Molecular identification and evaluation of genetic diversity of Kenyan *Dioscorea* species using universal DNA barcodes of *rbcL* and *matK* regions

Nicholas Cheptarus Kipkiror

Reg. Number: I56/88873/2016

A thesis submitted in partial fulfillment for the award of the Degree of Master of Science in Bioinformatics in the Centre for Biotechnology and Bioinformatics (CEBIB) of the University of Nairobi

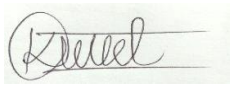
May, 2021

DECLARATION

This thesis is my original work and has not been presented for a degree in any university

Nicholas Kipkiror

Registration No: I56/88873/2016

Signature:  Date: …….18th May, 2021……

This thesis has been submitted for examination with our approval as university supervisors:

Dr. Edward Muge

Department of Biochemistry

University of Nairobi

Signature:  Date: ……18th May, 2021……

Dr. Evans Nyaboga

Department of Biochemistry

University of Nairobi

Signature:  Date: ……18th May, 2021……

DEDICATION

I wholeheartedly dedicate this work to my parents Mr. and Mrs. James Kiptoo and my siblings.

ACKNOWLEDGEMENT

I am sincerely grateful to my supervisors Dr. Edward Muge and Dr. Evans Nyaboga for their mentorship, guidance and invaluable support offered during the research process. I also express my greatest appreciation to CEBIB staff led by Dr. George Obiero, Dr. Rosaline Macharia and Dr. Bernard Kulohoma for their continuous support and encouragement during the entire academic period. I extend my gratitude to Anne Owiti, Abraham Parwos, Edwin Rono, Rosemary Mutindi and Tiberius Aboge for the professional support offered. Special thanks to my classmates for the inspiration, motivation and the shared experience. I am indebted to you all.

Thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

6-PGD:	6-Phosphogluconate dehydrogenase
AAT:	Alpha-1 antitrypsin
AFLP:	Amplified Fragment Length Polymorphism
BLAST:	Basic Local Alignment Search Tool
CBOL:	Consortium for the Barcode of Life
cpDNA:	Chloroplast DNA
CTAB:	Cetyl trimethylammonium Bromide
EDTA:	Ethylene diaminetetraacetic acid
FAO:	Food Agriculture Organization
IITA:	International Institute of Tropical Agriculture
INDELS:	Insertion and Deletion
ITS2:	Internal Transcribed Spacer 2
matK:	Maturase K
ML:	Maximum Likelihood
PCR:	Polymerase Chain Reaction
PtDNA:	Plastid DNA
RAPDs	Random Amplification of Polymorphic DNA
rbcL:	Ribulose 1, 5-bisphosphate carboxylase/oxygenase
SCARS	Sequence Characterized Amplified Region
SDS	Sodium Dodecyl Sulfate
SKDH:	Shikimate 5-dehydrogenase
SNP	Single Nucleotide Polymorphism
SSRs	Simple Sequence Repeats
trnK:	Lysine tRNA

