V. K. and Karihaloo, J. L. 2009. Development of EST–SSRs in finger millet (*Eleusine coracana* ssp *coracana*) and their transferability to pearl millet (*Pennisetum glaucum*). *Journal of Plant Biochemistry and Biotechnology*, **18**: 97–100.

Ashley M. V. 2010. Plant parentage, pollination, and dispersal: How DNA microsatellites have altered the landscape. *Critical Reviews in Plant Sciences*, **29**: 148-161.

Castoe, T. A., Poole, A. W., de Koning, A. P. J. et al. 2012. Rapid microsatellite identification from Illumina paired-end genomic sequencing in two birds and a snake. *PLoS ONE*, **7**: e30953.

Craig, J., Fowler, S., Burgoyne, L. A., Scott, A. C. and Harding, H. W. J. 1988. Repetitive deoxyribonucleic acid (DNA) and human genome variation: A concise review relevant to forensic biology. *Journal of Forensic Sciences*, **33**: 1111-1126.

da Maia, L. C., Palmieri, D. A., de Souza, V. Q., Kopp, M. M., de Carvalho, F. I. F. and de Oliveria, A. C. 2008. SSR Locator: Tool for simple sequence repeat discovery integrated with primer design and PCR simulation. *International Journal of Plant Genomics*, **2008**: doi:10.1155/2008/412696.

Dow, B. D., Ashley, M. V., and Howe, H. F. 1995. Characterization of highly variable $(GA/CT)_n$ microsatellites in the bur oak, *Quercus macrocarpa*. *Theoretical and Applied Genetics,* **91**: 137–141.

Ellengren, H. 2000. Heterogeneous mutation processes in human microsatellite DNA sequences. *Nature Genetics,* **24**: 400–402.

Ellengren, H. 2004. Microsatellites: Simple sequences with complex evolution. *Nature Reviews. Genetics*, **5**: 435–445.

Ellis, J. R. and Burke, J. M. 2007. EST–SSRs as a resource for population genetic analyses. *Heredity*, **99**: 125–132.

Elliott, C. P., Enright, N. J., Allcock, R. J. N., Gardner, M. G., Meglecz, E., Anthony, J. and Krauss, S. L. 2014. Microsatellite markers from the Ion Torrent: A multi-species contrast to 454 shotgun sequencing. *Molecular Ecology Resources,* **14**: 554–568.

Excoffier, L. and Lischer, H. E. L. 2010. Arlequin suite ver3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**: 564–567.

Faircloth, B. C. 2008. MSATCOMMANDER: Detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular Ecology Resources*, **8**: 92–94.

Gimenes, M. A., Hoshino, A. A., Barbosa, A. V., Palmieri, D. A. and Lopes, C. R. 2007. Characterization and transferability of microsatellite markers of the cultivated peanut (*Arachis hypogaea*). *BMC Plant Biology*, **7**: 9.

Glenn, T. 2011. Field guide to next-generation DNA sequencers. *Molecular Ecology,* **11**: 759–769.

Gou, S.W. and Thompson, E. A. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics,* **48**: 361−372.

Grist, S. A., Figaira, F. A. and Morley, A. A. 1993. Dinucleotide repeat polymorphisms isolated by the polymerase chain reaction. *BioTechniques,* **15**: 304 – 309.

Hamada, H., Petrino, M. G., and Kakunaga, T. 1982. A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. *Proceedings of the National Academy of Svciences* (USA) **79**: 6465–6469.

Hedrick, P. W. 2005. A standardized genetic differentiation measure. *Evolution,* **59**: 1633-1638.

Hodel, R. G. J., Segovia-Salcedo, M., Landis, J. B. et al. 2016. The report of my death was an exaggeration: A review for researchers using microsatellites in the 21st century. *Applications in Plant Sciences*, **4**: doi:10.3732/apps.1600025.

Ince, A. G., Karaca, M. and Onus, A. N. 2010. Polymorphic microsatellite markers transferable across capsicum species. *Plant Molecular Biology Reporter*, **28**:285–291.

Jacob, H. J., Lindpaintner, K., Lincoln, S. E. et al. 1991. Genetic mapping of a gene causing hypertension in the stroke-prone spontaneously hypertensive rat. *Cell*, **67**: 213-224.

Jost, L. 2008. G_{ST} and its relatives do not measure differentiation. *Molecular Ecology*, **17**: 4015-4026.

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

Kearse, M., Moir, R., Wilson, A. et al. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**: 1647–1649.

Liu, Y-P., Su, X., Luo, W-C., Lv, T., Chen, K-L., Harris, A. J. and Shah, S. A. 2017. Development of SSR markers from transcriptomes for *Orinus* (Poaceae), an endemic of the Qinghai-Tibetan Plateau. *Applications in Plant Sciences*, **5**: doi:10.3732/apps.1700029.

Meglecz, E., Pech, N., Gilles, A., Dubut, V., Hingamp, P., Trilles, A., Grenier, R. and Martin, J. F. 2014. QDD version 3.1: A user friendly computer program for microsatellite selection and primer design revisited: Experimental validation of variables determining genotyping success rate. *Molecular Ecology Resources*, **14**: 1302–1313.

Merrit, B. J., Culley, T. M., Avanesyan, A., Stokes, R. and Brzyski, J. 2015. An empirical review: characteristics of plant microsatellite markers that confer higher levels of genetic variation. *Applications in Plant Sciences*, **3**: doi:10.3732/apps.1500025.

Nazareno, A. G., Pereira, R. A. S., Feres, J. M., Mestriner, M. A. and Alzate-Marin, A. L. 2009. Transferability and characterization of microsatellite markers in two Neotropical *Ficus* species. *Genetics and Molecular Biology*, **32**: 568–571.

Nunome, T., Negoro, S., Miyatake, K., Yamaguchi, H. and Fukuoka, H. 2006. A protocol for the construction of microsatellite enriched genomic library. *Plant Molecular Biology Reporter*, **24**: 305–312.

Parida, S.K., Kalia, S. K., Sunita, K. et al. 2009. Informative genomic microsatellite markers for efficient genotyping applications in sugarcane. *Theoretical and Applied Genetics*, **118**: 327–338.

Peakall, R. and Smouse, P. E. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics Applications Note*, **28**: 2537–2539.

Powell, W., Machray, G. C. and Provan, J. 1996. Polymorphism revealed by simple sequence repeats. *Trends in Plant Sciences*, **1**: 215-222.

Ramsay, L., Macaulay, M., degli Ivanissevich, S. et al. 2000. A simple sequence repeat-based linkage map of barley. *Genetics*, **156**: 1997–2005.

Riser, J. P., Schwabe, A. L. and Neale, J. R. 2017. Novel microsatellite development and characterization for *Phacelia formosula* (Hydrophyllaceae). *Applications in Plant Sciences*, **5**: doi:10.3732/apps.1700030.

Rozen, S. and Skaletsky, H. 1999. Primer3 on the WWW for general users and for biologist programmers. In: *Methods in molecular biology, vol. 132: Bioinformatics methods and protocols* (Misener, S. and Krawetz, S. A. eds.), pp. 365–386. Humana Press, Totowa.

Saha, M. C., Cooper, J. D., Mian, M. A. R. and Chekhovskiy, K. 2006. Tall Fescue genomic SSR markers: development and transferability across multiple grass species. *Theoretical and Applied Genetics*, **113**:1449–1458.

Schoebel, C. N., Brodbeck, S., Buehler, D. et al. 2013. Lessons learned from microsatellite development for nonmodel organisms using 454 pyrosequencing. *Journal of Evolutionary Biology,* **26**: 600–611.

Shi, X., Wu, H., Xie, H., Zheng, Y., Huang, Y. and Yu, S. 2017. Development of microsatellite markers for *Eurya acuminatissima* (Theaceae). *Applications in Plant Sciences*, **5**: doi:10.3732/apps.1700037.

Stagel, A., Portis, E., Toppino, L., Rotino, G. L. and Lanteri, S. 2008. Gene-based microsatellite development for mapping and phylogeny studies in eggplant. *BMC Genomics,* **9**: 357.

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

Tautz, D. and Renz, M. 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Research,* **12**: 4127–4138.

Tautz, D. and Schlotterer, C. 1994. Simple sequences. *Current Opinion in Genetics & Development,* **4**: 832 – 837.

Thiel, T., Michalek, W., Varshney, R. K. and Graner, A. 2003. Exploiting EST databases for the development and characterization of gene derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics,* **106**: 411–422.

Tosso,F. N., Doucet, J-L., Migliore, J., Daïnou, K., Kaymak, E., Kameni, F. S. N. and Hardy, O. J. 2017. Characterization of mimrosatellite markers in the African tropical tree species *Guibourtia ehie* (Fabaceae, Detarioideae). *Applications in Plant Sciences*, **5**: doi:10.3732/apps.1700023.

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

Viard, F., Franck, P., Dubois, M. P., Estoup, A. and Jarne, P. 1998. Variation of microsatellite size homoplasy across electromorphs, loci, and populations in three invertebrate species. *Journal of Molecular Evolution*, **47**: 42–51.

Wang, M. L., Barkley, N. A. and Jenkins, T. M. 2009. Microsatellite markers in plants and insects. Part I. Applications of biotechnology. *Genes, Genomes and Genomics*, **3**: 54–67.

Wang, M. L., Gillaspie, A. G., Newman, M. L., Dean, R. E., Pittman, R. N., Morris, J. B. and Pederson, G. A. 2004. Transfer of simple sequence repeat (SSR) markers across the legume family for germplasm characterization and evaluation. *Plant Genetic Resources*, **2**: 107–119.

Wang, X., Lu, P. and Luo, Z. 2013. GMATo: A novel tool for the identification and analysis of microsatellites in large genomes. *Bioinformation*, **9**: 541–544.

Weber, J. L. and May, P. L. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics,* **44**: 388–396.

Zalapa, J. E., Cuevas, H., Zhu, H., Steffan, S., Senalik, D., Zelden, E., McCown, B., Harbut, R. and Simon. P. 2012. Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. *American Journal of Botany,* **99**: 193–208.

