ALAIN RENE ATANGANA

PHENOTYPIC DIVERSITY IN FRUIT AND SEED TRAITS, AND NEUTRAL GENETIC DIVERSITY IN *ALLANBLACKIA FLORIBUNDA*

Thèse présentée
à la Faculté des études supérieures de l'Université Laval
dans le cadre du programme de doctorat en sciences forestières
pour l'obtention du grade de Philosophiae Doctor (Ph. D.)

DÉPARTEMENT DES SCIENCES DU BOIS ET DE LA FORÊT
FACULTÉ DE FORESTERIE, DE GEOGRAPHIE ET DE GEOMATIQUE
UNIVERSITÉ LAVAL
QUÉBEC

2010

Résumé

Allanblackia floribunda est un arbre des forêts denses humides tropicales valorisé pour la teneur élevée en acides gras de ses graines, essentiellement constitués d'acides stéarique et oléique dont l'efficacité dans la réduction du mauvais cholestérol de l'organisme humain a été prouvée. Pour cette raison, les graines de *A. floribunda* collectées en milieu naturel sont commercialisées. Toutefois, les travaux sur la culture de cette espèce sont encore à leur phase initiale. Nous avons déterminé la possibilité d'amélioration génétique de cette espèce en échantillonnant 17 à 40 fruits par arbre de 70 arbres distribués sur quatre sites en milieu naturel. La matière grasse a été extraite des graines, et les teneurs en acides stéarique et oléique estimées à l'aide de méthodes développées au cours de cette étude. La variation phénotypique des traits des fruits et des graines a été caractérisée dans et entre les arbres, et entre les sites. Les estimations de répétabilité des caractères mesurés ont été effectuées. Des corrélations phénotypiques entre les traits étudiés ont aussi été estimées, et quatre traits ont été retenus pour effectuer la sélection multi-caractère de 20 arbres-plus qui constitueront la population d'amélioration de cette espèce pour la production des graines. Nous avons par la suite isolé 10 marqueurs moléculaires de type microsatellite polymorphes à partir de *A. floribunda*, et sept de ces marqueurs étaient polymorphes à la fois chez *Allanblackia gabonensis* et *Allanblackia stanerana.* La variation de huit loci microsatellites a permis de caractériser la structure génétique neutre de 10 populations de *A. floribunda* de zone de forêt naturelle du Cameroun, puis d'inférer l'histoire récente des forêts humides d'Afrique Centrale. Aucune différence significative n'a été observée entre les paramètres génétiques de la population d'amélioration et celle existant en milieu naturel indiquant qu'une amélioration de cette espèce à partir des 20 arbres sélectionnés ne réduirait pas sa diversité génétique neutre. Toutefois, une légère augmentation du taux de consanguinité a été observée dans la population d'amélioration, et des recommandations sont formulées pour la conservation des ressources génétiques durant l'amélioration de *A. floribunda*.

Abstract

Allanblackia floribunda or tallow tree is a tropical forest-tree species that is valued for its seeds, which are rich in hard fat consisting mostly of stearic and oleic acids, reported to lower plasma cholesterol levels, thus reducing the risks of heart attack. Owing to this fat profile, *Allanblackia* oil is used for margarine production and in soap and ointments manufacture, and seeds extracted from *Allanblackia* fruits by local communities are traded. We determined whether the species could be genetically improved for fruit/seed production by sampling 17 to 40 fruits from each of 70 trees that were distributed among four sites in wild stands. Fat was extracted from the seeds, and stearic and oleic acid content of the fat was estimated using methods developed in this study. Phenotypic variation in fruit/seed traits was assessed within- and among-trees, and among sites. Repeatabilities were estimated for measured characters, and relationships between these characters investigated. Twenty "plus trees" were selected for breeding, and implications for improvement discussed. Then we isolated and characterized ten microsatellite primer pairs for *A. floribunda*. Seven of these microsatellite loci were polymorph for both *Allanblackia gabonensis* and *Allanblackia stanerana* species as well. Using eight informative microsatellite loci, we have characterized the genetic structure of *A. floribunda* natural populations from Cameroon, and inferred the recent history of rainforests from Central Africa. No significant difference was identified in genetic parameters between wild stands and the breeding population, indicating that breeding *A. floribunda* from 20 trees would not reduce nuclear genetic diversity. However, a slight increase in inbreeding was observed in the breeding population, and recommendations for genetic diversity conservation during tree improvement in the species are made.

Foreword

First, I would like to express my sincere gratitude to my supervisor, Prof. Damase Khasa, for undertaking the task of supervising this research project, his invaluable guidance, and for full support during this research. He also acted as an excellent advisor and mentor. Next, I would like to thank my co-supervisor, Dr Jean Beaulieu, for taking me on as a student, and his invaluable contribution to this project.

I am indebted to Dr Tony Simons (World Agroforestry Centre, ICRAF) for his support at the beginning of this project. Also acknowledged is Prof. Roger Leakey for sharing his broad experience on tree domestication, and the many hours spent reading drafts. I feel owe a sincere thank to my advisory committee members for their valuable examinations and constructive suggestions which have improved this work. I would like to thank Drs. Zac Tchoundjeu, Ann DeGrande, and M. Ebenezar Asaah (ICRAF Cameroon) for their support during leaf material collection in Cameroon. Also acknowledged are Drs Harrie Hendrickx, Eric van der Vlis and Dennis van Houten (Unilever Netherlands), and all ICRAF staff in Cameroon.

I would like to acknowledge André Gagné (CEF, Université Laval) for volunteering his time and sharing his experience on molecular lab experiments. Also acknowledged are Dr Craig Newton (ATG Genetics, Vancouver), Sauphie Senneville (Canada Chair in Forest and Environmental Genomics), Vincent Bourret (Université Laval) and Dr Nathalie Carisey (Faculty of Forestry and Geomatics, Université Laval). Especially, I thank my colleagues (Sébastien, Benjamin, Frank Bedon, Julie, Pat, Julien, Karine, Damien) in the Molecular Biology laboratory of the Institut de Biologie Intégrative et des Systèmes (IBIS) from Université Laval for their friendship and kind support during my studies. Also, I owe innumerable thanks to family members (Ludwine, Lydie, Flore, Ginette, Séraphin, Bertrand, Annick, Thierry and Martiale) and friends (Daniel Tonye, Marie Ngono, Dr Marie-Laure Ngo Mpeck-Nyemeck) for their support.

The Natural Sciences and Engineering Research Council of Canada (NSERC) through the Fonds Général pour les Etudes Supérieures (FGES), from Université Laval, provided grant for doctoral studies.

To Dad, and in loving memory of Mom

Table of contents

List of tables

List of figures

Chapter 1: Introduction

1.1 Background

Congo basin forests are ranked second as global biodiversity hotspot among major tropical wilderness areas (Mitterneier *et al*. 1998), and they harbor important carbon stock (Justice *et al*. 2001). Consequently, they are amongst top conservation priority areas (Olson & Dinerstein 1998). Sixty-five million people live in or near these forests of Central Africa (Aubé 1996). These people mainly rely on agriculture and the forest for food and income, and are amongst the poorest in the region. In Cameroon for example, 55% of people in the forest zone live below the poverty line, farmers being the poorest occupational group with 57% below the poverty line (GoC 2003). Pressures on natural forests have increased with the decline in cocoa and coffee prices since the 1980's. As a result, the annual rate of deforestation in the humid tropical forest in Africa between 1990 and 1997 has reached 0.43%, with an annual deforested area of 0.85 ± 0.30 million hectares (ha), whereas the annual regrowth rate was estimated at 0.07% in the same period (Achard *et al*. 2002). The Food and Agricultural Organization (FAO) stated that Africa suffered a loss of forest of 7,317,000 ha from year 2000 to year 2005 (FAO 2006 p. 20-21). Most of this tropical forest has been lost as a result of changes in land use from forest to agriculture (WCFSD 1999). One of the urgent actions to address this problem, according to WCFSD (World Commission on Forests and Sustainable Development), is to provide more extensive support to community-based agroforestry in order to reduce the exploitation of primary forests for subsistence products. This poverty-reduction and forest-protection strategy could be achieved through the development and cultivation of marketable and under-utilized "new crops" from these forests (Leakey *et al*. 2005). Consequently, new initiatives in tropical forest tree improvement aiming at developing cultivars of trees with desired fruit, nut and medicinal characteristics (Leakey and Newton 1994; Franzel *et al*. 1996, Simons and Leakey 2004) are underway. The 'Cinderella species' overlooked by science (Leakey and Newton 1994) now forming this worldwide program of indigenous tree domestication

include *Allanblackia* species, with a special focus on Cameroon, Nigeria, Equatorial Guinea, Gabon, Democratic Republic of Congo, Ghana, Guinea Conakry, Liberia and Sierra Leone (Leakey and Simons 1998; Tchoundjeu *et al*. 2006). These multi-purpose species have traditionally provided communities with their everyday needs for food and medicinal products (Leakey 1999), and often have high market value regionally and internationally (Ndoye *et al*. 1998; Tabuna 1999), as well as important social and livelihood benefits in local communities (Schreckenberg *et al*. 2002).

Allanblackia (Guttiferae Juss. 1789 vs. Clusiaceae Lindl. 1836) are medium-sized tree species of humid forest zone of Africa producing berry-like fruits that are suspended on long pedicels, and seeds of the genus are recalcitrant (germination success less than 5 %, Vivien and Faure 1996). Nine species in the genus have been recorded (Bamps 1969), namely *Allanblackia floribunda* Oliver or tallow-tree, *Allanblackia gabonensis* (Pellegr.) P. Bamps, *Allanblackia kimbiliensis* Spirlet, *Allanblackia kisonghi* Vermoesen, *Allanblackia marienii* Staner, *Allanblackia parviflora* A. Chev., *Allanblackia stanerana* Exell & Mendonça, *Allanblackia stuhlmanii* (Engl.) Engl. and *Allanblackia ulugurensis* Engl. *A. parviflora* is distributed in Upper Guinea domain of the Guineo-Congolian region of Africa (Denys 1980), from Sierra Leone and Guinea to Ivory Coast and Ghana (Bamps 1969), and is very similar to *A. floribunda*, which is distributed in the lower Guinea Domain and the Congo Basin (Bamps 1969; Denys 1980) from Benin to Democratic Republic of Congo (DRC) and North Angola. *A. kisonghi* and *A. marienii* are found in DRC, mainly in Bas-Congo and District Forestier Central, the former being also found in Kasai, whereas *A. kimbiliensis* is found in DRC and Bwindi forests in Uganda (Bamps 1969; Cunningham 1992). *A. ulugurensis* and *A. stuhlmannii* occur in the Eastern Arc mountains of Tanzania. *A. gabonensis* is distributed from Cameroon to Democratic Republic of Congo, from 500 to 1750 m elevation above sea level, while *A. stanerana* is found from Cameroon to Angola and Democratic Republic of Congo in evergreen littoral forests (Vivien and Faure 1996). Mature *A. floribunda* trees reach 30 m in height, 80 cm girth with bole straight and horizontal branches (Vivien and Faure 1996). Male and female flowers occur on the same tree in *A. floribunda* but usually on separate branches (Keay 1989).

1.2. Tree domestication in the humid tropics of Africa

Crop domestication has been practiced for millennia (Simmonds 1987; Diamond 2002), and commonly refers to selection, breeding and adaptation in agricultural systems. Tree domestication, which is more recent (Dafni 1992), can be simply defined as the selection and management of trees to increase their benefits to humankinds. These benefits include forest timber products (Zobel and van Buijtenen 1989), non-timber forest products also known as Agroforestry Tree Products (sap e.g., *Acer saccharum*, Larochelle *et al*. 1998; resins e.g., *Pinus jeffreyi*, Savage *et al*. 1996; gums e.g., *Acacia longifolia*, Baldwin *et al*. 1998; fruits e.g., *Irvingia gabonensis*, Leakey *et al*. 2000, nuts e.g., *Ricinodendron heudelotii*, Ngo Mpeck *et al*. 2003), and ecological services (soil improvement, e.g., *Sesbania sesban*, Hartemink *et al*. 1996; shade Sparks *et al*. 1996; erosion control Wilkinson 1999; carbon stock Burrows *et al*. 2002).

In the humid tropics of Africa, tree domestication for non-timber forest products using agroforestry techniques has been defined as an iterative procedure involving the identification, production, management and adoption of high quality germplasm (Simons 2003). Indeed, agroforestry tree domestication is executed in the humid tropics of Africa as a farmer-driven and market-led process, which matches the intra-specific diversity of locally important trees to the needs of subsistence farmers, product markets and agricultural environments (Simons and Leakey 2004). Participatory tree domestication as implemented in the region since 1998 by the World Agroforestry Centre (ICRAF) and its partners uses the variability within species, by selecting trees with desirable phenotypic traits (Leakey *et al*. 2000; Atangana *et al*. 2002), and propagating them asexually (Tchoundjeu *et al*. 2006) using inexpensive, robust and simple methods (Leakey *et al*. 1990). Germplasms of selected trees are introduced in existing land-use systems such as cocoa and coffee farms. Farmers are trained on how to select best trees and reproduce them using robust and inexpensive vegetative propagation methods such as rooting of leafy stem cuttings and

marcotting. High-scale monoculture of priority species are not promoted by ICRAF and his local partners, as this would reduce forest cover. Cultivation of *Allanblackia* species in existing farming systems is an efficient alternative of extraction for the sustainable supply of *Allanblackia* seeds, and for a better management of genetic resources in this genus. Therefore, ICRAF and its local partners are promoting diversity in mixed farming systems, so as to diversify and increase farmer's diet and their revenues. For this reason, *Allanblackia* domestication strategy is different from that of *Elaeis guineensis*. According to FAO (1996), human being depend on few crops (maize, wheat and rice provide over 50 per cent of the daily global requirement in calories and proteins, and 150 crops are commercialized on a significant global scale). Tree domestication would promote underutilized species. Although tree domestication includes tree breeding, very little has been done to assess the level of genetic control in traits of interest in the species undergoing tree domestication in the humid tropics of Africa, including *I. gabonensis*, *Dacryodes edulis*, *R. heudelotii*, *Chrysophyllum albidum*, *Garcinia kola*, *Cola acuminata, Cola anomala* and *Cola nitida*. Studies are in progress to develop and fine-tune rooting of leafy stem cuttings and marcotting in these species, and quantitative descriptors of variation *I. gabonensis* have been described by Leakey *et al*. (2000). These descriptors have been used to assess phenotypic variation in fruits and seeds of *I. gabonensis* and *Dacryodes edulis* in Cameroon (Atangana *et al*. 2001; Wahuriu 1999), and plus trees selected for domestication (Atangana *et al*. 2002). However, genetic resources preservation in the species under domestication in West and Central Africa has been restricted to the establishment of genebanks in Cameroon and Nigeria for *I. gabonensis*.

1.3 Forest tree breeding

Forest tree breeding aims at solving some specific problem or producing a specially desired product (Zobel and Talbert 1991 p. 6), by using genetic information to increase performance in certain traits. Indeed, tree breeding refers to plant material submitted to

evolution forces (Simmonds 1987), and has two objectives: 1) breeding, and 2) conservation of genetic resources.

1.3.1. Breeding

Traits are influenced by one gene (Mendelian traits) or many genes (polygenic traits). Polygenic (or quantitative) traits are of utmost importance in tree breeding and are of three types (Hartl 1988 p. 216-217):

- Continuous traits for which there is a continuum of phenotypes (e.g., growth rate, weight);
- Meristic traits for which the phenotype is expressed in discrete or integral classes (e.g., number of flowers on a petal); when the number of possible phenotypes is large, there is no distinction between a meristic and a continuous trait;
- Threshold traits (e.g., diabetes).

The variation of continuous traits is used in tree breeding, which includes four main steps: 1) economic and ecological analysis of needs, 2) determination of the amount, kind and causes of genetic variability of characters of interest within the species, 3) selection, and 4) production of improved propagules.

The first step consists in defining the priorities and objectives of breeding, assessing reforestation needs, ecologic optima of the species of interest, and economic costs and time with respect to tree breeding in the species as well. The second step consists in partitioning the phenotypic variability of traits of interest into components attributable to different causes, for the determination of the degree of resemblance between relatives (Falconer 1989 p.139; Falconer and McKay 1996 p. 145-146). In wild stands, phenotypic (or total) variance for each metric character is partitioned into geographic (provenance), sites within provenance, stands within sites, individual trees within stands, and within-tree variances, respectively (Zobel and Talbert 1991 p. 58). Within-tree variance allows estimation of repeatability (Falconer and McKay 1996 p. 136-137), which expresses the proportion of the variance of single measurement that is due to permanent differences between individuals,

both genetic and environmental. No conclusion about the genetic control of a character can be drawn from studies in wild stands, as trees are of unknown pedigrees. Common gardens are multi-site experiments in which several genetic materials (provenances, families and individuals) are assessed for their performance in traits of interest, so as to quantify the effects of genotype, environment and genotype x environment interaction. Randomization and replication of treatments in each site is essential for such experiments, so as to estimate heritability values per site and overall. Common gardens allow the partition of phenotypic variance into genetic and environment components using the formula:

$$V_P = V_G + V_E + V_{GxE}$$

where V_P, V_G, V_E and V_{GXE} are, respectively, phenotypic, genetic, environmental, and genetic x environmental variances. Genetic variance is the sum of additive (V_A or breeding value: part of the genetic material is transmitted from parent to offspring through sexual reproduction) and non-additive (V_{NA}) variances, the latter being divided into dominance (due to interaction of specific alleles at gene locus) and epistatic (due to interactions among gene loci; Zobel and Talbert 1991 p.51). Therefore, the degree of resemblance between relatives, which expresses the extent to which phenotypes are determined by the genes transmitted from the parents, can be obtained as the ratio V_A/V_P (Falconer and Mackay 1996 page p. 123) also called narrow-sense heritability (h^2). The degree of genetic determination, also called broad-sense heritability (H^2), is the ratio V_G/V_P (the whole genetic material is transmitted from parent to offspring: asexual reproduction) also obtained from common gardens.

Several ways of estimating variance components exist; however, maximum likelihood (ML), restricted maximum likelihood and minimum variance quadratic unbiased estimation (MIVQUE) methods provide estimators with better properties (Milliken and Johnson 2009 p. 309) when the design is unbalanced. The MIVQUE method does not require iteration, whereas the ML method maximizes the probability of occurrence of the sample results; however, the variance component estimates using ML are biased, and the REML is preferred, owing to its versatility. Indeed, REML does not require balanced designs, allows

unlimited hypotheses to be tested, and missing data present no special problem. The REML approach is powerful in estimating variance components in random effects model (Littell *et al*. 2006 p. 7) using the mixed model procedure of the Statistical Analysis System (SAS Institute 2006). Mixed models are statistical models (description of how the data can be produced; Littel *et al*. 2006 page 2) including random and fixed effects. Mixed models are appropriate for block clustered designs, repeated measures designs, nested (hierarchical) designs and split-plot designs.

Artificial selection is a technique aiming at choosing on the basis of their performance in certain characters in base populations, individuals to be used for breeding (breeding population); the best individuals in the breeding population are sometimes used to constitute a production population as well. Genetic gain is expected from selection, as the performance of progenies from the breeding population is expected to be higher than that observed in base populations. The expected response to selection is the product of selection intensity, heritability and phenotypic standard deviation (Zobel and Talbert 1991 p. 133, Nanson 2001 p. 155). The best breeding method to be used for achieving gains is determined by heritability. High narrow-sense heritability indicates that phenotypic variation provides a good estimate of genetic variation; hence individual selection is appropriate, whereas low heritability indicates that selection should be done among families. Single-trait selection methods include individual, provenance, family, individual intra-family and combined family and intra-family, progeny-test, clonal, and multi-stage selections. Multiple-trait selection systems include tandem, independent culling and index selections (Nanson p. 186-188). Information from relatives can be used to perform combined selection (Nanson 2001 p. 179) so as to increase the efficiency of selection. Other methods of selection exist, including indirect selection (in juveniles' individuals), marker-assisted selection, and recurrent selection allowing cumulative gains through many breeding cycles. Once selection is done, specific crosses can be done as to pack desired traits into improved individuals for germplasm production in seed or clonal orchards.

1.3.2. Conservation of genetic resources

The main evolutionary factors (mutation, selection, gene flow and genetic drift) shaping the genetic structures of populations are theoretically in equilibrium for each species in wild stands; any significant external influence on one of these factors will disrupt the existing equilibrium, and affect the evolutionary trajectory, hence diminishing the species' adaptive potential of these populations. Selecting and breeding a species for a character will certainly increase in the population the frequency of alleles controlling the character of interest, provided this character is under genetic control. Applying repeatedly artificial selection (recurrent selection) in each breeding cycle would disrupt the equilibrium between forces driving evolution, and bottlenecks can occur. Therefore, any tree breeding program should develop strategies aimed at preserving whole or majority of source's genetic diversity in the species both at functional and neutral levels during tree improvement.

Preserving genetic resources in a species should be based on a detailed inventory of the genetic diversity of that species. This inventory is done using common gardens for species undergoing breeding programs. However, these tests are time-consuming, although they are useful in dissecting phenotypic variation. Molecular markers allow rapid assessment of the total genetic diversity in the species, and guide the sampling of populations representing this genetic diversity for preservation. These markers can be functional (when genes controlling the characters of interest for breeding are known) or neutral. Molecular markers can also be used to infer signatures of past events, therefore allowing better strategies for genetic resources preservation. Population genetic tools are efficient to address this issue.

Population genetics is a science that deals with Mendel's laws and other genetic principles as they apply to entire populations of organisms, including the study of the various forces that result in evolutionary changes in species through time (Hartl and Clark 2007 p. 3-4). Population genetics is essential in conservation, plant breeding, evolutionary biology, ecology, genetics, genomics, natural history, and one of its goals is to understand the evolutionary and biological significance of genetic variation, then recommendations can be made about how to best manage genetic resources. A key element in population genetics

is the understanding of genetic diversity, which is the variation in allelic and genotypic composition in a given species. Markers used to investigate genetic diversity in forest trees to date have been of biochemical and molecular genetic types (Strauss *et al*. 1992).

1.3.2.1. Biochemical markers

Strauss *et al.* (1992) define biochemical markers as 'those derived from study of the chemical products of gene expression, such as proteins sequences or net charges, and composition of secondary chemicals such as terpenoids'. Terpenoid compounds of conifers include monoterpenes and their derivations, sesqui- and diterpenes, polymeric carotenoid and chlorophyll pigments, growth regulators, gibberellic and abscisic acids. Suitability of terpenes as genetic markers in forest genetics has been reviewed by Hanover (1992). The number of potential terpene markers in trees is large owing to thousands of derivatives of isoprene molecule. However, the number of statistically independent dimensions they provide is less than the number of chemical species observed, owing to the strong correlations among terpene concentrations. However, terpenoids have proved to be valuable tools to assess geographic variation in forest trees in the years 60 and 70 (Forde and Blight 1964; Tobolski and Hanover 1971; Wilkinson *et al*. 1971; Coyne and Keith 1972; Wilkinson and Hanover 1972; Hanover 1974; Zavarin *et al*. 1975), before the widespread of numerical methods for production of phylogenetic trees. Most of terpenoids studies have then been restricted to present maps or matrice distances. However, Hanover (1992) stated that terpenoid substances in forest trees are versatile biochemical systems for use as genetic markers.

Biochemical markers also include allozymes. Enzymatic electrophoresis, in which allozymes in solution move in response to an electric field, is the oldest molecular genetic characterization method. Application of this simple and inexpensive method is universal, and the number of analyses per unit of time is high. However, the polymorphism is limited to 2-5 alleles per locus, the number of loci is also limited, a large proportion of the variation is not detected, the nature of genetic difference between alleles is unknown, and possibility of non-neutral polymorphism exists. This method has been widely used in population

structure in tree species (Cheliak *et al*. 1985; Moran *et al*. 1989; Perry and Knowles 1989; Coates and Sokolowski 1989; Rajora 1989; Hamrick and Murawski 1991; Ledig *et al*. 2006; Mylecraine *et al*. 2004; Parks *et al*. 1994), hybridization (Maki and Murata 2001) and selection studies (Riginos *et al*. 2002).

1.3.2.2. Molecular genetic DNA markers

Molecular genetic markers are 'those derived from direct analysis of genetic polymorphism in DNA sequences' (Strauss *et al*. 1992), and include non Polymerase Chain Reaction (PCR)-based markers (e.g., RFLPs) and PCR-based markers.

