DEVELOPMENT OF MICROSATELLITE MARKERS AND ANALYSIS OF GENETIC DIVERSITY AND POPULATION STRUCTURE OF SANDALWOOD (*Osyris lanceolata* Hochst. & Steud.) IN KENYA

BY

OTIENO JOHN OCHIENG

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF DEGREE OF MASTER OF SCIENCE IN PLANT BREEDING AND BIOTECHNOLOGY IN THE SCHOOL OF AGRICULTURE AND BIOTECHNOLOGY

UNIVERSITY OF ELDORET, KENYA

APRIL, 2024

DECLARATION

Declaration by the Student

This research thesis is my original work and has not been submitted for any academic award in any institution and shall not be reproduced in part or full, or in any format without prior written permission from the author and/ or University of Eldoret.

Otieno John Ochieng

_____ _____

AGR/PBG/001/15 **Date**

Declaration by the Supervisors

This research thesis has been submitted with our approval as the university supervisors.

_____ _____

Dr. Oliver Kiplagat, **Date**

Department of Biotechnology, University of Eldoret,

P.O Box 1125-3011 Eldoret, Kenya.

_____ _____

Dr. Kennedy Pkania, **Date**

Department of Biotechnology, University of Eldoret,

P.O Box 1125-3011 Eldoret, Kenya.

_____ _____

Dr. David Odee, **Date**

Chief Research Scientist/Visiting Scientist

UK Centre for Ecology and Hydrology

DEDICATION

To Joseph Nyaguti Otieno, my late grandfather, who actively supported me in my determination to find and realize my potential, and to make this contribution to our world.

ABSTRACT

African sandalwood (*Osyris lanceolata* Hoscht and Steud) is a multipurpose and drought tolerant, African tropical hemi-parasitic tree species belonging to *Santalaceae* family. It is an evergreen dioecious shrub or small tree growing to a height of 1-7 meters depending on soil-type, climate and genetics of the tree. The species is valued globally for its essential oils used in perfumery and pharmaceutical industries. The increased demand for its essential oils and other products is bringing pressure to bear on the dwindling *O. lanceolata* populations and habitats in Kenya and East African region. due to overexploitation through anthropogenic activities that include illegal trade, overgrazing bush burning and destruction of host plant species for fuel wood, timber, charcoal burning and building materials. Consequently, the Convention on International Trades in Endangered Species (CITES) recently issued notification to review and gather information on the conservation status of *O. lanceolata* among other concerns. Although protected under CITES, the species continued to be heavily smuggled and overexploited. However, knowledge regarding the genetic diversity and population structure of the extant Kenyan populations, which is vital in informing conservation and sustainable management strategies of the species is still limited. Therefore, the aim of this work was to develop microsatellite (SSR) markers and use them to evaluate the genetic diversity and population structure of the species across the geographical distribution range in Kenya. A set of 12 polymorphic and five monomorphic microsatellite markers were developed and characterised using standard genome assembly, SSR identification and primer design protocols. Ten highly polymorphic microsatellite loci were used to characterise 288 individuals over ten natural populations, namely Baringo, Embu, Gachuthi, Gwasi, Kibwezi, Kitui, Makueni, Meru, Mau and Mt Elgon. The loci produced 178 alleles with a high Shannon's Information index (I) values ranged from 0.805 to 1.6. The average observed heterozygosity across all loci varied from 0.112 to 0.815. A high level of genetic diversity was inferred from the genetic diversity parameters (*He* = 0.587, *I* = 1.302 and *PPL* = 97 %). The unweighted pair group method of arithmetic averages (UPGMA) and population structure analysis grouped these 288 individuals into two major groups. The AMOVA results indicated that 62% of the total genetic variation was found within populations, while only 38% was observed among populations. Evaluating genetic diversity is vital for identifying populations for conservation priority and establishing baseline data for informed conservation strategies at the local level. This study represents the first examination of the genetic diversity and population structure of O. lanceolata using SSR markers. The newly developed microsatellite markers will be valuable for future breeding programs and genetic studies aimed at formulating effective conservation plans.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF PLATES

LIST OF ABBREVIATIONS, ACRONYMS AND SYMBOLS

ABI:	Applied Biosystems Inc.
AFLP:	Amplified Fragment Length Polymorphism
AFLP:	Amplified fragment length polymorphism
AMOVA:	Analysis of Molecular Variance
ASAL:	Arid and Semi-Arid Land
BLAST:	Basic Local Alignment Search Tool
CTAB:	Cetyl Trimethyl Ammonium Bromide
CTAB:	Cetyltrimethyl ammonium bromide
EDTA:	Ethylene diamine tetra-acetic acid
EST:	Express Sequence Tags
GPS:	Global Positioning System
HWE:	Hardy Weinberg Equilibrium
IAM:	Infinite Alleles Model
ICRAF:	International Centre for Research in Agroforestry
KEFRI:	Kenya Forestry Research Institute
KWS:	Kenya Wildlife Service
LD	Linkage Disequilibrium
LSD:	Least Significant Difference
PCoA:	Principal Coordinate Analysis
PIC:	Polymorphic Information Content
RAPD:	Random Amplified Polymorphic
RFLP:	Restriction Fragment Length Polymorphism Seq. Sequence
SD:	Standard Deviation
TPM:	Two-Phase Mutation
UPGMA:	Unweighted Pair Group Method with Arithmetic mean

ACKNOWLEDGEMENT

I sincerely thank Kenya Forestry Research Institute (KEFRI) and UK Centre for Ecology and Hydrology (UKCEH), under whose collaboration my research work was funded through, International Timber Tracking Organization (ITTO) project contract No.: 2005-34135-16007 and EU TRUE HORIZON 2020 project contract No. 727973.

My sincere gratitude and appreciation goes to my supervisors; Dr. Oliver Kiplagat, Dr. Kennedy Pkania of the University of Eldoret and Dr. David Odee formerly of the UK Centre for Ecology and Hydrology (UKCEH), for their exemplary guidance, support, patience, and encouragement throughout my research studies.

I also extend my gratitude to Dr. Stephen Cavers and Dr. Perry Anika of the UKCEH for the short training opportunity in the UK through which we gain the capacity to develop the SSR markers used in this study.

I also extend my gratitude to Dr. Stephen Omondi (KEFRI) for the guidance during my laboratory work and data analysis. Dr. Stephen Indieka of Egerton University for his encouragement and support during proposal development and final thesis writing. My colleagues Emmanuel Makatiani, John Gicheru, Charles Oduor and Milton Esitubi for technical assistance during collection and processing of field samples and all the staff of the Biotechnology laboratory (KEFRI) for their moral support during my research.

Finally, my heart-felt thanks also go to my wives, Gladys Atieno (the late) and Ann Kabeke, siblings Jullie Akinyi, William Kimaya, Truphena Achieng, Victoria Munyiva and Nereah Auma for their love and moral support.

CHAPTER ONE

INTRODUCTION

1.1 Background of the study

African sandalwood (*Osyris lanceolata* Hochst. & Steud.)(*Santalaceae*) is a multipurpose, drought-tolerant and hemi- parasitic tree, well known for its essential oils used in perfumery industries (Teixeira da Silva *et al*., 2019). It emerged as a potential commercial species in Africa due to significant decline in original sources of sandalwood oil, e.g., *Santalum album* L. (Indian subcontinent) and *Santalum spicatum* (R.Br.) A. DC. (Australia) in the 1990s, and the increasing demand for sandalwood oil over the years (Mugula *et al*., 2021). The genetic diversity of a species is important to its current persistence and long-time evolutionary potential (Bernatchez, 2016). Endangered species generally have small or declining populations, and often these populations suffer from inbreeding and erosion of genetic diversity resulting in elevated extinction risks (Weeks, Stoklosa & Hoffmann, 2016).

A crucial element of biodiversity is genetic variety. The precise identification of their accessions is necessary for the preservation and sustainable use of plant genetic resources. Morphological traits have been the foundation of traditional approaches for determining the genetic diversity or relatedness among groups of plants. Environmental circumstances, however, have the potential to affect these traits. To characterise variety for *in situ* and *ex situ* conservation, molecular markers avoid the difficulties of environmental impacts acting upon characteristics (Idrees & Irshad, 2014).

Molecular markers are frequently employed to detect genetic differences and similarities across species, varieties, or accessions to inform the screening and selection of germplasm for agricultural improvement programs and conservation (Govindaraj, Vetriventhan & Srinivasan, 2015). Additionally, molecular markers offer quicker and more accurate methods for identifying known and unidentified plant species as well as for addressing taxonomic and evolutionary queries (Jose & Harikrishnan, 2019). High levels of polymorphism, codominant inheritance, frequent occurrence in the genome, and reproducibility are crucial characteristics of excellent molecular markers (Idrees & Irshad, 2014). There are several molecular markers, and new ones are constantly being created. For diversity analysis and fingerprinting, molecular markers like Amplified fragment length polymorphism (AFLP) and Simple sequence repeat (SSR) are useful (Rauscher & Simko, 2013).

Particularly for the investigation of population genetic variation, molecular markers like SSRs have been shown to be useful because they are highly informative and reproducible. They have also been employed to develop endangered species conservation plans (Wu *et al*., 2020; Wang, 2020). Their distribution is highly dense, uneven, and ostensibly non-random across the genomes of numerous completely sequenced animals (Iftikhar & Carney, 2016). In certain circumstances, SSR markers created for one species can be used on closely related species. According to Miah *et al*. (2013), they frequently exhibit significant levels of intra- and inter-specific polymorphism.

Sequences that are short tandemly repeated are widely used in biological research because of their polymorphisms and high mutation rates. In comparison to the noncoding

proportion in plant species, all SSR forms, with the exception of trinucleotides and hexanucleotides, are much less common (Vieira, Santini, Diniz, & Munhoz, 2016). It is simple to tell the difference between heterozygotes for various segments. Because of this, individual loci that correspond to certain primer pairs co-dominate and may have several alleles. According to Wu *et al*. (2020), microsatellites are frequently found in regulatory regions, such as untranslated regions and introns. Their placement in the genome affects how usefully they may be used as molecular markers. Microsatellite markers from noncoding areas, for instance, have enough polymorphism to distinguish between species that are closely related (Mathi *et al*., 2016).

For the protection and use of genetic resources, it is crucial to identify the genetic structure of germplasm collections. Relationships between individuals, genetic similarity and difference within and between genotype groupings, as well as their evolutionary history, are all inferred from population structure (Bellucci *et al*., 2021). Model-based clustering approaches, distance-based phylogenetic methods (Yusefi *et al*., 2019), and multivariate consensus representation of genetic relationship among populations (Esfandani Bozchaloyi & Sheidai, 2018), are methods for reconstructing population structure.

For determining genetic similarity or dissimilarity, a variety of coefficients are utilized, including Dice, Jaccards, and Squared Euclidean distance (Esfandani Bozchaloyi & Sheidai, 2018), Additionally, it is possible to partition variance using molecular variance analysis in order to infer population structure (Liang *et al*., 2015). The marker system that is employed and the experiment's goal influence the selection of suitable coefficients and

procedures. Cluster analysis and other descriptive techniques are effective for highlighting intriguing populations. However, it is necessary to assess the reliability using the proper statistical techniques, such as bootstrapping. The populations have been decreasing since 2002 because of the heavy exploitation for international trade, including illegal harvesting and trade (CITES, 2013). The sharp rise in the extraction of *O. lanceolata* in Kenya and Tanzania is believed to be linked to a decline of the resource in the global primary sources of sandalwood in Australia and India (CITES, 2013).

Despite restrictive controls on illegal trade on sandalwood, the tree continues to be illegally harvested at an alarming rate annually. The tree has become the most illegally traded flora in Kenya and the East African region as a whole for more than two decades (Bunei, 2017). Notwithstanding a presidential protection under the laws of Kenya through Gazette Notice No. 3176 on April 04, 2007 (Republic of Kenya, 2007), recent anecdotal evidence suggests that there is an increasing trend in overexploitation and illegal trade of the species (Bunei, 2017). Dwindling of *O. lanceolata* populations in Africa is attributed to overexploitation and lack of robust management strategies (CITES 2013). *O. lanceolata* is assigned an automated status of least concern (LC) (Bunei, 2017) with an unknown population trend but acknowledging decline in East Africa where it's considered endangered due to over exploitation.

Studies of natural genetic variation in *O. lanceolata* are scarce, and only a few genomic resources are available despite its endangered status and economic importance. Initially, isozymes served as reliable markers for genetic analysis in the Indian sandalwood species, *Santalum album* (Patel *et al*., 2016) but this biochemical marker was relatively low in

abundance and therefore low polymorphism. Understanding the genetic diversity and structure of plant species is a prerequisite for formulating effective breeding and conservation strategies (Vanavermaete *et al*., 2021). In recent years, with advances in molecular biological technologies, molecular marker-based analysis of genetic variation has avoided the limitations of morphological and biochemical indicators that are readily affected by environmental factors, and are more accurate for evaluation of population genetic parameters (Nadeem *et al*., 2018).

A number of studies have been conducted to understand the genetic diversity of sandalwood using random amplification of polymorphic DNA (RAPD) (Patel *et al*., 2016) and RFLP (Byrne *et al*., 2003). However, owing to the lack of highly informative genetic information, to date no SSR markers are available for evaluation of the genetic diversity within *O. lanceolata*, which limits its breeding and commercial development. Among the various types of molecular markers, microsatellites (or simple sequence repeats; SSRs) have become the preferred choice because of their high degree of reproducibility, ability to identify high levels of genetic polymorphism, codominant inheritance, and abundance in plant genomes (Grover *et al*., 2016). Currently, SSR markers have been widely used in assessments of population genetic diversity (Gadissa *et al*., 2018), cultivar identification (Sulu *et al*., 2020), and association mapping (Wang, 2020) as effective tools to describe genetic variation in plants. Microsatellites have increasingly become the markers of choice for endangered and threatened species due to their codominant, highly polymorphic nature, and cost efficiency when processing high numbers of samples.

With the rapid development and continuous improvement of techniques in molecular biology, molecular markers have been widely used for population genetic diversity analysis, breeding programs and evolutionary research (Rasheed *et al*., 2014). Among the molecular marker types, simple sequence repeat (SSR) markers, including genomic SSR (gSSR) and expressed sequence tag SSR (EST-SSR), are highly regarded because of their rich content distribution, high polymorphism, ease of use, and repeatability (Wei *et al*., 2011). Compared with gSSR, EST-SSR has a higher level of transferability to related species, are more evolutionarily conserved, have lower development costs, and closer relationships with genes of known functions (Meyer *et al*., 2017).

1.2 Statement of the problem

In Kenya, the survival of O. lanceolata is endangered due to overexploitation for its valuable essential oils, which are used in the production of perfumes, high-end cosmetics, medicines, and drugs (Andiego et al., 2019). The mode of harvesting sandalwood is very destructive and unsustainable as the whole tree is uprooted to get the heartwood from stem, stump and roots (Ruffo *et al*., 2002). Major threats to the survival of this species in the wild include anthropogenic activities, lack of inadequate data on species population status, poor regeneration rate, and unsustainable harvesting (Frick, Kingston, & Flanders, 2020). Despite restrictive controls on the illegal trade of sandalwood, the tree continues to be poached at an alarming rate annually (Bunei, 2017). This species is protected in Kenya and United Republic of Tanzania under Presidential decrees. In Kenya, Legal Notice No 3176 of 2007 under the Forests Act, 2005 placed East African Sandalwood under Presidential Protection to allow for development of mechanisms for sustainable harvesting of the species.

The African sandalwood is among the most prized and heavily exploited medicinal plants in Kenya (Edna et al., 2009). It has been illegally traded in Kenya and across East Africa for more than two decades (Ruffo et al., 2002; Mwang'ingo et al., 2007; Bunei 2017). Illegal harvesters often uproot the entire tree, including its trunk and roots (see Plate 1.1 & 1.2), as these parts are believed to contain more oil per unit weight. Female trees are particularly targeted due to the perception that they possess higher oil content than males. This practice disrupts the species' ability to reproduce and threatens its genetic diversity (Otieno et al., 2016), potentially leading to extinction unless appropriate conservation measures are implemented.

Plate 1.1: Photograph showing *O. lanceolata* illegal harvesting from natural habitat (Courtesy: Suleiman Mbatiah; Nation media group).

Plate 1.2: Photograph showing a whole stump of *O. lanceolata* harvested (Courtesy: Suleiman Mbatiah; Nation media group).

1.3 Justification of the study

Microsatellites, also known as simple sequence repeats or SSRs, are the most widely used kind of molecular markers due to their high reproducibility, capacity to detect high levels of genetic polymorphism, codominant inheritance, and abundance in plant genomes.