4. A MOUNTAIN RANGE IS A STRONG GENETIC BARRIER BETWEEN POPULATIONS OF *AFZELIA QUANZENSIS* (POD MAHOGANY) WITH LOW GENETIC DIVERSITY

4.1 Introduction

The spatial distribution of genetic variation has important evolutionary and conservation implications (Allendorf and Luikart 2007; Frankham 2010; Budd et al. 2015). Analysis of population genetic structure can help identify genetic discontinuities caused by human activities or landscape features, as well as identifying genetically depauperate populations that might have higher probabilities of inbreeding depression and lower probabilities of persistence (Suni and Whiteley 2015). Restricted gene flow among disjunct plant populations may lead to increased genetic differentiation and an accompanying reduction in genetic diversity following genetic drift. Geographic separation may lead to reproductive isolation (Ghazoul 2005; Lowe et al. 2005). Reproductive isolation may cause mating of related individuals, and the resultant inbreeding depression which decreases fitness. However, these outcomes are not inevitable in some tree species that can maintain high levels of long-distance seed dispersal and pollen-mediated gene flow (Ashley 2010).

Generally, many tree species appear to be outcrossing, showing high levels of gene flow and genetic diversity, and low genetic differentiation at the local and regional spatial scales (Ahmed et al. 2009; Craft and Ashley 2010; Omondi et al. 2010; Abraham et al. 2011; Addisalem et al. 2016). Successful long-distance dispersal of both seeds and especially pollen in

many tree species may maintain genetic connectivity and prevent structuring over large geographic ranges. However, in some phylogeographic studies of tropical African tree species, high regional differentiation has been detected, and these studies have suggested that landscape features play a major role in structuring genetic variation in east, central and west Africa. The African rift valley (Muchugi et al. 2006), the Dahomey Gap (Fontaine et al. 2004; Karan et al. 2012; Hardy et al., 2013), the Mega Chad Lake (Tsy et al. 2009), and the Adamawa Mountains (Allal et al. 2011) have been identified as African landscape features that restrict gene flow.

In southern Africa, several mountain ranges have appeared since the break-up of Gondwana (Moore 1999; Moore and Larkin 2001; Goudie 2005). These mountain ranges severed drainage systems and are potentially disrupting genetic connectivity among low-altitude plant populations. One of the longest mountain ranges in southern Africa, the Kalahari-Zimbabwe (KZ) axis (Moore et al. 2009), stretches across Namibia, Botswana, and central Zimbabwe. The KZ axis, caused by a late Palaeogene event (~43-33 Ma) (Moore et al. 2009), separated the northerly Zambezi from the southerly Limpopo drainage systems in Zimbabwe. The KZ axis severed the range of low-altitude plant species, and shaped those that re-established following the climatic oscillations of the Pleistocene, into northern and southern distributions in Zimbabwe (Moore, 1988). The impact of the southern African mountain ranges, including the KZ axis, in shaping genetic diversity in the subcontinent has not been studied. Plant species separated by the KZ axis into northern and southern distributions include *Commiphora merkei*, *Xanthocercis zambesiaca*, *Exoecaria bussei*, *Hippocratea parvifolia*, *Cordia goetzei* (Moore 1988), and *Afzelia quanzensis* (van Wyk and van Wyk 1997), the subject of our research.

Afzelia quanzensis is an economically important species that is increasingly being threatened by anthropogenic activities in its native range. It is a medium to large deciduous tree (up to 35 m height) whose native range is eastern and southern Africa, including Kenya, Angola, Tanzania, Democratic Republic of Congo, Malawi, Mozambique, Zimbabwe, Botswana, Zambia, and South Africa (Germishuizen et al. 2005). It is a low altitude species, and as such, populations are separated into northern and southern distributions by the KZ axis in Zimbabwe.

The IUCN has regionally listed *A. quanzensis* as vulnerable in Malawi, while in Mozambique and Zimbabwe it is listed as near threatened and at lower risk, respectively. Information on the level of genetic diversity, genetic structure of disjunct populations, and how landscape features affect gene flow, are critical for genetic resource management. Genetic rescue, demarcation of management units, and identification of source populations for germplasm collection are informed by genetic structure and levels of genetic diversity.

The aim of this study was to genetically characterize *A. quanzensis* populations separated by the KZ axis. I specifically wanted to (1) determine the levels of genetic diversity of the spatially separated populations (2) determine genetic differentiation among the populations and (3) identify genetic discontinuities, if they exist, and correlate them to landscape features. To achieve these objectives, I used nuclear microsatellites for genotyping. The hypervariability, codominant inheritance, transferability to congeners, and putative neutrality of microsatellite loci make them ideal markers for population genetics studies (Selkoe and Toonen 2006; Ashley 2010). Genetic characterization of the spatially separated *A. quanzensis* populations will help to inform local conservation management decisions.

4.2 Materials and Methods

4.2.1 Study species

Afzelia quanzensis is one of seven species in the genus *Afzelia* that occur naturally in Africa. The genus belongs to the pea family, Fabaceae, and subfamily Caesalpinioideae. *Afzelia quanzensis* is a diploid species ($2n = 24$) (Turner and Fearing 1959) which has hermaphrodite flowers that undergo petal and stamen suppression during development (Turker 2002). Flowers are sweet-scented with red-pink petals which attract honey bees. Bees have been identified as pollination agents in *Afzelia* species (Kato et al. 2008; Ariwaodo and Harry-Asobara 2015). Seeds have a nutritious red aril that is sought after by animals such as baboons, monkeys, and squirrels (Gathua 2000). Monkeys (Gathua 2000) and hornbills (Germishuizen et al. 2005) are seed dispersers, while other animals are frugivores.

In Zimbabwe, the species is used heavily in woodcarving (Braedt and Standa-Gunda 2000). The failure of local and national institutions to regulate harvesting of the species has made it a *de facto* open access resource in unprotected forests (Braedt and Schroeder 2003). Combined with the slow growth rate of the species, the rate of harvesting may be unsustainable (Gerhardt and Todd 2009).

4.2.2 Sampling and microsatellite genotyping

Leaf material of *A. quanzensis* was collected at nine sites in Zimbabwe, five north of the KZ axis (N = 135), hereafter referred to as northern sites, and four in the south (N = 57), hereafter referred to as southern sites (Figure 4.1). Leaves were collected from mature individuals (at least 31 cm at DBH). I sampled around homesteads in eight collection sites as we could not find the species in the wild due to illegal logging. Only one sampling locality, N5, was in a protected forest managed by the Forestry Commission of Zimbabwe. There was no evidence of recruitment at any of the sites. All individuals available at a collection locality were sampled, and these individuals were at least 100 meters apart.

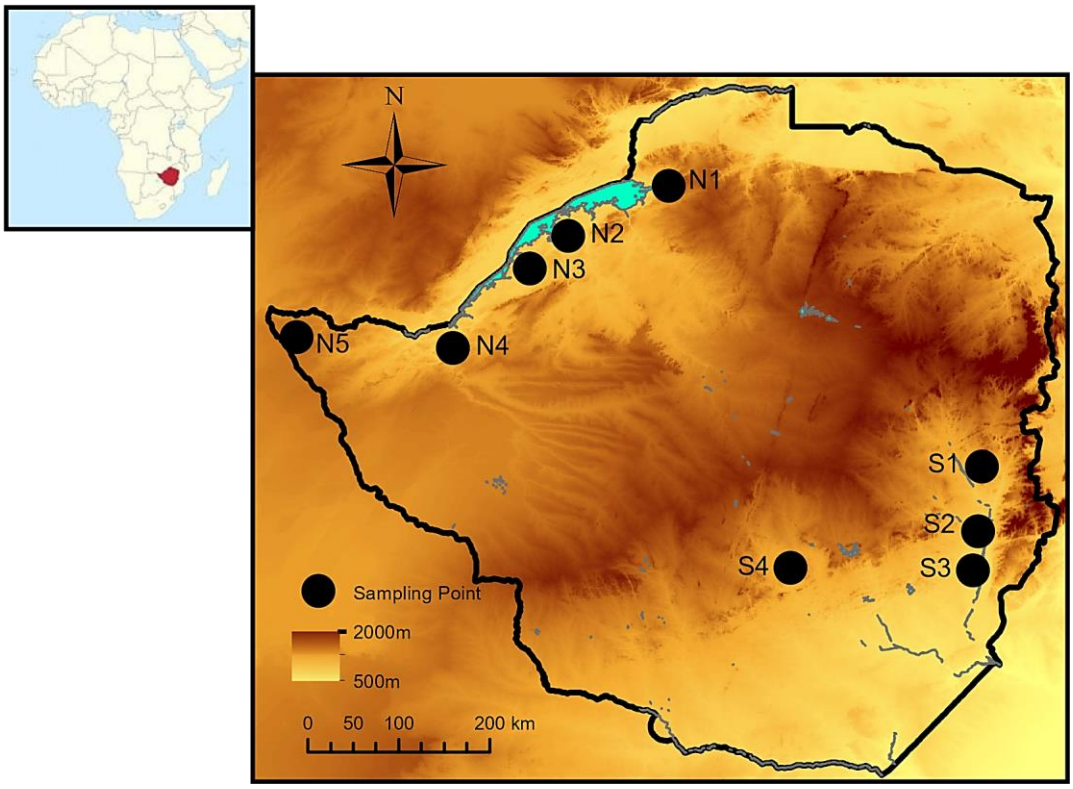


Figure 4.1. Location of collection sites of *Afzelia quanzensis* in Zimbabwe. The Kalahari-Zimbabwe axis is the high-altitude area which stretches from the south-west to the north-east of the country, and separates the northern (N1-N5) from the southern (S1-S4) collection sites

Leaves were dried in silica gel and stored at -20 ˚C until DNA extraction. Number of samples per site ranged from 8 to 32, and geographic coordinates of collection localities are shown in Table VI. In total, 192 samples were collected for genotyping.