Restriction Fragment Length Polymorphisms (RFLPs; Beckmann and Soller, 1983) is a technique based on the difference in length of a restriction fragment found segregating in natural populations. The technique involves restricting DNA with endonucleases, separating the resulting fragments by gel electrophoresis, and visualizing the size-sorted fragments (Avise 2004). Indeed, RFLPs result from the presence/absence of particular restriction sites in DNA, and they are abundant in the genomes of most organisms. They have been used in genetic studies of natural populations, as they are highly discriminating, and only 5-6 loci need to be tested. However, this inexpensive technique requiring high quality and large amount of DNA, is slow and cannot be automated, and statistical analysis is difficult. Also, larger restriction fragments in plants are more likely to be polymorphic than smaller fragments owing to the high frequency of insertions/deletions observed at these loci (McCouch *et al*. 1988; Miller and Tanskley 1990). However, RFLPs analyses have been used to infer the genetic structure in plants (Raybould *et al*. 1996) including *Thuja plicata* (Glaubitz *et al*. 2000). Restriction analyses techniques are also used for plant organelle DNA (e.g., plant mitochondrial DNA and chloroplast DNA).

Invention of PCR technology by Karry Mullis in 1984 facilitated the development and use of various molecular genetic markers, including random amplified polymorphic DNA (RAPDs), Amplified-Fragment Length Polymorphism (AFLPs), Single-Strand Conformation Polymorphisms (SSCPs), minisatellites, simple sequence repeats (SSRs) or

microsatellites, Expressed Sequence Tag Polymorphism (ESTP), Expressed Sequence Tags-Simple Sequence Repeats (EST-SSR) and single nucleotide polymorphisms (SNPs). These markers can be of nuclear (Avise 2004) or organel origin (e.g., mtDNA, Palumbi and Baker 1994). Polymerase Chain Reaction consists in the amplification of specific DNA fragments using oligonucleotide primers (small number of nucleotides joined together in a short stretch of single-stranded DNA). This technique has three steps, namely denaturation, annealing and extension/elongation, and allows targeting regions of interest in the genome, together with the possibility of using small amounts of tissue and old DNA.

RAPDs (Williams *et al*. 1990; Welsh and McClelland 1990) is a technique involving the use of short PCR primers of arbitrary sequence to amplify anonymous genomic sequences. RAPDs have been widely used in population biology studies during the 1990s (Lowe *et al*. 2000; Dawson *et al*. 1995; Rieseberg 1996); however, RAPD data are hampered by a lack of complete genotypic information resulting from dominance (Lynch and Milligan 1994).

Amplified-Fragment Length Polymorphism (Vos *et al.* 1995) is based on the selective PCR amplification of restriction fragments from a total digest of genomic DNA. This inexpensive method allows co-amplification of high numbers of restriction fragments but is constrained by (i) allelic dominance, (ii) allelic diversity reduced to 2 alleles per locus, (iii) the logistic complexity of its application, and (iv) the fact that it requires good quality DNA. AFLPs have been used in population biology studies in Oak species (Dodd and Kashani 2003; Dodd and Afzal-Rafii 2004; Coart *et al*. 2002), in *Pinus oocarpa* (Diaz *et al*. 2001), in *Cedrela odorata* (de la Torre *et al*. 2008) and in *Metrosideros bartlettii*, a rare tree endemic to New Zealand (Drummond *et al*. 2001).

Differences in fragment sizes owing to differences in the number of tandem repeats that are short (2-5 bp in the case of microsatellites, Tautz 1989; Weber and May 1989; Hearne *et al.* 1992) or longer (approximately 20 bp for minisatellites, Armour and Jeffreys 1992) can also be assayed as genetic markers. These markers also called Variable Number of Tandem Repeats (VNTRs) are widely distributed in the nuclear genome, and found in many

loci. Advantages of minisatellites are (i) their high variability, (ii) the fact that many loci can be analyzed in short time, (iii) the use of universal probes and (iv) no elaboration of methodological development is needed. However, minisatellites are not suited for population studies owing to complex banding patterns, occurrence of genotyping errors, and the fact that information cannot be separated between individual loci. Owing to their high variability, co-dominance, repeatability, ease and reliability of scoring, microsatellites, which are short tandem repeats, are ideal molecular PCR-based markers for population genetic studies (Powell *et al*. 1996). Indeed, microsatellites require less DNA, they are widely distributed in the genome and can be assessed using old or degraded DNA; they can be standardized and automation is possible, together with multiplexing (analysis of many loci). Microsatellites are usually neutral, and polymorphism is variable. However, microsatellites should be developed for a taxon previously unstudied, and the technique is costly. Microsatellites have been widely used in population biology studies in tree species (Dayanandan *et al*. 1999; Collevatti *et al*. 2001; Novick *et al*. 2003; Lemes *et al*. 2003; Khasa *et al*. 2006; Born *et al*. 2008).

Expressed sequence tags are used for gene discovery, as they are short sub-sequence of a transcribed complementary DNA (cDNA) sequence. Expressed sequence tags polymorphisms (ESTPs; Harry *et al*. 1998) are co-dominant markers developed from sequenced cDNA clones using PCR-RFLP techniques. However, the use of gel-based techniques such as denaturing gradient gel electrophoresis (DGGE; Fischer and Lerman 1983) also reveals ESTPs, and the power of DGGE over that of the PCR-RFLP method has been demonstrated (Temesgen *et al*. 2000; 2001). ESTPs are used for comparative mapping, to map genes of known function, and to identify candidate genes (Temesgen *et al*. 2001).

EST databases can be used to develop EST-SSRs, which detect variation in the expressed portion of the genome. EST-SSRs are inexpensive compared to genomic SSRs, and can be transferable to related species. EST-SSRs are used for population genetic analyses and conservation (Ellis and Burke 2007) as well as study of genetic diversity (Gupta *et al*. 2003).

Single Nucleotide Polymorphisms are genetic markers consisting in the variation of a particular nucleotide in a DNA sequence within species or between related species. Owing to their large number and evenly-distribution in the genome, low assay cost in large experiments and ease of automation, SNPs can be quickly assayed in a large number of DNA samples for a small cost per assay. However, the number of alleles at a given locus is small as a consequence of the low rate of base substitution. Indeed, although in theory each SNP marker can display up to four alleles, namely adenine, cytosine, guanine and thymine, in practice only two alleles are present most of the time at any given SNP. SNPs are used to understand the relationship between naturally occurring genotypic and phenotypic diversity in forest trees (Neale and Ingvarsson 2008).

Single-strand conformation polymorphisms (Orita *et al*. 1989) assays DNA variation on the basis of variation in conformation. This approach is based on the observation that changes in sequence alter the folding of single-stranded DNA which, in turn, can affect electrophoretic mobility, resulting in SSCP. This effective and inexpensive method has some limitations, as the maximum fragment size for analysis is 400 kb, and some polymorphisms could not be identified.

To date, known genomes (nuclear DNA, mitochondrial DNA or mtDNA and plant chloroplast DNA or cpDNA) are used in population genetic studies. As nuclear DNA is transmitted biparentally, it provides information on the evolutionary history of two loci. Mitochondrial DNA in animals is widely used in population genetic studies, owing to its small and constant size (15.7-19.5 kb) and ease of manipulation, ease of survey for polymorphism by restriction, digestion, and rapid rate of evolution (mutation rate higher than that of nuclear DNA). Mitochondrial DNA is present in multiple copies per cell conferring robustness to degradation; it is usually maternally inherited (through the egg cytoplasm) and exhibits a lack of apparent recombination (recombination has however been reported in hominid mtDNA, Awadalla *et al*. 1999). This apparent lack of recombination indicates that conclusions drawn from single-locus studies rely on only one realization of gene genealogy. Also, the presence of functional linked genes in the mitochondrion

(Anderson *et al*. 1982) makes it difficult to distinguish between population expansions and natural selection. Another constraint in using mtDNA is the fact that its size in plants is bigger and variable (200-2500 kb) and the rate of nucleotides' substitution slower than in animals. Also, frequent rearrangements in genes' position occur within the mitochondrion, and transmission mode (maternal or paternal) varies per species in plants. Plant mtDNA evolves more slowly than animal mtDNA with respect to nucleotide sequence (Palmer and Herbon 1988), limiting the utility of this genome in population biology studies. However, mitochondrial DNA has been used in population biology studies in tree species, namely *Pinus banksiana* (Godbout *et al*. 2005), *Pinus sylvestris* L. (Naydenov *et al*. 2007), *Pinus albicaulis* (Richardson *et al*. 2002) and in *Picea mariana* (Jaramillo-Correa *et al*. 2004).

Plant chloroplast DNA is transmitted maternally (Havey *et al*. 1998), paternally (Yang *et al*. 2000) and biparentally (Shore and Triassi 1998). Its size varies greatly among species, and the rate of cpDNA evolution is slow in terms of primary nucleotide sequence and gene rearrangements (Avise 2004), with possible exceptions (Milligan *et al*. 1989). Owing to this slow rate of evolution cpDNA data have proved valuable for estimating plant phylogeny (Zurawski and Clegg 1987) and population structures in tree species (Jones *et al*. 2006; Jaramillo-Correa *et al*. 2006, Muloko-Ntutume *et al*. 2000; Leong Pock Tsy *et al*. 2009).

1.3.2.3. Markers and conservation genetics

Genetic markers are important in conservation genetics that has to do with the preservation of variability in rare or threatened populations, or within species under breeding programs. Indeed, it is assumed that higher heterozygosity and low inbreeding within populations enhances a population's survival probability over time (Avise 2004). Observed heterozygosity and expected heterozygosity (estimated proportion of loci expected to be heterozygous under Hardy-Weinberg equilibrium) are the most common diversity indices, together with number of alleles. Inbreeding level in rare or threatened populations is also of utmost importance for genetic management, as is the genetic structure of the species of interest. While the observed heterozygosity corresponds to the frequency of heterozygous loci in a population, the expected heterozygosity is twice the product of

allele frequencies of heterozygous loci. Population structure is assessed with *F*-statistics (Weir 1996). Populations are said to be genetically different if their estimated F_{ST} is significantly positive. Within a population, the inbreeding level also known as F_{IS}, which represents the amount of discrepancy from Hardy-Weinberg equilibrium (HWE), measures the reduction in heterozygosity of an individual due to non-random mating or mating within its population (or sub-population). The HWE consists in the stability of allelic and genotypic frequencies among generations in populations of large size on which there is no or negligible action of main evolutionary forces.

The goals of conservation genetics are preservation of genetic diversity and promotion of ecology and evolutionary processes fostering and sustaining biodiversity (Crandall *et al*. 2000). To do so, an assessment of genetic diversity, parentage, kinship, is important. This assessment is done using markers. Neutral molecular genetic markers provide information of evolutionary forces on the impact on a given species. Therefore, neutral molecular genetic markers are efficient tools for conservation applications. However, these markers tell nothing about the adaptative potential of a species, although neutral genetic variants in one situation may be adaptative in another. For example, microsatellites commonly known as neutral molecular genetic markers have been found to code for certain diseases (Woerner *et al*. 2001; Chung *et al*. 2008). Conservation of genetic resources can then be done by combining neutral markers and adaptative traits (Tripiana *et al*. 2007).

1.4. Importance of *Allanblackia floribunda*

The bark extracts of *A*. *floribunda* are locally used against coughs, dysentery, diarrhea, toothache (Raponda-Walker and Sillans 1961), and as an aphrodisiac and pain reliever. Consequently, the species is one of the most commonly used medicinal plants in Cameroon (Laird 1996) and possibly has even greater potential. For example, Guttiferone F, an HIV-inhibitor (Fuller *et al*. 1999) was found in the extracts of the heartwood from *A*. *floribunda* (Locksley and Murray 1971). The active compounds of its bark contain prenylated xanthone, a natural product acting against human epidermoid carcinoma of the nasopharynx cancer line (Nkengfack *et al*. 2002). *Allanblackia* fruits contain many seeds rich in a hard

white fat (67.6 – 73 % of seed mass, Foma and Abdala 1985). This fat consists mostly of oleic and stearic acids (39-45 % and 52-58 % respectively, Hilditch *et al*. 1940), which are reported to lower plasma cholesterol levels (Bonanome and Grundy 1988), thus reducing the risks of heart attack. Owing to this fat profile, *Allanblackia* oil is used for margarine production and in soap and ointments manufacture. The seeds extracted from *Allanblackia* fruits by local communities are traded, as a supply chain has been established in the Humid Tropics of Africa (http://www.allanblackia.info/) through Novella projects. Unilever PLC has created a guaranteed market for the product, and other companies are increasingly enrolled in this industry. As a consequence of increasing demand in *Allanblackia* seed fat, seed extraction from fruits collected in wild stands is not sustainable. Cultivating *Allanblackia* trees is an alternative, as a public-private partnership has been established between the United Nations Development Programme, Unilever, the World Agroforestry Centre (ICRAF), a Dutch Non-Governmental Organization (SNV) and local NGOs to domesticate trees of the species for fruit/seed production. Currently, attempts to improve *Allanblackia* trees have been restricted to the development of appropriate protocols for rooting of leafy stem cuttings (Atangana *et al*. 2006), and to surveys of clonal variation in the rooting of cuttings (Atangana and Khasa 2008) in *A. floribunda*. The present thesis is intended to initiate tree breeding in *A. floribunda*, as a model for tree improvement in the genus *Allanblackia*.

1.5. Research questions

This study will seek to answer the following questions:

- Does phenotypic variation exist in fruit/nut traits in tallow-tree? If so, what is the amount and pattern of phenotypic diversity in *A. floribunda* from wild stands, and is part of it under genetic control?
- Do "plus-trees", which combine desirable attributes of several traits, exist in the studied species for fruit/nut production, so as to constitute a breeding population?
- What is the level and amount of neutral genetic diversity in populations from wild stands, and in the putative first-cycle breeding population of tallow-tree? Does

"plus-trees" selection's for fruit/nut production in *A. floribunda* in wild stands affect neutral genetic diversity in the species?

1.6. Hypothesis

The main hypothesis underlying this study is that phenotypic variation does exist in *A. floribunda* for fruit/seed production. Therefore "plus-trees" selection for breeding is possible, because part of this variation is under genetic control, and that breeding tallow-tree from the minimum breeding population-size would not reduce wild nuclear neutral genetic diversity.

1.7. Objectives

The main objective of this study is to initiate *A. floribunda* tree breeding and genetic resource's preservation throughout tree improvement in Cameroon. Specific objectives are:

a. to develop methods for assessing the stearic and oleic acid contents in the seed fat of *A. floribunda* trees, and estimate the pattern and amount of variation in wild stands with respect to fruit/seed characters in the species, thereby selecting "plus-trees" for breeding (chapter 2);

b. to develop and characterize microsatellite markers in *A. floribunda*, and test cross-species amplification in *Allanblackia gabonensis* and *Allanblackia stanerana* (chapter 3);

c. to assess nuclear genetic diversity, gene flow and temporal size-fluctuations in *A. floribunda* natural populations, the effects of these fluctuations on genetic diversity, and the impact of "plus-tree" selection for breeding on genetic diversity in *A. floribunda* (chapter 4).

1.8. Literature cited

Achard, F., H.D. Eva, H-J. Stibig, P. Mayaux, J. Gallego, T. Richards and J-P. Malingreau. 2002. Determination of deforestation rates of the world's humid tropical forests. *Science* **297**: 999-1002. Doi: 10.1126/science.1070656.

Armour, J. A. L., and A. J. Jeffreys. 1992. Biology and applications of human minisatellite loci. *Current Opinion in Genetics and Development* **2**: 850-856.

Atangana, A.R., Z. Tchoundjeu, J.M. Fondoun, E. Asaah, M. Ndoumbe, and R.R.B. Leakey. 2001. Domestication of *Irvingia gabonensis*: 1. Phenotypic variation in fruits and kernels in two populations from Cameroon. *Agroforestry Systems* 53: 55-64.

Atangana, A.R., V. Ukafor, P. Anegbeh, E. Asaah, Z. Tchoundjeu, J-M. Fondoun, M. Ndoumbe and R.R.B. Leakey. 2002. Domestication of *Irvingia gabonensis*: 2. The selection of multiple traits for potential cultivars from Cameroon and Nigeria. *Agroforestry Systems*: **55**(3): 225-229.

Atangana, A.R., Z. Tchoundjeu, E.K. Asaah, A.J. Simons, and D.P. Khasa 2006. Domestication of *Allanblackia floribunda*: Amenability to vegetative propagation. *Forest Ecology and Management* **237**: 246-251. Doi: 10.1016/j.foreco.2006.09.081.

Atangana, A.R., and D.P. Khasa. 2008. Preliminary survey of clonal variation in rooting of *Allanblackia floribunda* leafy stem cuttings. *Canadian Journal of Forest Research* **38**: 10-15.

Aubé, J. 1996. Etude pour favoriser le développement des produits forestiers non-ligneux dans le cadre du Central African Regional Program for the Environment (CARPE). Washington, DC, Forestry Support Program, USAID. http://carpe.umd.edu/Products/

Avise, J.C. 2004. Molelar markers, natural history and evolution. Second edition, Sinauer Associates, Inc. Publishers, Sunderlan, Massachussetts. 684 pp.

Awadalla, P., A. Eyre-Walker, and J. M. Smith. 1999. Linkage disequilibrium and recombination in hominid mitochondrial DNA. *Science* **286** (5449): 2524-2525.

Baldwin, T.C., P.E. Quah, and A.R. Menzies. 1998. A serotaxonomic study of *Acacia* gums exudates. *Phytochemistry* **50**(4): 599-606.

Bamps, P. 1969. Notes sur les Guttiferae d'Afrique tropicale. *Bulletin du Jardin botanique national de Belgique* **39**(4): 345-372.

Beckmann, J.S., and M. Soller. 1983. Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping and costs. *Theoretical and Applied Genetics* 67: 35-43.

Bonanome, A., and S.M. Grundy. 1988. Effect of dietary stearic acid on plasma cholesterol and lipoprotein levels. *New England Journal of Medicine* 318(19): 1244-1248.

Born, C., O.J. Hardy, M.H. Chevallier, S. Ossari, C. Attéké, J. Wickings, and M. Hossaert-McKey. 2008. Small-scale spatial genetic structure in the Central African rainforest tree species *Aucoumea klaineana*: a stepwise approach to infer the impact of limited gene dispersal, population history and habitat fragmentation. *Molecular Ecology* 17: 2041-2050.

Burrows, W.H., B.K. Henry, P.V. Back, M.B. Hoffmann, L.J. Tait, E.R. Anderson, N. Menke, T. Danaher, J.O. Carter, and G.M. McKeon. 2002. Growth and carbon stock change in eucalypt woodlands in northeast Australia: ecological and greenhouse sink implications. *Global Change Biology* 8(8): 769-784.

Cheliak, W.M., J.A. Pitel, and G. Murray. 1985. Population structure and the mating system of white spruce. *Canadian Journal of Forest Research* 15(2): 301-308.

Chung, H., D.J. Young, C.G. Lopez, T.A.T. Le, J.K. Lee, D. Ream-Robinson, S.C. Huang, and J.M. Carethers. 2008. Mutation rates of TGFBR2 and ACVR2 coding microsatellites in human cells with defective DNA mismatch repair. *PLos ONE* 3(10): 1-9.

Coart, E., V. Lamote, M. De Loose, E. van Bockstaele, P. Lootens, and I. Roldan-Ruiz. 2002. AFLP markers demonstrate local genetic differentiation between two indigenous Oak species [*Quercus robur* L. and *Quercus petraea* (Matt.) Liebl.] in Flemish populations. *Theoretical and Applied Genetics* 105: 431-439.

Collevatti, R.G., D. Grattapaglia, and J.D. Hay. 2001. Population genetic structure of the endangered tropical tree species *Caryocar braziliense*, based on variability at microsatellite loci. *Molecular Ecology* 10: 349-356.

Coyne, J.F. and G.C. Keith. 1972. Geographic survey of monoterpenes in loblolly and shortleaf pines. USDA For. Serv. Res. Paper SO-79. 12 p.

Crandall, K.A., O.R.P. Belinda-Emonds, G.M. Mace, and R.K. Wayne. 2000. Considering evolutionary processes in conservation biology. *Trends in Ecology and Evolution* 15: 363-366.

Cunningham, A.C. 1992. People, park and plant use: research and recommendations for multiple use zones and development alternatives around Bwindi Impenetrable National Park, Uganda. CARE Intl., Kampala, Uganda.

Dafni, A. 1992. Pollination ecology: a practical approach. Oxford University Press, Oxford, 250 pp.

Dawson, I.K., A.J. Simons, R. Waugh, and W. Powell. 1995. Diversity and genetic differentiation among subpopulations of *Gliricidia sepium* revealed by PCR-based assays. *Heredity* 74: 10-18.

Dayanandan, S., J. Dole, K. Bawa, and R. Kesseli. 1999. Population structure delineated with microsatellite markers in fragmented populations of tropical tree, *Carapa guianensis* (Meliaceae). *Molecular Ecology* 8: 1585-1592.

de la Torre, A., C. Lopez, E. Yglesias, and J.P. Cornelius. 2008. Genetic (AFLP) diversity of nine *Cedrela odorata* populations in Madre de Dios, southern Peruvian Amazon. *Forest Ecology and Management* 255: 334-339.

Denys, E. 1980. A Tentative Phytogeographical Division of Tropical Africa Based on Mathematical Analysis of Distribution Maps. *Bulletin du Jardin botanique national de Belgique* 50(3/4): 465-504.

Diaz, V., L.M. Munis, and E. Ferrer. 2001. Random amplified polymorphic DNA and amplified fragment length polymorphism assessment of genetic variation in Nicaraguan populations of *Pinus oocarpa*. *Molecular Ecology* 10: 2593-2603.

Diamond, J. 2002. Evolution, consequences and future of plant and animal domestication. *Nature* 418: 700-707.

Dodd, R.S., and N. Kashani. 2003. Molecular differentiation and diversity among the California red oaks (Fagaceae; *Quercus* sect. Lobatae). *Theoretical and Applied Genetics* 107: 884-892.

Dodd, R.S., and Z. Afzal-Rafii. 2004. Selection and dispersal in a multispecies oak hybrid zone. *Evolution* 58: 261-269.

Drummond, R.S.M., D.J. Keeling, T.E. Richardson, R.C. Gardner, and S.D. Wright. 2001. Genetic analysis and conservation of 31 surviving individuals of a rare New Zealand tree, *Metrosideros bartlettii* (Myrtaceae). *Molecular Ecology* 9(8): 1149-1157.

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

Falconer, D.S. 1989. Introduction to quantitative genetics. 3rd edition. Longman Scientific and Technical, Essex, England. 432 p.

Falconer, D.S., and T.F.C. Mackay. 1996. Introduction to quantitative genetics 4th edition. Longman, Essex, England.

FAO 1996. Report on the State of the World's Plants Genetic Resource for food and Agriculture, Rome, Italy, 511 p.

FAO 2006. Evaluation des resources forestières mondiales 2005: Progrès vers la gestion forestière durable. FAO, Rome, Italie 320 p.

Fischer, S.G., and L.S. Lerman. 1983. DNA fragments differing by single-pair substitutions are separated in denaturing gradient gels: corresponding with melting theory. *PNAS* 80: 15791583.

Foma, M., and T. Abdala. 1985. Kernel oils of seven species of Zaire. *Journal of the American Oil Chemists Society* 62(5): 910-911.

Forde, M.B., and M.M. Blight. 1964. Geographical variation in the turpentine of Bishop pine. *New Zealand J. Bot*. 2: 44-52.

Franzel, S., H. Jaenicke, and W. Janssen. 1996. Choosing the Right Trees: Setting Priorities for Multipurpose Tree Improvement. *Research Report* **8**, ISNAR, The Hague, The Netherlands. 87 p.

Fuller, R.W., J.W. Blunt, J.L. Boswell, J.H. 2nd Cardellina, and M.R. Boyd. 2003. Guttiferone F, the first prenylated benzophenone from *Allanblackia stuhlmanni*. *Planta Med*. 69(9): 864-866.

Glaubitz, J.C., Y.A. El-Kassaby, and J.E. Carlsson. 2000. Nuclear restriction fragment length polymorphism analysis of genetic diversity in in western redcedar. *Canadian Journal of Forest Research* 30: 379-389.

GoC. 2003. Poverty Reduction Strategy Paper. Government of Cameroon, Yaoundé.

Godbout, J., J.P. Jaramillo-Correa, J. Beaulieu, and J. Bousquet. 2005. A mitochondrial DNA minisatellite reveals the postglacial history of jack pine (Pinus banksiana), a broad-range North American conifer. *Molecular Ecology* 14: 3497-3512.

Hamrick, J.L., and D.A. Murawski. 1991. Levels of allozyme diversity in populations of uncommon Neotropical tree species. *Journal of Tropical Ecology* 7: 395-399.

Hanover, J.W. 1974. Biochemical analysis of tree speciation. *Proc. Third North Amer. For. Biol. Workshop*, Fort Collins, CO. pp. 106-131.

Hanover, J.W. 1992. Applications of terpene analysis in forest genetics. *New Forests* 6: 159-178.

Hartemink, A.E., R.J. Buresh, B. Jama, and B.H. Janssen. 1996. Soil nitrate and water dynamics in *Sesbania sesban* fallows, weeds fallows, and maize. *Soil Science Society of America Journal* 60: 568-574.

