MOLECULAR ECOLOGY AND POPULATION GENETICS OF TWO MULTIPURPOSE TROPICAL TREES, *MORINGA OLEIFERA* LAM. AND *M. STENOPETALA* (BAK. F) CUF.

Geoffrey Mwanza Muluvi, BSc (For.) Hons.

A thesis presented to the University of Dundee in fulfilment of the requirement for the degree of doctor of philosophy.

AUGUST 1998

TABLE OF CONTENTS.

CHAPTER TWO

CHAPTER THREE

LIST OF FIGURES

LIST OF TABLES.

ACKNOWLEDGEMENTS.

I would like to thank my supervisors Professor Janet Sprent and Professor Wayne Powell for their advice and support when this work was being carried out and for their comments during the writing process. I am particularly grateful to Professor Sprent, Dr Geoff Folkard (University of Leicester) and Dr David Odee, Kenya Forestry Research Institute (KEFRI), for their participation and guidance during *Moringa* field surveys and in collection of plant material in Kenya. I am also grateful to the following staff at KEFRI; Sheila Mude, Wyclif Mauta, John Gicheru, Emmanuel Makatiani and John Sutherland (University of Leicester) for assistance in seed collection and procurement. I would like to thank the European Union (contract No. ERBTS3-CT94-0309) for funding my studentship. I am also grateful to the Director KEFRI and the Kenya Government for granting me three years study leave and for additional financial support. Additional sources of funding for this work were provided by Professor Janet Sprent and Professor Wayne Powell to whom I am very grateful.

Valuable advice and support came from a number of additional sources at University of Dundee including Dr Joan Sutherland, Dr Patrick Whitty, Shona McInroy, Dr Hilary Young, Dr Clare Halpin, Dr Will Whitfield, Dr Richard Parsons, Dr Wen-Jun Wang, Gail Alexander, Bob Ferrier, Iain Tennant and at SCRI, Dr Joan Russell, Dr David Marshall and Jim McNicol for which I am very grateful. Thanks to all my friends at SCRI and University of Dundee who helped me in many ways; Nicole Soranzo, Gemma White, Dr Jim Provan, Lalith Perera, Allan Booth, Dr Keith Skene and Dr Abdellah Barakate. Finally, I would like to thank my family and friends in Kenya for the support, patience and understanding during the three years this work was being carried out.

UNIVERSITY OF DUNDEE

ACCESS TO THESES PRESENTED FOR HIGHER DEGREES

1. AUTHOR'S DECLARATION

The University Library receives many requests from other libraries to supply photocopies or microfilms of theses of this University. Such requests are normally complied with, provided that the consent of the author can be obtained. Occasionally, however, it is difficult or even impossible to contact an author, especially in the case of graduates from overseas.

It would greatly assist the Library if authors of theses would complete and return the certificate below, indicating whether or not they are prepared to sanction the reproduction of their theses, subject to appropriate safeguards. Further reference to an author in this connection would then be unnecessary.

It is emphasised that, if this permission is given, it in no way affects an author's interests in his or her work under copyright law.

NAME ..... GEOFFREY M. MULUVI .....

TITLE OF THESIS ..... MOLECULAR ECOLOGY AND POPULATION GENETICS OF TWO MULTIPURPOSE TROPICAL TREES MORINGA OLEIFERA LAM. AND M. STENOPETALA (BAK F) CUF.

(a) I am willing that my thesis should be available for reproduction at the discretion of the Librarian of the University of Dundee and on the understanding that users are made aware of their obligations under copyright, or

(b) I am not willing that my thesis should be available for reproduction.

(Please delete whichever is inapplicable)

Signed ..... Date ..... 4/9/98 .....

2. SUPERVISOR'S DECLARATION

The University Library will restrict access to a thesis where a formal agreement exists on the confidentiality of the research. By a decision of Senate, requests for restricted access in the absence of a formal agreement will be referred to the Library Committee.

To assist Library staff, you are requested to complete the certificate below.

SUPERVISOR ..... J. I. SPRENT .....

DEPARTMENT ..... BIOL SCI .....

(a) The subject matter of the thesis above is not confidential and there should be no restrictions on access, or

(b) ~~Access to the above thesis should be restricted for reasons of confidentiality and I will furnish the Librarian with further details in writing.~~

(Please delete whichever is inapplicable)

Signed ..... Date ..... 4/9/98 .....

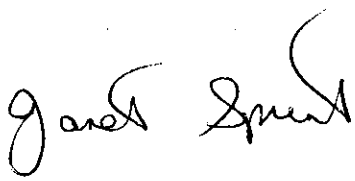
DECLARATION.

This thesis is my own composition the results presented are of investigations conducted by myself, work other than my own is clearly indicated by reference to relevant workers or their publications and it has not been presented in any previous application for a higher degree.

Geoffrey Mwanza Muluvi.

STATEMENT

I certify that Geoffrey Mwanza Muluvi has spent 9 terms of full time research work under my supervision and has fulfilled the conditions of ordinance No 14 (University of Dundee) and is thus qualified to submit this thesis for the degree of Doctor of Philosophy.



Professor Janet Irene Sprent.

ABSTRACT.

M. oleifera Lam.and *M. stenopetala* (Bak. f) Cuf. are multipurpose tropical tree species native to India and East African lowlands respectively. In tropical developing countries, both species provide a variety of useful products such as coagulants for water clarification, edible oil, vegetable, fodder, traditional medicine and use in agroforestry systems. *M. oleifera* was introduced to Kenya about 100 years ago, but until recently, the species was considered of marginal value and details on these introductions, the genetic diversity and relationships within and among the introduced populations have not been documented. In addition, information on genetic structure in *M. stenopetala* and its relationship with *M. oleifera* and their potential for improvement is currently lacking.

RAPDs and AFLPs were used in genetic analyses in *M. oleifera* (chapter 3) and in order to further understand the resulting genetic structure and relationships, work was carried out to investigate clonality (chapter 4) and estimate outcrossing rates (chapter 5). To gain insights into agromorphological relationships, their variability and potential for improvement, an *M. oleifera* provenance trial was established in Kenya (chapter 6). The genetic structure in *M. stenopetala* populations was investigated using RAPDs and AFLP markers and the relationship between the two species determined using AFLP (chapter 7). Data revealed hierarchical genetic structure, extensive sexual reproduction, self-compatibility, reduced genetic variation and possible multiple origins among Kenyan *M. oleifera* provenances. Most of the phenotypic variation in quantitative traits among Kenyan *M. oleifera* was found within provenances. A wide divergence and different genetic structures in both species was found. The importance of data for the conservation, utilisation, management and improvement of genetic resources within both species are discussed.

DEDICATION.

I dedicate my thesis to Jayne, Linda, Kelvin and my parents.

CHAPTER ONE

1.0 General introduction.

1.1 Background information on *Moringa oleifera* and *M. stenopetala.*

1.1.1 Taxonomy.

Moringa oleifera Lam (syn. *M. Pterygosperma* Gaertn.) and *M. stenopetala* (Bak.f.). Cuf (syn. *Donaldsonia stenopetala* Bak.f.,) are members of the Moringaceae, a monogeneric family of shrubs and trees (Ramachandran et al 1980, Jahn et al 1986, Morton 1991). Although very similar to the Capparidaceae, the Moringaceae is distinguished by parietal placentation, 3-valved fruits, elongated, non-dehiscent berries and winged seeds (Ramachandran et al 1980). The genus was previously credited with only four species (Philips 1951), but today it numbers 14 (Verdcourt 1985). Important characteristics of this genus include coagulants for water clarification, edible oil and antimicrobial substances (Jahn 1984, Jahn et al 1986). *M. oleifera* and *M. stenopetala* are the most important species in the genus because of the quantity and quality of their flocculating substances and the two species will be discussed further in subsequent sections. Classification of the two species based on morphological characters has been described by Beentje (1994). Mature trees of *M. oleifera* and *M. stenopetala* are shown in Fig. 1.1.

Fig. 1.1 Mature trees of *M. oleifera* and *M. stenopetala*.

A *M. oleifera* tree growing in Mbololo.

B *M. stenopetala* tree growing in Island camp (Kokwa Island), Baringo.

1.1.2 Origin and distribution.

M. oleifera (common name; annual drumstick) is indigenous to North west India (Ramachandran et al 1980). The species has a wide distribution range from the warm tropical coastal climate to subtropical climate. Although it prefers alluvial sands, it can be grown in a variety of soils (Nautiyal and Venkataraman 1987). It is cultivated throughout India, the Middle East and in almost the whole tropical belt (Ramachandran et al 1980, Jahn et al 1986, Kantharajah and Dodd 1991). In Kenya, the species grows as a shrub or small tree attaining a height of 2.5-10 m (Beentje 1994).

M. oleifera was introduced to Kenya at the turn of the century by Indian workers who came to Africa to build the Mombasa to Kampala railway line. These workers planted their cherished "drumstick" trees in their new settlements along this line. The trees flourished in the area between Mombasa and Mtito Andei (altitude 17-600 m) and fruiting was observed up to Sultan Hamud (altitude 1500 m), however, further north in Nairobi (altitude 1660 m) *M. oleifera* failed to develop flowers (Jahn 1991).

M. stenopetala is a multipurpose tree, 6-12 m high and is indigenous to northern Kenya and Ethiopia (Jahn 1991) where it grows wild up to 1000-1800 m. It is found in riverine and lake sides in dry areas (Beentje 1994) and also on rocky ground (Teketay 1995). *M. stenopetala* trees can achieve breast height diameters of 60 cm, are drought and high temperature tolerant and grow in Kenya in areas with mean annual temperatures of 24 °C-30 °C and in southern Ethiopia with mean annual rainfall of 500-1400 mm (Teketay 1995). *M. oleifera* produces fruit within 6 months of planting while *M. stenopetala* requires at least 2.5 years for fruit production.

1.1.3 Propagation.

M. oleifera is propagated through seeds, cuttings (Ramachandran et al 1980, Jahn et al 1986, Nautiyal and Venkataraman 1987), tissue culture (Kantharajar and Dodd 1991) and air layering (Nautiyal 1987). *M. stenopetala* is propagated through direct seed sowing in the field or nursery raised seedlings (Teketay 1995). Conventional cultivation of *M. stenopetala* from cuttings is unknown (Jahn 1991). There is no requirement for pre-treatment of seeds but shading, especially during hot weather, is needed for optimum germination (Jahn et al 1986). Seeds exhibit no dormancy and germinate readily within 7-10 days (Nautiyal and Venkataraman 1987). The optimum temperature for germinating *M. stenopetala* seeds is 25 °C (Teketay 1995). Seedlings are susceptible to drought and should not be transplanted too early if good survival is to be expected, however, established saplings and poles are found to be drought resistant (Nautiyal and Venkataraman 1987). Established trees coppice and pollard well (Nautiyal and Venkataraman 1987). *M. oleifera* trees are fast growing and have been reported to attain 2.5 m in 3 months in Zaire (Morton 1991) and 4 m in 4 months in western Kenya (unpublished data). Fig. 1.2 illustrates fast growing, six month old, *M. oleifera* trees growing at Siaya, western Kenya.

Fig. 1.2 Six month old *M oleifera* trees growing in Siaya, western Kenya.

The trees were planted and managed by school pupils within their compound.

1.1.4 Diseases and pests.

A root-rot caused by fungus *Diplodia* sp and the bark-eating caterpillar (*Indarbela quadrinotata* Wlk) have been observed in India (Ramachandran et al 1980, Morton 1991). Bark damage by insect pests on *M. oleifera* trees has been observed in Kibwezi and Mbololo regions of Kenya (Fig. 1.3 B). Other insect pests of *M. oleifera* found in India include *Eupterote molifera* Wlk, *Tetragonia siva., Diaspidotus* sp and *Ceroplastodes cajani* (Ramachandran et al 1980, Morton 1991). Defoliation of *M. stenopetala* by insect pests has been observed by the author at Baringo, Kenya (Fig. 1.3 A). The insect damage is serious during the dry season but trees recover naturally during the rainy season. However, this defoliation has been shown to reduce growth and productivity of the plants (unpublished field data). Susceptibility of *M. oleifera* to termite attack was observed in Puerto Rico (Morton 1991) and by the author in certain dry areas of Kenya such as Kitui and Kibwezi.

Defoliation by insect pests in *M. oleifera* has been observed in India on trees planted in sorghum or maize fields during the rainy season (Jahn 1991). Small green caterpillers that caused defoliation during the rainy season were reported in the southern Rift Valley (Kenya) with trees recovering the following dry season (Mayer and Stelz 1993). Seeds sown in deep generic ferrasols were also attacked by insects (Mayer and Stelz 1993). A plant parasite attacking young *M. oleifera* trees and other members of related families was observed by the author in western Kenya (Fig. 1.3 C).

Fig. 1.3 *M. oleifera* and *M.stenopetala* trees attacked by pests and plant parasite.

A Insect defoliated *M. stenopetala* in Marigat, Kenya.
B Insect defoliated and termite attacked *M. oleifera* in Kibwezi, Kenya.
C Six month old *M. oleifera* in Siaya, Kenya, attacked at the tip by a plant parasite.

1.1.5 Uses.

The uses of *M. oleifera* and *M. stenopetala* and other members in this genus are similar and are outlined below.

1.1.5.1 Water purification.

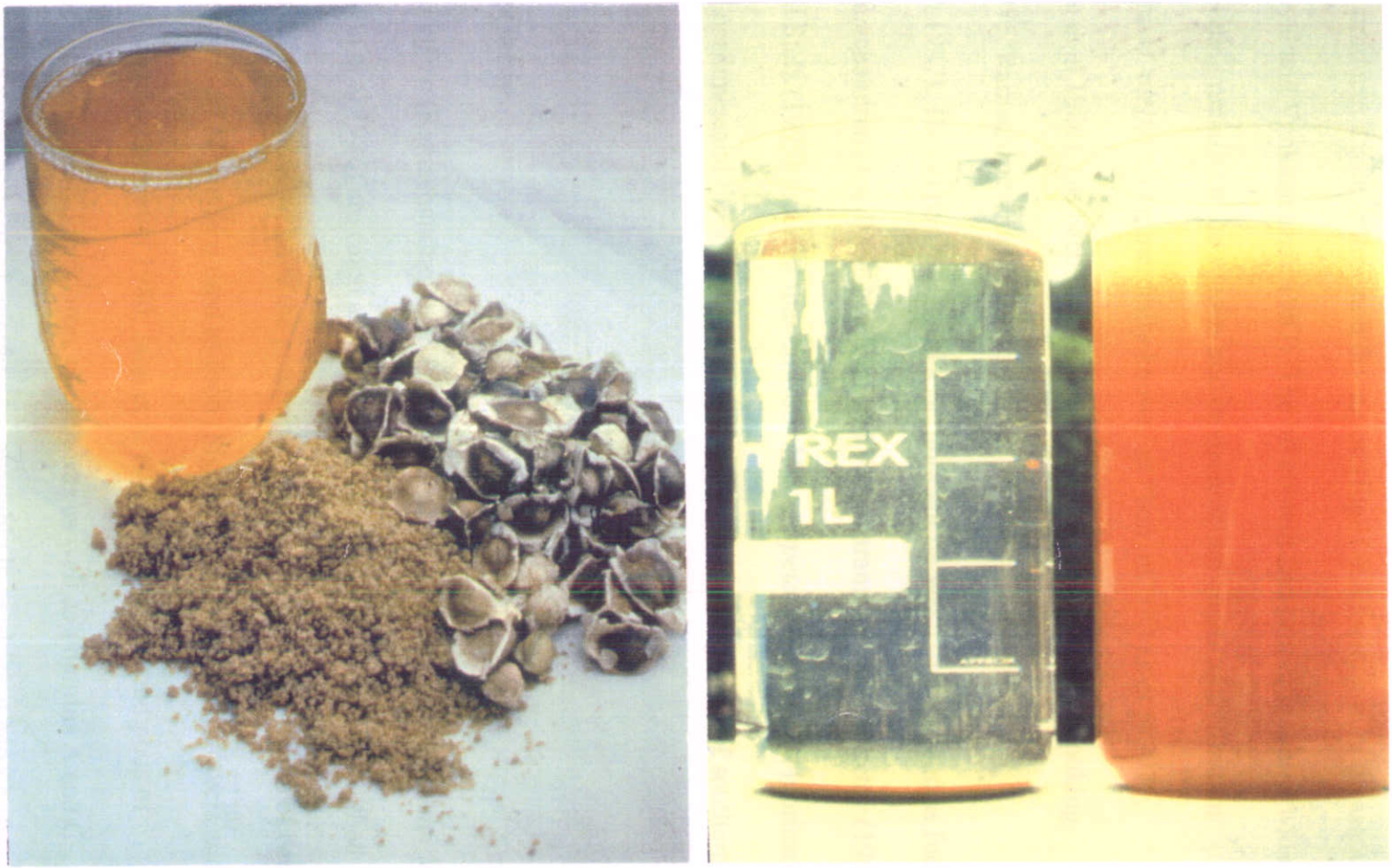
Seeds of *Moringa* have flocculation and antimicrobial properties (Jahn 1984, Muyibi and Evanson 1995, Teketay 1995, Ndabigengesere & Narasiah 1998). Powder from the seeds is used as a natural coagulant for low-cost water purification (Fig. 1.4 B) with whole crushed seed kernels of *M. stenopetala* reported to reduce bacterial contamination by 90-99.9 % (Jahn 1989). Coagulants in *Moringa* seeds are comparable to the conventional coagulant alum (aluminium sulphate) and when applied in doses of 30 to 200 mg/l in accordance with raw water quality, the *Moringa* seed powder suspension could clarify different types of tropical surface waters with low, medium and high turbidities to tap-water quality within one to two hours (Jahn 1984). *Moringa* seeds have traditionally been used as a water coagulant in Sudan (Jahn et al 1986) and by the Njemps people living around Lake Baringo (Kenya) since 1973 (unpublished field data).

The active agents in *Moringa* seed coagulants are water soluble, highly cationic polypeptides with dimers of 13 KDa and sub units of about 6.5 KDa, and isoelectric point above pH 10 with both monomer and dimer retaining coagulation properties (Gassenschmidt et al 1995, Ndabigengesere et al 1995). The predominant mechanism of the coagulation with *Moringa* seed powder is adsorption and charge neutralisation (Ndabigengesere et al 1995). As opposed to aluminium sulphate, *Moringa* coagulants require no pH adjustment, show low values of sludge volume and the products are

Fig. 1.4 Oil extracted from seeds and water clarified with *M. oleifera* coagulants.

A *M. oleifera* seeds, presscake and edible oil.

B Muddy and clarified water.



Photographs courtesy of Dr. G. Folkard.

organic and biodegradable (Ndabigengesere et al 1995, Ndabigengesere and Narasiah 1996).

1.1.5.2 Oil.

Moringa seed produces a pale yellow non drying oil (Fig. 1.4 A) with a mild pleasant flavour (Nautiyal and Venkataraman 1987). The oil content and properties show wide variation depending mainly on the species and environmental conditions (Ibrahim 1974). Oil from *M. oleifera*, commercially known as "ben oil" (Khan et al 1975, Ramachandran et al 1980), has been used in the cosmetic industry and for cooking (Khan et al 1975, Ramachandran et al 1980, Jahn et al 1986, Nautiyal and Venkataraman 1987). The oil has been used as a lubricant by watch makers and as fuel for lamps (Ramachandran et al 1980, Nautiyal and Venkataraman 1987). Teketay (1995) and Mayer and Stelz (1993) reported that oil from *M. stenopetala* is used as a lubricant, for cosmetics/ perfumes and for soap production by farmers in some areas of southern Ethiopia.

1.1.5.3 Vegetable.

Leaves of *M. oleifera* (Ramachandran et al 1980, Jahn et al 1986, Beentje 1994) and *M. stenopetala* (Jahn et al 1986) are eaten as vegetables. Subadra et al (1997) found that dehydrated leaves of *M. oleifera* have the potential to serve as a valuable source of beta-carotene. Other parts of *M. oleifera* used as food (Fig. 1.5) are green pods, flowers and roasted seeds (Ramachandran et al 1980, Jahn et al 1986). The leaves and pods of *Moringa* are rich in proteins, calcium, iron, phosphorus as well as vitamins A and C (Teketay 1995). In Kenya, pods of M. *oleifera* are available in markets at Mombasa and

are also exported to Europe. The coastal and Burji communities of Kenya eat leaves of *M. oleifera* and *M. stenopetala* respectively while the Kamba people eat the flowers (Fig. 1.5B) which they claim taste like fried eggs. In both cases, the leaves or flowers are boiled then fried before consumption (unpublished field observations). In southern Ethiopia, *Moringa* leaves are eaten and sold in markets during the rainy season, but farmers prefer to eat ordinary cabbage and other vegetables and keep *M. stenopetala* leaves for the dry season, when other vegetables are not available (Jahn 1991, Mayer and Stelz 1993).

1.1.5.4 Fodder.

Leaves and young shoots of *Moringa* are used as fodder for livestock (Jahn et al 1986, Teketay 1995). Makkar and Becker (1997) found that leaves and the residues obtained after the recovery of oil and coagulants from *Moringa* can be good sources of protein for animal feed. In some dry areas of Kenya, leaves and dry pods of *M. stenopetala* are popular as goat feed. The livestock of the Turkana and the Njemps people (Kenya) browse on *M. stenopetala* leaves (Jahn 1991) while in Konso (Ethiopia), young pods are fed to goats and sheep (Mayer and Stelz 1993).

Fig. 1.5 Flowers and fruits of *M. oleifera* which are used as vegetables.

A Sliced and canned immature fruits for export.

B Flowers

Photographs courtesy of Dr. G. Folkard.

1.1.5.5 Medicinal value.

Jahn (1991) reported that the Turkana of Kenya take leaf extracts of *M. oleifera* orally as a remedy for leprosy, while the Njemps chew the bark to relieve coughs. Caceres et al (1991), in an ethnobotanical survey of the uses of *M. oleifera* in Guatemala, reported that the main medicinal uses are for the treatment of infectious diseases of the skin and mucosa (boils, spots, ringworm rash), digestive system (stomach pains, diarrhoea) and respiratory tracts (fever, influenza, cold).

The bitter-tasting water poured off after cooking leaves of *Moringa* is drunk by the Konso people of Ethiopia to prevent colds and anaemia, while the Gidole and Burji (Kenya) use it as a cure for digestive problems and dysentery. In Arba Minch (Ethiopia), the treatment of hypertension and diabetes with leaves of *Moringa* is advocated by physicians (Jahn 1991). The leaves are used to treat high blood pressure, diarrhoea, diabetes and heart problems (Mayer and Stelz 1993). In many parts of the Rift valley, liquid squeezed from roots of *M. stenopetala* is used to treat malaria.

Caceres et al (1991) confirmed that the popular utilisation of the seeds and fresh leaves of *Moringa* as an antibacterial agent in skin and respiratory infections has some scientific justification. Caceres and Lopez (1991) demonstrated (*in vivo*) the effectiveness of *M. oleifera* ointments in the treatment of experimentally induced *Staphylococcus aurea* pyodermia in mice. Recent *in vivo* experiments by Makonnen et al (1997) have shown that extracts from *M. stenopetala* can lower blood glucose concentration. Pal et al (1995) identified antiulcer activity in *M. oleifera* leaf extracts while Ezeamuzie et al (1996) identified anti-inflammatory substances in *M.oleifera* that can be used to treat both acute and chronic inflammatory conditions.

1.1.5.6 Spice.

Roots of *M. oleifera* and bark of *M. stenopetala* are used as food spices. The roots smell like horseradish and were used by British colonists in Kenya as a substitute for horseradish (Jahn et al 1986).

1.1.5.7 Rayon, paper pulp and gum.

Moringa wood was found to be suitable for the production of both rayon and paper pulp which was of higher quality than eucalyptus (Nautiyal and Venkataraman 1987). The corky, grey bark yields a coarse fibre and is used in making mats, paper and cordage, while gum exuding from the stem is used in calico printing (Ramachandran et al 1980).

1.1.5.8 Agroforestry.

Both *M. stenopetala* and *M. oleifera* are intercropped with staple food crops such as sorghum, maize, finger millet, barley, pulses, tubers, cotton, sugar cane, coffee, pepper and papaya (Babu and Rajasekaran 1991, Mayer and Stelz 1993, Teketay 1995). They can also be used as a live fence, as ornamentals and in windbreaks (Jahn et al 1986). *M. oleifera* is cultivated as an important agroforestry crop at Mbololo and Kibwezi.

1.1.6 Genome size and chromosome number.

M. oleifera is diploid with 2n = 28 (Puri 1941, Ramachandran et al 1980). The genome size of *M. oleifera* (DNA amount of the unreplicated haploid genome of an individual) is 0.6 pg (579 Mb) (NB $1pg=10^{-12}$ g; 1 Mb = 10^6 nucleotide base pairs; 1 pg = 965 Mb) (Bennet and Leitch 1995). The DNA content among the nuclei of flowering plants is known to differ by more than 600-fold, ranging from less than 0.2 pg in *Arabidopsis*

thaliana to 127.4 pg in *Fritillaria assyriana* (Bennett and Smith 1976, 1991). Therefore, the genome size of *M. oleifera* is very small compared to most angiosperms (Bennet and Leitch 1995). Information on genome size of *M. stenopetala* is currently lacking.

1.2.0 Genetic and phenotypic variation.

1.2.1 Genetic variation.

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

1.2.2 Phenotypic variation.

Several phenotypes of *M. oleifera* have been reported in India (Ramachadran et al 1980, Jyoth et al 1990, Morton 1991) based on length of the pods and includes short pods (15-23 cm), medium sized pods (23-60 cm) (Fig.1.6 C), long pods (60-90 cm), and very long pod phenotype (90-120 cm). Other phenotypes reported are those which flower throughout the year and yield heavy crops (Fig. 1.6 B); those which do not flower (grown mainly for foliage); wide stem diameter phenotypes suitable for pulping (Fig.

1.6 A) and the wild trees which bear small and inferior fruits. Phenotypic obervations in the form of replicated trials have not been reported for *M. stenopetala.*

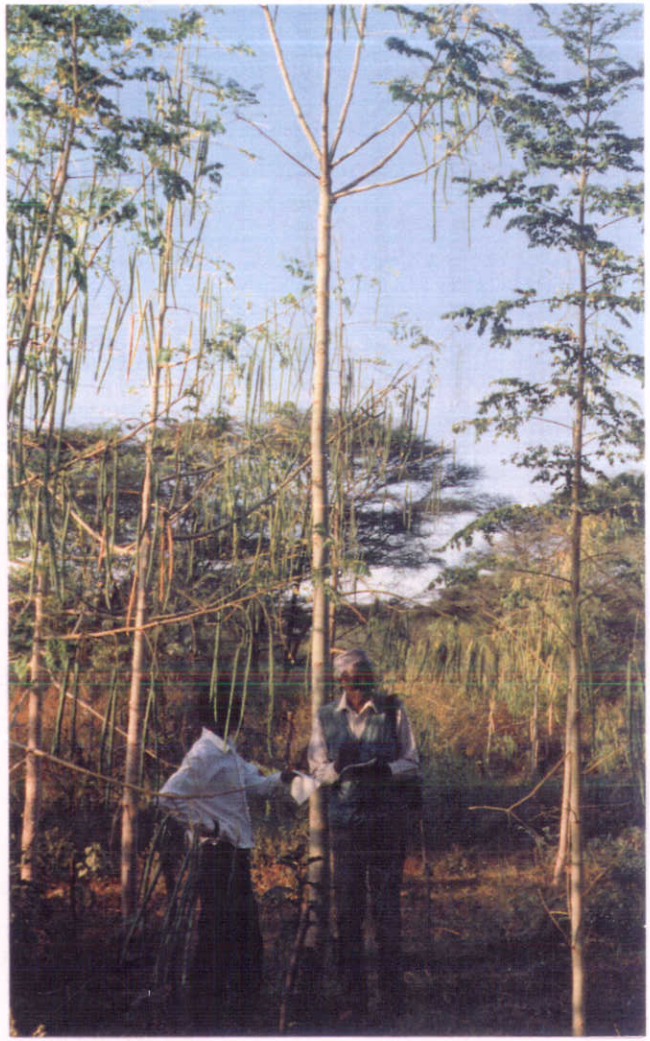
1.2.3 Genetic structure.

The genetic structure refers to the distribution of genetic variation within and between populations. A detailed understanding of the extent and distribution of the genetic variation within a genus can be exploited for genetic improvement of any organism (Chalmers et al 1992, Sharma et al 1995). Population geneticists have long recognised that the genetic diversity present in a species is hierarchically structured. In addition to differences among individuals within any one population, there may be differences among populations within a given geographical region, differences among populations from different geographical regions, and differences among entire geographic regions (Holsinger and Mason-Gamer 1996).

The distribution of variation among populations is the product of interactions among several evolutionary factors such as selection, effective population size, and the ability of the species to disperse pollen and seeds (Hamrick 1989). In general, selection should increase population differentiation, as would genetic drift. On the other hand, species with more pollen or seed movement should have less differentiation than species with restricted gene flow. Levels of genetic variation differ significantly among species with different geographic ranges, life forms, and taxonomic affinities (Loveless 1992). Cultivated taxa maintain higher percentage of polymorphic loci and higher mean heterozygosity than native tropical species while woody perennials possess more variation at marker loci than short-lived plants (Loveless 1992).

Fig. 1.6 Some phenotypes and fruits of *M. oleifera*.

A Clear straight bole and relatively wide diameter tree (1 year old) growing in Marigat, Kenya.



B High yielding *M. oleifera* (6 month old) tree growing in Siaya, Kenya.

C Medium size (about 45 cm) fruits from *M. oleifera* growing in Kitui, Kenya.

Reviews of the levels of variation detected in a wide range of plant species based on isozyme data, reveal that outcrossing tropical tree species maintain most of their variation within populations (Hamrick, 1989). However, genetic variation in selfing plants is expected to be larger among, rather than within populations (Bonnin et al 1996). Morphological data have provided valuable insights into genetic relationships within and among natural populations in other species such as *Pinus banksiana* (Maley and Parker 1993), *Lens* (Ahmad et al 1997), *Acacia* (Harrier et al 1997). More recently, molecular techniques have provided an additional tool for examining variation (Orozco-Castillo et al 1994).

1.3.0 Molecular markers.

Genetic diversity has been studied in plant species using a variety of morphological, chemical and molecular descriptors. Traditionally, measurements of genetic diversity rely upon the ability to resolve differences in morphological characters (Dawson et al 1993, Tingey and del Tufo 1993). The range of characters available may be increased by the use of electron microscopy or biochemical and phytochemical assays (Karp et al 1996). Although these approaches are extremely powerful, DNA-based procedures for the detection of polymorphism reveal an immense number of characters (Karp et al 1996). Molecular markers, which includes protein and DNA based assays (Powell 1992), have provided unprecedented power to explore the genomes of diverse plant species, including trees (Neale and Harry 1994).

1.3.1 Protein markers.

The most widely used protein markers in plant breeding are isozymes (Powell 1992) and seed proteins (Obara-Okeyo and Kako 1998). Isozymes are enzymes that share a

common substrate but differ in electrophoretic mobility, and are revealed when tissue extracts are subjected to electrophoresis in various types of gels and subsequently immersed in solution containing enzyme specific stains (Powell 1992). Isozyme analysis is a valuable tool for determining genetic relationships (Heun et al 1994), estimating genetic diversity and evaluating population differentiation (Orozco-Castillo et al 1994).

However, isozyme polymorphisms deal with a limited group of functional enzymes and mainly concern variations in the coding sequence of the gene (Bachmann 1994), thus limiting the number of polymorphic loci detected (Orozco-Castillo et al 1994), while DNA polymorphisms can occur anywhere in the genome including coding and non coding, single copy or repetitive DNA (Bachmann 1994, Szmidt et al 1996).

1.3.2 DNA marker systems.

Various methods from molecular biology reveal sequence polymorphisms in organellar and nuclear DNA that can be used as highly informative markers for the structure and dynamics of genomes at the level of populations and individuals (Bachmann 1994). The developed DNA markers which can be used to study various aspects of the plant genome include random amplified polymorphic DNA (RAPD; Williams et al 1990), simple sequence repeat polymorphism (SSR; Tautz 1989), cleavable amplified polymorphic sequences (CAPS; Lu et al 1993), amplified fragment length polymorphism (AFLP; Zabeau and Vos 1993), and restriction fragment length polymorphism (RFLP; Botstein et al 1980). These molecular techniques vary in the way that they resolve genetic differences, in the type of data they generate and in the taxonomic levels at which they can be most appropriately applied (Karp et al 1996). A summary of the different characteristics and how they compare for the different

molecular screening technologies is given in Table 1.1 as reported by Karp and Edwards (1997).

Table 1.1 Comparison of different molecular screening techniques.

Characteristics	RFLPs	RAPDs	SSRs	AFLPs	PCR-Seq[1]
Developmental costs ($ per probe)	Medium (100)	Low (none)	High (500)	Low (none)	High (500)
Level of polymorphism	Low-medium	Medium	High	Medium	Medium
Automation	No	Yes/ No	Yes/ No	Yes/ No	Yes
Cost of automation	High	Medium	High	High	High
Reliability	High	Low	High	Medium	High
Level of skills required	Low	Low	Low-medium	Medium	High
Cost ($ per assay)	High (2.00)	Low (1.00)	Low (1.00)	Medium (1.50)	High (2.00)
Radioactivity	Yes/No	No	Yes/ No	Yes/ No	Yes/ No

[1] PCR-sequencing.

1.3.2.1 Random Amplified polymorphic DNA (RAPD).

The Random Amplified Polymorphic DNA (RAPD) assay was developed independently by two different laboratories (Welsh and McClelland 1990. Williams et al 1990). The RAPD assay detects nucleotide sequence polymorphism in a DNA amplification based assay using only a single primer of arbitrary nucleotide sequence (Williams et al 1990, Bowditch et al 1993, Tingey and del Tufo 1993, Williams et al 1993).

The RAPD principle and reaction conditions were described in detail by Williams et al

(1993) and Bowditch et al (1993). To perform a RAPD assay, a single oligonucleotide of an arbitrary DNA sequence is mixed with genomic DNA in the presence of a thermostable DNA polymerase and a suitable buffer and then subjected to temperature cycling conditions typical of the polymerase chain reaction. At an appropriate annealing temperature during the thermal cycle, the single primer binds to sites on opposite strands of the genomic DNA that are within an amplifiable distance of each other (from 100 bp to a few thousand nucleotides), and a discrete DNA segment is produced.

The DNA amplification reaction is repeated with several different primers resulting in several amplified bands from each primer. Amplification products are separated electrophoretically in agarose gels giving banding patterns that can be scored for genetic variation (Sigurdsson et al 1995). RAPD polymorphisms result from either a nucleotide base change that alters the primer binding site or an insertion or deletion within the amplified region (Williams et al 1990). Polymorphisms are usually noted by the presence or absence of an amplification product from a single locus.

RAPD markers can be used for genetic mapping applications (Williams et al 1990), genetic variation studies (Chalmers et al 1992, Abo-elwafa 1995, Margale et al 1995; Nesbitt et al 1995, Salimath et al 1995, Ramser et al 1996, Gillies et al 1997, Palacios and Gonzalez-Candelas 1997, and Bustos et al 1998), cultivar or clone identification (Hu and Quiros 1991, Samec and Nasinec 1995, Sigurdsson et al 1995), phylogenetic and pedigree studies (Heun and Helentjaris 1993; Salimath et al 1995), estimation of outcrossing rates (Fritsh and Reiseberg 1992, Gaiotto et al 1997) as well as for genetic diagnostics (Williams et al 1993). This technique has also been used for the identification of associations between various quantitative traits and RAPD molecular

markers (Virk et al 1996).

The advantages of the RAPD assay include its simplicity and speed (Rafalski et al 1991, Margale et al, 1995, Salimath 1995, Ramser et al 1996), low cost (Salimath 1995, Ramser et al 1996) and the requirements for small amounts of genomic DNA (Rafalski et al 1991, Waugh and Powell 1992, Ramser et al 1996). The assay is non radioactive and is applicable to a broad range of species. However, most RAPD markers show dominant / recessive inheritance in diploid organisms (Williams et al 1993, Thormann et al 1994, Szmidt et al 1996) and transferability of the method between laboratories is often limited (Joshi and Nguyen 1993, Williams and St. Clair 1993, Micheli et al 1994). RAPD markers were used in chapters 3 and 7 for genetic analysis of *M.oleifera* and *M. stenopetala,* respectively.