Genomic DNA was extracted from approximately 20 mg of leaf tissue using the Qiagen Plant Mini Kit (Qiagen, Valencia, CA, USA). Using a Nanodrop spectrophotometer (Nanodrop Technologies, Wilmington, DE, USA), DNA concentrations were measured, and diluted samples to approximately 10 ng/µL. Samples were genotyped at 10 nuclear microsatellite loci developed by Jinga et al. (2016) for *A. quanzensis*, following their PCR conditions. The 10 loci were Afq12, Afq35, Afq43, Afq44, Afq45, Afq49, Afq62, Afq67, Afq68, and Afq69. PCR amplicons were analysed using an ABI 3730*xl* DNA analyser (Applied Biosystems, Foster City, CA, USA), with ALEXA-725 ladder (Maddox and Feldheim 2014). The microsatellite genotypes were scored using GeneMapper software version 3.7 (Applied Biosystems).

4.2.3 Genetic diversity estimates

The 10 loci were checked for presence of null alleles using the program Micro-Checker version 2.2.3 (van Oosterhout et al. 2004), while the program Arlequin version 3.5 (Excoffier and Lischer 2010) was used to perform exact tests with a Markov chain (1 000 000 Markov chain steps and 100 000 dememorization steps) for deviations from Hardy-Weinberg equilibrium for each collection site. Genetic diversity estimates were calculated using the program GenAlEx version 6.5 (Peakall and Smouse 2006), including mean number of alleles per locus (*A*), effective number of alleles (A_E), observed heterozygosity (H_O), and expected heterozygosity (H_E).

GenAlEx version 6.5 was also used to calculate fixation indices (F_{IS}). Allelic richness (A_R), and private allelic richness (PA_R) were estimated using the rarefaction method implemented in the program HP-Rare version 1.1 (Kalinowski 2005). The northern populations were pooled, as well as the southern populations, and number of private alleles were determined in each pool. Population declines of northern and southern sites were estimated separately using the Garza-Williamson index (M), where $M = k / r$, when k is the total number of alleles, and r is the overall range in allele size (Garza and Williamson 2001). When a population declines, genetic drift is likely to occur, resulting in a loss of alleles. The loss of any allele contributes to a reduction in k, while the loss of the largest and smallest alleles contributes to a reduction in r, thus k is expected to decline quicker than r. The ratio $M = k / r$ is therefore expected to be smaller in recently reduced populations. The Garza-Williamson index was calculated using Arlequin version 3.5. The index can identify populations that have undergone recent declines from those that have been historically small (Garza and Williamson 2001).

4.2.4 Population differentiation

Population differentiation statistics, F_{ST} (Weir and Cockerham 1984), G_{ST} (Nei and Chesser 1983), G'_{ST} (Hedrick 2005), and D_{JOST} (Jost 2008), were calculated together with their 95% confidence intervals, using the R (R Development Core Team 2016) package diveRsity (Keenan et al. 2013). G'_{ST} and D_{JOST} overcome the drawback associated with F_{ST} and G_{ST} of their maximum possible values being dependent on within-population diversity (Jost 2008; Hedrick 2005). Population differentiation statistics were plotted against number of alleles using the corPlot function in diveRsity to determine whether the microsatellite loci are useful in inferring

demographic processes. If differentiation estimates are approximately equal, this suggests that demographic processes have similar effects across neutral loci differing in mutation rates (Keenan et al. 2013). Pairwise differentiation estimates of D_{JOST} and F_{ST} were also calculated using the R package diveRsity. Isolation by distance (IBD) was tested on the total sample set, and separately on northern, and southern sites by Mantel tests in GenAlEx based on distance matrices of pairwise F_{ST} and geographic distance between sample locations. The number of permutations for all Mantel tests was 10 000.

4.2.5 Genetic clustering and identification of genetic barriers

A Bayesian clustering method, implemented in the program STRUCTURE version 2.3.4, was used to delineate genetic clusters based on the analysis of individual multilocus genotypes and to assign individuals to the cluster where their posterior probability is highest (Pritchard et al. 2000). The STRUCTURE approach is a non-spatial model that does not consider sampling locations. The number of assumed genetic clusters (K) was set from one to nine. Ten runs with 100 000 MCMC iterations were performed for each K following a burn-in of 50 000 iterations. The admixture model and correlated allele frequencies were selected. An optimum K value was selected by comparing delta K values (Evanno et al. 2005), and a plot of ln Pr (X|K) using STRUCTURE HARVESTER version 0.6.94 (Earl and vonHoldt 2012). The two plots were included to infer the optimum value of K because the delta K method does not allow assessment at K = 1 (Janes et al. 2017). CLUMPP version 1.1.2 (Jakobsson and Rosenberg 2007) was used to perform consensus analyses on the average scores for the inferred K value. STRUCTURE output was visualized using STRUCTURE PLOT (Ramasamy et al. 2014).

A principal components analysis (PCA) was performed using the R package adegenet version 2.0.1 (Jombart 2008) to summarize genetic similarities among individuals. Missing genotypes were replaced by mean allele frequency of the respective collection site. A Monmonier's function in adegenet version 2.0.1 identified paths of strongest genetic distances between populations. The coordinates of the sites were used to build the initial connection network. F_{ST} values were used in a matrix of pairwise genetic distances for populations, and random noise was removed by principal coordinates analysis using the adegenet function dudi.pco. The Monmonier's function then generated a genetic boundary based on maximum pairwise genetic distances. Estimates of population differentiation statistics, genetic clustering, and identification of genetic discontinuities are distinct but complementary analyses that can increase confidence to conclusions derived using different methods.

4.2.6 Estimation of time of divergence and migration rates

STRUCTURE analysis and PCA revealed two highly distinct gene pools, one made up of northern samples, and the other of southern samples (see Results). To estimate time of divergence between the two gene pools, samples from the five northern sites were pooled to form a single large northern population. Likewise, samples from the four southern sites were pooled into one southern population. The time of divergence of the two gene pools was estimated by approximate Bayesian computation (ABC) under the scenario shown on Figure 4.2, implemented in the program *DIY ABC* 2.1.0 (Cornuet et al. 2008). A reference table, which forms the basis of parameter estimation, consisted of 1 000 000 simulated data sets. Each simulation was based on

a set of one-sample and two-sample summary statistics. One-sample summary statistics were mean number of alleles, mean genic diversity, mean allele size variance, and mean Garza-Williamson's index. Two-sample summary statistics were mean number of alleles per locus, mean genic diversity, mean allele size variance, F_{ST} between the two gene pools (Weir and Cockerham, 1984), mean index of classification (Pascual et al., 2007), shared allele distance, and genetic distance [$(d\mu)^2$] (Goldstein et al., 1995) between the gene pools. The generalized stepwise mutation model (Estoup et al. 2002) and the default mean mutation rate across all loci of 10^{-4} to 10^{-3} were used.

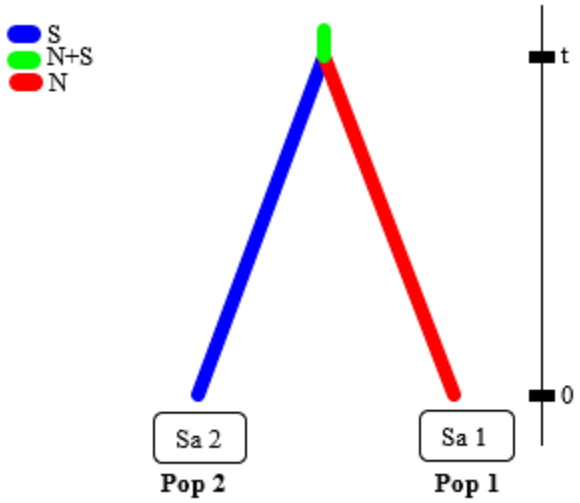


Figure 4.2. Evolutionary scenario implemented in approximate Bayesian computation using the software *DIY ABC* 2.1.0 (Cornuet et al. 2008) to estimate time since divergence of two gene pools of *Afzelia quanzensis*

Migration rates were estimated using a Bayesian method implemented in the program BayesAss edition 3.0.4 (Wilson and Rannala 2003). The number of iterations for the MCMC was set at 10 000 000, with a burn-in length of 1 000 000, and sampling after every 1 000 iterations.

4.3 Results

The 10 microsatellite loci were moderately polymorphic, with number of alleles per locus ranging from 3 to 15. A total of 79 alleles were identified among the 192 individuals, with a mean of 4.367 per locus. There was no evidence of null alleles across all loci, and all loci showed no departure from Hardy-Weinberg equilibrium, except for loci Afq44 and Afq49 at S3. We attributed this departure to low sample size at S3. The mean estimates of genetic diversity among all sites were 4.367, 2.917, 2.208, 0.197, 0.466 and 0.452 for A, A_R, A_E, PA_R, H_O and H_E, respectively (Table VI). The mean estimates of genetic diversity in the northern sites were 4.500, 2.798, 2.130, 0.178, 0.443, and 0.432 for A, A_R, A_E, PA_R, H_O and H_E, respectively. In southern sites, the mean genetic diversity estimates were 4.200, 3.065, 2.304, 0.220, 0.493, and 0.478 for A, A_R, A_E, PA_R, H_O and H_E, respectively. There was no significant difference in H_E (t test, P = 0.078), A (t test, P = 0.359), PA_R (t test, P = 0.119) and A_E (t test, P = 0.062) between northern and southern sites. Allelic richness was significantly different between northern and southern sites (t test, P = 0.007). Private allele analysis revealed 17 alleles that were only observed in the northern gene pool and nine that were only observed in the southern gene pool. The M for the pooled northern sites was 0.370, while that of the southern sites was 0.354. A value of $M < 0.68$ suggests a recent population decline, particularly for data with seven or more loci (Garza and Williamson 2001).