Microsatellites are the most popular and versatile genetic markers with a myriad of applications in population genetics, conservation biology and evolutionary biology. Since they are a practical molecular marker for determining genetic divergence and diversity both within and between populations, microsatellites have gained widespread acceptance.

To establish robust management strategies for *O. lanceolata*, studies to generate scientific data on the species population status and genetics are necessary. The African sandalwood populations are reported to be declining but information on population dynamics across its entire range of distribution is lacking. Inadequate information on genetic diversity and population, structure for *O. lanceolata* makes it very difficult to implement informed strategies for *in situ* and *ex situ* conservation (Andiego *et al*., 2019). There is need to study the genetic diversity within and among natural populations to enhance informed strategies for conservation and utilization. This is especially important for the Kenyan populations where illegal logging is most prevalent. Quantifying the overall diversity of *O. lanceolata* is a critical first step to managing its conservation and developing plans for genetic improvement and commercial development of the species.

Assessment of genetic diversity is crucial in identification of populations for conservation priority and creating baseline data for informed conservation strategies. Effective conservation of species requires data on the levels and distribution of genetic diversity within and among populations. To implement an effective conservation programme, it is essential to understand the genetic structure of endangered populations and the dynamics of genetic variation over space and time. A population that is genetically diverse merits to be given a high priority for conservation because of the expected high adaptive potential in a rapidly changing environment (Otieno *et al*., 2016; Bunei, 2017). There is need to determine the extent of genetic variability in *Osyris lanceolata* populations that would be desirable for conservation programs.

Information on the genetic diversity of *O. lanceolata* would therefore be essential for formulation of conservation strategies to mitigate the effects of anthropogenic activities, reproductive failures, low regeneration and recruitment. The genetic information will provide information such as level of genetic variation (high to low) in each population and the genetic differences among population that will be useful to evaluate good stock for conservation and restoration programmes.

1.4 Objectives

1.4.1 Broad objective

To develop and characterise microsatellites markers to use in analysing the genetic diversity and population structure of *O. lanceolata* populations in Kenya.

1.4.2 Specific objectives

i. To develop and characterise microsatellite markers for *O. lanceolata*.
ii. To assess genetic diversity and population structure of *O. lanceolata* using the developed microsatellite markers

1.4.3 Research hypotheses

Ho_1: Microsatellite markers are not randomly distributed along the genome of *O. lanceolata*.

Ho_2: Populations of *O. lanceolata* are genetically diverse and differentiated across the distribution range in Kenya.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

2.1.1 Taxonomic description of *Osyris lanceolata*

Osyris lanceolata, also known as African sandalwood, watta bush or Camwood, is used for its scented wood and to extract essential oil. It belongs to the angiosperm order Santalales, family *Santalaceae*. The family *Santalaceae* comprises about 35 genera and 400 species widely distributed but common in tropical or sub-tropical regions (Da Silva, Soner, & Nataraj, 2016). Three genera in this family are found in Africa, namely *Thesium, Osyridicarpus* and *Osyris* (CITES, 2013). The largest genus in this family is *Thesium* with more than 200 species and mainly native to Africa and Mediterranean regions (CITES, 2013). . The species is highly variable in morphology, especially leaf size and shape (Figure 2.4) depending on climate, altitude, edaphic variables and sex type which may account for the occurrence of various synonyms.

Table 2.1: Scientific Classification

Kingdom:	Plantae
Clade:	Tracheophytes
Clade:	Angiosperms
Clade:	Eudicots
Order:	Santalales
Family:	Santalaceae
Genus:	*Osyris*
Species:	*O. lanceolata*
Binomial name:	*Osyris lanceolata* Hochst & Steud.
Synonyms:	*Osyris abyssinica* Hochst. ex A.Rich., etc.

2.1.2 Morphological description

Osyris lanceolata is an evergreen tree, ranging in height from 2 to 6 meters. It is characterized by multiple stems and a spreading growth habit, forming a rounded to irregular canopy as depicted in Plate 2.1. The young branches of this tree exhibit a blue-green hue, while its bark appears smooth and dark brown to blackish in color. It functions as a partial root parasite, relying on the roots of other plants for nutrients while still producing its own chlorophyll. Consequently, it is commonly found growing in close association with shrubs of various woody species. In Kenya, it is known to form associations with several species from genera such as *Acacia, Albizia*, *Rapanea, Bridellia*, *Cordia*, and *Teclea.*

Plate 2.1: Photograph of *O. lanceolata* growing together with a host.

(Source: Author, 2023)

The leaves of *Osyris lanceolata* are characterized by their smooth, thick, and leathery texture, arranged alternately along the stem (as seen in Plate 2.2). They are typically 13–50 mm long, sharply pointed, and have a blue-green color, often exhibiting a waxy bloom. The small, yellowish-green flowers are produced in axillary clusters on long, slender stalks. These leaves are small, simple, and rigid, with a characteristic upward orientation. They range in size from 30-45 × 10-25 mm, although their length can vary from 13-50 mm. They are lance-shaped or occasionally egg-shaped, with a broad tapering apex and a fine, sharp tip. The base of the leaf also tapers broadly. The leaf margin is smooth and rolled, and the leaves are typically hairless, with a grey-green or blue-green color and a waxy bloom that can be rubbed off. Often, the leaves display an orange-colored margin. The attachment of the petiole to the stem forms ridges running down the stem. *Osyris*

*lanceolat*a exhibits significant morphological variation, particularly in leaf size and shape, influenced by factors such as climate, altitude, soil conditions, and sex type. This variability may explain the occurrence of various synonyms, as depicted in Plate 2.2.

Plate 2.2: Photograph showing different leaf morpho types of *O. lanceolata* growing within KEFRI, Muguga from different plants. (Source: Author, 2023)

Osyris lanceolata typically has multiple stems (as seen in Plate 2.3) with dark brown to blackish or grey, smooth bark (see Fig. 1.6). The branches are upright, angular, and stiff, while the branchlets are greenish-blue, angular, or square in cross-section. The upper twigs of the canopy spread out from a single point, and both the twigs and leaves tend to point upwards.

Plate 2.3: Photograph of multi stemmed *O. lanceolata* growing at KEFRI Headquarters Muguga Planted: 2008; (Source: Author, 2023).

The plant's flowers are unisexual and grow in short, auxiliary panicles with small bracteoles. They are tiny, yellowish-greenish, and appear in clusters of 2-3 in the leaf axils. These very small, greenish-yellow flowers are on long, slender stalks, as depicted in Plate 2.4 (Schmidt et al., 2002). All parts of the flower are in groups of four, with stamens attached to the base of the fleshy perianth lobes. Male and female flowers are on separate trees (Thomas & Grant, 2002), and O. lanceolata flowers throughout the year, with the highest flowering activity observed between January to April and August to December (Kamondo et al., 2014; Beentje, 1994). Fruits usually ripen between May and September (Aoko, 2009).

The fruit is a small, one-seeded drupe, about 15 × 10 mm in size, as described by Palmer and Pitman (1972). These fleshy, egg-shaped fruits start off green, then change to yellow and finally become bright red when fully ripe. They are adorned with a persistent calyx, as depicted in Plate 2.4 (Schmidt et al., 2002).

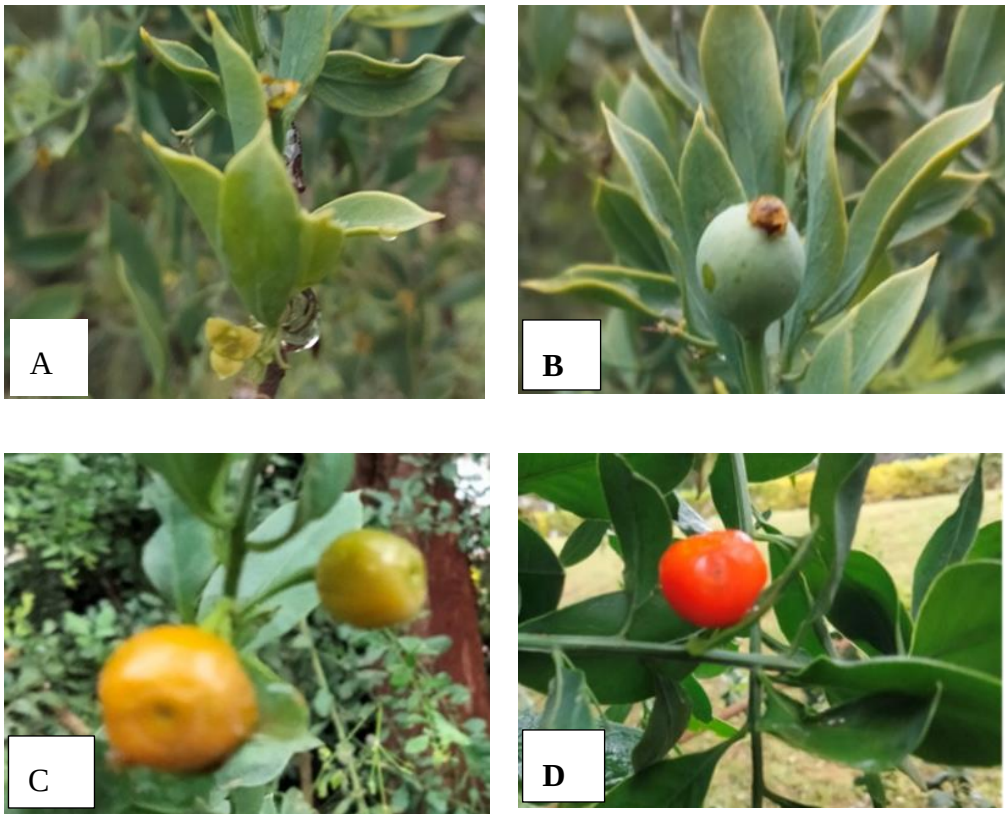

Plate 2.4: Photograph of *O. lanceolata* different fruits stages (A) flower, (B) immature, (C) mature and (D) ripe; (Source: Author, 2023)

2.1.3 Distribution of *O. lanceolata* in Kenya

Osyris lanceolata naturally thrives in a variety of habitats, including rocky outcrops, mountain slopes, forested grasslands, woodlands, and forest borders. It is found in coastal, eastern, Rift Valley, Nyanza, central, and western regions of the country. This species is

well adapted to arid and semi-arid areas (ASALs), characterized by a mean annual rainfall of 500-700 mm (Andiego et al., 2019). It typically grows in scattered patches in rocky terrain, evergreen bushlands, grasslands, thickets, and along the edges of dry forests, at altitudes ranging from 900 to 2550 meters above sea level (masl). Its distribution spans various hilly areas, including Amboseli, Loitoktok, Kibwezi (Chyullu Hills), Taita Hills, Narok, Mbeere, Kitui, Mau, Kajiado, West Pokot, Makueni, Gwasi Hills, Ngong Hills, Turkana, Laikipia, Marsabit, Koibatek, and Meru North Districts (CITES, 2013; Gathara et al., 2014).

2.1.4 Ecology

Numerous beetles feed their larvae on the rock tannin-bush, yet no significant injury has ever been seen. The dotted border butterfly, *Mylothris chloris agathina*, consumes this plant. It is commonly described as a partial root parasite that utilizes the root systems of other plants while still producing chlorophyll on its own (Andiego *et al*., 2019). This shrub is typically closely linked to the bushes of woody species, particularly *Grewia flavescens, Burkea africana*, and *Combretum molle*. *Osyris lanceolata* is associated with other plant species such as *Harrisonia abbysinica, Euclea divinivorum, Lantana camara, Cajan cajanis, Rhus natalensis, Rhus vulgaris, Maytenus acuminate, Croton megalorcapus, Acacia kirkii, Grewia similis* and *Dandonea viscosa* amongst others (Mumbu *et al.*, 2020)

2.1.5 Reproductive biology of *O. lanceolata*

In the natural field, *O. lanceolata* develops from seeds. Fresh seeds must be put directly into the earth in the garden. Although no preparation is required, nicking the seed's base

enhances germination (Andiego *et al*., 2019). After six weeks, seed germination can reach 60%. When still in the early phases of growth, plants prefer shaded, well-drained clays or deep loam soil. They are drought and cold resilient, and they grow extremely slowly.

The species is dioecious, which means that the trees that bear the male and female seeds are two separate species. Although the species mating system has not been thoroughly investigated, other family studies have shown that sandalwood is an outcrossing species (CITES, 2013; Gathara *et al*., 2014). However, *S. album* is found in China using a hybrid mating strategy that includes both outcrossing and selfing (Baskorowati, 2011). Although the species' flower shape was intended for self-pollination, the low rate of self-pollination (2%) observed indicates that this species is predominantly outcrossing (Ma *et al*., 2005; Baskorowati, 2011). *O. lanceolata* pollination method and seed dissemination have not been thoroughly investigated, despite some research depictions to the contrary (Bunei, 2017; Andiego *et al*., 2019).

Although the species mating system has not been investigated, other family studies have shown that sandalwood is an outcrossing species (CITES, 2013; Gathara *et al*., 2014). However, *S. album* was found to have a mixed mating system in China that included both outcrossing and selfing (Bunei, 2017). Due to either inadequate pollen grain production or restricted pollen grain transportation, the species frequently faces the risk of having a limited reproductive success (Andiego *et al*., 2019). The over-harvesting of female trees (CITES, 2013), the male and female trees differing budding, blooming, and pollination seasons, and the species pollination system are all factors that have been linked to the failure of the species to reproduce.

2.1.6 Socio-economic importance of *O. lanceolata*

The wood of this species is tough and long-lasting; it is utilized for building materials, fuel, and kitchenware (pestles). The root yields a strong red dye, whilst the early Cape residents utilized the bark for tanning leather. The majority of the root fibers are utilised in basketry. Tea made from roots and bark is also a tonic for soup (CITES, 2013; Gathara *et al*., 2014).

Only as a last resort for herders, ripe fruits are consumed uncooked along with the seed. *Osyris lanceolata* has historically been used by several Kenyan groups to store milk in gourds for an extended length of time. In Kenya, the root decoction is used to cure diarrhoea. In Tanzania, anaemia and various STDs are treated using a decoction made from the bark and heartwood. There are claims that the plant's extracts can treat illnesses like Hepatitis B. To create cosmetics and perfume, roots and wood are perfumed (Mweru, 2018).

The essential oils from *O. lanceolata* are extracted for use in perfumery. The major branches, roots, and heartwood of the trunk are used to refine the oil. This oil mixes effectively with a variety of aroma components, making it a popular blender-fixative used in many fragrances (CITES, 2013; Gathara *et al*., 2014). Due to the significant decline in major sandalwood oil sources, such as *Santalum album L.* (Indian subcontinent) and *Santalum spicatum* (R.Br.) A. DC. (Australia) in the 1990s, as well as the rising demand for sandalwood oil over time (Page *et al*., 2012), it became a potential commercial species in Africa. The oils contain antibacterial and blending qualities that make them acceptable for use as fixatives in other scents.

Essential oils are utilised in the cosmetic business to create premium lotions, expensive soaps, and scented candles, and they are also employed as flavourings in culinary products (Mweru, 2018). The same oils have chemo-preventive properties used to manage eruptive and inflammatory skin diseases (Page *et al*., 2012). Sandalwood oils are also effective in treating gonorrhoea, bronchitis, and other illnesses. Products from *O. lanceolata* are used in Africa to treat chest discomfort, fever, diarrhoea, candidiasis, and malaria. During religious and ceremonial rituals, Muslims, Hindus, and Buddhists burn wood and oil from *O. lanceolata* (Gupta *et al*., 2011). Its stems contain antipyretic properties for cattle in Africa, while the bark and roots make a red dye for skin tanning. The root system is helpful in phytoremediation efforts because it may be used to store heavy metals (Page *et al*., 2012).

Different Kenyans utilise *O. lanceolata* for various purposes. While the Kamba use burnt roots to heal snake bites and root fibers in basketry, the Pokot who live in Kenya's North Rift area utilise a cocktail of roots to treat diarrhoea and other stomach disorders. The Meru utilise charred stems as a milk preservative, while the Maasai and Tugens employ a bright red colour from the roots in the beverage as a herbal tea alternative (Mukonyi *et al*., 2011).

2.1.7 Threats towards loss of *O. lanceolata* populations

Due to growing overexploitation and illegal logging for the past thirty years, the drought-tolerant and multipurpose African sandalwood populations and subpopulations, especially has long been considered endangered. The most recent assessment for *The IUCN Red List*

of Threatened Species in 2017 lists *Osyris lanceolata* as Least Concern, but this is in part due to lack of data (Wilson, 2018).