1.3.2.2 Amplified fragment length polymorphism (AFLP).

Amplified fragment length polymorphism (AFLP), is a novel PCR based technique for DNA fingerprinting (developed by Keygene, Inc.) which selectively amplifies DNA fragments (Vos et al 1995, Zabeau and Vos 1993). Prior to PCR amplification, genomic DNA is digested by two restriction endonucleases and site-specific adaptors are then ligated to the resulting DNA fragments to generate template DNA for PCR. Template DNA is amplified using PCR primers complementary to the adaptor sequences. Addition of extra nucleotides at the 3' end of the PCR primers allows the selective amplification of only those restriction fragments starting with nucleotides homologous to those of the primers. In this way AFLP fingerprints can be tailored to produce patterns of desired complexity. In principle, both the restriction enzymes and the nature and number of selective nucleotides can be varied. Consequently, a virtually unlimited

number of markers can be generated. Resolution of the resulting DNA fragments on standard sequencing gels allows for the detection of amplified fragment length polymorphisms (AFLPs).

The reproducibility of AFLP profiles is assured by using primer sequences of at least 16 nucleotides at stringent amplification conditions. The reproducibility of the AFLP technique is further demonstrated by the consistency in relative band intensities. However, the quantitative differences in band intensities between the lanes, are easily used to recognise the homo- or hetero-zygous presence of an AFLP marker allele (Van Eck et al 1995).

Similar to RAPD analysis, an AFLP assay requires no prior sequence knowledge, but detects a 10-fold greater number of loci (20-100) than those detected by RAPD analysis (Maughan et al 1996, Powell et al 1996). Thus, AFLPs have the capacity to analyse many markers in a short time (Hartl & Seefelder 1998). They are particularly good for mapping and fingerprinting, and genetic distances can be calculated between genotypes (Karp et al 1996). AFLP markers appear to be inherited in a stable Mendelian manner and have immediate and specific applications in gene tagging when combined with near-isogenic line and bulk segregant analysis (Maughan et al 1996). Moreover, AFLP primers can be easily distributed among laboratories by publishing primer sequences (Maughan et al 1996). They do, however, share many of the limitations, with respect to band homologies and identities as found in RAPDs (Karp et al 1996). AFLP markers were used for genetic analysis on the work reported in all the chapters with the exception of chapter 6.

1.4.0 Comparison between RAPDs and AFLPs.

With the development of molecular markers and their many perceived advantages, it is now common for these techniques to be applied to assess genetic diversity in germplasm collections and to supplement and refine the morphological-based classification. However, in recent years, the number of molecular techniques available for application in this area has increased dramatically, with each method differing in principle, in application, in the type and amount of polymorphism detected and in cost and time requirements (Russell et al 1997, Powell et al 1996). Faced with this wealth of marker technology, it is appropriate to determine if the same patterns of variability and genetic relationships are revealed by the RAPDs and AFLP markers (chapter 3 and 7).

1.5.0 Investigation of clonality.

Clonality in plants can arise through production of asexual propagules such as bulbils and cuttings and selfing of asexually produced seeds (agamospermy) (Ellstrand and Roose 1987). Because of the special problems clonal plants present for analysis of populations (Parks and Werth 1993), it is important to discriminate among genets (single genetic individuals) in clone-forming species in order to allow ramets (morphologically distinct individuals) belonging to each clone to be recognised and the spatial configuration of the clones to be visualised (Cook 1983). Techniques such as excavations have been used to study clonality in some plants but for a majority of species, genet recognition is facilitated by molecular markers (Ellstrand and Roose 1987). Molecular techniques represent the only reasonable approach for confident assignment of ramets to genets in rapidly expanding and fragmenting clonal plants (Parks and Werth 1993) (chapter 4.0).

1.6.0 Mating systems.

Characterisation of mating systems is of prime concern in the study of plant population genetics and evolution because it determines the genetic composition of a population (Milligan 1996, Premoli 1996). Tropical trees have diverse and, in many cases, poorly understood pollination and seed dispersal mechanisms and studies of pollination and reproductive biology, as well as genetic breeding systems, are critical to their management, genetic improvement and conservation (Loveless 1992).

The interaction between pollinators and plant population density or other environmental factors may influence the level of selfing in the population and affect the expression of inbreeding depression (Norman et al 1997). For self-compatible species, insect pollinator may lead to higher outcrossing rates than wind pollination, particularly when individual plants are highly dispersed (Norman et al 1997). With wind pollination, the concentration of pollen grains decreases rapidly with distance from the plant, making outcrossing unlikely (Lemen 1980). In contrast, insect pollinators may fly long distances, carrying large quantities of pollen directly to a receptive stigma (Norman et al 1997).

Stand density is one of the environmental factors that may play a major role in determining outcrossing rates in plant populations (Morgante et al 1991). It had been hypothesised that in wind-pollinated species, outcrossing rates should decrease as stand density decreases, because pollen dispersal decreases exponentially with distance (Farris and Mitton 1984). In low-density stands, the pollen cloud of any individual tree contains

a relatively higher proportion of its own pollen than in high-density stands and that there is an increase in the local mating competition among sib pollen (Morgante et al 1991). However, contrasting experimental results have been obtained concerning this hypothesis. Positive correlations between outcrossing rate and plant density have been found in *Pinus pondorosa* Laws (Farris and Milton 1984) but Morgante et al (1991) did not find stand density in two Norway spruce (*Picea abies*) populations to have any impact on outcrossing rates. Measurements of outcrossing rates have been further complicated by the fact that single generation estimates of selfing rates made over multiple years demonstrate temporal variation (Hamrick 1982, Holtsford and Ellstrand 1990, Doyle and Ritland 1993). This lead Milligan (1996) to develop a method for estimating long-term mating systems using DNA sequences. The mating system in *M. oleifera* was investigated and is described in chapter 5.

1.7.0 Morphological and quantitative variation.

Morphological data have provided valuable insight into genetic relationships within and among populations (Maley & Parks 1993), as recently demonstrated by Harrier et al (1997) in the genus *Acacia* and Ahmad et al (1997) in *Lens* species. Studies of quantitative genetic variation using traditional quantitative genetics design where environmental variance is minimized can reveal not only genetic relationships between and within populations under study but also the amount of variation available for manipulation in breeding programmes. However, morphological markers have limited practical applications in plant breeding and real challenges are in identification of specific regions of the genome that enhance the expression of quantitatively controlled characters (Powell 1992). Because most economically important traits in forest trees are

quantitatively inherited (Neale and Harry 1994), the dissection and manipulation of these quantitatively controlled characters are important objectives in both basic and applied genetic research (Powell 1992). The relationship between three provenances of M. oleifera based on agromorphological characters was investigated and is reported in chapter 6.

1.8.0 General aims of the project.

Information on genetic relationships and diversity within and between the introduced Kenyan populations of *M. oleifera* and those of *M. stenopetala* is not available. For many years, the species were considered of marginal commercial value and consequently, clonal identities or seed sources from these introductions have not been maintained. Analysing existing genetic and quantitative variation between and within populations is important for the efficient management, conservation, utilisation and improvement of *Moringa* genetic resources in Kenya.

The specific objectives of the present study were to:

i) optimise procedures for DNA isolation from *Moringa* species;

ii) examine the utility of RAPDs and AFLP markers in determining molecular genetic variation in *M. oleifera;*

iii) determine the extent of clonal propagation in two Kenyan populations (Mbololo and Kibwezi) of *M. oleifera;*

iv) estimate minimum rates of outcrossing in *M. oleifera* by use of AFLPs;

v) examine the provenance variation in quantitative and agromorphological traits in *M. oleifera*;

vi) compare genetic diversity and relationships among Kenyan populations of *M. stenopetala* and

vii) estimate the relationship between *M. oleifera* and *M. stenopetala* by the use of AFLP markers.

CHAPTER TWO

2.0 General materials and methods.

2.1 Plant material.

For genetic diversity and relationships between and within populations of *M. oleifera*, DNA was isolated from 2 week old seedlings originating from India (2 populations, one natural and the other, Pkm1, a selection from the wild population), Nsanje (southern Malawi), and four Kenyan provenances (Likoni, Mbololo, Kibwezi, and Kitui) (Fig. 2.1). Pkm1 is a short stem variety which is cultivated as a crop in India. The two Indian populations were obtained from Tamil Nadu (South India). The four Kenya populations are distributed from the coast westwards and each population is separated by a distance of at least 150 km from its nearest neighbour. Kibwezi, Mbololo and Likoni are the major sites of this species in Kenya while the Kitui population consists of very few individuals.

In order to determine the clonal identity of *M. oleifera* in Kenya, DNA was isolated from mature trees sampled from Kibwezi and Mbololo. In order to estimate the minimum rates of outcrossing in *M. oleifera*, DNA was also isolated from 4 mother trees and their 86 progenies sampled from a seed orchard at Mbololo.

For genetic variation in *M. stenopetala*, DNA was isolated from individuals sampled

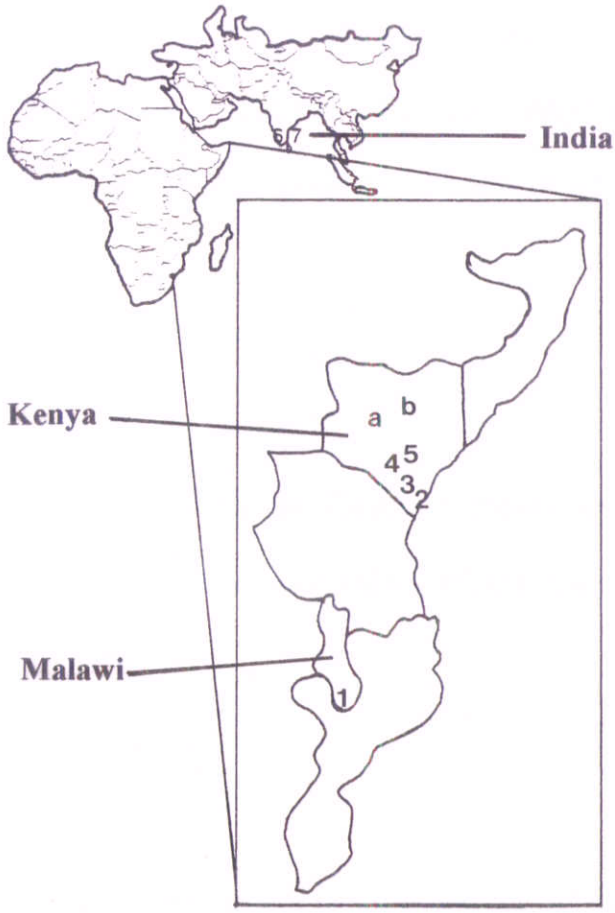
Fig. 2.1 The location of provenances of the two *Moringa* species used for genetic analysis.

M. oleifera

1 Malawi
2 Likoni
3 Mbololo
4 Kibwezi
5 Kitui
6 Pkm1
7 India

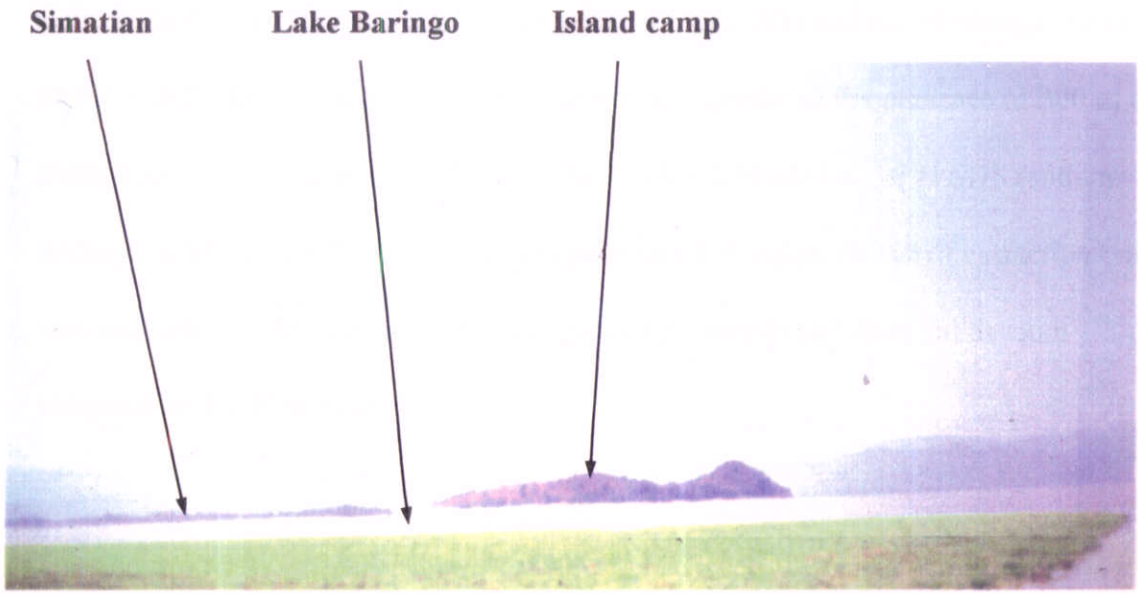
M. stenopetala

a Lake Baringo
b Isiolo



Scale: 1:30,800,000

Fig. 2.2 The two islands within Lake Baringo where *M. stenopetala* individuals were sampled.



from three Kenyan populations, Island Camp (Kokwa Island), Isiolo and Simatian. Kokwa and Simatian are small islands located in Lake Baringo (Fig. 2.2) while Isiolo is situated in the northern part of eastern Kenya (Fig.2.1). In order to examine interspecific relationship between *M. oleifera* and *M. stenopetala*, 70 individuals belonging to 7 populations of *M. oleifera* and 24 individuals from *M stenopetala* were analysed using AFLP.

2.2.0 DNA isolation.

Two published methods (Edwards et al 1991, Murray and Thompson 1980) classified on the basis of the detergent (SDS or CTAB) used were tested and modified to suit the species as described in the subsequent sections.

2.2.1 SDS method.

The method of Edwards et al (1991) was used with major modifications to isolate total DNA from *Moringa.* Two leaves were obtained from 2 weeks old seedlings which had been grown in a greenhouse. The leaves were folded and a lid of a microcentrifuge tube (1.5 ml) shut on them to obtain leaf discs of about 0.1 g. A little sterile sand, insoluble polyvinypolypyrrolidone (PVPP) (also called polyclar AT) and liquid nitrogen were added and the leaf discs ground with a sterile micro pestle in the presence of 200 µl of extraction buffer containing 1 M Tris, 5 M NaCl, 0.5 M EDTA, 10 % SDS (sodium dodecyl sulphate) and 0.0007 % 2 mercaptoethanol. Another 500 µl of extraction buffer was then added. The samples were vortexed for 7 seconds and then left at room temperature for 45 minutes.

The samples were centrifuged at 10,000 rpm for 10 minutes and 500 µl of the supernatant transferred to a clean microcentrifuge tube. An equal volume of chilled chloroform : isoamyl alcohol (24:1) was added. The samples were mixed well to emulsify and then centrifuged at 10,000 rpm for 10 minutes. 300µl of the aqueous phase was transfered to another clean microcentrifuge tube and an equal volume of chilled isopropanol (propan-2-ol) added and the contents mixed well by inversion. The samples were then left at room temperature for 2 minutes before centrifuging at 10,000 rpm for 10 minutes to precipitate the nucleic acids. The supernatant was poured off and the DNA pellets dried under vacuum before resuspending in 200 µl of TE buffer (10 mM Tris, 1 mM EDTA).

In order to remove RNA from the nucleic acid, additional steps were included and involved addition of 2 µl of *Rnase*A (10 mg/ml) to each DNA sample and then incubating at 65°C for 15 minutes. Two volumes (400 µl) of chilled absolute ethanol was added to each sample and then centrifuged at 10,000 rpm for 10 minutes to reprecipitate the DNA. The supernatant was poured off and the DNA pellets dried under vacuum before resuspending in 1000 µl of TE buffer [10 mM Tris (pH 7.5), 1 mM EDTA] and storing at either - 20°C or 4°C.

2.2.2 1.5 X CTAB method.

Total DNA was extracted using the CTAB (hexadecyltrimethylammonium bromide) method (Murray and Thompson 1980). Leaf samples were obtained as described in section 2.2.1. A little sterile sand, insoluble PVPP and liquid nitrogen were added. 200 µl of pre-heated (55 °C) CTAB extraction buffer [1 M Tris (pH 7.5), 1.5 % CTAB, 5 M NaCl, 0.5 M EDTA (ethylenediaminetetra acetic acid), 0.15 % 2 mercaptoethanol (14 M

stock)] were also added and the samples ground using a sterile micro pestle. Another 400 µl of the preheated CTAB extraction buffer was added and the mixture vortexed for 5 seconds before incubating at 55 ^{o} C for 10 minutes.

The supernatant (500 µl) was transferred into a clean microcentrifuge tube and the extract allowed to cool to room temperature before adding an equal volume of chilled chloroform: isoamyl alcohol (24: 1). The samples were mixed well by inversion to emulsify and then centrifuged at 10,000 rpm for 10 minutes. 300 µl of the aqueous phase (the top layer) was transferred to a clean microcentrifuge tube and 1.2 volume of chilled isopropanol (propan-2-ol) added. The contents were mixed by inversion and then centrifuged at 10,000 rpm for 8 minutes to pellet the nucleic acids.

The supernatant was poured off and the DNA pellets washed with 400 µl of chilled 70 % (v/w) ethanol by centrifuging at 10,000 rpm for 1 minute. The DNA pellets were dried under vacuum and then resuspended in 200 µl of TE buffer (10 mM Tris, 1 mM EDTA) then transferred to clean microcentrifuge tubes and stored at -20 °C.

2.3.0 DNA quantitation.

DNA concentration and purity were determined using the following spectrophotometric and ethidium bromide fluorescence method.

2.3.1 Spectrophotometric determination.

A 1:100 dilution of DNA in 1 × TE buffer was made by mixing 5 µl of the DNA solution with 495µl of TE buffer. The absorbance (optical density) of the DNA samples

was measured at 260 nm. DNA concentration of the sample was calculated using the following equation:-

DNA concentration of sample (μg/ml) = A_{260} × Dilution factor × 50.

where A_{260} is the absorbance of the DNA sample at 260 nm.

2.3.2 Ethidium bromide fluorescence.

2 μl of DNA sample was mixed with 0.4 μl of gel loading buffer containing bromophenol blue and then loaded into a slot in a 0.8 % agarose minigel containing ethidium bromide (0.5 μg/ml). 2 μl of each of the series of uncut lambda DNA (0, 2.5, 5, 20, 30, 40, 50 μg/ml) were each mixed with 0.4 μl of the gel loading buffer and loaded into individual wells of the agarose gel. Electrophoresis was carried out until the bromophenol blue migrated approximately 1-2 cm. The gels were photographed using short wavelength ultraviolet irradiation. The intensity of fluorescence of the unknown DNA was compared with that of the DNA standards and the quantity of DNA in the sample estimated.

2.3.3 DNA purity based on optical density (OD).

Because the ratio OD_{260} / OD_{280} for pure DNA preparations is equal to 1.8, the purity for the sample DNA was estimated by taking its absorbance at 260 nm and 280 nm and comparing the resulting ratio with that of pure DNA preparation.

2.4.0 Random amplified polymorphic DNA.

2.4.1 RAPD Reaction.

RAPD reactions were carried out in a 50µl volume containing 15 ng of genomic DNA, 1 µl of 50 X dNTP mix (12.5 mM of each of the dNTPs from Bioline Cat. No. S40200), 1 µl of 0.2µM primer, 0.5 units Taq polymerase (Bioline M95801B), 5 µl of 10 X NH_4 reaction buffer containing 160 mM $(NH_4)_2SO_4$, 670 mM Tris-HCl (pH 8.8 at 25°C), 0.1 % Tween-20 and 1.0 mM $MgCl_2$ (Bioline M95801B). A control that contained all the components of the RAPD reaction except the template DNA was included to check that no self-amplification or DNA contamination occurred. Reactions were overlaid with 50 µl of mineral oil (Sigma M 5904) to prevent evaporation. The RAPD reactions were assembled in a laminar flow hood.

DNA was amplified in a thermal cycler (Techne thermal cycler PHC-1A) using the following programme: 1 minute at 94 °C (denaturation), 1 min at 36 °C (annealing), and 2 min at 72 °C (extension) for 45 cycles followed by an additional extension step of 5 min at 72 °C. The amplified samples were placed at 4 °C until gel analysis was carried out.

2.4.2 Gel electrophoresis and documentation.

Amplification products were separated on a 1.4 % agarose gel using 0.5 X TBE electrophoresis buffer (0.045 M Tris-borate and 0.001 M EDTA). DNA samples (15 µl) were mixed with 6 X gel loading buffer [0.25 % bromophenol blue and 40 % (w/v) sucrose in water]. 3 µl of Lambda EcoR I / Hind III size marker (50 µg DNA / 250 µl)

in a loading buffer containing 10 mM EDTA, 0.04 % bromophenol blue and 5 % glycerol (Advanced Biotechnologies Ltd Cat. No. AB-0199) were loaded to assist in determining the sizes of unknown DNA samples. Electrophoresis was carried out at 117 V. Reproducibility was checked for each primer and DNA sample by running three replicates. The gels were visualised on a long wavelength (310nm) ultraviolet transilluminator and photographed.

2.5.0 Amplified fragment length polymorphism.

The AFLP assays were performed as described by Vos et al (1995) using *Pst*I and *Mse*I restriction enzymes as previously used by Powell et al (1997). A detailed experimental protocol showing the reaction mixture and primer and adaptor sequences used is outlined below.

2.5.1 Restriction digestion of DNA.

30 µl (0.5 µg) of high quality genomic DNA isolated as described in section 2.2.1 was double digested with *Mse*I (frequent 4-base cutter) and *Pst*I (methylation sensitive, rare 6-base cutter) restriction enzymes (New England Biolab) by incubating overnight at 37 ^{0} C with a 10 µl consisting of 0.5µl of *Pst* I (10 U / µl), 1.25 µl of *Mse*I (4 U / µl), 8 µl of 5 × restriction ligation (RL) buffer [1 M Tris- HAc, (pH7.5), 1 M MgAc, 1 M KAc, 0.1 M dithiothreitol, bovine serum albumin (10 mg/ml) and 0.25 µl of sterile distilled water]. 5 µl of each DNA sample and 1 µl of 10 × dye and 4 µl of water were run on a 1.5 % agarose gel to check for complete digestion.

2.5.2 Ligation of adaptors.

Equal volumes of forward and reverse adaptors comprising *Pst*I (5 µM) and *Mse*I (50 µM) were prepared and annealed to make them double stranded by incubating at 65 °C for 10 minutes, 37 °C for 10 minutes and 25 °C for 10 minutes before transferring them to -20 °C. The sequences of the adaptors used were as follows;

*Mse*I adaptors -Forward adaptor 5'-GACGATGAGTCCTGAG-3'

-Reverse adaptor 5'-TACTCAGGACTCAT-3'

*Pst*I adaptors -Forward adaptor 5'-CTCGTAGACTGCGTACATGCA-3'

-Reverse adaptor 5'-TGTACGCAGTCTAC-3'

The ends of the restriction DNA fragments were ligated with the double stranded adaptors. 10 µl of ligation mix consisting 1 µl *Mse* I adapter mix, 1µl *Pst* I adapter mix, 1 µl 10 mM ATP, 2.0 µl of restriction ligation (5x) buffer, 1 µl T4 DNA ligase and 4.0 µl of sterile distilled water was prepared and added to each DNA sample and incubated for 3 hours at 37 °C.

2.5.3 Pre-amplification of DNA.

The PCR preamplification of adaptor-ligated restriction fragments was performed to generate large quantities of PCR products for subsequent selective amplification. This consisted of a PCR reaction using nonselective (zero base pair extension) primers (M00 and P00). The PCR reactions were performed in a 20 µl volume consisting of 2.5 µl of 10 x PCR buffer, 2.5µl of 2 mM dNTPs, 1.0 µl (75 ng)of *Pst*I Primer [P00 (5'GACTGCGTACATGCAG)], 1.0 µl (75ng) of *Mse*I Primer [M00 (5'GATGAGTCCTGAGTAA)], 0.2 µl of Taq polymerase (5 U / µl), 12.8 µl of sterile distilled water and 5 µl of ligated DNA sample. Pre-amplification was carried out at

94 °C for 30 sec, 60 °C for 30 sec and 72 °C for 1 min for 30 cycles.

Preamplification products were diluted and used as starting material for the selective radioactive amplification by adding 50 µl of 0.1 x TE buffer to each sample and mixed gently before storing at -20 °C. 5 µl of pre-amplified DNA was mixed with 1 µl (10 x) loading dye and 4 µl of water and run on 1.5 % agarose gel to check the success of the pre-amplification reaction.

2.5.4 Selective amplification reaction.

Selective amplification was conducted using various combinations of two AFLP primers specific for *Pst*I (P primer) and *Mse*I (M primer) adaptors. The P primers contained two selective nucleotides while M primers contained three selective nucleotides at the 3' end. One of the two primers (M primer) was radioactively labelled using γ ^{33}P ATP and T4 Kinase.

2.5.4.1 AFLP primers.

AFLP analysis used six *Pst*I (P11, P12, P14, P15, P17, P18) and two *Mse*I primers (M51 and M40) as follows:

PstI 11 (P11): 5' GAC TGC GTA CAT GCA G **AA**

PstI 12 (P12): 5' GAC TGC GTA CAT GCA G **AC**

PstI 14 (P14): 5' GAC TGC GTA CAT GCA G **AT**

PstI 15 (P15): 5' GAC TGC GTA CAT GCA G **CA**

PstI 17 (P17): 5' GAC TGC GTA CAT GCA G **CG**

PstI 18 (P18): 5' GAC TGC GTA CAT GCA G **CT**

MseI 151 (M51): 5' GAT GAG TCC TGA GTA A **CCA**

MseI 40 (M40): 5' GAT GAG TCC TGA GTA A **AGC**

The selective nucleotides are shown in bold. An *Mse*I and one *Pst*I primer were selected and the former radioactively labelled. The primer pair combinations used are indicated in each relevant chapter.

2.5.4.2 Labelling primers.

Labelling reactions per sample were performed in a total volume of 1.1 µl consisting 0.1 of µl (0.1 µCi) of (γ) [^{33}P]dATP, 0.134 µl (50 ng/µl) of the primer to be labelled (M primer), 0.025 µl of T4 Kinase, 0.2 µl of the corresponding 5 X Forward reaction buffer and 0.641 µl of sterile distilled water. The resulting mixture was incubated at 37 °C for 1 hour and then heated to 70 °C for 10 min to stop the labelling reaction.

2.5.4.3 Selective PCR.

The *Pst*I primer and *Mse*I primers contained the same sequences as those used in the preamplification but with two and three selective nucleotides, respectively, at the 3' end. The PCR reaction was performed in 20 µl volumes consisting of 1 µl of ^{33} P labelled M primer, 0.5 µl of the unlabelled M primer, 2 µl of 10 x PCR buffer, 2 µl 2 mM dNTPs, 0.6 (50 ng/µl) of the P primer, 0.1 µl of Taq polymerase (5 U / µl), 10.8 µl of sterile distilled water and 3 µl of the preamplified DNA sample.

A Perkin Elmer 9600 PCR machine was used to amplify the mixture for 1 cycle of 94°C for 30 s, 65 °C for 30 s, 72 °C for 60 s, then lowering annealing temperature 0.7 ^{0}C each

cycle for 11 cycles; and then 23 cycles of 94 °C for 30 s, 56 °C for 60 s, 72 °C for 60 s. The start at a very high annealing temperature allows optimal primer selectivity. Gradually decreasing the annealing temperature increases the efficiency of primer binding. At the end of the selective radioactive PCR, the samples were denatured by adding an equal volume (20 µl) of the sequencing loading buffer (98% formamide, 0.2m EDTA, 50mg xylene cyanol, and 50 mg bromophenol blue) and the mixture heated at 94 °C for 5 min then placed directly on ice.

2.5.5 Gel analysis.

Two clean glass plates (small and large) were cleaned well with alcohol. The small plate was treated with a repelcote (BDH laboratory supplies) and the large plate with a binding solution. Spacers were placed on the large plate and the small plate placed on top before tightly clipping them together. 6% polyacrylamide gel was prepared by mixing 75 ml of 6 % easigel (Scotlabs), 500 µl of 10% ammonium persulphate and 50 µl of TEMED (N, N, N', N' tetramethyl ethylenediamine). The gel was poured evenly (avoiding bubbles) at least 2 hours before running.

The gel was prerun at 80W for 30 minutes. 5 µl of the selectively amplified denatured DNA samples were loaded into each track and run for 1 hour 45 minutes at 80W. Gels bound to the plate were fixed in a 10 % acetic acid solution for 20 minutes, washed thoroughly under a running tap and then dried in an oven at 80 ^{0}C for 2 hours or 50 ^{0}C overnight, before exposing to X-ray film for about 5 days. Gels not bound to a plate were transferred to a 3 MM chromatography paper and then dried in the gel dryer (Biorad) for approximately 2 hours at 80 ^{0}C.

2.6.0 Data analysis.

2.6.1 Scoring RAPD and AFLP data.

Amplification products were scored as discrete character states (present / absent). Each polymorphic band was considered as a locus, so that every locus has two alleles, identified by the presence and absence of the band.

2.6.2 Shannon's Index.

The degree of polymorphism was quantified using Shannon's index of phenotypic diversity:

$H_o = - \sum P_i \ln P_i$

where P_i is the frequency of phenotype *i* (King and Schaal 1989). H_o can be calculated and compared for different primers as demonstrated by Wachira et al (1995).

Let

$H_{pop} = 1/n \sum H_o$

be the average diversity over the *n* different populations, and let

$H_{sp} = - \sum p \ln p$

be the diversity calculated from the phenotypic frequencies *n* in all the populations considered together. Then the proportion of diversity present within, H_{pop}/H_{sp}, can be compared with that between populations, $(H_{sp} - H_{pop})/H_{sp}$.

2.6.3 Nei's unbiased statistic.

The AFLP data were transformed into allele frequencies and diversity values were calculated using Nei's unbiased statistic (1987):

$$H = n(1-\Sigma[p_i^2])/(n-1)$$

where n = number of individuals analysed and p_i is the frequency of the ith allele. Diversity values were averaged across the loci.

2.6.4 Nei and Li similarity matrix.

Estimates of similarity between genotypes based on the number of shared amplification products were used to generate a similarity matrix (Nei & Li 1979). The calculations were based on the probability that an amplified fragment from one accession will also be present in another according to the following equation:

Similarity(S_{ab}) = $2.N_{ab} / (N_a + N_b)$; where

N_{ab} = number of shared fragments between genotypes a and b;

N_a = number of fragments in genotype a; and

N_b = number of fragments in genotype b.

2.6.5 Principal co-ordinates, dendrograms and AMOVA analysis.

A matrix of genetic distances between genotypes based on the number of shared amplification products was calculated using the metric of Nei & Li (1979). Average genetic distances within populations were calculated from the distance matrix, which was also used to construct a neighbour-joining phylogram using the NEIGHBOR and DRAWTREE options in the PHYLIP package (V3.57c:Joe Felsenstein, University of Washington, U.S.A).

Principal coordinate analysis based on the similarity matrix was performed with GENSTAT V 5.31 using group average linkage to produce a 3-D plot showing the

relationships between the individuals studied. The similarity matrix was also used to perform a hierarchical analysis of molecular variance (AMOVA; Excoffier et al, 1992) using the ARLEQUIN software (Schneider et al 1997). The same programme was used to generate a matrix of pairwise Fst values which was then used to construct a dendrogram showing the relationships between the populations.

2.6.6 Average heterozygosity, effective multiplex ratio and marker index.

The utility of each marker class was compared by calculating average heterozygosity, (H_{av}), effective multiplex ratio (E), and marker index (MI), as described by Powell et al (1996) for each population and all the populations considered together (i.e species). The expected heterozygosity, H_p, was calculated for each locus from the sum of the squares of allele frequencies. Let

$$H_p = 1 - \Sigma P_i^2$$

where pi is the allele frequency for the *i*-th allele, then

$H_{av} = \beta \Sigma H_p / n_p$; where

$\Sigma H_p / n_p$ = average heretozygosity of polymorphic loci; and

β = fraction of loci that are polymorphic.

The effective multiplex ratio (E), which is the number of polymorphic loci analysed per experiment was calculated for each population and the species as a product of the fraction of polymorphic loci (β) and the number of loci analysed per experiment (n). The marker index (MI), was expressed as the product of the effective multiplex ratio (E) and the average expected heterozygosity for the polymorphic loci $H_{av(p)}$.

2.6.7 Multiband fingerprint similarity.

The probability that two trees have an identical multiband fingerprint was calculated using X^m, where X is the proportion of shared bands and *m* the number of bands amplified (Jeffreys et al 1985, Georges et al 1988).

2.6.8 Estimating outcrossing rates.

Outcrossing rates were estimated using MLDT programme (Ritland 1990) for dominant markers.

2.6.9 ANOVA analysis

Analysis of variance (ANOVA) for quantitative and morphological traits was carried out using the Genstat statistical package (1995).

CHAPTER THREE

3.0 Genetic variation and relationships in *M. oleifera* populations revealed by multi-locus assays (RAPDs and AFLPs).

3.1 Introduction.

M. oleifera is indigenous to North West India, (Jahn et al 1986, Kantharajah and Dodd 1991), but was introduced to Kenya from India (exact location unknown) at the beginning of this century (Jahn 1991). Although detailed studies of the distribution of genetic variability are limited, considerable variation in quantitatively inherited traits has been reported in natural populations from India (Ramachadran et al 1980) and introduced populations in Kenya (unpublished data), indicating considerable potential for improvement. However, little information is available on the genetic base of Kenyan populations. Such information would facilitate tree improvement programmes and the conservation and exploitation of *M. oleifera* genetic resources.

Genetic diversity within populations is of great concern to conservation biologists because paucity of genetic variation is thought to reduce the ability of populations to adapt to changing environments, thereby causing their extinction (Beardmore 1983, Maki et al 1996). The patterns of genetic variation within and among populations can be influenced by mutation, genetic drift, the mating system, gene flow, and selection (Slatkin 1987). Colonising species usually exhibit an extreme pattern of population

structure, with very low within and high between population variation (Hamrick 1989). The pattern of genetic structure probably results from a variety of phenomena, including the founding of populations by a few individuals, low levels of repeated migration due to geographic isolation, novel selection in new habitats, and a propensity to self-fertilise (Barret and Shore 1989, Kercher and Conner 1996). Low levels of genetic variability are common for rare or geographically restricted plant species (Smith and Pham 1996).

Traditionally, morphological and agronomic traits have been used to measure genetic diversity (Wilde et al 1992, Karp et al 1996). Unfortunately, these characters may be influenced by environmental factors and may therefore not reflect true genetic similarity or differences (Dawson et al 1993, Tingey and del Tufo 1993). The long generation time of most perennial crops also indicates that many of the morphological descriptors can only be assessed at maturity (Wilde et al 1992). Molecular markers that are not subject to environmental influences have proved to be powerful tools in the assessment of genetic variation both within and between plant populations by analysing large number of loci distributed throughout the genome (Sharma et al 1995, Loarce et al 1996). In plant breeding programmes, information concerning the genetic diversity within a species is essential for efficient use of genetic resources (Sharma et al 1995, Loarce et al 1996), characterisation of individual accessions, in detecting duplication of genetic material in germplasm collections, and as a guide in the choice of parents for breeding programmes. High levels of genetic variation are important in safeguarding against biotic and non-biotic factors, especially in long lived crops such as trees (Van de Ven and McNicol 1995).

Random amplified polymorphic DNA (RAPD) (Welsh and McClelland 1990, Williams

et al 1990) and amplified fragment length polymorphism (AFLP) (Vos et al 1995) methods were chosen in this study. The RAPD technique has been used successfully to analyse genetic variation in a number of woody species (Chalmers et al 1992, Nesbitt et al 1995, Wachira et al 1995). The AFLP technique has been used successfully in diversity studies (Maughan et al 1996, Sharma et al 1996, Travis et al 1996, Ellis et al 1997, Hongtrakul et al 1997, Paul et al 1997, Perera et al 1998, Winfield et al 1998).