Table VI. Collection sites, latitude and longitude coordinates, number of individuals sampled (*N*), mean number of alleles per locus (*A*), allelic richness (A_R) effective number of alleles (A_E), private allelic richness (PA_R), observed heterozygosity (H_O), expected heterozygosity (H_E), and fixation indices (F_{IS}) for *Afzelia quanzensis* samples

Collection Site	Latitude, Longitude	*N*	*A*	A_R	A_E	PA_R	H_O	H_E	F_{IS}
N1	16°58.415′, 29°20.043′	32	4.800	2.810	2.133	0.200	0.405	0.424	0.068
N2	17°07.965′, 28°20.315′	27	4.700	2.950	2.249	0.150	0.450	0.448	-0.009
N3	17°40.027′, 27°28.198′	22	3.800	2.710	2.050	0.120	0.445	0.417	-0.089
N4	18°19.147′, 27°06.528′	24	4.400	2.690	2.100	0.220	0.452	0.412	-0.091
N5	18°07.836′, 25°51.853′	30	4.800	2.830	2.119	0.200	0.464	0.461	-0.012
S1	19°36.056′, 32°30.084′	15	4.000	3.050	2.482	0.200	0.527	0.507	-0.033
S2	20°00.379′, 32°25.953′	18	4.700	3.180	2.331	0.200	0.507	0.507	-0.009
S3	20°38.809′, 32°21.140′	8	3.400	2.920	2.223	0.230	0.448	0.415	-0.104
S4	20°36.529′, 30°40.087′	16	4.700	3.110	2.181	0.250	0.491	0.481	-0.019
Total or Mean		192	4.367	2.917	2.208	0.197	0.466	0.452	-0.032

Global estimates of population differentiation statistics were F_{ST} = 0.0936, G_{ST} = 0.1001, G'_{ST} = 0.1982, and D_{JOST} = 0.0598. Bootstrapping estimates of the lower 95% confidence interval for the population differentiation statistics did not go below zero, indicating significant population differentiation. Pairwise estimates of D_{JOST} and F_{ST} showed significant differentiation between northern and southern sites (Figure 4.3). Slopes of F_{ST} and G_{ST} against number of alleles were slightly negative, while that of G'_{ST} was zero. A slope of D_{JOST} against number of alleles was slightly positive. However, all the slopes that showed a gradient were not significant, suggesting that these population differentiation statistics should be useful in inferring past demographic processes, including past gene flow trends.

Results of the Mantel test using the total sample set showed a significant positive correlation between genetic and geographic distances, Rxy = 0.431 (P = 0.028). For the northern sites only, there was a positive but not significant correlation between genetic and geographic distances, Rxy = 0.366 (P = 0.182), and for the southern sites, the correlation was negative and not significant, Rxy = -0.226 (P = 0.495).

A.

	N1	N2	N3	N4	N5	S1	S2	S3	S4
N1									
N2	0.0080								
N3	0.0118	0.0073							
N4	0.0063	0.0033	0.0036						
N5	0.0165	0.0042	0.0151	0.0044					
S1	0.0771	0.0600	0.0580	0.0694	0.0836				
S2	0.0767	0.0657	0.0718	0.0913	0.0921	0.0146			
S3	0.1222	0.1183	0.1131	0.1324	0.1741	0.0154	0.0210		
S4	0.0339	0.0267	0.0411	0.0449	0.0476	0.0160	0.0091	0.0114	

B.

	N1	N2	N3	N4	N5	S1	S2	S3	S4
N1									
N2	0.0183								
N3	0.0228	0.0184							
N4	0.0258	0.0192	0.0131						
N5	0.0356	0.0149	0.0398	0.0264					
S1	0.1599	0.1222	0.1290	0.1526	0.1357				
S2	0.1815	0.1518	0.1679	0.1977	0.1704	0.0283			
S3	0.2513	0.2212	0.2284	0.2679	0.2519	0.0534	0.0545		
S4	0.1084	0.0814	0.1004	0.1279	0.1009	0.0376	0.0273	0.0443	

Figure 4.3. Pairwise population differentiation estimates with color gradients (A) D_{JOST} (Jost, 2008), and (B) F_{ST} (Hedrick, 2005), of *Afzelia quanzensis* samples from nine sites in Zimbabwe

STRUCTURE analysis revealed the highest delta *K* occurred at *K* = 2 at 10 simulations when *K* was ranged from 1 to 9 (Figure 4.4, Appendix C). A plot of ln Pr (X|*K*) shows the curvature plateauing at *K* = 2 and *K* = 3 (Figure 4.5, Appendix D). The delta *K* methods has been reported to frequently identify *K* = 2 as the optimum number of clusters even when more subpopulations are present (Janes et al. 2017). *K* = 2 was retained taking into consideration results of the PCA, the Monmonier's function, and the delta *K* method. The consensus membership coefficients at *K* = 3 are shown on Figure 4.6, Appendix E. At *K* = 2, one cluster was mostly made up of individuals from the northern sites, while most individuals from the southern sites made up the other cluster (Figure 4.7). Although there were a few admixed individuals, membership coefficients were generally high, with mean membership coefficient in the northern cluster of 0.9381, and for the southern cluster, the mean was 0.9322.

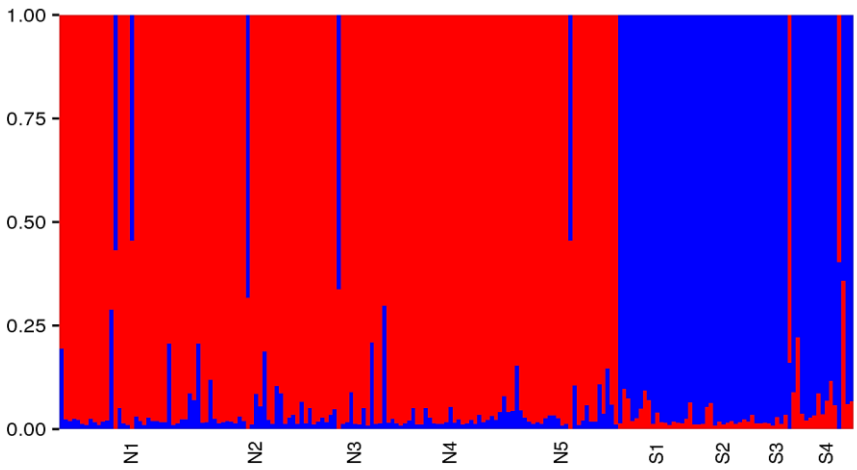


Figure 4.7. The proportion of membership coefficient for each individual in nine *Afzelia quanzensis* sites for inferred clusters when *K* = 2 according to STRUCTURE analysis (Pritchard et al., 2000) (red for northern, and blue for southern sites)

PCA identified two groups, with individuals from the northern and southern sites largely separated along PCA 1. One group was made up of individuals from southern sites, and the other of individuals from northern sites (Fig. 4.8). The Monmonier's function implemented in the R package adegenet identified a single genetic barrier between northern and southern sites (Figure 4.9, Appendix F). The identified genetic barrier coincided with the KZ axis.

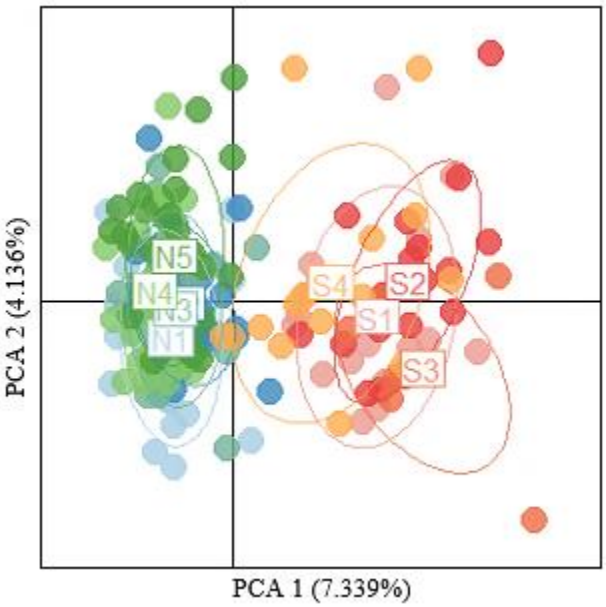


Figure 4.8. Principal components analysis of 192 *Afzelia quanzensis* individuals, based on multilocus genotypes, collected from nine sites in Zimbabwe

ABC, run in the program *DIY ABC* 2.1.0, estimated number of generations since divergence of the northern and southern gene pools at 323. A mean generation time of 150 years was assumed for *A. quanzensis* given that the mean generation time of most tropical, long-lived tree species range from 100 to 200 years (Baker et al. 2014). Therefore, the northern and southern gene pools have been isolated for over 45 000 years. The estimated effective population size of the northern gene pool was 4 060, while that of the southern gene pool was 4 310. The prior and posterior probability plots generated by ABC analysis are shown on Figure 4.10, Appendix G. Migration rate from the northern to the southern gene pool was estimated at 0.0041, that is the fraction of individuals in the southern gene pool that are migrants derived from the northern gene pool per generation. The migration rate from the southern to the northern gene pool was estimated at 0.0148.

4.4 Discussion

Genetic diversity is critical for adaptation and resistance to environmental changes, and consequently for the long-term survival of species. The microsatellite data in this study indicate a relatively low level of genetic diversity (H_E = 0.452, A = 4.367, A_E = 2.225, A_R = 2.917, PA_R = 0.197) for *A. quanzensis* among all sites compared to other long-lived tree species. Higher levels of genetic diversity have been observed in a congeneric species, *A. xylocarpa* (H_E = 0.620, A_R = 3.400) (Pakkad et al. 2014), and in a number of other trees, such as *Ficus sycomorus* (H_E = 0.663) (Ahmed et al. 2009), *Koompassia malaccensis* (H_E = 0.843 – 0.854) (Noreen and Webb 2013), *Dysoxylum malabaricum* (H_E = 0.626) (Bodare et al. 2013), and *Cabralea canjerana* (H_E

= 0.732, A_R = 6.32) (Melo and Franceschinelli 2016), all of which were analysed using nuclear microsatellite markers.