Harry, D.E., B. Temesgen, and D.B. Neale. 1998. Codominant PCR-based markers for *Pinus taeda* developed from mapped cDNA clones. *Theoretical and Applied Genetics* 97: 327-336.

Hartl, D. 1988. A primer of population genetics. 2nd edition. Sinuauer Associates, Inc., Sunderland, USA. 295 p.

Hartl, D., and A. G. Clark. 2007. Principles of population genetics, Fourth edition. Sinauer Associates, Sunderland, Massachussetts. 652 p.

Havey, M.J., J.D. McCreight, B. Rhodes, and G. Taurick. 1998. Differential transmission of the *Cucumis* organellar genomes. *Theor*. *Appl*. *Genet*. 97: 122-128.

Hearne, C., S. Ghosh, and J. Todd. 1992. Microsatellites for linkage analysis of genetic traits. *Trends in Genetics* 8: 288-294.

Hilditch, T.P., M.L. Marea, and Y.A.H. Zaky. 1940. The fatty acids and glycerides of seed fats of *Allanblackia floribunda* and *Allanblackia parviflora*. *Journal of the Society of Chemical Industry* 59: 25.

Jaramillo-Correa, J.P., J. Beaulieu, and J. Bousquet. 2004. Variation in mitochondrial DN reveals multiple distant glacial refugia in black spruce (*Picea mariana*), a transcontinental North American conifer. *Molecular Ecology* 13(9): 2735-2747.

Jaramillo-Correa, J.P., J. Beaulieu, F.T. Ledig, and J. Bousquet. 2006. Decoupled mitochondrial and chloroplast DNA population structure reveals Holocene collapse and population isolation in a threatened Mexican-endemic conifer. *Molecular Ecology* 15(10): 2787-2800.

Jones, M.E., M. Shepherd, R.J. Henry, and A. Delves. 2006. Chloroplast DNA variation and population structure in a widespread forest tree species. *Conservation Genetics* 7(5): 691-703.

Justice, C., D. Wilkie, Q. Zhang, J. Brunner, and C. Donoghue. 2001. Central African forests, carbon and climate change. *Climate Research* 17: 229-246.

Keay, R.W.J. 1989. Trees of Nigeria. Clarendon Press Oxford, UK.

Khasa, P.D., J.P. Jaramillo-Correa, B. Jacquish, and J. Bousquet. 2006. Contrasting microsatellite variation between subalpine and western larch, two closely-related species with different distribution patterns. *Molecular Ecology* 15: 3907-3918.

Laird, S. 1996. *Medicinal Plants of Limbe Botanical Garden Cameroon*. Limbe Botanical Garden, Limbe, Cameroon.

Larochelle, F., E. Forget, A. Rainville, and J. Bousquet. 1998. Sources of temporal variation in sap sugar content in a mature sugar maple (*Acer saccharum*) plantation. *Forest Ecology and Management* 106: 307-313.

Leakey, R.R.B. 1999. Potential for novel food products from agroforestry trees: A review. *Food Chemistry* **66**(1):1-14.

Leakey, R.R.B., and A.C. Newton. 1994. Tropical Trees: the Potential for Domestication and the Rebuilding of Forest Resources. HMSO, London, UK, 284.

Leakey, R.R.B., J.F. Mesén, Z. Tchoundjeu,, K.A. Longman, J.Mc.P. Dick., A. Newton, A. Matin, J. Grace, R.C. Munro, and P.N. Muthoka. 1990. Low-technology techniques for the vegetative propagation of tropical trees. *Commonwealth Forestry Review* 69: 247-257.

Leakey, R.R.B., and A.J. Simons. 1998. The domestication and commercialization of indigenous trees in agroforestry for the alleviation of poverty. *Agroforestry Systems* 38: 165-176.

Leakey, R.R.B., J-M. Fondoun, A. Atangana, and Z. Tchoundjeu. 2000. Quantitative descriptors of variation in the fruits and seeds of *Irvingia gabonensis*. *Agroforestry Systems* 50**:** 47-58.

Leakey, R.R.B., Z. Tchoundjeu, K. Schreckenberg, S. Shackleton, C. and Shackleton. 2005. Agroforestry Tree Products (AFTPs): Targeting Poverty Reduction and Enhanced Livelihoods. *International Journal of Agricultural Sustainability* 16: 5-16.

Ledig, F.T., P.D. Hodgskiss, and D.R. Johnson. 2006. The structure of of genetic diversity in Engelmann spruce and a comparison with blue spruce. *Can. J. Bot*. 84: 1806-1828.

Lemes, M.R., R. Gribel, J. Proctor, and D. Grattapaglia. 2003. Population genetic structure of mahogany (*Swietenia macrophylla* King, Meliaceae) across the Brazilian Amazon, based on variation at microsatellite loci: implications for conservation. *Molecular Ecology* 12**:** 2875-2883.

Leong Pock Tsy, J-M., R. Lumaret, D. Mayne, A. Ould Mohammed Vall, Y.I.M Abutaba, M. Sagna, S. N. O. Rakotondralambo Ra Oseta, and P. Danthu. 2009. Chloroplast DNA phylogeography suggests a West African centre of origin for the baobab, *Adansonia digitata* (Bombacoideae, Malvaceae). *Molecular Ecology* 18: 1707-1715.

Littel, R. C., G. A. Milliken, W. W. Stroup, R. D. Wolfinger, and O. Schabenberger. 2006. SAS for Mixed models, Second edition. SAS Institute, Cary, NC, USA. 814 pp.

Locksley, H.D., and I.G. Murray. 1971. Extractives from Guttiferae. Part XIX. The isolation of two benzophenones, six xanthones and two biflavonoids from the heartwood of *Allanblackia floribunda* Oliver. *J. Chem. Soc* (C), 1332-1340.

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

Lynch, M., and B.G. Milligan. 1994. Analysis of population genetic structure with RAPD markers. *Molecular Ecology* 3: 91-99.

Maki, M., and J. Murata. 2001. Allozyme analysis of the hybrid origin of *Arisaema ehimense* (Araceae). *Heredity* 86: 87-93.

McCouch, S.R., G. Kochert, Z.H. Yu, Z.Y. Wang, G.S. Clush, W.R. Coffman and S.D. Tanksley. 1988. Molecular mapping of rice chromosomes. *Theor. Appl. Genet.* 76: 815-829.

Miller, J.C. and S.D. Tanksley. 1990. Effect of different restriction enzymes probe source, and probe length, on detecting restriction fragment length polymorphism in tomato. *Theor. Appl. Genet.* 80: 385-389.

Milligan, B.G., J.N. Hampton, and J.D. Palmer. 1989. Dispersed repeats and structural reorganization in subclover chloroplast DNA. *Mol. Biol. Evol*. 6: 355-368.

Milliken, G. A., and D. E. Johnson. 1992. Analysis of messy data Vol. I: designed experiments. Chapmann & Hall, New York. 473 p.

Mittermeier, R.A., N. Myers, J.B. Thomsen, G.A.B. da Fonseca, and S. Olivieri, 1998. Biodiversity hotspots and major tropical wilderness areas: areas to setting conservation priorities. *Conservation Biology* 12(3): 516-520.

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

Mylecraine, K.A., J.E. Kuser, P.E. Smouse, and G.L. Zimmermann. 2004. Geographic allozyme variation in Atlantic white-cedar, *Chamaecyparis thyoides* (Cupressaceae). *Can. J. For. Res*. 34: 2443-2454.

Nanson, A. 2001. *Génétique et amélioration des arbres forestiers*. Gembloux: Les presses agronomiques de Gembloux.

Naydenov, K., S. Senneville, J. Beaulieu, F. Tremblay, and J. Bousquet. 2007. Glacialo vicariance in Eurasia: mitochondrial DNA evidence from Scots pine for a complex heritage involving genetically distinct refugia at mid-northern latitudes and in Asia minor. *BMC Evolutionary Biology* 7: 233

Ndoye, O., M. Ruiz-Perez, and A. Eyebe, 1998. The markets of non-timber forest products in the Humid Forest Zone of Cameroon. Network Paper Rural Development Forestry Network 22c. Rural Development Forestry Network, Overseas Development Institute (ODI), London.

Neale, D.D., and P.K. Ingvarsson. 2008. Population, quantitative and comparative genomics of adaptation in forest trees. *Current Opinion in Plant Biology* 11(2): 149-155.

Ngo Mpeck, M.L., E.K. Asaah, Z. Tchoundjeu, and A.R. Atangana**.** 2003. Strategies for the domestication of *Ricinodendron heudelotii*: Evaluation of variability in natural populations from Cameroon. *Food*, *Agriculture* and *Environment* 1 (3&4): 257-262.

Nkengfack, A.E., G.A. Azebaze, J.C. Vardamides, Z.T. Fomum, and F.R. van Heerden. 2002. A prenylated xanthone from *Allanblackia floribunda*. *Phytochemistry* 60: 381-384.

Novick, R.R., C.W. Dick, M.R. Lemes, C. Navarro, A. Caccone, and E. Bermingham. 2003. Genetic structure of Mesoamerican populations of Big-leaf mahogany (*Swietenia macrophylla*) inferred from microsatellite analysis. *Molecular Ecology* 12: 2885-2893.

Olson, D., and D. Dinerstein. 1998. The global 200: a representation approach to conserving the earth's most biologically valuable ecoregions. *Conservation Biology* 12: 502-515.

Orita, M., H. Iwahana, H. Kanazawa, K. Hayashi, and T Sekiya. 1989. Detection of polymorphisms of human DNA by gel electrophoresis and single-strand conformation polymorphisms. *PNAS* USA 86: 2766-2770.

Palmer, J.D. and L.A. Herbon. 1988. Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *J. Mol. Evol*. 28: 87-97.

Palumbi, S.R., and C.S. Baker. 1994. Contrasting population structure from nuclear intron sequence and mtDNA in humpback whales. *Molecular Biology and Evolution* 11(3): 426-435.

Parks, C.R., J.F. Wendel, M. Sewell, and Y-L. Qiu. 1994. The significance of allozyme variation and introgression in the *Liriodendron tulipifera* complex (Magnoliaceae). *American Journal of Botany* 81: 878-889.

Raponda-Walker, A., and R. Sillans. 1961. Les plantes utiles du Gabon. Paul Lechevalier, Paris VI.

Raybould, A.F., J. Goudet, R.J. Mogg, C.J. Gliddon, and A.J. Gray. 1996. Genetic structure of a linear population of *Beta vulgaris* ssp. *Maritime* (sea beet) revealed by isozyme and RFLP analysis. *Heredity* 76: 111-117.

Richardson, B.A., S.J. Brunsfeld, and N.B. Klofenstein. 2002. DNA from bird-dispersed seed and wind-disseminated pollens provides insights into postglacial colonization of whitebark pine (*Pinus albicaulis*). *Molecular Ecology* 11: 215-227.

Rieseberg, L. 1996. Homology among RAPD fragments in interspecific comparisons. *Molecular Ecology* 5: 99-105.

Riginos, C., K. Sukhdeo, and C.W. Cunningham. 2002. Evidence for selection at multiple allozyme loci across a mussel hybrid zone. *Molecular Biology and Evolution* 19: 347-351.

SAS. 2006. *SAS 9.1.3 Service Pack 4 for Windows*, Cary, North Carolina: SAS Institute.

Savage, T.J., B.S. Hamilton, and R. Croteau. 1996. Short-chain alkanes (tissue-specific biosynthesis of h-heptane in *Pinus jeffreyi*). *Plant Physiology* 110(1): 179-186.

Schreckenberg, K., A. Degrande, C. Mbosso, Z. Boli Baboule, C. Boyd, L. Enyong, J. Kanmegne, and C. Ngong. 2002. The social and economic importance of *Dacryodes edulis* (G.Don) H.J. Lam in southern Cameroon. *Forests, Trees and Livelihoods* 12: 15-40.

Shore, J.S., and M. Triassi. 1998. Paternally biased cpDNA inheritance in *Turnera ulmifolia* (Turneraceae). *Am. J. Bot*. 85: 328-332.

Simmonds, N.W. 1987. Principles of crop improvement. Longman, New York, 408 pp.

Simons, A.J., and R.R.B. Leakey. 2004. Tree domestication in tropical agroforestry. *Agroforestry Systems* 61: 167-181.

Snajberk, K. and E. Zavarin. 1976. Mono- and sesqui-terpenoid differentiation of *Pseudotsuga* of the United States and Canada. *Biochem. Systematics and Ecology* 4: 159-163.

Sparks, T.H., J.N. Greatorex-Davies, J.O. Mountford, M.N. Hall, and R.H. Marrs. 1996. The effects of shade on the plant communities of rides implantation woodland and implications for butterfly conservation. *Forest Ecology and Management* 80(1&3): 197-207.

Squillace, A.E. and O.O. Wells. 1981. Geographic variation of monoterpenes in cortical oleoresin of loblolly pine. *Silvae genetica* 30: 127-135.

Strauss, S.H., J. Bousquet, V.D. Hipkins, and Y.-P. Hong. 1992. Biochemical and molecular genetic markers in biosystematic studies of forest trees. *New Forests* 6: 125-158.

Tabuna, H. 1999. The market for Central African non-wood forest products in Europe. In: Sunderland, T. C. H., Clark L. E. & Vantomme P. (eds), Current research issues and prospects for conservation and development, FAO, Rome, pp 251-263.

Tautz, D. 1989. Hypervariability of simple sequence as a source of polymorphic DNA markers. *Nucleic Acid Research* 17: 6463-6471.

Tchoundjeu, Z., E.K. Asaah, P. Anegbeh, A. Degrande, P. Mbile, C. Facheux, A. Tsobeng, A.R. Atangana, M.L. Ngo-Mpeck, and A.J. Simons. 2006. Putting participatory domestication into practice in West and Central Africa. *Forests, Trees and Livelihoods* **16**: 53-69.

Tripiana, V., M. Bourgeois, D. Verhaegen, P. Vigneron, and J-M. Bouvet. 2007. Combining microsatellites, growth, and adaptative traits for managing in situ genetic resources of *Eucalyptus urophylla*. *Canadian Journal of Forest Research* 37 (4): 773-785.

Temesgen, B., D.B. Neale, and D.E. Harry. 2000. Use of haploid mixtures and heteroduplex analysis enhance polymorphism revealed by denaturing gradient gel electrophoresis. *BioTechniques* 28: 114-122.

Temesgen, B., G.R. Brown, D.E. Harry, C.S. Kinlaw, M.M. Sewell, and D.B. Neale. 2001. Genetic mapping of expressed sequence tag polymorphism (ESTP) markers in loblolly pine (*Pinus taeda* L.). *Theoretical and Applied Genetics* 102: 664-675.

Tobolski, J.J., and J.W. Hanover. 1971. Genetic variation in the monoterpenes of Scotch pine. *Forest Science* 17: 293-299.

Vivien, J., and J.J. Faure. 1996. Fruitiers sauvages d'Afrique: Espèces du Cameroun. *Nguilou-Kerou Edition, Saint-Cénéré,* France, 416 p.

Vos, P., R. Hogers, M. Bleeker, M. Reijians, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acid Research* 23(21): 4407-4414.

Wahuriu, A.N. 1999. Characterization of fruit traits towards domestication of an indigenous fruit tree of West and Central Africa: A case study of *Dacryodes edulis* in Cameroon. MSc Thesis, University of Edinburg, 58 pp.

WCFSD. 1999. Our forests – Our future. *Summary Report: World Commission on Forests and Sustainable Development, WCFSD, Winnipeg,* 38 p.

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

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

Wilkinson, A.G. 1999. Poplars and willows for soil erosion control in New Zealand. *Biomass and Bioenergy* 16(4): 263-274.

Wilkinson, R.C. and J.W. Hanover. 1972. Geographical variation in the monoterpene composition of red spruce. *Phytochem*. 11: 2007-2010.

Wilkinson, R.C., J.W. Hanover, J.W. Wright, and R.H. Flake. 1971. Genetic variation in the monoterpene composition of white spruce. *Forest Science* 17: 83-90.

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

Woerner, S.M., J. Gebert, Y.P. Yuan, C. Sutter, R. Ridder, P. Bork, and M. von Knebel Doeberitz. 2001. Systematic identification of genes with coding microsatellites mutated in DNA mismatch repair-deficient cancer cells. *International Journal of Cancer* 93: 12-19.

Yang, T.W., Y.A. Yang, and Z. Xiong. 2000. Paternal inheritance of chloroplast DNA in interspecific hybrids in the genus *Larrea* (Zygophyllaceae). *Am*. *J*. *Bot*. 87: 14521458.

Zavarin, E., K. Snajberk, and J. Fisher. 1975. Geographic variability of monoterpenes from cortex of *Abies concolor*. *Biochem*. *Syst*. and *Ecol*. 3: 191-203.

Zobel, B., and J.P. van Buijtenen. 1989. Wood variation: its causes and control. Springer-Verlag Berlin Heidelberg, Germany.

Zobel, B., and J. Talbert. 1991. *Applied Forest Tree Improvement*, Waveland, New York, USA.

Zurawski, G. and M.T. Clegg. 1987. Evolution of higher-plant chloroplast DNA-encoded genes: Implications for structure-function and phylogenetic studies. *Annu*. *Rev*. *Plant Physiol*. 38: 391-418.

Chapter 2: Phenotypic variation in fruits and seeds and tree selection in *Allanblackia floribunda* for fruit/seed production

This chapter was submitted for publication to Food Chemistry with the title **Phenotypic variation in fruits and seeds and tree selection in *Allanblackia floribunda* (Clusiaceae) for fruit/seed production** by Alain R. Atangana, Erik van der Vlis, Damase P Khasa, Dennis van Houten, Jean Beaulieu and Harrie Hendrickx

2.1 Résumé

Nous avons déterminé la possibilité d'amélioration génétique de *Allanblackia floribunda* pour la production des fruits et des graines en échantillonnant 17 à 40 fruits de 70 arbres chacun distribués dans 4 sites en zone de forêt naturelle du Cameroun. La matière grasse a été extraite des graines, et les teneurs en acides stéarique et oléique déterminées à l'aide de méthodes développées au cours de cette étude. Des variations hautement significatives ont caractérisé les caractères des fruits et des graines dans et entre les arbres, alors qu'aucune différence significative n'a été observée entre les sites. Les caractères du fruit et la masse moyenne des graines ont révélé des répétabilités modérées. Des corrélations fortes et positives ont été observées entre les teneurs en acides stéarique et oléique. Les profils d'acides gras ne variaient pas avec les autres caractères mesurés. Vingt 'arbres-plus' ont été sélectionnés pour l'élevage, et les implications pour l'amélioration sont discutées.

2.2 Abstract

Allanblackia floribunda is a tropical forest-tree species that is valued for its seeds, which are rich in hard fat consisting mostly of stearic and oleic acid. We determined whether the species could be genetically improved for fruit/seed production by sampling 17 to 40 fruits from each of 70 trees that were distributed among four sites in wild stands. Fat was extracted from the seeds, and stearic and oleic acid content of the fat was estimated using methods developed in this study. Highly significant between- and within-tree variation characterized fruit and seed characters, while between-site phenotypic variation was not significant. Moderate repeatabilities were identified in fruit characters and mean seed mass. Strong positive relationships were found between stearic and oleic acid contents of seed fat. Seed fat profiles were not found to vary with other fruit characters. Twenty "plus trees" were selected for breeding, and implications for improvement are discussed.

2.3 Introduction

Forest tree breeding activities, which aim at producing a specially desired product (Zobel and Talbert 1991 p. 6), rely heavily on intraspecific variation in characters of importance for improvement. The reason is that genetic progress is partly determined by the amount of phenotypic variation existing in the species to be improved, because for a given selection intensity, the larger the phenotypic variation the larger the selection differential to the population average (Zobel and Talbert 1991 p. 131). Therefore, tree breeding programs usually start with the determination of the amount and pattern of variability in wild stands using a nested sampling procedure (Zobel and Talbert 1991 p. 58), followed by the establishment of provenance trials (Nanson 2001 p. 457-458) allowing estimation of additive genetic variance (Falconer and Mackay 1996 p. 1) which is the variance of breeding values (Falconer and Mackay 1996 p. 125). However, genotypic gains (Nanson 2001 p. 154) can be achieved through the selection and propagation of "plus trees" from wild stands using clonal approaches, provided that the phenotypic variance for the character of interest is under genetic control.

Although forest trees provide communities with various products, wood has traditionally been the product of interest that has been targeted in tree improvement programs (Zobel and van Buijtenen 1989). Forests are also the sources of some valuable non-timber products, including sap (e.g., *Acer saccharum*, Larochelle *et al.* 1998), resins (e.g. *Pinus jeffreyi*, Savage *et al.* 1996), gums (e.g., *Acacia longifolia*, Baldwin *et al.* 1998), fruits (e.g., *Irvingia gabonensis*, Leakey *et al.* 2000) and nuts (e.g., *Ricinodendron heudelotii*, Ngo Mpeck *et al.* 2003), but they have not been valued as much as wood for timber production and fiber for pulp and paper.

Forest tree breeding for fruit and seed production is being increasingly implemented in the humid tropics of Africa as part of the domestication of high-value indigenous multi-purpose trees (Leakey and Simons 1998; Leakey *et al.* 2005a; Tchoundjeu *et al.* 2006). Candidate species that have been targeted for improvement include *Allanblackia*

floribunda Oliver, commonly known as tallow-tree or ouotera, which is a member of the mangosteen family (Guttiferae Juss. 1789 vs. Clusiaceae Lindl. 1836). Mature *A. floribunda* is a medium-sized tree species, about 30 m in height and 80 cm in diameter at breast height, with a natural distribution range that extends from Benin to the Democratic Republic of Congo and North Angola (Bamps 1969). Tallow-tree occurs in mixed stands in evergreen lowland and deciduous forests, which have two wet and two dry seasons, and an annual rainfall ranging between 1900 mm and 4000 mm (Letouzey 1985). Trees bear many-seeded, berry-like brown fruits that are suspended on long pedicels (Vivien and Faure 1996). The bark of *A. floribunda* is locally used against coughs, dysentery, diarrhea, toothache, and as an aphrodisiac and pain reliever. Moreover, its seeds are rich in a hard white fat (67-73%, Foma and Abdala 1985) consisting mostly of stearic and oleic acids (Hilditch *et al.* 1940), hence the species' common name. Oleic and stearic acids are reported to lower plasma cholesterol levels (Bonanome and Grundy 1988), thus reducing the risks of heart attack. Owing to this profile, *A. floribunda* seed fat is used for margarine production and in the manufacture of soap and ointments. The seeds that are extracted from *A. floribunda* fruits by local communities are purchased by Unilever PLC. The company has created a guaranteed market for the product, which is expected to grow into a new US$120 million industry for Africa. Currently, attempts to improve *A. floribunda* trees have been restricted to the development of appropriate protocols for rooting of leafy stem cuttings (Atangana *et al.* 2006), and to surveys of clonal variation in the rooting of cuttings (Atangana and Khasa 2008). The germination rate of *A. floribunda* seeds is very low (less than 5%, Vivien and Faure 1996). The objectives of the present study, which is part of a wider program aimed at improving trees in the genus *Allanblackia*, are two-fold: 1) to develop methods for assessing the stearic and oleic acid content in the seed fat of *A. floribunda* trees, and 2) to estimate the pattern and amount of variation in wild stands with respect to fruit/seed characters in the species using a nested sampling procedure, thereby selecting "plus trees" for breeding.