Selection of a DNA marker system depends on project objectives, population structure, the genomic diversity of the species under investigation, marker system availability, time required for analysis, and the cost per unit information (Staub et al 1994). Clearly, each marker system has advantages and disadvantages, and therefore it is critical to evaluate each marker system for its potential utility before use. Marker systems also differ in their utility across populations, species, and genera, and their efficacy in the detection of polymorphism (Staub et al 1994, Karp et al 1996). Two important aspects of a marker system's utility are the information content and multiplex ratio (Powell et al 1996). Standard measures of diversity may be used to evaluate information content and the multiplex ratio is the number of loci simultaneously analysed per experiment. These two metrics have been used to compare RFLPs, AFLPs, SSRs and RAPDs in common soybean, barley and potato germplasm (Powell et al 1996, Russell et al 1997, and Milbourne et al 1997). A combination of multiplex ratio and the diversity index provides an overall measure of marker utility defined as the Marker Index (Powell et al 1996).

To my knowledge molecular markers have not been previously utilised to analyse genetic variation in *Moringa* species. The specific objectives of the present study were

to: 1) establish the extent and distribution of genetic variation within and between populations of *M. oleifera*; 2) assess genetic relationships between introduced African populations and native material from India and; 3) examine and compare the utility of RAPDs and AFLP markers for genetic analysis in *M. oleifera.*

3.2 Materials and methods.

3.2.1 Plant material.

Seventy individuals of *M. oleifera* from 7 different populations described in section 2.1 were used for both AFLP and RAPD analyses. The number of individuals per population was then doubled and the AFLP technique deployed to analyse genetic diversity using a reduced number of primer pairs. Genomic DNA was isolated from young seedlings obtained from trees randomly chosen from each population.

3.2.2 DNA isolation technique.

Genomic DNA was isolated following modification of Edwards et al (1991) method as described in section 2.2.1

3.2.3 RAPD protocol.

The RAPD amplification reaction and gel electrophoresis and documentation were carried out as described in section 2.4.0.

3.2.4 The AFLP procedure.

The AFLP assays were performed as described d in section 2.5.0.

3.2.5 Data analysis.

Amplification products were scored as discrete character states (present / absent) and analysed as described in section 2.7.0.

3.3 Results.

3.3.1 RAPD polymorphism and genetic relationships using 70 individuals.

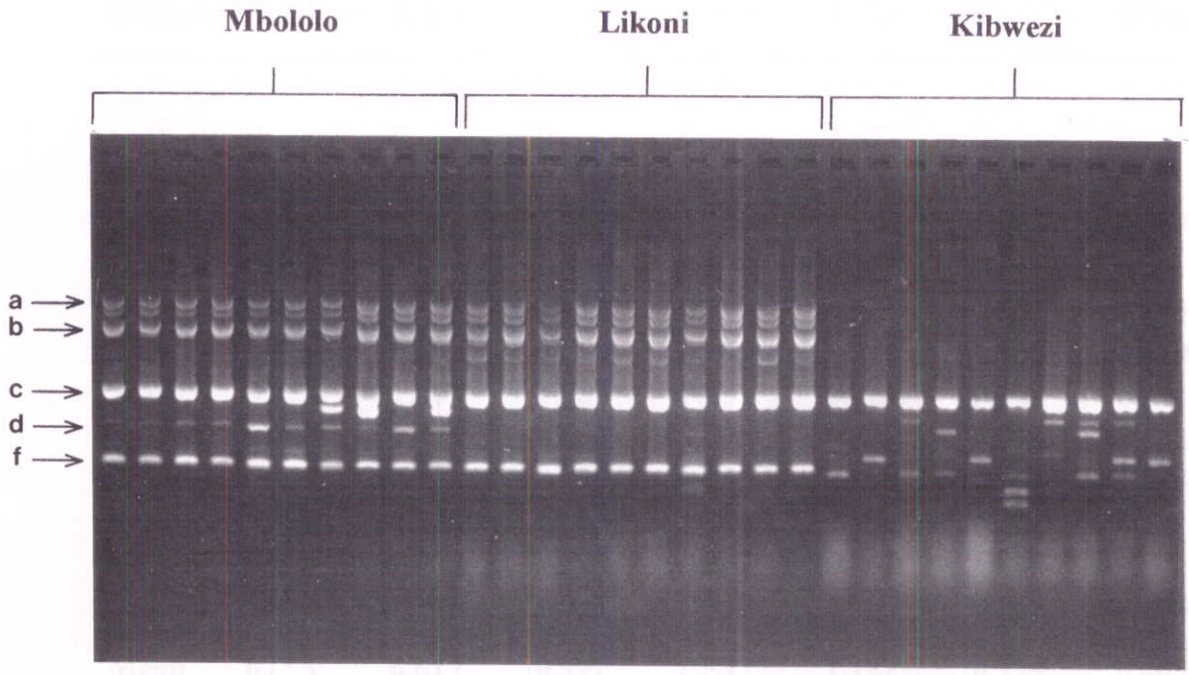
The 13 RAPD primers used generated a total of 137 amplification products, of which 127 (92.7 %) were polymorphic (Table 3.1). The number of polymorphic fragments generated by each primer ranged from 5 (Sc10-17) to 13 (SC10-19) with an average of 9.77 polymorphic fragments per primer (Table 3.1). An example of the molecular profile generated by one primer (SC10-25) is shown in Fig 3.1. Primers differed in their capacity to detect polymorphism (Table 3.1). Measures of intrapopulation variability based on the proportion of polymorphic products scored in a single population ranged from 18.1 % for the Malawi population to 49.6 % for the Indian (natural) collection. The accessions derived from the Kenyan collections showed intrapopulation variability ranging from 21.3 % (Kitui) to 25.2 % (Kibwezi).

The phenotypic frequencies detected with the 13 primers were calculated and used in estimating genetic diversity (H_o) within populations (Table 3.2). The Indian sample (natural) had the highest (1.152) mean estimate of genetic diversity (H_o) while Malawi had the lowest (0.408).

Table 3.1 **Amplification products of 7 provenances of *M. oleifera* generated with RAPD primers.**

	Primer		Amplification products			No. of polymorphic products						
No.	Code	Sequence (5'-3')	Total	Polymorphic	%	Kibwezi	Likoni	Mbololo	Kitui	India	Pkm1	Malawi
1	Sc10-23	GGCTCGTACC	9	8	89	3	1	1	2	4	3	5
2	Sc10-19	CGTCCGTCAG	13	13	100	0	0	0	0	9	0	0
3	Sc10-17	GTTAGCGGCG	6	5	83	0	0	0	0	4	2	0
4	Sc10-25	CGGAGAGTAC	13	12	92	4	0	1	4	5	4	1
5	Sc10-16	CCTGGCGAGC	9	8	89	1	0	0	2	0	0	0
6	Sc10-14	TCCCGACCTC	8	8	100	2	4	6	2	3	3	1
7	Sc10-64	CCAGGCGCAA	11	9	82	1	2	1	1	7	4	4
8	Sc10-59	GCATGGAGCT	12	11	92	4	7	3	3	5	1	1
9	Sc10-30	CCGAAGCCCT	12	12	100	7	7	3	4	4	3	3
10	Sc10-24	ACCCATGCGG	12	12	100	1	0	1	1	8	9	4
11	Sc10-5	TCGGAGTGGC	11	11	100	3	3	9	4	7	0	0
12	Sc10-58	CGGGAGACCC	10	7	70	0	3	3	1	1	7	0
13	Sc10-22	GTAGGCGTCG	11	11	100	6	4	3	3	6	3	4
Total products			**137**	**127**		**32**	**31**	**31**	**27**	**63**	**39**	**23**
% polymorphism					**93**	**25.2**	**24.4**	**24.4**	**21.3**	**49.6**	**30.7**	**18.1**

Fig. 3.1 An example of a molecular profile generated from *M oleifera* populations using RAPD primer Sc10-25.



Product sizes (bp).

a 5,000	d 2,000
b 4,000	e 1,750
c 2,900	f 1,500

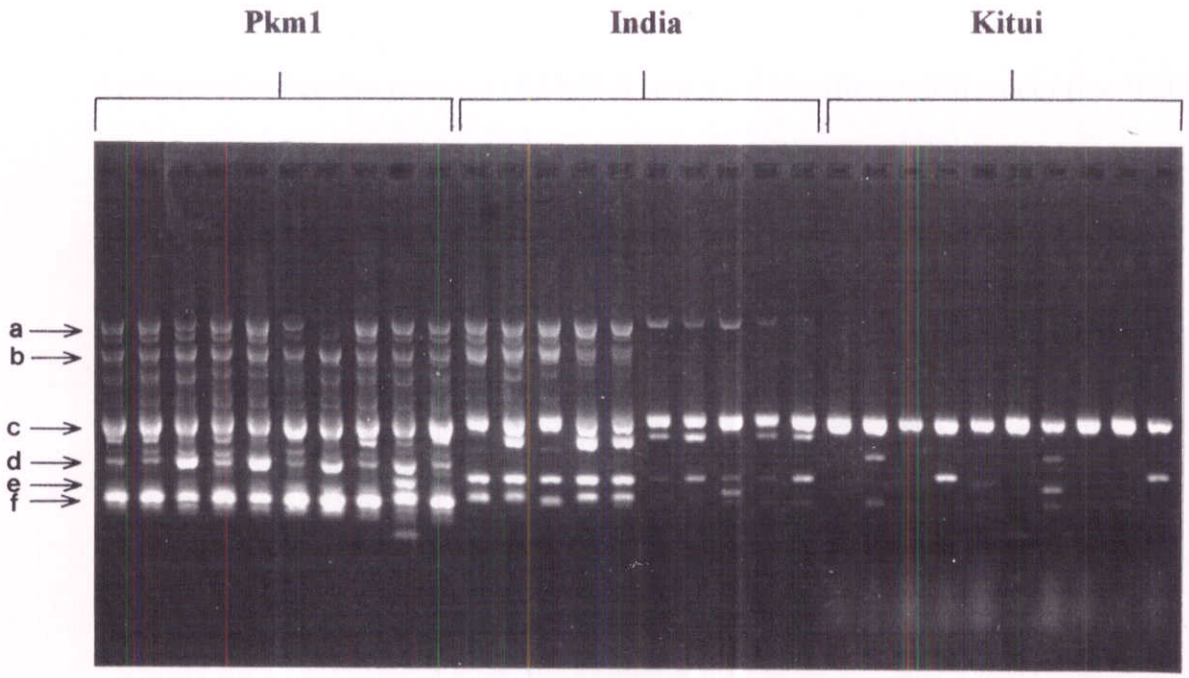


Table 3.2 **Estimates of genetic diversity (H_0) within populations of *M. oleifera* based on RAPD markers.**

Primer	Provenances						
	Kibwezi	Likoni	Mbololo	Kitui	India	Pkm1	Malawi
Sc10-23	0.787	0.179	0.322	0.728	0.734	0.691	1.231
Sc10-19	0	0	0	0	2.473	0	0
Sc10-17	0	0	0	0	0.735	0.556	0
Sc10-25	1.356	0	0.361	1.235	1.596	1.067	0.367
Sc10-16	0.307	0	0	0.273	0.273	0	0
Sc10-14	0.480	1.125	1.412	0.577	0.775	0.551	0.095
Sc10-64	0.361	0.556	0.250	0.367	1.576	0.463	0.870
Sc10-59	0.715	1.803	1.089	0.651	0.784	0.322	0.321
Sc10-30	1.808	1.476	0.595	1.096	0.618	0.918	0.632
Sc10-24	0.367	0	0.230	0.250	2.090	2.967	1.096
Sc10-5	0.878	0.788	2.106	1.324	1.916	0	0
Sc10-58	0	0.980	0.285	0.307	0.367	1.878	0
Sc10-22	1.978	1.141	1.035	0.950	1.035	0.715	0.6901
Mean	**0.695**	**0.619**	**0.591**	**0.597**	**1.152**	**0.779**	**0.408**

Partitioning of the phenotypic diversity into within and between population components using Shannon's index of phenotypic diversity (Table 3.3) indicates that on average, most of the variation in *M. oleifera* (69 %) occurs between populations.

Analysis of molecular variance (AMOVA) detected significant differences ($p < 0.0001$) in variation between the African and Indian groups (12.8%), as well as between populations within groups (49.2%) (Table 3.4). The between population variation accounted for 59.1% while the within population variation was 40.9%. Most of the between provenance variation is therefore accounted for by differences between provenances within groups.

Table 3.3 **Partitioning of genetic diversity between and within populations of *M.oleifera* based on 13 RAPD primers.**

Primer	H_{pop}	H_{sp}	Within populations $[H_{pop}/H_{sp}]$	Between Populations $[(Hsp-H_{pop})/H_{sp}]$
Sc10-23	0.667	1.2175	0.548	0.452
Sc10-19	0.353	2.992	0.118	0.882
Sc10-17	0.184	0.911	0.202	0.798
Sc10-25	0.855	3.768	0.227	0.773
Sc10-16	0.122	2.332	0.052	0.948
Sc10-14	0.716	1.820	0.394	0.606
Sc10-64	0.635	1.575	0.403	0.597
Sc10-59	0.812	1.978	0.412	0.589
Sc10-30	1.020	2.740	0.373	0.628
Sc10-24	1.000	3.27	0.305	0.695
Sc10-5	1.002	3.722	0.269	0.731
Sc10-58	0.545	1.255	0.434	0.566
Sc10-22	1.078	3.323	0.324	0.676
Mean	**0.691**	**2.377**	**0.312**	**0.688**

Table 3.4 **Analysis of molecular variance (AMOVA) based on RAPDs.**

Source of variation	df	Variance components: Variance	% Variation	Probability
Between provenances	6	11.98	59.1	$P < 0.00001$
Between groups (India Vs Africa)	1	2.82	12.8	$P < 0.00001$
Between provenances within groups	5	10.81	49.0	$P < 0.00001$
Within provenances	63	8.42	40.9	$P < 0.00001$

The principal co-ordinate analysis (PCOa) (Fig. 3.2A) separated Kenyan and the Indian plus Malawi populations into two distinct groups and further grouped Kibwezi with Kitui and Mbololo with Likoni. A PCOa based on cluster analysis for all the 70 individuals (Fig. 3.2B) grouped Mbololo and Likoni, and Kibwezi and Kitui individuals very closely.

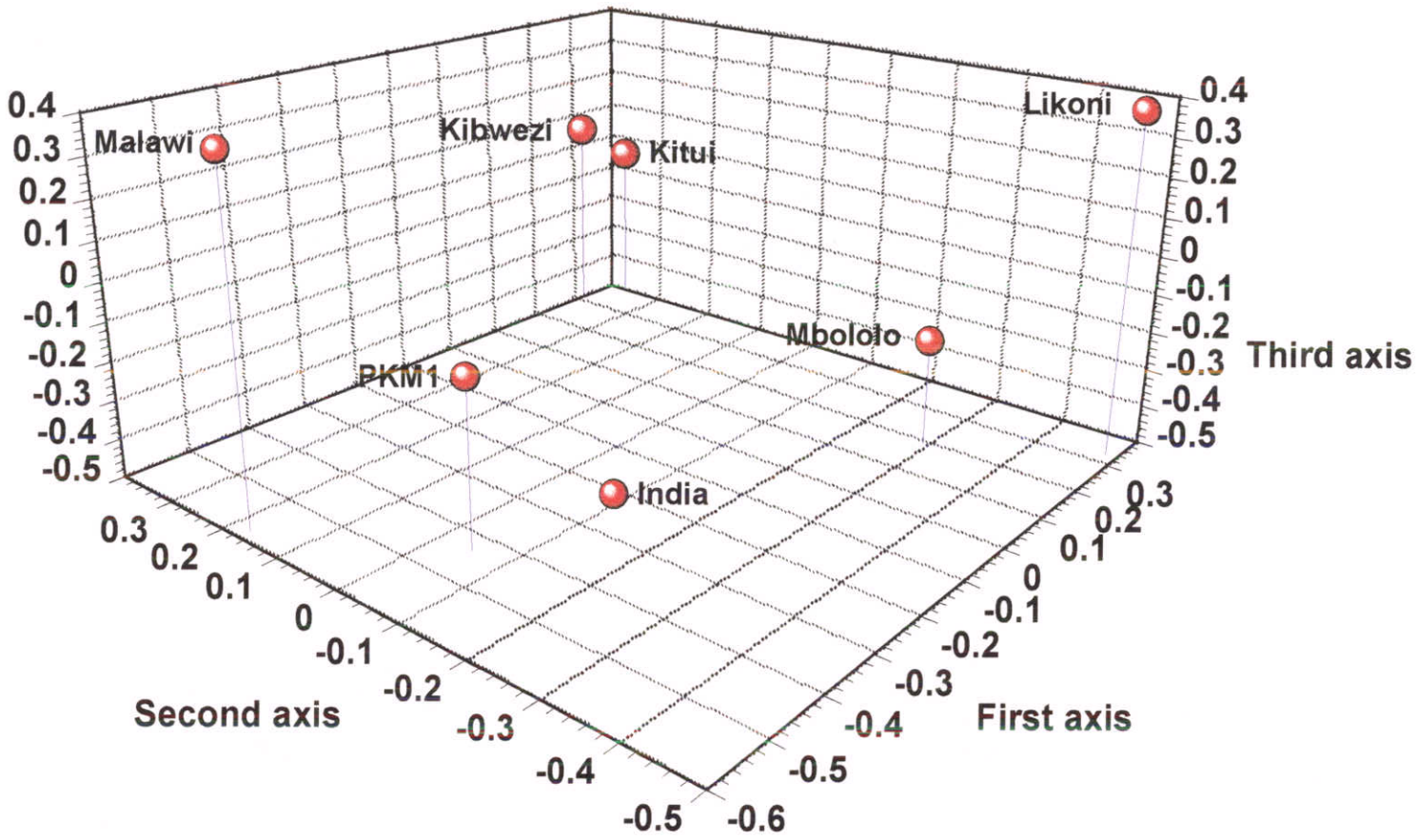


Fig. 3.2 (A) 3-Dimensional principal co-ordinate analysis based on RAPDs showing the clusteing of the 7 *M. oleifera* provenances.

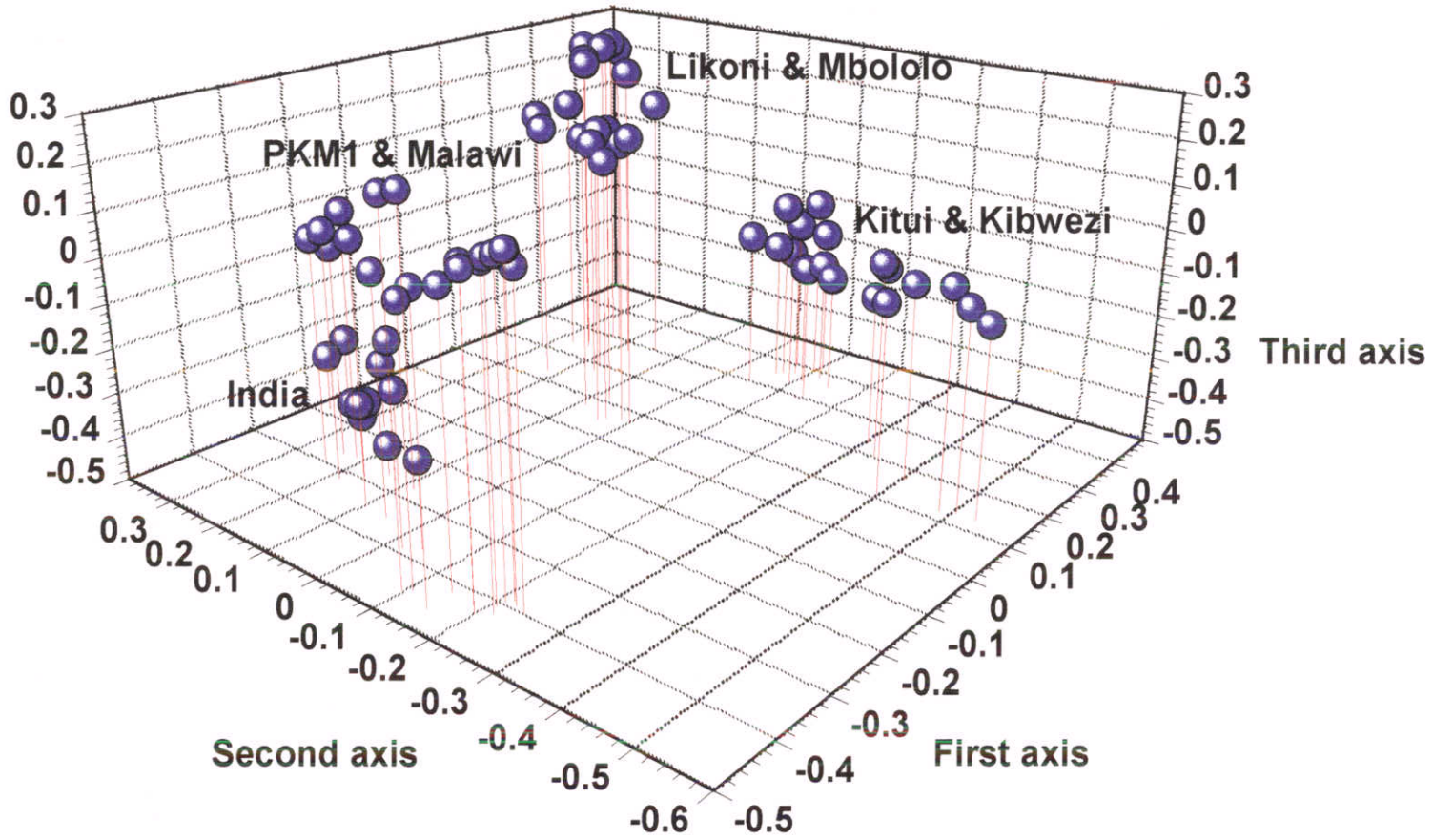


Fig. 3.2 (B) 3-Dimensional principal co-ordinate analysis based on RAPDs showing the relationship between the 70 individuals from seven provenances of *M. oleifera.*

Dendrograms based on pairwise Fst comparisons of 278 AFLPs and 127 RAPDs among the seven populations are shown in Fig. 3.3 (A) and 3.3 (B), respectively. The RAPD dendrogram separated Kenyan and Indian plus Malawi populations into two distinct groups and further clustered Kibwezi with Kitui and Mbololo with Likoni. Generally, individuals were grouped according to their respective populations (Fig. 3.4). However, Pkm1 individuals formed two clusters while one individual from Kitui was grouped with those from Kibwezi.

3.3.2 AFLP polymorphism and genetic relationships using 70 individuals.

Analysis of the 70 accessions of *M. oleifera* with 7 AFLP primer pairs identified a total of 463 fragments, of which 278 (60.0%) were polymorphic between two or more accessions (Table 3.5). Examples of typical AFLP variation for sections of two primer pairs in samples of *M. oleifera* individuals is shown in Fig. 3.5. Polymorphic fragments were generated by each of the 7 primer pairs. The number of fragments detected by individual primer pairs ranged from 32 (for the primer pair P17-M51) to 88 (P15-M51) (Table 3.5). The number of polymorphic fragments for each primer pair varied from 17 (P17-M51) to 54 (P15-M51) with an average of 39.7 polymorphic fragments per primer pair (Table 3.5).

Different primer combinations detected different levels of polymorphism ranging from 40.6% (P17-M51) to 77.6% (P12-M51) (Table 3.5). Measures of intrapopulation variability ranged from 16.2% (Kitui) to 50% (India). Phenotypic frequencies were calculated and mean estimates of genetic diversity (Table 3.6) within populations

Fig. 3.3 A dendrogram based on AFLP (A) and RAPD (B) analyses showing the clustering of *M. oleifera* provenances.

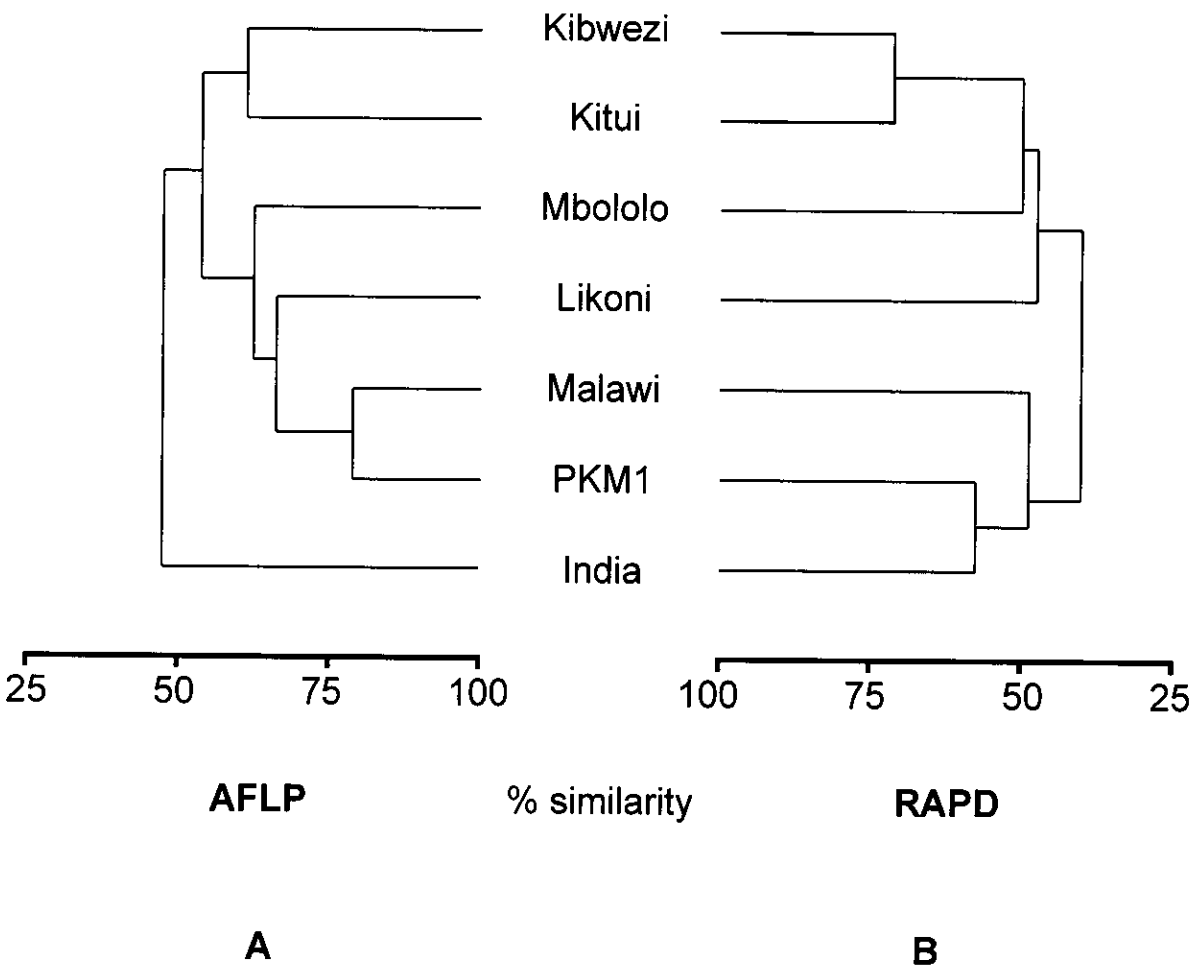
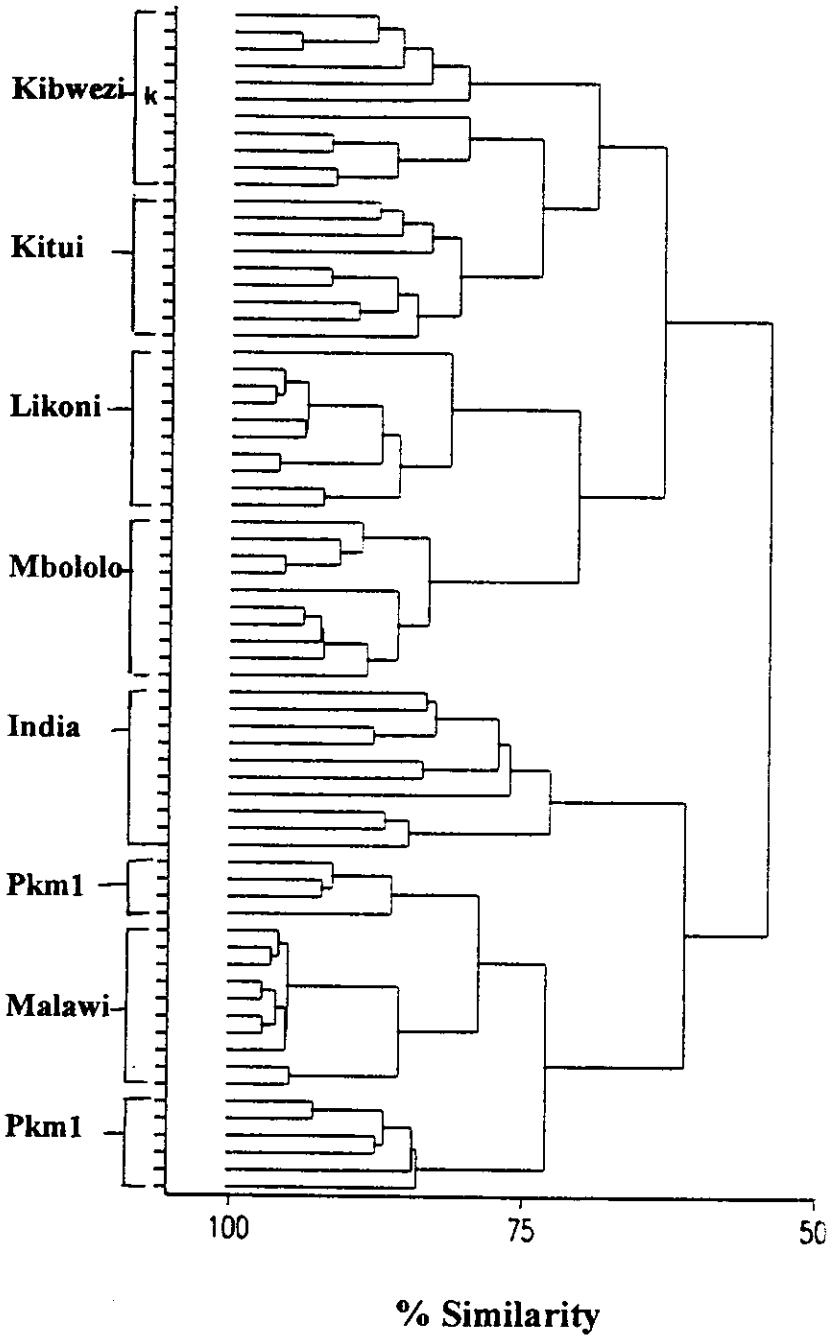


Fig. 3.4 A dendrogram based on RAPD showing clustering of 70 individuals from the seven provenances of *M. oleifera*.

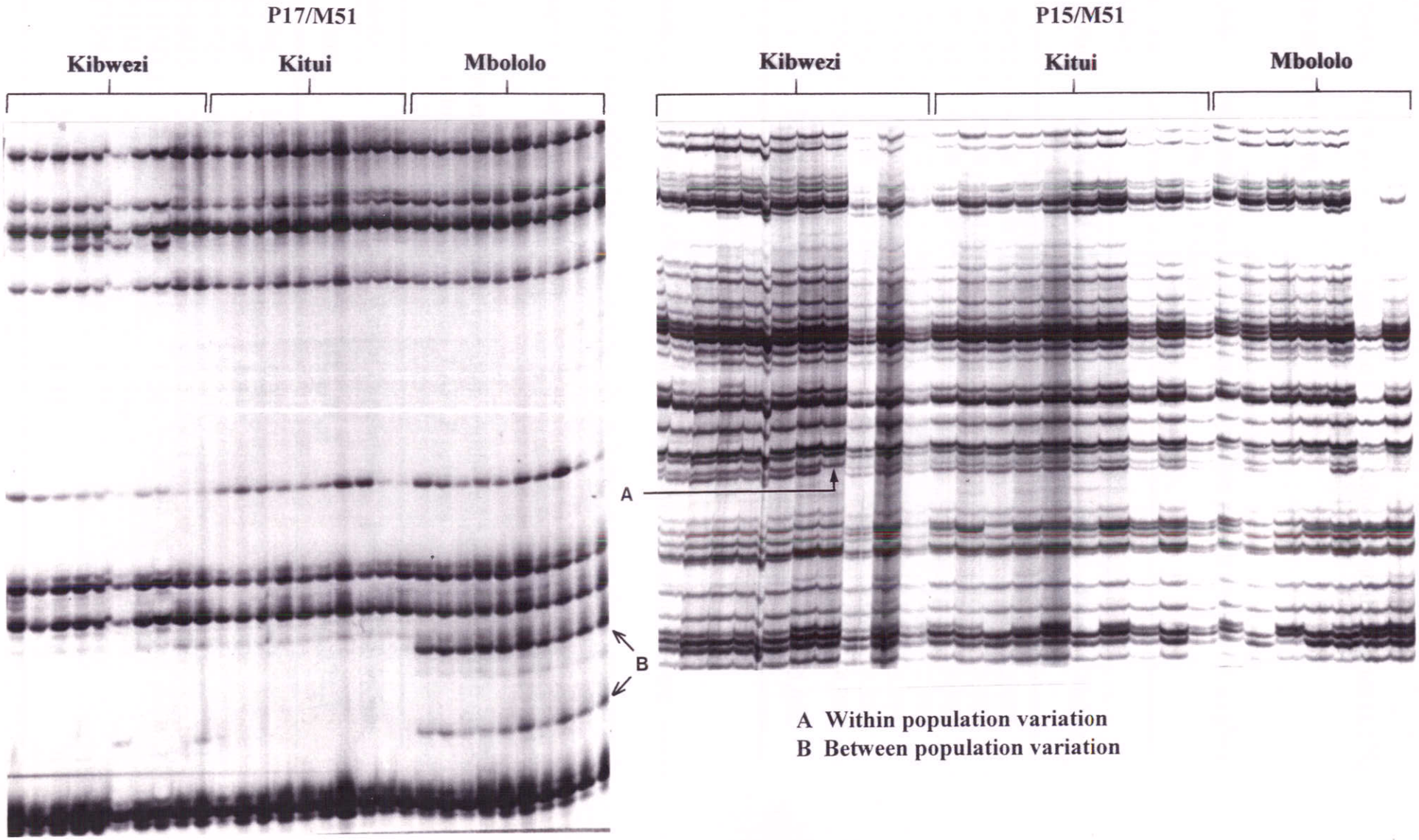


K shows an individual from Kitui provenance which was grouped with those from Kibwezi.

Table 3.5 **Amplification products from 7 provenances of *M. oleifera* based on 7 AFLP primers.**

Primer		Amplification products			Provenances						
No	**Pair**	**Total**	**Polymorphic**	**%**	**Kibwezi**	**Likoni**	**Mbololo**	**Kitui**	**India**	**Pkm1**	**Malawi**
1	P14M51	58	36	62.1	13	3	5	4	13	21	7
2	P12M51	58	45	77.6	9	9	16	5	31	15	12
3	P17M51	32	13	40.6	2	6	0	1	4	6	4
4	P18M51	87	41	47.1	16	12	9	9	27	13	8
5	P14M40	57	44	77.2	15	20	17	5	26	22	15
6	P11M51	83	45	54.2	14	12	16	17	21	24	18
7	P15M51	88	54	61.4	24	18	8	4	17	17	17
Total		**463**	**278**	**60.0**	**93**	**80**	**71**	**45**	**139**	**118**	**81**
% Polymorphism					**33.5**	**28.8**	**25.5**	**16.2**	**50**	**42.4**	**29.1**

Fig. 3.5 Examples of typical AFLP polymorphism generated from *M. oleifera* using primer pair P17/M51 and P15/M51.



A Within population variation
B Between population variation

determined. Overall, the Indian natural population had the highest level (4.593) of within population variability while among the Kenyan, Kibwezi had the highest (3.171) and Kitui the least (1.501).

Table. 3.6 Estimates of genetic diversity (H_o) within populations of *M. oleifera* based on AFLP markers.

Primer		Provenances						
No	Pair	Kibwezi	Likoni	Mbololo	Kitui	India	Pkm1	Malawi
1	P14M51	4.156	0.976	1.439	1.037	2.252	5.383	2.027
2	P12M51	2.340	1.896	4.897	0.868	7.486	4.651	2.557
3	P17M51	0.500	0.951	0.000	0.095	0.715	1.337	0.909
4	P18M51	4.252	3.369	2.664	1.822	7.217	3.459	2.389
5	P14M40	3.338	3.231	2.476	1.463	4.303	5.169	3.751
6	P11M51	3.770	3.106	4.410	4.530	5.824	6.096	5.258
7	P15M51	3.840	3.000	1.590	0.690	4.355	4.065	4.260
Mean		**3.171**	**2.361**	**2.496**	**1.501**	**4.593**	**4.309**	**3.022**

Partitioning of the phenotypic diversity into within and between population components indicated that most of the AFLP variation was due to between population variation with an average value of 57.6% (Table 3.7).