Low levels of genetic diversity, however, have been reported in some tree species, such as *Firmiana danxiaensis* (H_E = 0.364, A_E = 1.684, A_R = 2.444) (Chen et al. 2014), and *Fontainea picrosperma* (H_E = 0.407, A_R = 2.480, PA_R = 0.076) (Lamont et al. 2016). In these studies, low genetic diversity was attributed to effects of small and isolated populations (Chen et al. 2014; Lamont et al. 2016). The low values of *M* in both the northern and southern gene pools suggest the populations might have suffered a recent decline. Indeed, *Afzelia quanzensis* has been heavily logged, mainly for woodcarving, in the study area such that it was rare to find individuals in unprotected forests away from homesteads. Extreme deforestation reduces the representation of the original gene pool, and fragmented populations may continue to lose genetic diversity due to genetic drift. The long-term persistence of *A. quanzensis* in both the northern and southern sites in a changing environment may be threatened because deleterious effects of low genetic diversity may decrease reproductive success. However, if the individuals are well adapted to their microhabitats, low genetic diversity may not negatively impact the populations, and thus we recommend long-term monitoring of reproductive success because the populations likely suffered a recent decline.

Global estimates of population differentiation statistics showed significant differentiation among all sites. This is in contrast to many examples of spatially separated populations of long-lived tree species that typically show genetic uniformity because of insufficient length of time since isolation, extensive and recurrent gene flow, and high outcrossing rates. For example, in

the African fig tree, *F. sycomorus*, pollen was shown to be transferred for more than 160 km by the fig wasp, *Ceratosolen arabicus* (Ahmed et al. 2009). Such long-distance gene flow and high outcrossing rates result in low population differentiation across wide geographic areas in both temperate (Craft and Ashley 2010; Budd et al. 2015; Lumibao et al. 2016; Pazouki et al. 2016) and tropical (Dick et al. 2007; Fuchs and Hamrick 2010; Jones et al. 2013) forest tree species.

While genetic diversity studies of African vegetation have lagged behind work in temperate regions and the Neotropics, a number of landscape features in west, central, and east Africa have been shown to genetically structure plant populations. For example, in a phylogeographical study of the baobab tree, *Adansonia digitata*, an economically important and culturally significant tree, genetic differentiation was observed between populations separated by the Mega Chad Lake in Central Africa (Tsy et al. 2009). Population differentiation was observed between east and west African populations of the African mahogany, *Khaya senegalensis*, a species harvested for its excellent timber and veneer qualities (Karan et al. 2012). The *K. senegalensis* populations were separated by the Dahomey Gap, a dry corridor of mainly savannah that splits the west African rainforest into eastern and western parts. The Dahomey Gap also separates differentiated populations of the shea tree, *Vitellaria paradoxa*, into eastern and western gene pools (Fontaine et al. 2004). The African rift valley in east Africa and the Adamawa Mountains between Nigeria and Cameroon have been identified as genetic barriers for the African cherry, *Prunus africana* (Muchugi et al. 2006), and *V. paradoxa* (Allal et al. 2011), respectively.

In this study, pairwise estimates of population differentiation statistics and IBD analyses strongly indicate that there is gene flow among southern sites, and among northern sites, but very restricted between these two groups of sites. STRUCTURE analysis and PCA clearly delineate two gene pools, one composed of southern sites, and the other of northern sites. There was a relatively substantial number of private alleles in the northern and southern gene pools. The Monmonier's function detected a genetic barrier that coincided with the KZ axis, ABC analysis of time since divergence of the northern and southern gene pools post-dates the age of the axis, and migration rates between gene pools are very low. Collectively, these results indicate that the KZ axis is possibly a strong barrier to gene flow between southern and northern sites.

Landscape features restrict gene flow by hindering the movement of pollination agents and seed dispersers. The African honeybee, *Apis mellifera scutellata,* has been shown to transfer pollen up to 3.2 km (Dick et al. 2003). It is highly unlikely that honeybees may transfer pollen between southern and northern sites given the width of the KZ axis, which spans more than 300 km at some sections. Dispersal of *A. quanzensis* seeds is generally limited to short distances because the seeds are not swallowed (Gathua 2000). Seed dispersers eat the nutritious red aril and discard the viable seed, thus spending little time carrying the seed. It is also highly unlikely that seed dispersers, such as monkeys and hornbills, may travel across the KZ axis while carrying seeds. Thus, genetic connectivity is strong among *A. quanzensis* individuals on either side of the KZ axis, and highly restricted across the uplift. Genetic connectivity of other low-altitude plant species that are separated into southern and northern populations by the KZ axis, such as *C. merkei*, *X. zambesiaca*, *E. bussei*, *H. parvifolia* and *C. goetzei*, may also be restricted similarly.

STRUCTURE analysis identified a few admixed individuals (Fig. 4.4). Although the reason for this admixture is uncertain, a plausible explanation could be human-mediated seed dispersal across the KZ axis. *Afzelia quanzensis* seeds are often collected and strung into necklaces or made into trinkets that are sold as curios (Germishuizen *et al*., 2005). Long-distance seed dispersal is possible if these collected seeds escape and germinate in a different locality.

Apart from the KZ axis, other mountain ranges have emerged in southern Africa due to tectonic movements and volcanism since the break-up of Gondwana. These mountain ranges include the Escarpment, Bushmanland Harts, and Etosha-Griqualand-Transvaal axes (Moore et al. 2009). The uplifts potentially affected the distribution of many low-altitude plant populations, and may be barriers to gene flow among disjunct populations. The effect of these southern African landscape features on genetic differentiation of low-altitude plant species has not been investigated, but results of this study suggest that they might be having impacts similar to the KZ axis. Plant communities may be highly structured in the subcontinent, and this has implications for conservation of genetic resources. To conserve all extant genetic variation in this region, there is need to identify distinct gene pools that likely arose as a result of spatial separation.

4.5 Conclusions

This study indicates that there is low genetic diversity in *A. quanzensis* in both northern and southern sites, and that the KZ axis is possibly a strong genetic barrier between southern and northern individuals. The two groups of individuals have differentiated into distinct gene pools. The KZ axis appears to strongly impede the movement of pollination agents and seed dispersers,

and thus largely avoiding admixture between the gene pools. Considering these results, we make the following recommendations: (1) Individuals from all sites should be monitored in long-term studies for deleterious effects of low genetic diversity, (2) Genetic material from the distinct gene pools should separately be used in establishment of seed banks, and *in situ* and *ex situ* preservation sites, (3) Restoration should be done across the landscape with seeds coming from local trees, and (4) Anthropogenic threats, such as illegal logging, should be mitigated across the entire landscape to avoid the continued decline of genetic diversity in *A. quanzensis*.

4.6 References

Abraham, S. T., Zaya, D. N., Koenig, W. D. and Ashley, V. A. 2011. Interspecific and intraspecific pollination patterns of valley oak, *Quercus lobata*, in a mixed stand in coastal central California. *International Journal of Plant Science*, **172**: 691-699.

Addisalem, A. B., Bongers, F., Kassahun, T. and Smulders, M. J. M. 2016. Genetic diversity and differentiation of the frankincense tree (*Boswellia papyrifera* (Del.) Hochst) across Ethiopia and implications for conservation. *Forest Ecology and Management*, **360**: 253-260.

Ahmed, S., Compton, S. G., Butlin, R. K. and Gilmartin, P. M. 2009. Wind-borne insects mediate directional pollen transfer between desert fig trees 160 kilometers apart. *Proceedings of the National Academy of Sciences*, **106**: 20342-20347.

Allal, F., Sanou, H., Millet, L. et al. 2011. Past climate changes explain the phylogeography of *Vitellaria paradoxa* over Africa. *Heredity*, **107**: 174-186.

Allendorf, F. W. and Luikart, G. 2007. *Genetics and the conservation of populations*. Blackwell Publishing, Hoboken.

Ariwaodo, J. O. and Harry-Asobara, J. L. 2015. Preliminary investigation on flowering and fruiting pattern in a plantation grown *Afzelia africana* Sm stand in Umuahia, Nigeria. *American Journal of Plant Science*, **6**: 219-227.

Ashley, V. M. 2010. Plant parentage, pollination, and dispersal: How DNA microsatellites have altered the landscape. *Critical Reviews in Plant Sciences*, **29**: 148-161.

Baker, R. T., Pennington, R. T., Magallon, S. et al. 2014. Fast demographic traits promote high diversification rates of Amazonian trees. *Ecology Letters*, **17**: 527-536.

Bodare, S., Tsuda, Y., Ravikanth, G., Shaanker, U. R. and Lascoux, M. 2013. Genetic structure and demographic history of the endangered tree species *Dysoxylum malabaricum* (Maliaceae) in Western Ghats, India: implications for conservation in a biodiversity hotspot. *Ecology and Evolution*, **3**: 3233-3248.

Braedt, O. and Schroeder, J. M. 2003. Institutions governing the use of forest products: woodcraft commercialisation in southern Africa. *Forests, Trees and Livelihoods*, **13**: 59-78.

Braedt, O. and Standa-Gunda, W. 2000 Woodcraft markets in Zimbabwe. *International Tree Crops Journal*, **10**: 367-384.

Budd, C., Zimmer, E. and Freeland, R. J. 2015. Conservation genetics of *Magnolia acuminata*, an endangered species in Canada: Can genetic diversity be maintained in fragmented, peripheral populations? *Conservation Genetics*, **16**: 1359-1373.

Chen, S., Li, M., Hou, R., Liao. W., Zhou, R. and Fan, Q. 2014. Low genetic diversity and weak population differentiation in *Firmiana danxiaensis*, a tree species endemic to Danxia landform in northern Guangdong, China. *Biochemical Systematics and Ecology*, **55**: 66-72.

Cornuet, J-M., Santos, F., Beaumont, M. A. et al 2008. Inferring population history with *DIY ABC*: a user-friendly approach to approximate Bayesian computation. *Bioinformatics*, 24: 2713-2719.

Craft, J. K. and Ashley, V. M. 2010. Pollen-mediated gene flow in isolated and continuous stands of bur oak, *Quercus macrocarpa* (Fabaceae). *American Journal of Botany*, **97**: 1999-2006.

Dick, W. C., Bermingham, E., Lemes, R. M. and Gribel, R. 2007. Extreme long-distance dispersal of the lowland tropical rainforest tree *Ceiba pentandra* L. (Malvaceae) in Africa and the neotropics. *Molecular Ecology*, **16**: 3039-3049.