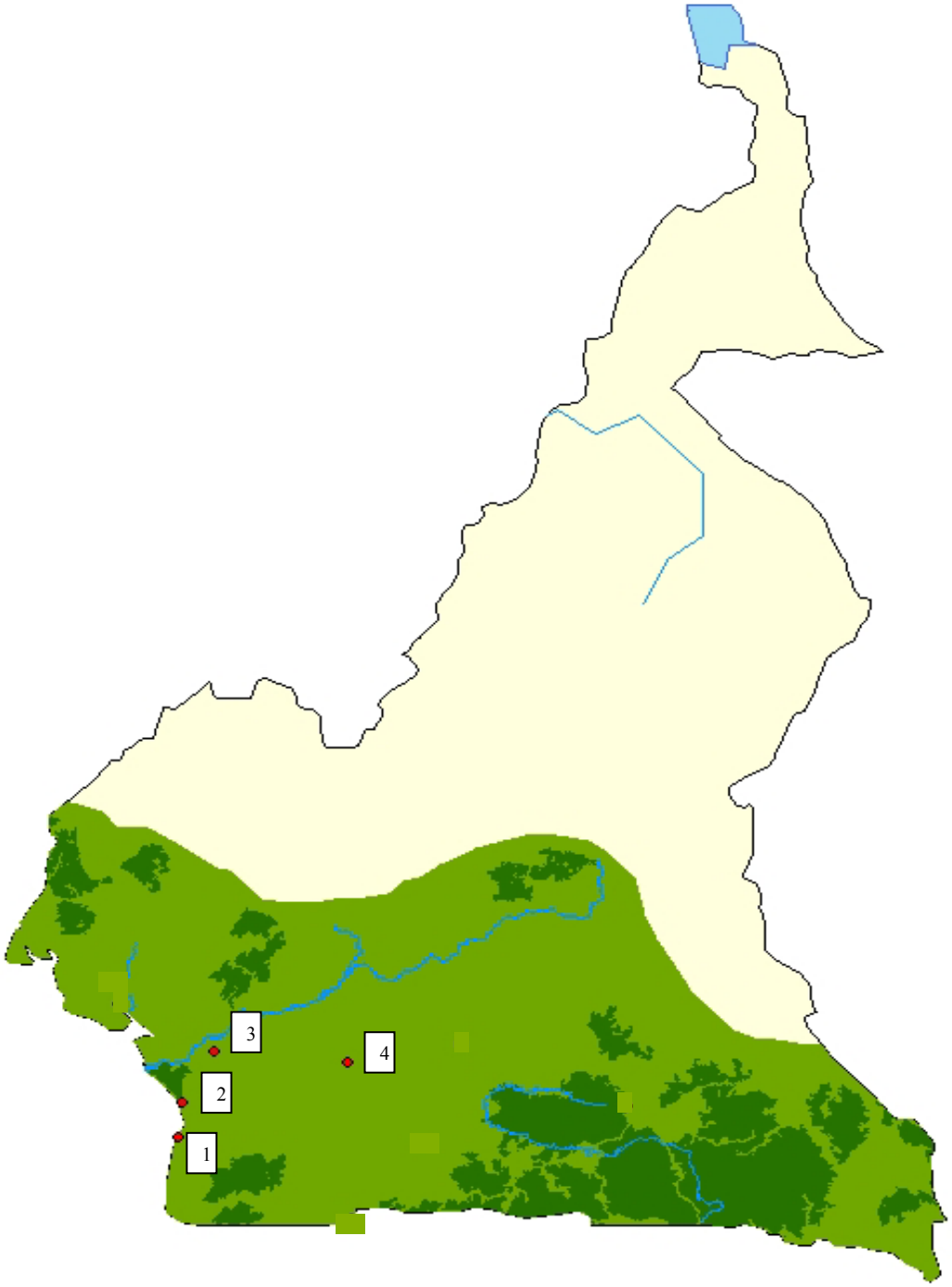

Figure 2.1. Location of sampling *Allanblackia floribunda* sites in Cameroon (1: Bikondo; 2: Mouanko; 3: Yalpenda; 4: Ngoumou)

2.4 Materials and methods

2.4.1 Morphological characterization: plant and fruit sampling, fruit and seed measurement

Four sites (Fig. 2.1) were randomly sampled in the natural range of *A. floribunda* in Cameroon, and 70 mature fruiting trees (6, 23, 8 and 33 trees in Bikondo, Mouanko, Ngoumou and Yalpenda, respectively) randomly selected were visited during the fruiting season spanning from August to October, in year 2005. Unless otherwise stated, forty fruits were randomly collected on each tree, weighed using scales graduated to 2 kg, and measured (length and diameter) using calipers graduated to 0.1mm (Leakey *et al*. 2000). Seeds were immediately extracted, weighed using portable scales graduated to 2 g and counted. Seeds from all fruits collected on the same tree were bulked, kept separate in a collection bag, and brought to the World Agroforestry Centre (ICRAF) laboratory in Yaoundé for fat extraction.

2.4.2 Fat extraction

Seeds were sun-dried for two days before milling in a domestic grinder. The resulting powder was heated for 30 min and immediately introduced into a manual press for fat extraction. The hot liquid fat became solid at ambient temperature (average temperature in Yaoundé: 25^{o}C; Letouzey 1985), and fat from seeds of each tree was weighed using an electronic balance (Mettler Toledo PB 3002) to the nearest 100 mg. Each fat sample was packed in a thin plastic bag, and shipped to the Unilever Research and Development Laboratory, Vlaardingen (Netherlands), for fat analyses.

2.4.3 Analytical method for the determination of stearic- and oleic acid contents in hard fat samples

2.4.3.1 Gas chromatography (GC) analysis

Approximately 100 mg of each fat sample was weighed into a 50 ml Falcon tube to which 5 ml of *tert*-butylmethylether (LiChrosolv 99.8%, Merck, Darmstadt, Germany) was added. After the fat was dissolved, 0.5 ml of the solution was transferred to a 2 ml Eppendorf vial, together with 100 µl of a 3 % sodium methoxide and 400 µl of *n*-heptane, (LiChrosolv 99.3% Merck, Darmstadt, Germany) to derivatize the fatty acids. The 3% methoxide solution was prepared by combining 100 µl of 30 % sodium methoxide solution (Fluka, Buchs, Switzerland) with 900 µl methanol. After 2 min of vortex mixing, the solution was centrifuged for 3 min at 13,000 rpm. Any water remaining was then removed by adding 100 mg of dry Na_2SO_4 (Merck, Darmstadt, Germany), after which the solution was again vortex-mixed for 2 min and centrifuged for an additional 3 min at 13,000 rpm. Finally, 500 µl was transferred to a GC vial along with 500 µl of the internal standard (IS) solution which consisted of 650 mg methyl tricosanoate diluted in 100 ml with *n*-heptane.

The fatty acids, which had been derivatized as methyl esters, were identified and quantified by gas chromatography (Hewlett-Packard (HP) 6890 series GC system), followed by FID (flame ionization detection). A 1 µl volume of each sample was injected into the GC system, which consisted of an Agilent 7683 series autosampler, coupled to a HP 7683 series sample injector (275 ^{o}C), and a Varian WCOT fused silica capillary column with a CP-WAX 52CB coating (10 m x 0.10 mm ID; film thickness 0.2 µm; 275 ^{o}C detection temperature). Helium was used as the carrier gas (0.5 ml/minute, 350 kPa). Total run time for each sample was 21 minutes, which included ramping of the column temperature from 100 ^{o}C to 165 ^{o}C (30 ^{o}C/minute), followed by 165 ^{o}C to 255 ^{o}C (10 ^{o}C/minute).

2.4.3.2 Peak integration and processing

Characterization of the FAME (Fatty Acid Methyl Ester) profile of each sample was accomplished, using TotalChrom software (Perkin Elmer, MA, USA) to integrate the peaks (Fig. 2.2). A FAME reference mixture (#189-19, Sigma, product No. L9405, C4 to C24) was used to assign all individual FAME peaks and to establish their retention times. Once peaks found in the samples and reference mixture were identified and integrated, all peak areas were adjusted to compensate for the FID response for each FAME profile, as follows:

CFAME (%) = AreaFAME*100 / AreaTOTAL

where:
CFAME (%) = FAME content, normalized to 100% total FAME
AreaFAME = Corrected peak area for FAME (Area/FIDx)
AreaTOTAL = Sum of the corrected peak areas
FIDx = FID factor for FAME x

FID correction factors were calculated from the molecular weight of the FAME using the following formula:

FIDx = Mwx / (nx – 1) * 12.01

where:
Mwx = molecular weight for FAME x
nx = the number of carbon atoms in FAME x
12.01 = carbon atomic weight.

The FID factor of C16:0 was taken as the reference (1.00). All other factors used in the calculation were relative to this value (Table 2.1). For isomers, response factors were identical to the response factor of the corresponding saturated fatty acid. The adjusted

areas were used to determine the amount of stearic and oleic acids in the corresponding sample. The stearic methyl ester content (mg) was calculated using the internal standard according to:

stearic acid (mg) = adjusted peak area stearic acid / area IS * (x ml IS * [IS])

Sample oleic acid content was similarly calculated.

Table 2.1. List of response factors for peak integration in *Allanblackia floribunda* seed fat chromatography (FA: Fatty Acid; Mw: molecular weight; n: number of carbon atoms; FID: Flame Ionization Detector; ref: reference)

FA methylester	Mw	n-1	FID correction factor	Response factor	Calibration factor
4:0	102.13	4	2.126	1.51	0.661806
6:0	130.19	6	1.807	1.28	0.779034
8:0	158.24	8	1.647	1.17	0.854588
9:0	172.27	9	1.594	1.13	
10:0	186.30	10	1.551	1.10	0.907346
11:0	200.32	11	1.516	1.08	0.928221
12:0	214.35	12	1.487	1.06	0.946326
13:0	228.37	13	1.463	1.04	0.962249
14:0	242.40	14	1.442	1.02	0.976289
15:0	256.42	15	1.423	1.01	0.988832
16:0	270.46	16	1.407	1.00 (ref)	1.00 (ref)
17:0	284.49	17	1.393	0.99	1.010101
18:0	298.52	18	1.381	0.98	1.019287
19:0	312.52	19	1.370	0.97	
20:0	326.57	20	1.360	0.97	1.035236
21:0	340.57	21	1.350	0.96	1.042308
22:0	354.62	22	1.342	0.95	1.048679
23:0	368.62	23	1.334	0.95	1.054723
24:0	386.68	24	1.328	0.94	1.060129

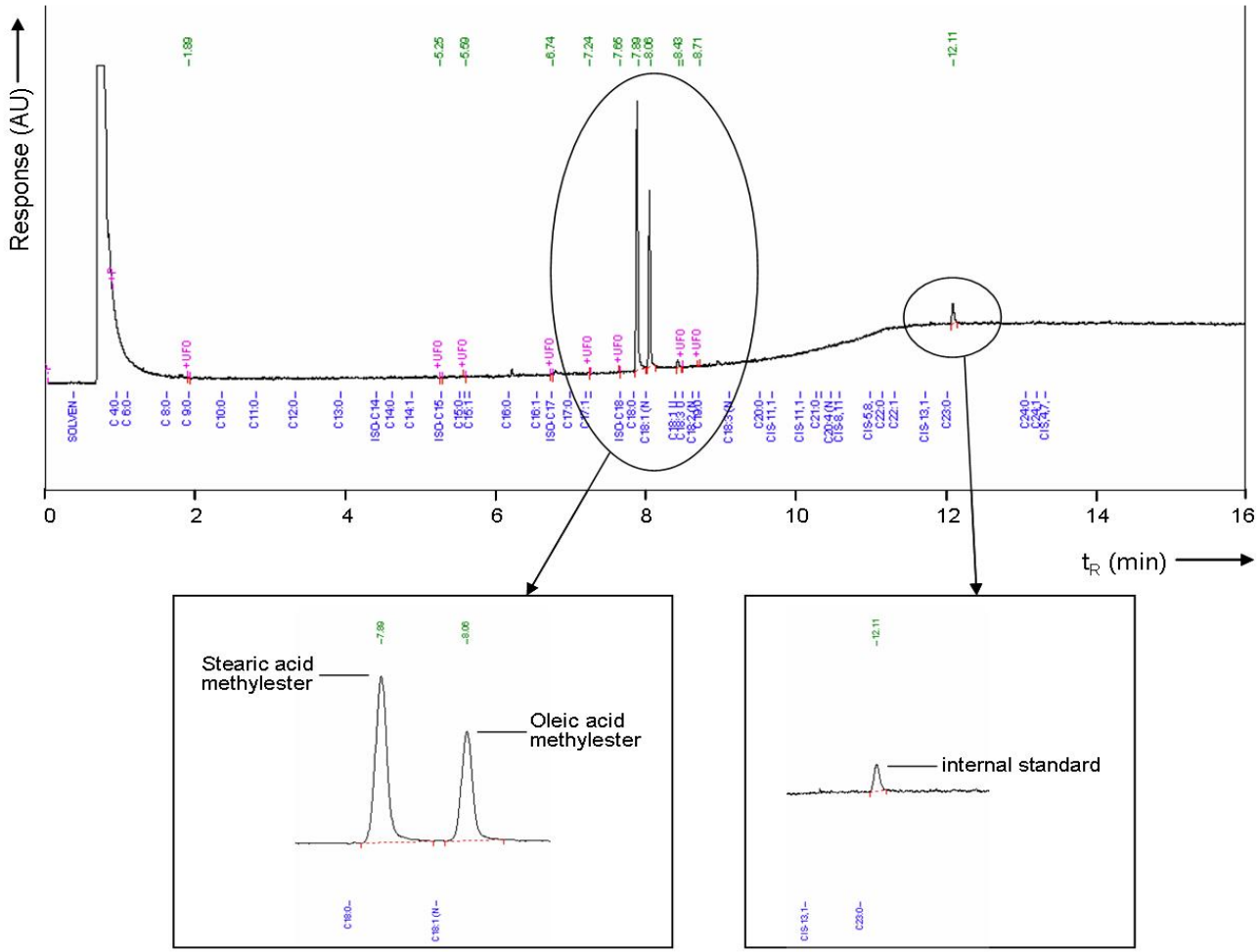

Figure 2.2. Typical chromatogram of an *Allanblackia floribunda* hard fat sample.

2.4.4 Statistical analysis

2.4.4.1 Variation

Variance components were estimated in SAS (PROC MIXED, SAS Institute Inc. 2006) using the restricted maximum likelihood method. All treatment effects were assumed to be random, and when fruit mass, fruit length, fruit diameter, number of seeds per fruit, and total and mean seed mass per fruit were used as dependent variables, data analyses were performed following the linear statistical model for a two-stage nested design (Montgomery 1991 p. 440):

$$Y_{ijk} = \mu + \tau_i + \beta_{j(i)} + e_{(ij)k}$$

where Y_{ijk} is the average value for the dependent variable for the *i*th site, for the *j*th tree nested in the *i*th site, and for the *k*th fruit nested in the *j*th tree nested in *i*th site; μ is the overall mean; τ_i, $\beta_{j(i)}$, and e_{ijk} corresponded, respectively, to site effect, tree nested within site (i.e., the between-individual variance), and the fruit nested within tree nested within site effect (i.e., the within-individual variance, Falconer 1989 p. 139). The last term was the source of random error in the model. Repeatability was estimated using the following formula (Falconer 1989 p. 140):

$$R = V_G + V_{Eg} / V_P$$

where R is the repeatability, and V_G, V_{Eg} and V_P correspond to genotypic, general environmental and total phenotypic variances, respectively. Genotypic and general environmental variances constituted the between-tree variance.

As seeds from fruits of each tree were bulked before fat extraction, variance components for the stearic and oleic acid contents were estimated following the nested model:

$Y_{ij} = \mu + \tau_i + e_{j(i)}$

where Y_{ij} is the average value for the dependent variable for the *i*th site and *j*th tree; μ is the overall mean; τ_i and $e_{j(i)}$ corresponded, respectively, to the site effect and tree nested within-site effect (i.e., between-tree variation). The latter term corresponded to the random error component of the model.

2.4.4.2 Selection

As trees that were visited in this study occurred within natural stands (unknown pedigrees), individual (mass) selection (Nanson 2001 p. 176) was used to identify potential individuals to be included in breeding or production populations. Relationships between measured fruit and seed traits were examined using Pearson product-moment correlations (PROC CORR, SAS Institute Inc. 2006). These correlation coefficients were also useful in identifying potential correlated responses to selection (Falconer and Mackay 1996 p. 165 and p. 318). Further, they allowed reduction of the initial number of characters to a set of characters to be used for individual multiple-trait selection, using the independent culling method (Zobel and Talbert 1991 p. 136-137), since economic weights for the traits of interest have not yet been studied.

2.5 Results

2.5.1 Variation

Among the variance component estimates that are summarized in Tables 2.2 and 2.3, site had no effect on any of the measured characters, while tree-to-tree variation was highly significant ($P < 0.0001$) for each dependent variable. Highly significant ($P < 0.0001$) within-tree variation was also observed in the measured characters, except for stearic and oleic acid content in seed fat, which could not be estimated as seeds were bulked per tree sample before fat extraction. Stearic and oleic acid contents in seed fat ranged from 2.4 mg to 10.3 mg, and from 1.5 mg to 6.7 mg per tree sample, respectively.

Repeatability estimates of mean seed mass per fruit (R = 0.649), fruit length (R = 0.593), fruit mass (R = 0.563), and fruit diameter (R = 0.559) were all moderate, whereas repeatability in total seed mass per fruit and number of seeds per fruit was low (Tables 2.2 and 2.3).

Table 2.2. Variation in *Allanblackia floribunda* fruit characters from four natural stands in Cameroon (SE = Standard Error; computed from the inverse of the estimation information matrix; DF = Degrees of Freedom).

Character	Parameter	DF	Ratio	Variance ± SE	*Z*-value	Pr > Z
Fruit mass (kg)	Site	3	0.285	0.0681 ± 0.079	0.86	0.1944
	Between-trees	276	1.6514	0.4008 ± 0.072	5.60	< 0.0001
	Within-trees	11485	1.0000	0.2427 ± 0.003	35.39	< 0.0001
	Total phenotypic			0.7116		
	Repeatability			0.563		
Fruit length (cm)	Site	3	0	0		
	Between-trees	276	1.4596	18.6622 ± 3.265	5.72	< 0.0001
	Within-trees	11485	1.0000	12.7857 ± 0.361	35.39	< 0.0001
	Total phenotypic			31.4479		
	Repeatability			0.593		
Fruit diameter (cm)	Site	3	0.3264	0.3671 ± 0.370	0.99	0.1607
	Between-trees	276	1.6787	1.8881 ± 0.335	5.63	< 0.0001
	Within-trees	11485	1.0000	1.1247 ± 0.032	35.39	< 0.0001
	Total phenotypic			3.3799		
	Repeatability			0.559		

2.5.2 Phenotypic correlations among traits

Pearson correlations among measured fruit and nut characters of *A. floribunda* trees are summarized in Table 2.4. Strong and positive relationships were found between stearic and oleic acid contents in *A. floribunda* seed fat, and most of the measured characters were moderately to highly correlated with fruit mass (Table 2.4). Moderate relationships were found among seed characters, whereas seed fat characters were weakly related to other measured characters (Table 2.4). From these correlations, only the stearic acid content of seed fat, fruit mass, mean seed mass per fruit, and the number of seeds per

fruit were used to investigate tree selection in *A. floribunda* for fruit/seed production. Oleic acid content was not retained because it was highly positively correlated (r = 0.9903; P < 0.0001) to stearic acid content. Moreover, as mass was positively correlated to dimensional traits of fruits, only the former was used in the selection process.

Table 2.3. Variation in *Allanblackia floribunda* seed characters from four natural stands in Cameroon (SE = Standard Error; computed from the inverse of the estimation information matrix; DF = Degrees of Freedom)

Character	Parameter	DF	Ratio	Variance ± SE	*Z*-value	Pr > Z
Number of seeds/fruit	Site	3	0.06328	10.857 ± 15.611	0.70	0.234
	Between-trees	276	0.4697	80.589 ± 14.987	5.38	<0.0001
	Within-trees	11485	1.0000	171.560 ± 4.847	35.39	<0.0001
	Total phenotypic			236.006		
	Repeatability			0.341		
Total seed mass/fruit (g)	Site	3	0.04943	311.87 ± 423.23	0.74	0.2306
	Between-trees	276	0.6916	4363.58 ± 793.14	5.50	<0.0001
	Within-trees	11485	1.0000	6309.34	35.39	<0.0001
	Total phenotypic			10984.79		
	Repeatability			0.397		
Mean seed mass/fruit (g)	Site	3	0.03398	0.02924 ± 0.148	0.20	0.4217
	Between-trees	276	1.9098	1.6431 ± 0.295	5.56	<0.0001
	Within-trees	11485	1.0000	0.8603 ± 0.024	35.39	<0.0001
	Total phenotypic			2.5326		
	Repeatability			0.649		
Stearic acid content in seed fat (mg)	Site	3	0.04026	0.1528 ± 0.469	0.33	0.3724
	Between-trees	276	1.0000	3.7964 ± 0.671	5.66	<0.0001
	Total Phenotypic	11485		3.9492		
Oleic acid content in seed fat (mg)	Site	3	0.02086	0.0406 ± 0.196	0.21	0.4182
	Between-trees	276	1.0000	1.9439 ± 0.344	5.65	<0.0001
	Total Phenotypic	11485		1.9845		

Table 2.4. Pearson correlations among *Allanblackia floribunda* fruit and seed characters in Cameroon (N = 70; *P*-values are in parentheses)

	Fruit length	Fruit diameter	Total seed mass/Fruit	Number of seeds/Fruit	Mean seed mass/Fruit	Stearic acid content of seed fat	Oleic acid content of seed fat
Fruit mass	0.6351 (<0.0001)	0.8865 (<0.0001)	0.7585 (<0.0001)	0.2315 (0.0538)	0.6746 (<0.0001)	0.0408 (0.7373)	0.0696 (0.5672)
Fruit length		0.3198 (0.0070)	0.5416 (<0.0001)	0.1656 (0.1707)	0.4487 (<0.0001)	0.0919 (0.4494)	0.0847 (0.4857)
Fruit diameter			0.6846 (<0.0001)	0.2767 (0.0204)	0.5611 (<0.0001)	-0.0177 (0.8847)	0.0163 (0.8934)
Total seed mass/Fruit				0.6074 (<0.0001)	0.6109 (<0.0001)	-0.0643 (0.5970)	-0.0612 (0.6151)
Number of seeds/Fruit					-0.2367 (0.0485)	-0.0163 (0.8938)	-0.0471 (0.6987)
Mean seed mass/Fruit						-0.0603 (0.6200)	-0.0228 (0.8515)
Stearic acid content of seed fat							0.9903 (<0.0001)

2.5.3 Tree selection

High tree-to-tree variation was observed in fruit mass, the number of seed per fruits, the mean seed mass, and the stearic acid content of seed fat (Figs. 2.2, 2.3, 2.4 and 2.5) in *A. floribunda*, which allowed us to select "plus trees" for fruit and seed production. Application of the independent culling method to the surveyed trees, which used the four characters retained for selection, identified 20 trees that could constitute a potential breeding population (Table 2.5). Most of the trees selected from this survey were located in Yalpenda and Mouanko (56% and 28%, respectively). Moreover, 42.4%, 37.5% and 30.4% of the surveyed trees in Yalpenda, Ngoumou and Mouanko, respectively, were selected. Trees Y024, Y19, Y007, B1, M30 and NG7 seemed to combine the best characteristics for fruit/seed production (Table 2.5), as determined from this survey.

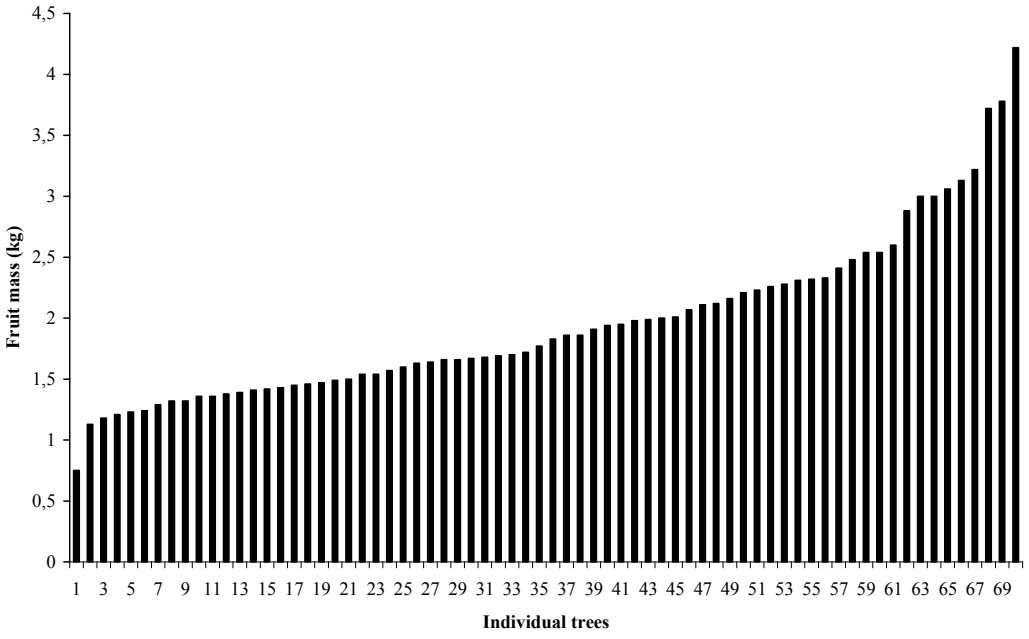

Figure 2.3. Tree-to-tree variation in fruit mass in *Allanblackia floribunda* from wild stands in Cameroon.