Climate change, forest fires, insects' pests, illnesses, natural catastrophes, and human activities including logging, farming, and settlement are the main challenges to the survival of many forest tree species. Despite numerous government interventions, *O. lanceolata* populations in Kenya, Tanzania, Uganda, and other countries in East Africa have declined due to overexploitation and poaching from the wild (CITES, 2013; Wilson, 2018). For example, it is reported that *O. lanceolata* populations in Handeni district of Tanga region of Tanzania, have nearly vanished because of ongoing unlawful harvesting (Bunei, 2017).

2.2 Genetic resource conservation

The entire number of genes and traits that make up a plant's genetic resource are found in that species. Due to habitat loss and ecosystem damage, the biodiversity of forests is seriously under danger (Mba, 2013). In order to preserve genetic variety and prevent extinction, it is crucial to conserve tree germplasm (Kasso & Balakrishnan, 2013). Management of tree species and genetic resource conservation are different from those of food crops. The lengthy generation interval, numerous genera, and outbreeding reproductive characteristics of trees necessitate a more adaptable and intricate strategy for genetic resource preservation (Cai *et al*., 2021).

Tree genetic resources can be managed and conserved using one of two basic methods: *in situ* conservation, which entails preserving the genetic diversity of the species at hand in the environment where it was first encountered, such as in the wild, or conventional

farming systems. The goal of this strategy is to preserve the evolutionary processes that create species variety. *Ex situ* conservation refers to the preservation of a plant's genetic material outside of its natural habitat (Hawkes, Maxted & Ford-Lloyd, 2012).

Ex situ includes putting seeds or other planting materials from a species in a niche that is distinct from its natural habitat but has similar ecological and climatic circumstances. The benefit of *in situ* conservation over *ex situ* is that it preserves the genetic interactions that naturally exist between crops, their wild relatives, and the local environment. *Ex situ* methods, on the other hand, impede adaptive evolutionary development, particularly when it comes to disease and pest resistance (Hawkes, Maxted & Ford-Lloyd, 2012).

According to scientific data for Kenya, the majority of the surviving plant species of *O. lanceolata* may be found in gazetted and protected regions, while most of the unlawful harvesting occurs in private and community forests (Mumbu, 2020). It has been reported by Kenya Forest Service (2010) that it is highly challenging to implement the presidential prohibition on sandalwood harvesting in Kenya. To make sure that the sandalwood is effectively preserved and conserved inside the protected regions, Kenyan Wildlife Service (KWS) was granted an extra responsibility (Karanja, 2012). At Muthale and Chuluni, Kenya Forestry Research Institute (KEFRI) has set up demonstration trials to track the growth trends of the *O. lanceolata* species. The demonstration plots act as centres for trainings and community awareness (Gathara *et al*., 2022).

2.3 Genetic diversity and its importance

In forest ecosystems, maintaining genetic diversity is critical for both short-term and long-term productivity. Threatened and fragmented plant species can preserve their genetic

diversity if they have a broad genetic base and exhibit outcrossing mating preferences. This allows them to retain genetic diversity, reduce inbreeding, and facilitate gene flow within and among populations, enhancing their ability to adapt to environmental changes. Maintaining genetic diversity also enhances pest resistance and mitigates the negative impacts of inbreeding. Genetic diversity in plants can be measured using metrics such as heterozygosity and the percentage of polymorphic loci. Heterozygosity, as defined by Hoffmann, Miller & Weeks (2021), is the average percentage of loci in a single tree that possess two distinct alleles at a given location. The maximum heterozygosity level, indicating a wide genetic variety, is 0.5.

Observed heterozygosity (Ho) can be calculated using co-dominant molecular markers. Expected heterozygosity (He) can be estimated using both dominant and co-dominant molecular markers, assuming certain population characteristics such as structure, size, and inheritance mode. The proportion of polymorphic loci, which represents the number of polymorphic loci in an individual's genome as a percentage, is another measure of genetic diversity (Gathara et al., 2022). Genetic polymorphism occurs when more than two alleles are present at a locus, each with frequencies greater than 1% (Kasso & Balakrishnan, 2013). Rare alleles are those found in a population with a frequency of less than 1%.

2.4 Measurement of genetic diversity

Utilizing several markers, such as morphological markers, biochemical markers, and DNA molecular markers, genetic diversity in tree species is measured. Because they may be found in an infinite number and are unaffected by environmental or developmental variables, DNA markers are the most trustworthy of them. Several studies have examined

the genetic diversity in sandalwood to date using molecular tools and methodologies such and isozymes, RAPD, RFLP, SSR, and ISSR markers (Sandeep *et al*., 2020).

2.4.1 Morphological markers

These are a tree or plant species' outward characteristics. Since there are so few Mendelian physical features found in tree species, it is challenging to employ such markers in research on forest diversity. Traditional methods for assessing the genetic diversity of tree species have employed morphological characterisation; however, these morphometric markers are unreliable because they are easily altered by environmental factors (Briggs & Walters, 2016).

2.4.2 Isozymes

Since the 1950s, isozymes have been used to characterize plant genetic resources for management and to assess genetic diversity, research population genetics, and establish phylogenetic connections (de Carvalho *et al*., 2013). Different enzymes come in the form of isozymes, yet they all have the same catalytic activity. They arise from amino acid modifications that result in net charge shifts or conformational (spatial structure) changes that modify the electrophoretic properties of the proteins. Individual samples' isozyme profiles can be shown by adding a particular enzyme stain (Hilsdorf *et al*., 2017).

Since no DNA extraction, primers, probes, or sequence information is needed, isozymes are useful. They are scarce, exhibit modest levels of polymorphism, and are susceptible to environmental influences. Despite these significant drawbacks, isozymes have been

employed to evaluate the outcrossing rate of crops and to determine the population divergence of certain crops (Burow & Blake, 2019).

2.5 DNA (molecular) markers

The genomes of sandalwood contain PCR-based molecular/DNA markers such as RAPD, ISSR, and SSR. Randomly amplified polymorphic DNA (RAPD) markers (Briggs & Walters, 2016), simple sequence repeat markers (SSRs), and restriction fragment length polymorphism (RFLP) markers (Edwards et al., 2020) have been employed to examine the genetic diversity of sandalwood. Studies on sandalwood genetic diversity have predominantly utilized RAPD and ISSR analyses conducted independently.

2.5.1 Random Amplified Polymorphic DNA (RAPD)

Random DNA sequences are amplified using a PCR-based technique called RAPD (Kumar & Gurusubramanian, 2011). It is a dominant marker that binds to several locations dispersed across the plant genome by using a single short primer (8–10 base pairs) under low temperature circumstances. Then, on an agarose gel, amplicons produced from the DNA of various populations are separated. The benefit of this method is that, unlike AFLP, it does not require prior knowledge of the DNA sequence of the species being researched (Hilsdorf *et al*., 2017). The procedure is also inexpensive since there is a likelihood that the primer used may produce PCR products (amplicons).

Analysis of randomly amplified polymorphic DNA (RAPD) is based on targets that are not specified. The use of RAPD markers do not need a significant quantity of DNA, previous knowledge of the DNA sequences, or the use of radioisotopes (Kumar &

Gurusubramanian, 2011). According to Russell *et al*. (1993), RAPD offer a precise measurement of the genetic variation's distribution across populations and geographic regions. RAPD markers have been used to evaluate the genetic diversity in sandalwood species, for example Santalum album and S. spicatum (Shashidhara *et al*., 2003), and wide range of other taxa, *Breonadia salicina* , *Rheum officinale* , *Eucalyptus globulus* (Nesbitt *et al*., 1995), *Populus* .*Prunus africana*, a species that is endangered in Africa, has had its genetic diversity analyzed using RAPD (Muchugi *et al*., 2006). Using RAPD, the genetics of *Vitex fischeri*, a significant hard wood tree in central Kenya, has been investigated. A medicinal tree called *Commiphora wightii* has recently been shown to contain variation thanks to the application of RAPD (Suthar *et al*., 2008). However, RAPD markers have several limitations: foremost being low reproducibility, making them unsuitable for transference or comparison of results in similar species.

2.5.2 Amplified Fragment Length Polymorphism (AFLPs)

High levels of polymorphism can be found using the reliable, safe, and highly repeatable AFLP marker approach. It is not necessary to be familiar with an organism's genome beforehand (Bog *et al*., 2013). The genetic variation of numerous medicinal plants has been studied using this approach, namely *Zingiber cassumunar* Roxb, *Adhatoda vasica* and *Andrographis paniculata* (Varma & Shrivastava, 2018), *Croton antisyphiliticus Mart* and *Croton antisyphiliticus* Mart.

2.5.3 Inter Simple Sequence Repeats (ISSRs)

Microsatellite segments surround genomic DNA segments known as ISSRs has been used in *O. lanceolata* (Andiego *et al*., 2019). When compared to RAPDs and AFLPs, the

markers have superior repeatability thanks to their wide distribution across the genome and high degree of variation. ISSR and RAPDs markers have been widely utilised in combination in the studies of the genetic diversity of medicinal plant species including *Osyris lanceolata* (Andiego *et al.,* 2019) *Costus Pictus*, (Naik *et al*., 2017), *Canthium parviflorum* (Kala *et al*., 2017), *Penthorum chinense* (Mei *et al*., 2017), *Hedychium coronarium* (Parida *et al*., 2017) and *Stevia rebaudiana* (Sharma *et al*., 2016).

2.5.4 Microsatellites (SSR)

Microsatellites, also known as simple sequence repeat (SSR) markers, are highly effective molecular tools for assessing genetic variation in plants due to their pronounced polymorphism (Kala et al., 2017). SSR markers are characterized by repetitive sequences consisting of mono-, di-, tri-, tetra-, penta-, or hexa-nucleotide units arranged in tandem repeats of 1-10 nucleotide motifs. They demonstrate locus-specific codominance, high heterozygosity, wide distribution across the genome, and easier detectability compared to other molecular markers. These microsatellite markers have been successfully employed in genetic assessments of various plant species, including Vicia faba (El-Esawi, 2017) and *Campomanesia adamantium*. They have also contributed to studies on the genetic diversity of Santalum album in India (Fatima et al., 2019). Overall, SSRs stand out as one of the most informative molecular markers for plant genetic research.

Advancements in sequencing and bioinformatics have facilitated the creation of new SSR markers. Next-generation sequencing (NGS) technology, for instance, enables rapid and cost-effective identification of SSRs. Numerous plant species, including *Medicago sativa, Elymus sibiricus, Onobrychis viciifolia, Angelica gigas, Lentinula edodes*, and *Spondias*

tuberosa, have seen the development of a substantial number of SSR markers through high-throughput sequencing.

Due to their inheritance pattern characterized by codominance, high polymorphism, stability, repeatability, widespread distribution, and cost-effectiveness, microsatellites (SSRs) are widely favored molecular markers for assessing genetic diversity and population structure. In this study, we utilized ten highly polymorphic SSR markers to evaluate the genetic diversity and population structure across ten natural populations within the distribution range.

CHAPTER THREE

METHODOLOGY

3.1 Study area

The study was carried out across the natural distribution range of the species in Kenya (Figure 3.1). Populations of *O. lanceolata* occurring in Gwasi Hills, Kitui, Chyulu hills (Kibwezi), Kikuyu escarpment forest (Gachuthi), Mau, Mt. Elgon, Makueni, Baringo, Embu and Laikipia were sampled based on survey data from the National Museum of Kenya (What is the article cited and year of publication)). Ten populations were selected for the study across the distribution range. GPS was used to geo-reference each individual sample and population sampled (Table 3.1).

Table 3.1: Geographical locations of *Osyris lanceolata* populations sampled

Population	Latitude	Longitude	Altitude (masl)	Eco-climatic zone
Baringo	0.3309	35.7993	2048	Semi-humid (III)
Embu	-0.7259	37.656	1159	Semi-humid to semi-arid (IV)
Gachuthi	-1.2082	36.6305	1933	Semi humid (III)
Gwasi	-0.677	34.1675	1349	Semi humid (III)
Kibwezi	-2.5358	37.9024	1085	Semi-arid (V)
Kitui	-1.4434	37.9949	1167	Semi-arid (V)
Makueni	-1.9454	37.4929	1369	Semi-humid to semi-arid (IV)
Mau	-0.677	34.1675	2105	Semi-humid (III)
Meru	0.7259	37.656	1159	Arid (VI)
Mt. Elgon	1.1374	34.8306	1952	Semi humid (III)

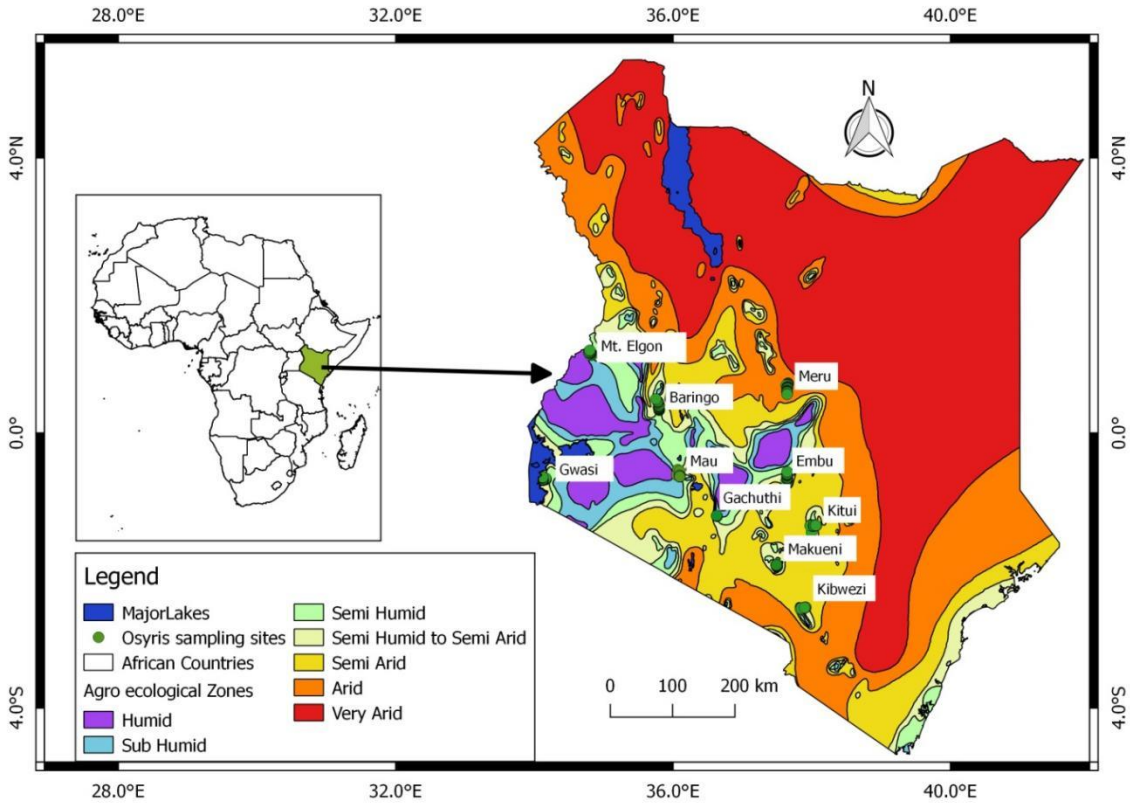

Figure 3.1: The distribution of sampled natural populations of *O. lanceolata* in Kenya. (Source: Author, 2023)

3.2 Plant materials

A total of 288 fresh leaf samples were gathered from ten natural populations across the distribution area, with each population comprising a sample size ranging from 26 to 30 individuals (refer to Table 3.2). To guarantee adequate geographical coverage, individuals were selected with a minimum distance of over 100 meters between them. The sampled populations ranged in altitude from 1085 to 2105 meters. Leaf specimens were carefully preserved in zip-lock polythene bags containing silica gel and stored at -20 ℃ in the laboratories of KEFRI Headquarters Muguga for subsequent DNA isolation and analysis.

Table 3.2: Sample size per population

Population	Sample size
Baringo	30
Embu	30
Gachuthi	28
Gwasi	26
Kibwezi	30
Kitui	30
Makueni	29
Mau	30
Meru	27
Mt. Elgon	28

3.3 DNA extraction

3.3.1 Qiagen DNeasy plant mini kit

DNA extraction was conducted using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany), following the manufacturer's instructions. Briefly, sample material was placed into 2ml microcentrifuge tubes, and these tubes were then inserted into the Tissuelyser Adaptor set 2 x 24 and secured in the clamps of the TissueLyser II (Qiagen). The samples underwent grinding for 1 minute at 30Hz, with the process repeated by reversing the tube positions within the adaptor set.