Dick, W. C., Etchelecu, G., Austerlitz, F. 2003. Pollen dispersal of tropical trees (*Dinizia excelsa*: Fabaceae) by native insects and African honeybees in pristine and fragmented Amazonian rainforest. *Molecular Ecology*, **12**: 753-764.

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

Estoup, A., Jarne, P. and Cornuet, J-M. 2002. Homoplasy and mutation model at microsatellite loci and their consequences for population genetic analysis. *Molecular Ecology*, **11**: 1591-1604.

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

Excoffier, L. and Lischer, H. E. L. 2010. Arlequin suite ver3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**: 564–567.

Frankham, R. 2010. Challenges and opportunities of genetic approaches to biological conservation. *Biological Conservation*, **143**: 1919-1927.

Fontaine, C., Lovett, P. N., Sanou, H., Maley. J. and Bouvet, J. M. 2004. Genetic diversity of the shear tree (*Vitellaria paradoxa* C. F. Gaertn), detected by RAPD and chloroplast microsatellite markers. *Heredity*, **93**: 639-648.

Fuchs, E. J. and Hamrick, L. M. 2010. Spatial genetic structure within size classes of the endangered tropical tree *Guaiacum sanctum* (Zygophyllaceae). *American Journal of Botany*, **97**: 1200-1207.

Gathua, M. 2000. The effects of primates and squirrels on seed survival of a canopy tree, *Afzelia quanzensis*, in Arabuko-Sokoke forest, Kenya. *Biotropica*, **32**: 127-132.

Garza, J. C. and Williamson, E. G. 2001. Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology*, **10**: 305-318.

Gerhardt, K. and Todd, C. 2009. Natural regeneration and population dynamics of the tree *Afzelia quanzensis* in woodlands in Southern Africa. *African Journal of Ecology*, **47**: 583-591.

Germishuizen, G., Crouch, N. R. and Condy, G. 2005. *Afzelia quanzensis*. *Flowering Plants of Africa*, **59**: 74–83.

Ghazoul, J. 2005. Pollen and seed dispersal among dispersed plants. *Biological Reviews*, **80**: 413-443.

Goldstein, D. B., Linares, A. R., Cavalli-Sforza, L.L. and Feldman, M. W. 1995. An evaluation of genetic distances for use with microsatellite loci. *Genetics,* **139**: 463-471.

Gourdie, A. S. 2005. The drainage of Africa since the Cretaceous. *Geomorphology,* **67**: 437-456.

Hardy, O. J. Born, C., Budde, K. et al. 2013. Comparative phylogeography of African rain forest trees: A review of genetic signatures of vegetation history in the Guineo-Congolian region. *Comptes Rendus Geoscience*, **345**: 284-296.

Hedrick, P. W. 2005. A standardized genetic differentiation measure. *Evolution*, 59: 1633–1638.

Jakobsson, M. and 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**: 1801-1806.

Janes, J. K., Miller, J. M., Dupuis, J. R., Malenfant, R. M., Gorrell, J. C., Cullingham, C. I. and Andrew, R. L. 2017. The *K* = 2 conundrum. *Molecular Ecology*, **26**: 3594-3602.

Jinga, P., Palagi, J. and Ashley, V. A. 2016. Development of microsatellite loci of pod mahogany, *Afzelia quanzensis* (Fabaceae), by Illumina shotgun sequencing, and cross-amplification in *A. africana*. *Applications in Plant Sciences*, **4**: 16000010. doi:10.3732/apps.1600010.

Jombart, T. 2008. *adegenet*: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24: 1403-1405.

Jones, F. A., Cerón-Souza, I., Hardesty, B. D. and Dick. C. W. 2013. Genetic evidence of Quaternary demographic changes in four rain forest tree species sampled across the Isthmus of Panama. *Journal of Biogeography*, **40**: 720-731.

Jost, L. 2008. G_{ST} and its relatives do not measure differentiation. *Molecular Ecology*, 17:4015–4026.

Kalinowski, S. T. 2005. HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*, **5**:187–189.

Karan, M., Evans, D. S., Reilly, D. et al. 2012. Rapid microsatellite marker development for African mahogany (*Khaya senegalensis*, Maliaceae) using next-generation sequencing and assessment of its intra-specific genetic diversity. *Molecular Ecology Resources*, **12**: 344-353.

Kato, M., Kosaka, Y., Kawatika, A. et al. 2008. Plant-pollinator interactions in tropical monsoon forests in southeast Asia. *American Journal of Botany*, **95**: 1375-1394.

Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W. and Prodöhl, P. A. 2013. diveRsity: An R package for the estimation and exploration of population genetic parameters and their associated errors. *Methods in Ecology and Evolution*, **4**: 782–788.

Lamont, W. R., Conroy, C. G., Reddell, P. and Ogbourne, M. S. 2016. Population genetic analysis of a medicinally significant Australian rainforest tree, *Fontainea picrosperma* C.T. White (Euphorbiaceae): biogeographic patterns and implications for species domestication and plantation establishment. *BMC Plant Biology*, **16**: 57. doi:10.1186/s12870-016-0743-2.

Lowe, A. J., Boshier, D., Ward, M., Bacles, C. F. E. and Navarro, C. 2005. Genetic resource impacts of habitat loss and degradation; reconciling empirical evidence and predicted theory for neotropical trees. *Heredity*, **95**: 255-273.

Lumibao, Y. C., Gaskill, M., Flood, K. and Mclachlan, S. J. 2016. Persistence of genetic variation in recolonized *Tsuga canadensis* (Eastern hemlock) populations following historic forest clearance. *Plant Species Biology*, **31**: 73-79.

Maddox, D. J. and Feldheim, K. 2014. A cost-effective size standard for fragment analysis that maximizes throughput on five dye set platforms. *Conservation Genetics Resources*, **6**: 5-7.

Melo, O. D. T. A. and Franceschinelli, V. E. 2016. Gene flow and fine-scale spatial genetic structure in *Cabralea canjerana* (Maliaceae), a common tree species from the Brazilian Atlantic forest. *Journal of Tropical Ecology*, **32**: 135-145.

Moore, A. E. 1988. Plant distribution and the evolution of the major river systems in southern Africa. *South African Journal of Geology*, **91**: 346-349.

Moore, A. E. 1999. A reappraisal of apeirogenic flexure axes in southern Africa. *South African Journal of Geology*, **102**: 363-376.

Moore, A. E., Cotterill, F. P. W., Broderick, T. and Plowes, D. 2009. Landscape evolution in Zimbabwe from the Permian to present, with implications for kimberlite prospecting. *South African Journal of Geology*, **112**: 65-88.

Moore, A. E. and Larkin, P. A. 2001. Drainage evolution in south-central Africa since the breakup of Gondwana. *South African Journal of Geology*, **104**: 47-68.

Muchugi, A., Lengkeek, A. G., Kadu, C. A. C., Muluvi, G. M., Njagi, E. N. M.and 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**: 498-506.

Nei, M. and Chesser, R. K. 1983. Estimation of fixation indices and gene diversities. *Annals of Human Genetics*, **47**: 253-259.

Noreen, A. M. E. and Webb, L. E. 2013. High genetic diversity in a potentially vulnerable tropical tree species despite extreme habitat loss. *PLoS One,* **8**: e82632. doi:10.1371/journal.pone.0082632.

Omondi, S. F., Kireger, E., Dangasuk, O. G. et al. 2010. Genetic diversity and population structure of *Acacia senegal* (L) Willd. in Kenya. *Tropical Plant Biology,* **3**: 59-70.

Pakkad, G., Kanetani, S. and Elliott, S. 2014. Genetic diversity and differentiation of an endangered tree species, *Afzelia xylocarpa* (Kurz) craib in Thailand revealed by nuclear microsatellite markers. *African Journal of Biotechnology*, **13**: 366-377.

Pascual, M., Chapuis, M. P., Mestres, F. et al. 2007. Introduction history of *Drosophila subobscura* in the New World: a microsatellite based survey using ABC methods. *Molecular Ecology*, **16**: 3069-3083.

Pazouki, L., Shanjani, S. P., Fields, D. P. et al. 2016. Large within-population genetic diversity of the widespread conifer *Pinus sylvestris* at its soil fertility limit characterized by nuclear and chloroplast microsatellite markers. *European Journal of Forest Research*, **135**: 161-177.

Peakall, R. and Smouse, P. E. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**: 288−295.

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

R Development Core Team. 2016. *R*: *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available from http://www.r-project.org/.

Ramasamy, R. K., Ramasamy, S., Bindroo, B. B. and Naik, V. G. 2014. STRUCTURE PLOT: A program for drawing elegant STRUCTURE bar plots in user friendly interface. *SpringerPlus*, **3**: 431.

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

Suni, S. S. and Whiteley, A. R. 2015. Genetic structure of a montane perennial plant: the influence of landscape and flowering phenology. *Conservation Genetics*, **16**: 1431-1442.

Tsy, J. L. P., Lumaret, R., Mayne, D. et al. 2009. Chloroplast DNA phylogeography suggests a West African centre of origin for the baobab, *Adansonia digitata* L. (Bombacoideae, Malvaceae). *Molecular Ecology*, **18**: 1707-1715.

Turker, S. C. 2002. Comparative floral ontogeny in the Detarieae (Leguminosae: Caesalpinioideae). 2. Zygomorphic taxa with petal and stamen suppression. *American Journal of Botany*, **89**: 888-907.

Turner, B. L. and Fearing, O. S. 1959. Chromosome numbers in the Leguminosae II. African species, including phyletic interpretations. *American Journal of Botany*, **46**: 49-57.

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

van Wyk B, and van Wyk P (1997) *Field guide to trees of southern Africa*. Struik Publishers, Cape Town.

Weir, B.S. and Cockerham, C. C. 1984. Estimating F-statistics for the analysis of population structure. *Evolution*, **38**: 1358-1370.

Wilson, A. G. and Rannala, B. 2003. Bayesian inference of recent migration rates using multilocus genotypes. *Genetics,* **163**: 1177-1191.