Table 3.7 Partitioning of genetic variation within and between populations using Shannon's index for 7 AFLP primers.

Primer				Within Populations	Between Populations
No	Pair	H_{pop}	H_{sp}	$[H_{pop}/H_{sp}]$	$[(H_{sp}-H_{pop})/H_{sp}]$
1	P14M51	2.467	7.448	0.331	0.669
2	P12M51	3.528	8.430	0.419	0.582
3	P17M51	0.644	2.427	0.265	0.735
4	P18M51	3.596	7.017	0.513	0.488
5	P14M40	3.390	6.248	0.543	0.457
6	P11M51	4.713	8.549	0.551	0.449
7	P15M51	3.114	9.002	0.346	0.654
Mean		**3.065**	**7.017**	**0.424**	**0.576**

AMOVA analysis (Table 3.8) showed significant differences ($p < 0.0001$) in variation between the African and Indian populations (14.3%), as well as between populations (provenances) within groups (37.6%). The between population component accounted for 48.1% while the within population variation was 51.9%.

Table 3.8 Analysis of molecular variance (AMOVA) based on AFLP markers.

		Variance components		
Source of variation	**df**	**Variance**	**% variation**	**Probability**
Between provenances	6	5.35	48.1	P < 0.00001
Between groups (India vs. Africa)	1	1.73	14.3	P < 0.00001
Between provenances within groups	5	4.53	37.6	P < 0.00001
Within provenances	63	5.78	51.9	P < 0.00001

A three dimensional principal co-ordinate analysis (Fig. 3.6A) grouped together the two Indian (India and Pkm1) populations. The Kenyan populations formed two groups, one containing Kibwezi and Kitui and the other Likoni and Mbololo populations (Fig. 3.6A). The Malawi population was grouped closer to the Indian than the other African populations. A PCO based on cluster analysis for all the seventy genotypes (Fig. 3.6B) did not clearly separate Mbololo individuals from Likoni, Kibwezi from Kitui or Pkm1 from Malawi.

A dendrogram (Fig. 3.3A, p56) generated by population pairwise Fst comparisons of 278 polymorphic AFLPs scored over the 7 populations grouped the Indian (natural) as a unique group and then split the other populations into two groups, one with Kibwezi and Kitui, and the other with the rest of the populations. Pkm1 was further associated

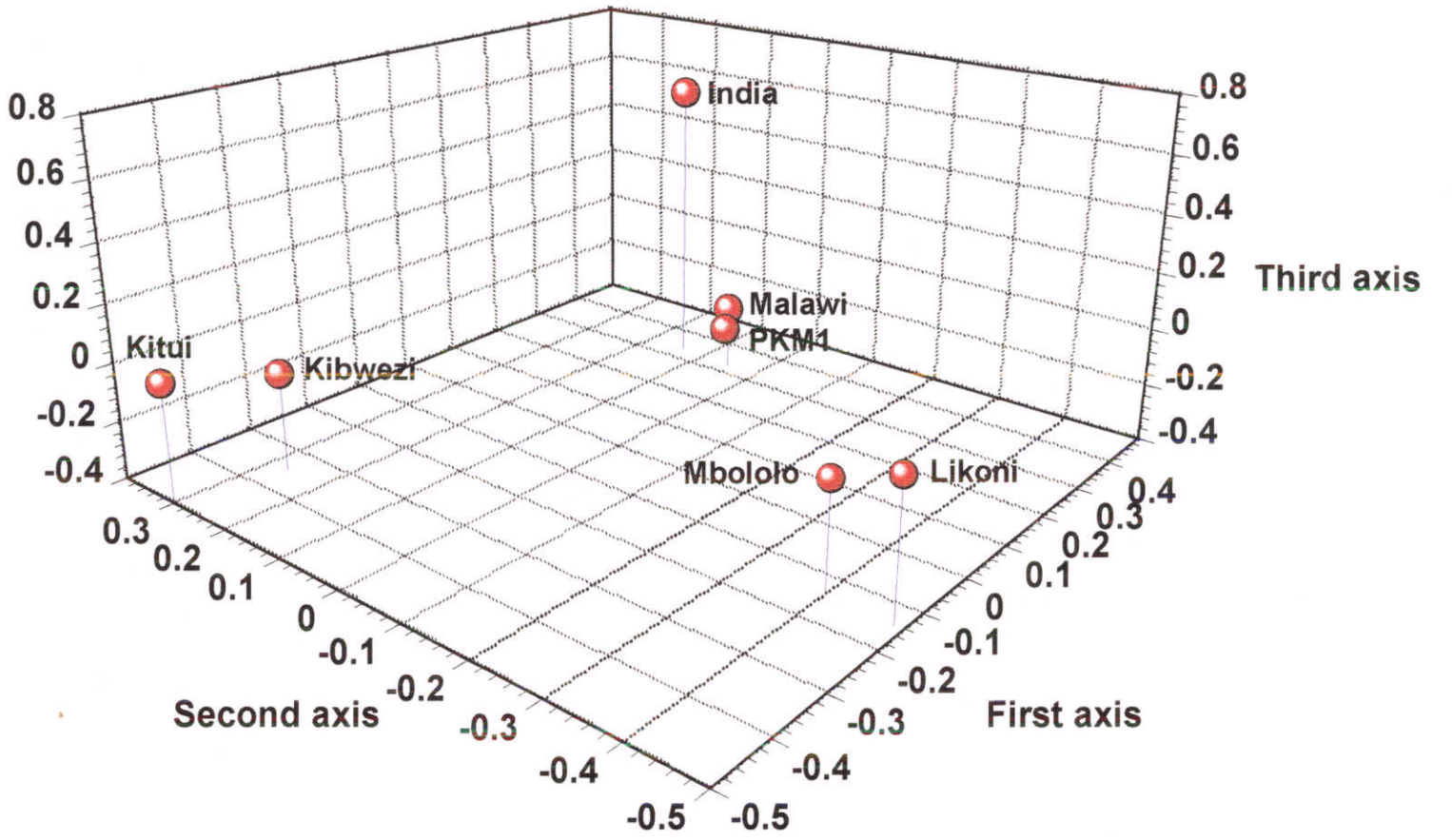


Fig. 3.6 (A) 3-Dimensional principal co-ordinate analysis based on AFLPs showing clustering of the 7 provenances of *M. oleifera.*

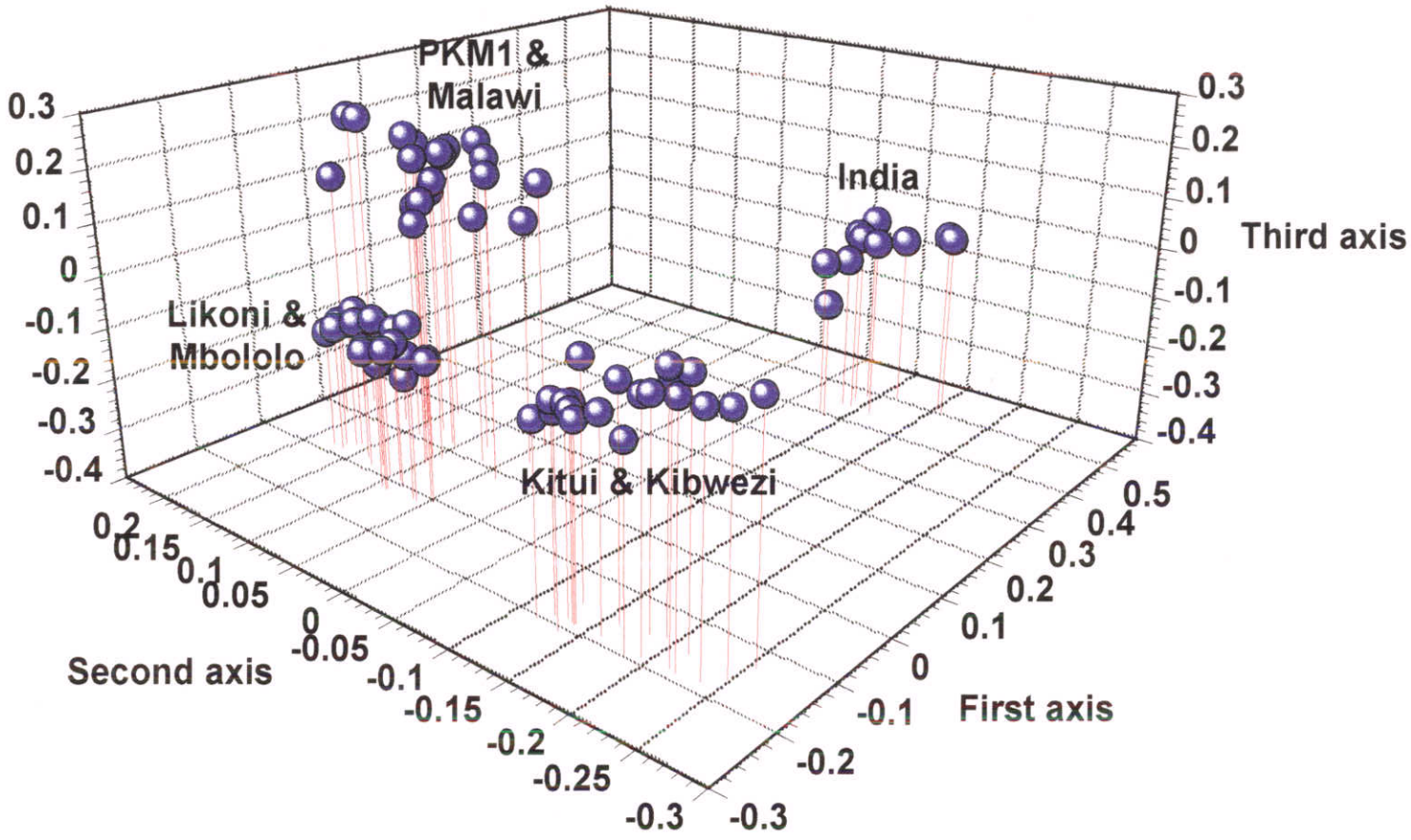


Fig. 3.6 (B) 3-Dimensional principal co-ordinate analysis based on AFLPs showing clustering of the 70 individuals of *M. oleifera* from the 7 provenances.

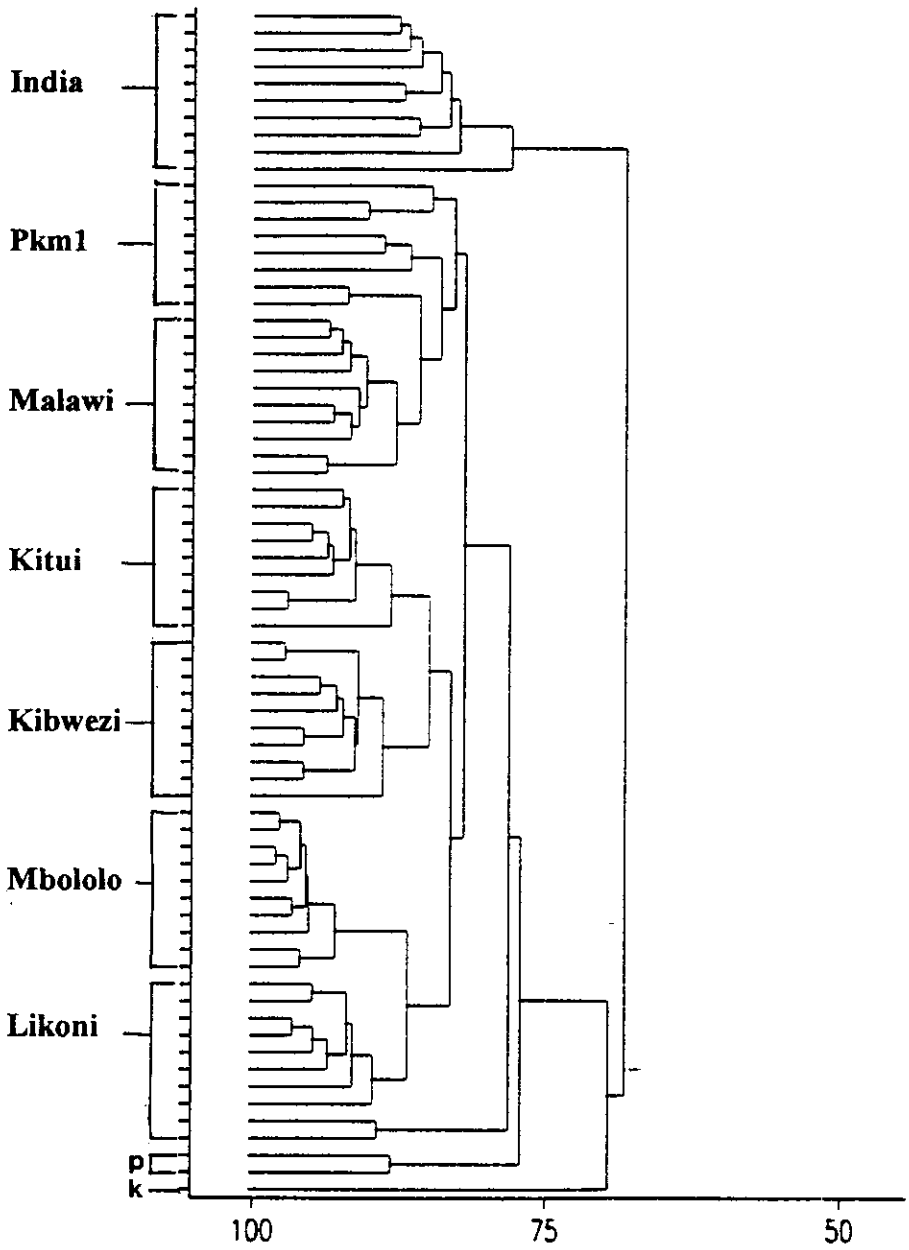
with Malawi while Likoni was close to Mbololo. A dendrogram based on cluster analysis for all the seventy individuals grouped them according to their respective populations (Fig. 3.7). However, one individual from Kitui and two from Pkm1 were not grouped with members of their respective populations.

3.3.3 Comparison between RAPDs and AFLPs.

The average heterozygosity (H_{av}), the marker index (MI) and the effective multiplex ratios (E) were calculated for each primer and assay and averaged for comparison for each population and all the populations considered together. The highest levels of polymorphism (mean average heterozygosity) were detected with RAPDs (Table 3.9) for all the populations (Table 3.10). The average expected heterozygosity value for the species was 0.273 for RAPDs while AFLPs had 0.139 (Table 3.9 and 3.10 respectively). AFLP had the highest values of MI in all the populations and the species (3.656) compared to 2.612 for RAPD. However, three populations (Likoni, Mbololo and Kitui) had higher MI values with RAPD than AFLP. The relatively higher MI values for AFLP compared to RAPD, was due to the high effective multiplex ratio (E) (Table 3.10).

To compare the RAPDs and AFLP patterns of variation, procrustes rotation (Chatfield and Collins 1980) was performed on the first three principal co-ordinate dimensions of each assay. At the individual level (n=70), rotating the AFLP scores onto the RAPD scores accounted for 69.5 % of the variation which suggests a good degree of congruence between the results obtained using RAPDs and those obtained with AFLPs. When the analysis was repeated using the means of pairwise distances between populations (n=7) this value rose to 84.2 % indicating a high level of similarity between the molecular assays at the population level.

Fig. 3.7 A dendrogram based on AFLP showing the clustering of 70 individuals from the seven provenances of *M. oleifera*.



P and K show individuals from Pkm1 and Kitui provenances respectively.

Table 3.9 **Average heterozygosity (H_{av}), effective multiplex ratio (E), and marker index (MI) for 7 populations of *M. oleifera* based on RAPD.**

	Kibwezi			Likoni			Mbololo			Kitui			India			Pkm1			Malawi			Species		
Primer	**Hav**	**E**	**MI**	**Hav**	**E**	**MI**	**Hav**	**E**	**MI**	**Hav**	**E**	**MI**	**Hav**	**E**	**MI**	**Hav**	**E**	**MI**	**Hav**	**E**	**MI**	**Hav**	**E**	**MI**
P23	0.127	1	0.127	0.04	0.11	0.004	0.107	0.11	0.012	0.107	0.44	0.047	0.096	1.78	0.170	0.060	1	0.06	0.331	2.78	0920	0.139	7.11	0.987
P19	0	0	0	0	0	0	0	0	0	0	0	0	0.271	6.23	1.687	0	0	0	0	0	0	0.234	13	3.043
P17	0	0	0	0	0	0	0	0	0	0	0	0	0.193	2.67	0.516	0.150	0.67	0.1	0	0	0	0.191	4.17	0.796
P25	0.131	1.23	0.161	0	0	0	0.001	0.08	0.000	0.095	1.23	0.117	0.185	1.92	0.355	0.120	1.23	0.148	0.07	0.08	0.005	0.369	11.08	4.093
P16	0.071	0.11	0.008	0	0	0	0	0	0	0.056	0.44	0.025	0	0	0	0	0	0	0	0	0	0.307	7.11	2.180
P14	0.075	0.5	0.038	0.178	2	0.355	0.243	4.5	1.091	0.085	0.5	0.043	0.123	1.13	0.138	0.098	1.13	0.110	0.024	0.13	0.003	0.293	8	2.342
P64	0.083	0.09	0.008	0.082	0.36	0.030	0.046	0.09	0.004	0.076	0.09	0.007	0.016	4.46	0.073	0.078	1.46	0.114	0.126	1.46	0.184	0.205	7.36	1.507
P59	0.098	1.33	0.131	0.425	4.08	1.735	0.150	0.75	0.113	0.09	0.75	0.068	0.115	2.08	0. 240	0.08	0.08	0.007	0.076	0.08	0.083	0.226	10.08	2.284
P30	0.132	4.08	0.538	0.105	4.08	0.429	0.068	0.75	0.051	0.147	1.33	0.196	0.092	1.33	0.122	0.103	0.75	0.078	0.07	0.75	0.053	0.328	12	3.937
P24	0.07	0.08	0.006	0	0	0	0.083	0.08	0.007	0.043	0.08	0.004	0.223	5.33	1.191	0.313	0.67	2.115	0.137	1.33	0.182	0.272	12	3.259
P5	0.111	0. 82	0.091	0.107	0.82	0.088	0.265	7.36	1.955	0.165	1.46	0.241	0.24	4.46	1.070	0.000	0	0	0	0	0	0.408	11	4.486
P58	0	0	0	0.144	0.9	0.130	0.054	0.9	0.049	0.064	0.1	0.006	0.084	0.1	0.008	0.264	4.9	1.294	0	0	0	0.210	4.9	1.029
P22	0.224	3.27	0.732	0.156	1.46	0.227	0.118	0.82	0.097	0.102	0.82	0.083	0.145	3.27	0.476	0.089	0.82	0.073	0.091	1.46	0.132	0.364	11	4.007
Mean	**0.086**	**0.96**	**0.142**	**0.095**	**1.06**	**0.231**	**0.087**	**1.19**	**0.260**	**0.079**	**0.56**	**0.111**	**0.137**	**2.67**	**0.465**	**0.226**	**0.98**	**0.315**	**0.071**	**0.62**	**0.120**	**0.273**	**9.14**	**2.612**

Table 3.10 **The average heterozygosity (H_{av}), effective multiplex ration (E) and marker index (MI) for 7 populations of *M. oleifera* based on AFLP analysis.**

Primer	Kibwezi			Likoni			Mbololo			Kitui			India			Pkm1			Malawi			Species		
Pair	Hav	E	MI	Hav	E	M I	Hav	E	MI	Hav	E	MI	Hav	E	MI	Hav	E	MI	Hav	E	MI	Hav	E	MI
P14M51	0.101	2.91	0.293	0.023	0.16	0.004	0.032	0.43	0.014	0.018	0.28	0.005	0.056	2.91	0.162	0.109	7.60	0.831	0.041	0.85	0.035	0.186	22.35	4.159
P12M51	0.057	1.40	0.079	0.060	1.40	0.084	0.103	4.41	0.454	0.026	0.43	0.011	0.157	16.58	2.600	0.117	3.89	0.452	0.055	2.48	0.137	0.164	34.91	5.719
P17M51	0.02	0.13	0.003	0.046	1.13	0.051	0	0	0	0.006	0.03	0.000	0.037	0.5	0.018	0.058	1.13	0.065	0.036	0.5	0.018	0.089	5.28	0.471
P18M51	0.061	2.94	0.180	0.051	1.66	0.084	0.038	0.93	0.036	0.030	0.93	0.028	0.116	8.28	0.958	0.057	1.92	0.109	0.031	0.74	0.022	0.106	19.32	2.047
P14M40	0.080	3.95	0.314	0.094	7.02	0.660	0.074	5.07	0.377	0.036	0.44	0.016	0.082	7.68	0.629	0.131	8.49	0.110	0.092	3.95	0.363	0.163	33.97	5.54
P11M51	0.060	2.36	0.142	0.049	1.74	0.085	0.074	3.08	0.229	0.065	3.48	0.225	0.093	5.31	0.495	0.093	6.94	0.642	0.084	3.90	0.328	0.134	24.40	3.271
P15M51	0.07	6.55	0.458	0.061	3.68	0.225	0.034	0.73	0.025	0.011	0.18	0.002	0.081	3.28	0.267	0.069	3.28	0.227	0.066	3.28	0.217	0.132	33.14	4.385
Mean	0.064	2.89	0.210	0.055	2.40	0.170	0.051	2.09	0.180	0.027	0.83	0.041	0.089	6.36	0.733	0.091	4.75	0.348	0.058	2.24	0.160	0.139	24.77	3.656

3.3.4 Genetic diversity and relationships based on 140 individuals using AFLP.

Genetic diversity and relationships between and within populations of *M. oleifera* was investigated using four pairs of AFLP primers. One hundred and forty individuals were used, 20 from each population. Each of the four primer combinations revealed between one and 30 polymorphic loci in individual populations and between 16 to 60 polymorphic loci across all populations. In total, 157 polymorphic loci were generated, with between 19 (Kitui) and 61 (India) amplified in individual populations. Diversity values for primer pairs in individual populations ranged from 0.006 (P17/M51 in Kitui and Mbololo) to 0.145 (P14/M51 in Kibwezi), with average values for all loci ranging from 0.026 in Kitui to 0.099 in India. Values of Nei's average diversity within populations ranged from 0.040 in Kitui to 0.122 in the Pkm1 population (Table 3.11).

Partitioning the variation within and between populations using an analysis of molecular variance (AMOVA) showed that 59.15% of the genetic variability existed as variation between populations (Φ_{ST}=0.5915; P<0.00001: Table 3.12). A comparison of African and Indian populations showed that 18.59% of the variation existed between the two groups, but this was non-significant (P=0.0538). 14.44% of the variation was found to exist between the Kenyan populations and the rest (Φ_{ST}=0.1444; P=0.0342).

Figure 3.8 shows the neighbour-joining phylogenetic tree based on Nei and Li's genetic distance. The India and Pkm1 populations form single, distinct clades which are separate from the four Kenyan populations. The individuals from the Malawi population do not form a single monophyletic group but are dispersed, with two individuals

Table 3.11 **Diversity statistics for 4 AFLP primer pairs in 140 individuals (7 populations) of *M. oleifera* (n = number of polymorphic bands; H = genetic diversity; D = Nei and Li average genetic distance).**

	P12		P14		P15		P17		All loci		
Population	n	H	n	H	n	H	n	H	n	H	D
India	30	0.144	13	0.098	15	0.111	3	0.046	61	0.099	0.115
Pkm1	16	0.108	18	0.109	16	0.084	5	0.074	55	0.094	0.122
Malawi	17	0.068	8	0.056	17	0.078	3	0.023	45	0.056	0.074
Likoni	11	0.055	3	0.041	18	0.077	8	0.073	40	0.061	0.083
Mbololo	19	0.104	5	0.041	7	0.043	1	0.006	32	0.049	0.064
Kitui	10	0.057	4	0.023	4	0.018	1	0.006	19	0.026	0.040
Kibwezi	13	0.087	12	0.145	22	0.090	1	0.011	48	0.083	0.102
All pop^{n}s	48		33		60		16		157		

(Malawi 9 and Malawi 10) associated with the India population. The remaining Malawi individuals are closer to the Pkm1 population than they are to the Kenyan populations, which is consistent with the Malawi population being a recent introduction from India. The Kenyan populations form single, monophyletic clades and cluster in two groups, the first containing the Likoni and Mbololo populations and the second containing the Kitui and Kibwezi populations.

Table 3.12 Analysis of molecular variance (AMOVA) in 140 individuals (7 populations) of *M. oleifera*.

		Variance component		
Source of variation	**d.f**	**Variance**	**% variation**	**Probability**
Between provenances	6	5.87	59.15	P<0.00001
Between groups (India vs. Rest)	1	2.04	18.59	P=0.0538
Between provenances within groups	5	4.89	44.53	P<0.00001
Within provenances	133	538.8	36.88	P<0.00001
Between groups (Kenya vs. Rest)	1	1.53	14.44	P=0.0342
Between provenances within groups	5	4.99	47.23	P<0.00001
Within Provenances	133	538.80	38.33	P<0.00001

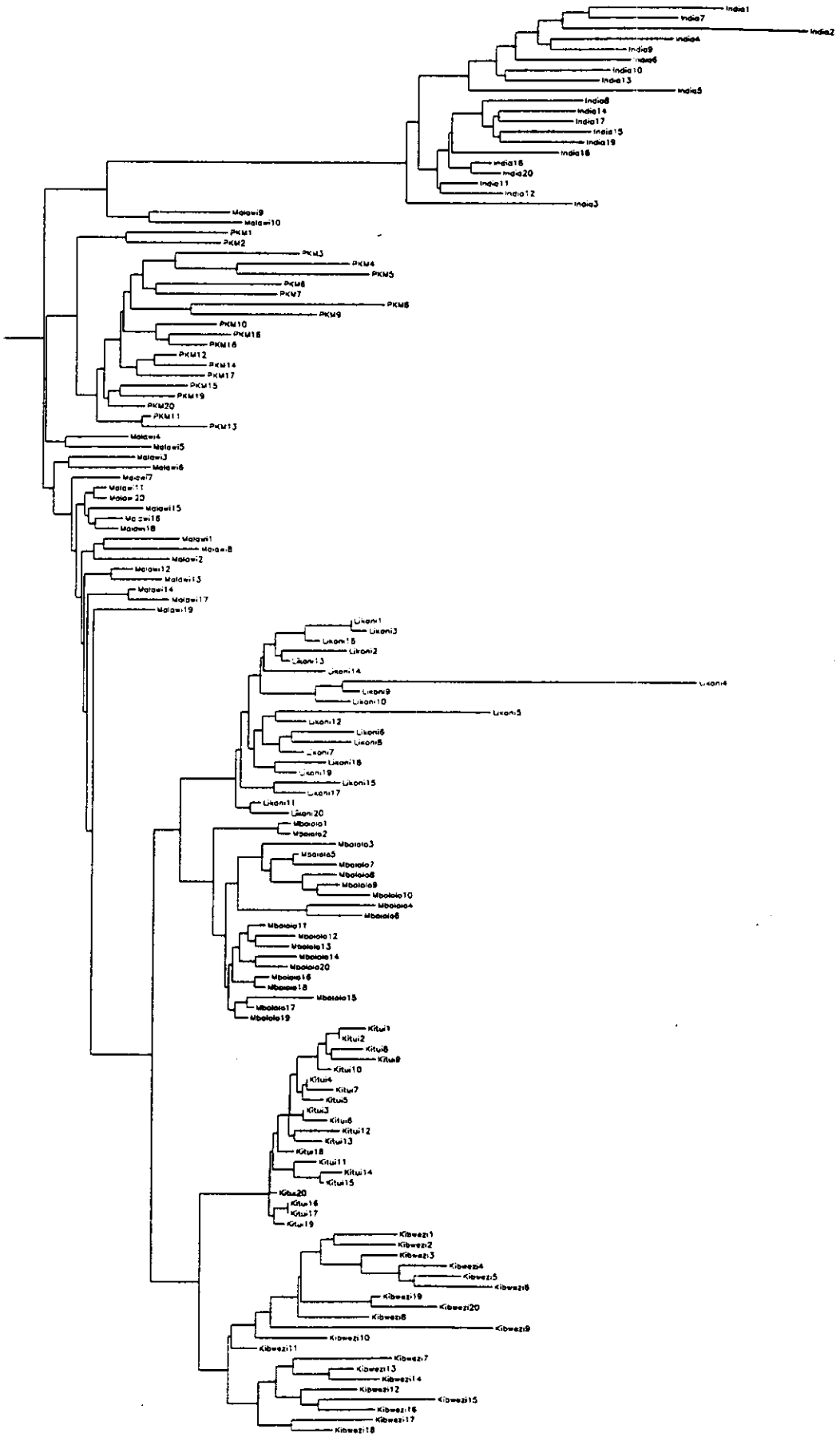


Fig. 3.8 Neighbour-joining phylogenetic tree of relationships between 140 individuals of *M. oleifera* based on Nei and Li genetic distance.

3.4 Discussion.

In the present work, 13 RAPD primers and 7 AFLP primer combinations were used to study genetic diversity within and between 7 geographically isolated populations of *M. oleifera*. When the sample size was doubled to 140 and AFLP markers deployed to analyse genetic diversity and relationships, there was no major difference between this analysis and the one obtained with 70 individuals. In order to provide a balanced comparison between RAPD and AFLP markers in genetic analysis of *M. oleifera*, the results referred in the discussion were those obtained with 70 individuals. Technical limitations such as the number of samples which can be amplified at once in a PCR machine contributed to a reduced sample size in RAPD analysis.

RAPD primers rich with GC sequences have been shown to generate a large number of amplified bands in higher plant genomes (Williams et al 1993); hence primers were selected with a GC content ranging from 60 - 80 %. As expected, the pattern and the number of amplified bands varied with the primers used for both assays. The genome size, constitution of amplification mix, thermal cycling and the sensitivity of gel staining system have been identified as some of the factors affecting the number of RAPD amplification products generated (Samec and Nasinec 1995, Van De Ven and McNicol 1995). However, in AFLPs, the number of selective nucleotides in the *Pst*I and *Mse*I primers used for the selective amplification, as well as the complexity of the genomic DNA, determine the number of resulting amplified DNA fragments (Vos et al 1995). The percentage of polymorphic loci was high with RAPDs (93 %) compared to AFLPs (60 %) in agreement with similar work on barley (Russell et al 1997) and pepper (Paran et al 1998). However, AFLP primers were four times as effective as RAPD primers in detecting polymorphism. The average number of polymorphic products per

primer were 9.77 and 39.71 for RAPD and AFLP primers respectively. A similar ratio was obtained when RAPDs were compared with AFLPs in detecting polymorphism in pepper (Paran et al 1998).

Significant levels of population differentiation were detected with both RAPDs and AFLPs. However, AFLPs detect more variation within populations (51.9%) than is detected with RAPDs (40.9%). High levels of between population variation in tropical tree species using RAPDs have been previously found in *Gliricidia sepium* (Chalmers et al. 1992) and *Cedrela odorata* L., (Gillies et al 1997). However, these results are in contrast to expectations for woody, perennial, predominantly outcrossing species which maintain most variation within populations (Hamrick 1989). For example, in *Camellia sinensis* (Wachira et al 1995) and *Theobroma cacao* (Russell et al 1993) most variation was found within populations. Several studies have compared diversity estimates obtained by RAPDs and isozymes (Dawson et al 1993, Peakall et al 1995) and indicated that RAPDs tend to be particulary useful for discriminating between populations. Most selfing species are characterised by higher variation among populations whereas predominantly outcrossed wind pollinated species exhibit less variation among populations (Hamrick and Godt 1989). *M. oleifera* is insect (bee) pollinated (Puri 1941, Chand et al 1994) and adapted to a mixed mating system (Jyoth et al 1990). Work reported in chapter 5 suggests about 26 % selfing and this may account for the observed high between population differentiation (Cardoso et al 1998). However, the founder effect and restricted gene flow between populations because of geographical isolation (Schaal et al 1998) may account for the significant differences between population variation and regions.

Both RAPDs and AFLPs are multilocus dominant markers capable of detecting single nucleotide mutations as well as insertions / deletions. However, the differences in the levels of variation detected may be due to their relative sensitivity to these types of mutations (Powell 1996). AFLP polymorphism are due to point mutation in the restriction sites or in the selective nucleotides and small insertions/ deletions within the restriction fragment (Vos et al 1995) while RAPD polymorphism results from either a nucleotide base change that alters the primer binding sites or an insertion or deletion within the amplified region (Williams et al 1990). In addition, the higher levels of between population variation detected in this study (and possibly others) may simply reflect the accuracy with which RAPD products are scored when they represent population specific polymorphisms. There is also more confidence when scoring high frequency RAPD products than scoring rare products.

The highest level of genetic diversity was within the Indian natural population (1.152 and 4.309 for RAPD and AFLP, respectively) followed by Pkm1, a selection from the natural population. Abo-elwafa et al (1995) using RAPD markers found that the level of intraspecific variation in cultivated lentil is lower than that in wild species. Among the Kenyan populations, the highest level of genetic variation was observed among the widely dispersed Kibwezi population for both RAPDs (0.695) and AFLPs (3.171) while the more restricted Kitui populations had the least. Distribution range and population size have been identified as major correlates of within population genetic variation in tropical tree species with restricted populations showing significantly less variation than those with broader distribution (Loveless 1992, Hamrick et al 1992) while genetic variation varies directly with population size (Travis et al 1996). The relatively low levels of genetic variation among all introduced populations may suggest that these

populations were based on a small number of genetically related individuals at the time of introduction.

The clustering obtained in this study shows a direct relationship between populations and their geographic origin in agreement with similar work by Hormaza et al (1994) and Pakniyat et al (1997) in *Pistacia vera* L cultivars and wild barley, respectively. The grouping closely of Kibwezi and Kitui populations and Mbololo with Likoni by both PCOs may suggest two sources of introduction of Kenyan *M. oleifera*. Historical relationships have been found to contribute to the genetic structure of most plant populations with those having recent common ancestry being genetically more similar than those having more distant common ancestry (Schaal et al 1998). Kitui and Kibwezi fall within the same climatic zone while Mbololo is intermediate between Likoni and Kibwezi. Loveless (1992) identified habitat heterogeneity to have an effect on population structure. Therefore, if the Kenyan populations were introduced from a common source, then climatic factors, short rotation and restricted gene flow may have played a role in influencing genetic differentiation between the two Kenyan groups. Jahn (1991) suggested that introduction of *M. oleifera* in Kenya may have started at the coast (actual location not specified) and later spread to other areas. The clustering of the Malawi population close to Indian populations suggests an Indian rather than an African origin for this population.

The present data on patterns of genetic differentiation suggest that the Kenyan populations should be considered distinct from each other for the purposes of seed collection, planting and management. In the case of conservation of genetic resources seed collection should be done across the species range to ensure a more representative

sampling of the genetic variation. The significant levels of genetic differentiation observed in this study may be related to adaptive variation and structured progeny trials are required to assess the performance of the various populations for traits of interest. Although there is a risk in using neutral genetic markers for making inferences about adaptive processes unless selection is still acting or there is a very close linkage between the selected locus and the neutral marker (Palacios and Gonzalez-Candelas 1997), work on *Cedrela odorata* by Gillies and co-workers (1997) found a correlation between molecular differences between populations and adaptive characters.

Given the proliferation of genome marker technologies, a comparison of different marker techniques using the same DNA samples is timely (Sharma et al 1996). Although the magnitude differed, both assays detected significant differences between populations and between regions. Generally, comparable patterns of clustering individuals and populations were revealed. The procrustes rotation analyses confirmed that similar patterns of variation were detected using both assays. As an overall measure of marker efficiency in detecting polymorphism, the average expected heterozygosity, the multiplex ratio and the marker index for both assays were calculated (Powell et al 1996). The AFLP assay gave high values for effective multiplex ratio and marker index in agreement with earlier work on soybean by Powell et al (1996) while RAPDs had higher values for average expected heterozygosity which is in agreement with similar analyses conducted on potato (Milbourne et al 1997). The high marker index of the AFLP assay derives from its high effective multiplex ratio rather than from high levels of detected polymorphism (Powell et al 1996). To date, all comparative studies concur in identifying AFLP as an unique technology with high marker utility (Powell et al 1997). The multiplex ratio of the AFLP assay can be varied by changing both the

restriction enzymes and the nature of selective nucleotides at the 3` end of the PCR primers virtually generating unlimited number of markers (Van Eck et al. 1995). The simultaneous amplification of large numbers of polymorphic DNA fragments represents perhaps the greatest single advantage of the AFLP technique. Furthermore, the increased robustness of this technique provides accurate estimate of genetic distance between individuals and populations (Travis et al 1996). High levels of polymorphism detected in this species are indicative of high levels of outcrossing among individuals within populations (Loveless and Hamrick 1984, Karron 1991).