To induce membrane lysis and release DNA from nuclei, 400 µl of Lysis buffer AP1 and 20 µl of proteinase K were added to the ground samples. The mixture was thoroughly mixed and incubated in a thermo-shaker at 65 °C and 12000 rpm for 30 minutes, followed by cooling to room temperature (20 °C).

The entire content of the Eppendorf tube was then transferred to a pre-filter (Analytik Jena) and centrifuged at 13,000×g for 5 minutes. Subsequently, 700 µL of supernatant was transferred to a new 2-mL Eppendorf tube. After the addition of 500 µL of chloroform, the samples were vigorously mixed and centrifuged at 13,000×g for 15 minutes at 4 °C. The upper phase was collected, mixed with 500 µL chloroform, and centrifuged again at 13,000 rpm for 15 minutes at 4 °C. The resulting supernatant (in 250 µL batches) was collected in a 2-mL Eppendorf tube and carefully mixed with 1 mL phosphate buffer.

Following this, 600 µL of this solution was transferred to a QIAquick spin column and centrifuged at 17,900×g for 1 minute, with the filtrate discarded. The remaining sample batches underwent the same procedure with their respective columns. Subsequently, 500 µL of AW2 (70% ethanol) was added to the QIAquick spin column and centrifuged at 17,900×g for 1 minute. After discarding the filtrate, the column was centrifuged again at 14,000×g for 1 minute.

Finally, DNA was eluted into a fresh 1.5 mL Eppendorf tube by adding 30–100 µL elution buffer to the QIAquick spin column, incubating it for 5 minutes at room temperature, and then centrifuging at 14,000×g for 1 minute.

3.3.2 DNA extraction by Cetyltrimethyl Ammonium Bromide (CTAB) method

Plant genomic DNA extraction was conducted following the modified cetyletrimethyle ammonium bromide (CTAB) method outlined by Hanaoka et al. (2013). Dry leaf tissues weighing approximately 0.2 g were ground into a fine powder using a mixer mill (Model Retsch M M 400) at 60 Hz for five minutes. The ground material was then transferred to a

2.0 ml microfuge tube containing 1,000 µl of isolation buffer (IB, consisting of 10% polyethylene glycol, 0.35 M sorbitol, 0.1 M Tris-HCl-pH 8.0, and 0.5% β-mercaptoethanol). The solution was thoroughly mixed by vortexing and then centrifuged at 12,000 rpm and 4 °C for 3 minutes. The supernatant was carefully removed, and another 1,000 µl of IB was added, vortexed, and centrifuged. This step was repeated three times until the supernatant became less viscous.

Following this, the supernatant was discarded, and 300 µl of lysis buffer (0.35 M sorbitol, 0.1 M Tris-HCl-pH 8.0, and 0.5% β-mercaptoethanol) was added and mixed by vortexing. Subsequently, 30 µl of sarcosine solution was added, and the solution was incubated at room temperature for 10 minutes. Then, 300 µl of 2x CTAB buffer (2% CTAB, 0.1 M Tris-HCl, 20 mm EDTA, 1.4 M NaCl, 0.5% β-mercaptoethanol) was added, and the solution was incubated at 65 °C for 30 minutes in an oscillating water bath.

Next, an equal volume (600 µl) of chloroform isoamyl alcohol (CIA; 24:1 ratio) was added, and the solution was mixed by gentle inversion for 20 minutes and then centrifuged (14,000 rpm at room temperature) for 10 minutes. The resulting supernatant was transferred to a new 1.5 ml microfuge tube, and 600 µl of isopropanol was added. After mixing by gentle inversion, the solution was centrifuged (15,000 rpm at 4 °C) for 5 minutes. The supernatant was discarded, and 500 µl of TE (10 mM Tris-HCl and 1 mM EDTA) was added to dissolve the precipitated nucleic acid.

Subsequently, an appropriate quantity of RNase (as per the reagent supplier's protocol) was added to the solution, mixed by inversion, and then incubated at 37 °C for 60 minutes. An equal volume (500 µl) of TE-saturated phenol was added, and the solution

was mixed by gentle inversion for 10 minutes and then centrifuged (8,000 rpm at room temperature) for 10 minutes. The supernatant was transferred to a new 1.5 ml microfuge tube. This step was repeated with an equal volume (500 µl) of TE-saturated phenol, followed by gentle inversion for 20 minutes and centrifugation (8,000 rpm at room temperature for 10 minutes).

The resulting supernatant was transferred to a new 1.5 ml microfuge tube, and 1/10 volume (50 µl) of 3 M ammonium acetate was added. Then, an equal volume (550 µl) of isopropanol was added, and the solution was mixed by gentle inversion and centrifuged (15,000 rpm at 4 °C) for 5 minutes. The supernatant was discarded, and 800 µl of 70% ethanol was added. After gentle inversion, the solution was centrifuged (15,000 rpm at 4 °C) for 5 minutes to wash the DNA pellet. The supernatant was discarded, and the DNA pellet was dried for 15 minutes and dissolved in 200 µl of DNase-free water for subsequent use.

3.4 DNA quantification and quality analysis

3.4.1 Spectrophotometry

The concentration and purity of the DNA were assessed using a Biospec-nano spectrophotometer (Thermo Scientific, USA). The purity of the DNA was determined by calculating the ratio of absorbance at 260 nm and 280 nm, with a ratio of approximately 1.8 considered indicative of "pure" DNA. Subsequently, the DNA was diluted to a concentration of 20 ng/µL and stored at 4 °C until further use.

3.4.2 Agarose gel electrophoresis

DNA quantification and quality analysis were also carried out by electrophoresis on a 1 % agarose gel. Volumes of 3 µl of sample and 3 µl of loading dye (bromophenol blue, xylene cyanol and sucrose) were mixed gently by pipetting up and down before loading in the wells of submerged gels in an electrophoresis unit. The outer wells One microliter of a 40 ng/µL standard molecular weight marker (Lambda DNA Invitrogen, USA) mixed with 2 µL of loading dye for comparison. The gel was run at 100V for 50 min in 0.5X TBE buffer (0.1M Tris base, 0.1M boric acid and 0.02M EDTA; pH 8.0). The DNA bands were visualised, and images were acquired using the Gel logic 200 Imaging system (Kodak, Germany).

3.5 Processing of SSR Motifs and Primer Design

The genomic DNA obtained from the Silica gel-dried leaf sample collected from the Gachuthi population served as the DNA source for constructing the genomic library. This DNA sample was dispatched to the Gene Pool Institute of Evolutionary Biology at the University of Edinburgh for sequencing purposes. The development of SSR markers was based on Illumina paired-end DNA sequence reads obtained from the young leaves of a single O. lanceolata individual. These reads were imported into PAL Finder version 0.02.04 software (Castoe et al., 2012) to extract simple sequence repeats (SSRs) and to design primer pairs for amplification (Table 3.3).

The identified microsatellites and their corresponding primers were organized using QDD (http://net.imbe.fr/~emeglecz/qdd.html) with the settings specified in the set_qdd_default.ini file (see Appendix III). Any gaps that appeared during the scaffolding

process, conducted with SOAPdenovo (http://soap.genomics.org.cn/soapdenovo.html) using the configuration file in Appendix II, were filled using Gap Closer (version 1.12). The contigs longer than 1000 base pairs from the initial assembly were examined and annotated for functionality using Blast2GO. Utilizing this data, 48 primer pairs for loci containing either di- or trinucleotide repeats were chosen.

Table 3.3: Microsatellite primer detection and design process

Name of file	Size of file	Number of reads
140812_M01145_0150_000000000-A9HCP_1_IL-TP-002_1.sanfastq.gz	1.23 Gb 4.89 Gb	8,624,438
140812_M01145_0150_000000000-A9HCP_1_IL-TP-002_2.sanfastq.gz	1.40 Gb 4.89 Gb	8,624,438
140825_M00689_0164_000000000-AA67H_1_IL-TP-002_1.sanfastq.gz	0.92 Gb 3.60 Gb	6,371,173
140825_M00689_0164_000000000-AA67H_1_IL-TP-002_2.sanfastq.gz	1.01 Gb 3.60 Gb	6,371,173
TOTAL	4.56 Gb 16.98 Gb	29,991,222

3.6 Screening and validation of SSR Markers

The 17 selected primer pairs were tested on eight individuals from various populations of O. lanceolate (refer to Table 3.4). Each locus was individually amplified in 10-µL PCR reactions, containing 1 µL of a 1:50 dilution of template DNA (5–50 ng), 0.02 µM forward primer, 0.45 µM reverse primer, and 0.45 µM M13 primer (labelled with PET, FAM, NED, and VIC). A touchdown thermal cycling program was applied for all loci, with an annealing temperature ranging between 57-55°C. The cycling profile included an initial denaturation at 95°C for 15 minutes, followed by 10 cycles at 94°C for 30 s, 57°C for 90 s, and 72°C for 60 s (with the annealing temperature decreasing by 1°C per cycle);

and 22 cycles at 94°C for 30 s, 55°C for 90 s, and 72°C for 60 s, with a single final cycle at 60°C for 30 minutes, using a Verity 96-well thermocycler (Applied Biosystems).

The PCR analysis was conducted in a final volume of 10 µl containing 2x Multiplex PCR Master Mix (Qiagen), 0.15 µM non-fluorescently labelled forward primer (Applied Biosystems), 0.01 µM fluorescently labelled forward primer, and 0.15 µM reverse primer. Approximately 10 ng of template DNA was added. The reactions followed a touch-down procedure, with the initial cycle starting with an initial denaturation at 95°C for 15 minutes, followed by 8 cycles at 94°C for 30 s, 65°C for 1.5 minutes (with a 2°C decrease after every cycle), and 72°C for 1 minute, followed by 24 cycles at 94°C for 30 s, 51°C for 1.5 minutes, and 72°C for 1 minute, and a single final extension at 60°C for 30 minutes. The reaction was conducted using a Verity 96-well thermocycler (Applied Biosystems).

The amplified fragments were analyzed via capillary electrophoresis against an internal standard (Liz 600 size standard) on an ABI 3500 genetic analyzer (Applied Biosystems). The alleles were manually scored using Gene Mapper Software 5 (Applied Biosystems). Highly polymorphic loci were further tested using the same PCR conditions on all individuals in this study (n=288) from the 10 populations representing the Kenyan distribution range of the species.

Table 3.4: Summary of primer information used in the study and their Gene bank accession numbers

DDBJ Gene Bank accession no.	Locus	Primer sequences (5′ -3′)	Repeat motif		Allele size range (bp)
LC126834	KFOL2	F:AGAATGTCATTTGAAGGCTCGA	CGTC	57	178-194
		R:CCTTTCCTCCGTTCTCCTCG			
LC154965	KFOL7	F: CTGTGCAATGGAGAAGGCCA	ATT	61	115-120
		R:CGCGGGATTGGGATGTCATA			
LC154966	KFOL8	F:GCTGCTTCTACGGTCACTGT	CCG	58	120-130
		R:GTGGTGGATATGGAGGTGGC			
LC126835	KFOL13	F:TCCGAGGAACAGGGACTCTT	AC	60	139-165
		R:AGCGAAGAACTCATGAGCGAA			
LC154967	KFOL15	F:CATTGACGAATTGCATCCCGT	CGC	60	145-150
		R:CGTGAAGTTCAGTGCAAACC			
LC154968	KFOL16	F:TGGAGCCCATTCTCTTTCCTT	GT	59	130-160
		R:TGCACGTATTCCACATTTCCA			
LC126836	KFOL17	F:CATTGACGAATTGCATCCCGT	AG	60	178-220
		R:CGTGAAGTTCAGTGCAAACC			
LC154969	KFOL19	F:GGTAGCGAGCGGTGATATGT	TC	57	200-230
		R:ACCTAACAACTTGAAGCTCTCCC			
LC126838	KFOL24	F:CAACTCGATCGTGCATTGGC	CT	61	219-263
		R:TCCGCATATCCATTTGGCCG			
LC154970	KFOL27	F:CTAAACTGTCAGGGCTTGCT	ATG	61	225-230
		R:ATACCTTAGCTCCCGTTGCG			

LC126839	KFOL28	F:ATAAAGGCCCACGAGCTCAG	CT	60	245-255
		R:AACATCGCCATGCAGAACAG			
LC154961	KFOL29	F:GCTGAATCAGGGACAGGCAT	GA	56	230-250
		R:GGCCTCGAACAAAGTGCATG			
LC126840	KFOL30	F:CTAAACTGTCAGGGCTTGCT	TC	55	270-306
		R:ATACCTTAGCTCCCGTTGCG			
LC126841	KFOL37	F:TTTCTAGAGCTAACATACCTCTGAA	TG	59	300-340
		R:ATGACCTGGGTGCTTTGCTG			
LC126843	KFOL42	F:AGGTCCTCCTGCCTGAGAAT	TG	61	315-337
		R: CATAGGGCTGTGATGCGTCA			
LC126844	KFOL47	F:TTTGATCGTAAATTATAGATGTCCACA	CA	59	353-387
		R:CCCTTGCTTGATCTCCAGGTA			
LC126845	KFOL48	F:GAGTGCATGGAATTATGTGCGT	TC	55	369-393
		R:TCGCCATGAGAAGGGTTACT			

5′M13 tail: TGTAAAACGACGGCCAGT; F, forward sequence; R, reverse sequence;

3.7 Population analyses using the microsatellite markers

To assess the utility of the microsatellite markers developed in this research for elucidating population genetic structure and diversity, a subset of 10 microsatellite primers were chosen based on their polymorphic nature. These microsatellite markers were then examined across 288 individuals of *O. lanceolata* from 10 populations spanning the distribution range in Kenya (refer to Table 3.1).

PCR products generated by the polymorphic microsatellite primers underwent a second round of PCR. In this subsequent PCR, the forward primer of each microsatellite was 5′ labeled with four different fluorescent dyes (6-FAM, VIC, NED, and PET from Applied Biosystems). Before fragment analysis, 1 µL of PCR product from four microsatellite loci, each labeled with a different fluorescent dye, was combined and diluted at a ratio of 1:100 with sterile dH_2O. To accurately identify variations in microsatellite length among isolates, a genotyping reaction was conducted by suspending 1µL of the diluted PCR product in 8.5µl HiDi Formamide and 0.5µL of Gene Scan 600LIZ dye Size Standard (Applied Biosystems, Foster City, CA). Genotyping reactions were carried out on an ABI 3500 Genetic Analyzer (Applied Biosystems). Alleles were subsequently identified and scored using Gene Mapper version 5 software (Applied Biosystems).

3.8 Statistical analysis

For each SSR locus, important indices of genetic diversity were calculated using GenAlEx version 6.5 (Peakall and Smouse, 2012) software. The indices comprised number of alleles (N_A), number of effective alleles (N_E), Shannon's information index (I), observed heterozygosity (H_O), expected heterozygosity (H_E), inbreeding coefficient (F_{is}),

gene flow (N_m), and genetic differentiation index (F_{st}). The polymorphic information content (PIC) value and Hardy–Weinberg equilibrium (HWE) for the loci were calculated using PowerMarker version 3.25.

Analysis of molecular variance (AMOVA) was conducted using GenAlEx version 6.5 (Peakall & Smouse, 2012) to estimate the variance components of genetic variation among and within populations. Bayesian analysis was performed to evaluate the population genetic structure and detect the most likely number of population genetic clusters of *O. lanceolata* using STRUCTURE version 2.3.4.

For each simulated value of *K* (range from 1 to 10), 15 independent runs were performed with a burn-in period of 200,000 iterations followed by 1,000,000 Markov chain Monte Carlo repetitions. The Δ*K* method was implemented in the Structure Harvester (Earl & VonHoldt, 2012) program based on the STRUCTURE results to determine the optimum *K* value. The percentage membership of each individual in every cluster (*Q* value) was determined; an individual with a *Q* value higher than 0.80 was considered to have a single genetic component (a pure individual). Nei's genetic distance (DA) (Nei & Takezaki, 1983) among the 288 individuals and 10 populations was calculated using PowerMarker version 3.25 (Liu and Muse, 2005).

To explore the genetic relationships among all samples, the DA matrix was used to construct a dendrogram by hierarchical clustering with the unweighted pair group method with arithmetic mean (UPGMA) using PowerMarker version 3.25 (Liu & Muse, 2005). The dendrograms were visualised, manipulated, and annotated with the Interactive Tree of Life online tool (Letunic and Bork, 2007). Based on the standardized covariance of

genetic distance, a principal coordinate analysis (PCoA) was conducted with GenAlEx version 6.5 (Peakall & Smouse, 2012).