ABSTRACT

Yams (*Dioscorea* species) are tuber crops that are grown in tropical regions of Africa, the Caribbean, South America, Asia and South Pacific islands. In Kenya, yams are grown in western, coastal and central highland regions as an important food security crop. The crop has also gained commercial interest due to the potent medicinal/pharmacological properties as well as industrial use. However, many *Dioscorea* species have similar morphology and easily confused and consequently their phylogenetic relationships is poorly understood, which hinders their full utilization and genetic conservation. Therefore, the aim of this study was to identify and evaluate the genetic diversity of *Dioscorea* species cultivated in Kenya using ribulose 1, 5-bisphosphate carboxylase/oxygenase (*rbcL)* and Maturase K (*matK)* barcoding markers. Leaf/tuber samples of 22 yam accessions were collected from farmers' fields and the National Genebank of Kenya (GBK) followed by DNA extraction and PCR amplifications using *rbcL* and *matK* barcoding markers. The purified PCR products were bidirectionally sequenced and sequences generated assembled in Geneious Prime 2019.2. Sequences alignment was conducted on Muscle version 3.8., and subsequent analysis done using PhyML package implemented in Geneious prime 2019.2, BEAST version 1.10, DnaSP version 6.12.03 and Arlequin 3.5.2.2. Phylogenetic analysis of *rbcL* and *matK* sequences revealed four strongly supported distinct species that included *Dioscorea bulbifera, Dioscorea alata, Dioscorea minutiflora* and *Dioscorea cayenensis.* The yam accessions explored were maximally differentiated into two distinct clades. The specific clade of each of the yam accession was informed by its geographic location and species. DNA polymorphism in the yam species exhibited both synonymous and non-synonymous mutation. *RbcL* sequences had nucleotide diversity and average number of nucleotide differences of 0.00392 and 2.668, respectively. *MatK* gene sequences had a nucleotide diversity of 0.00632 and average number of nucleotide differences of 4.87273. There were seven haplotypes within the *rbcL* gene with a diversity index of 0.800 and variation of 0.00374. There were four haplotypes within the *matK* gene with a diversity index of 0.745 and variation of 0.00956. Tajima neutrality indicated that *rbcL* gene of the yam accessions is under sweep selection while *matK* gene is under balanced selection suggesting that the population is growing. The findings from this study demonstrate that *rbcL* and *matK* barcoding markers can identify and differentiate *Dioscorea* species as well as establish their phylogenetic relationships. These results provide a scientific basis for molecular identification and discrimination of *Dioscorea* species which is useful in the efficient utilization of the genetic resources as well as in designing their breeding strategies.

CHAPTER ONE

INTRODUCTION

1.1 Background to the study

The tropical regions of Africa, South America, the Caribbean, Asia and South Pacific islands, are home to yams (*Dioscorea* species) cultivation (Andres *et al*., 2016). In Africa, yam is the second most important tuber crop after cassava and therefore constitutes a major source of starch for the growing population where food security is continuously becoming a challenge. Globally, West Africa is the largest growing zone for yams with Nigeria, Ghana, Côte d'Ivoire, Benin, and Togo being the main growing belt (Fu *et al*., 2011). Annually, approximately 90% of the entire global yam production is grown in this region in about four million hectares with production of fifty million tons (Asala and Ebukiba, 2016). In Kenya, yams are grown mainly in three regions namely western, coastal and central highlands (Kariuki, 2012).

Yams are important source of nutritional supplement. They also serves to provide medicinal/pharmacological and industrial applications to consumers (Mignouna *et al*., 2008). Yams contain proteins and micronutrients that are important in improving human health; vitamin D and potassium are available in *Dioscorea rotundata* (Obidiegwu *et al.*, 2020). The crop also contains thiocyanate, which is an agent counteracting sickle cell anemia in humans (Cornago *et al*., 2011). Bitter yams (*Dioscorea dumetorum*) are prospective in lowering blood sugar levels thus used to treat diabetes. It also provides treatment to conditions such as stomach colic, rheumatoid arthritis, menstrual disorders and *Schistosomiasis* (Sonibare *et al.*, 2010). In the industry, yams can be processed to starch or chips which can be converted to yam flour (Andriamparany *et al*., 2014). Yam consumers consider taste, the color of flesh after cooking and texture as some of their criteria for consumption. Other preferences considered include the price, origin and size of the yam tuber (Barlagne *et al*., 2017).

Yams are angiosperms belonging to the monocotyledonous order *Dioscoreales;* however, they express a vestigial cotyledon that places them at intermediate

phylogenetic relationships between monocotyledon and dicotyledonous plants (Mignouna *et al*., 2008). They are herbaceous plants with aerial storage called bulbil and with a perennial season characterized by growth cycle. Its genetic improvement has been affected by a long growth cycle (approximately 8 months), non-uniform flowering of male and female, dioecy, no flowering and sometimes poor flowering. The dioecy feature of yams limits their efficiency with regard to breeding (Tamiru *et al*., 2017). The other factors that impede the improvement of yams genetics are polyploidy, vegetative propagation methods and inadequate knowledge on their genetic diversity (Balogun and Gueye, 2013).