For non-industrial agroforestry species, obtaining quick, accurate estimates of the distribution of genetic variation in a cost effective manner is particularly important. To date, most studies of this nature have employed RAPDs. In the present study, both RAPDs and AFLPs have been used to detect variability in common genotypes. The AFLP technology is extremely robust and proficient at revealing intra-population diversity and estimating genetic distance between individuals and populations (Travis et al 1996, Arens et al 1988, Winfield et al 1998). Furthermore, three times the number of data points (amplification products) were generated with AFLPs compared with RAPDs over an equivalent period of time. These factors contribute to the conclusion that AFLPs provide a cost-effective procedure to monitor the extent and distribution of diversity in *Moringa* and other agroforestry species. Although similar patterns of relationships among the populations and individuals were revealed by both methods, slight differences were evident. Milbourne et al (1997) and Paran et al (1998) found similar differences in the way dendrograms from RAPDs and AFLPs grouped genotypes from potato and pepper respectively, possibly due to the higher level of polymorphism revealed by AFLP markers (Paran et al 1998).

3.5 Conclusion.

To my knowledge this is the first report of molecular techniques being applied to genetic analysis of *M. oleifera*. It is also the first study that has deployed both RAPDs and AFLPs on common genotypes for a perennial tropical tree species. This study has shown that polymorphism based on RAPD and AFLP molecular markers can be useful in determining genetic relationships between genotypes and populations of *M. oleifera* and that there is a considerable amount of genetic variation available between and within the geographically isolated populations of the species. The results will have important practical implications in planning conservation programmes and utilisation of *M. oleifera* genetic resources. It will also be important to establish whether the observed genetic differences between the Kenyan populations is reflected in differences in important agronomic traits such as yield (Chapter 6). One of the biological factors affecting the genetic structure of populations is the mating system and the report of this investigation is given in chapter 5.

CHAPTER FOUR.

4.0 AFLP multilocus phenotypes do not suggest clonality in two commonly used seed stands of *M. oleifera* from Eastern Kenya.

4.1 Introduction.

Since its introduction to Kenya at the beginning of this century (Jahn 1991), *M. oleifera* was considered of marginal commercial value and, consequently, no seed sources or clonal identities from these introductions have been maintained. However, the importance of *M. oleifera* has now been recognised and its wide scale planting in East Africa has gained momentum over the last three years. Seeds for planting in Kenya and neighbouring countries, particularly Tanzania and lately Uganda, are being obtained from a seed stand in Mbololo and occasionally from Kibwezi. These seed sources have been established and maintained by farmers and information is scarce on their genetic quality and relatedness. *M. oleifera* can be easily propagated vegetatively (Ramachandran et al 1980, Nautiyal and Venkataraman 1987, Kantharajah and Dodd 1991) and it has been difficult to ascertain to what extent this mode of propagation may have been carried out among the Kenyan provenances. Lack of genetic information on these seed stands coupled with the possibility of vegetative propagation resulted in the need to establish the genetic integrity of trees as a priority. DNA fingerprinting of trees sampled from these two seed sources was carried out in the present work. In plants, the

degree of fingerprint similarity mainly depends on the level of heterozygosity and this is influenced by the means of pollination and the propagation method (Sigurdsson et al 1995).

Clonality in plants can have important implications for selection and conservation of populations because single genetic individuals (genets) may comprise numerous morphological units (ramets) that appear distinct (Maddox et al 1989, Kennington et al 1996). Determinants of population characteristics such as population size, recruitment, mortality, levels of polymorphism, and conformance to Hardy-Weinberg equilibrium are all contigent upon observations on genetically distinct individuals (Park and Werth 1993). Field observations (morphological features) have been used to obtain information on clonal extent (Parks and Werth 1993). A census based on counts of recognisable individuals (ramets) will overestimate the number of genetically distinct individuals (genets) and could make the true conservation status of a taxa go undetected (Sydes and Peakall 1998). However, for a majority of species, genet recognition is facilitated by molecular markers such as isozymes (Parks and Werth 1993), DNA restriction digestion and hybridization (Kennington et al 1996) and RAPDs (Sigurdsson et al 1995, Sydes and Peakall 1998). Allozyme studies however, sometimes underestimate the number of genets within a population (e.g. Lui and Furnier 1993, Peakall et al 1995, Smouse and Chevillon 1998) while several RAPD primers would be required to achieve good results (Sigurdsson et al 1995).

Clonal populations are often expected to become dominated by one or a few large genets over time due to the low levels of recruitment through seed (Ellstrand and Roose 1987). An understanding of the extent of clonality is therefore critical for the

implementation of most appropriate conservation management of plants suspected of being clonal. The objective of the present work was to use AFLPs (Vos et al 1995) to investigate the extent of clonality arising from vegetative propagation in two popular seed stands (Mbololo & Kibwezi) in order to allay fears that the seeds we are using for massive planting in Kenya and elsewhere may be from a single or very few clones.

4.2 Materials and methods.

4.2.1 Plant material.

Leaf samples were collected from a total of 20 *M. oleifera* trees, 11 from Mbololo and 9 from Kibwezi. The trees were randomly selected from the two widely used seed sources of this species in Kenya.

4.2.2 DNA isolation and the AFLP procedure.

Genomic DNA was isolated from leaf samples from the chosen trees following a modification of the Edwards et al (1991) protocol as described in chapter 2. The method to obtain amplified fragment length polymorphism (AFLP) markers followed Vos et al (1995) exactly as described in chapter 2. Three pairs of AFLP primers whose sequences are shown in section 2.5.4.1 were used.

4.2.3 Data analysis.

Patterns of amplified DNA fragments were examined visually and scored as present (1) and absent (0). The number of different banding patterns (fingerprints) produced by

each primer was identified. The differentiation between individuals is based on the proportion of AFLP fragments that were unambiguously polymorphic (present or absent) between the individuals and was strengthened in some cases by individual-specific AFLP markers thus increasing the accuracy of data acquisition. The average probability that two trees have an identical multiband fingerprint was computed for each primer and the two populations considered together as proposed by Jeffreys et al (1985), exactly as used by Georges et al (1988), Russell et al (1993) and Yeh et al (1995), using X^m, where X is the proportion of shared bands and *m* was the number of bands amplified. This probability was the lower limit of the affinity between trees because it assumed random association among the DNA markers (Yer et al 1995).

4.3 Results.

The three primers chosen for analysis generated 203 amplification products scored across the populations with a grand average of 67.6 products per primer. The total number of polymorphic products scored across the two populations was 159 with a mean value of 53 products per primer. The number of products generated by each primer in each single population is shown in Table 4.1. Primers differed in the ability to identify unique multiband phenotypes among the 20 sampled trees (Table 4.1). Primer pair P14-M51 easily discriminated the 20 sampled trees as distinct multiband phenotypes while primer pair P11-M51 could not resolve any difference between three individuals sampled from Kibwezi. Similarly, primer pair P17-M51 could not distinguish between two individuals from each population. However, trees sampled within and between populations were distinct phenotypes representing different

genotypes.

The probability (X^m) that two trees have identical AFLP phenotypes for each and across the three primers was computed (Table 4.2). Values of X^m decreased exponentially as the number of primer (or bands) increased arithmetically. When all the 3 primer pairs were combined, there was a 2.001×10^{-135} probability that two trees of the same multiband phenotype were miss-classified as being different. When the power of discriminating among trees was defined as P=1-*Xm* (Yeh et al 1995), primers did not differ in their power to discriminate among the sampled trees as the value was one for all the three primers. Although primer pair P17-M51 produced 2 pairs of identical multiband phenotypes, when the pairwise comparisons of the trees between populations (20X19)/2, was considered the proportion 4/190 was not significantly different from the zero expectation under the null hypothesis that every AFLP multiband phenotype was unique.

Table 4.1 Amplification products and phenotypes identified in 20 *M. oleifera* trees based on 3 pairs of AFLP primers.

Primer		Mbololo			Kibwezi			Combined		
No.	**Code**	**Total**	**Poly.[1]**	**mult-f.[2]**	**Total**	**Poly.**	**mult-f.**	**Total**	**Poly.**	**mult-f.**
1	P11-M51	79	53	11	60	28	7	85	59	17
2	P14-M51	58	53	11	37	32	9	65	60	20
3	P17-M51	41	28	10	30	17	8	50	37	16
Total		**178**	**134**		**136**	**86**		**203**	**159**	
Mean		**59.3**	**44.7**		**45.3**	**28.6**		**67.6**	**53**	

[1] Polymorphic loci
[2] Multiband fingerprints.

Table 4.2 Proportion of band sharing and the probability (X^m) of two trees having an identical multiband phenotype.

Primer		Mbololo		Kibwezi		Species	
No.	Code	Similarity	X^m	Similarity	X^m	Similarity	X^m
1	P11-M51	0.329	7.219×10^{-39}	0.533	4.015×10^{-17}	0.306	1.933×10^{-44}
2	P14-M51	0.086	1.589×10^{-62}	0.135	6.643×10^{-33}	0.077	4.187×10^{-75}
3	P17-M51	0.317	3.495×10^{-21}	0.433	1.243×10^{-11}	0.26	5.606×10^{-30}
Total		**0.247**	**7.944×10^{-109}**	**0.368**	**9.022×10^{-60}**	**0.217**	**2.001×10^{-135}**

The distribution of individual specific AFLP markers is shown in Table 4.3. There was a higher number of individual specific bands among the Kibwezi individuals than Mbololo.

Table 4.3 **Distribution of genotype specific bands among the 20 individuals sampled.**

	Primer	Mbololo											Kibwezi								
	pair	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10	b11	k1	k2	k3	k4	k5	k6	k7	k8	k9
Individual	P11-M51	1	1	1	0	0	0	2	1	0	1	0	3	0	0	0	0	4	0	0	0
specific loci	P14-M51	0	2	0	0	0	0	0	0	1	0	0	1	0	1	0	2	0	0	0	0
	P17-M51	0	0	0	2	0	1	0	4	2	0	0	1	0	0	1	1	5	2	0	0

4.4 Discussion.

The AFLP technique is a powerful DNA fingerprinting method as shown by the primer pair (P14-M51) which could discriminate between all the 20 sampled trees. However, primer pair P11-M51 could not distinguish between three individuals from Kibwezi (k7, k8, k9), while primer pair P17-M51 identified only 18 multiband phenotypes as it failed to distinguish between two Kibwezi individuals (k2 & k9) and also two individuals from Mbololo (b3 & b5). There was a higher number of individual specific bands among Kibwezi individuals in agreement with the high diversity found in this population in chapter 3. The primer pair which identified all the 20 individuals as distinct multiband phenotypes also produced the highest percentage of polymorphic loci (92.3%) compared to 75.5% for P17-M51 and 69.4 % for P11-M51. The results suggest that the more informative a given primer, the more suitabile it is for DNA fingerprinting work. A similar study in *Populus tremuloides* Michx. by Yeh et al (1995), using RAPDs, found the power of the primers to discriminate among the sampled trees ranged from 92% to 69%. Earlier work on 27 clonal plants by Ellstrand and Roose (1987), using isozymes, morphological and quantitative characters revealed a significant correlation between the number of characters scored and the number of genotypes detected, suggesting that sensitivity is generally enhanced with the addition of characters, but not with increased sample size. Therefore, careful selection of the best primer pairs could reduce the cost and time associated with AFLP fingerprinting and allow large samples to be easily screened. However, current work suggests that use of multiple primer pairs increases the reliability and sensitivity of the AFLP method since the probability of matching fingerprints is much lower when combining data from two or more primers. The need to use multiple primers was supported by the possibility of conflict between the predicted probability of band sharing and actual value (Table 4.2).

The efficiency of the method combined with its sensitivity allows AFLP fingerprints to be obtained routinely and on a large scale, which is a requirement for the practical management of clonal selection programmes.

When the power of discriminating among trees was defined as $P=1-X^m$, the value was 1.0, that is, the chance match probability of observing identical banding patterns in two unrelated phenotypes is statistically very small. It can therefore be concluded that we do not have members of the same clone among the sampled trees and by extension, the possibility of clones in *M. oleifera* populations is very limited. Lack of common multiband phenotypes between individuals from the two populations by the least discriminating primer pairs confirms the results obtained in chapter 3 of high divergence between Mbololo and Kibwezi populations. Although there are many reports of clonal individuals that contain discrete genetic clones (e.g Ellstrand and Roose 1987, Berg and Hamrick 1994, Kennington et al 1996), current results suggest occurrence of sexual reproduction in *M.oleifera* as a common event.

The occurrence of extensive sexual reproduction in *M. oleifera* implies that the choice of populations for protection and the collection of material for ex-situ programmes has to be considered similarly to that of other sexually reproducing species. The present study will be a relief to those who have been using these seed sources without knowing their genetic quality. Because of the high divergence between populations (Chapter 3), it will be essential to represent individuals from all populations in any ex-situ collection in order to obtain representation of the genetic variation in the species.

4.5 Conclusion.

The occurrrence of multilocus genotypes among all 20 sampled individuals, and the extremely low probabilities of observing multiple copies of the same genotype clearly indicate the dominance of sexual reproduction in *M. oleifera* from the two populations studied.

CHAPTER FIVE.

5.1 An investigation of the mating system in an *M. oleifera* seed orchard using AFLP markers reveals self-compatibility.

5.1 Introduction.

Detailed knowledge and understanding of the mating system in natural and artificial plant populations are fundamental prerequisites for understanding their genetic architecture and evolutionary potential (Epperson 1992, El-Kassaby and Jaquish 1996, Mitchell and Marshall 1998). Proper estimates of the outcrossing rates are often needed for planning breeding programmes (Ritland and Jain 1981), conservation and management of tropical trees (Loveless 1992). Aside from the importance in understanding population genetic structure, significant inbreeding depression due to selfing decreases the survival and growth of seedling progeny in some species such as conifers (Furnier and Adams 1986).

The mating system in plant populations is influenced by genetic and environmental factors (Clegg 1980). The majority of outcrossing angiosperms have bisexual flowers, a condition from which self-pollination can evolve directly through the modification of self-incompatibility or other floral traits that prevent self-pollination (Schoen et al 1997). Patterns of floral variation among the angiosperms suggest that characters that prevent self-pollination have been lost independently in many widely separated evolutionary lineages (Schoen et al 1997) because traits promoting self-pollination are selectively advantageous under a wide variety of circumstances (Lande and Schemske

1985).

In addition to the role of the variable floral architectures in determining mating systems of the plant populations (Ennos 1981), the mating system may be sensitive to plant density and population size (Clegg 1980, Ennos and Clegg 1982, Goodell et al 1997), type of pollination vector and abundance (Aide 1986), flower colour (Brown and Clegg 1984), size of floral display (Dudash and Barret 1989) and anther-stigma separation (Karron et al 1997). Temporal changes in quality or quantity of pollinator service or variation in the timing of flowering can lead to seasonal changes in the mating patterns and composition of the outcross pollen pool (Moran and Brown 1980, Fripp et al 1987, Goodell et al 1997, Mitchell and Marshall 1998). As such it is reasonable to expect that outcrossing rates could vary extensively both spatially within and between populations, and temporally within a single population (Wolfe and Shore 1992).

Traditional methods used for the measurement of mating systems have been based on the analysis of floral morphology, greenhouse crossing experiments, and (where appropriate) the observation of pollinator behaviour (Clegg 1980). The practical use of phenotypic markers in trees is limited by a number of factors such as long time required for progeny to reach maturity for the markers to be scored and lack of consistency between phenotypic markers and outcrossing (Gjuric and Smith 1996). The development and application of isozymes provided numerous genetic markers which can be used to measure mating systems in plant populations (Brown and Allard 1970, Holtsford and Ellstrand 1990, Cottrell and White 1995, Premoli 1996, Schoen et al 1997). In recent years DNA based methods such as RAPDs (Gjuric and Smith 1996) and AFLPs (Gaiotto et al 1997) have been used to estimate outcrossing rates. However,

due to their dominance behaviour, RAPD and AFLP markers provide less information per locus than co-dominant markers (Gaiotto et al 1997). This is particularly relevant for applications that require genotype discrimination, as in the case of outcrossing-rate estimation (Gaiotto et al 1997). However, Ritland and Jain (1981) demonstrated, that this limitation could be readily overcome by multilocus estimation of outcrossing with dominant markers with intermediate gene frequencies.

M. oleifera is adapted to selfing (geitonogamy) and outcrossing (xenogamy) with larger fruit set, seed set and fecundity in the latter mode (Jyoth et al 1990). The flowers produce both pollen and nectar with bees as the main pollinators (Puri 1941, Jyoth et al 1990, Chand et al 1994). The two anterior petals form a suitable landing place for the bees (Puri 1941).

The main aims of this study were to: (1) test the utility of dominant AFLP markers in estimating outcrossing rates in *M. oleifera*; (2) use AFLP genetic markers to obtain estimates of the outcrossing rate using progeny genotypes in family arrays in an *M. oleifera* seed orchard from Mbololo, Kenya.

5.2 Materials and methods.

5.2.1 Plant material.

Single tree collection was carried out in an *M. oleifera* seed orchard in Mbololo, Kenya. The seeds were grown under greenhouse conditions. DNA was isolated as described in section 2.2.1 from the seedlings (leaves) and also from leaves collected from each

maternal plant. A random sample of 4 families of open pollinated progeny arrays of 20-23 individuals, giving a total of 86 individuals, were used for this study.

5.2.2 AFLP procedure.

The AFLP technology (Vos et al 1995) was used following the procedure described in section 2.5, employing *Pst*I and *Mse*I as rare and frequent cutter enzymes, respectively. The nucleotide sequence of the two AFLP primer pairs used, *Pst*I (P12, P11) and *Mse*I (M51) are shown in section 2.5.4.1. AFLP markers were identified by the first primer code followed by the locus number, e.g., P12-lo01.

5.2.3 Data analysis.

Scoring of bands was carried out considering only two possible alleles: band presence or absence. The mating system was analysed using the multilocus mixed mating program (MLDT) of Ritland (1990). From progeny array data, the programme simultaneously estimated (i) the multilocus outcrossing rates (t_m) by the Newton-Raphson method; (ii) the mean single-locus outcrossing rate (t_s); (iii) single locus inbreeding coefficient (Wright's fixation index) of the maternal parents (F); (iv) the pollen and ovule allele frequencies (p and o) by the expectation-maximisation method; (v) variances of the above quantities using the bootstrap method where the progeny array (within families) is the unit of resampling (100 bootstraps used) and (vi) the maternal genotype. For each locus, a χ^2 statistic was calculated to test the null hypothesis that the number of observed progeny individuals for each genotype class from each maternal genotype plant did not differ from the expected number under the mixed-mating model. Assumptions of the model are as described in Ritland and Jain (1981). In particular, the

model specifies that both selfing and outcrossing occur in the population (Shaw and Allard 1982). Multilocus estimation is statistically more efficient than single-locus estimation because multilocus data sets contain more information about outcrossing than is available at any one single locus (Furnier and Adams 1986).

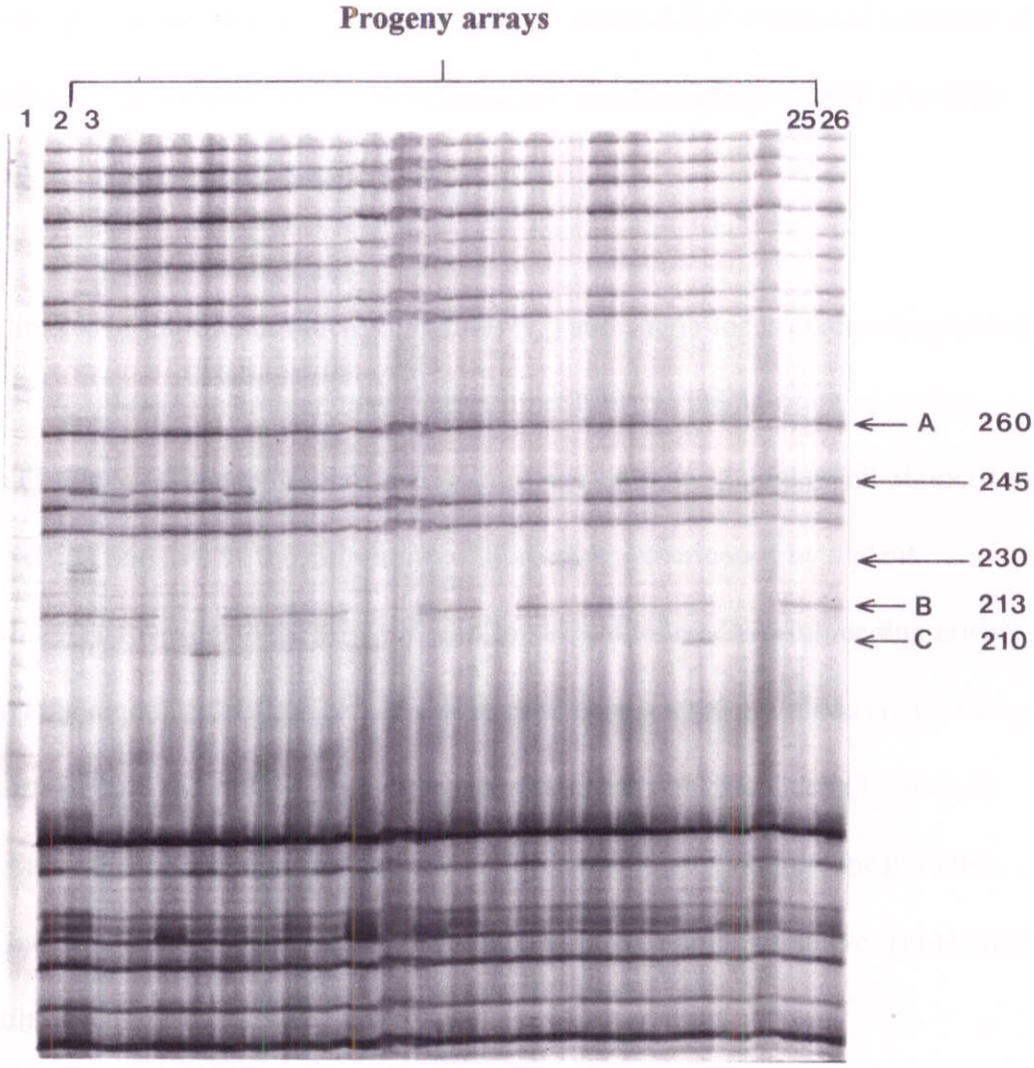
5.3 Results.

An AFLP band present in all offsprings (e.g., Fig. 5.1, A) indicates that the maternal plant could be homozygous (+ / +) for this marker or heterozygous (+ / -) or homozygous null (- / -), and that the marker could be at a very high frequency in the pollen pool, i.e., in the orchard. However, because the maternal DNA was included in the analysis (Fig. 5.1), it was possible to determine when the maternal plant was homozygous null at a particular locus. Markers present in the maternal genotype and absent in a few offspring (e.g., Fig.5.1, B) suggest that the maternal plant is heterozygous with the marker at high frequency. In this case offspring originating from outcrossing cannot be discriminated against based solely on this marker. Finally, the most informative configuration is when a marker is absent in the maternal plant (e.g.Fig.5.1, C). In this case it can be inferred that the maternal plant has a homozygous genotype (- / -) for the marker and that progeny individuals that have the marker are most surely heterozygous (+ / -) and a product of outcrossing. Therefore, outcrossing events could be readily identified if the maternal plant has a homozygous null genotype. Evidently even in this case, not all of the outcrossing events can be identified based only on this marker, as progeny individuals that were outcrossed but received the "band absent" allele from the paternal plant could not be directly detectable. It was by accumulating data from several markers and estimating allele frequencies in all the

Fig. 5.1 An example of a molecular profile generated by the primer pair P12/M51 showing segregation of AFLP markers used to identify outcrossing events.

Lane 1, molecular weight marker (bp); lanes 2 and 26, two replicates of maternal genotype; lanes 3 to 25, family progeny arrays.

Marker A: Maternal plant homozygous (+/+);
Marker B: Maternal plant is either homozygous (+/+) or heterozygous (+/-);
Marker C: Maternal plant homozygous null (-/-), presence of this marker in progenies indicates an outcrossing event.



families for the AFLP markers used, that all the outcrossing events could be detected. Markers present in the maternal plant and all progenies were non-informative in detecting outcrossing events and were excluded from the analysis.

Thirty seven out of 50 loci assayed had significant differences between the allele frequencies of ovules and incoming pollen at the 95% level (Table 5.1). This suggests that the maternal trees did not represent all the local pollen pools. This violation of the assumption of the equivalence of pollen allele frequencies received by the maternal trees has an unmeasurable but relatively minor effect on the estimate of the true population outcrossing rate (Ritland and Jain 1981). A χ^2 statistic to test the conformity of marker loci to the mixed-mating model, indicated for seventeen AFLP markers the number of observed progeny individuals for each genotype class from each maternal genotype departed from the expected numbers (Table 5.1).

The estimates of multilocus outcrossing rates (t_m) and single-locus outcrossing rates (t_s) obtained from MLDT clearly indicate self compatibility in the *M. oleifera* mating system (Table 5.2). The multilocus outcrossing rate estimates based on all 50 loci was 74 %. Selfing may not be the only form of inbreeding experienced by a plant population. To investigate the possibility of biparental inbreeding (inbreeding arising from mating among related plants), the difference between the multilocus outcrossing rate estimates and the mean of the single locus estimate was calculated (Table 5.2). There was significant difference between multilocus estimates and the mean of the single locus estimates, suggesting the existence of mating among relatives (biparental inbreeding).

Table 5.1 **Allele frequencies, their respective standard deviations (σ), χ^2 statistics for agreement with the mixed-mating model and the maternal genotypes (M.G) (known or inferred).**

Locus	Gene frequency		χ^2	M.G[1]				Locus	Gene frequency		χ^2	M.G[1]			
	Pollen (σ)	Ovule		1	2	3	4		Pollen(σ)	ovule		1	2	3	4
P12-1	0.09 (0.07)	0.2	3.36	2	3	3	2	P11-26	0.05 (0.01)	0.5	10.43*	3	3	3	3
P12-2	0.06 (0.03)	0.1	0.86	3	3	3	2	P11-27	0.05 (0.03)	0.0	0.05	3	3	3	3
P12-3	0.30 (0.07)	0.1	2.60	3	2	3	3	P11-28	0.06 (0.03)	0.0	0.10	3	3	3	3
P12-4	0.03 (0.03)	0.1	6.38*	3	3	3	2	P11-29	0.13 (0.04)	0.0	0.04	3	2	3	3
P12-5	0.04 (0.04)	0.1	13.54*	3	3	3	2	P11-30	0.06 (0.04)	0.1	5.71*	3	3	3	3
P12-6	0.04 (0.04)	0.1	6.75*	3	3	3	2	P11-31	0.19 (0.06)	0.0	1.22	2	3	3	3
P12-7	0.08 (0.05)	0.1	10.42*	3	3	3	2	P11-32	0.08 (0.05)	0.1	7.21*	2	3	3	1
P12-8	0.49 (0.09)	0.2	0.76	2	3	3	2	P11-33	0.18 (0.08)	0.3	1.25	2	3	3	3
P12-9	0.09 (0.04)	0.2	0.07	3	3	1	3	P11-34	0.00 (0.00)	0.1	0.58	3	3	2	3
P12-10	0.17 (0.06)	0.2	2.12	3	3	2	2	P11-35	0.12 (0.05)	0.1	3.02	3	3	3	3
P12-11	0.21 (0.09)	0.3	6.84*	2	2	3	2	P11-36	0.03 (0.02)	0.0	0.00	3	2	3	3
P12-12	0.00 (0.00)	0.1	6.37*	3	2	3	3	P11-37	0.06 (0.04)	0.1	10.49*	3	2	3	3
P12-13	0.00 (0.00)	0.2	2.08	2	2	3	3	P11-38	0.16 (0.06)	0.1	8.55*	3	3	3	3
P12-14	0.17 (0.07)	0.1	11.18*	3	2	3	3	P11-39	0.02 (0.02)	0.0	0.01	3	3	3	3
P12-15	0.01 (0.00)	0.5	2.76	1	1	2	3	P11-40	0.02 (0.02)	0.0	0.01	3	3	3	3
P12-16	0.07 (0.04)	0.4	0.07	1	1	3	3	P11-41	0.19 (0.04)	0.0	0.16	1	2	1	1
P12-17	0.02 (0.00)	0.5	9.01*	1	1	2	3	P11-42	0.20 (0.08)	0.3	3.93*	1	3	1	1
P12-18	0.04 (0.03)	0.4	0.02	1	1	3	3	P11-43	0.98 (0.03)	0.9	0.02	3	3	1	3
P12-19	0.03 (0.00)	0.5	11.85*	1	1	2	3	P11-44	0.65 (0.11)	0.8	0.05	3	3	1	3
P12-20	0.01 (0.00)	0.5	0.10	1	1	2	3	P11-45	0.02 (0.02)	0.2	0.00	3	3	1	3
P12-21	0.77 (0.07)	0.9	0.00	3	3	3	3	P11-46	0.02 (0.02)	0.2	0.00	3	3	1	3
P12-22	0.02 (0.02)	0.0	0.00	3	3	3	3	P11-47	0.02 (0.02)	0.2	0.00	3	3	1	3
P12-23	0.02 (0.02)	0.0	0.00	1	2	1	3	P11-48	0.02 (0.02)	0.2	0.00	1	3	1	1
P12-24	0.06 (0.02)	0.5	13.03*	1	2	1	3	P11-49	0.02 (0.02)	0.2	0.00	1	3	3	1
P12-25	0.05 (0.01)	0.5	10.43*	1	2	1	3	P11-50	0.18 (0.08)	0.8	0.04	3	3	3	3

* Marker locus with significant deviation at the 0.05 level.

[1]Maternal genotypes (known or inferred) for the 4 open pollinated families of *M. oleifera*. Genotype 1, homozygous for the "band-presence" allele (+/+); genotype 2, heterozygous (+/-); genotype 3, homozygous (- /-). Standard error for paternal gene frequencies were not computed because sampling was done within families.

Table 5.2 **Multilocus (t_m) and single-locus (t_s) outcrossing rates and Wrights fixation index (F) (standard errors in parentheses).**

Site	Families	Offspring	t_m	t_s	t_m-t_s	F
Mbololo	4	86	0.740	0.522	0.219	0.376
			(0.065)	(0.050)	(0.043)	(0.000)*

* Standard error for F was not computed because sampling was done within families.

5.4 Discussion.

Dominant AFLP markers have been previously used to estimate outcrossing rates in *Eucalyptus urophylla* (Gaiotto et al 1997). The present work shows the use of AFLP markers in a mating system study of *M. oleifera* and demonstrates that AFLP markers, though dominant with a lower information content than co-dominant markers are adequate for the study of the mating system in plant populations. This high-throughput marker technology allows the analysis of a large number of individuals with a large number of markers in a relatively short time as only a single AFLP pair could allow the generation of sufficient markers to obtain a robust estimate of outcrossing rate. However, when using dominant markers for estimating outcrossing rate, it is imperative that an adequate screening of primers be done so as to maximise the probability of amplifying a large number of polymorphic markers in the progeny array (Gaiotto et al 1997). Only by amplifying a large number of markers does one have the flexibility of selecting highly reproducible ones that can be readily scored across families and which display low-to-intermediate frequency for the dominant marker allele in the population. The predominance of heterozygous (2) and homozygous null (3) classes for the inferred maternal genotypes (Table 5.2) reflects the selection of low-to-

intermediate frequency markers necessary to carry out such a study.

The estimates of outcrossing rates obtained indicate that *M.oleifera* seeds from the Mbololo seed source are a product of both selfing and outcrossing events. The mixed mating system (t_m = 0.74) described for this species is consistent with the observations of self-compatibility in India (Puri 1941, Jyoth et al 1990). In addition, growing evidence suggests that many tree species with bisexual flowers seem to have at least limited self-compatibility (Bawa and Ashton 1991). Comparable levels of outcrossing have been observed in some species such as *Schiedea lydgatei* (t_m =0.694-0.874), *Hydrophyllum appendiculatum* (t_m=0.62-0.81) by Norman et al (1997) and Wolfe and Shore (1992) respectively. However, high outcrossing rates (t_m > 0.9) have been observed in a majority of conifers (Furnier and Adams 1986, Morgante et al 1991, Cottrell and White 1995). Because *M. oleifera* is not exclusively outcrossing, expected levels of genetic variation for predominantly outcrossing species are not likely to be maintained (Gaiotto et al 1997).

Single-locus estimation is more sensitive to related matings other than selfing (Furnier and Adams 1986). Thus, if inbreeding other than selfing occurs, t_s will generally underestimate outcrossing to a much greater extent than t_m. In the current work, multilocus estimate differed significantly from the single-locus estimate, suggesting significant biparental inbreeding (Ritland 1990). Possible reasons for significant differences between the allele frequencies of ovules and incoming pollen (Table 5.2) have been advocated (Murawski and Hamrick 1992, Furnier and Adams 1986). Of these, the immigrant pollen from outside the sample population or from an unrepresented sample of maternal trees due to the small number of families sampled

may account for the significant differences detected between the allele frequencies of ovules and incoming pollen in the present work.

The fixation index, F, in the progeny was higher than expected based on the estimate of t_m. Taking $t_m = 0.740$, the expected fixation index was $[F = (1-t)/(1+t)] = 0.149$, while the estimated F was 0.376. A higher than expected F suggests more inbreeding than expected in the progeny population used to carry out the study. Since the mating system in *M. oleifera* involves some selfing, an excess of homozygotes in progenies would be expected if the populations are in mating-system equilibrium (Furnier & Adams 1986). Mating-system studies of natural populations of *Eucalyptus*, reported an F higher than expected based on the estimated t_m (Peters et al 1990, House and Bell 1994). A χ^2 test indicated that observed progeny genotype frequencies did not conform to those expected under mixed mating for some marker loci. Several factors can contribute to such violations: selection against homozygous genotypes, genotype-dependent outcrossing rate, and unbalanced frequencies of pollen in the population (Ritland 1983).

Although many population genetic models assume random mating or random outcrossing there is increasing evidence for non-random outcrossing in many tree species (O'Malley and Bawa 1987, Murawski and Hamrick 1991, Murawski and Hamrick 1992). The effect of nonrandom outcrossing could manifest itself in a reduction of the effective population size, genetic drift, and increased population structuring (Murawski and Hamrick 1991). Although models for the evolution of selfing as well as empirical studies of selfing rates in natural populations (reviewed in Barret and Eckert 1990) suggest that partial selfing is not unusual, Lande and Schemske (1985) predicted partial selfing as a rare "transitory" condition that occurs when species are

evolving from outcrossing to selfing.

In estimating heritability and genetic gains, the assumption that the relationships among the progeny is 0.25 leads to inaccurate estimation of the additive variance if the relationships among the progeny are not entirely half sib (Falconer 1960, Mousseau 1987, Askew and El-Kassaby 1994). The additive variance coefficient can be greatly reduced by selfing and increased by the presence of foreign pollen (Falconer 1989, Gaiotto et al 1997, Askew and El-Kassaby 1994). Askew and El-Kassaby (1994) working with *Pinus taeda*, plotted the relationship between the selfing rate and the additive variance coefficient and the relationship between the % overestimation of the additive genetic variance and the outcrossing rates. Using this relationship for *M. oleifera*, a 26% selfing results in an increase of the coefficient of relationship from the idealized 0.25 to 0.32 which can lead to overestimation of the proportion of additive genetic variance by 30%.

According to El-Kassaby et al (1994), great overestimation will be achieved if selfing is ignored because selfing contributes additional factors to the covariance between relatives, including dominance and inbreeding depression effects and appropriate adjustments to estimators of quantitative genetic parameters are required. In studies carried out with *Eucalyptus regnans*, Griffin and Cotteril (1988), suggested that with an outcrossing rate of 0.57 to 0.76, a coefficient equal to 0.4 should be used to control inflation of heritability and genetic gain. Adjustments should be done to the *M. oleifera* seed orchard to avoid over-estimation of this parameter.