Bottleneck 1.2 software (Piry *et al.*, 1999) was employed to test for the recent bottleneck events using the one-tailed Wilcoxon sign-rank test with the infinite alleles model (IAM, 30%), the stepwise mutation model (SMM, 70%) and two-phase model (TPM), and a further mode-shift test was also performed. The models compare observed and expected gene diversities based on the observed number of alleles under mutation-drift equilibrium to test for compatibility with an equilibrium model. Tests for mode-shift (change in an allele frequency distribution) were also applied.

CHAPTER FOUR

RESULTS

4.1 DNA Quality and Quantity

The concentration of genomic DNA (gDNA) in the samples ranged from 330.80 ng/µl to 15,711.80 ng/µl, with a mean (M) of 2940.34 ng/µl and a standard deviation (SD) of 2027.46 (refer to Fig. 4.1). The large variance in DNA yield could be attributed to variations in the leaf samples used for DNA extraction and potential DNA loss during phase separation. The assessment of DNA purity based on the absorbance ratio of 260nm/280nm revealed a maximum ratio of 2.01 and a minimum of 1.53, with a mean (M) of 1.79 and a standard deviation (SD) of 0.098. Samples with a ratio of ≥1.80 were considered to be of high quality and purity, and were subsequently utilized in PCR analyses.

4.2 Screening and validation of SSR Markers

The summary statistics from the initial screening of 17 selected primer pairs on 84 O. lanceolata samples from three natural populations (Mt. Elgon with 28 samples, Gachuthi with 27 samples, and Kitui with 29 samples) are presented in Table 4.1. The number of alleles per locus varied across the three populations, ranging from one (KFOL27) to 21 (KFOL17). Expected heterozygosity showed a range from 0.00 (KFOL15, KFOL27, and KFOL29) to 0.902 (KFOL24) in the Mt. Elgon population, 0.00 (KFOL28) to 0.824 (KFOL7, KFOL15, KFOL17, KFOL19, and KFOL27) in the Gachuthi population, and 0.00 (KFOL7, KFOL15, KFOL19, KFOL27, and KFOL42) to 0.863 (KFOL17) in the Kitui population (refer to Table 1). The total paternity exclusion probability (Pe) across all

loci was calculated as 0.989. Following Bonferroni correction, only one pair of loci (KFOL16 - KFOL37) showed significant Linkage Disequilibrium (LD) at the 5% level. Deviation from Hardy-Weinberg equilibrium (HWE) was observed for one locus (KFOL47) in the Mt. Elgon population (refer to Table 4.1). Among the 17 markers developed, 12 were found to be polymorphic, while five (KFOL7, KFOL8, KFOL15, KFOL27, and KFOL29) were monomorphic.

In total, 163 alleles were amplified by the 10 polymorphic SSR markers among 288 *O. lanceolata* individuals, with a mean of 7.000 observed alleles per locus, and ranging in length from 132 to 350 bp. The number of alleles per marker varied from 5 (KFOL 2) to 25 (KFOL 17). Among the 10 markers, the observed heterozygosity (H_O) of nine markers was lower than the expected heterozygosity (H_E), and the average inbreeding coefficient (F_{is}) was positive. Nine SSR markers showed moderate or high polymorphism levels (PIC > 0.25) and two markers (ssps344 and KNUPF67) showed low polymorphism levels (PIC < 0.25) among the tested *O. lanceolata* populations. All SSR genotyping data for the 10 loci showed strongly significant deviation from the HWE ($p < 0.01$).

Table 4.1 Descriptive statistics over all loci for three natural population.

Locus	Na	H_O Mt. Elgon	H_O Gachuthi	H_O Kitui	H_E Mt. Elgon	H_E Gachuthi	H_E Kitui
KFOL2	5	0.393	0.556	0.483	0.572	0.626	0.569
KFOL7	2	0.043	0.000	0.000	0.043	0.000	0.000
KFOL8	2	0.000	0.200	0.462	0.073	0.184	0.434
KFOL13	7	0.556	0.148	0.069	0.552	0.139	0.067
KFOL15	2	0.000	0.000	0.000	0.000	0.000	0.000
KFOL16	5	0.107	0.333	0.107	0.103	0.352	0.166
KFOL17	21	0.893	0.741	0.793	0.879	0.824	0.863
KFOL19	3	0.259	0.000	0.000	0.338	0.000	0.000
KFOL24	15	0.821	0.192	0.276	0.902	0.286	0.452
KFOL27	1	0.000	0.000	0.000	0.000	0.000	0.000
KFOL28	5	0.714	0.000	0.069	0.605	0.000	0.067
KFOL29	2	0.000	0.074	0.034	0.000	0.073	0.034
KFOL30	12	0.643	0.333	0.483	0.614	0.471	0.663
KFOL37	17	0.889	0.185	0.517	0.853	0.278	0.609
KFOL42	6	0.308	0.037	0.000	0.277	0.036	0.000
KFOL47	15	0.393*	0.731	0.759	0.791	0.771	0.826
KFOL48	12	0.357	0.519	0.621	0.343	0.666	0.519

Na, number of observed alleles per locus, H_O observed heterozygosity, H_E expected heterozygosity

4.3 Test for deviation from Hardy – Weinberg equilibrium (HWE)

Each population's single locus deviation from HWE was determined at a confidence level ranging from 95% to 99.9 % ($P<0.005$) to ($P<0.001$). The Kitui and Mt Elgon populations had only 1 locus out of the total 10 loci used in this study KFLO24 and KFLO47, respectively, that deviated significantly from the HWE. Kibwezi, Meru and Baringo showed the highest number of loci that deviated significantly both at 80 % and 70 %, respectively, out of the total loci studied (Appendix II). In the Gwasi, Mau and Embu populations, a total of 6 loci (which was equivalent to 60% of the total loci) were in HWE below the expected 95% confidence level ($P<0.005$). In this case, 4 loci (40% of the total

loci) deviated significantly from HWE. Two loci namely, KFOL28 and KFOL42 were monomorphic with Gachuthi and Kitui populations (Appendix IV)

4.4 Population genetic diversity

The genetic diversity indices for the ten populations are summarized in Tables 4.2 and 4.3. Across the 288 individuals, a total of 178 alleles were detected at the ten microsatellite loci. The average number of alleles (A), allelic richness (AR), observed heterozygosity (Ho), and expected heterozygosity (H_E) across all loci ranged from 5 to 25, 4.5 to 18.5, 0.112 to 0.815, and 0.226 to 0.869, respectively (refer to Table 4.2). The diversity parameters exhibited variation among the populations, as detailed in Table 4.3. The observed number of alleles (Na), effective number of alleles (Ne), Shannon's information index (I), observed heterozygosity (Ho), expected heterozygosity (HE), and coefficient of inbreeding (FIS) averaged across the populations ranged from 4.0 to 9.8, 2.3 to 4.7, 0.805 to 1.600, 0.339 to 0.635, 0.406 to 0.669, and 0.026 to 0.350, respectively (refer to Table 4.3). The mean number of alleles per locus per population was 7.2, ranging from 4.0 in Gachuthi to 9.8 in Baringo. The highest HE (0.702) was observed in the Meru population, while the lowest value (0.406) was found in the Gachuthi population.

Table 4.2: Statistical values of microsatellite markers on 288 samples across 10 populations of *O. lanceolata* in Kenya

Locus	A	AR	H_O	H_E	F_{IS}	F_{ST}	Nm
KFOL24	21	12.5	0.667	0.694	0.081	0.202	0.985
KFOL28	10	6.6	0.465	0.454	0.017	0.373	0.421
KFOL42	9	4.7	0.112	0.226	0.406	0.606	0.163
KFOL47	19	14.1	0.587	0.779	0.268	0.113	1.955
KFOL48	15	10	0.387	0.476	0.194	0.349	0.467
KFOL2	5	4.5	0.477	0.53	0.118	0.165	1.263
KFOL37	22	14.4	0.623	0.685	0.131	0.213	0.924
KFOL30	22	11.5	0.612	0.703	0.152	0.162	1.292
KFOL17	25	18.5	0.815	0.869	0.08	0.069	3.383
KFOL13	15	7.9	0.318	0.449	0.268	0.359	0.446
Mean	16.3	10.47	0.5063	0.5865	0.1715	0.2611	1.1299

For each locus: Average number of alleles (A), Allelic richness (AR), Observed heterozygosity (*Ho*), Expected heterozygosity (*He*), Coefficient of inbreeding (F_{is}), Genetic differentiation coefficient (F_{st}), Gene flow (*Nm*).

Table 4.3: Genetic diversity parameters for the 10 *Osyris lanceolata* populations over all loci.

Population	N	Na	Ne	I	N_P	H_O	H_E	PPL %	F_{IS}
Gwasi	26	5.1	2.8	1.132	0.8	0.480	0.576	100	0.21
Mt. Elgon	28	8.6	4.3	1.488	0.3	0.626	0.651	100	0.03
Baringo	30	9.8	4.4	1.600	0.5	0.539	0.669	100	0.21
Mau	30	8.2	4.7	1.527	0.1	0.635	0.659	100	0.05
Gachuthi	28	4.0	2.3	0.805	0.3	0.339	0.406	90	0.15
Embu	30	6.0	3.2	1.02	0.2	0.440	0.465	90	0.11
Meru	27	8.3	4.5	1.589	0.3	0.462	0.702	100	0.35
Kibwezi	30	7.8	4.1	1.425	0.2	0.523	0.627	100	0.24
Kitui	30	5.4	2.8	1.014	0.1	0.417	0.479	90	0.14
Makueni	29	8.3	3.6	1.419	0.1	0.603	0.631	100	0.07
Overall Mean	**29**	**7.2**	**3.7**	**1.302**	0.4	**0.506**	**0.587**	**97**	**0.16**

For each locus: Number of alleles observed (Na), Effective number of alleles (Ne), Shannon's Information index (I), Observed heterozygosity (Ho), Expected heterozygosity (He), Percentage polymorphism Loci (PPL %), Coefficient of inbreeding (F_{is}).

4.5 Molecular variance among *O. lanceolata* populations

Analysis of molecular variance (AMOVA) based on 10 populations of *O. lanceolata* indicated that huge variation was within populations (62%, $P < 0.001$), whereas variation among populations was minimal at 38 % ($P < 0.001$) (Table 4.4).

Table 4.4: Analysis of molecular variance (AMOVA) within and among *O. lanceolata.*

Source of variation	d.f.	SS	Est. Var	%	P
Among populations	9	45715.034	166.785	38%	<0.001
Within populations	278	77090.855	277.305	62%	<0.001
Total	287	122805.889	444.090	100	<0.001

d.f. – Degrees of freedom, SS - Variance, MS – Mean square, Est. Var. – Estimated Variance

4.6 Population structure

The analysis of molecular variance indicated significant genetic differentiation within populations ($P<0.001$). Among the total genetic diversity, 166.785 variance components were observed among populations, while 277.305 variance was within populations, with 62% of the diversity occurring within populations (refer to Table 4.4). Additionally, the results of the AMOVA analysis were in line with the mean Fst value, suggesting low genetic differentiation among populations. Comparing these findings with the UPGMA dendrogram and the Bayesian STRUCTURE ($K = 2$) bar plot revealed a general agreement between population subdivisions and genetic relationships. The populations were predominantly grouped into two main clusters.

The inbreeding coefficient (F_{IS}) per locus ranged from 0.04 (KFOL24) to 0.51 (KFOL42), with an average of 0.194 alleles per locus. Additionally, genetic differentiation (FST) among individual loci ranged from 0.069 at KFOL17 to 0.437 at KFOL28, with an average value of 0.275 alleles per locus, suggesting a low level of genetic differentiation among the populations. Moreover, gene flow (Nm) ranged from 0.162 at KFOL42 to 3.351 at KFOL17, with an average of 1.108 (see Table 4.3). The results of the AMOVA analysis indicated that 62% of the variation was within populations, while 38% was among populations (refer to Table 4), possibly due to relatively high gene flow (Nm = 1.108) between O. lanceolata populations.

The unweighted Neighbor-joining method clustered the 288 individuals from the ten populations into two major clusters (see Fig. 4.1). These two primary clusters were further subdivided into two smaller clusters. Cluster A comprised the Gwasii, Makueni, Baringo, Mau, and Mt Elgon populations. Within this cluster, the Gwasi and Makueni populations were situated at relatively lower altitudes, while the Mau, Mt Elgon, and Baringo populations, also within the same cluster, occupied higher altitudes. Cluster B consisted of the Kitui, Embu, Gachuthi, Kibwezi, and Meru populations.

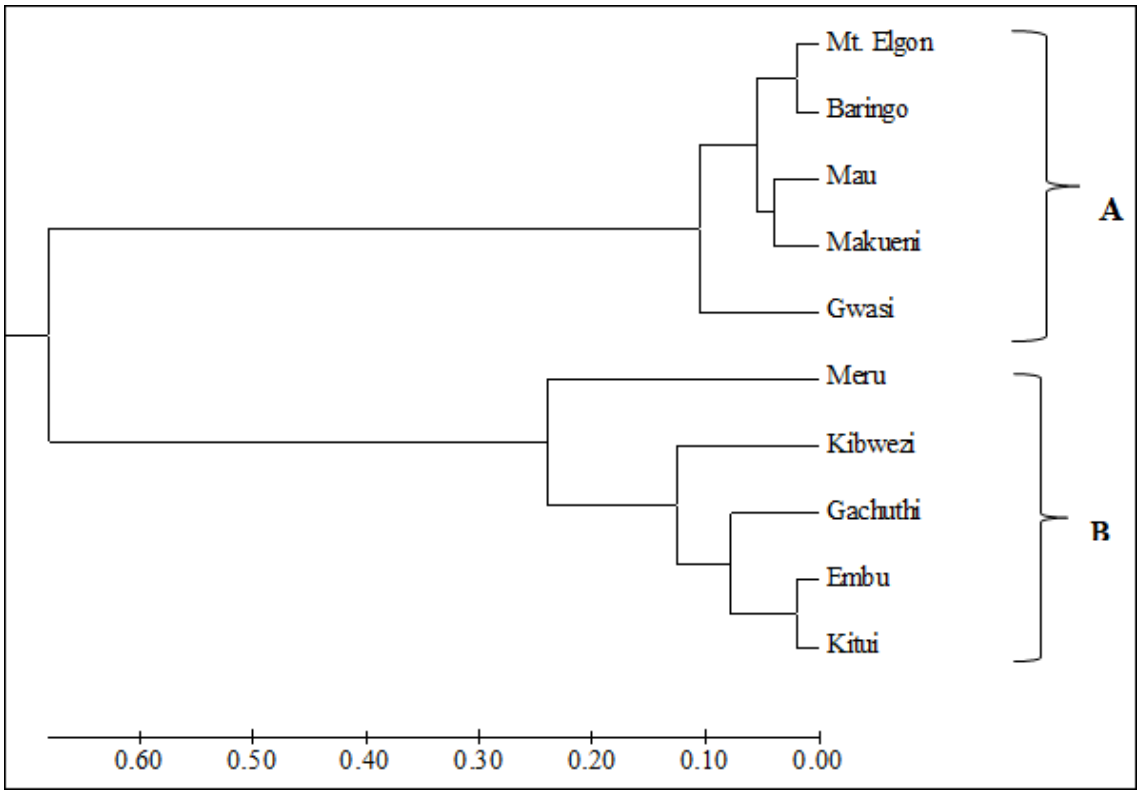

Figure 4.1: Genetic divergence of *O. lanceolata* based on UPGMA cluster analysis clustering among ten populations into two main groups, A and B.

The individuals underwent further assessment for population stratification using the STRUCTURE program. Comparing these results with the UPGMA dendrogram (see Fig. 4.1) and the Bayesian STRUCTURE (K = 2) bar plot (see Fig. 4.2) revealed a general agreement between population subdivisions and genetic relationships among populations. Notably, several genetically similar populations were found in the same geographic region. The Mantel test indicated a significant correlation between genetic distance and geographic distance among populations (r = 0.419, P = 0.005). Thus, it is suggested that the genetic structure of the O. lanceolata population in Kenya may be influenced by

geographic distance. The clustering pattern resembled the principal coordinate analysis (PCoA) based on the 10 microsatellite loci (refer to Fig. 4.3).