APPENDICES

A

Bioclimatic variables used in modeling the distribution of *Afzelia quanzensis*

Bio1 = Annual Mean Temperature
Bio2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))
Bio3 = Isothermality (Bio2/Bio7) (* 100)
Bio4 = Temperature Seasonality (standard deviation *100)
Bio5 = Maximum Temperature of Warmest Month
Bio6 = Minimum Temperature of Coldest Month
Bio7 = Temperature Annual Range (Bio5-Bio6)
Bio8 = Mean Temperature of Wettest Quarter
Bio9 = Mean Temperature of Driest Quarter
Bio10 = Mean Temperature of Warmest Quarter
Bio11 = Mean Temperature of Coldest Quarter
Bio12 = Annual Precipitation
Bio13 = Precipitation of Wettest Month
Bio14 = Precipitation of Driest Month
Bio15 = Precipitation Seasonality (Coefficient of Variation)
Bio16 = Precipitation of Wettest Quarter
Bio17 = Precipitation of Driest Quarter
Bio18 = Precipitation of Warmest Quarter
Bio19 = Precipitation of Coldest Quarter

B

Table 11. Contribution of variables to a Maxent model of *Afzelia quanzensis*

Variable	Percent contribution	Permutation importance
Prec12	36.1	19.8
Bio4	10.1	9.7
Prec9	8.3	1.8
Bio14	6.4	0.6
Prec11	6.1	4.9
Prec10	5.2	5
Tmin6	3.7	0.5
Tmin1	3.4	0
Bio2	3.2	5.7
Prec1	2.3	18.8
Prec3	2	8.1
Bio12	1.9	2.3
Bio6	1.5	0.6
Tmax10	1.2	2.4
Tmin4	1.1	1
Prec2	0.9	2.8
Prec5	0.8	1.4
Prec6	0.7	0.1
Bio9	0.6	0.7
Bio15	0.6	1.8
Tmin10	0.5	1.2
Tmin8	0.4	0.2
Tmin12	0.3	0.1
Prec7	0.3	3.2
Bio7	0.3	0.8
Tmax8	0.2	0.7
Prec4	0.2	0.2
Bio1	0.2	0
Tmax3	0.2	0.3
Tmax2	0.2	1.2
Bio16	0.1	0.9
Tmax5	0.1	0.1
Bio18	0.1	0.7
Bio5	0.1	0.5
Tmin9	0.1	0.2
Tmax6	0.1	0.4

C

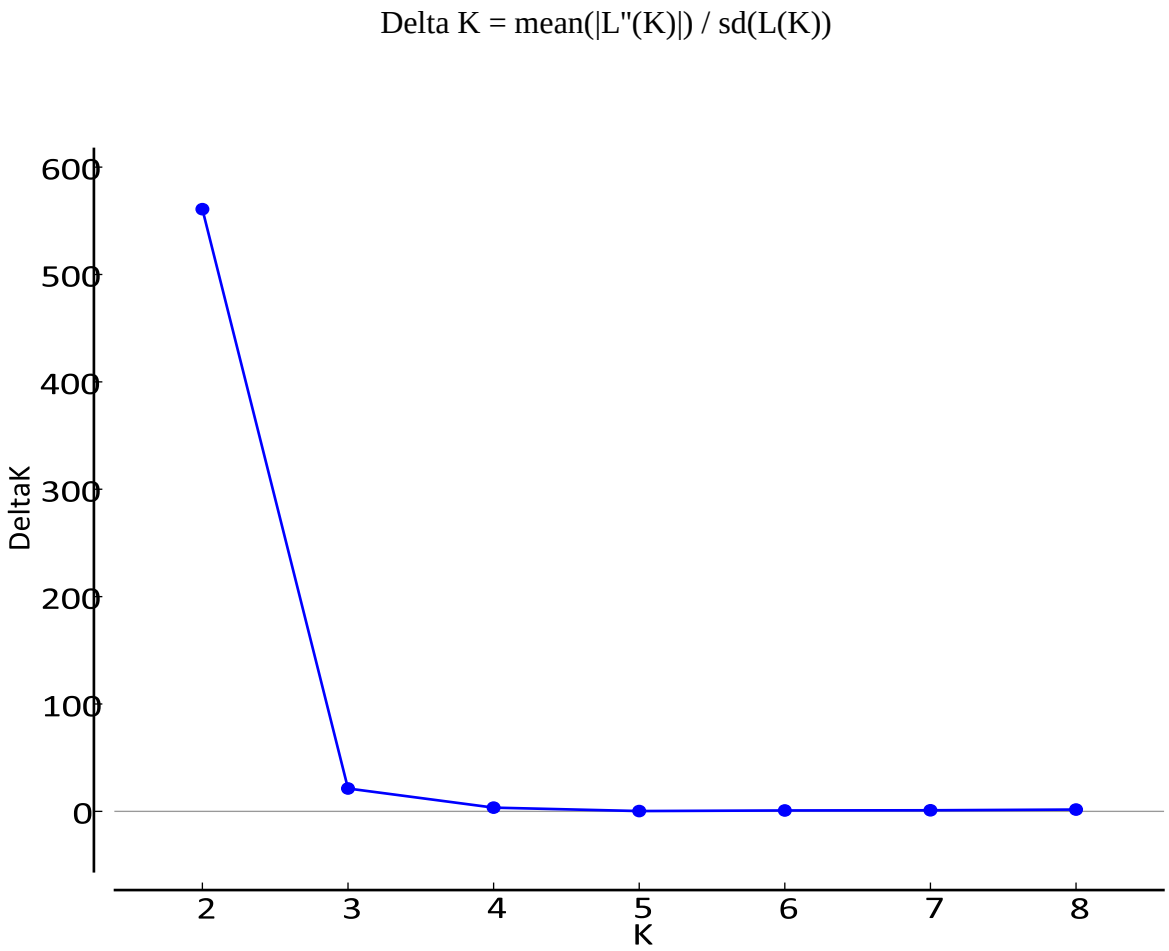


Figure 4.4. Values of delta *K* (Evanno et al. 2005) as a function of *K* (number of clusters) associated with results of STRUCTURE analysis (Pritchard et al., 2000) of *Afzelia quanzensis* individuals collected from nine sites in Zimbabwe

D

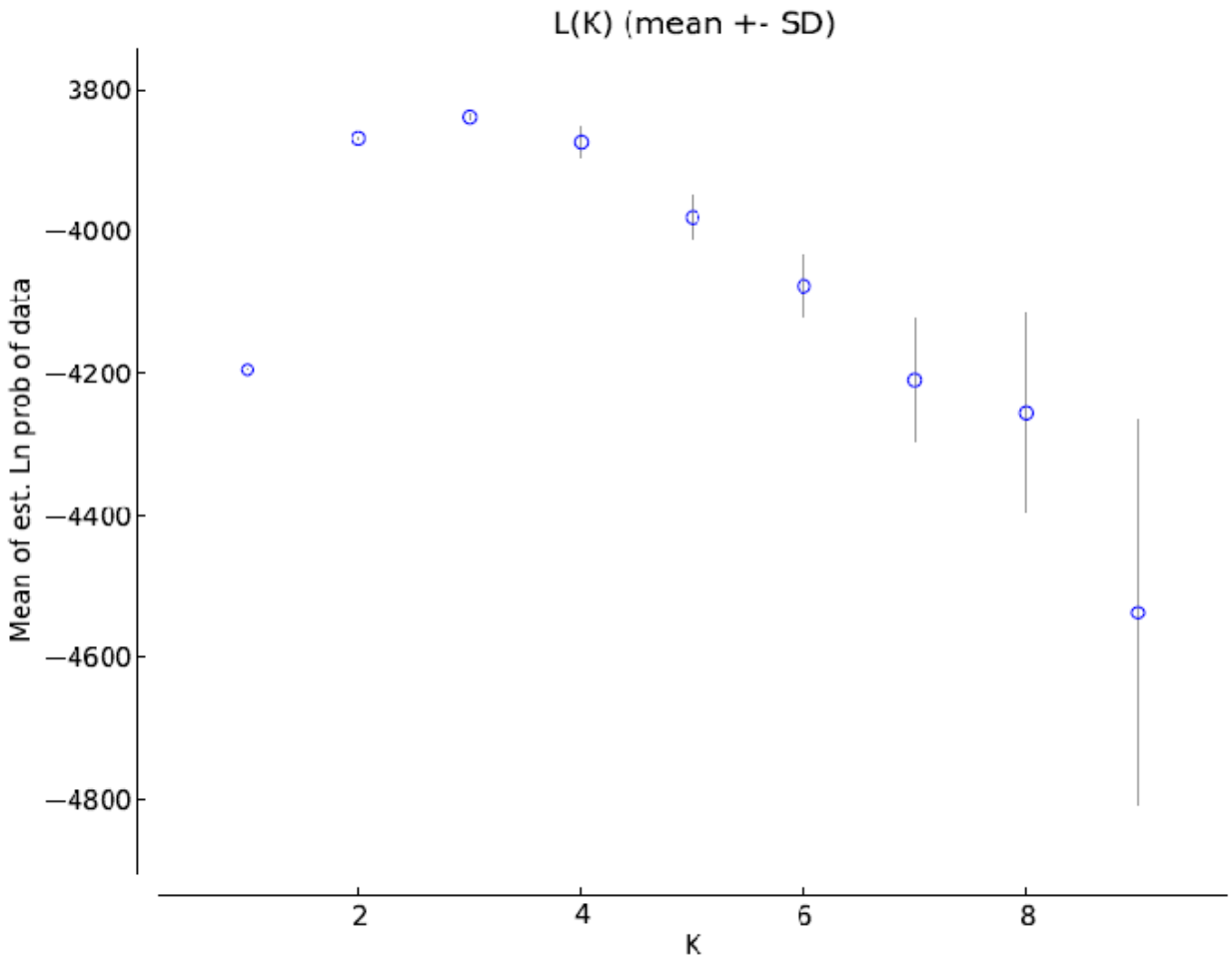


Figure 4.5. A plot of the estimated natural logarithm of the probability of data (ln Pr (X|*K*)) and standard deviations derived from 10 runs for each value of *K* = 1-9 in STRUCTURE analysis (Pritchard et al. 2000) of *Afzelia quanzensis* individuals collected from nine sites in Zimbabwe