The presence of selfing as well as early sexual maturity (6 months to 1 year) might

provide an alternative breeding programme with this species as compared with traditional directional selection. If inbreeding depression is weak, then it should be possible to use breeding schemes involving inbred lines and hybridisation (El-Kassaby et al 1994). In designing seed orchards, randomisation and minimum distance between related individuals will need to be worked out to maximise cross-fertilisation among unrelated clones and to minimise selfing or mating among related ramets. The above suggestions are further strengthened by previous observations in India where hand-pollination with xenogamous pollen gave 100% fruit set, 81 % seed set and 9 % fecundity, while with geitonogamous pollen the respective rates were 62, 64 and 6 % (Jyoth 1990).

5.5 Conclusion.

The present work has helped provide information on the genetic quality of seeds from the Mbololo seed orchard and the importance of randomisation and minimum distance requirement in seed orchard layouts for this species and provides an opportunity for developing inbred lines and hybridisation because of the ability to self. However, future studies should focus on outcrossing rates of individuals and populations in relation to mechanisms (or environmental parameters) that favour either outcrossing or selfing.

CHAPTER SIX

6.0 Variation and interrelationships in three Kenyan provenances of *M. oleifera* based on agronomic and morphological traits.

6.1 Introduction.

An increasing interest in the quality and yield of the fruits of *M. oleifera* is at present shared by scientists and organisations concerned with improved nutrition, hunger-aid and water supplies in rural areas of tropical developing countries (Jahn 1989). Geographical variation in quantitative characters in *M. oleifera* such as pod and seed size has been reported (Ramachandran et al 1980). To make utilisation of *Moringa* trees economic by optimising production, it is essential that fast-growing trees with highest possible fruiting qualities are planted in the right environment and silvicultural / horticultural operations optimised (Jahn 1988). Quantitative characters have been a major area of genetical study for over a century because they are a common feature of natural variation in populations and are typical of commercially important traits in crops (Kearsey & Farquhar 1998). First attempts at studying quantitative characters stem from the work of Galton (1889) on man before the rediscovery of Mendel, through the pioneering work of Fisher (1918), Wright (1934), Mather (1949) and Falconer (1960) to the new era opened up by recent developments in molecular biology (Tanksley 1993).

Most of the important agronomic characters like yield and yield components (grain number, grain weight), plant height and days to flowering are controlled by polygenes (Mather 1949) but the number of genes and their interactive effects controlling the expression of these traits are poorly understood (Powell 1992, Waugh & Powell 1992, Bezant 1997, Mohan et al 1997). However, estimation of variation of quantitative genetic characters can provide valuable information for making conservation decisions and give an indication of how a species responds to selection (Storfer 1996). A successful selective breeding programme will depend on the availability of variation in quantitative characters for manipulation.

Allozymes and DNA based techniques have been used to assess genetic variation in many plant species. However, correlation between allozyme heterozygosity and additive genetic variance in quantitative traits has been found to be weak and surveys of DNA frequently analyse non-coding regions that are assumed to be selectively neutral or regions of unknown function (Avise 1994). In addition, use of morphological and agronomic traits in characterisation of genetic resources is limited by the phenomenon of phenotypic plasticity (Schmalhausen 1949, Pigliucci 1996). However, it is possible to obtain some basic information about the genetical architecture of quantitative characters from relatively simple experiments in randomised experimental field trials. Work described in this chapter, therefore, attempts to: i) evaluate and characterise *M. oleifera* germplasm from three Kenyan provenances (Mbololo, Likoni and Kibwezi) based on quantitative characters; ii) establish whether genetic differences and relationships determined in chapter three (based on DNA markers) could be reflected in agronomic and morphological characters; iii) estimate distribution of variation in quantitative characters among the three provenances.

6.2 Materials and methods.

6.2.1 Experimental design.

An *M. oleifera* provenance trial was planted in May 1996 at a spacing of 2.5 m x 2.5 m and assessed in May 1997. The experiment was established at Marigat, Baringo, in the Kenyan Rift Valley. It was a completely randomised design with four blocks and three treatments (provenances). Nine plants per provenance were planted in each block (36 trees per provenance) and a guard row surrounded the blocks. Seven quantitative and morphological characters of each experimental tree were measured. The characters measured and methods of measuring them (Ahmad et al 1997) are shown in Table 6.1.

6.2.2 Data analysis.

Morphological data were analysed using the Genstat statistical package (1995). Analysis of variance (ANOVA) and standard errors were used to test differences between the provenances. Variance components were estimated as proposed by Falconer and Mackay (1996).

6.3 Results.

Variation displayed by the seven characters among the three provenances is shown in Table 6.2. The highest coefficient of variation was shown in total yield, followed by single seed weight and the plant diameter. The lowest value was shown in fruit diameter, followed by fruit length, number of seeds per fruit and plant height. The Kibwezi provenance had the highest means for five out of seven characters measured

followed by Mbololo. The Mbololo provenance had the highest mean for total yield per tree only while the mean for fruit diameter was highest in Likoni provenance. Assessment of the pairs of means using standard error (Table 6.2) indicated that there were significant differences between Kibwezi and Mbololo for four characters (plant height and diameter, fruit length and number of seeds) and between Likoni for three traits (plant height and diameter and fruit length). However, there were no significant differences between the means for Mbololo and Likoni for any of the seven characters investigated. The analysis of variance is shown in Table 6.3. Significant differences were detected between the three provenances for six out of the seven quantitative characters measured, the exception being fruit diameter.

Table 6.1 Quantitative characters used to evaluate performance of three Kenyan provenances of *M. oleifera*.

Character	Unit	Where/ When measured
Plant height	m	From the ground level to the shoot tip after 1 year
Plant diameter	cm	At the ground level after 1 year
Fruit diameter	cm	Diameter of the widest girth of mature fruit
Fruit length	cm	From the base of mature fruits to the tip
No. of seeds / fruit		At maturity
Single seed weight	g	weight of 100 seeds from dry fruits chosen at random and averaged for a single seed.
Yield	g	Weight of all the dry seeds produced by each plant during the first year

The form of analysis for estimating the variance components is shown in Table 6.4 as proposed by Falconer and Mackay (1996). There are supposed to be p provenances with an average of k individuals per provenance. The values of mean squares are denoted by Ms_p and Ms_w. The mean square within provenance is itself the estimate of the within-provenance variance component but the other mean square is not the variance

component. The variance component of between provenance variation was estimated as shown in Table 6.4.

Table 6.2 Means, standard error (s.e) and coefficient of variation (C.V) for seven quantitative characters measured in 3 provenances of *M.oleifera.*

Character	Unit	Provenances			Grand mean	s.e	C.V
		Mbololo	Likoni	Kibwezi			
Plant height	m	4.91	4.86	5.90	5.22	0.25	21.94
Plant diameter	cm	8.83	8.60	10.88	9.44	0.52	23.2
Fruit diameter	cm	1.55	1.64	1.56	1.58	0.05	14.3
Fruit length	cm	35.70	35.4	45.0	38.7	1.56	17.1
No. of seeds / fruit		14.69	15.52	18.05	16.09	0.75	19.7
Single seed weight	g	0.21	0.20	0.26	0.22	0.02	36.8
Total yield per tree	g	294	113	197	202	62.2	131.0

Table 6.3 Analysis of variance (ANOVA) between and within the three provenances of *M. oleifera* based on seven quantitative characters.

Character	df		Mean square		V.r[1]	Prob.[2]
	Between	Within	Between	Within		
Plant height	2	91	12.40	1.10	11.24	P < 0.001
Plant diameter	2	91	56.45	4.81	11.74	P < 0.001
Fruit diameter	2	69	0.08	0.05	1.55	P = 0.219
Fruit length	2	71	1062	43.70	24.3	P < 0.001
No. of seeds / fruit	2	63	110.4	10.00	10.94	P < 0.001
Single seed weight	2	70	0.04	0.01	5.39	P = 0.007
Total yield per tree	2	70	295766	69740	4.24	P = 0.018

[1] Variance ratio
[2] Probability

Table 6.4 Form of analysis for estimating the variance components.

Source	df	Mean square	Estimate of variance components
Between provenance	p-1	Ms_p	$(1/k)(Ms_p-Ms_w)$
Within Provenance	k-1	Ms_W	Ms_w

Partitioning of the phenotypic variation into between and within provenances indicated that 78.1% of the total variation was due to within provenance variation (Table 6.5). However, the fruit length had almost equal components of between and within provenance variation. The highest percentage of within provenance variation was associated with the fruit diameter followed by the weight of a single seed while the fruit length accounted for the least.

Table 6.5 Estimation of variance components and its distribution between and within provenances.

Character	k	Ms_p-Ms_w	Between [$1/k(Ms_p-Ms_w)$]	Within [Ms_w]	Total	Distribution (%)	
						Between	Within
Plant height	31	11.30	0.365	1.1	1.465	24.9	75.1
Plant diameter	31	51.64	1.666	4.81	6.476	25.7	74.3
Fruit diameter	24	0.03	0.001	0.05	0.051	2.0	98.0
Fruit length	25	1018.25	40.73	43.75	84.480	48.2	51.8
No. of seeds / fruit	22	100.40	4.564	10.00	14.564	31.3	68.7
Single seed weight	24	0.03	0.001	0.01	0.011	9.1	90.9
Total yield / tree	24	226026	9418	69740	79158	11.9	88.1

6.4 Discussion.

Comparison of means for the three provenances using standard errors indicated a wide divergence between Kibwezi and the other two provenances. In addition, lack of significant differences between the means for Mbololo and Likoni for all the seven quantitative characters suggests a close relationship between the two provenances. A nucleotide sequence variation survey based on RAPDs and AFLPs (Chapter 3) grouped Mbololo with Likoni but grouped Kibwezi separately in agreement with the results found in the present work. Waugh & Powell (1992) argued that individuals expressing high or low phenotypic scores for the measured traits are expected to differ at most of the loci controlling the characters. The present work therefore suggests that differences found between provenances in chapter 3 can also be reflected in their quantitative and morphological characters.

Most of the phenotypic variation was found within the provenances supporting results found in chapter 3 and 4, based on nucleotide sequence survey, which revealed high levels of within provenance differentiation and lack of uniformity within provenances. The variability and high coefficient of variation observed in most of the quantitative characters examined offer ample scope for a selective breeding programme for the species. Results suggest that higher genetic gains may be achieved through selection within the Kenyan provenances for all the characters investigated. The exceptionally high yield found within the Mbololo provenance suggest that a selection programme within this provenance can dramatically improve *M. oleifera* seed yield. Although the overall level of between provenance variation based on quantitative characters is low, a substantial component of between provenance variation was found in fruit length and the number of seeds per fruit. This observation coupled with the significant differences

in means for most of the characters between Kibwezi and the other provenances suggests that provenances should be treated separately for the purpose of conservation, management and utilisation. Templeton (1986) urged caution in management programmes that entail the mixture of populations (with fixed allelic differences) because, while such management strategies have the potential to increase genetic variation and offset inbreeding depression, they can result in the breakdown of coadapted gene complexes leading to a decline in population fitness (outbreeding depression) when populations are adapted to different local conditions. The relationships based on agromorphological traits might be due to long term local adaptation, restricted gene flow, possible multiple origins (Chapter 3) and biased selection for specific traits (field observation by the author) and will be of value in exploitation and management programmes for the Kenyan provenances.

Seed weight of *M. oleifera* observed in this study (0.20 - 0.26g) is slightly higher than (0.18 - 0.2 g) reported by Jahn (1989) for Kenyan *M. oleifera* and 0.19 - 0.2 (Jahn 1988) for *M. oleifera* growing in Sudan. However, seeds with cotyledons of an average weight of 0.3 - 0.32 g have been reported in Guatemala (Jahn 1989), Haiti (0.25 - 0.27 g) and South India (0.26 - 0.28 g) (Jahn 1988). Results suggest that although heavy seeds can be realised through selection within the Kibwezi provenance, they can also be achieved through introduction from suitable clones from other countries. Comparative water treatment experiments have shown that equal amounts of powder from small and large seeds of *M. oleifera* of different origin has almost the same clarification effect (Jahn 1989). The total amount of flocculating material which a single tree can provide per year will therefore depend on the seed size (weight) and annual seed yield. Additionally, large seeds have the practical advantage of being easier to clean and if at the same time

less seeds are needed, the preparation of seed suspension will require less time. Ibrahim et al (1974) reported that seeds of *M. oleifera* contain 25-34 % edible oil. Therefore, the heavier the seeds of *M. oleifera* the higher the quantity of oil that can be produced.

Fruit length of *M. oleifera* observed in this study (35.4 - 45.0 cm) can be compared with that reported in an earlier work by Morton (1991) and Jahn (1989) for Kenyan *M. oleifera* (25 - 40 cm). The fruits of Kenyan *M. oleifera* can be described as medium sized according to the classification proposed by Jahn (1989) and Ramachandran et al (1980). Great variability in fruit length in *M. oleifera*, ranging from 15 - 120 cm, has been reported in the homeland India (Ramachandran et al 1980). Fruits of equal length do not necessarilly have the same fruit quality (weight and number of seeds) (Jahn 1989). Morton (1991) reported that fruits of *M. oleifera* could achieve a diameter of 2 cm, which is slightly wider than means for this trait found in the present work.

Jahn (1988) reported that *M. oleifera* trees grown from seeds can reach heights of 3 - 5m during the first year, even if planted on marginal soils and can fruit within this period, which is in agreement with observations made in the present work. The Kibwezi provenance, though lower yielding than Mbololo, is fast growing compared to the other Kenyan provenances, achieving a mean height of 5.9 m within one year, which was above the average reported for *M. oleifera* in other regions (Jahn 1988). Apart from genetic qualities, disturbances in pollination and rare diseases affecting flowers and fruits, poor seed yields are mainly due to drought and horticultural / silvicultural shortcomings (close spacing, inadequate soil drainage) and socio-cultural practices (careless stripping of edible leaves and flower panicles, collection of flowers and exposure to browsing animals) (Jahn 1989). Application of farm yard manure and

ammonium sulphate have been found to considerably increase yield of *M. oleifera* (Jahn 1991).

6.5 Conclusion.

This study gives a clear assessment of *M. oleifera* germplasm in Kenya and generally reveals that within provenance selection for quantitative characters is the key to a successful selective breeding programme to improve the yield of Kenyan *M. oleifera.* However, to guarantee high quality products along with speedy fruit production, large seeds and high yields, cultivation should be co-ordinated by global exchange of seed materials from the most suitable clones.

CHAPTER SEVEN

7.0 Population genetic structure in *M. stenopetala* revealed by RAPD and AFLP markers and its relationship with *M. oleifera.*

7.1 Introduction.

Seed from all species of *Moringa* contains flocculant components (Jahn 1984, Jahn et al 1986, Muyibi and Evanson 1995), but those of *Moringa stenopetala* have the strongest flocculating effect per unit volume because of the quality and quantity of the flocculating (Jahn 1989, Jahn 1991) and anti-microbial substances (Jahn 1989, Jahn 1991, Mayer and Stelz 1993). The quantitative differences can be associated with seed size which are twice as heavy in *M. stenopatala* compared to *M. oleifera*. However, details on the amount and distribution of genetic variation and the genetic relationship between the various Kenyan provenances of *Moringa* species are currently lacking. Such information will be necessary for making conservation, utilisation and improvement decisions.

The genetic structure of populations refers to the distribution of genetic variation within and between populations (Alvarez-Buylla and Garay 1994) and is largely a consequence of the mating system, the rate at which genes move between habitat patches (McCauley 1997) and selection acting at the microhabitat level (Berg & Hamrick 1995). Many studies have shown the nonrandom distribution of genetic variation in natural

populations and emphasised the importance of understanding its spatial structure (Loveless and Hamrick, 1984, Jain, 1990). When gene flow between spatial units is restricted, short distance genetic differentiation will occur, either because selection is heterogeneous, or because of genetic drift, or both (Endler 1977). Even when gene flow is not restricted, genetic differentiation may still occur if selection is heterogeneous and strong (Bonnin et al 1996). Population genetic structure within a species has recently received even more attention, with increasing concern in the area of conservation biology, in which the interplay of population dynamics and genetic variability is of special interest (Bonnin et al 1996). Despite this intensified interest, studies of the population genetics of tropical agroforestry trees remain limited (Schierenbeck et al 1997). Moreover, the knowledge of both within and among population genetic differentiation will help develop efficient sampling strategies (Bonnin et al 1996), utilisation (Sharma et al 1995) and improvement / breeding (Chalmers et al 1992, Nesbitt et al 1995, Salimath et al 1995) of plant genetic resources.

Beentje (1994) gave detailed classification separating *M. stenopetala* and *M. oleifera* on morphological characters. The two species have similar uses, chemotaxonomical relationships and ecological requirements but their relationship based on molecular markers has not been reported. Molecular markers provide great potential for resolving interspecific structures in plant populations (Arnold et al 1991). The genetic relationship between *M. oleifera* and *M. stenopetala* was investigated using AFLP markers. These markers provide an efficient method of estimating genetic relationships and have been used successfully in phylogeny (Sharma et al 1996). In a plant-breeding programme, estimates of genetic relationships can be useful for organising germplasm, for the identification of cultivars, assisting in the selection of parents for hybridisation, and

reducing the number of accessions needed to ensure sampling a broad range of genetic variability (Thormann et al 1994, Wachira et al 1997).

Random amplified polymorphic DNA (RAPD) (Welsh and McClelland1990, Williams et al 1990) markers have been used successfully to describe population genetic structure in many tropical tree species, e.g., *Gliricidia* sp. (Chalmers et al 1992), *Eucalyptus globulus* (Nesbitt et al 1995), *Caesalpinia echinata* (Cardoso et al 1998), *Cedrela ordorata* L (Gillies et al 1997). RAPD polymorphisms result from either a nucleotide base change that alters the primer binding site or an insertion or deletion within the amplified region (Williams et al 1990). The advantages and limitations of using RAPD were discussed in section 1.4.0.

The principle, advantages and disadvantages of the AFLP technique have already been discussed in section 1.5.0. The AFLP technique has been successfully deployed in DNA fingerprinting (Hongtrakul et al 1997) and genetic variation studies (Travis et al 1996, Pakniyat et al 1997, Perara et al 1998, Hartl and Seefelder 1998).

Given the proliferation of genetic markers, comparisons between techniques are inevitable in order to determine which technique is best suited to the issues being examined (Powell et al 1996, Russell et al 1997). Since each technique not only differs in principle but also in the way they can resolve genetic differences and in the taxonomical levels at which they can be most appropriately applied (Karp et al 1996), the utility of AFLPs and RAPDs in genetic analysis of *M. stenopetala* was investigated. These markers provide invaluable tools to study the structure of genetic variability of

natural populations (Bonnin et al 1996) and have not previously been used in genetic analysis of *M. stenopetala.*

Seeds of *M. stenopetala* trees growing on the Island Camp (Kokwa) and Simatian Islands of Lake Baringo have been used by the Njemps community for water clarification for many years, while the Burji living around Isiolo have traditionally been using leaves of *M. stenopetala* as a vegetable. With increasing population and awareness of the potential products from this species, pressure has been exerted on the genetic resources of *M. stenopetala* thus calling for its urgent conservation, proper management and utilisation. The *M. stenopetala* populations targeted in the present work are also the major seed sources in Kenya.

The specific objectives of the present study were to: (i) examine the utility of RAPDs and AFLP markers in assessing the level and distribution of genetic variation within and between these threatened populations of *M. stenopetala,* ii) assess the extent of genetic differentiation between populations and iii) establish the relationship between *M. oleifera* and *M. stenopetala* using AFLPs.

7.2.0 Materials and methods.

7.2.1 Plant material.

Three populations of *M. stenopetala* (see section 2.1) were used. For RAPD analysis, DNA was isolated from 10 individuals, from each population. Eight individuals from each population were included for AFLP analysis. Same DNA samples were used for

both investigations. For interspecific relationships, the seventy individuals of *M. oleifera* from 7 provenances (see chapter 3) together with the 24 individuals of *M. stenopetala* described above were analysed using AFLPs.

7.2.2 DNA isolation.

Genomic DNA was isolated following a modification of the procedure of Edwards et al (1991) described in section 2.2.1.

7.2.3 RAPD procedure.

The RAPD amplification reaction, gel electrophoresis and documentation were carried out as described in section 2.4.

7.2.4 AFLP procedure.

The AFLP assays were carried out as described in section 2.5 using six pairs of AFLP primers. Sequences of the primers and adaptors used are described in sections 2.5.4.1 and 2.5.2, respectively.

7.2.5 Data analysis.

Amplification products were scored as discrete character states (present / absent) and analysed as described in section 2.7.0.

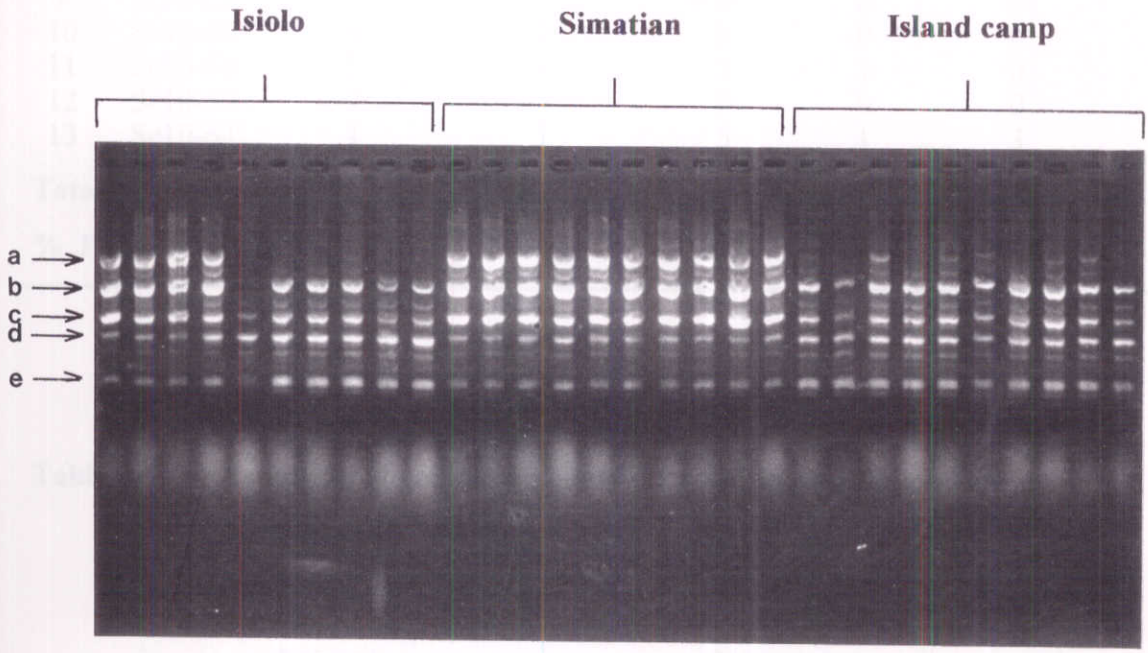
7.3.0 Results.

7.3.1 Population genetic structure and relationships in *M. stenopetala* based on RAPDs.

Although all 13 RAPD primers tested produced amplification products in all three populations, 4 primers (Sc10-16, Sc10-17, Sc10-30, Sc10-59) did not produce polymorphic products and were excluded from the statistical analysis. Analysis of the 30 accessions of *M. stenopetala* from the three populations using 13 primers, identified a total of 71 bands of which 40 (57.1%) were polymorphic. An example of a typical RAPD profile generated by primer Sc10-22 is shown in Fig. 7.1. The average number of amplification products per primer was 5.5. The number of polymorphic loci produced per primer ranged from 2 (for primer SC10-19) to 10 (for primer Sc10-25) (mean of 3.1). The percentage of polymorphic loci found in *M. stenopetala* ranged from 55 (Isiolo) to 67.5 (Simatian), Table 7.1.

The distribution of variability between and within populations was calculated using the Shannon's index of phenotypic diversity. Estimates of diversity (H_o) were calculated for each population (Table 7.2) and high diversity was detected in the Simatian population (0.841). However, some primers detected more variation in either Island Camp or Isiolo populations rather than Simatian populations (Table 7.2).

Fig. 7.1 An example of a molecular profile from *M. stenopetala* populations generated by RAPD primer Sc10-22.



Product sizes (bp).

a 5,150
b 3,500
c 1,870
d 1,700
e 1,580

Table 7.1 **Amplification products from *M. stenopetala* based on 13 RAPD primers.**

Primer No.	Primer Code	Amplification products Total	Polymorphic	%	Provenance I. Camp	Simatian	Isiolo
1	Sc10-5	5	5	100	2	3	4
2	Sc10-14	6	5	83	5	5	0
3	Sc10-16	4	0	0	0	0	0
4	Sc10-17	4	0	0	0	0	0
5	Sc10-19	4	2	50	2	1	2
6	Sc10-22	7	5	71	3	4	1
7	Sc10-23	6	3	50	2	0	0
8	Sc10-24	5	3	60	2	2	3
9	Sc10-25	10	10	100	6	9	6
10	Sc10-30	5	0	0	0	0	0
11	Sc10-58	7	4	57	1	0	4
12	Sc10-59	4	0	0	0	0	0
13	Sc10-64	4	3	75	1	3	2
Total		**71**	**40**		**24**	**27**	**22**
% Polymorphism				**56.3**	**60**	**67.5**	**55**

Table 7.2 **Estimates of genetic diversity (H_o) within populations of *M stenopetala* based on RAPD.**

Primer No.	Primer Code	Provenance I. Camp	Simatian	Isiolo
1	Sc10-5	0.485	0.859	1.208
2	Sc10-14	1.693	1.511	0.000
3	Sc10-19	0.552	0.322	0.644
4	Sc10-22	0.958	1.239	0.367
5	Sc10-24	0.628	0.722	0.802
6	Sc10-25	1.330	1.958	1.909
7	Sc10-58	0.307	0.000	1.072
8	Sc10-64	0.230	0.960	0.499
9	Sc10-23	0.325	0.000	0.000
Mean		**0.723**	**0.841**	**0.722**

The H_{pop} provides a measure of the average diversity within populations while H_{sp} provides a

measure of diversity within all populations considered together (Table 7 3). Primer Sc10-25 detected highest within population variation (1.718) and primer Sc10-58 the least (0.459) (Table 7.3). Primer Sc10-25 also detected high variability (2.171) within all populations considered together (species) while Sc10-19 detected the least (0.552). An examination of the relative proportion of diversity present within (H_{pop}/H_{sp}) and between (H_{sp}-H_{pop})/H_{sp} populations indicates that on average a higher level of diversity (61.8 %) is maintained within rather than between (38.2 %) populations (Table 7.3). Only primers Sc10-58 and Sc10-23 detected more variation between populations.

Table 7.3 Partitioning of genetic diversity between and within populations of *M. stenopetala* using RAPD markers.

Primer No.	Code	H_{pop}	H_{sp}	Within [H_{pop}/H_{sp}]	Between [(H_{sp}-H_{pop})/ H_{sp}]
1	Sc10-5	0.851	1.308	0.650	0.350
2	Sc10-14	1.068	1.748	0.611	0.389
3	Sc10-19	0.506	0.552	0.916	0.084
4	SC10-22	0.854	1.441	0.593	0.407
5	Sc10-24	0.718	0.800	0.897	0.103
6	Sc10-25	1.732	2.171	0.798	0.202
7	Sc10-58	0.459	0.993	0.462	0.538
8	Sc10-64	0.563	1.065	0.529	0.471
9	Sc10-23	0.108	1.024	0.106	0.894
Mean		**0.762**	**1.234**	**0.618**	**0.382**

Distribution of genetic variation using AMOVA found almost equal components of within (53.2%) and between (46.8 %) population differentiation (Table 7.4). A dendrogram based on 40 RAPD products first separated the three *M. stenopetala* populations into two main groups, one containing the Isiolo population and the other the two populations from Lake Baringo. Some three genotypes from Island Camp were further clustered with those from Simatian

Island (Fig.7 2). The rest of the Island camp genotypes formed two small clusters. The PCOa axis grouped each of the 3 populations separately with Simatian and Island Camp accessions being more closely associated (Fig.7.3).

Table 7.4 Analysis of molecular variance (AMOVA) between and within the three populations of *M. stenopetala* based on RAPD markers.

Source of variation	df	Variance components		Probability
		Variance	% variation	
Between populations	2	10.2	40.4	$P < 0.00001$
Within populations	27	15.1	59.6	$P < 0.00001$
Total	29	25.3		

Fig. 7.2 A dendrogram based on RAPDs showing the clustering of *M. stenopetala* individuals from the three populations investigated.

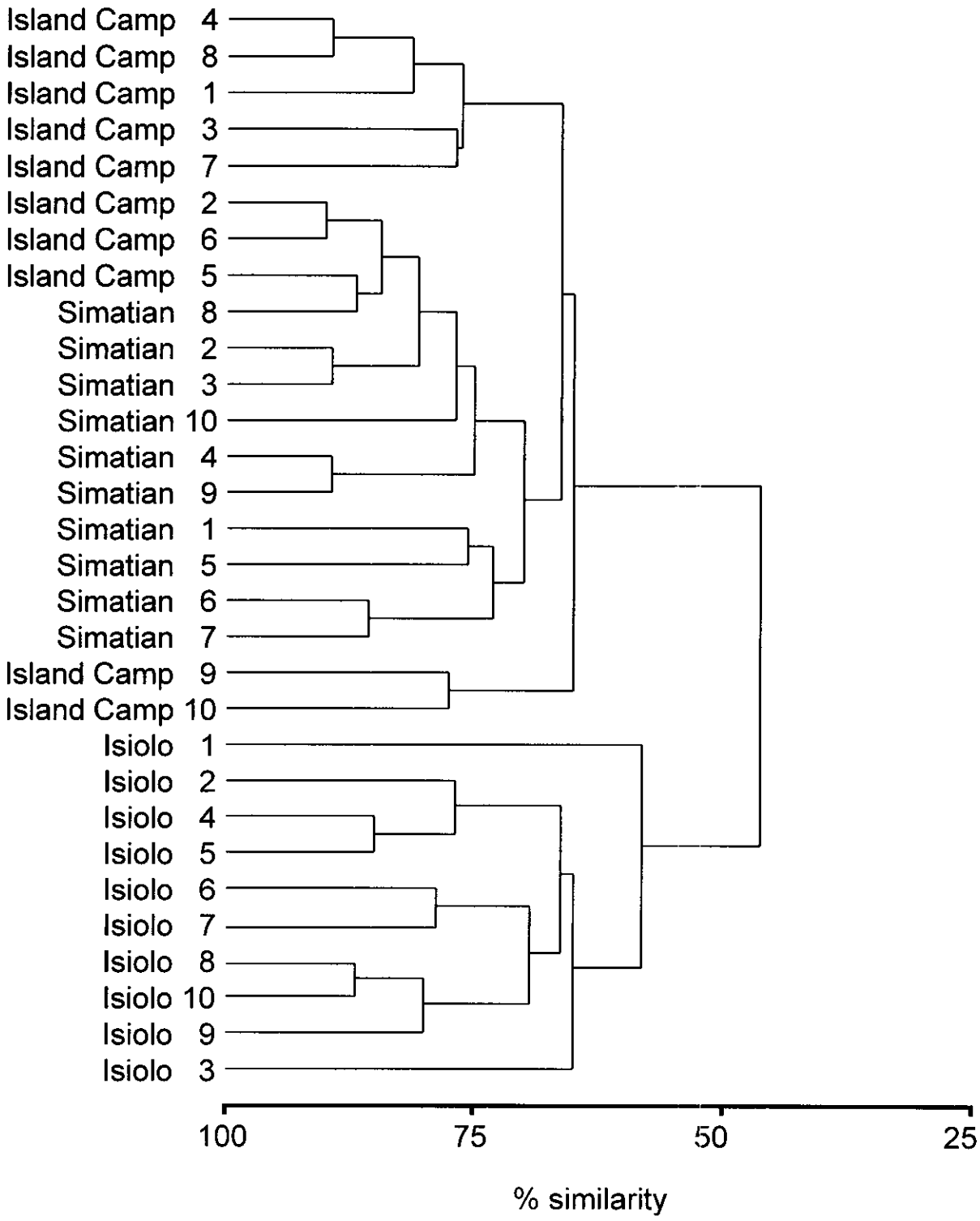
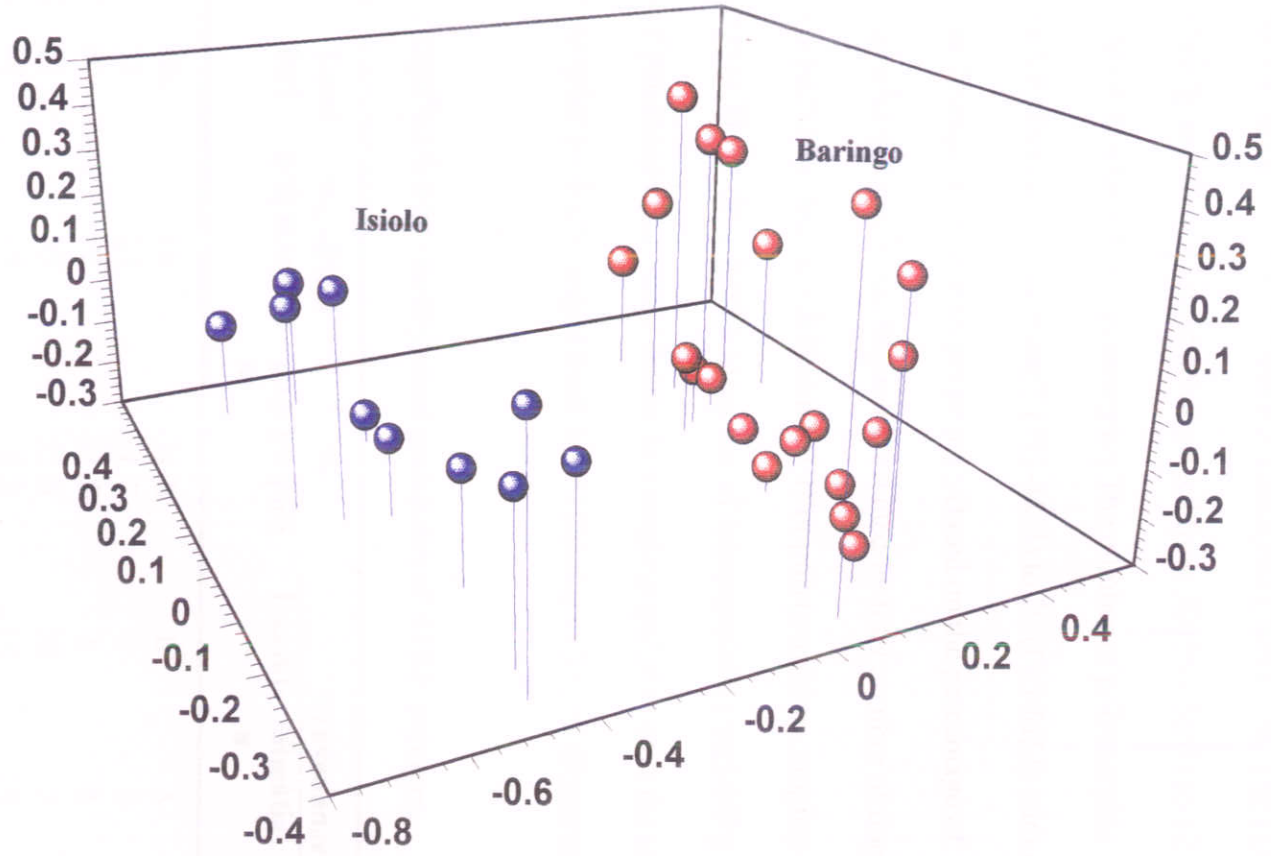


Fig. 7.3 3-Dimensional principal co-ordinate analysis based on RAPDs showing separation between Baringo and Isiolo populations of *M. stenopetala.*



7.3.2 Population genetic structure and relationships in *M. stenopetala* based on AFLP markers.

Analysis of 24 accessions of *M. stenopetala* with 6 AFLP primer pairs identified a total of 300 fragments, of which 175 (58.3%) were polymorphic (Table 7.5). The number of fragments detected by individual primer pairs ranged from 30 (P17-M51) to 82 (P15-M51) with a mean of 50 products per primer pair. The number of polymorphic fragments for each primer pair varied from 7 (P17-M51) to 59 (P15-M51) with an average of 29.2 polymorphic fragments per primer. Based on the percentage of polymorphic fragments (polymorphic fragments divided by total number of fragments observed), primer pairs also detected different levels of polymorphism, ranging from 23.3% (P17-M51) to 78.3% (P11-M51). Measures of intrapopulation variability based on the number of polymorphic products scored in a single population over the total number of polymorphic products ranged from 26.9% (Isiolo) to 37.7% (Simatian).