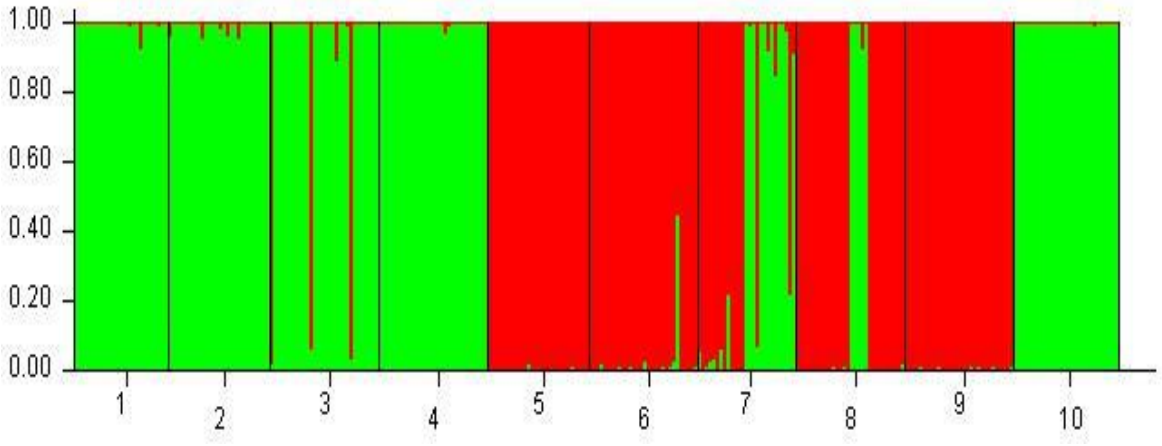

Figure 4.2: Estimated genetic structure for *K* = 2 obtained with the STRUCTURE program for 10 populations of *O. lanceolata*. 1 Gwasi, 2 Mt. Elgon, 3 Baringo, 4 Mau, 5 Gachuthi, 6 Embu, 7 Meru, 8 Kibwezi, 9 Kitui, 10 Makueni. Red represents Groups A and green represents Groups B.

4.7 Principal Coordinate Analysis

The level of genetic similarity among the samples from the different populations was visualised using principal coordinate analysis (PCoA). Principal coordinates analysis (PCoA) plot for the samples of *O. lanceolata* in Kenya exhibited partial population differentiation. As evident in the PCoA plot, samples from the same geographical location were mostly clustered together, especially for Meru and Kiambu isolates. From the principal co-ordinates analysis, it can be inferred that the number of genetic clusters is two (Fig 4.3).

PCoA analysis provides a confirmation of the results obtained by STRUCTURE and AMOVA indicating that *O. lanceolata* from Kenya exhibits partial population differentiation. To a greater extent samples from Gwasi clustered separately, occupying their space matrix and slightly mixing with isolates from Makueni, Mau, Baringo and Mt Elgon. Samples from Meru were ordinated closer to the samples from Kibwezi, Gachuthi, Embu and Kitui indicating more similarity. The occurrence of samples from Meru in clusters where the Baringo and Embu samples exist suggests some level of gene flow across different populations. The results from this particular study suggest that there was a slight exchange of genetic material between geographical locations.

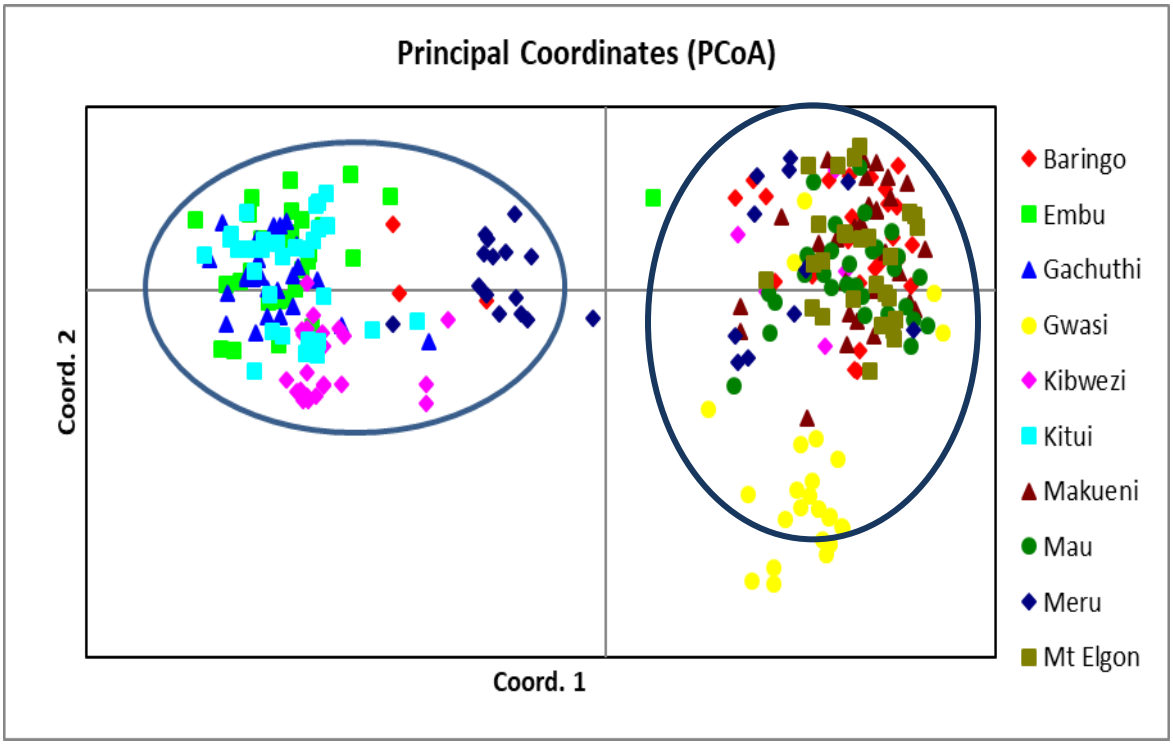

Figure 4.3: Principal coordinates analysis (PCoA) of all individuals from ten *O. lanceolata* populations. Different colours represent the individuals of different populations**.**

The pairwise matrix presents Nei's 1973 measure of genetic distance among the ten populations of *O. lanceolata*. The genetic distances between these populations varied, ranging from 0.06 between the Kitui and Embu populations to 1.681 between the Embu and Mau populations, with an overall population mean genetic distance of 0.2241. Notably, the Mt. Elgon, Baringo, and Gwasi populations appeared to be closely related, exhibiting genetic distances ranging from 0.253 to 0.275. Conversely, the Gachuthi population showed the highest level of differentiation from the other populations, with genetic distances ranging from 1.650 to 0.169, followed by Kitui, with distances ranging from 1.645 to 1.493 (Table 4.5).

Table 4.5: Nei's genetic distance of *O. lanceolata* populations in Kenya

Population	Gwasi	Mt. Elgon	Baringo	Mau	Gachuthi	Embu	Meru	Kibwezi	Kitui	Makueni
Gwasi	0.000									
Mt. Elgon	0.253	0.000								
Baringo	0.275	0.078	0.000							
Mau	0.219	0.114	0.161	0.000						
Gachuthi	1.650	2.207	1.604	2.090	0.000					
Embu	1.404	1.632	1.187	1.681	0.177	0.000				
Meru	0.774	0.601	0.496	0.624	0.560	0.414	0.000			
Kibwezi	1.085	1.439	1.133	1.386	0.361	0.239	0.598	0.000		
Kitui	1.493	1.736	1.244	1.902	0.169	0.060	0.468	0.215	0.000	
Makueni	0.216	0.135	0.150	0.114	2.023	1.608	0.557	1.454	1.653	0.000

4.8 The population bottleneck test

The population bottleneck analysis identified evidence of recent population bottlenecks in one population (Gwasi) out of the ten populations, using the infinite allele mutation model (IAM) (see Table 4.6). While some populations exhibited lower genetic diversity, there was no indication of recent genetic bottlenecks or reductions in effective population size under the two-phased mutation model (TPM), infinite allele mutation model (IAM), or stepwise mutation model (SMM). Additionally, the mode-shift test, based on the frequency distribution of alleles, revealed a typical L-shaped mode for all populations, consistent with a non-bottleneck model.

Table 4.6: Analysis of recent genetic bottleneck based on Linkage disequilibrium and Hardy's-Weinberg equilibrium in 10 populations of *O. lanceolata.*

Population	He	P-IAM	P-TPM	P-SMM
Gwasi	0.587	0.005*	0.958	0.991
Mt. Elgon	0.663	0.615	0.993	1.000
Baringo	0.681	0.615	1.000	1.000
Mau	0.670	0.246	0.884	0.988
Gachuthi	0.413	0.367	0.752	0.898
Embu	0.473	0.248	1.000	1.000
Meru	0.716	0.097	0.935	0.988
Kibwezi	0.638	0.246	0.935	0.988
Kitui	0.487	0.248	0.918	0.976
Makueni	0.642	0.722	0.997	0.998

IAM Infinite allele model, TPM Simple mutation model, SMM Simple mutation model

CHAPTER FIVE

DISCUSSION

5.1 Test for deviation from Hardy – Weinberg equilibrium (HWE)

Most of the 10 SSR primers used in the present study deviated significantly from the HWE at various levels of significance ranging from ($p < 0.05$, $p < 0.01$ and $p < 0.001$), which might reflect that the samples were collected from natural populations affected by complex environmental factors and anthropogenic activities. The study also established that the species has challenges with regeneration and spatially distributed in small patches with a long distribution range. The mean PIC was 0.429, indicating that the genetic diversity of the species in Kenya is generally moderate, which was lower than that reported for *Scutellaria baicalensis* (0.72) and *Perilla frutescens* (0.582) (Zhang *et al*., 2022).

5.2 Genetic diversity

In this study, we estimated the genetic diversity parameters of 288 *O. lanceolata* populations at 10 loci. Overall genetic diversity among the populations was high Na=7.2 and He = 0.587 (Table 4.2). Similar results were obtained in previous studies of *Santalum* species like S. *spicatum* (Millar *et al*., 2012), *S. lanceolatum* (Jones *et al*., 2010), *S. leptocladum* (Jones *et al*., 2010) and *S. album* (Fatima *et al*., 2019). This study demonstrated that the genetic diversity level of *O. lanceolata* was moderate (Ho = 0.506, He = 0.587) even though it is a rare and endangered species in Kenya/East Africa. Expected heterozygosity ranged from 0.406 in Gachuthi and 0.702 in Meru with a mean value of 0.587. Genetic diversity analysis using ISSR markers indicated a level for *O.*

lanceolata that was not high (Andiego *et al*., 2019) and lower than the results of this study. The low degree of genetic variability within *O. lanceolata* populations might be due to the fragmentation, discrimination of attributions due to random genetic drift and minimum amount of gene flow between the populations (GBIF, 2021). The average of Nm was 0.156 < 1 implying that the low rate of gene flow occurred within populations than among the populations. The possible reason for this result is that the sporadic and narrow distribution range, as well as the small sizes of populations and large spatial distances between populations, limit pollination among populations, resulting in selfing and inbreeding and potentially leading to low genetic diversity. A reduction in genetic variation might suggest a decline in adaptation to a changing environment, leading to an increased danger of extinction and increased inbreeding (CITES, 2021). The mean positive inbreeding coefficient (F_{IS}) values (0.051) indicated an excess of homozygotes in *O. lanceolata*. The results strengthened the assumption that endangered plants within a narrow distribution are generally aplastic.

5.3 Population structure and differentiation

In this study, the UPGMA cluster analysis grouped 10 populations from distinct regions into two clades (see Figure 4.1), indicating the presence of two genetic groups within the Kenyan population. Notably, the Wundanyi and Kitui populations were situated at relatively lower altitudes within this cluster. Moreover, the results of the principal coordinate analysis (PCoA) were consistent with those of the STRUCTURE analysis and aligned with the UPGMA clustering. This observation was further supported by the AMOVA analysis, which revealed that 38% of the genetic variation was among populations, while 62% was within populations.

These findings are in line with previous research on O. lanceolata using ISSR markers and are consistent with studies on other tropical tree species, such as Acacia senegal, Acacia tortilis, and Jatropha curcas. Geographical barriers are known to impede plant dispersal, and the relatively lower among-population variation observed in O. lanceolata (compared to within-population variation) may be attributed to the small sizes of its populations, which primarily consist of scattered individual trees in the wild.

Natural adaptation and geographical barriers likely contribute to the differentiation of the ten populations of *O. lanceolata* in Kenya. Additionally, the analysis indicated some level of gene flow between populations, similar to observations in *Tectona grandis*. These populations exhibited relatively high gene introgression, suggesting rich genetic backgrounds and potential centers of diversification for this species.

While pollen-mediated gene flow is expected to be limited in *O. lanceolata* due to its insect-pollination mechanism, seed dispersal through bird ingestion could facilitate gene flow over long distances, as reported in the congeneric Mediterranean species *Osyris quadripartita*. The AMOVA analysis indicated that variation within populations was 62%, while variation among populations was 38%, possibly due to moderate gene flow between *O. lanceolata* populations. Our study also revealed a significant heterozygote deficit in each population, with high values of FIS, particularly in the Gwasi population.

In this study, the UPGMA cluster analysis grouped 10 populations from distinct regions into two clades (see Figure 4.1), indicating the presence of two genetic groups within the

Kenyan population. Notably, the Wundanyi and Kitui populations were situated at relatively lower altitudes within this cluster. Moreover, the results of the principal coordinate analysis (PCoA) were consistent with those of the STRUCTURE analysis and aligned with the UPGMA clustering. This observation was further supported by the AMOVA analysis, which revealed that 38% of the genetic variation was among populations, while 62% was within populations.

These findings are in line with previous research on *O. lanceolata* using ISSR markers and are consistent with studies on other tropical tree species, such as *Acacia senegal*, *Acacia tortilis*, and *Jatropha curcas*. Geographical barriers are known to impede plant dispersal, and the relatively lower among-population variation observed in *O. lanceolata* (compared to within-population variation) may be attributed to the small sizes of its populations, which primarily consist of scattered individual trees in the wild.

Natural adaptation and geographical barriers likely contribute to the differentiation of the ten populations of *O. lanceolata* in Kenya. Additionally, the analysis indicated some level of gene flow between populations, similar to observations in *Tectona grandis*. These populations exhibited relatively high gene introgression, suggesting rich genetic backgrounds and potential centers of diversification for this species.

While pollen-mediated gene flow is expected to be limited in *O. lanceolata* due to its insect-pollination mechanism, seed dispersal through bird ingestion could facilitate gene flow over long distances, as reported in the congeneric Mediterranean species *Osyris quadripartita*. The AMOVA analysis indicated that variation within populations was 62%, while variation among populations was 38%, possibly due to moderate gene flow

between *O. lanceolata* populations. Our study also revealed a significant heterozygote deficit in each population, with high values of FIS, particularly in the Gwasi population.

5.4 Population bottlenecks

The bottleneck comparison between the observed and expected heterozygosity under a mutation-drift model using the Wilcoxon signed-rank test found none of the populations deviating significantly from mutation-drift equilibrium ($P > 0.05$) under all the mutation model's assumptions. All populations showed normal L-shape distribution of rare allele frequencies under the mode-shift test. This is an indication that the anthropogenic disturbances and the illegal poaching of the species have not left any population bottleneck signatures. Furthermore, high gene flow values found in this study suggest that there is genetic connectivity among the population that would prevent genetic drift. Effective and efficient conservation practices should be put in place as soon as possible to reduce the population decline, more so for populations like Mt Elgon Makueni and Mau, which have shown signs of reduced gene diversity. Since genetic variation mainly occurs within populations, preferential *in situ* conservation must be implemented for populations with high genetic diversity and individuals with unique genes.

CHAPTER SIX

CONCLUSION AND RECOMMENDATIONS

6.1 Conclusions

1. This study successfully developed 17 microsatellites markers, 12 polymorphic and 5 monomorphic which were used in the current study for the determination of genetic diversity and population structure of *O. lanceolata* across the distribution range in Kenya. The markers developed are the first reported for *O. lanceolata* and are suitable for population genetics studies due to their high polymorphic characteristics.
2. The genetic data obtained from this study show that Meru and Mt Elgon have the highest genetic diversity. The findings serve as a guide for enhancing germplasm and parental selection for purpose of future breeding programs and conservation efforts.
3. Two clusters were created from the ten natural populations; these groups/clusters can serve as the two management units for the conservation goal. Genetic linkages between populations discovered by SSR analysis will be useful.

6.2 Recommendations

1. The conservation strategy for O. lanceolata should focus on preserving the identified genetic clusters. To achieve this, tactics for seed collection need to be developed to establish an ex-situ seed germplasm bank. It's essential to collect as

many samples as possible from each population across its entire natural geographical distribution range.