There are more than six hundred species of yams with only six of them being socioeconomically beneficial with regards to food, income and medication (Verter and Bečvářová, 2015). These species include white yam (*D. rotundata*), water yam (*D. alata*), air potato (*D. bulbifera*), yellow yam (*D. cayenensis*), bitter yam *(D. dumetorum)* and Chinese yam (*Dioscorea esculant*) (Fig. 1.1). However, distinguishing these species and mapping their location is largely based on morphological features (Anokye *et al*., 2014). Use of molecular markers especially chloroplast and mitochondrial markers can help distinguish the yam species and identify mutations over an evolutionary time.

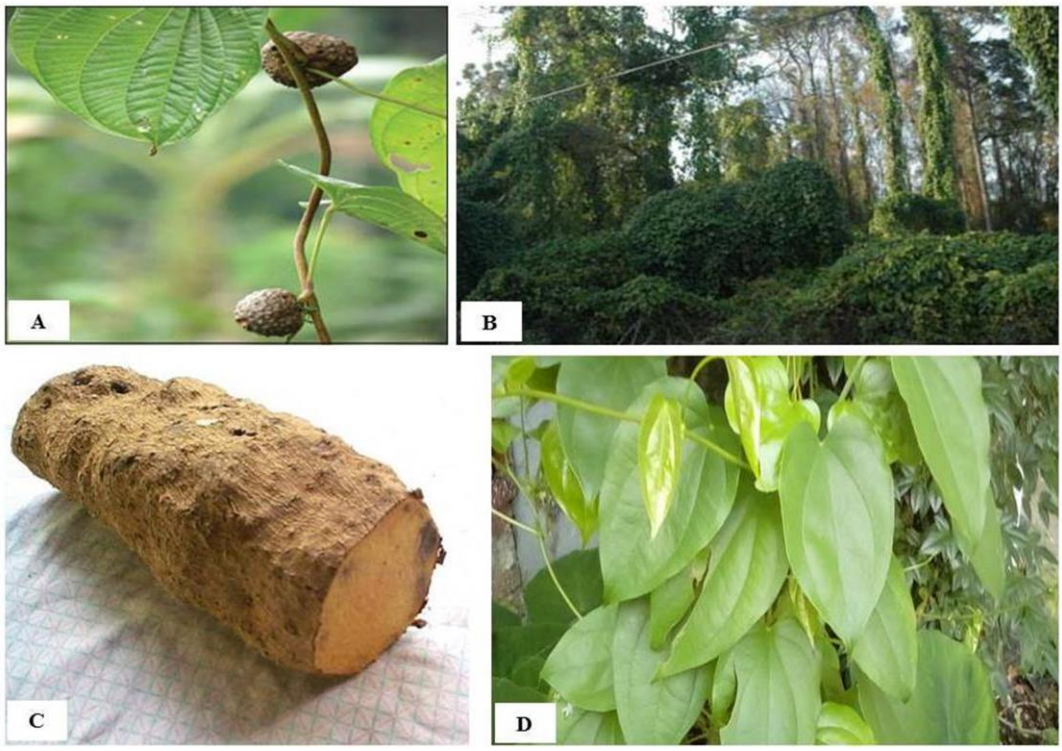


Figure 1.1: *Dioscorea* species

A: Represents a leaf and stem of *D. bulbifera*; **B**: Represents *D. alata* plantation; **C**: Shows *D. cayenensis* tuber; and **D**: Represents *D. rotundata* leaves (Adapted from Global Invasive Species Database).