Table 7.5 Amplification products based on 6 pairs of AFLP primers.

No.	Primer Pair	Total loci	No. of loci polymorphic	% polymorphism	Provenance		
					I. camp	Simatian	Isiolo
1	P14M51	46	16	34.8	0	5	0
2	P12M51	33	22	66.7	4	5	2
3	P17M51	30	7	23.3	1	4	3
4	P18M51	63	35	55.6	30	7	11
5	P11M51	46	36	78.3	14	19	19
6	P15M51	82	59	72.0	5	26	12
Total			**300**	**175**	**54**	**66**	**47**
% Polymorphism				**58.3**	**30.9**	**37.7**	**26.9**

The phenotypic frequencies detected with the 6 primer pairs were calculated and used in estimating genetic diversity (H_o) within populations (Table 7.6). The average genetic diversity values ranged from 1.810 (Isiolo) to 2.590 (Simatian).

Table 7.6 Estimates of genetic diversity (H_o) within populations of *M. stenopetala* based on AFLP.

Primer		Provenance		
No.	Pair	I. Camp	Simatian	Isiolo
1	P14M51	0.000	1.602	0.000
3	P12M51	1.279	0.946	0.736
3	P17M51	0.117	0.527	0.736
4	P18M51	8.842	1.711	2.263
5	P11M51	3.160	4.568	4.363
6	P15M51	1.461	6.183	2.763
Mean		**2.477**	**2.590**	**1.810**

Partitioning of the phenotypic diversity into within and between population components using Shannon's index indicated that a higher component of the AFLP variation was due to within population variation, with an average value of 55.9% (Table 7.7). Further analysis of the population genetic structure based on AMOVA found also a higher component of within population variation (58.1%) in *M. stenopetala* (Table 7.8).

The AFLP dendrogram generally grouped individuals according to their respective populations with a few exceptions (Fig. 7.4). For example, although all the 8 genotypes from Simatian were grouped together, 5 genotypes from Island Camp were more closely association with the Simatian population while 3 genotypes from Island camp (1, 4 & 8) formed a cluster of their own.

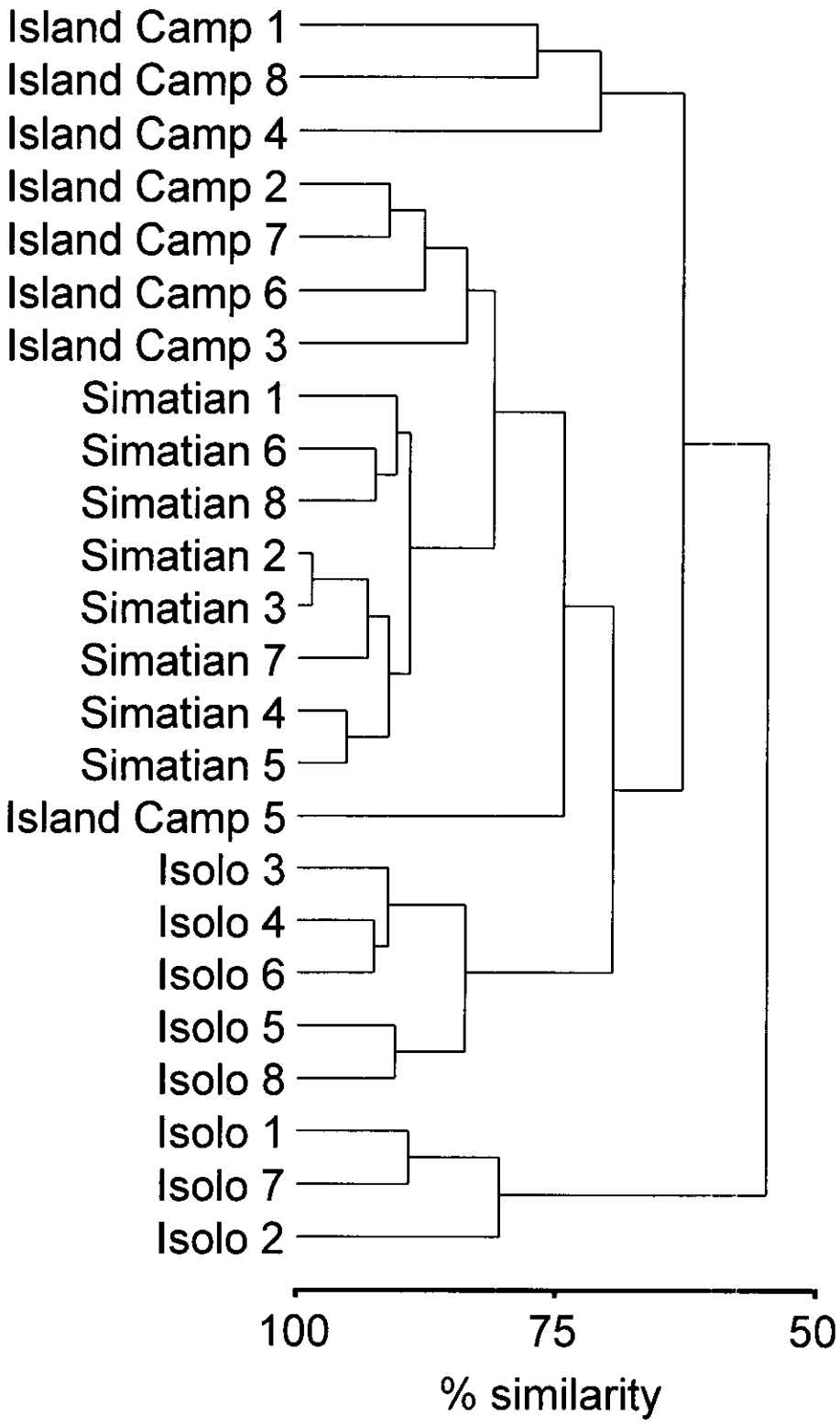
Table 7.7 Partitioning of genetic variation within and between populations using Shannon's index based on AFLP markers.

No.	Primer Pair	H_{pop}	H_{sp}	Within $[H_{pop}/H_{sp}]$	Between $[(H_{sp}-H_{pop})/H_{sp}]$
1	P14M51	0.534	0.798	0.669	0.331
2	P12M51	0.987	1.299	0.760	0.240
3	P17M51	0.460	1.767	0.260	0.740
4	P18M51	4.272	6.851	0.624	0.377
5	P11M51	4.030	5.600	0.720	0.280
6	P15M51	3.469	10.770	0.322	0.678
	Mean	**2.292**	**4.515**	**0.559**	**0.441**

Table 7.8 Analysis of molecular variance between and within populations of *M. stenopetala* based on AFLP.

Source of variation	df	Variance components		Probability
		Variance	% variation	
Between populations	2	4.5	41.9	$P < 0.00001$
Within populations	21	6.2	58.1	$P < 0.00001$
Total	23	10.8		

Fig. 7.4. A dendrogram based on AFLPs showing the relationship between 24 *M. stenopetala* individuals from three populations.



7.3.3 Comparison between RAPDs and AFLPs in genetic analysis of *M. stenopetala.*

The mean average heterozygosity, effective multiplex ratio and marker index were calculated for each population and all the populations considered together (Tables 7.9 & 7.10). The level of polymorphism (mean average heterozygosity) was highest in Simatian and least in Isiolo for both assays. The Simatian population also had the highest mean effective multiplex ratio and marker index for both assays while Isiolo had the least (except the effective multiplex ratio based on RAPDs which was the same as that of Island Camp). When all the populations were considered together, the RAPD assay had the highest average heterozygosity while the AFLP had a higher effective multiplex ratio and marker index.

Table 7.9 Average expected heterozygosity (H_{av}), effective multiplex ratio (E) and marker index (MI) based on RAPD analysis.

Primer		Island camp			Simatian			Isiolo			Species		
No.	Code	H_{av}	E	MI	H_{av}	E	MI	H_{av}	E	MI	H_{av}	E	MI
1	Sc10-5	0.16	0.8	0.128	0.196	3.0	0.588	0.196	3.2	0.627	0.331	5.0	1.656
2	Sc10-14	0.413	4.2	1.722	0.293	4.2	1.222	0.000	0.0	0.000	0.314	4.2	1.309
3	Sc10-19	0.125	1.0	0.125	0.08	0.3	0.02	0.16	1.0	0.160	0.125	1	0.125
4	Sc10-22	0.191	1.3	0.246	0.234	2.3	0.536	0.069	0.1	0.010	0.253	3.6	0.904
5	Sc10-23	0.060	0.7	0.04	0.000	0.0	0.000	0.000	0.0	0.000	0.229	1.5	0.343
6	Sc10-24	0.160	0.8	0.128	0.168	0.8	0.134	0.184	1.8	0.331	0.196	1.8	0.354
7	Sc10-25	0.178	3.6	0.641	0.222	8.1	1.798	0.212	3.6	0.763	0.251	10.0	2.511
8	Sc10-58	0.069	0.1	0.010	0.000	0.0	0.000	0.214	2.3	0.490	0.217	2.31	0.496
9	Sc10-64	0.045	0.3	0.011	0.365	2.3	0.821	0.210	1.0	0.210	0.372	3.0	0.836
Mean		**0.156**	**1.4**	**0.339**	**0.173**	**2.3**	**0.569**	**0.142**	**1.4**	**0.288**	**0.254**	**3.6**	**0.948**

Table 7.10 Average expected heterozygosity, H_{av}, effective multiplex ratio, E, and marker index, MI, based on AFLP analysis.

Primers		Simatian			Isiolo			Island Camp			Species		
No.	Pair	H_{av}	E	MI	H_{av}	E	MI	H_{av}	E	MI	H_{av}	E	MI
1	P14M51	0.425	5.0	2.125	0.000	0.0	0.000	0.000	0.0	0.000	0.203	5	1.01
2	P12M51	0.232	4.2	5.332	0.156	0.7	0.104	0.327	2.7	0.871	0.311	6	1.86
3	P17M51	0.161	2.3	0.367	0.116	1.3	0.149	0.031	0.1	0.004	0.348	7	2.43
4	P18M51	0.074	1.4	0.104	0.093	3.5	0.322	0.372	25.7	9.574	0.314	35	10.99
5	P11M51	0.272	13.9	3.781	0.267	13.9	3.703	0.202	7.5	1.524	0.324	26	8.42
6	P15M51	0.210	15.7	3.30	0.102	3.4	0.343	0.050	0.6	0.029	0.351	43	15.09
Mean		**0.229**	**7.1**	**2.486**	**0.122**	**3.8**	**0.770**	**0.164**	**6.1**	**2.000**	**0.309**	**20.3**	**6.63**

7.3.3 Relationship between *M. stenopetala* and *M. oleifera* based on AFLP markers.

The 6 pairs of AFLP primers generated a total of 470 amplification products, of which 391 (83.2 %) were polymorphic (Table 7.11). Each primer generated 38-104 amplification products with an average of 78.3 fragments per primer. The number of polymorphic fragments generated by each primer ranged from 25 (P17-M51) to 93 (P15-M51) with an average of 65.2 polymorphic products per primer. The percentage of polymorphic loci calculated for each species as a ratio of the number of polymorphic loci scored in that species to the total number of loci polymorphic between the two species was 50.9% for *M. oleifera* and 38.0% for *M. stenopetala.* Amplification products generated from each species and the corresponding polymorphism have already been described in section 7.3.2 (*M. stenopetala*) and in chapter 3 (*M. oleifera*). Examples of species specific markers are shown in Fig. 7.5.

Table 7.11 Amplification products generated with 6 pairs of AFLP primers from *M. oleifera* and *M. stenopetala.*

Primer No.	Pair	Amplification products Total	polymorphic	%
1	P14M51	64	43	67.2
2	P12M51	62	57	91.9
3	P17M51	38	25	65.8
4	P18M51	103	84	81.6
5	P11M51	104	89	85.6
6	P15M51	99	93	93.9
	Total	470	391	
	Mean	**78.3**	**65.2**	**83.2**

Analysis of molecular variance (AMOVA) detected significant differences (P<0.00001) between *M. oleifera* and *M. stenopetala* species (65.31%) as well as among populations within species (15.63%) (Table 7.12). The within population component accounted for 19.06% of the total variation.

Table 7.12 Analysis of molecular variance (AMOVA) between *M. oleifera* and *M. stenopetala* based on AFLP.

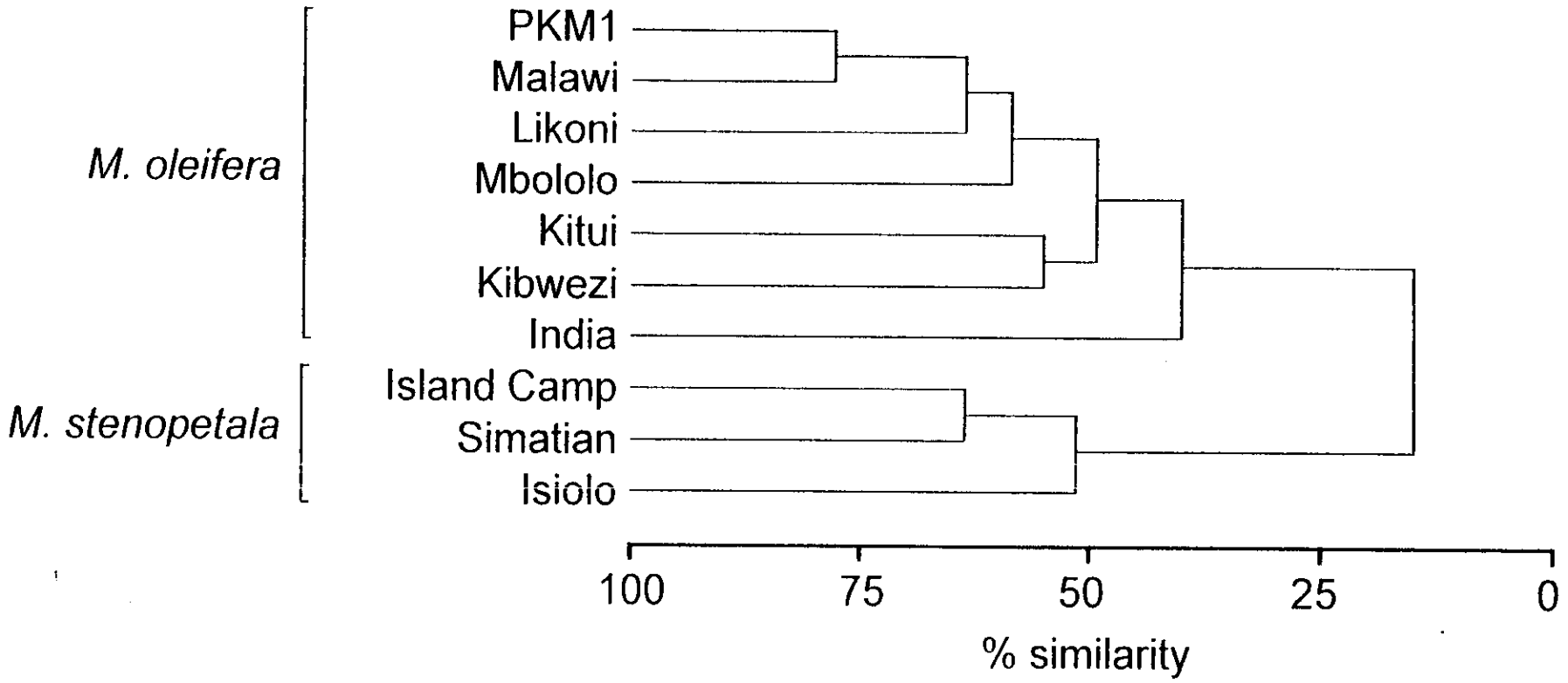
Source of variation	df	Variance components variance	% variation	Probability
Between the species.	1	14.34	65.3	P<0.00001
Among populations within species	8	3.43	15.6	P<0.00001
Within populations	84	4.18	19.0	P<0.00001
Total	93	21.95		

A dendrogram based on pairwise Fst comparisons of 391 AFLPs among the seven populations of *M. oleifera* and three of *M. stenopetala* is shown in Fig. 7.6. The dendrogram effectively separated the two species with each population forming a unique cluster.

Fig. 7.5 An example of an AFLP molecular profile generated with primer pair P15/M51 showing some markers specific to *M stenopetala*.

A *M. stenopetala* specific loci.

Fig. 7.6 A dendrogram based on AFLPs showing inter- and intraspecific relationships between *M. oleifera* and *M stenopetala.*



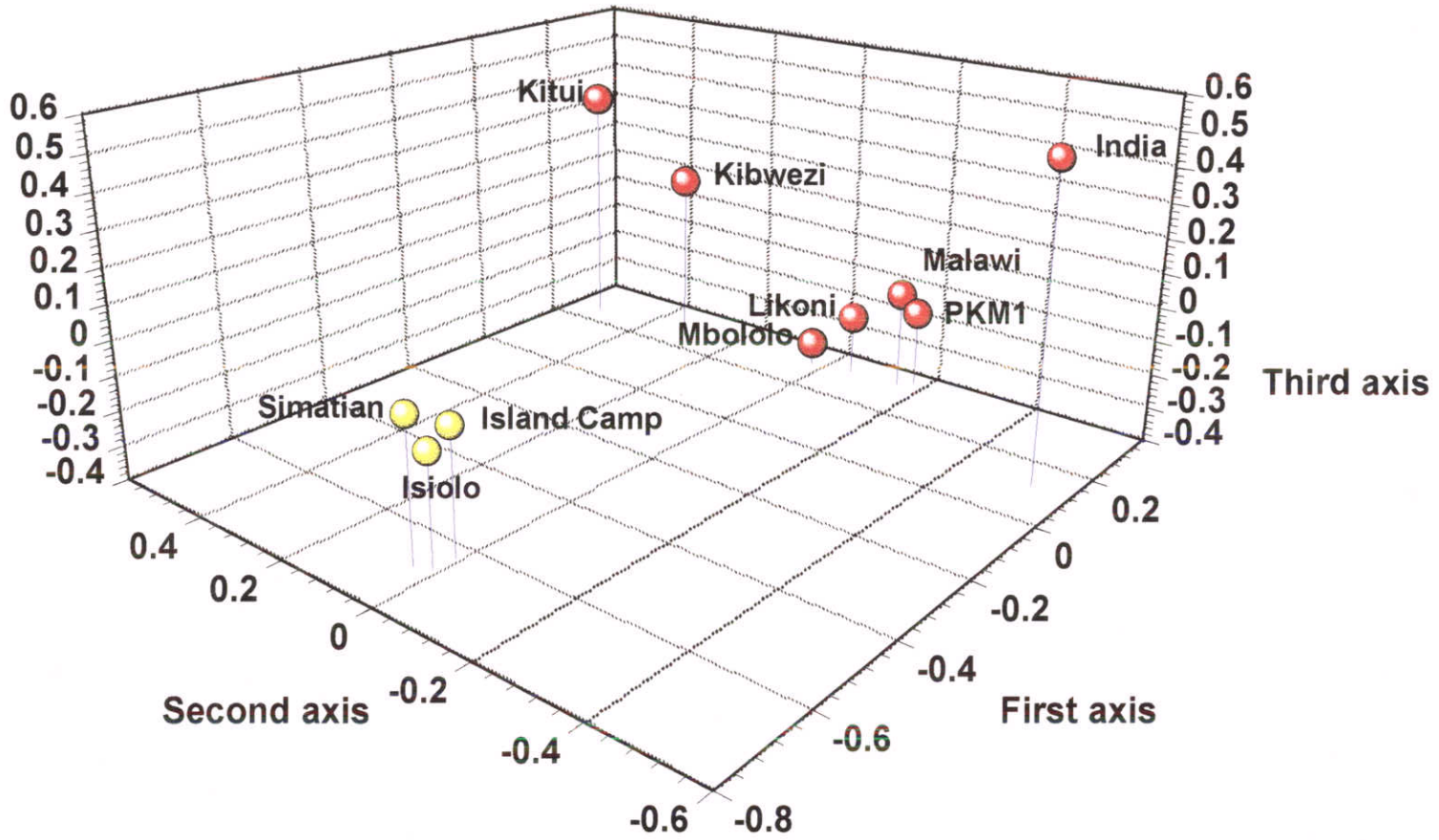


Fig. 7.7 3-Dimensional principal co-ordinate analysis based on AFLPs showing the relationship between 7 provenances of *M. oleifera* and *3 of M. stenopetala.*

In order to assess whether the grouping of populations could be further resolved, principal co-ordinate analysis (PCOa) based on similarity matrix was used to examine the shared fragment data available among individual genotypes for each population. Populations grouped according to their respective species. The PCOa (Fig. 7.7) provided a clear separation of the two species and supported the previous forms of analysis.

7.4.0 Discussion.

The majority of RAPD primers (9 out of 13) used in this study generated polymorphic fragments in the individuals studied. The failure by some RAPD primers to generate polymorphism has already been reported in the genus *Vigna* (Kaga et al 1996). All the six pairs of AFLP primers used generated polymorphism demonstrating the usefulness of AFLPs in discriminating between accessions within populations of *M. stenopetala* compared to RAPDs.

The distribution of genetic variation in *M. stenopetala* using both RAPD and AFLP based on two forms of analysis, AMOVA and Shannon's index, indicates that most of the variation is found within populations. For example AMOVA analysis gave 59.1% and 58.1% of within population variation for RAPD and AFLP repectively. Similar studies using isozymes (Hamrick 1990) on genetic variation among outcrossing tropical tree species have found most of the variation within populations. However, in *M. oleifera* (chapter 3) there was a high component of between provenance variation. The extent and distribution of variation revealed by RAPD and AFLP in *M. stenopetala* is in agreement with the conclusion that long-lived woody species with a wide and

continuous range have higher genetic diversity within populations and lower differentiation among populations (Hamrick & Godt 1989). *M. stenopetala* is a long-lived woody perennial with a wide distribution in Northern Kenya and this may account for the higher values of within population variation and lower population differentiation in this species than *M. oleifera*. The breeding system is a major determinant of a species' genetic structure (Ohara et al 1996) but information on the breeding system of *M. stenopetala* is completely lacking. The high component of within population variation revealed by this analysis suggest that one would be able to capture a wide range of diversity available in *M. stenopetala* by sampling from one locality.

The pattern of polymorphism revealed by RAPDs and AFLPs in the three populations of *M. stenopetala* using Shannon's index was generally the same. Similar observations were made in chapter 3 when both techniques were deployed in genetic analysis of *M. oleifera*. Both assays showed that the Simatian population was more polymorphic followed by the Island Camp and Isiolo the least. Some trees have been cleared to pave the way for settlement on Island Camp and this may account for the reduced diversity in this population compared to Simatian. The association of accessions from Simatian and Island Camp were not surprising since the two populations are located on two Islands situated within Lake Baringo thus increasing chances of gene flow between them. The Isiolo population was quite distinct from the other two and that could be related to restricted gene flow due to increased distance of separation (> 200 km) between it and the other populations. These results are consistent with the expectation that increased geographical distance results in an increase in population level differentiation for tropical tree species (Hamrick et al 1982, Schierenbeck et al 1997). While there was clear separation between individuals belonging to the same population some individuals

within a population appeared distinct. These divergent individuals may be useful for screening and inclusion in a core collection.

The present work has shown that polymorphism based on RAPD and AFLP markers can be useful in determination of genetic relationships in *M. stenopetala.* Knowledge of genetic relationships when complemented with phenotypic data can reveal a source of desirable characteristics in closely related individuals (Obara-Okeyo and Kako 1998). To the breeder this is important since variation ultimately determines the potential for making gains from selection. It is also an important consideration for efficient rationalisation and utilisation of germplasm resources (Russell et al 1997). The exploitation of RAPDs and AFLPs would assist in the elimination of duplicates, allow rationalisation of existing seed sources and enable collecting expeditions to be targeted towards geographical areas possessing maximum levels of genetic diversity.

The use of RAPDs for comparative purposes relies on the assumption that similarity of fragment size is a dependable indicator of homology (Rieseberg 1996). Work on *Helianthus* (Rieseberg 1996), *Glycine* (Williams et al 1993), *Brassica* and *Raphanus sativa* (Thormann et al 1994) suggests that fragment size is a more reliable predictor of homology among closely related individuals or populations than at higher taxonomic levels. However, recent work by de Bustos et al (1998) indicates that similar sized fragments in the genus *Hordeum* demonstrated homology in samples for different individuals, populations or taxonomic units. RAPDs have also been criticised for lack of reproducibility (Neale and Harry 1994) but experience from the present work revealed that optimisation and strict adherence to similar experimental reactions and conditions can greatly increase reproducibility of RAPDs.

The current availability of several powerful PCR-based techniques for the detection of DNA polymorphism has made comparisons between them necessary. Each technique not only differs in principal, but also in the type and amount of polymorphism detected. In the present work, the efficiency of RAPD and AFLP markers in genetic analysis of *M. stenopetala* was compared. The AFLPs are the most efficient because they can reveal many polymorphic bands per single lane. The average number of polymorphic bands per lane or per PCR for AFLPs was 29.2, compared to 3.1 for RAPDs. When the overall diversity indices of the two techniques was compared, AFLP was the highest (0.309) compared to 0.254 for RAPD. Powell et al (1996) introduced the concept of Marker Index as an overall measure of marker efficiency, and they demonstrated that, in *Glycine,* AFLPs had the highest Marker Index compared to other available marker systems. In the present work, AFLPs had the highest values for Marker Index (6.63) and effective multiplex ratio (20.3) compared to 0.948 and 3.6 respectively for RAPDs. Similar studies on potato (Milbourne et al 1997), barley (Russell et al 1997) and soybean (Powell et al 1996) found a high marker index and effective multiplex ratio associated with AFLPs compared to RAPDs. The high Marker Index (diversity index) is a reflection of the efficiency of AFLPs to simultaneously analyse a large number of bands rather than the levels of polymorphism detected (Russell et al 1997, Powell et al 1996).

Current results describing the patterns of genetic variation within and between populations of *M. stenopetala* have a number of important implications for the conservation and sustainable use of this species. In particular, the highly significant differences recorded between the Baringo and Isiolo populations indicates that

populations occurring in both of these areas need to be adequately conserved if the full breadth of genetic variation across the species is to be maintained. In the event of large-scale plantations of *M. stenoptala*, care should be taken to ensure provenance identity and utilisation of broad based seeds. In general, future germplasm-collecting strategies, whether for genetic conservation, tree improvement programmes or agroforestry purposes, should take into account of the occurrence of population differentiation between localities. As indicated by the current investigation, a high degree of genetic variation is evident within individual populations and it must be maintained by sustainable forest management strategies.

Increased human population activity on the two islands within Lake Baringo, especially the Island Camp (Kokwa) calls for urgent conservation measures to be put in place before this unique genetic resource is lost. Furthermore, only a few wild *M. stenopetala* trees are found within these populations and on the borders of the lake (Baringo) and Lake Turkana (Morgan 1981, Jahn 1991). Current work reports decreased diversity on Island Camp compared to Simatian island and that could be due to disturbance resulting from expansive hotel and human settlements.

AMOVA analysis based on AFLP data detected significant levels of differentiation within populations (19%), between populations (15.6%) and between species (65.3 %). Populations from each species were distinct and formed separate clusters. Despite the two species having basically the same uses and chemotaxonomical similarities (Jahn 1986), a wide divergence between them was clearly demonstrated in the present work, further confirming the present classification between the two species based on morphological data (Beentje 1994) and seed differences (Jahn et al 1986).

7.5.0 Conclusion

Evidence that much of the RAPD and AFLP variation in *M. stenopetala* is found within localities suggests that sampling from a few localities may capture a large proportion of variation within this species. Nevertheless, sampling from a wide range of localities is still advisable as substantial population differentiation was found. The current work is the first attempt to gain some insights into the genetics of *M. stenopeta* and may stimulate more detailed studies such as understanding the breeding system and variation in important agronomic characters. This is also to my knowledge the first time work on interspecific relationship between the two species using molecular markers has been reported.

CHAPTER EIGHT

8.0 Principal findings, implications and proposed future research.

8.1 Introduction.

The importance of *Moringa* has recently been recognised in East Africa and wide scale planting is currently being undertaken. Seeds are being acquired from existing Kenyan provenances whose genetic nature, potential for improvement and suitability as seed sources for massive planting is unknown. The species can be easily propagated vegetatively and because information is scarce regarding to what extent this may have been practised, a selective breeding and conservation programmes will be hampered by lack of information on whether selected ramets are members of the same genet or represent different clones. Therefore, the purpose of this thesis was to use molecular markers to generate information addressing specific questions of relevance for the overall management, utilisation and conservation of genetic resources within *M. oleifera* and *M stenopetala.*

8.2 Principal findings.

When RAPDs and AFLPs markers were used to determine population genetic structure, it was found that in *M. oleifera* most of the variation is found between populations (chapter 3) while in *M. stenopetala* most variation is found within populations (chapter 7). Distribution of phenotypic variation among three Kenyan provenances of *M. oleifera*

based on agronomic and morphological characters found most of the variation (78.1 %) within provenances. Partitioning of genetic variation between the two species using AFLPs detected 65.3% of the total variation between species. The practical implication of this finding is that although seeds of *M. stenopetala* may be pooled, provenance collection is recommended for *M. oleifera*. An *ex-situ* conservation program involving *M. oleifera*, would require that seed collection be done across the species range in order to achieve adequate sampling. Results suggest that high genetic gains may be realised through selection within individual provenances than between provenances for the various important traits. However, there was a substantial component of between provenance variation in fruit length and number of seeds per pod and this should be taken into account in any improvement programme. The pattern of genetic structure in *M. stenopetala* implies that adequate sampling can be achieved from a few localities thus drastically reducing costs associated with the establishment of a core collection in a conservation programme. However, care should be taken to ensure that the high variability within provenances is not eroded.

In order to determine to what extent vegetative propagation has been practised and its associated genetic effects, DNA fingerprinting was carried out using individuals sampled from two populations frequently used as seed sources (chapter 4). Data revealed extensive sexual reproduction implying that conservation and improvement programmes can be considered similar to that of other sexually reproducing species. Lack of common multiband phenotypes among individuals sampled suggest that the numerous morphological units (ramets) that appear distinct are most likely members of different genets. This observation implies that conservation or selective breeding programme targeting agromorphological characters of economic importance is viable

and it has alleviated fears that in Kenya there may be single clones resulting from many years of vegetative propagation.

Agromorphological characters provided further evidence that *M. oleifera* provenances may require different management schemes to optimise yield and that they should continue to be considered distinct from each other during conservation and exploitation. Based on variability and high coefficient of variation, specific traits were identified which can be manipulated in a selective breeding programme to achieve high genetic gains. When Kenyan germplasm of *M. oleifera* was compared with that existing in other countries (Ramachadran et al 1980, Jahn 1988, Jahn 1989), recommendations were given on specific traits which can be greatly improved by introductions from other countries (chapter 6). Among Kenyan provenances, the Kibwezi recording high and significant differences in means for most of the characters except total seed yield which was higher in Mbololo provenance. This observation suggests that a breeding programme designed to capture and combine desirable qualities evident in individual populations should be considered.

For the first time, the proportion of selfing and outcrossing in *M. oleifera* was estimated using AFLP markers (chapter 5). Data revealed a mixed mating system with 26% selfing suggesting that randomisation and minimum distance requirements in layout of *M. oleifera* seed orchards need to be properly worked out. This ability to self, coupled with early maturity, may provide an opportunity for developing inbred lines and hybridisation thus producing high quality genotypes of *Moringa.*

8.3 Other findings.

Procedures suitable for isolation of high quality genomic DNA from *Moringa* were developed through modification of existing protocols (chapter 2). The modified SDS method produced a high quality and concentrated DNA (17.5 μg/ml from 0.1 g of leaf tissue) compared to other methods and this DNA was suitable for RAPDs and AFLPs. This protocol may be useful for future molecular work for both species and save time associated with optimisation of DNA isolation.

The AFLP technology is extremely robust and proficient at revealing intra-population diversity and estimating genetic distance between individuals and populations (Travis et al 1996, Arens et al 1988, Winfield et al 1998). Furthermore, four times the number of data points (amplification products) were generated with AFLPs compared with RAPDs (chapter 3) over an equivalent period of time. These factors contribute to the conclusion that AFLP provides a cost-effective procedure to monitor the extent and distribution of diversity in *Moringa* and other agroforestry species. However, the inherent simplicity, efficiency and non requirement for radioactive materials of RAPD assay coupled with the observed high levels of polymorphism and genetic relationships in this work indicate the potential of this technique in genetic and breeding studies in *Moringa* in tropical developing countries where most laboratories are not equipped with radioactive facilities.

Historical factors and restricted gene flow between populations (Schaal et al 1998), self-compatibility (Cardoso et al 1998), and multiple introductions may explain the observed genetic structure in *M. oleifera* which was contrary to what is expected for woody, perennial, predominantly outcrossing species (Hamrick 1990). On the other hand,

continuous distribution and a relatively longer sexual maturity (Loveless 1992) could possibly account for the relatively high within population genetic component in *M. stenopetala.* A relatively low level of genetic diversity observed among introduced populations may be due to a founder effect at the time of introduction. Distribution range and population size were identified as major correlates of within population variation among the Kenyan populations. This observation may imply that cultivation of *M oleifera* over a wide area (increasing population size) may help safeguard the existing genetic diversity. The relationships based on agromorphological traits might be due to long term local adaptation, restricted gene flow, possible multiple origins, and biased selection for specific traits and can be exploited in breeding programmes to improve the Kenyan *M. oleifera.* The high degree of genetic variation evident within individual populations must be maintained by appropriate forest management strategies.

8.4 Future research.

With availability of resources, future research in *Moringa* could proceed as summarised below.

First, the association between the DNA marker(s) and quantitative traits should be determined and genes controlling these characters mapped to provide a better understanding of the genetical nature of agronomically and economically important traits in both species. Heritability of important agronomic traits such as yield and yield components in both species could be estimated by taking advantage of the short generation time or recently developed molecular techniques. Because of the ability to self and possible lack of inbreeding depression, the potential to hybridise and produce inbred lines with a view to developing high yielding *Moringa* plants should be

investigated. Additionally, selective breeding programmes targeting specific traits should be initiated to improve the yield of existing *M. oleifera* germplasm in Kenya.

Second, determination of the mating system and other biological and environmental factors affecting distribution of genetic diversity in *M. stenopetala* should be attempted. In *M. oleifera*, future work should examine how outcrossing rates vary over space and time by measuring outcrossing rates in different populations and years. Studies of gene flow, especially between the fragmented populations of *M.oleifera* in Kenya, may further help to unravel the population genetic structure found in the current work.

Third, examination of restriction-site variation of cpDNA and its implication for the evolution and taxonomy of *Moringa.* Genetic variation studies and subsequent conservation should be extended to other *Moringa* species found in the East African Lowlands as some of them may be threatened or endangered and require urgent attention and may possess desirable traits for inclusion in a *Moringa* improvement programme. The cpDNA, because of its low rate of sequence evolution, may be the most appropriate to use in phylogenetic studies in genus *Moringa.*

9.0 REFERENCES.

Abo-elwafa A, Murai K, Shimada T (1995) Intra- and inter-specific variations in *Lens* revealed by RAPD markers. *Theoretical and Applied Genetics*, 90: 335-340.

Ahmad M, McNeil DL and Sedcole JR (1997) Phylogenetic relationships in *Lens* species and their inter-specific hybrids as measured by morphological characters. *Euphytica*, 94: 101-111.

Aide TM (1986) The influence of wind and animal pollination on varion in outcrossing rates. *Evolution*, 40: 434-435.

Alvarez-Buylla ER and Garay AA (1994) Population genetic structure of *Cecropia obtusifolia*, a tropical pioneer tree species. *Evolution*, 48: 437-453.

Arens P, Coops H, Jansen J, Vosman B (1998) Molecular genetic analysis of black poplar *(Populus nigra* L) along Dutch rivers. *Molecular Ecology*, 7: 11-18.

Arnold ML, Buckner CM and Robinson JJ (1991) Pollen-mediated introgression and hybrid speciation in *Louisiana irises. Proceedings of the National Academy of Sciences of the USA*, 88: 1398-1402.

Askew GR and El-Kassaby AL (1994) Estimation of relationship coefficients among progeny derived from wind-pollinated orchard seeds. *Theoretical and Applied Genetics*, 88: 267-272.

Avise JC (1994) Molecular markers, Natural History and Evolution, Chapman and Hall.