2. The conservation strategy should integrate both in-situ and ex-situ methods to safeguard the species' valuable genetic resources, especially considering the current scenario marked by a rapid decline in population numbers and the severe endangerment of their natural habitats.
3. All populations, particularly those exhibiting high levels of genetic diversity or significant genetic distinctions, should receive protection. In-situ conservation methods are deemed most effective for safeguarding endangered plants, as they ensure the preservation of the entire gene pool within their natural habitats.
4. During ex-situ conservation efforts, artificial hybridization can be employed among populations with notable genetic differences to swiftly enhance heterozygosity. Following the cultivation of seeds collected from the field in ex-situ settings, seedlings should be reintroduced into their source sites.

REFERENCES

Andiego, K. P., Dangasuk, O. G., Odee, D. W., Omondi, F. S., Otieno, D. F., & Balozi, B. K. (2019). Genetic diversity of endangered sandalwood (*Osyris lanceolata*) populations in Kenya using ISSR molecular markers. *East African Agricultural and Forestry Journal*, *83*(2), 80-93.

Aoko E.A (2014). Evaluation of Anti-microbial activity of *Osyris lanceolate* . Unpublished Master's thesis, Jomo Kenyatta University of Science and Technology.

Balakrishnan, S., Dev, S. A., Sakthi, A. R., Vikashini, B., Bhasker T, R., Magesh, N. S., & Ramasamy, Y. (2021). Gene-ecological zonation and population genetic structure of Tectona grandis Lf in India revealed by genome-wide SSR markers. *Tree Genetics & Genomes*, *17*, 1-14.

Baskorowati, L. (2011). Flowering intensity and flower visitors of *Santalum album* L. at ex-situ conservation plot, Watusipat, Gunung Kidul, Yogyakarta. *Indonesian Journal of Forestry Research*, *8*(2), 130-143.

Bellucci, E., Mario Aguilar, O., Alseekh, S., Bett, K., Brezeanu, C., Cook, D., ... & Papa, R. (2021). The INCREASE project: Intelligent Collections of food-legume genetic resources for European agrofood systems. *The Plant Journal*, *108*(3), 646-660.

Bernatchez, L. (2016). On the maintenance of genetic variation and adaptation to environmental change: considerations from population genomics in fishes. *Journal of Fish Biology*, *89*(6), 2519-2556.

Bhandari, P. S., Bhandari, R., Sah, B. K., Gyawali, S., Bhusal, M., Shrestha, S., & Shakya, S. (2017). Antibacterial activity of methanolic extract of *Mangifera indica* (bark) and *Osyris lanceolata* (leaves) from western region of Nepal. *International Journal of Pharmacognosy*, *4*(6), 200-207.

Bhattacharyya, P., Kumaria, S., Job, N., & Tandon, P. (2015). Phyto-molecular profiling and assessment of antioxidant activity within micropropagated plants of Dendrobium thyrsiflorum: a threatened, medicinal orchid. *Plant Cell, Tissue and Organ Culture (PCTOC)*, *122*, 535-550.

Bog, M., Schneider, P., Hellwig, F., Sachse, S., Kochieva, E. Z., Martyrosian, E., ... & Appenroth, K. J. (2013). Genetic characterization and barcoding of taxa in the genus *Wolffia* Horkel ex Schleid.(Lemnaceae) as revealed by two plastidic markers and amplified fragment length polymorphism (AFLP). *Planta*, *237*, 1-13.

Briggs, D., & Walters, S. M. (2016). *Plant variation and evolution*. Cambridge University Press.

Bunei, E. K. (2017). The hunt for the precious wood: Illegal trade of sandalwood as an international criminal enterprise in Kenya. *Society and Business Review*.

Burow, M. D., & Blake, T. K. (2019). Molecular tools for the study of complex traits. In *Molecular dissection of complex traits* (pp. 13-30). CRC Press.

Cai, C., Xiao, J., Ci, X., Conran, J. G., & Li, J. (2021). Genetic diversity of *Horsfieldia tetratepala* (Myristicaceae), an endangered Plant Species with Extremely Small Populations to China: Implications for its conservation. *Plant Systematics and Evolution*, *307*, 1-12.

CITES (2013). Convention on International Trade in Endangered Species of Wild Fauna and Flora: Consideration of proposals for amendment of appendices i and ii. Sixteenth meeting of the Conference of the Parties Bangkok (Thailand), 3-14 March 2013.CoP16 Prop. 69.

Da Silva, J. A. T., Soner, D., & Nataraj, M. (2016). African sandalwood or Nepalese sandalwood: a brief synthesis. *Notulae Scientia Biologicae*, *8*(1), 57-61.

Das, S. C., & Pullaiah, T. (2021). Sustainable Use, Threats, and Conservation of Sandalwood. *Sandalwood: Silviculture, Conservation and Applications*, 267-276.

de Carvalho, M. A. P., Bebeli, P. J., Bettencourt, E., Costa, G., Dias, S., Santos, T. M. D., & Slaski, J. J. (2013). Cereal landraces genetic resources in worldwide GeneBanks. A review. *Agronomy for Sustainable Development*, *33*, 177-203.

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

Edwards, T. P., Trigiano, R. N., Ownley, B. H., Windham, A. S., Wyman, C. R., Wadl, P. A., & Hadziabdic, D. (2020). Genetic diversity and conservation status of *Helianthus verticillatus*, an endangered sunflower of the Southern United States. *Frontiers in Genetics*, *11*, 410.

El-Esawi, M. A. (2017). SSR analysis of genetic diversity and structure of the germplasm of faba bean (*Vicia faba* L.). *Comptes Rendus Biologies*, *340*(11-12), 474-480.

Esfandani Bozchaloyi, S., & Sheidai, M. (2018). Molecular diversity and genetic relationships among Geranium pusillum and G. pyrenaicum with inter simple sequence repeat (ISSR) regions. *Caryologia*, *71*(4), 457-470.

Fatima, T., Srivastava, A., Somashekar, P. V., Hanur, V. S., Rao, M. S., & Bisht, S. S. (2019). Assessment of morphological and genetic variability through genic microsatellite markers for essential oil in Sandalwood (Santalum album L.). *3 Biotech*, *9*, 1-16.

Feitosa-Alcantara, R. B., Bacci, L., Blank, A. F., Alves, P. B., Silva, I. M. D. A., Soares, C. A., ... & Arrigoni-Blank, M. D. F. (2017). Essential oils of *Hyptis pectinata* chemotypes: isolation, binary mixtures and acute toxicity on leaf-cutting ants. *Molecules*, *22*(4), 621.

Frick, W. F., Kingston, T., & Flanders, J. (2020). A review of the major threats and challenges to global bat conservation. *Annals of the New York Academy of Sciences*, *1469*(1), 5-25.

Gadissa, F., Tesfaye, K., Dagne, K., & Geleta, M. (2018). Genetic diversity and population structure analyses of *Plectranthus edulis* (Vatke) Agnew collections from diverse agro-ecologies in Ethiopia using newly developed EST-SSRs marker system. *BMC genetics*, *19*(1), 1-15.

Ganie, S. H., Upadhyay, P., Das, S., & Sharma, M. P. (2015). Authentication of medicinal plants by DNA markers. *Plant Gene*, *4*, 83-99.

Gathara, M., Kamondo, B., Mwadalu, R., Omondi, S., Kivuti, M., Wairagu, N., ... & Gachie, P. (2022). Influence of Soil Nutrients, Tree Age, and Sandalwood Provenances on Sandalwood Oil Yield and Quality. *International Journal of Forestry Research*, *2022*.

Gathara, M., Mackenzie, P., Kimondo, J., & Muturi, G. (2014). Prediction of *Osyris lanceolata* (Hochst. & Steud.) site suitability using indicator plant species and edaphic factors in humid highland and dry lowland forests in Kenya. *Journal of Horticulture and Forestry*, *6*(11), 99-106.

Govindaraj, M., Vetriventhan, M., & Srinivasan, M. (2015). Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. *Genetics research international*, *2015*.

Grover, N., Sharma, J., Sengupta, S., Singh, S., Singh, N., & Kaur, H. (2016). Long-term effect of tobacco on unstimulated salivary pH. *Journal of oral and maxillofacial pathology*, *20*(1), 16.

Gupta, H. K., Patnaik, S., Saxena, K. G., & Sharma, E. (2011). PS Ramakrishnan, KS Rao, UM Chandrashekara, N. Chhetri. *Traditional Forest-Related Knowledge: Sustaining Communities, Ecosystems and Biocultural Diversity*, *12*, 315.

Hawkes, J. G., Maxted, N., & Ford-Lloyd, B. V. (2012). *The ex situ conservation of plant genetic resources*. Springer Science & Business Media.

Herrera, C.M. (1988). The fruiting ecology of *Osyris quadripartita*: individual variation and evolutionary potential. *Ecology*, 69, 233 - 249.

Hilsdorf, A. W., Hallerman, E. M., Hilsdorf, A. W., & Hallerman, E. M. (2017). Characterization of genetic resources. *Genetic Resources of Neotropical Fishes*, 55-117.

Hoffmann, A. A., Miller, A. D., & Weeks, A. R. (2021). Genetic mixing for population management: From genetic rescue to provenancing. *Evolutionary Applications*, *14*(3), 634-652.

Idrees, Muhammad & Irshad, Muhammad. (2014). Molecular markers in plants for analysis of genetic diversity: a review. *European academic research*, *2*(1), 1513-1540.

Iftikhar, H., & Carney, G. E. (2016). Evidence and potential in vivo functions for biofluid miRNAs: From expression profiling to functional testing: Potential roles of extracellular miRNAs as indicators of physiological change and as agents of intercellular information exchange. *BioEssays*, *38*(4), 367-378.

IUCN. (2009). The 2009 IUCN Red List of Threatened Species. The World Conservation Union (IUCN), Switzerland.

Jones, B. L., Waycott, M., Robson, H. L., Calladine, A., & Page, T. (2010). Isolation and characterization of microsatellite loci in *Santalum lanceolatum* and *Santalum leptocladum* (Santalaceae). *American Journal of Botany*, *97*(10), e97-e98.

Jose, D., & Harikrishnan, M. (2019). Molecular markers and their optimization: Addressing the problems of nonhomology using decapod COI gene. In *Biochemical Analysis Tools-Methods for Bio-Molecules Studies*. IntechOpen.

Kala, M., Kala, S. C., Ratnakar, K. R., Sekhar, A. C., & Reddy, P. C. O. (2017). Assessment of genetic diversity of *Canthium parviflorum* (Lam.) by RAPD and ISSR markers. *Ann. Plant Sci*, *6*(11), 1775-1783.

Karanja, D. (2012, January). The role of the Kenya Wildlife Service in protecting Kenya's wildlife. In *The George Wright Forum* (Vol. 29, No. 1, pp. 74-80). George Wright Society.

Kasso, M., & Balakrishnan, M. (2013). *Ex situ* conservation of biodiversity with particular emphasis to Ethiopia. *International Scholarly Research Notices*, *2013*.

Kumar, N. S., & Gurusubramanian, G. (2011). Random amplified polymorphic DNA (RAPD) markers and its applications. *Sci Vis*, *11*(3), 116-124.

KWS & TWRI (2010). Aerial Total Count Amboseli – West Kilimanjaro and Magadi-Natron Cross Border Landscape. Wet Season, March 2010. Kenya Wildlife Service (KWS) and Tanzania Wildlife Research Institute. (TWRI), Nairobi.

Lagally, E. T., Emrich, C. A., & Mathies, R. A. (2001). Fully integrated PCR-capillary electrophoresis microsystem for DNA analysis. *Lab on a Chip*, *1*(2), 102-107.

Letunic, I., & Bork, P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics*, *23*(1), 127-128.

Liang, W., Dondini, L., De Franceschi, P., Paris, R., Sansavini, S., & Tartarini, S. (2015). Genetic diversity, population structure and construction of a core collection of apple cultivars from Italian germplasm. *Plant Molecular Biology Reporter*, *33*, 458-473.

Liu, K., & Muse, S. V. (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*, *21*(9), 2128-2129.

Mathi Thumilan, B., Sajeevan, R. S., Biradar, J., Madhuri, T., N. Nataraja, K., & Sreeman, S. M. (2016). Development and characterization of genic SSR markers from Indian mulberry transcriptome and their transferability to related species of Moraceae. *PloS one*, *11*(9), e0162909.

Mba, C. (2013). Induced mutations unleash the potentials of plant genetic resources for food and agriculture. *Agronomy*, *3*(1), 200-231.

Mei, Z., Zhang, X., Khan, M. A., Imani, S., Liu, X., Zou, H., ... & Fu, J. (2017). Genetic analysis of Penthorum chinense Pursh by improved RAPD and ISSR in China. *Electronic Journal of Biotechnology*, *30*, 6-11.

Meyer, L., Causse, R., Pernin, F., Scalone, R., Bailly, G., Chauvel, B., ... & Le Corre, V. (2017). New gSSR and EST-SSR markers reveal high genetic diversity in the invasive plant *Ambrosia artemisiifolia* L. and can be transferred to other invasive Ambrosia species. *PloS One*, *12*(5), e0176197.

Miah, G., Rafii, M. Y., Ismail, M. R., Puteh, A. B., Rahim, H. A., Islam, K. N., & Latif, M. A. (2013). A review of microsatellite markers and their applications in rice breeding programs to improve blast disease resistance. *International Journal of Molecular Sciences*, *14*(11), 22499-22528.

Millar, M. A., Byrne, M., & Barbour, E. (2012). Characterisation of eleven polymorphic microsatellite DNA markers for Australian sandalwood (*Santalum spicatum*) (R. Br.) A. DC.(Santalaceae). *Conservation Genetics Resources*, *4*, 51-53.

Mugula, B. B., Kiboi, S. K., Kanya, J. I., Egeru, A., Okullo, P., Curto, M., & Meimberg, H. (2021). Knowledge gaps in taxonomy, ecology, population distribution drivers and genetic diversity of African sandalwood (*Osyris lanceolata* Hochst. & Steud.): A scoping review for conservation. *Plants*, *10*(9), 1780.

Mukonyi, K. W., Oeba, V. O., Chiteva, R., Lelon, J., & Gathara, M. W. (2011). Germination and growth performance of *Aloe turkanensis* and *Aloe secundiflora* under different substrates. *African Journal of Agricultural Research*, *6*(10), 2132-2138.

Mumbu, D. M. (2020). *Assessing the utilization and socio–economic impacts of Osyris lanceolata (*hochst & steudel*) and associated environmental degradation in Kitui county* (Doctoral dissertation).

Mweru, K. (2018). *Use and Conservation of Wild Medicinal Food Plants in Loita, Narok County Kenya* (Doctoral dissertation, University of Nairobi).

Nadeem, M. A., Nawaz, M. A., Shahid, M. Q., Doğan, Y., Comertpay, G., Yıldız, M., ... & Baloch, F. S. (2018). DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. *Biotechnology & Biotechnological Equipment*, *32*(2), 261-285.

Naik, A., Prajapat, P., Krishnamurthy, R., & Pathak, J. M. (2017). Assessment of genetic diversity in Costus pictus accessions based on RAPD and ISSR markers. *3 Biotech*, *7*, 1-12.

Nei, M., & Takezaki, N. (1983). Estimation of genetic distances and phylogenetic trees from DNA analysis. *Proceedings of the 5th World Congress on Genetics Applied to Livestock* , *21*(21), 405-412.

Nsibo, D. L., Barnes, I., Omondi, D. O., Dida, M. M., & Berger, D. K. (2021). Population genetic structure and migration patterns of the maize pathogenic fungus, Cercospora zeina in East and Southern Africa. *Fungal Genetics and Biology*, *149*, 103527.

Otieno, J. O., Omondi, S. F., Perry, A., Odee, D. W., Makatiani, E. T., Kiplagat, O., & Cavers, S. (2016). Development and characterization of microsatellite markers for *Osyris lanceolata* Hochst. & Steud., an endangered African sandalwood tree species. *Tropical Plant Research, 3*, 701-703.

Page, T., Tate, H., Potrawiak, A., Berry, A., & Bled, C. (2012). Planted sandalwood developments in Vanuatu.

Parida, R., Mohanty, S., & Nayak, S.. (2017). Molecular characterization of endangered medicinal plant species *Hedychium coronarium* from Eastern India. *International Journal of Pharmacy and Pharmaceutical Sciences*, *9*(1), 173-178.

Patel, D. M., Fougat, R. S., Sakure, A. A., Kumar, S., Kumar, M., & Mistry, J. G. (2016). Detection of genetic variation in sandalwood using various DNA markers. *3 Biotech*, *6*, 1-11.