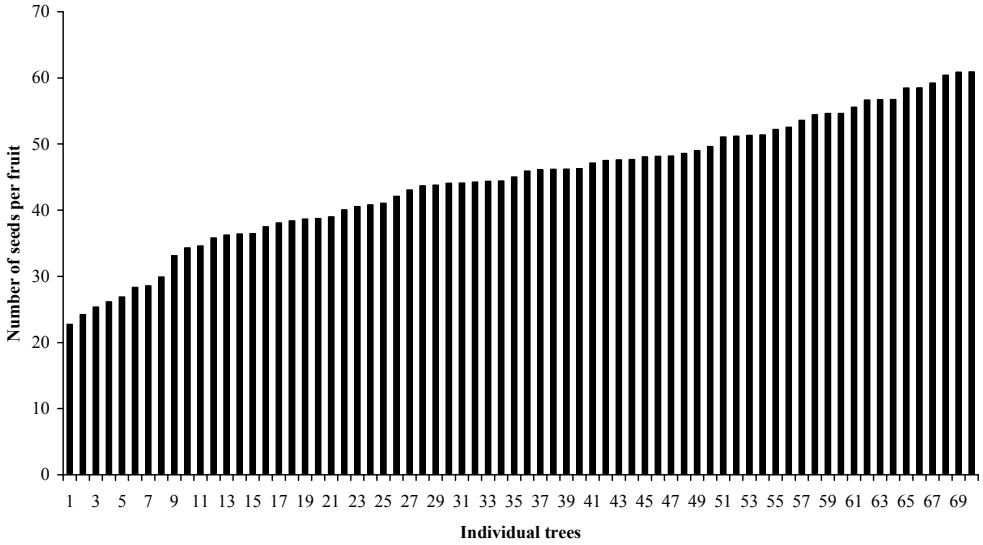

Figure 2.4. Tree-to-tree variation in seed number per fruit in *Allanblackia floribunda* from wild stands in Cameroon.

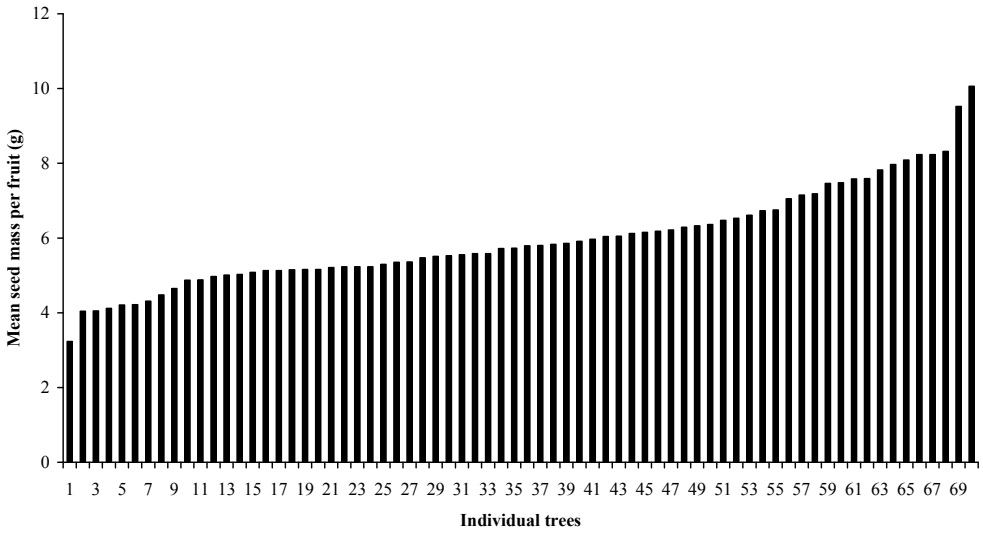

Figure 2.5. Tree-to-tree variation in mean seed mass per fruit in *Allanblackia floribunda* from wild stands in Cameroon.

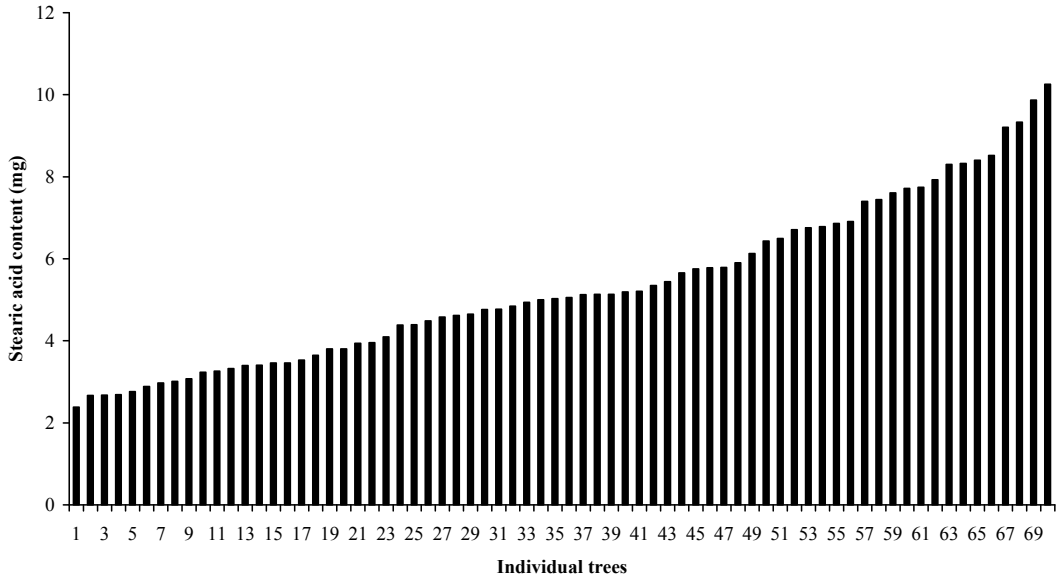

Figure 2.6. Tree-to-tree variation in stearic acid content in seed fat in *Allanblackia floribunda* from wild stands in Cameroon.

2.6 Discussion

The study of natural phenotypic variation within a species using a nested sampling procedure is the first step in any tree improvement program. It permits partitioning of phenotypic variation into among-site, between-tree and sometimes, within-tree sources. The latter is useful for estimating the repeatability of a given trait, which expresses the proportion of the variance of single measurements that is due to permanent differences between individuals, both genetic and environmental (Falconer 1989 p. 139). We have demonstrated for the first time the importance of variation in chemical and physical properties of *A. floribunda* fruits regarding fruit/seed production from trees in wild stands.

In the present study, no significant differences were found among sites for any fruit and seed characteristics that were analyzed. This lack of among-site phenotypic variation

contrasts with the results of Atangana *et al.* (2001) on *Irvingia gabonensis* and could likely be explained by the relative homogeneity of ecological conditions of the sites that were surveyed (Letouzey 1985). Information on the reproductive biology of the species could also provide insights into this lack of among-site phenotypic variation in the measured characters. The broad variability found between- and within-trees in this study is typical of that found in populations of an out-breeding species, and indicates that extensive

Table 2.5. *Allanblackia floribunda* identified "plus trees" for fruit/seed production using the independent culling method on trees surveyed in four wild stands in Cameroon.

Site	Tree number	Stearic acid content of seed fat (mg)	Fruit mass (kg)	Number of seeds per fruit	Mean seed mass (g)	Rank
Yalpenda	Y024	9.204	3.78	49.6	6.29	1
Yalpenda	Y19	7.743	3.06	60.4	6.22	2
Yalpenda	Y007	7.443	2.48	56.65	6.15	3
Bikondo	B1	7.399	2.11	48.15	5.86	4
Mouanko	M30	6.760	2.41	51.2	8.09	5
Ngoumou	NG7	6.714	2.12	46.2	5.97	6
Yalpenda	Y016	5.655	3.22	56.7	7.46	7
Yalpenda	Y41	5.445	4.22	56.7	8.23	8
Mouanko	M41	5.784	1.99	59.2	5.23	9
Yalpenda	Y021	7.609	1.86	45.9	5.72	10
Yalpenda	Y011	6.782	1.91	45.0	5.91	11
Yalpenda	Y89	7.929	2.16	38.4	5.83	12
Mouanko	M34	9.333	1.24	43.8	5.36	13
Mouanko	M7	5.128	1.86	54.6	5.13	14
Mouanko	M14	8.521	1.98	36.4	7.58	15
Mouanko	M5	6.433	2.21	38.7	5.02	16
Yalpenda	Y112	5.899	3.00	44.1	7.82	17
Yalpenda	Y10	5.786	2.26	43.1	5.58	18
Mouanko	M15	5.756	1.49	37.5	5.79	19
Yalpenda	Y98	8.402	1.36	54.6	5.15	20
Minimum value retained for tree selection		5.1	1.2	36	5	
Mean tree selected		6.99	2.34	48.3	6.22	
Overall mean		5.33	1.96	44.5	5.98	

phenotypic diversity exists in fruits and seeds of *A. floribunda* within wild stands. Similar results were found for *I. gabonensis* (Atangana *et al.* 2001; Leakey *et al*. 2005b) and *Dacryodes edulis* (Leakey *et al*. 2002). The high amounts of stearic and oleic acid found in *A. floribunda* seed fat in this study are attractive for the food industry for the design of nutritive and *trans* free products (shortenings, margarine, frying oil, Wassel and Young 2007). With the exception of seed number per fruit and total seed mass per fruit, the majority of the variation was found among trees (Tables 2.2 and 2.3), as is the case for most commercial traits in forest trees (Zobel and Talbert 1991 p. 50). From this result, we can infer that, if those traits are under genetic control, a portion of potential genetic gains could be achieved for most of them through mass selection of the best individuals. Indeed, fruit mass, fruit width fruit length and flesh were found to be heritable in *Olea europa* L. (Zeinanloo *et al*. 2009). Also, fruit mass, shape, flavor and colour were found to be highly heritable in *Mangifera indica* L. (Brettell *et al*. 2004). We assumed in this study that the characters under study are moderately heritable. Repeatability, which sets an upper limit to broad- and narrow-sense heritability values (Falconer 1989 p. 140), was found to be moderate and weak for the fruit and seed physical traits, respectively, indicating that these characteristics are weakly to moderately heritable. From these results, it appears that *A. floribunda* tree improvement activities for fruit/seed production might combine breeding and silvicultural practices (Zobel and Talbert 1991 p. 6). Unfortunately, no repeatability estimates could be obtained for stearic and oleic acid contents seed fat in the present study because seeds collected on the same tree were bulked before fat extraction.

The expected response to selection is the product of intensity of selection, heritability and phenotypic standard deviation (Zobel and Talbert 1991 p. 133). As the heritability of the fruit and seed traits assayed appears to be low, based on the repeatability estimates obtained, high genetic gains in *A. floribunda* fruit/seed characters through tree breeding would be achievable only by applying a high intensity of selection. However, to do so, a large number of trees would have to be assessed for fruit/seed characters. Biophysical characterization of *A. floribunda* trees at a large scale in the natural range of the species would also be advisable to increase potential gains from wild stands, as significant differences might be found among stands with a larger survey of natural stands. This survey

should also be implemented so as to have many fat samples per tree for the assessment of repeatability in stearic and oleic acid content of seed fat.

Prediction of seed characteristics is of practical value in tree improvement programs, as tree selection for seed production could be done indirectly from fruit traits. Strong relationships found between fruit mass and total seed mass per fruit, and between fruit diameter and total seed mass per fruit indicate that total seed mass per fruit could be predicted from fruit mass or fruit diameter, provided that these characters are genetically correlated. Fruit mass could also integrate fruit diameter, although moderate relationships were identified between these two characters. The weak relationships found between fruit characters and seed fat profiles indicate that seed fat characters could not be predicted from fruit characteristics in *A. floribunda*. Weak relationships found between stearic and oleic acid content of seed fat and other seed characters indicate that tree selection for seed fat production should be done separately for fat profiles and other seed characters.

Multiple-trait selection or selection for several traits in forest trees can be implemented using tandem selection, independent culling or by employing a selection index. The latter is based on genetic correlations between traits of interest, which could not be obtained from this study, and the economic weights of these traits, which have not yet being assessed. Multiple-trait selection could be done using just a few measured characteristics in case strong and positive relationships exist between the traits of interest for improvement, through the process of correlated response to selection. A strong and positive phenotypic relationship found between stearic and oleic acid contents in *A. floribunda* seed fat suggests that improvement in one of these traits might lead to improvement in the other. Also, strong and positive relationships that were identified between fruit mass and fruit diameter, and between fruit mass and total seed mass per fruit in wild stands seem to indicate that an improvement in one character might cause a simultaneous increase in the other. This is why we have decided to limit the characters for *A. floribunda* tree selection to fruit mass, number of seeds per fruit, mean seed mass and stearic acid content in seed fat. Those characters are of main interest in the studied species for fruit/seed production. This is only a first step in a *A. floribunda* breeding program that we are setting up and genetic tests are

needed to estimate genetic correlations between the characters of interest and to confirm that phenotypic correlations are representative of genetic correlations. Clonal tests allowing the assessment of genotype by environment interactions and estimation of broad-sense heritabilities and associated genetic gains in the characters of interest for fruit/seed production in *A. floribunda* are advisable, as recent studies have demonstrated that vegetative propagation of *A. floribunda* is possible using leafy stem cuttings (Atangana *et al*. 2006), and clonal variation in rooting in this species (Atangana and Khasa 2008). Interestingly, application of independent culling to the pooled surveyed sites identified trees from each site that could constitute a breeding population, which should minimize inbreeding during *A. floribunda* tree improvement for fruit/seed production. The "top six" trees identified in this study (trees Y024, Y19, Y007, B1, M30 and NG7) could have constituted ideotypes for fruit/seed production in *A. floribunda*, whether traits of interest were under genetic control, and regardless whether they constitute a production population.

The 20 mature trees that were selected for breeding from this study (out of 70) appeared to be high, given the low- to moderate repeatability values that we estimated. However, this number lies between the effective sizes of the number of founders required to maintain 90% of genetic variation for quantitative traits in the source population for 200 years (Ralls and Ballou 1986). Studies are in progress to compare the level of genetic diversity in the selected breeding population with what exists in wild stands in *A. floribunda* using nuclear markers.

In conclusion, it is clear from this study that extensive phenotypic variation exists in fruit and seed traits of *A. floribunda*. As no genetic information is available, expected progress towards improvement of the species for fruit/seed production is low. However, this study represents the first step, and it is worth implementing tree improvement in *A. floribunda* considering the time it will take to set up a breeding program including progeny tests, controlled crosses, and eventually, clonal tests. Recommendations on how best to improve *A. floribunda* for fruit and seeds production from this study include optimization of silvicultural practices for an increase in fruit/seed production, and assessment of fruit

mass, mean seed mass per fruit, number of seeds per fruit, and stearic acid content of the seed mass.

2.7 Acknowledgements

This work was jointly supported by Unilever PLC, together with grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) through the Fonds Général pour les Etudes Supérieures (FGES), from Laval University to Atangana, through a NSERC discovery grant to Khasa, and from a Canadian Wood Fibre Centre grant to Beaulieu. We also thank the World Agroforestry Centre (ICRAF) for partially funding this work. Drs. Tony Simons, Zac Tchoundjeu, and Ramni Jamnadass, and Messrs. E. Asaah (ICRAF), Stéphane Daigle (CEF) and Dr. Bill Parsons are thanked for their help.

2.2. Literature cited

Atangana, A.R., Z. Tchoundjeu, J.-M. Fondoun, E.K Asaah, M. Ndoumbe, & R.R.B. Leakey. 2001. Domestication of *Irvingia gabonensis*: 1. Phenotypic variation in fruit traits in 52 trees from two populations in the humid lowlands of Cameroon. *Agroforestry Systems* 53: 55-64.

Atangana, A. R., Z. Tchoundjeu, E.K. Asaah, A.J. Simons, and D.P. Khasa. 2006. Domestication of *Allanblackia floribunda*: Amenability to vegetative propagation. *Forest Ecology and Management* 237: 246-251. doi: 10.1016/j.foreco.2006.09.081.

Atangana, A. R., and D.P. Khasa. 2008. Preliminary survey of clonal variation in rooting of *Allanblackia floribunda* leafy stem cuttings. *Canadian Journal of Forest Research*. 38: 10-15 doi:10.1139/X07-151.

Baldwin, T. C., P.E. Quah, and A.R. Menzies. 1998. A serotaxonomic study of *Acacia* gums exudates. *Phytochemistry* 50(4), 599-606.

Bamps, P. 1969. Notes sur les Guttiferae d'Afrique tropicale. *Bulletin du Jardin botanique national de Belgique* 39(4): 345-372.

Bonanome, A., and S.M. Grundy. 1988. Effect of dietary stearic acid on plasma cholesterol and lipoprotein levels. *New England Journal of Medicine* 318(19): 1244-1248.

Bretell, R.I.S., P.R. Johnson, V.J. Kulkarni, W. Muller, and I.S.E. Bally. 2004. Inheritance of fruit characters in hybrid mangoes produced through controlled pollination. *Acta Horticulturae* 645: 319-326.

Falconer, D. S. 1989. *Introduction to quantitative genetics*. *3rd edition*. Essex: Longman Scientific and Technical.

Falconer, D. S., and T.F.C. Mackay. 1996. *Introduction to quantitative genetics 4th edition*. Essex: Longman.

Foma, M., and T. Abdala. 1985. Kernel oils of seven species of Zaire. *Journal of the American Oil Chemists Society* 62(5): 910-911.

Hilditch, T. P., M.L. Marea, and Y.A.H. Zaky. 1940. The fatty acids and glycerides of seed fats of *Allanblackia floribunda* and *Allanblackia parviflora*. *Journal of the Society of Chemical Industry* 59: 25.

Larochelle, F., E. Forget, A. Rainville, and J. Bousquet. 1998. Sources of temporal variation in sap sugar content in a mature sugar maple (*Acer saccharum*) plantation. *Forest Ecology and Management* 106: 307-313.

Leakey, R. R. B., and A.J. Simons. 1998. The domestication and commercialisation of indigenous trees in agroforestry for the alleviation of poverty. *Agroforestry Systems* 38: 165-176.

Leakey, R. R. B., J.-M. Fondoun, A. Atangana, and Z. Tchoundjeu. 2000. Quantitative descriptors of variation in the fruits and seeds of *Irvingia gabonensis*. *Agroforestry Systems* 50**:** 47-58.

Leakey, R. R. B., A.R. Atangana, E. Kengni, A.N. Wahuriu, C. Usoro, P.O. Anegbeh, and Z. Tchoundjeu. 2002. Domestication of *Dacryodes edulis* in West and Central Africa: characterisation of genetic variation. *Forests, Trees and Livelihoods* 12: 57-71.

Leakey, R.R.B., Z. Tchoundjeu, K. Schreckenberg, S. Shackleton, and C. Shackleton. 2005a. Agroforestry Tree Products (AFTPs): Targeting Poverty Reduction and Enhanced Livelihoods. *International Journal of Agricultural Sustainability* 16: 5-16.

Leakey, R.R.B., P. Greenwell, M.N. Hall, A.R. Atangana, C. Usoro, P.O. Anegbeh, J.-M. Fondoun, and Z. Tchoundjeu. 2005b. Domestication of *Irvingia gabonensis*: 4. Tree-to-tree variation in food-thickening properties and in fat and protein contents of dika nut. *Food Chemistry* 90(3): 365-378.

Letouzey, R. 1985. Notice de la carte phytogéographique du Cameroun au 1/500000. 5-B I. In *Bibliographie et index des noms scientifiques* (pp. 143-240). Toulouse: Institut de la carte Internationale de la Végétation.

Montgomery, D.C. 1991. *Design and analysis of experiments. 3rd edition*. New York: John Wiley and Sons.

Nanson, A. 2001. *Génétique et amélioration des arbres forestiers*. Gembloux: Les presses agronomiques de Gembloux.

Ngo Mpeck, M.L., E.K. Asaah, Z. Tchoundjeu, and A.R. Atangana, A. R. 2003. Strategies for the domestication of *Ricinodendron heudelotii*: Evaluation of variability in natural populations from Cameroon. *Food*, *Agriculture* and *Environment* 1(3&4): 257-262.

Ralls, K., and J. Ballou. 1986. Captive breeding programs for populations with a small number of founders. *Trends in Ecology and Evolution* 1(1): 19-22.

SAS. 2006. *SAS 9.1.3 Service Pack 4 for Windows*., Cary, North Carolina: SAS Institute.

Savage, T. J., B.S. Hamilton, and R. Croteau. 1996. Short-chain alkanes (tissue-specific biosynthesis of h-heptane in *Pinus jeffreyi*). *Plant Physiology* 110(1): 179-186.

Tchoundjeu, Z., E.K. Asaah, P.O. Anegbeh, A. Degrande, P. Mbile, C. Facheux, A. Tsobeng, A.R. Atangana, M.-L. Ngo Mpeck, & A.J. Simons. 2006. Putting

participatory domestication into practice in West and Central Africa. *Forests, Trees and Livelihoods* 16: 53-69.

Vivien, J., and J.J. Faure. 1996. *Fruitiers sauvages d'Afrique (espèces du Cameroun)*. Clohars Carnoet: Nguila-Kerou.

Wassel, P., and N.W.G. Young. 2007. Food applications of *trans* fatty acid substitutes. *International Journal of Food Science and Technology* 42: 503-517.

Zeinanloo, A., A. Shahsavari, A. Mohammadi, and M.R. Naghavi. 2009. Variance component and heritability of some fruit characters in olive (*Olea europa* L.). *Scientia Horticulturae* 123: 68-72.

Zobel, B., and van J.P. Buijtenen. 1989. *Wood variation: its causes and control*. Berlin Heidelberg: Springer-Verlag.

Zobel, B. and J. Talbert. 1991. *Applied Forest Tree Improvement*. Illinois: Waveland Press.

Chapter 3: Microsatellite DNA markers for *Allanblackia floribunda* Oliv. (Clusiaceae) and cross-species amplification in *Allanblackia gabonensis* and *A. stanerana*

This chapter was submitted for publication to Molecular Ecology Resources with the title **Microsatellite DNA markers for *Allanblackia floribunda* Oliv. (Clusiaceae) and cross-species amplification in *Allanblackia gabonensis* and *A. stanerana*** by Alain R. Atangana, André Gagné, Craig H. Newton, Jean Beaulieu and Damase P. Khasa

3.1 Résumé

Dix loci de type microsatellite ont été isolés et caractérisés pour l'espèce forestière tropicale *Allanblackia floribunda*. A partir de 194 clones séquencés, 62 paires d'oligonucléotides ont été conçues et évaluées à l'aide de l'ADN génomique d'arbres matures de populations naturelles du Cameroun. Neuf, huit et sept loci ont révélé des polymorphismes de 2 à 10, 2 à 11 et 2 à 8 allèles potentiels pour *A. floribunda*, *A. gabonensis* et *A. stanerana*, respectivement. Les hétérozygoties observées et attendues variaient de 0.267 à 0.900, et de 0.244 à 0.764, respectivement. Ces marqueurs isolés à partir de *A. floribunda* sont utiles pour des études de génétique des populations chez d'autres espèces du genre *Allanblackia*.

3.2 Abstract

Ten microsatellite loci were isolated and characterized for the tropical tree *Allanblackia floribunda*. From 194 sequenced clones, we designed 62 oligonucleotide pairs, which were tested on genomic DNA from mature trees from natural populations. Nine, eight and seven loci provided markers with polymorphism of 2 to 10, 2 to 11 and 2 to 8 putative alleles for *A. floribunda*, *A. gabonensis* and *A. stanerana*, respectively. Observed and expected heterozygosities ranged from 0.267 to 0.900, and from 0.244 to 0.764, respectively. These markers isolated from *A. floribunda* are valuable for population genetic studies in other species in the genus *Allanblackia*.

3.3 Introduction

Allanblackia floribunda Oliver, *Allanblackia gabonensis* (Pellegr.) P. Bamps and *Allanblackia stanerana* Exell & Mendonça (Clusiaceae) are forest tree species distributed in rainforests across Central Africa. *A. floribunda* occurs in evergreen lowland and deciduous forests, and its natural distribution range extends from Benin to the Democratic Republic of Congo and North Angola (Vivien and Faure 1996), whereas *A. gabonensis* is distributed from Cameroon to Democratic Republic of Congo, from 500 to 1750 m elevation above sea level. *A. stanerana* is found from Cameroon to Angola and the Democratic Republic of Congo in evergreen littoral forests (Vivien and Faure 1996).

Allanblackia trees are valued for their seeds, which are rich in hard fat that is used for margarine production. Consequently, *Allanblackia* seeds are purchased by agro-industrial companies, raising issues about conservation of genetic resources in this genus (Atangana *et al*. 2010). Indeed, *A. gabonensis* is included on the 2008 IUCN Red List of Threatened Species as vulnerable, with declining populations (http://www.iucnredlist.org/). We have developed microsatellite DNA markers for *A. floribunda* to investigate gene flow and population structure in *A. floribunda* and have examined the transferability of these markers to *A. gabonensis* and *A. stanerana*.