E

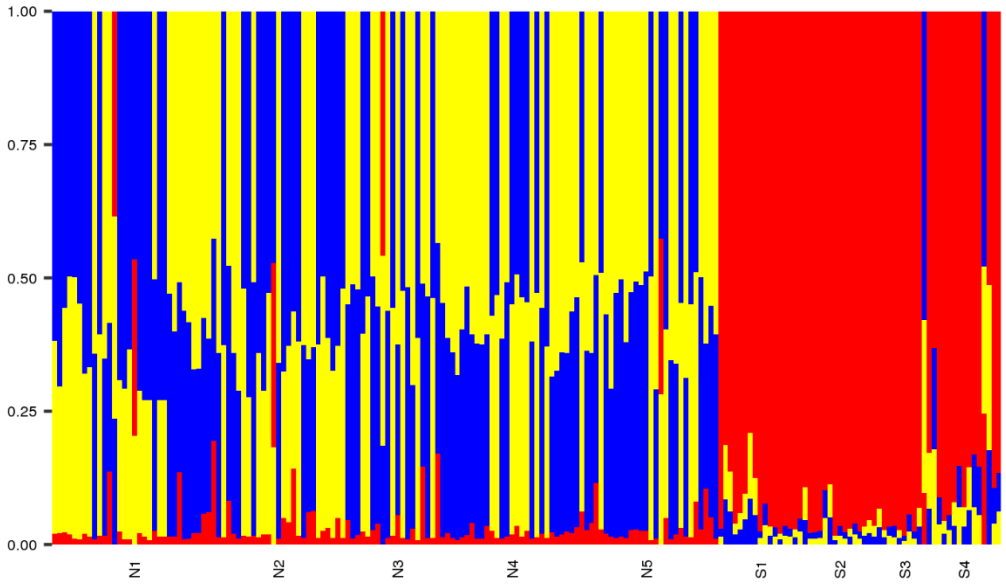


Figure 4.6. The proportion of membership coefficient for each individual in nine *Afzelia quanzensis* sites for inferred clusters when $K = 3$ according to STRUCTURE analysis (Pritchard et al. 2000)

F

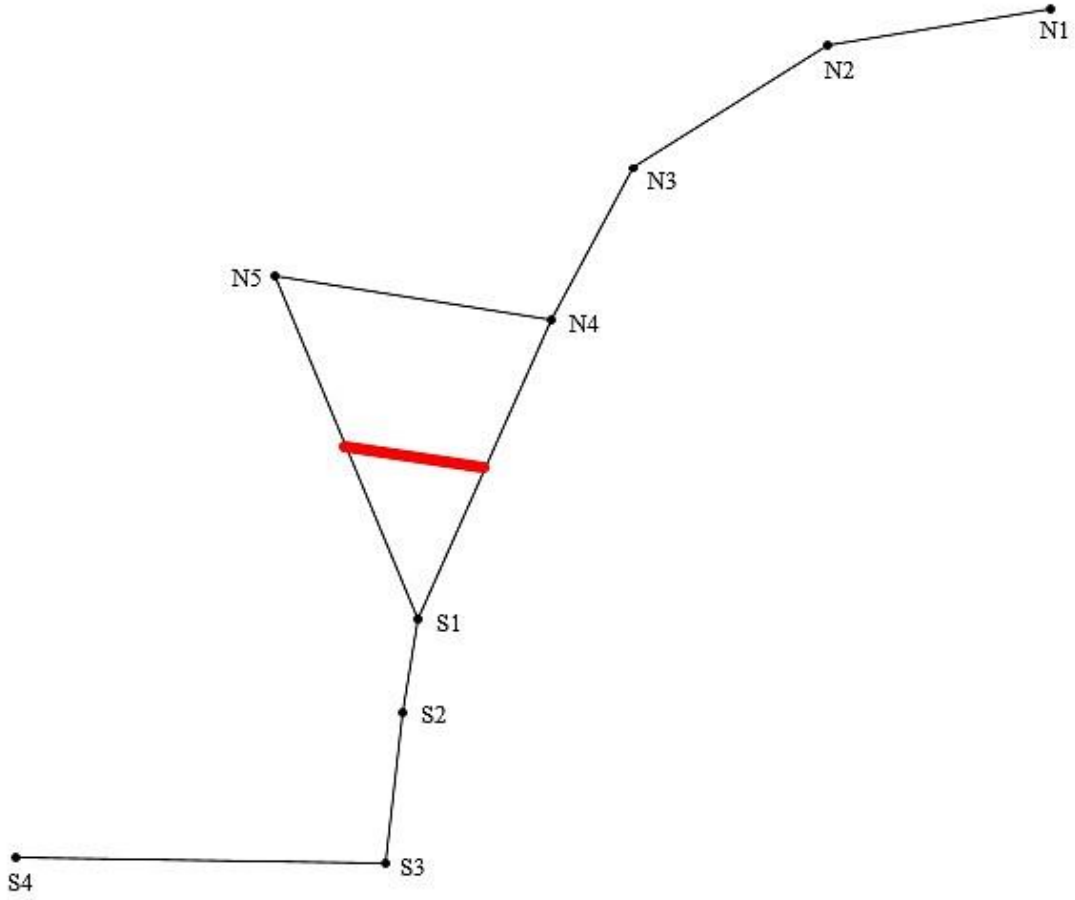
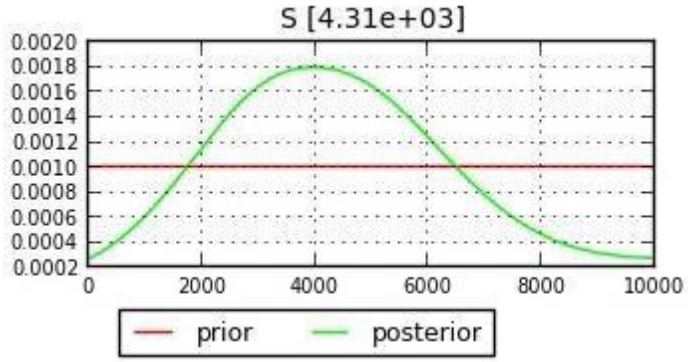


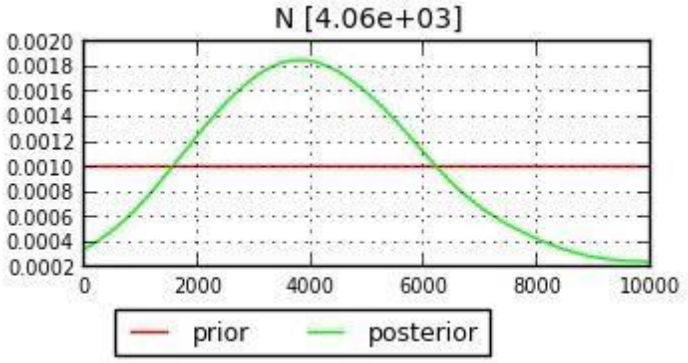
Figure 4.9. A connection network generated by a Monmonier's function among *Afzelia quanzensis* collection sites, and the position of a genetic barrier (shown in red) between northern and southern collection sites

G

A.



B



C

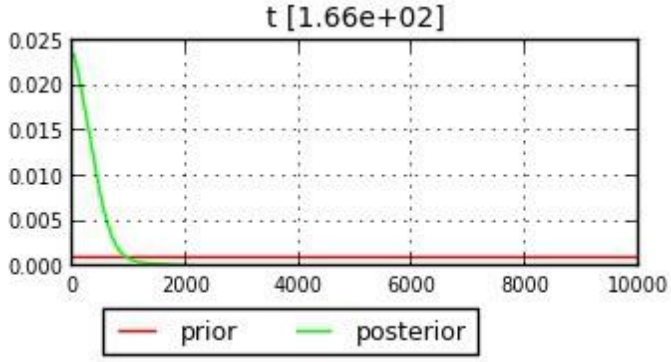


Figure 4.10. Prior and posterior probability plots derived from approximate Bayesian computation of the estimation of effective population size and time since divergence of two *Afzelia quanzensis* gene pools, A-effective population size of the southern gene pool, B-effective population size of the northern gene pool, and C-number of generations since divergence

H

Applications in Plant Sciences Copyright and Licensing Information

VITA

NAME:	Percy Jinga
EDUCATION:	B.S., Biology, Bindura University of Science Education, Bindura, Zimbabwe
	M.S., Conservation Biology, Bindura University of Science Education
TEACHING:	Department of Biological Sciences, University of Illinois at Chicago, Genetics Laboratory for Undergraduates, 2015-2017
HONORS:	2013 Fulbright Junior Staff Development Scholarship, Fulbright
	2014 Dr. Benjamin L. Van Duuren Travel Grant, International Institute of Education
	2015 Chicago Consular Corps Scholarship, University of Illinois at Chicago
	2015 Elmer Hadley Award, Biological Sciences Department, University of Illinois at Chicago
PROFESSIONAL: MEMBERSHIP	Botanical Society of America
PUBLICATIONS:	Jinga, P., Palagi, J. and Ashley V. M. 2016. Development of microsatellite loci of pod mahogany, *Afzelia quanzensis* (Fabaceae), by Illumina shotgun sequencing, and cross-amplification in *A. africana*. *Applications in Plant Sciences*, **4**: doi:10.3732/apps.1600010.
	Jinga, P., Munosiyei, P., Bobo, D. E. and Tambura, L. 2012. Seasonality of parasitism in free range chickens from a selected ward of a rural district in Zimbabwe. *African Journal of Agricultural Research*, **7**: 3626-3631.
POSTER: PRESENTATIONS	Development of microsatellite loci in *Afzelia quanzensis*. Botany 2016, Savanna, Georgia, USA
	A mountain range is a strong genetic barrier between populations of *Afzelia quanzensis*, Botany 2017, Fort Worth, Texas, USA