Babu SC and Rajasekaran B (1991) Agroforestry, altitude towards risk and nutrient availability: a case study of south Indian farming systems. *Agroforestry Systems*, 15: 1-15.

Bachmann K (1994) Molecular markers in plant ecology. *The New Phytologist*, 126: 403-418.

Barret SCH and Eckert CG (1990) Variation and evolution of mating systems in seed plants. In: Biological Approaches and Evolutionary Trends in Plants (ed. S Kawano), 229.254. Academic Press, London

Barret SCH and Shore JC (1989) Isozyme variation in colonizing plants. In: Isozymes in Plant Biology (eds. DE Soltis and PS Soltis), 127-145. Dioscorides press, Portland, OR.

Bawa KS and PS Ashton (1991) Conservation of rare trees in tropical rain forests: a genetic perspective. In: Genetics and Conservation of Rare Plants (eds. DA Falk and KE Holsinger), 62-71. Oxford University Press, New York.

Beardmore JA (1983) Extinction, survival and genetic variation. In: Genetics and Conservation (eds. CM Schonewald-Cox, SM Chambers, B MacBryde and WL Thomas, 125-151. Benjamin/Cummings, Menio Park, CA.

Beentje HJ (1994) Kenya Trees, Shrubs and Lianas. National Museums of Kenya, Nairobi, Kenya.

Bennett MD and Leitch IJ (1995) Nuclear DNA amounts in angiosperms. *Annals of Botany,* 76:113-176.

Bennet MD and Smith JB (1976) Nuclear DNA amounts in angiosperms. *Philosophical Transactions of the Royal Society of London* B 274: 227-274.

Bennet MD and Smith JB (1991). Nuclear DNA amounts in angiosperms. *Philosophical Transactions of the Royal Society of London* B 334: 309-345.

Berg EE and Hamrick JL (1994) Spatial and genetic structure of the two scandhills oaks: *Quercus laevis* and *Quercus margaretta* (Fagaceae). *American Journal of Botany*, 81: 7-14.

Berg EE and JL Hamrick (1995) Fine scale genetic structure of a turkey oak forest, *Evolution* 49: 110-120.

Bezant J, Laurie D, Pratchett N, Chojeckj J and Kearsey M (1997) Mapping QTL controlling yield and yield components in a spring barley (*Hordeum Vulgare* L.) cross using marker regression. *Molecular Breeding*, 3: 29-38.

Bowditch BM, Albrright DJ, Williams GJK and Broun MJ (1993) Use of randomly amplified polymorphic DNA markers in comparative genome studies. *Methods in Enzymology,* 224: 294-309.

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

Botstein D, White RL, Skolnick M and Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, 32: 314-331.

Brown AHD and Allard RW (1970) Estimation of the mating system in open-pollinated maize populations using isozyme polymorphisms. *Genetics*, 66: 135-145.

Brown BA and Clegg MT (1984) Influence of flower color polymorphism on genetic transmission in a natural population of the common morning glory, *Ipomoea purpurea. Evolution*, 38: 769-803.

Caceres A, Cabrera O, Morales O, Mollinedo P and Mendia P (1991) Phamacological properties of *Moringa oleifera.* 1: preliminary screening for antimicrobial activity. *Journal of Ethnopharmacology*, 33: 213-216.

Caceres A and Lopez S (1991) Pharmacological properties of *M. oleifera*. 3. effects of seed extracts in the treatment of experimental pyodermia. *Fitoterapia* volume LXII: 449-450.

Cardoso MA, Provan J, Powell W, Ferreira PCG and De Oliveira DE (1998) High genetic differentiation among remnant populations of the endangered *Caesalpinia echinata* Lam. (Leguminosae-Caesalpinioideae). *Molecular Ecology*, 7: 601-608.

Chalmers KJ, Waugh R, Sprent JI, Simons AJ and Powell W (1992) Detection of genetic variation between and within populations of *Gliricidia sepium* and *G. maculata* using RAPD markers. *Heredity*, 69: 465-472.

Chand H, Singh R and Hameed SF (1994) Studies on the flowering period, density of Indian mustard and competing flora for honey bees in and around Pusa, Bihar (India). *Journal of Entomology Research*. 18: 361-367.

Chatfield C and Collins AJ (1980) Introduction to Multivariate analysis. Chapman & Hall London.

Clegg MT (1980) Measuring plant mating systems. *Bioscience*, 30: 814-818.

Cook R (1983) Clonal plant populations. *American Scientist*, 71:244-253.

Cottrell JE and White IMS (1995) The use of isozyme genetic markers to estimate the rate of outcrossing in a Sitka spruce (*Picea sitchensis* (Bong.) Carr.) seed orchard in Scotland. *New Forests*, 10: 112-122.

Dawson IK, Chalmers KJ, Waugh R and Powell W (1993) Detection and analysis of genetic variation in *Hordeum spontaneum* populations from Israel using RAPD markers. *Molecular Ecology*, 3: 151-159.

De Bustos A, Casanova C, Soler C and Jouve N (1998) RAPD variation in wild populations of four species of the genus *Hordeum* (Poaceae). *Theoretical and Applied Genetics,* 96: 101-111.

Doyle JA and Ritland K (1993) Inbreeding depression in two *Mimulus* taxa measured by multigenerational changes in the inbreeding coefficient. *Evolution*, 47: 361-373.

Dudash MR and Barret SCH (1989) The influence of environmental and demographic factors on the mixed mating system of *Sabatia angularis*. *American Journal of Botany*, 76: 99-100.

Edwards K, Johnstone C and Thompson C (1991) A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Research*, 19:1349.

El-Kassaby YA and Jaquish B (1996) Population density and mating system pattern in western larch. *Journal of Heredity*, 87: 438-443.

El-Kassaby YA, Russell J and Ritland K (1994) Mixed mating system in an experimental poulation of western red ceder, *Thuja plicata*. *Journal of Heredity*, 85: 227-231.

Ellis RP, McNicol JW, Baird E, Booth A, Lawrence P, Thomas B and Powell W (1997) The use of AFLPs to examine genetic relatedness in barley. *Molecular Breeding*, 3: 359-369.

Ellstrand AJ and Roose ML (1987) Patterns of genotypic diversity in clonal plant species. *American Journal of Botany*, 74: 123-131.

Endler JA (1977) Geographic Variation, Speciation and Clines. Princeton University Press, Princeton, New Jersey.

Ennos RA (1981) Quantitative studies of mating system in two symatric species of Ipomoea (Convolvulaceae). *Genetica*, 57: 93-98.

Ennos RA and Clegg MT (1982) Effects of population substructuring on estimates of outcrossing rate in plant populations. *Heredity*, 48: 283-292.

Epperson BK (1992) Spatial structure of genetic variation within populations of forest trees. *New Forests*, 6: 257-278.

Excoffier L, Smouse PE and Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*, 131: 479-491.

Ezeamuzie IC, Ambakederemo AW, Shode FO and Ekwebelem SC (1996) Antiinflammatory effects of *Moringa-oleifera* root extract. *International Journal of Pharmacognosy*, 34: 207-212.

Falconer DS (1960) Introduction to Quantitative Genetics, First Edition. Oliver and Boyd, Edinburgh.

Falconer DS (1989) Introduction to Quantitative Genetics, Third Edition. Longman Scientific and Technical, England.

Falconer DS and Mackay TFC (1996) Introduction to Quantitative Genetics, Fourth Edition. Addison Wesley Longman Limited, Harlow, England.

Farris MA and Mitton JB (1984) Population density, outcrossing rate, and heterozygote superiority in ponderosa pine. *Evolution*, 38: 1151-1154.

Fisher RA (1918) The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society Edinburgh*, 52: 399-433.

Fripp YJ, Griffin AR and Morran GF (1987) Variation in allele frequencies in the outcross pollen of *Eucalyptus regnans* F Muell throughout a flowering season. *Heredity, 59*: 161-171.

Fritsch P and Rieseberg LH (1992) High outcrossing rates maintain male and hermaphrodite individuals in populations of the flowering plant *Datisca glomerata*. *Nature*, 359: 633-636.

Furnier GR and Adams WT (1986) Mating system in natural populations of Jeffrey pine. *American Journal of Botany*, 73: 1002-1008.

Gaiotto FA, Bramucci M and Grattapaglia D (1997) Estimation of outcrossing rate in a breeding population of *Eucalyptus urophylla* with dominant RAPD and AFLP markers. *Theoretical and Applied Genetics*, 95: 842-849.

Galton F (1989) Natural Inheritance. Macmillan, London.

Gassenschmidt U, Jany K.D, Tauscher B and Niebergall H (1995) Isolation and characterization of a flocculating protein from *Moringa oleifera* Lam. *Biochemica et Biophysica ACTA-GENERAL SUBJECTS*, 1243: 477-481.

Genstat Statistical Package (1995) Version 5.32. Lawes Agricultural Trust (Rothamsted Experimental Station).

Georges M, Lequare AS, Castelli M, Handset R and Vassart G (1988) DNA fingerprinting in domestic animals using four different minisatellite probes. *Cytogenetics Cell Genetics*, 47: 127-131.

Gillies ACM, Cornelius JP, Newton AC, Navarro C, Hernandez M and J Wilson (1997) Genetic variation in Costa Rican populations of the tropical timber species C*edrela ordorata* L. assessed using RAPDs. *Molecular Ecology*, 6: 1133-1145.

Gjuric R and Smith SR, Jr (1996) Identification of cross-pollinated and self-pollinated progeny in Alfalfa through RAPD nulliplex loci analysis. *Crop Science*, 36: 389-393.

Goodell K, Elam DR, Nason JD and Ellstrand NC (1997) Gene flow among small populations of a self-incompatible plant:an interaction between demography and genetics. *American Journal of Botany*, 84:1362-1371.

Griffin AR and Cotteril PP (1988) Genetic variation in growth of outcrossed, selfed and open-pollinated progenies of *Eucalyptus regnans* and some implications for breeding strategy. *Silvae Genetica* 37: 124-131

Hamrick JL (1982) Plant population genetics and evolution. *American Journal of Botany*, 10: 1685-1693.

Hamrick JL (1989) Isozymes and the analysis of genetic structure in plant populations. In: Isozymes in Plant Biology, (eds. DE Soltis and PS Soltis, 87-105. Dioscorides Press, Portland, OR.

Hamrick JL and Godt MJ (1989) Allozyme diversity in plant species. In: Plant population genetics, breeding and genetic resources. (eds. AHD Brown, M Clegg, A Kahler, B Weer) 43-63. Sunderland: Sinauer Ass. Inc.

Hamrick JL, Godt MJW and Sherman-Broyles SL (1992) Factors influencing levels of genetic diversity in woody plants. *New forests*, 6: 95-124.

Harrier LA, Whitty PW, Sutherland JM and Sprent JI (1997) Phenetic investigation of non-nodulating African species of *Acacia* (*Leguminosae*) using morphological and molecular markers. *Plant Systematics and Evolution*, 205: 27-51.

Hartl L and Seefelder S (1998) Diversity of selected hop cultivars detected by fluorescent AFLPs. *Theoretical and Applied Genetics*, 96: 112-116.

Heun M and Halentjaris T (1993) Inheritance of RAPDs in F_1 hybrids of corn. *Theoretical and Applied Genetics,* 85: 961-968.

Heun M, Murphy JP and Phillips TD (1994) A comparison of RAPD and isozyme analysis for determining the genetic relationships among *Avena sterilis* L. accessions. *Theoretical and Applied Genetics*, 8: 689-696.

Holsinger KE and Mason-Gamer RJ (1996) Hierarchical analysis of nucleotide diversity in geographically structured populations. *Genetics*, 142: 629-639.

Holtsford TP and Ellstrand NC (1990) Inbreeding effects in *Clarkia tembloriensis* (Onagraceae) populations with different natural outcrossing rates. *Evolution*, 44: 2031-2046.

Hongtrakul V, Huestis GM and Knapp SJ (1997) Amplified fragment length polymorphisms as a tool for DNA fingerprinting sunflower germplasm: genetic diversity among oilseed inbred lines. *Theoretical and Applied Genetics*, 95: 400-407.

Hormaza JI, Dollo L and Polito VS (1994) Determination of relatedness and geographical movements of *Pistacia vera* L. (pistachio; Anacardiaceae) germplasm by RAPD analysis. *Economic Botany*, 48: 349-358.

House APN and Bell JC (1994) Isozyme variation and mating system in *Eucalyptus urophylla* S.T. Blake. *Silvae Genetica*, 43: 167-176.

Hu J and Quiros CF (1991) Identification of broccoli and cauliflower cultivars with RAPD markers. *Plant Cell* reports, 10: 505-511.

Ibrahim, S.S, Ismail M, Samuel G, Kamel E and El-Azahari T (1974). Ben seeds: A potential oil source. *Agricultural Research Review* (Egypt), 52, 47-50.

Jain SK (1990) Variation and selection in plant populations. In: Population Biology, (eds. K Wohrman and SK Jain), 199-230. Springer Verlag, Berlin.

Jahn SAA (1984) Effectiveness of traditional flocculants as primary coagulants and coagulant aids for the treatment of tropical raw water with more than a thousand fold fluctuation in turbidity. *Water Supply*, 2: 8 -10.

Jahn SAA (1988) Using *M. oleifera* seeds as coagulants in developing countries. *Journal AWWA*, pp 43-50.

Jahn SAA (1989) *Moringa oleifera* for food and water purification - selection of clones and growing of annual short term. *Entwicklung and Landlicher Raun*, 23: 22-25.

Jahn SAA, Mussad HA and Burgstaller H (1986) The tree that purifies water. Cultivating Multipurpose Moringaceae in the Sudan. *Unasylva*, 152: 23-28.

Jahn SAA (1991) The traditional domestication of a multipurpose tree *Moringa stenopetala* (Bak.f.) Cuf. in the Ethiopian Rift valley. *AMBIO*, 20: 244-47.

Jeffreys AJ, Wilson V and Thein SL (1985) Individual-specific fingerprints of human DNA. *Nature*, 316: 76-79.

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

Joshi, CP and Nguyen HT (1993) Application of the random amplified polymorphic DNA techmique for the detection of polymorphism among wild and cultivated tetraploid wheats. *Genome,* 36: 602-609.

Jyoth PV, Atluri JB and Reddi CS (1990) Pollination ecology of *Moringa oleifera* (Moringaceae). *Proceedings of the Indian Academy of Sciences*, (Plant science), 100: 33-42.

Kaga A, Tomooka N, Egawa Y, Hosaka K and Kamijima O (1996) Species relationships in the subgenus *Ceratotropis* (genus *Vigna*) as revealed by RAPD analysis. *Euphytica*, 88: 17-24.

Kantharajah AS and Dodd WA (1991) Rapid clonal propagation of *M. oleifera* Lam., using tissue culture. *South Indian Horticulture*, 39: 224-228.

Karp A and Edwards KJ (1997) Molecular techniques in the analysis of the extent and distribution of genetic diversity. In: Molecular Genetic Techniques for Plant Genetic Resources. (eds. WG Ayad, T Hodgkin, A Jaradat and VR Rao) Report of an IPGRI workshop, 9-11 October 1995, Rome, Italy. International Genetic Resources Institute, Rome, Italy.

Karp A, Serberg O, and Buiatti M (1996) Molecular techniques in the assessment of botanical diversity. *Annals of Botany*, 78: 143-149.

Karron JD (1991) Patterns of genetic variation and breeding systems in rare plant species. In: Genetics and Conservation of Rare Plants, (eds. DA Falk and KE Holsinger), 87-98. Oxford University Press, New York, NY.

Karron JD, Jackson RT, Thumser NN and Schlicht SL (1997) Outcrossing rates of individual *Mimulus ringens* genets are correlated with anther-stigma separation. *Heredity*, 79: 365-370.

Kearsey MJ and Farquhar GL (1998) QTL analysis in plants; where are we now ?. *Heredity*, 80: 137-142.

Kennington WJ, Waycott M and James SH (1996) DNA fingerprinting supports notions of clonality in a rare mallee, *Eucalyptus argutifolia*. *Molecular Ecology*, 5: 693-696.

Kercher S and Conner JK (1996) Patterns of genetic variability within and among populations of wild Radish, *Raphanus raphanistrum* (Brassicaceae). *American* Journal of Botany, 83: 1416-1421.

Khan FW, Gul P and Malik MN (1975) Chemical composition of oil from *Moringa oleifera*. *The Pakistan Journal of Forestry*, pp 100-102.

King LM, Schaal BA (1989) Ribosomal DNA variation and distribution in *Rubeckia missouriensis*. *Evolution*, 42:1117-1119.

Lande R and Schemske DW (1985) The evolution of self-fertilization and inbreeding depression in plants. I. Genetic models. *Evolution* ,39: 24-40.

Lemen C (1980) Allocation of reproductive effort to the male and female strategies in wind-pollinated plants. *Oecologia,* 45: 156-159.

Loarce Y, Callego R and Ferrer E (1996) A comparative analysis of the genetic relationships between rye cultivars using RFLP and RAPD markers. *Euphytica*, 88: 107-115.

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

Loveless MD and Hamrick JL (1984) Ecological determinants of genetic structure in plant populations. *Annual Review of Ecology and Systematics*, 15: 65-95.

Lu Z, Douthitt MP, Taffs RE, Ran Y, Norwood LP and Chumakov KM (1993). Quantitative aspects of the mutant analysis by PCR and restriction enzyme cleavage (MAPREC). *PCR Methods and Applications*, 3: 176-180.

Lui Z and Furnier GR (1993) Comparison of allozyme. RFLP and RAPD markers for revealing genetic variation within and trembling aspen and bigtooth aspen. *Theoretical and Applied Genetics*, 87: 97-105.

Maddox G, Cook R, Wimberger P and Gardescu S (1989) Clone structure in four solidago altissima (Asteraceae) populations: rhizome connections within genotypes. *American Journal of Botany,* 76: 318-326.

Maki M, Masuda M and Inoue K (1996) Genetic diversity and hierarchical population structure of a rare autotetraploid plant, *Aster kantoensis* (Asteraceae). *American Journal of Botany*, 83: 296-303.

Makkar HPS and Becker K (1997) Nutrients and antiquality factors in different morphological parts of the *Moringa oleifera* tree. *Journal of Agricultural Science*, 128: 311-322.

Makonnen E, Hunde A and Damecha G (1997) Hypoglycaemic effects of *Moringa stenopetala* aqueous in rabbits. *Phytotherapy Research*, 11: 147-148.

Maley ML and Parker WH (1993) Phenotypic variation in cone and needle characters of *Pinus banksiana* (jack pine) in North-Western Ontario. *Canadian Journal of Botany*, 71: 43-51.

Margale E, Herve Y, Hu J and Quiros CF (1995) Determination of genetic variability by RAPD markers in Cauliflower, Cabbage and Kale local cultivars from France. *Genetic Resources and Crop Evolution*. 42: 281-289.

Mason-Gamer RJ, Holsinger KE and Jansen RK (1995) Chloroplast DNA haplotype variation within and among populations of *Coreopsis grandiflora* (Asteraceae). *Molecular Biology and Evolution*, 12: 371-381.

Mather K (1949) Biometrical Genetics, First Edinburgh Methuen, London.

Mather K and and Jinks JL (1982) Biometrical Genetics. Chapman and Hall, London, UK.

Maughan PJ, Maroof MAS, Buss GR and Huestis GM (1996) Amplified fragment length polymorphism (AFLP) in soybean: species diversity, inheritance, and near-isogenic line analysis. *Theoretical and Applied Genetics*, 93: 392-401.

Mayer FA and Stelz E (1993) *Moringa stenopetala* provides food and low cost water purification. *Agroforestry Today*, 5: 16-18.

McCauley DE (1997) The relative contribution of seed and pollen movement to the local genetic structure of *Silene alba*. *Journal of Heredity*, 88: 257-263.

Micheli MR, Bova R, Pascale E and D'Ambrosio E (1994) Reproducible DNA finger printing with the random amplified polymorphic DNA (RAPD) method. *Nucleic Acids Research*, 22: 1921-1922.

Milbourne D, Meyer R, Bradshaw JE, Baird E, Bonar N, Provan J, Powell W and Waugh R (1997) Comparison of PCR based marker systems for the analysis of genetic relationships in cultivated potato. *Molecular Breeding*, 3: 127-136.

Milligan BG (1996) Estimating long-term mating systems using DNA sequences. *Genetics*, 142: 619-627.

Mitchel RJ and Marshall D (1998) Nonrandom mating and sexual selection in a desert mustard: an experimental approach. *American Journal of Botany*, 85: 48-55.

Mohan MS, Nair S, Bhagwat A, Krishna TA, Yano M, Bhatia CR and Sasaki T (1997) Genome mapping, molecular markers and marker-assisted selection in crop plants. *Molecular Breeding*, 3: 87-103.

Moran GF and Brown HD (1980) Temporal heterogeneity of outcrossing rates in alpine ash (*Eucalyptus delagatensis* R.T. Bak.). *Theoretical and Applied Genetics*, 57: 101-105.

Morgan WTW (1981) Ethnobotany of the Turkana: use of plants by a pastoral people and their livestock in Kenya. *Economic Botany*, 35: 96-130.

Morgante M, Vendramin GG and Rossi P (1991) Effects of stand density on outcrossing rate in two Norway spruce (*Picea abies*) populations. *Canadian Journal of Botany*, 69: 2704-2708.

Morton JF (1991) The horseradish tree, *Moringa pterygosperma* (moringaceae)- a boon to Arid lands ?!. *Economic Botany*, 45: 318-333.

Mousseau TA and Roff DA (1987) Natural selection and the heritability of fitness components. *Heredity* 59: 181-197.

Murawski DA and Hamrick JL (1991) The effect of the density of flowering individuals on the mating systems of nine tropical tree species. *Heredity* 67, 167-174.

Murawski DA and Hamrick JL (1992) Mating system and phenology of *Ceiba pentandra* (Bombacaceae) in central Panama. *Journal of Heredity*, 83: 401-404.

Murray MG and Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research*, 8: 4321-4325.

Muyibi SA and Evanson LM (1995) Optimising physical parameters affecting coagulation of turbid waters with *Moringa oleifera* seeds. *Water Research*, 29: 2689-2695.

Nautiyal BP and Venkataraman KG (1987) *Moringa* (drumstick) - An ideal tree for social forestry: Growing conditions and uses - Part 1. *MYFOREST*, 23: 53 -58.

Ndabigengesere A and Narasiah KS (1996) Influence of operating parameters on turbidity removal by coagulation with *Moringa oleifera* seeds. *Environmental Technology*, 17: 1103-1112.

Ndabigengesere A and Narasiah KS (1998) Quality of water treated by coagulation using *M. oleifera* seeds. *Water Research*, 32: 781-791.

Ndabigengesere A, Narasiah KS and Talbot BG (1995) Active agents and mechanism of coagulation of turbid waters using *Moringa oleifera*. *Water Research*, 29: 703-710.

Neale DB and Harry DE (1994) Genetic mapping in forest trees: RFLPs, RAPDs and beyond. *AgBiotech News and Information*, 6: 107N-114N.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York, Ny, USA

Nei M and Li W-H (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences, U.S.A,* 76: 5269-5273.

Nesbitt KA, Potts BM, Vaillancourt RE, West AK and Reid JB (1995) Partitioning and distribution of RAPD variation in a forest tree species, *Eucalyptus globulus* (Myrtaceae). *Heredity*, 74: 628-637.

Norman JK, Weller SG and Sakai AK (1997) Pollination biology and outcrossing rates in hermaphroditic *Schiedea Lydgatei* (Caryophyllaceae). American Journal of Botany, 84:641-648.

Obara-Okeyo and Kako S (1998) Genetic diversity and identification of *Cymbium* cultivars as measured by random amplified DNA (RAPD) markers. *Euphytica*, 99: 95-101.

Ohara M, Takeda H, Ohno Y and Shimamoto Y (1996) Variations in the breeding system and the population genetic structure of *Trillium kamtschaticum* (Liliaceae). *Heredity*, 76: 476-481.

O'Malley DM and Bawa KS (1987) Mating system of a tropical rain forest tree species. *American Journal of Botany,* 74: 1143-1149.

Orozco-Castillo C, Chalmers KJ, Waugh R and Powell W (1994) Detection of genetic diversity and selective gene introgression in coffee using RAPD markers. *Theoretical and Applied Genetics*, 87: 934-940.

Pakniyat H, Powell W, Baird E, Handley LL, Robinson D, Scrimgeour CD, Nevo E, Hackett CE, Caligari PDS and Forster BP (1997) AFLP variation in wild barley (*Hordeum spontaneum* C. Koch) with reference to salt torelance and associated ecogeography. *Genome,* 40: 332-341.

Pal SK, Mukherjee PK and Saha BP (1995) Studies on the antiulcer activity of *Moringa oleifera* leaf extract on gastric ulcer models in rats. *Phytotherapy Research*, 9: 463-465.

Palacios C and Gonzalez-Candelas F (1997) Analysis of population genetic structure and variability using RAPD markers in the endemic and endangered *Limonium dufouri* (Plumbaginaceae). *Molecular Ecology*, 6: 1107-1121.

Paran I, Aftergoot E and Shifriss C (1998) Variation in *Capsicum annuum* revealed by RAPD and AFLP markers. *Euphytica*, 99: 167-173.

Parks JC and Werth CR (1993) A study of spatial features of clones in a population of Bracken fern, *Pteridium aquilinum* (Dennstaedtiaceae). *American Journal of Botany*, 80: 537-544.

Paul S, Wachira FN, Powell W and Waugh R (1997) Diversity and genetic differentiation among populations of Indian and Kenyan tea (*Camellia sinensis* (L.) O. kuntze) revealed by AFLP markers. *Theoretical and Applied Genetics*, 94: 255-263.

Peakall R, Smouse PE and Huff DR (1995) Evolutionary implications of allozyme and RAPD variation in diploid populations of dioecious buffalo grass *Buchloe dactyloides*. *Molecular Ecology*, 4: 135-147.

Perera L, Russell JR, Provan J, McNicol JW and Powell W (1998). Evaluating genetic relationships between indigenous coconut (*Cocos nucifera* L.) accessions from Sri Lanka by means of AFLP profiling. *Theoretical and Applied Genetics*, 96: 545-550.

Peters BG, Lonie JS and Moran GF (1990) The breeding system, genetic diversity and pollen sterility in *Eucalyptus pulcerulenta*, a rare species with disjunct populations. *Australian Journal of Botany*, 38: 59-70.

Philips EP (1951) The genera of South African flowering plants. Second edition, p358. Government printer, Pretoria,

Pigliucci M (1996) Modelling phenotypic plasticity.II. Do genetic correlations matter?. *Heredity*, 77: 453-460.

Powell W (1992) Plant genomes, gene markers, and linkage maps. In: Biotechnology and Crop Improvement in Asia (ed. JP Moss), 297-322. International Crops Research Institute for the Semi-Arid Tropics, Patancheru, A.P, 502324, India

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

Powell W, Thomas WTB, Baird E, Lawrence P, Booth A, Harrower B, McNicol JW and Waugh R (1997) Analysis of quantitative traits in barley by the use of amplified fragment length polymorphisms. *Heredity*, 79: 48-59.

Premoli A (1996) Allozyme polymorphisms, outcrossing rates, and hybridization of South American Nothofagus. *Genetica*, 97: 55-64.

Puri V (1941) The life history of *Moringa* Lam. *Journal of Indian Botanical Society*, 20: 263-283.

Rafalski JA, Tingey SV and Williams GJK (1991) RAPD markers - a new technology for genetic mapping and plant breeding. *Biotechnology News and Information*, 3: 645-648.

Ramachandran C, Peter KV and Gopalakrishnan PK (1980) Drumstick (*Moringa oleifera*): a multipurpose Indian vegetable. *Economic Botany*, 34: 276-283.

Ramser J, Lopez-Peralta C, Wetzel R, Weising R, and Kahl (1996) Genomic variation and relationships in aerial yam (*Dioscorea bulbifera* L.) detected by random amplified polymorphic DNA. *Genome*, 39: 17-25.

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

Rieseberg LH, Beckstrom-Sternberg S, Doan K (1990) *Helianthus annuus* ssp. *texanus* has chloroplast DNA and nuclear ribosomal RNA genes of *Helianthus debilis* ssp *cucumerifolius*. *Proceedings of the National Academy of sciences of the USA*, 87: 593-597.

Ritland K (1983) Estimation of mating systems. In: Isozymes in Plant Genetics and Breeding, Part A. (eds. SD Tanksley, TJ Orton), 289-302. Elsevier Science Publishers, Amsterdam.

Ritland K (1990) A series of FORTRAN computer programs for estimating plant mating systems. *Heredity*, 47: 37-54.

Ritland K and S Jain (1981) A model for the estimation of outcrossing rate and gene frequencies using *n* indepent loci. *Heredity*, 47: 35-52.

Russell JR, Husein F, Johnson E, Waugh R and Powell W (1993) Genetic differentiation of Cocoa (*Theobroma cacao* L) populations revealed by RAPD analysis. *Molecular Ecology*, 2: 89-97.

Russell JR, Fuller JD, Macaullay M, Hatz BG, Jahoor A, Powell W and Waugh R (1997) Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. *Theoretical and Applied Genetics*, 95: 714-722.

Salimath SS, Oliveira AC, Godwin ID and Bennetzen JL (1995). Assessment of genome origins and genetic diversity in the genus *Eleusine* with DNA markers. *Genome*, 38: 757-763.

Samec P and Nasinec V (1995) Detection of DNA polymorphism among Pea cultivars using RAPD technique. *Biologia Plantarum*, 37: 321-327.

Schaal BA, Hayworth DA, Olsen KM, Rauscher T and Smith WA (1998) Phylogenetic studies in plants: problems and prospects. *Molecular Ecology*, 7: 465-474.

Schneider S, Kueffer J-M, Roessli D, Excoffier L (1997) Arlequin Ver. 1.1: A software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Schierenbeck KA, Skopski M, Lieberman D and Lieberman M (1997) Population structure and genetic diversity in four tropical tree species in Costa Rica. *Molecular Ecology*, 6: 137-144.

Schmalhausen II (1949) Factors of Evolution. Blakiston, Philadelphia.

Schoen DJ, Johnston MO, L'Heureux A-M and Marsolais JV (1997) Evolutionary history of the mating system in *Asinckia* (Boraginaceae). *Evolution,* 51: 1090-1099.

Sharma SK, Dawson IK and Waugh R (1995) Relationships among cultivated and wild lentils revealed by RAPD analysis. *Theoretical and Applied Genetics*, 91: 647-654.

Sharma SK, Knox MR and Ellis THN (1996) AFLP analysis of the diversity and phylogeny of *Lens* and its comparison with RAPD analysis. *Theoretical and Applied Genetics*, 93: 91-96.

Shaw DV and Allard RW (1982) Estimation of outcrossing rates in Douglas-fir using isozyme markers. *Theoretical and Applied Genetics*, 62: 113-120.

Sigurdsson V, Anamthawat-Jonsson K and Siurgeirsson A (1995) DNA fingerprinting of *Populus trichocarpa* clones using RAPD markers. *New Forests*, 10: 197-206.

Slatkin M (1987) Gene flow and the geographic structure of populations. *Science,* 236: 787-792.

Smith JF and TV Pham (1996) Genetic diversity of the narrow endemic *Allium aaseae* (Alliaceae). *American Journal of Botany*, 83: 717-726.

Smouse PE and Chevillon C (1998) Analytical aspects of population-specific DNA fingerprinting for individuals. *Journal of Heredity*, 89: 143-150.

Staub JE, Serquen FC and Gupta M (1994) Genetic maps, map construction, and their application in plant breeding. *Horticultural Science*, 31: 729-740.

Storfer A (1996) Quantitative genetics: a promising approach for the assessment of genetic variation in endangered species. *Trends in Ecology and Evolution*, 11: 343-348.

Subadra S, Monica J and Dhabhai D (1997) Retention and storage stability of beta-carotene in dehydrated drumstick leaves (*Moringa oleifera*). *International Journal of Food Sciences and Nutrition*, 48: 373-379.

Sydes MA and Peakall R (1998) Extensive clonality in the endangered shrub *Haloragodendron lucasii* (Haloragaceae) revealed by allozymes and RAPDs. *Molecular Ecology*, 7: 87-93.

Szmidt AE, Wang X and Lu M (1996) Empirical assessment of allozyme and RAPD variation in *Pinus sylvestris* (L.) using haploid tissue analysis. *Heredity,* 76: 412-420.

Tanksley SD (1993) Mapping polygenes. *Annual Review of Genetics*, 27: 205-233.

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

Teketay D (1995) The effect of temperature on the germination of *Moringa stenopetala*, multipurpose tree. *Tropical Ecology*, 36: 49- 57.

Templeton AR (1986) Coadaptation and outbreeding depression, in Conservation Biology: The Science of Scarcity and Diversity (ed. ME Soule), 105-116, Sinauer.

Thormann CE, Ferreira ME, Camargo LEA, Tivang JG and Osborn TC (1994) Comparison of RFLP and RAPD markers to estimate genetic relationships within and among crucifererous species. *Theoretical and Applied Genetics*, 88: 973-980.

Tingey SV and del Tufo JP (1993) Genetic analysis with random amplified polymorphic DNA markers. *Plant Physiology*, 101: 349-352.

Travis SE, Maschinski J and Keim P (1996) An analysis of genetic variation in *Astragalus cremnophlax* var. *cremnophylax*, a critically endangered plant, using AFLP markers. *Molecular Ecology*, 5: 735-745.

Van Eck HJ, van der Voort JR, Draaistra J, van Zandvoort P, van Enckevort E, Segers B, Pelman J, Jacobsen E, Helder J and Bakker J (1995) The inheritance and chromosomal localization of AFLP markers in a non-inbred potato offspring. *Molecular Breeding*, 1: 397-410.

Van de Ven WTG and McNicol RJ (1995) The use of RAPD markers for the identification of sitka spruce (*Picea sitchensis*) clones. *Heredity*, 75: 126-132.

Van Ham RCHJ and Hart H (1998) Phylogenetic relationships in the crassulaceae inferred from chloroplast DNA restriction-site variation. *American Journal of Botany*, 85: 123-134.

Verdcourt B (1985) A synopsis of the Moringaceae. *Kew Bulletin*, 40:1.

Virk PS, Ford-Lloyd BV, Jackson MT, Pooni, HS, Clemeno TP and Newbury JH (1996) Predicting quantitative variation within rice germplasm using molecular markers. *Heredity*, 76: 296-304.

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

Wachira FN, Waugh R, Hackett CA and Powell W (1995) Detection of genetic diversity in tea (*Camellia sinensis*) using RAPD markers. *Genome*, 38: 201-210.

Wachira FN, Powell W and Waugh R (1997) An assessment of genetic diversity among *Camellia sinensis* L. (cultivated tea) and its wild relatives based on randomly amplified polymorphic DNA and organelle-specific STS. *Heredity*, 78: 603-611.

Waugh R and Powell W (1992) Using RAPD markers for crop improvement. *Tibtech*, 101: 186-191.

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

Wilde J, Waugh R and Powell W (1992) Genetic fingerprinting of *Theobroma* clones using randomly amplified polymorphic DNA markers. *Theoretical and Applied Genetics*, 83: 871-877.

Williams C and St-Clair DA (1993) Phenetic relationships and levels of variability detected by restriction fragment length polymorphism and random amplified polymorphic DNA analysis of *Lycopersicon esculentum*. *Genome*, 36: 619-630.

Williams JGK, Hanafey MK, Rafalski JA and Tingey SV (1993) Genetic analysis using random amplified polymorphic DNA markers. *Methods in Enzymology*, 218: 704-740.

Williams JGK, Kubelik AR, Livak K, Rafalski JA and Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*, 18: 6531-6535.

Winfield MO, Arnold GM, Cooper F, Leray M, White J, Karp A and Edwards KJ (1998) A study of genetic diveristy in *Populus nigra* ssp. *betulifolia* in the upper Severn area of the UK, using AFLP markers. *Molecular Ecology*, 7: 3-10.

Wolfe LM and Shore JS (1992) The mating system of *Hydrophyllum appendiculatum*, a protandrous species. *Sex Plant Reproduction*, 5: 239-245.

Wright S (1934) An analysis of variability in number of digits in an inbred strain of guinea pigs. *Genetics*, 19: 506-536.

Yeh FC, Chong DKX and Yang R.-C (1995) RAPD variation within and among natural populations of Trembling Aspen (*Populus tremuloides* Michx.) from Alberta. ***Journal of Heredity***, 86: 454-460.

Zabeau M and Vos P (1993) Selective restriction fragment amplification: a general method for DNA fingerprints. European Patent Application. Publication 0534858A1.