3.4 Materials and methods

Twigs were collected from thirty each of *A. floribunda* and *A. gabonensis* and thirteen *A. stanerana* wild trees in Lomié (03^{o} 13'N/13^{o} 35'E), Sangmelima (02^{o} 55'N/11^{o} 59'E) and Mouanko (03^{o} 15'N/09^{o} 57'E), respectively, in Cameroon, and DNA was extracted from 100 milligrams of frozen tissue using the miniplant DNA kit (QIAGEN). An

enriched genomic library was constructed (Khasa *et al*. 2000) using DNA extracted from a single individual of *A*. *floribunda*. Approximately 0.5-1.0 µg of total genomic DNA were digested with *Alu*I, *Hae*III, *Rsa*I and *Psb*AI restriction enzymes (New England Biolabs) in the presence of T4 DNA ligase (New England Biolabs), and equimolar mixture of the oligonucleotides M28 (5'-CTCTTGCTTGAATTCGGACTA-3') and M29 (5'-pTAGTCCGAATTCAAGCAAGAGCACA-3') linkers. The above mixture was then denatured and hybridised with biotinylated oligonucleotides AC_{12} and TC_{12} at 60 ^{o}C and biotinylated $GATA_6$ at 50 ^{o}C, then immobilized onto dynabeads M-270 streptavidin magnetic beads (Dynal, Invitrogen). These products were washed extensively at hybridization and room temperature, respectively, and amplified with M28 forward and reverse primers. The amplified mixture was then purified before digestion with *Eco*RI, and the digested reaction was purified and ligated into the plasmid vector pGEM3Z+ (Promega). Ligated DNAs were transformed into *E*. *coli* and selected on ampicillin agar plates, and screened by colony hybridization with ^{32}P-labelled $(AC/TC)_n$ or $(GATA)_n$ to identify microsatellite containing clones. Templates were amplified using universal M13 forward and reverse primers, and purified templates of positive colonies were sequenced using the 96-capillary 3730xl DNA Analyzer (Applied Biosystems).

Of 194 sequenced clones, 62 revealed microsatellite loci and sufficient flanking regions for primer design. Primers were designed using PRIMER 3 output program (Rozen and Skaletsky 2000) and were synthesized by Invitrogen, with an M13 tail (Oetting *et al*. 1995) at the 5' end of the forward primers. Primers were tested on four genomic *A*. *floribunda* DNA samples. These DNAs were amplified by Polymerase Chain Reaction (PCR) on a 96-well PTC Peltier thermal cycler (MJ Research) in a 10-µl reaction volume, with 25 ng DNA, 1 x PCR buffer (Invitrogen), 1.5 mM $MgCl_2$, 2.8 nmol dNTPs (Invitrogen), 0.2µl (5 mM) each forward, reverse and IRDye fluorescent-labelled M13 primer (700 or 800 nm), and 0.75 U of Altaenzymes Taq DNA polymerase. The PCR protocol consisted of an initial denaturation step for 3 min at 94 ^{o}C followed by 35 cycles of denaturation at 94 ^{o}C for 1 min, an annealing temperature for 1 min, an extension step at 72 ^{o}C for 1 min, and a final extension step at 72 ^{o}C for 10 min. However, two microsatellite loci were optimized using a touchdown program (Table 3.1)

as follows: 3 min of initial denaturation at 95 ^{o}C, followed by 19 cycles of denaturation at 95 ^{o}C for 1 min; annealing temperature for 30 s with a decrease of 0.5 ^{o}C per cycle; an extension step at 72 ^{o}C for 30 s, then 19 cycles of denaturation at 95 ^{o}C for 1 min; annealing temperature for 30 s; extension step at 72 ^{o}C for 30 s; and final extension at 72 ^{o}C for 10 min. The PCR products were electrophoresed in 1% agarose gel in 1 x TAE (Tris-acetate/EDTA) buffer, and staining was done in ethidium bromide solution (10 mg/ml). Of 62 pairs of primers tested, 20 amplified cleanly or were near the expected size and were tested for polymorphisms on 8% denaturing polyacrylamide gels (Genepage, Amresco) in 0.6 x TBE (Tris-borate/EDTA) buffer by electrophoresis on an IRD-DNA analyser (LI-COR). PCR products were denatured at 94 ^{o}C for 5 min, by adding an equal volume of stop buffer (95% formamide, 20 mM EDTA, 0.5 ml dd$H_2$0, 40 mg Bromophenol blue). Genotyping was performed using SAGA software. Wright's fixation index, observed and expected heterozygosities, and the number of alleles per locus were estimated using Genepop 4.0 (Raymond and Rousset 1995) software. Conformity with Hardy-Weinberg (HW) expectations and linkage disequilibrium was also tested using Genepop 4.0 and significance level adjusted by a sequential Bonferroni correction (Rice 1989). Fisher's exact tests for HW equilibrium were not performed on *A. stanerana* due to the number (13) of trees surveyed.

3.5 Results and discussion

Overall ten microsatellite loci were polymorphic (Table 1), and the number of alleles per locus ranged from 2 to 11, with seven markers showing polymorphism for the three species under study. Such transferability has been observed in *Milicia* species (Ouinsavi *et al*. 2006). Observed and expected heterozygosities are shown in Table 1. No significant gametic disequilibrium was observed for *A. floribunda* and *A. gabonensis*, respectively. Also, no HW deviation was observed for *A. floribunda* and *A. gabonensis* (Table 1). Presence of null alleles was tested using MICRO-CHECKER 2.2.3 software (Van Oosterhout *et al*. 2004) with a confidence level of 95% and there was no evidence for null alleles. Microsatellites developed in this study constitute an efficient tool to

investigate the genetic diversity in *A*. *floribunda* and other related species, and to implement strategies for the conservation of genetic resources in the genus Allanblackia.

Table 3.1. Characteristics of microsatellites for *Allanblackia floribunda*, *Allanblackia gabonensis* and *A. stanerana*

Locus	Acc. No	Repeat motif	Primer sequences (5'-3')	*Ta* (°C)	*Allanblackia floribunda* (*n*=30)						*Allanblackia gabonensis* (*n*=30)						*Allanblackia stanerana* (*n*=13)	
					Size range (bp)	Nb of alleles	H_O	H_E	*F*	H-W *P* values	Size range (bp)	Nb of alleles	H_O	H_E	*F*	H-W*P* values	Nb of alleles	H_O
*ULaflor*3	FJ429964	(GA)AA(GA)9	F : TGGAAAGTTAGGAAGGGATGAG R : TTTTCCCCTTTTTCTCCTTC	59	216-231	4	0.609	0.549	0.003^{ns}	0.873	204	1	0	0	-	-	1	0
*ULaflor*5	FJ429971	(GA)12(GAA)11	F : TAGTTAGATAAAACAAGAAATGGTTGG R : GAGGACAGCTTCACTTCCCTTT	62	162-185	10	0.633	0.7085	0.108^{ns}	0.160	163-205	11	0.49	0.72	0.30^{ns}	0.94	8	0.3
*ULaflor*57	FJ429967	(T/G/CAA)13	F : CCCTTTGACATGTTCATGGAT R : AAGGCTTGGTGGTGGTAGTG	63	144-153	4	0.833	0.671	-0.248^{ns}	0.990	144-165	7	0.90	0.74	-0.20^{ns}	0.09	5	0.92
*ULaflor*2	FJ429972	(TTG/A/C)13	F : TGGCTCTTCATTGTTTGAGG R : GGTATCCGATTTGGATGATTC	60-50^{td}	233-242	5	0.5667	0.591	0.042^{ns}	0.460	230-239	3	0.80	0.55	-0.4^{ns}	0.038	5	0.68
*ULaflor*7	FJ429965	(CAA)15	F : AGGTCGGTGCCTCTTACACT R : CCTTTGCCTCTTTGTTGCAT	61	178-201	5	0.5333	0.485	-0.103^{ns}	0.881	178-210	9	0.36	0.54	0.33^{ns}	1	3	0.45
*ULaflor*53	FJ429973	(CA)26	F : TAAGCCAAGGCAAAACCAAC R : TGGATCAAGTGAAA	61-51^{td}	102-114	5	0.5667	0.546	-0.038^{ns}	0.695	100-120	9	0.79	0.67	-0.18^{ns}	0.90	2	1

			ATCAAGGA															
*ULaflor*10	FJ42 9968	(C/GAA) 9	F : TTCGGGTTCAAACCT AGTCC R : TGTGGATCGGATTTG AGTGA	58	168-177	2	0.714	0.509	-0.414 ns	0.995	168-177	3	0.267	0.244	-0.08 ns	0.59	2	0.56
*ULaflor*58	FJ42 9966	(T/CAA)21	F : GGTCAAAAATGGTTT GGGAAC R : CACTTGTTCCCAAGT TGAGGT	61	205-211	3	0.517	0.527	-0.062 ns	0.713	205-211	3	0.483	0.610	0.24 ns	0.83	3	0.67
*ULaflor*56	FJ42 9970	(TTG/C/A)28	F : GGGACTTTGAAAAG CTTCTCTTT R : TCCAATTTGGGTGAT TCAAAA	58	214-220	3	0.607	0.507	-0.214 ns	0.961	-	-	-	-	-	-	-	-
*ULaflor*43	FJ42 9969	(GAAGG A)T(GAA GGA)GA TAGAGA A	F : TACAAATATCCAAA AGAAAATCTCC R : CATTCGGTCTAACTA GAGGAGCA	62	175	1	0	0	-	-	174-175	2	0.7	0.513	-0.33 ns	0.11	1	0

Observed (H_O) and expected (H_E) heterozygosities; annealing temperature (*Ta*) using a touchdown (^{td}) program; sampling size (*n*); Acc. No GenBank accession number; Wright's *F*; ns, non-significant (P < 0.05) departure from Hardy-Weinberg equilibrium; H-W Hardy-Weinberg; -, no amplification; /, or.

3.6 Acknowledgements

This work was jointly supported by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) through the Fonds Général pour les Etudes Supérieures (FGES), from Université Laval to Atangana, a NSERC discovery grant to Khasa, and a Canadian Wood Fibre Centre grant to Beaulieu. We also thank Sauphie Senneville (CEF, Université Laval) and Vincent Bourret (Université Laval) for their assistance during lab work. Dr. Bill Parsons (CEF) helped improve the editing of this manuscript.

3.7 Literature cited

Atangana AR, Beaulieu J, Khasa DP (2010) Wild genetic diversity preservation in a small-sized first generation breeding population of *Allanblackia floribunda* (Clusiaceae). *Tree Genetics & Genomes*, **6**:127-136.

Hilditch, T.P., M.L. Marea, and Y.A.H. Zaky. 1940. The fatty acids and glycerides of seed fats of *Allanblackia floribunda and Allanblackia parviflora. Journal of the Society of Chemical Industry* 59: 25.

Khasa, P.D., C.H. Newton, M.H. Rahman, B. Jaquish, and B.P. Dancik. 2000. Isolation, characterization, and inheritance of microsatellite loci in alpine larch and western larch. *Genome* 43: 439-448.

Nettel, A., F. Rafii, and R.S. Dodd. 2005. Characterization of microsatellite markers for the mangrove tree *Avicennia germinans* L. (Avicenniaceae). *Molecular Ecology Notes* 5: 103-105.

Oetting, W.S., H.K. Lee, D.J. Flanders, G.L. Wiesner, T.A. Sellers, and R.A. King. 1995. Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tailed primers. *Genomics* 30: 450-458.

Ouinsavi, C., N. Sokpon, J. Bousquet, C.H. Newton, and D.P. Khasa. 2006. Novel microsatellite DNA markers for the threatened African endemic tree species, *Milicia excelsa* (Moraceae), and cross-species amplification in *Milicia regia*. *Molecular Ecology Notes* 6: 480-483.

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

Rice, W.R. 1989. Analyzing tables of statistical tests. *Evolution* 43: 223-225.

Rozen, S., and H. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology*, *Edited by* S Krawetz, and S. Misener, pp. 365-386. Humana Press, Totowa, NJ.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535-538.

Vivien, J., and J.J. Faure. 1996. Fruitiers sauvages d'Afrique (espèces du Cameroun). Clohars Carnoet: Nguila-Kerou.

Chapter 4: Wild neutral genetic diversity preservation in first-cycle breeding population of *Allanblackia floribunda*

This chapter was published in Tree Genetics and Genomes with the title **Wild genetic diversity preservation in a small-sized first generation breeding population of *Allanblackia floribunda* (Clusiaceae)** by Alain R. Atangana, Jean Beaulieu and Damase P. Khasa (TGG 2010; 6: 127-136)

4.1 Résumé

La taille du groupe fondateur est de première importance dans les programmes d'amélioration des arbres. En théorie, la réduction de la taille du groupe du premier cycle d'amélioration augmenterait les gains génétiques ainsi que le taux de consanguinité, tout en retenant moins de diversité génétique naturelle. Nous avons déterminé si la sélection de 20 "arbres-plus" pour l'amélioration chez *Allanblackia floribunda* affecterait la diversité génétique neutre et le taux de consanguinité dans les populations d'amélioration et de production. A l'aide de huit marqueurs microsatellites informatifs, nous avons (i) estimé la diversité génétique nucléaire de 10 populations naturelles de *A. floribunda* de zone de forêt du Cameroun, et celle de la population d'amélioration, (ii) étudié les fluctuations temporelles de taille effective des populations naturelles de *A. floribunda*, et les effets de ces fluctuations sur la diversité génétique neutre actuelle, et (iii) testé l'hypothèse selon laquelle la diversité génétique de la population d'amélioration n'est pas différente de celle existant dans la nature. Des valeurs élevées de diversité génétique (A = 4.96; H_E = 0.59) et une différentiation modérée (θ = 0.048; R_{ST} = 0.061) ont été observées dans et entre les populations de zone de forêt naturelle. De grandes distances génétiques ont été observées entre les populations (D_C = 0.2769 ± 0.00554 en moyenne). Huit des populations sur dix ont montré des signes de déviation de l'équilibre-mutation-dérive, suggérant des goulots d'étranglement des populations lors du Pléistocène et des fluctuations de taille effective. Les tests de Mantel n'ont pas identifié de relation entre les distances génétiques et géographiques. Un dendrogramme Neighbor-Joining a révélé une structuration de populations qui pourrait être expliquée par des faits historiques. Des tests de Mann-Whitney n'ont pas identifié de différence significative entre les populations naturelles et la population d'amélioration, suggérant qu'une amélioration de *A. floribunda* à partir de la taille minimale du groupe fondateur ne réduirait pas la diversité génétique neutre. Toutefois, une légère augmentation de la consanguinité a été observée dans la population d'amélioration.

4.2 Abstract

Founder group size is of prime importance in tree breeding programs. In theory, reducing first-cycle breeding group size would increase expected genetic gains and inbreeding levels, and capture fewer wild genetic diversity. We determined whether sampling 20-plus trees for breeding in *Allanblackia floribunda*, a tropical forest tree species that has been recently enrolled in tree breeding program for fruit and seed production, would affect neutral genetic diversity and inbreeding level in both breeding and production populations. Using eight informative microsatellite loci, we: (a) assessed the nuclear genetic diversity of ten populations, and of the breeding population in the humid forest zone of Cameroon; (b) investigated temporal effective-size fluctuations in *A. floribunda* natural populations, with a view to identifying the role of past demographic events in the genetic structure of the studied species; and (c) tested the hypothesis that genetic diversity in a founder group of 20 individuals is not different from that existing in the wild. High levels of genetic diversity (A = 4.96; H_E = 0.59) and moderate differentiation (R_{ST} = 0.061) were found within and among populations in wild stands. High genetic distances existed between populations (average cord distance = 0.2769 ± 0.00554). Eight of the ten surveyed populations showed signs of deviation from mutation-drift equilibrium, suggesting Pleistocene population bottlenecks and fluctuations in effective population size. Mantel tests did not reveal any relationships between genetic and geographic distances. A neighbor-joining dendrogram showed a population structure that could be explained by historical factors. The hypothesis tested has been accepted. However, a slight increase in inbreeding was observed in the breeding population.

4.3 Introduction

Allanblackia floribunda Oliver (Clusiaceae) or tallow tree is a species distributed across rainforests of Central Africa, from Benin to the Democratic Republic of Congo and North Angola (Bamps 1969; Vivien and Faure 1996). Trees of the species produce berry-like fruits that are suspended on long pedicels, and seeds of the species are recalcitrants (germination success less than 5%, Vivien and Faure 1996). Mature A. floribunda trees reach 30 m in height, 80 cm girth with bole straight and horizontal branches (Vivien and Faure 1996). Tallow tree is valued for its seeds, which are rich in hard fat (67-73%, Foma and Abdala 1985). This fat consists mostly of stearic and oleic acid (Hilditch *et al*. 1940), which are reported to lower plasma cholesterol levels (Bonanome and Grundy 1988), thereby reducing the risks of heart attacks. Owing to this fat profile, *Allanblackia* oil is used for margarine production and in soap and ointments manufacture. The seeds extracted from *Allanblackia* fruits by local communities are traded, as a supply chain has been established in the humid tropics of Africa (http://www.allanblackia.info/). Studies on tree improvement in the species for fruit and nut production are underway, and in accordance with Soulé *et al*. (1986), 20 plus-trees have been selected from wild stands to constitute a first generation breeding population (Chapter 2; Atangana *et al*. unpublished data). This limited selection raises concerns about the maintenance of genetic diversity in both breeding and production populations.

Breeding programs have always sought to increase the frequency of desired traits in the breeding populations while maintaining sufficient genetic variation for continuous genetic gains over breeding cycles. Breeding programs also aim at preserving the whole or a part of a source's genetic diversity for unforeseen contingencies, and controlling inbreeding over multiple generations. Indeed, any loss of genetic variation in a breeding program would affect the maintenance of genetic fitness in a population (Soulé *et al*. 1986), as it reduces evolutionary potential (Frankham *et al*. 2004). Yet artificial selection, a major component of any breeding program, reduces the effective population size, and can thereby reduce genetic variation and increase inbreeding. Indeed, size bottlenecks

were found to reduce evolutionary potential in populations (Frankham *et al*. 1999). Therefore, care should be taken when establishing any breeding program, with regard to the number of individuals to be selected from wild stands for the first breeding cycle, as well as to the amount of genetic diversity displayed by the breeding population. Soulé *et al*. (1986) suggested that 90% of the genetic variation in the wild population be maintained over 200 years in captive breeding programs, with an effective founder group of at least 20 individuals, to minimize any founding effect. An assessment of genetic diversity and gene flow of wild populations of a species enrolled in a breeding program is therefore imperative for a better genetic management.

Life history and ecological traits are known to affect the amount and pattern of genetic diversity in woody species (Hamrick and Godt 1996). Past historical events, such as Pleistocene glaciations (Haffer 1969) or historical admixtures of gene pools (Born *et al*. 2008) also play an important role in shaping genetic diversity within and among populations (Walter and Epperson 2001, Hamrick *et al*. 1991). Although molecular markers are known to infer signatures of past history, few studies have examined Central African rainforest history from intraspecific population structure in tree species using molecular markers. These few studies have employed: random amplified polymorphic DNA (hereafter RAPD), as used by Dawson and Powell (1999) on *Prunus africana* and by Lowe *et al*. (2000) on *Irvingia gabonensis* and *I. wombolu* in Cameroon; chloroplast DNA, which was used by Muloko-Ntoutoume *et al*. (2000); and microsatellites, with which Born *et al*. (2008) analyzed populations of *Aukoumea klaineana* in Gabon. Indeed, Central African rainforest history has been more frequently inferred from pollen analysis (Taylor 1988; Maley 1991; Hamilton and Taylor 1991; Maley & Brenac 1998), species distribution (Hamilton 1976; Mayr and O'Hara 1986; Sosef 1994), sea sediments (Hamilton 1981), taxonomic criteria (Kingdon 1980; Jacobs 2004; Plana *et al*. 2004) and geomorphological data from satellite images (Nichol 1999), all of which provide evidence of forest refuges during Pleistocene glaciations. However, the precise location of these refuges is still controversial (Hamilton 1976; Maley 1991; Nichol 1999; Maley and Brenac 1998; Sosef 1994), as is their effect on species diversity (Plana *et al*. 2004). More data are needed to clarify on the demographic history of Central African

rainforests. As fewer geographical barriers to gene flow act on woody species than on mammals (e.g., rivers have been found to constitute genetic barriers to gorillas, Anthony *et al*. 2007), studies on tree species would be better able to detect temporal-size fluctuations effects on genetic variation in natural populations. Nuclear microsatellite DNA markers or simple sequence repeats (SSRs) are suited for such studies (Luikart *et al*. 1999). As SSR markers are codominant and widely distributed in the genome, they are an alternative to single-locus evolutionary studies for two reasons: (a) conclusions drawn from single-locus studies rely on only one realization of gene genealogy, and (b) the presence of functional linked genes in the mitochondrion (Anderson *et al*. 1982) makes it difficult to distinguish between population expansions and natural selection.

The aim of this study was three-fold: (a) to assess genetic diversity and gene flow in *A*. *floribunda* natural populations; (b) to investigate temporal size-fluctuations in *A*. *floribunda* natural populations, thereby shedding more light on the demographic history of Central African rainforests and putative forest refuges during past events, and (c) to assess the impact of plus-tree selection for breeding on genetic diversity. The hypothesis tested in this study was that genetic diversity in a breeding population of 20 individuals is not different from that existing in natural populations.

4.4 Materials and methods

4.4.1 Population sampling

A previous study aimed at characterizing phenotypic variation in fruit and nut traits in *A*. *floribunda* from wild stands allowed us to identify 20 plus-trees in the natural distribution area of the species in Cameroon (one tree in Bikondo: $2^0$57’ N $9^0$55’ E 5 m elevation above sea level (asl); seven trees in Mouanko: $3^0$05’ N 9077’E, 10 m elevation asl; one tree in Ngoumou: $3^0$37’ N 11018’ E, 650 m elevation asl; and eleven trees in Yalpenda: $3^0$43’ N $10^0$13’ E, 70 m elevation asl) that could constitute first-cycle breeding population (Chapter 2; Atangana *et al*. unpublished data). We subsequently collected

twigs from these plus-trees, and from each of 30 *A*. *floribunda* trees from ten natural populations (making a total of *n = 320*, Fig. 1) in Cameroon, respectively.

4.4.2 DNA Isolation and PCR protocols

We extracted DNA from 100 mg of frozen tissue using the miniplant DNA kit (QIAGEN Sciences, Germantown, MD, USA). Using eight microsatellite loci (*ULaflor*5, *ULaflor*57, *ULaflor*7, *ULaflor*2, *ULaflor*10, *ULaflor*53, *ULaflor*56 and *ULaflor*58, GenBank Accession Numbers FJ429964-FJ429973), we performed polymerase chain reaction (PCR) on a 96-well PTC-225 Peltier thermal cycler (MJ Research, Waltham, USA) in a 10-µl reaction volume following Atangana *et al*. (unpublished data). We denatured PCR products at 94 ^{o}C for 5 min, loaded these products on 8% denaturing polyacrylamide gels in 0.6 x TBE (Tris-borate/EDTA) buffer and electrophoresed the resulting products on an IRD-DNA analyzer (LI-COR, Lincoln, NE, USA).

4.4.3 Numerical analysis

Prior to genetic diversity estimation, we performed Fisher's exact tests using FSTAT 2.9.3.2 (Goudet 2002) to verify linkage disequilibrium between loci, with significance levels adjusted by a sequential Bonferroni correction (Rice 1989). We estimated descriptive statistics using FSTAT. These values included the mean number of alleles per locus (*A*), relatedness (*r*, Queller & Goodnight 1989) and observed and expected heterozygosities (H_O and H_E), together with fixation indices (Wright 1965) for each population and locus (Weir 1996), where the degree of inbreeding within populations (*f*), the overall inbreeding coefficient (*F*) and population differentiation (θ) correspond to Wright's F_{IS}, F_{IT}, and F_{ST}, respectively. We also used FSTAT to calculate significance and confidence intervals of fixation indices and relatedness by bootstrapping over loci (Weir 1996), together with unbiased estimates of population differentiation for microsatellites (R_{ST}, Slatkin 1995; Goodman 1997) and their significance using bootstrap and permutation tests. We based *P* values for F_{IS} on 8,000 randomizations. We based tests for population differentiation on 5,000 randomizations following the log-likelihood

goodness of fit statistics (Goudet *et al*. 1996), and adjusted the significance level of pairwise F_{ST} between populations by sequential Bonferroni correction. We derived an indirect estimate of outcrossing rate (*t*, which is generally biased (Ritland 2002)) from the equation (1-*f*) / (1+*f*). We performed Mann-Whitney tests (Zar 1984) on the mean diversity values (*A*, H_O and H_E) and on the degree of inbreeding within populations (*f*) to test our initial hypothesis of no genetic difference between the first-cycle breeding population and wild populations in *A. floribunda*.

As no study has yet reported the mutational model of microsatellites being surveyed in the present study, we used Cavalli-Sforza and Edwards' (1967) chord distance (D_C) to measure the genetic distance between populations. This metric relies on the geometric disposition of populations in a sphere delimited by allele frequencies, rather than on a given mutational model (Goldstein and Pollock 1997), and yields better tree topologies regardless of whether strong bottlenecks exist or not (Takezaki and Nei 1996). We applied neighbor-joining (NJ) analysis (Saitou and Nei 1987) to genetic distances using MEGA 4 software (Tamura *et al*. 2007) to expose evidence of any structure between populations. We calculated geographical distances between populations based on their latitude and their longitude using the software available at http://jan.ucc.nau.edu/~cvm/latlongdist.html. We performed Mantel tests using FSTAT software to correlate genetic and geographical distance between populations.