Pavoine, S., & Ricotta, C. (2014). Functional and phylogenetic similarity among communities. *Methods in Ecology and Evolution*, *5*(7), 666-675.

Peakall, R.; Smouse, P.E. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—An update. Bioinformatics 2012, 28, 2537–2539.

Rasheed, A., Xia, X., Yan, Y., Appels, R., Mahmood, T., & He, Z. (2014). Wheat seed storage proteins: Advances in molecular genetics, diversity and breeding applications. *Journal of Cereal Science*, *60*(1), 11-24.

Rauscher, G., & Simko, I. (2013). Development of genomic SSR markers for fingerprinting lettuce (Lactuca sativaL.) cultivars and mapping genes. *BMC plant biology*, *13*(1), 1-11.

Sandeep, C., Kumar, A., Rodrigues, V., Viswanath, S., Shukla, A. K., & Sundaresan, V. (2020). Morpho-genetic divergence and population structure in Indian Santalum album L. *Trees*, *34*, 1113-1129.

Shashidhara, G., Hema, M.V., Koshy, B., & Farooqi, A. A. (2003). Assessment of genetic diversity and identification of core collection in sandalwood germplasm using RAPDs. *The Journal of Horticultural Science and Biotechnology*, 78:4, 528-536, DOI: 10.1080/14620316.2003.11511659

Sharma, S., Walia, S., Singh, B., & Kumar, R. (2016). Comprehensive review on agro technologies of low- calorie natural sweetener stevia (*Stevia rebaudiana* Bertoni): a boon to diabetic patients. *Journal of the Science of Food and Agriculture*, *96*(6), 1867-1879.

Sülü, G., Kacar, Y. I. L. D. I. Z., Polat, I., KİTAPCI, A., Turgutoğlu, E., Şimşek, Ö. Z. H. A. N., & Satar, G. (2020). Identification of genetic diversity among mutant lemon and mandarin varieties using different molecular markers. *Turkish Journal of Agriculture and Forestry*, *44*(5), 465-478.

Teixeira da Silva, J. A., Kher, M. M., Soner, D., & Nataraj, M. (2019). Red sandalwood (*Pterocarpus santalinus* L. f.): biology, importance, propagation and micropropagation. *Journal of Forestry Research*, *30*, 745-754.

Vanavermaete, D., Fostier, J., Maenhout, S., & De Baets, B. (2021). Deep scoping: a breeding strategy to preserve, reintroduce and exploit genetic variation. *Theoretical and Applied Genetics*, *134*(12), 3845-3861.

Varma, A., & Shrivastava, N. (2018). Genetic structuring in wild populations of two important medicinal plant species as inferred from AFLP markers. *Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology*, *152*(5), 1088-1100.

Vieira, M. L. C., Santini, L., Diniz, A. L., & Munhoz, C. D. F. (2016). Microsatellite markers: what they mean and why they are so useful. *Genetics and Molecular Biology*, *39*, 312-328.

Wang, S. Q. (2020). Genetic diversity and population structure of the endangered species *Paeonia decomposita* endemic to China and implications for its conservation. *BMC Plant Biology*, *20*, 1-14.

Wang, J., & Shete, S. (2012). Testing departure from hardy–Weinberg proportions. *Statistical Human Genetics: Methods and Protocols*, 77-102.

Weeks, A. R., Stoklosa, J., & Hoffmann, A. A. (2016). Conservation of genetic uniqueness of populations may increase extinction likelihood of endangered species: the case of Australian mammals. *Frontiers in Zoology*, *13*, 1-9.

Wilson, B. (2018). *Osyris lanceolata*. The IUCN Red List of Threatened Species 2018: e.T200642A2675362. http://dx.doi.org/10.2305/IUCN.UK.2018-1.RLTS.T200642A2675362.en

Wu, Q., Zang, F., Xie, X., Ma, Y., Zheng, Y., & Zang, D. (2020). Full-length transcriptome sequencing analysis and development of EST-SSR markers for the endangered species *Populus wulianensis*. *Scientific Reports*, *10*(1), 16249.

Yusefi, G. H., Safi, K., & Brito, J. C. (2019). Network- and distance- based methods in bioregionalization processes at regional scale: An application to the terrestrial mammals of Iran. *Journal of Biogeography*, *46*(11), 2433-2443.

Zalapa, J. E., Cuevas, H., Zhu, H., Steffan, S., Senalik, D., Zeldin, E., & Simon, P. (2012). Using next- generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. *American journal of botany*, *99*(2), 193-208.

Zhang, C., Jia, C., Liu, X., Zhao, H., Hou, L., Li, M., ... & Li, Y. (2022). Genetic Diversity Study on Geographical Populations of the Multipurpose Species *Elsholtzia stauntonii* Using Transferable Microsatellite Markers. *Frontiers in Plant Science*, *13*.

Zheng, Y., Liu, J., Feng, X., & Gong, X. (2017). The distribution, diversity, and conservation status of Cycas in China. *Ecology and Evolution*, *7*(9), 3212-3224.

APPENDICES

Appendix I: Composition of ExoSap Master Mix

ExoSAP Mix	**x1**
Exo I (1U)	0.05µL
SAP (0.2 U)	0.2 µL
Multi Core Buffer	0.75 µL
DNA	5 µL
Final Volume	6 µL

Appendix II: Configurations file for SOAPdenovo.

max_rd_len=250

[LIB]

avg_ins=300

reverse_seq=0

asm_flags=3

pair_num_cutoff=3

map_len=32

q1=Osyris_Gachuthi_read1_files.fastq

q2=Osyris_Gachuthi_read2_files.fastq

Appendix III: General parameters.

#[0/1] 1 for funning QDD from Galaxy, 0 for running it from terminal

galaxy =0

#operating system [linux/win]

syst = linux

Full path to blast executables (including the bin folder). If the folder is in your path it can be left empty. (e.g. C:\Program Files\NCBI\blast-2.2.25+\bin or /home/pri/galaxy-dist/tools/qdd/ncbi-blast-2.2.27+/bin/) If the folder is in your path it can be left empty.

blast_path=

Full path to clustalw executables. If the folder is in your path it can be left empty. (e.g. C:\CLUSTALW2 or /home/pri/galaxy-dist/tools/qdd/clustalw-2.0.10-linux-i386-libcppstatic/)

clust_path =

Full path to Primer3_core executable. (e.g. C:\Primer3-release-2.3.6\ or C:\primer3-1.1.4-WINXP\bin or /usr/bin/ /usr/local/primer3-2.3.6/src/)

primer3_path = /home/annt/Tools/primer3-2.3.6/src/

Primer3 version [1/2] 1 for primer3-1.xxx, 2 for primer3-2.xxx

primer3_version = 2

Full path to qdd scripts. (e.g. /home/qdd/galaxy-dist/tools/qdd/ OR D:\QDD2.3\)

qdd_folder = /home/annt/Tools/qdd/

Output folder name with full path. Must be created before running qdd. If not specified output files are written in the current working directory

out_folder = /home/annt/Tools/qdd/qdd_output

#[0/1] (1 for deleting temporary files after the run)

del_files =1

string for naming outfiles. All outfile name starts with this string. If empty, the input filename is used for naming outfiles

outfile_string =

#[0/1] (1 for printing out supplementary information, only needed for debugging)

debug = 0

name of the local database including full path (e.g. /usr/local/nt/nt or D:\blastdb\nt) Only needed if local BLAST is used for contamination check

blastdb =

#number of threads for BLAST (if unsure, use 1)

num_threads = 1

[0/1] 1:run local blast for contamination check, 0:run remote blast for contamination check

local_blast = 0

Input sequences type for pipe1. 1 if sequences has been assembled (contigs, scaffolds, chromosomes), 0 if they are short sequencing reads. Same parameter is used for pipe3 to check the distance between nearest neigbours

contig = 1

#PIPE1 SPECIFIC PARAMETERS

#input file is in fastq format; [0/1] (1 for fatsq, 0 for fasta)

fastq = 0

if extracting microsatellites from contigs, get flank_length bp of flanking region on both sides of the microsatellite

flank_length = 300

#[0/1] (1 for running adapter/vector clipping step); Not applicable if contig =1

adapter= 0

#[integer] (sequences shorter then length_limit are eliminated)

length_limit= 80

#[fasta file with adaters] (can be empty if adapter=0) ; Not applicable if contig =1

adapter_file =

#PIPE2 SPECIFIC PARAMETERS

[0/1] Make consensus sequences (YES=1/NO=0)

make_cons=0

[integer] Minimum % of pirwise identity between sequences of a contig (80-100)

ident_limit =95

[floating] Proportion of sequences that must have the same base at a site to accept it as a consensus (0.5-1)

prop_maj =0.66

#PIPE3 SPECIFIC PARAMETERS

#Minimum size of PCR product

pcr_min = 90

#Maximum size of PCR product

pcr_max = 550

#PCR Product size interval

pcr_step = 50

PRIMER_GC_CLAMP = 0

PRIMER_MIN_SIZE = 18

PRIMER_MAX_SIZE = 27

PRIMER_OPT_SIZE = 20

PRIMER_OPT_TM = 60.0

PRIMER_MIN_TM = 57.0

PRIMER_MAX_TM = 63.0

PRIMER_MAX_DIFF_TM = 10.0

PRIMER_MIN_GC = 20.0

PRIMER_OPT_GC_PERCENT = 50.0

PRIMER_MAX_GC = 80.0

PRIMER_SELF_ANY = 8.0

PRIMER_SELF_END = 3.0

PRIMER_MAX_POLY_X = 3

PRIMER_NUM_RETURN = 3

Appendix IV: Summary of chi square tests for HWE for each population

Population	Locus	DF	ChiSq	P-value	Significance
Gwasi	KFOL24	28	43.396	0.032	*
Gwasi	KFOL28	10	8.682	0.563	ns
Gwasi	KFOL42	3	26.059	0.000	***
Gwasi	KFOL47	10	31.414	0.001	***
Gwasi	KFOL48	1	0.968	0.325	ns
Gwasi	KFOL2	3	5.723	0.126	ns
Gwasi	KFOL37	10	12.238	0.269	ns
Gwasi	KFOL30	36	29.717	0.761	ns
Gwasi	KFOL17	21	26.985	0.171	ns
Gwasi	KFOL13	6	26.954	0.000	***
Mt. Elgon	KFOL24	91	96.966	0.315	ns
Mt. Elgon	KFOL28	15	23.824	0.068	ns
Mt. Elgon	KFOL42	10	0.860	1.000	ns
Mt. Elgon	KFOL47	28	70.581	0.000	***
Mt. Elgon	KFOL48	10	4.024	0.946	ns
Mt. Elgon	KFOL2	6	9.930	0.128	ns
Mt. Elgon	KFOL37	105	102.765	0.543	ns
Mt. Elgon	KFOL30	36	22.782	0.958	ns
Mt. Elgon	KFOL17	91	109.382	0.092	ns
Mt. Elgon	KFOL13	15	4.457	0.996	ns
Baringo	KFOL24	78	88.994	0.185	ns
Baringo	KFOL28	28	56.292	0.001	**
Baringo	KFOL42	6	30.104	0.000	***
Baringo	KFOL47	66	221.895	0.000	***
Baringo	KFOL48	10	22.891	0.011	*
Baringo	KFOL2	10	2.081	0.996	ns
Baringo	KFOL37	120	128.879	0.273	ns
Baringo	KFOL30	78	119.290	0.002	**
Baringo	KFOL17	78	110.628	0.009	**
Baringo	KFOL13	36	57.235	0.014	*
Mau	KFOL24	45	52.151	0.216	ns
Mau	KFOL28	10	20.078	0.029	*
Mau	KFOL42	3	0.036	0.998	ns
Mau	KFOL47	36	53.777	0.029	*
Mau	KFOL48	3	11.302	0.010	*
Mau	KFOL2	10	9.392	0.495	ns
Mau	KFOL37	66	73.538	0.245	ns
Mau	KFOL30	66	59.636	0.696	ns
Mau	KFOL17	91	74.271	0.899	ns
Mau	KFOL13	36	74.537	0.000	***
Gachuthi	KFOL24	1	3.000	0.083	ns
Gachuthi	KFOL28	Monomorphic			

Gachuthi	KFOL42	1	0.009	0.923	ns
Gachuthi	KFOL47	15	15.542	0.413	ns
Gachuthi	KFOL48	10	15.462	0.116	ns
Gachuthi	KFOL2	6	20.813	0.002	**
Gachuthi	KFOL37	1	3.200	0.074	ns
Gachuthi	KFOL30	10	31.226	0.001	***
Gachuthi	KFOL17	45	52.425	0.208	ns
Gachuthi	KFOL13	3	0.166	0.983	ns
Embu	KFOL24	10	5.792	0.832	ns
Embu	KFOL28	Monomorphic			
Embu	KFOL42	3	30.009	0.000	***
Embu	KFOL47	66	132.388	0.000	***
Embu	KFOL48	3	0.646	0.886	ns
Embu	KFOL2	6	3.392	0.758	ns
Embu	KFOL37	6	4.901	0.557	ns
Embu	KFOL30	28	60.889	0.000	***
Embu	KFOL17	153	185.600	0.037	*
Embu	KFOL13	1	0.153	0.696	ns
Meru	KFOL24	78	121.025	0.001	**
Meru	KFOL28	10	38.033	0.000	***
Meru	KFOL42	6	46.986	0.000	***
Meru	KFOL47	36	85.063	0.000	***
Meru	KFOL48	21	62.384	0.000	***
Meru	KFOL2	6	3.618	0.728	ns
Meru	KFOL37	55	112.186	0.000	***
Meru	KFOL30	21	42.621	0.004	**
Meru	KFOL17	136	135.200	0.503	ns
Meru	KFOL13	15	57.000	0.000	***
Kibwezi	KFOL24	21	94.928	0.000	***
Kibwezi	KFOL28	10	60.122	0.000	***
Kibwezi	KFOL42	3	30.370	0.000	***
Kibwezi	KFOL47	66	96.747	0.008	**
Kibwezi	KFOL48	28	50.639	0.005	**
Kibwezi	KFOL2	3	2.542	0.468	ns
Kibwezi	KFOL37	66	103.745	0.002	**
Kibwezi	KFOL30	15	17.904	0.268	ns
Kibwezi	KFOL17	120	152.911	0.023	*
Kibwezi	KFOL13	15	91.200	0.000	***
Kitui	KFOL24	1	3.481	0.062	ns
Kitui	KFOL28	1	0.036	0.850	ns
Kitui	KFOL42	Monomorphic			
Kitui	KFOL47	45	39.155	0.717	ns
Kitui	KFOL48	15	6.895	0.960	ns
Kitui	KFOL2	6	14.148	0.028	*
Kitui	KFOL37	6	3.173	0.787	ns
Kitui	KFOL30	10	12.742	0.238	ns

Kitui	KFOL17	136	205.696	0.000	***
Kitui	KFOL13	3	11.731	0.008	**
Makueni	KFOL24	55	90.863	0.002	**
Makueni	KFOL28	10	22.797	0.012	*
Makueni	KFOL42	6	1.025	0.985	ns
Makueni	KFOL47	15	15.177	0.439	ns
Makueni	KFOL48	6	41.402	0.000	***
Makueni	KFOL2	1	0.627	0.428	ns
Makueni	KFOL37	66	107.041	0.001	**
Makueni	KFOL30	45	59.647	0.071	ns
Makueni	KFOL17	136	125.413	0.732	ns
Makueni	KFOL13	66	114.803	0.000	***

Key: ns=not significant, * P<0.05, ** P<0.01, *** P<0.001

Appendix V: Similarity Report

University of Eldoret

Certificate of Plagiarism Check for Thesis

Author Name	Otieno John Ochieng AGR/PBG/001/15
Course of Study	Type here...
Name of Guide	Type here...
Department	Type here...
Acceptable Maximum Limit	Type here...
Submitted By	titustoo@uoeld.ac.ke
Paper Title	DEVELOPMENT OF MICROSATELLITE MARKERS AND ANALYSIS OF GENETIC DIVERSITY AND POPULATION STRUCTURE OF SANDALWOOD (Osyris lanceolata Hochst. & Steud.) IN KENYA
Similarity	15%
Paper ID	1665403
Submission Date	2024-04-19 12:28:50

Signature of Student

University of Eldoret
19 APR 2024
UNIVERSITY LIBRARIAN
P. O. Box 1125, ELDORET-30100

Signature of Guide

Head of the Department

University Librarian

Director of Post Graduate Studies

* This report has been generated by DrillBit Anti-Plagiarism Software