As less than ten microsatellite loci were used for population genotyping, Wilcoxon sign-ranks tests (Luikart *et al*. 1998a), which were performed using the program BOTTLENECK (Piry *et al*. 1999), were used to determine whether natural populations that we surveyed in this study exhibit a significant number of loci with heterozygosity excess. Heterozygosity excess would be indicative of a genetic bottleneck when loci evolve under an infinite allele model (IAM; Kimura and Crow 1964; Maruyama and Fuerst 1985). However, we performed the discrimination of populations that have potentially experienced a recent fluctuation of their effective size from stable populations using the test for distortion of allele frequency distributions ('mode-shift' indicator, Luikart *et al*. 1998b), available in BOTTLENECK. This test is likely to detect an allele

frequency distortion when eight polymorphic loci are used in a sample of 30 individuals (Luikart *et al*. 1998b), which was the case in the present study, whenever these loci evolve under IAM or stepwise mutation model (SMM; Ohta and Kimura 1973; Luikart *et al*. 1998b).

4.5 Results

4.5.1 Within and among population genetic diversity and fixation indices

Fisher's exact tests did not reveal significant ($P < 0.05$) evidence for linkage disequilibrium between the markers that were surveyed in this study in each population, and overall, after Bonferroni correction. Also, no deficit or excess of heterozygotes were observed within-populations (Table 4.1). Randomizing alleles within populations identified no departure from HW equilibrium over all loci ($P = 0.9994$) and for each locus ($P = 0.8360$-0.9998), respectively, except for locus *ULaflor*5 ($P = 0.0002$; Table 4.2) for F_{IS}. Randomizing alleles across populations identified no deviation from HW equilibrium over all loci ($P = 0.5848$) and for each locus ($P = 0.81860$-0.9978) except for loci *ULaflor*5 ($P = 0.0002$) and *ULaflor*2 ($P = 0.0418$; Table 4.2) for F_{IT}.

All loci surveyed were variable in each natural population. Within-population genetic diversity indices are shown in Table 4.1. The total and mean number of alleles per locus ranged from 3 to 29 and from 4.5 to 5.38, respectively. Overall observed and expected estimates of heterozygosity across loci were 0.62 and 0.59, while H_O and H_E per population ranged from 0.58 to 0.66 and from 0.54 to 0.66, respectively. Lowest and highest diversity values were observed in Ngoumou ($A = 4.5$; $H_O = 0.58$ and $H_E = 0.54$) and Sangmelima ($A = 5.25$; $H_O = 0.60$ and $H_E = 0.66$; Table 4.1), respectively. The lowest values of average outcrossing rate (t) were found for loci *ULaflor*5 and *ULaflor*2 (Table 4.2), with that of locus *ULaflor*5 lower than $t = 1$, which was indicative of the occurrence of some inbreeding. This inbreeding is shown by the high values of

relatedness ($r > 0$; $P = 0.01$; Table 4.2) found for these loci, which are however lower than 0.25 ($P = 0.01$), indicating that individuals sampled in the surveyed populations are not half siblings.

Single-locus estimates of heterozygosities ranged from 0.564 to 0.733, from 0.497 to 0.815, and from 0.519 to 0.872 for H_O, H_S and H_T, respectively. Highest estimates of single-locus heterozygosities were observed for loci *ULaflor*5 and *ULaflor*57 (Table 4.2). Permutation tests identified highly significant ($P = 0.0001$) population differentiation over all loci and for each locus ($P = 0.0001 – 0.0009$), respectively. Also, bootstrapping identified highly significant positive θ estimate over all loci (Table 4.2). Single-locus estimates of population differentiation ranged from 0.013 to 0.109 for θ (mean = 0.048), and from 0.002 to 0.197 for R_{ST}, (mean = 0.063). The average values for both estimates (Table 4.2) indicated moderate differentiation (Wright 1978) between populations. Pairwise estimates of F_{ST} between populations were found to vary from weak (0.0086; $P = 0.05$) to moderate (0.0948 $P = 0.001$); the lowest estimate was found between Lomié and Mungo populations, which were among the most geographically distant populations (471.8 km, ranked second). The average Cavalli-Sforza and Edwards' chord distances between populations was high (0.2769 ± 0.00554), with values ranging between 0.187 and 0.354, indicating substantial divergence between populations. Mantel tests did not found any relationships between genetic and geographical distances ($r = -0.16$; $P = 0.927$), indicative of no evidence of geographical structure due to isolation by distance between surveyed populations.

The topology of the NJ dendrogram (Fig 4.1) identified three distinct groups, which showed no relationship with geographic proximity, evidence, and also for a lack of geographical structure. Mungo (Mung) and Lomié (Lomi), which were among the most geographically distant sampling sites, were genetically close (Fig 4.1) and constituted one cluster whereas other groups consisted of populations that were actually in close geographic proximity to one another.

4.5.2 Temporal size-fluctuations of populations

All populations exhibited significant number of loci with heterozygosity excess under the IAM assumption, while four populations exhibited behavior that was the same under SMM (Table 4.3). Results obtained under the SMM model are congruent with that of the 'mode-shift' indicator, suggesting that microsatellite loci surveyed in this study evolve under a strict SMM model. This result also explains the observed lack of congruence between deviations from mutation-drift equilibrium and mode-shift indicators for Mouanko, Ngoumou and Sangmelima populations (Cornuet and Luikart 1996). Two populations (Lomié and Mungo) were found to be in mutation-drift equilibrium, whereas eight populations showed significant signs of departure from mutation-drift equilibrium (Table 4.3), indicative of recent bottlenecks.

4.5.3 First-cycle breeding population genetic diversity and fixation indices and comparison with source's populations

The first-cycle breeding population genetic diversity and fixation indices are shown in Table 4.1. High levels of genetic diversity were observed in the putative breeding population, and Mann-Whitney tests did not detect any significant difference between breeding and natural populations for the number of alleles ($P = 0.0557$), H_O ($P = 0.0557$), H_E ($P = 0.0565$) and f ($P = 0.0569$). However, a slight increase in the fixation index was observed in the breeding population (Table 4.1), indicative of an increase in inbreeding compared to wild populations.

4.6 Discussion

4.6.1 Genetic diversity and fixation indices

To be sustainable, any breeding program should be based on the prior assessment of genetic diversity, both at phenotypic and molecular levels. Indeed, the assessment of magnitude and pattern of genetic diversity, as well as gene flow estimation between populations and investigations of relationships between gene flow, demographic connectedness and historical partitions within species (Avise 1996), constitute a prerequisite to any management of genetic resources of a species. In the present study, we have assessed the genetic diversity of a species that has only been recently chosen for inclusion in a breeding program, with a view to identifying the role of past demographic events in the genetic structure of its populations, and thereby shedding more light on the history of rainforests of Central Africa. Expected and observed within population heterozygosity estimates found in this study were slightly greater than those that were found in a microsatellite-based study of another forest tree species from Central Africa (*A. klaineana*; Born *et al*. 2008). As expected in woody species with a broad geographic range (Hamrick *et al*. 1992), high levels of genetic diversity were found over all loci and for each locus, respectively, in the present study. It is noteworthy that the mean number of alleles per locus (A = 4.96) was low compared with the findings of Novick *et al*. (2003) and Lemes *et al*. (2003) on big-leaf mahogany (*Swietenia macrophylla* King, Meliaceae, A = 13.0-18.4), by Aldrich *et al*. (1998) on Manil (*Symphonia globulifera*, Clusiaceae; A = 13.0-16.0 for adult forest trees), and by Collevatti *et al*. (2001) on *Caryocar brasiliense* (A = 8.0), using microsatellite markers. This could be due to the fact that SSR markers used in this study are mostly made of trinucleotide repeat motifs (Chapter 2; Atangana *et al*. unpublished data), which are known to have a low mutation rate compared with dinucleotide microsatellite loci (Chakraborty *et al*. 1997). The high levels of genetic diversity found in *A. floribunda* are characteristic of tropical rainforest tree species, and fall within the range of results reported by Novick *et al*. (2003) and Lemes *et al*. (2003) for big-leaf mahogany. Our results preclude low to moderate population differentiation (Hamrick *et al*. 1992) as a consequence of high amounts of gene flow between populations (Nybom 2004). Accordingly, estimates of R_{ST} and F_{ST} in the present study were significantly moderate. Indeed, the overall R_{ST} estimate in tallow tree was found to be greater than that of other rainforest tropical tree species such as *Swietenia humilis* (R_{ST} = 0.032; White *et al*. 1999), *Carapa guianensis* (R_{ST} = 0.041;

Dayanandan *et al*. 1999), and Manil (R_{ST} = 0.056; Aldrich *et al*. 1998), indicating the existence of moderate genetic structure between *A*. *floribunda* populations we surveyed, as revealed by high genetic distances found between these populations.

Three possible reasons may explain the high within population and moderate among-population genetic diversity indices in *A*. *floribunda*. First, deforestation activities in Cameroon for various reasons might have weakly impacted tallow tree habitats. Indeed, forests are fragmented in Cameroon via urbanization and agriculture. This fragmentation is weakened by the fact that, when clearing forests for farming, farmers usually spare a few trees with high nutritive or medicinal value (the bark of *A*. *floribunda* is locally used against coughs, dysentery, diarrhea, toothache, and as an aphrodisiac and pain reliever), an action that favors gene flow. Also, high-value tree species such as tallow tree are used for shading in cocoa and coffee (main cash crops in the area). Another reason is that *A*. *floribunda* trees in Cameroon are not yet subject to logging, which is selective and focuses on high-value timber trees such as *Entandrophragma* spp., *Lovoa trichilioides*, *Lophira alata*, and *Erythrophleum ivorense* (Global Forest Watch 2000; Hall *et al*. 2003). This hypothesis holds, as populations surveyed in this study were found to be in HW equilibrium. Second, except in the Mount Cameroon area, there is no apparent physical barrier to gene flow between populations, although no study has yet reported upon the mechanisms of pollen dissemination in *A*. *floribunda*. Indeed, tree species with outcrossing breeding systems tend to harbor high within population genetic diversity and low to moderate population differentiation (Hamrick *et al*. 1992), as a consequence of high gene flow between populations. The occurrence of male and female flowers on the same *A*. *floribunda* trees (Bamps 1969) could promote outcrossing (Allen & Hiscock 2008), although some selfing may occur, as indicated by the strong relatedness found in the locus *ULaflor*5. Also, a bird-pollinated and hermaphrodite monoecious member of the Clusiaceae (i.e., Manil) was found to be a predominantly outcrossing species (Degen *et al*. 2004); outcrossing has been reported in the genus *Clusia* (Vlasakova *et al*. 2008). A fine-scale study of the reproductive biology and mating system of *A*. *floribunda* is needed to shed more light on this. It is noteworthy that *A*. *floribunda* seeds are eaten by rodents, which probably play an effective role in dissemination of tallow tree seeds. However, the

lack of apparent physical barriers to gene flow between populations in *A. floribunda* suggests that past demographic events might have been the primary factor in shaping the genetic differentiation observed in the present study. Third, *A. floribunda* seeds are just in the early stages of commercial harvesting, and current pressure on genetic resources in the species is weak, which would probably not be the case when intensive seed harvesting will begin to occur. Consequently, the levels of genetic diversity found in this study could be used as benchmark for conservation of genetic resources in the species during tree improvement.

Forest refuges during past events, as they affect the distribution of genetic variation within and between populations, must be identified for the effective preservation of species at global and local scales. The lack of genogeographic structure between populations found in this study contrasts with the findings of Lowe *et al*. (2000) on bush mango (*Irvingia gabonensis* and *Irvingia wombolu*), and might be due to the fact that these authors used dominant markers (RAPD), few individuals (130), and a wider range of sampling sites (from Nigeria to Gabon). RAPD data are hampered by a lack of complete genotypic information resulting from dominance (Lynch and Milligan 1994). Populations were found to be genetically distant in the present study, such that three clusters indicative of a genetic structure were identified from NJ analysis. The clustering of Lomié and Mungo populations, which were the second-most geographically distant populations surveyed, have allowed us to speculate that past events might have played an important role in the distribution of genetic diversity in *A. floribunda*. Interestingly, the Lomié and Mungo populations were found to be in mutation-drift equilibrium under SMM, indicating that they were not subjected to any recent effective size-fluctuation, whereas other surveyed populations showed signs of expansions. Other clusters consisted of populations that can be geographically grouped. Results from this study suggest that Lomié and Mungo are located in Central African rainforest refuges, as postulated by Sosef (1994) and Kingdon (1980), respectively. The lack of temporal effective-size stability in populations from sites postulated as rainforest refuges from previous studies (Banga-Bakundu, Mouanko, Bikondo and Yalpenda; Hamilton 1976; Maley 1991)

questioned the location of these refuges in Cameroon, and other molecular data are needed to provide more insights in the debate.

4.6.2 First-cycle breeding population genetic diversity and comparison with source's populations

The smaller the breeding population size, the higher the expected genetic gain, other parameters being similar. Therefore, breeders face the challenge of identifying a breeding population size that increases genetic gains while preserving the whole or a part of genetic diversity of the source, and controlling inbreeding levels in subsequent generations. The lack of any significant difference between putative breeding and natural populations indicates that breeding *A. floribunda* from 20 trees would not reduce the nuclear neutral genetic diversity. Soulé *et al*. (1986) suggested that an effective founding group size of at least 20 individuals is necessary for the maintenance of 90% of wild genetic variation in a species over 200 years in a capture breeding program; results from the present study indicate that a founder group size of 20 individuals, which reflects an effective size of less than 20 individuals as $Ne < N$, might preserve wild genetic diversity in the breeding populations. However, data on other species are needed to confirm or refute this assumption. The slight increase in the *f*-estimate observed in the breeding population indicates that breeding a tree species from 20 individuals would increase the inbreeding level in breeding populations. Therefore, appropriate measures should be taken to reduce inbreeding levels in breeding populations throughout tree improvement in such small-sized first generation breeding groups. These measures include an increase in the breeding population size from a more widely sampled base population. As plus-trees are of unknown pedigrees, progeny trials need to be established, and neutral genetic variation assessed to ascertain if there has been preservation of genetic diversity in the breeding populations. Description of the mating system, along with assessment of gene flow between breeding and natural populations, would probably provide more insights into the strategy of genetic resource preservation in the species of interest during tree improvement. In order to obtain hybrids from trees from different groups to constitute the next cycle breeding population or production population, the application of a subgroup

breeding strategy consisting in using many breeding populations also would help in controlling inbreeding levels in breeding populations throughout tree improvement. This strategy should be applied, however, with an increase of the first-cycle breeding population size from a wider sampled base population.

4.7 Acknowledgments

Financial support was jointly provided by the Natural Sciences and Engineering Research Council of Canada (NSERC) to A.R.A. and D.P.K (Discovery grant), and by the Canadian Wood Fibre Centre to J.B. The World Agroforestry Centre assisted in leaf sample collection in Cameroon. We thank André Gagné, Sauphie Senneville (CEF) and Vincent Bourret (Université Laval) for their technical assistance during lab work. Dr. Bill Parsons (CEF) helped improve this manuscript.

4.8 Literature cited

Aldrich, P.R., J.L. Hamrick, P. Charriaga, and G. Kochert. 1998. Microsatellite analysis of demographic genetic structure in fragmented populations of the tropical tree *Symphonia globulifera*. *Molecular Ecology*, 7: 933-944.

Allen, A.M., and S.J. Hiscock. 2008. Evolution and phylogeny of self-incompatibility systems in angiosperms. In Franklin-Tong VE (ed) *Self-Incompatibility in Flowering Plants – Evolution, Diversity, and mechanism*. Springer-Verlag Berlin Heideberg pp. 73-101 (chap 4).

Anderson, S., M.H.L. Debruijn, A.R. Coulson, I.C. Eperon, F. Sanger, and I.G. Young. 1982. Complete sequence of bovine mitochondrial DNA conserved features of the mammalian mitochondrial genome. *Journal of Molecular Biology* 156: 683-717.

Anthony, N.M., M. Johnson-Bawe, K. Jeffery, S.L. Clifford, K.A. Abernethy, C.E. Tutin, S.A. Lahm, L.J.T. White, J.F. Utley, E.J. Wickings, and M.W. Bruford. 2007. The role of Pleistocene refugia and rivers in shaping gorilla genetic diversity in Central Africa. *PNAS* 104 (51): 20432-20436.

Avise, J.C. 1996. The scope of conservation genetics. In: Conservation genetics: case histories from nature (eds Avise JC, Hamrick JL), pp1-9. Chapman & Hall, New York, NY.

Bamps P .1969. Notes sur les Guttiferae d'Afrique tropicale. *Bulletin du Jardin botanique national de Belgique* 39, 345-372.

Bonanome, A., and S.M. Grundy. 1988. Effect of dietary stearic acid on plasma cholesterol and lipoprotein levels. *New England Journal of Medicine* 318(19): 1244-1248.

Born, C., O.J. Hardy, M.H. Chevallier, S. Ossari, C. Attéké, J. Wickings, and M. Hossaert-McKey. 2008. Small-scale spatial genetic structure in the Central African rainforest tree species *Aucoumea klaineana*: a stepwise approach to infer the impact of limited gene dispersal, population history and habitat fragmentation. *Molecular Ecology* 17: 2041-2050.

Cavalli-Sforza, L.L., and A.W.F. Edwards. 1967. Phylogenetic analysis models and estimation procedures. *American Journal of Human Genetics* 19**:** 233-257.

Chakraborty, R., M. Kimmel, D.N. Stivers, L.J. Davison, and R. Deka. 1997. Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. *PNAS* 94: 1041-1046.

Collevatti, R.G., D. Grattapaglia, and J.D. Hay. 2001. Population genetic structure of the endangered tropical tree species *Caryocar braziliense*, based on variability at microsatellite loci. *Molecular Ecology* 10: 349-356.

Cornuet, J.M., and G. Luikart. 1997. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* 144: 2001-2014.

Dayanandan, S., J. Dole, K. Bawa, and R. Kesseli. 1999. Population structure delineated with microsatellite markers in fragmented populations of tropical tree, *Carapa guianensis* (Meliaceae). *Molecular Ecology* 8: 1585-1592.

Dawson, I.K., and W. Powell. 1999. Genetic variation in the Afromontane tree *Prunus Africana*, an endangered medicinal species. *Molecular Ecology* 8: 151-156.

Degen, B., E. Bandou, and H. Caron. 2004. Limited pollen dispersal and biparental inbreeding in *Symphonia globulifera* in French Guiana. *Heredity* 93**:** 585-591.

Foma, M., and T. Abdala. 1985. Kernel oils of seven species of Zaire. *Journal of the American Oil Chemists Society* 62(5): 910-911.

Frankham, R., K. Lees, M.E. Montgomery, P.R. England, E.H. Lowe, and D.A. Briscoe. 1999. Do population size bottlenecks reduce evolutionary potential? *Animal Conservation* 2**:** 255-260.

Frankham, R., J.D. Ballou, and D.A. Briscoe. 2004. *A primer of conservation genetics*. Cambridge University Press, Cambridge, UK. 220 p

Global Forest Watch. 2000. An overview of logging in Cameroon. http://www.globalforestwatch.org/common/cameroon/ last accessed 02 December 2008.

Goldstein, D.B., and D.D. Pollock. 1997. Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. *Journal of Heredity* 88: 335-342.

Goodman, S.J. 1997. R_{ST} CALC: a collection of computer programs for calculating estimates of genetic differentiation from microsatellite data and determining their significance. *Molecular Ecology* 6: 881-885.

Goudet, J., M. Raymond, T. de Meeus, and F. Rousset. 1996. Testing differentiation in diploid populations. *Genetics* 144: 1933-1940.

Goudet, J. 2002. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3.2). Available at http://www.unil.ch/izea/softwares/fstat.html

Haffer, J. 1969. Speciation in Amazon forest birds. *Science* 1965: 131-137.

Hall, J.S., D.J. Harris, V. Medjibe, and P.M.S. Ashton. 2003. The effects of selective logging on forest structure and tree species composition in a Central African forest: implications for management of conservation areas. *Forest Ecology and Management* 183 (1&3): 249-264. doi: 10.1016/S0378-1127(03)00107-5.

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 palaecoenvironments: A review, *Palaeoecol*. *Afr*. 18: 283-304.

Hamilton, A.C. 1981. The quaternary history of African forests: its relevance to conservation. *African Journal of Ecology* 19: 1-6.

Hamilton, A.C., and D. Taylor. 1991. History of climate change and forests in tropical Africa during the last 8 millions years. *Climatic Change* 19**:** 65-78.

Hamrick, J.L., M.J.W. Godt, and S.L. Sherman-Broyles. 1992. Factors influencing levels of genetic diversity in woody plant species. *New Forests* 6: 95-124.

Hamrick, J.L., and M.J.W. Godt. 1996. Effects of life history on genetic diversity in plant species. *Philosophical Transactions of the Royal Society of London*. *Series B, Biological sciences* 351: 1291-1298.

Hilditch, T.P., M.L. Marea, and Y.A.H. Zaky. 1940. The fatty acids and glycerides of seed fats of *Allanblackia floribunda* and *Allanblackia parviflora*. *Journal of the Society of Chemical Industry* 59: 25.

Jacobs, B.F. 2004. Palaeobotanical studies from tropical Africa: relevance to the evolution of forest, woodland and savannah biomes. *Phil*. *Trans*. *R*. *Soc*. *Lond*. B 359: 1573-1583.

Kimura, M., and J.F. Crow. 1964. The number of alleles that can be maintained in a finite population. *Genetics* 49**:** 725-738.

Kingdon, J.S. 1980. The role of visual signals and face patterns in African forest monkeys (Guenons) of the genus *Cercopithecus*. *Transactions of the Zoological Society of London* 35: 425-475.

Lemes, M.R., R. Gribel, J. Proctor, and D. Grattapaglia. 2003. Population genetic structure of mahogany (*Swietenia macrophylla* King, Meliaceae) across the Brazilian Amazon, based on variation at microsatellite loci: implications for conservation. *Molecular Ecology* 12**:** 2875-2883.

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

Luikart, G., W.B. Sherwin, B.M. Steele, and F.W. Allendorf. 1998*a*. Usefulness of molecular markers for detecting population bottlenecks via monitoring genetic change. *Molecular Ecology* 7: 963-974.

Luikart, G., F.W. Allendorf, J.-M. Cornuet,, and W.B. Sherwin. 1998*b*. Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Genetics* **89:** 238-247.

Luikart, G., J.-M. Cornuet, and F.W. Allendorf. 1999. Temporal changes in allele's frequencies provides estimates of population bottleneck size. *Conservation Biology*, 13(3): 523-530.

Lynch, M., and B.G. Milligan. 1994. Analysis of population genetic structure with RAPD markers. *Molecular Ecology* 3: 91-99.

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

Maley, J., and P. Brenac. 1998. Vegetation dynamics, palaeoenvironments and climatic changes in the forests of western Cameroon during the last 28,000 years B.P. *Review of Palaeobotany and Palynology* 99: 157-187.

Maruyama, T., and P.A. Fuerst. 1985. Population bottlenecks and non equilibrium models in population genetics. II. Number of alleles in a small population that was formed by a recent bottleneck. *Genetics* 111: 675-689.

Mayr, E., and R.J. O'Hara. 1986. The biogeographic evidence supporting the Pleistocene forest refuge hypothesis. *Evolution* 40(1): 55-67.

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

Nei, M. 1078. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*. 89: 583-590.

Nichol, J.E. 1999. Geomorphological evidence and Pleistocene refugia in Africa. *The Geographical Journal* 165 (1): 79-89.

Novick, R.R., C.W. Dick, M.R. Lemes, C. Navarro, A. Caccone, and E. Bermingham. 2003. Genetic structure of Mesoamerican populations of Big-leaf mahogany (*Swietenia macrophylla*) inferred from microsatellite analysis. *Molecular Ecology* 12: 2885-2893.

Nybom, H. 2004. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Molecular Ecology* 13: 1143-1155.

Ohta, T., and M. Kimura. 1973. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genet. Res. Camb.* 22: 201-204.

Piry, S., G. Luikart, and J.-M. Cornuet. 1999. BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. *The Journal of Heredity* 90 (4): 502-503.

Plana, V., A. Gascoigne, L.L. Forrest, D. Harris, and R.T. Pennington. 2003. Pleistocene and pre-Pleistocene *Begonia* speciation in Africa. *Molecular Phylogenetics and Evolution* 31: 449-461.

Queller, D.C., and K.F. Goodnight. 1989. Estimating relatedness using genetic markers. *Evolution* 43 (2): 258-275.

Rice, W.R. 1989. Analyzing tables of statistical tests. *Evolution* 43: 223-225.

Ritland, K. 2002. Extension models for the estimation of mating systems using *n* independent loci. *Heredity* 88**:** 221-228.

Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4: 406-425.

Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139: 457-462.

Sosef, M.S.M. 1994 Refuge Begonias-Taxonomy, phylogeny and historical biogeography of *Begonia* sect. *Loasibegonia* and sect. *Scutobegonia* in relation to glacial rain forest refuges in Africa. (Studies in the Begoniaceae V.) *Belmontia* 26**:** 1-148.

Soulé, M., M. Gilpin, W. Conway, and T. Foose. 1986. The millennium ark: How long a voyage, how many staterooms, how many passengers? *Zoo Biology* 5: 101-113.

Takezaki, N., and M. Nei. 1996. Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* 144**:** 389-399.

Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA 4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* 24: 1596-1599.

Taylor, D.M. 1988. *The environmental history of the Rukiga highlands, South-West Uganda, during the last 40 000 to 50 000 years*, unpubl. D. Phil. Thesis, Univ. Ulster, UK.

Vlasakova, B., B. Kalinova, M.H.G. Gustafsson, and H. Teichert. 2008. Cockroaches as pollinators of *Clusia* aff. *Sellowiana* (Clusiaceae) on Inselbergs in French Guiana. *Annals of Botany* 102: 295-304.

Walter, R., and B.K. Epperson. 2001. Geographic pattern of genetic variation in *Pinus resinosa*: area of greatest diversity is not the origin of postglacial populations. *Molecular Ecology* 10: 103-111.

Weir, B.S. 1996. *Genetic Data Analysis II*. Sinauer Associates, Sunderland, Massachusetts.445 p

White, G., D.H. Boshier, and W. Powell. 1999. Genetic variation within a fragmented population of *Swietenia humilis* Zucc. *Molecular Ecology* **8**: 1899-1909.

Wright, S. 1965. The interpretation of population structure by *F*-statistics with special regard to systems of mating. *Evolution* 19: 395-420.

Wright, S. 1978. *Evolution and the genetics of populations. Vol 4: variability within and among natural populations*. University of Chicago Press, Chicago. 580 p.

Zar, J.H. 1984. *Biostatistical Analysis*, 2nd edition. Prentice Hall, Englewood Cliffs, New Jersey. 718 p

Table 4.1. Geographical locations and genetic diversity parameters estimated in 10 *Allanblackia floribunda* populations and one putative breeding population in the species.

Population	Code	Coordinates*	*n*	*A*	H_O	H_E	*F*
Mouanko	Moua	03° 15'/09° 57'	29.75	5.25	0.58	0.62	0.064 ns
Bikondo	Bikon	0257'/09° 55'	30	4.63	0.66	0.58	-0.140 ns
Banga-Bakundu	Bang	04° 18' 30''/09° 23' 42''	29.63	5.13	0.66	0.59	-0.115 ns
Ngoazick	Ngoa	01° 50'/11° 20'	30	5.25	0.60	0.66	0.085 ns
Ngoumou	Ngou	03° 37'/11° 18'	29.88	4.50	0.58	0.54	-0.081 ns
Akonolinga	Akon	03° 47'/12° 14'	29.38	4.75	0.59	0.56	-0.060ns
Lomie	Lomi	03° 13' 27''/13° 35' 15''	29.38	4.63	0.63	0.57	-0.107 ns
Mungo	Mung	04° 07' 48''/09° 26' 01''	29.75	5.00	0.64	0.59	-0.099 ns
Sangmelima	Sang	02° 55'/11° 59'	29.75	5.13	0.65	0.63	-0.026 ns
Yalpenda	Yalp	03° 43'/10° 13'	29.66	5.38	0.64	0.59	-0.086 ns
Mean			29.71	4.96	0.62	0.59	-0.056
Standard error			0.070	0.10	0.01	0.01	0.024
Breeding population		Mixed sites[1]	20	4.38	0.68	0.75	0.099

Abbreviations: *n*, mean sample size per locus; *A*, mean number of alleles per locus; H_O, observed heterozygosity and H_E, unbiased expected heterozygosity (Nei 1978); *f*, fixation index (Wright 1965). *Coordinates: latitude (°N)/longitude (°E); ns, non-significant (P< 0.05) using Fisher's exact tests; Mixed sites[1], 11 trees are located in Yalpenda, 7 in Mouanko, 1 in Bikondo and 1 in Ngoumou.

Table 4.2. Estimates of single-locus heterozygosity, genetic differentiation, and indirect estimates of outcrossing rates for 10 *Allanblackia floribunda* populations from Cameroon.

Locus	H_O	H_S	H_T	f	F	θ	R_{ST}	t	r
*ULaflor*5	0.597	0.815	0.872	0.268***	0.321***	0.072	0.002	0.577	0.109
*ULaflor*57	0.733	0.653	0.661	-0.123 ^{ns}	-0.107 ^{ns}	0.014	0.026	1.28	0.031
*ULaflor*2	0.593	0.567	0.630	-0.045 ^{ns}	0.069*	0.109	0.066	1.094	0.205
*ULaflor*7	0.597	0.507	0.526	-0.178 ^{ns}	-0.130^{ns}	0.041	0.033	1.433	0.094
*ULaflor*53	0.680	0.638	0.657	-0.067 ^{ns}	-0.032 ^{ns}	0.032	0.037	1.144	0.067
*ULaflor*10	0.616	0.497	0.526	-0.239^{ns}	-0.161 ^{ns}	0.063	0.057	1.628	0.151
*ULaflor*58	0.599	0.542	0.548	-0.106^{ns}	-0.092 ^{ns}	0.013	0.197	1.237	0.029
*ULaflor*56	0.564	0.509	0.519	-0.105^{ns}	-0.081^{ns}	0.022	0.070	1.235	0.047
Over all loci	0.622	0.591	0.617	-0.053^{ns}	-0.003 ^{ns}	0.048	0.0613	1.112	0.096
Confidence interval $^{\beta}$				(-0.171; 0. 114)	(-0.22; 0.166)	(0.021; 0.078)			(0.046; 0.151)

Abbreviations: H_O = mean observed heterozygosity within populations; H_S, mean expected heterozygosity within populations; H_T, mean expected heterozygosity in the total population; *f*, fixation index within populations; ns, non-significant; *, ***, significant ($P < 0.05$) and very highly significant ($P < 0.0001$), respectively; *F*, fixation index in the total population; θ, differentiation among populations (according to Weir 1996); R_{ST}, unbiased estimator (Slatkin 1995), calculated from Goodman (1997); *t*, outcrossing rate (derived from the *f*-values); *r*, relatedness following Queller and Goodnight (1989); $^{\beta}$ 99% confidence interval estimated bootstrapping over loci.

Table 4.3. One-tailed probabilities associated with Wilcoxon sign-rank tests used to detect deviations from expected levels of heterozygosities under the assumptions of Infinite Allele Model (IAM) or Stepwise Mutation Model (SMM), and mode-shift indicators of recently bottlenecked (shifted mode) or stable (normal distribution) populations.

Population	Code	Model		Mode-shift indicator
		IAM	SMM	
Mouanko	Moua	0.00586**	0.27344^{ns}	Shifted
Bikondo	Bikon	0.00391**	0.00391**	Shifted
Banga-Bakundu	Bang	0.00586**	0.02734*	Shifted
Ngoazick	Ngoa	0.00391**	0.01953*	Shifted
Ngoumou	Ngou	0.00977**	0.09766^{ns}	Shifted
Akonolinga	Akon	0.01367*	0.09766^{ns}	Shifted
Lomié	Lomi	0.00586**	0.15625^{ns}	Normal distribution
Mungo	Mung	0.01367*	0.19141^{ns}	Normal distribution
Sangmelima	Sang	0.00391**	0.09766^{ns}	Shifted
Yalpenda	Yalp	0.00586**	0.01953*	Shifted

*, **, significant ($P < 0.05$) and highly significant ($P < 0.001$); ns, non-significant

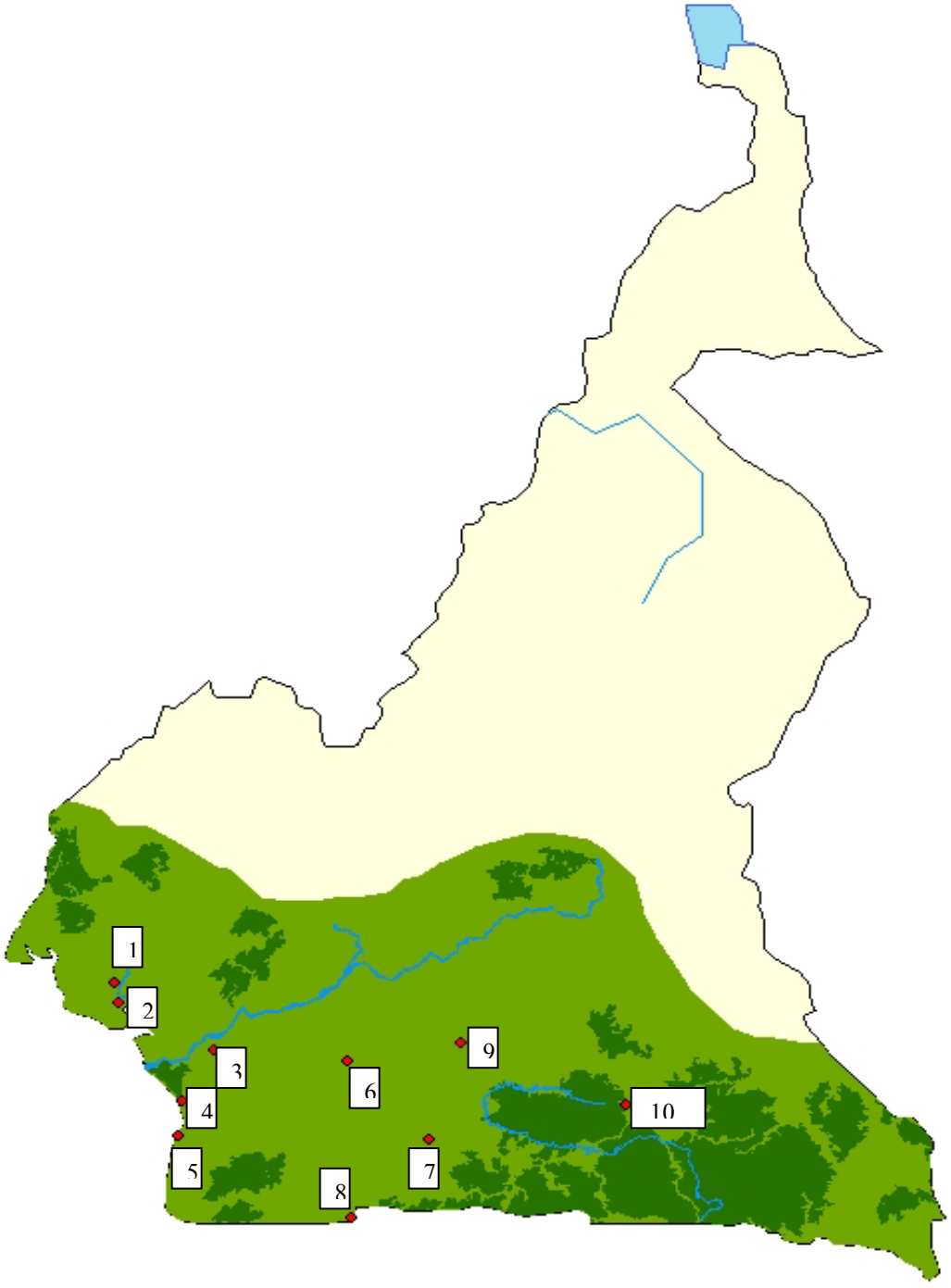

Figure 4.1. Sampling sites of *Allanblackia floribunda* populations (1: Banga-Bakundu; 2: Mungo; 3: Yalpenda; 4: Mouanko; 5: Bikondo; 6: Ngoumou; 7: Sangmelima; 8: Ngoazick; 9: Akonolinga; 10: Lomié) used in the present study.

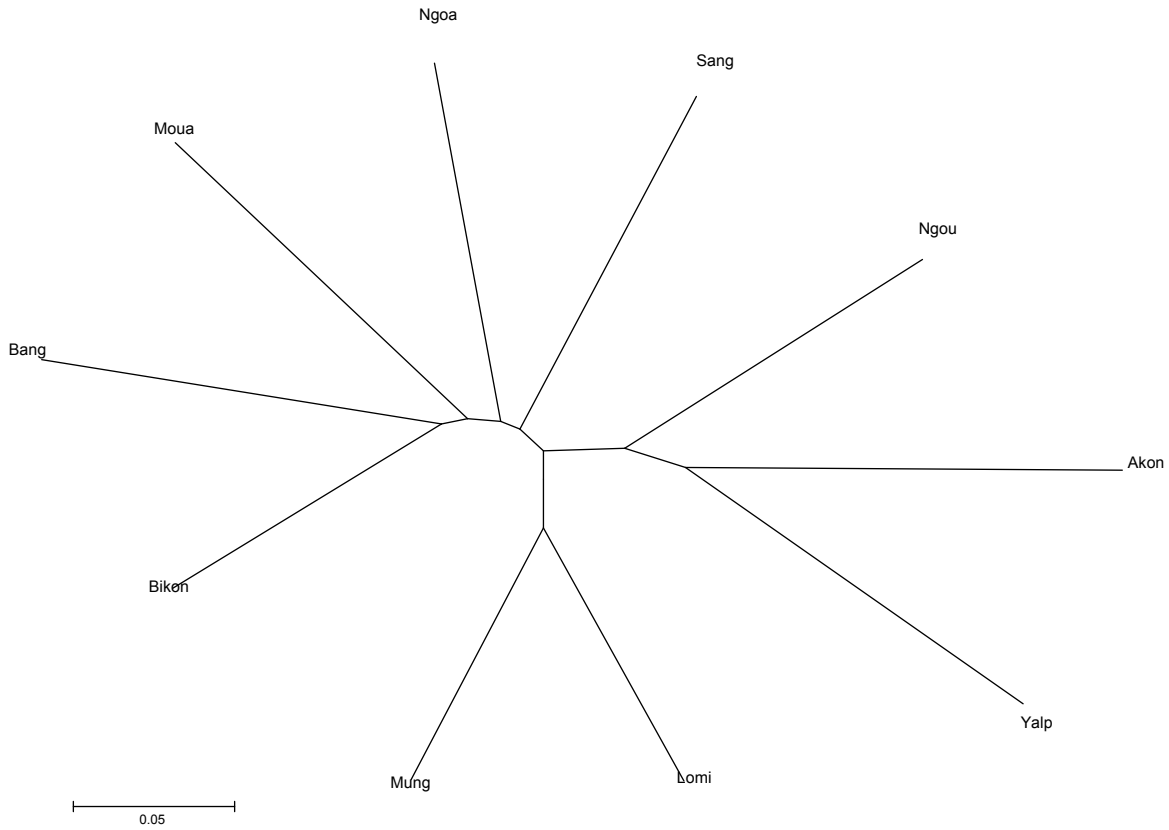

Figure 4.2. Neighbor-Joining dendrogram of pairwise chord distances (Cavalli-Sforza and Edwards 1967) between 10 *Allanblackia floribunda* populations from wild stands in Cameroon (Akon: Akonolinga; Bang: Banga-Bakundu; Bikon: Bikondo; Lomi: Lomié; Mung: Mungo; Ngoua: Ngoazick; Ngou: Ngoumou; Sang: Sangmelima; Yalp: Yalpenda).

Chapter 5: General conclusions

Forest tree breeding relies heavily on the amount and pattern of genetic variation in the characters of interest for improvement. Therefore, characterization of genetic variation in the traits of interest is of prime importance when planning for a tree breeding program. Also important are measures taken to minimize inbreeding level in breeding and production populations, and to preserve neutral and functional source's genetic diversity throughout tree improvement. The present study is innovative as quantitative and population genetic methods have been applied to provide necessary information for tree breeding for fruit / seed products, and for genetic resources preservation in the study species. The hypothesis tested in this study has been accepted, and some achievements made with regard to tree improvement in *A. floribunda*.

5.1 Phenotypic variation in fruit and seeds and tree selection

The first step in assessing phenotypic variation in fruit and seed traits consists in identifying quantitative descriptors of variation in these traits (Leakey *et al*. 2000). We described in chapter 2 quantitative methods for the assessment of fatty acid profiles in *A. floribunda* seed fat, and characterized the phenotypic variation in fruit and seed traits in the species. Indeed, substantial phenotypic variation exists in fruit and seed traits in *A. floribunda* in wild stands, allowing plus-tree selection for breeding. However, we did not quantify the amount of fatty acids yielded per tree, as well as repeatability estimates of fatty acid profiles, which are useful for tree selection in *A. floribunda* for fruit / seed production. Next step towards *A. floribunda* tree improvement consists in using methods developed in this study to assess phenotypic variation in fruit / seed traits in the whole distribution range of the species, together with between-species phenotypic variation in relevant traits for tree improvement in the genus *Allanblackia*. This survey should be implemented so as to assess repeatability of stearic and oleic acid contents in *A. floribunda* seed fat, as well as the amount of seed fat yielded per tree. Also, progeny tests are required to determine

heritability and associated genetic gains, breeding values in traits of interest in tallow-tree and genetic correlation of these traits as well, what would help implement strategies for tree selection and seedling production in orchards in the study species. Indeed, plus-trees identified in this study are of unknown pedigree, and no information is available with regard to their specific or general combining ability in fruit and seed traits of interest for improvement. Knowledge of mating system and of floral biology of *A. floribunda* is also required prior to estimation of specific and general combining ability in selected trees, which are required to better plan a breeding strategy in the species. Microsatellite markers developed in this study constitute valuable tools for the study of mating system in tallow-tree. Also, there is a need to breakdown tallow-tree seed dormancy, as the germination rate of *A. floribunda* seeds is very low (less than 5%, Vivien and Faure 1996), prior to the establishment of progeny tests. Further work is required to improve the rooting percentage in *A. floribunda* cuttings (Atangana *et al*. 2006; Atangana and Khasa 2008), in case clonal tests are to be carried out to estimate broad-sense heritability in fruit / seed traits, and to capture relevant characteristics for tree improvement in the species. No information is available pertaining to interspecific hybridization in the genus *Allanblackia* yet. Therefore, delimitation of *Allanblackia* species using molecular data and identification of putative interspecific hybrids in this genus is a prerequisite to *A. floribunda* tree breeding.

5.2 Genetic diversity preservation

Conservation of source's genetic diversity is of prime importance throughout breeding, as loss of genetic diversity in populations reduces their ability to adapt to changes in the environment, and leads to inbreeding depression. Neutral co-dominant molecular markers constitute valuable tools to assess genetic diversity and gene flow in populations of a given a species, and the inbreeding level in these populations. In chapter 3, we have developed and characterized from *A. floribunda* ten polymorphic microsatellite loci valuable for genetic population studies in tallow-tree. The first difficulty in this study arose from DNA extraction from leaf material, and we used twigs of leaves to that end. Further work is

required to fine-tune DNA extraction protocol from *Allanblackia* species, and to identify tissues yielding highest amounts of DNA of good quality. The second difficulty in developing microsatellites in the study species consisted in identifying the optimal conditions for each polymorphic primer pair. Indeed, *A. floribunda* oligonucleotide pairs are very sensitive to any slight temperature variation. Also, most microsatellite loci identified in this study are made of trinucleotide repeats, thus less polymorphic than dinucleotide repeats made loci. However, the fact that seven out of ten microsatellite loci developed in this study were polymorphic for the three species surveyed suggests a low level of mutation rate at neutral level between species in the genus *Allanblackia*. This high level of cross-amplification in microsatellite loci could facilitate resolution of taxonomic uncertainties in the genus *Allanblackia* using molecular data. However, these loci must first be assessed for polymorphism in the other species in the genus *Allanblackia*. Cross-amplification should also be assessed at functional level in *Allanblackia* species to address issues related to the evolution of genome in *Allanblackia* with regard to relevant traits for breeding. Indeed, within- and between-species variation in genes reported to play important role in fatty acid content in seeds in other species (such as *FAD2* in *Sesamum indicum*, Jung Kim *et al*. 2006; the homeobox gene *GLABRA2* in *Arabidopsis*, Shen *et al*. 2006) should be characterized in the genus *Allanblackia*. However, a screening of these genes should first be carried out in the genus of interest, with a view to identifying the genes shared by phyla, the genes that are unique to a phylum, and whether gene variation within- and between *Allanblackia* species is related to ecology.

The genetic structure of *A. floribunda* natural populations from Cameroon is described and compared with that of first-cycle breeding population in chapter 4. Techniques to reduce the risk of genetic diversity loss during tree improvement in the species are suggested in the same chapter, as a slight increase in inbreeding was found in the first-cycle breeding population. Using microsatellite variation, we have also explored evolutionary processes and demographic events in the past, thus shedding more light on the location of Central African forest refuges during Pleistocene glaciations. Patterns of DNA sequence divergence across rainforests of Cameroon observed in this study need to be confirmed in

other woody species to better locate Pleistocene forest refuges in Central Africa. However, high levels of genetic diversity found in this study should be maintained in *A. floribunda* populations throughout tree improvement in the species. To do so, forest fragmentation needs to be minimized in the course of tree improvement in *A. floribunda*. This is possible as tree domestication approach in the Humid Tropics of Africa is based on the integration of elite trees in existing land-use systems using agroforestry techniques. This enrichment-planting method would preserve tallow-tree's habitat, thus maintaining over generations neutral genetic diversity in the wild species. However, gene flow between trees from improved propagules and natural populations needs to be assessed, and investigations on potentially occurring hybridizations in sympatric zones carried out. To do so, knowledge of the reproductive biology of *Allanblackia* is imperative.

Indigenous tree domestication as implemented in the Humid Tropics of Africa involves tree selection and vegetative propagation using inexpensive, robust and simple technology (Leakey *et al*. 1990). Indeed, tree domestication utilizes the variation within species, by selecting trees with desirable traits (Atangana *et al*. 2002) and propagating them asexually, as a clonal approach aimed at cultivar development has been used (Tchoundjeu *et al*. 2006). Tree breeding has been overlooked, and genetic resources preservation limited to gene banks for bush mango. The present study set up the basis of tree breeding for fruit / nut production in tallow-tree, and suggested strategies for genetic resource preservation during tree improvement. Methods developed and tested in this study could be applied in other woody species under domestication elsewhere.

5.3 Literature cited

Atangana, A.R., V. Ukafor, P. Anegbeh, E. Asaah, Z. Tchoundjeu, J-M. Fondoun, M. Ndoumbe and R.R.B. Leakey. 2002. Domestication of *Irvingia gabonensis*: 2. The selection of multiple traits for potential cultivars from Cameroon and Nigeria. *Agroforestry Systems*: 55 (3): 225-229.

Atangana, A.R., Z. Tchoundjeu, E.K. Asaah, A.J. Simons, and D.P. Khasa 2006. Domestication of *Allanblackia floribunda*: Amenability to vegetative propagation. *Forest Ecology and Management* 237: 246-251. doi: 10.1016/j.foreco.2006.09.081.

Atangana, A.R., and D.P. Khasa. 2008. Preliminary survey of clonal variation in rooting of *Allanblackia floribunda* leafy stem cuttings. *Canadian Journal of Forest Research* 38: 10-15.

Jung Kim, M., H. Kim, J.S. Shin, C.-H. Chung, J.B. Ohlrogge and M.C. Suh. 2006. Seed-specific expression of sesame microsomal oleic acid desaturatse is controlled by combinatorial properties between negative *cis*-regulated elements in the *SeFAD2* promoter and enhancers in the 5'-UTR intron. *Mol. Gen. Genomics* 276: 351-368. doi 10.1007/s00438-006-0148-2.

Leakey, R.R.B., J.F. Mesén, Z. Tchoundjeu, K.A. Longman, J. McP. Dick, A. Newton, A. Matin, J. Grace, R.C. Munro, and P.N. Muthoka. 1990. Low-technology techniques for the vegetative propagation of tropical trees. *Commonwealth Forestry Rev.* 69 (3): 247-257.

Leakey, R.R.B., J.-M. Fondoun, A. Atangana, and Z. Tchoundjeu. 2000. Quantitative descriptors of variation in the fruits and seeds of *Irvingia gabonensis*. *Agroforestry Systems* 50: 47-58.

Shen, B., K.W. Sinkevicius, D.A. Selinger, and M.C. Tarczynski. 2006. The homeobox gene *GLABRA2* affects seed oil content in *Arabidopsis*. *Plant Molecular Biology* **60**: 377-387. doi 10.1007/s11103-005-4110-1.

Tchoundjeu, Z., E.K. Asaah, P. Anegbeh, A. Degrande, P. Mbile, C. Facheux, A. Tsobeng, A.R. Atangana, M.L. Ngo-Mpeck, and A.J. Simons. 2006. Putting participatory domestication into practice in West and Central Africa. *Forests, Trees and Livelihoods* 16: 53-69.

Vivien, J., and J.J. Faure. 1996. Fruitiers sauvages d'Afrique: Espèces du Cameroun. *Nguilou-Kerou Edition, Saint-Cénéré,* France, 416 p.