In order to improve the breeding systems and conservation of yams, its genetic diversity should be accounted (Banson and Danso, 2013). The use of DNA barcoding markers for identification and characterization of yams can be a useful tool because it is not subject to agro-climatic variations or environmental variations. It can also be able to provide requisite information and more insight on yam plant evolution (Deschamps *et al*., 2012). Therefore, DNA barcoding markers that are universal and have minimum evolution rates over time are considered suitable for elucidating the identity and characterization of yams or any other plant. Maturase K (*matK)* and ribulose 1, 5-bisphosphate carboxylase/oxygenase (*rbcL)* markers are considered candidate primers for identification and characterization of yam species because they are universal, and can provide better resolution/discrimination among species (Amit,

2014). It has also been reported that about 92% of the plants can be distinguished using these markers. Therefore, the use of DNA barcodes to ascertain plant species, mutations over evolutionary time and map their location is easy as it can be applied across species (Ngwe *et al*., 2015). In Kenya, there is limited information on the identity and characterization of cultivated *Dioscorea* species, useful information which is of importance for the preservation of their biodiversity. Therefore, the current study utilized *rbcL* and *matK* barcoding markers in the identification of species of *Dioscorea* cultivated in Kenya. In addition, the markers were used to resolve phylogenetic relationships among *Dioscorea* species as well as elucidating their evolutionary and taxonomic relationships.

1.2 Problem statement

The development of biodiversity resources requires that native plant species should be distinctively identified for their conservation (Ngwe *et al*., 2015). Depending exclusively on morphological characteristics of yam plants for their taxonomic classification is challenging due to the effect of heritable variations and growth environment. More so, it requires taxonomists with great expertise and skills to classify them. The propagation method of yams in the field makes it susceptible to pathogens and thus hinders accurate identification and characterization of the different species. Pests attack and propagation methods contribute to genetic loss of yam species. It is estimated that microbial rots cause about 15-40% of the crop lose after harvesting (Wumbei *et al*., 2019). The loss of yam diversity leads to loss of food to the population that depend on the crop (Wilkin *et al*., 2018). In Kenya there is limited molecular data on cultivated yam species, important information on the preservation of its biodiversity. Consequently, there is inadequate knowledge/information on diversity of yam species cultivated in Kenya thus impeding their conservation. Understanding the current status of yam species and their genetic diversity in Kenya can help in their full utilization and genetic resource conservation.

1.3 Justification

Yam is an important crop for consumption, industrial and medicinal purposes. Proper identification of the *Dioscorea* species will provide better understanding of the existing cultivars and enable the preservation of the crop which is experiencing genetic loss/erosion resulting from the vegetative propagation methods, desertification, drought and abandonment by farmers for other crops. DNA barcoding is used for identification as it can provide better resolution/discrimination among species (Amit, 2014). Equally, reports indicate that about 92% of the plants can be distinguished using DNA barcodes which can resolve the intergenic relatedness among species as it can be universally applied across species (Ngwe *et al*., 2015). For the preservation of yam species biodiversity, adequate information on their genetic diversity is fundamental for their full utilization and conservation of the genetic resource. Breeding of the species relies much on information on its genetic diversity. Being a neglected crop in Kenya, genetic diversity of yams needs to be elucidated for its conservation. Interestingly, the wild yam (*D. quadrata)* is the only species whose genetic diversity is known and conserved at the National Genebank of Kenya together with other yams whose species have not been elucidated. Phylogenetic analysis using chloroplast DNA markers *rbcL* and *matK* will help identify and discriminate the available yam species in the country. This is important for the preservation of the yam species from genetic loss/erosion and by extension their extinction as well designing their breeding strategies.

1.4 Hypothesis

i. There is no taxonomic classification of yam accessions cultivated in Kenya.
ii. There are no inter- and intra-species relationships of yam species cultivated in Kenya.

1.5 Objectives

1.5.1 Overall objective

To establish the inter- and intra-species relationships of yam species cultivated in Kenya using *rbcL* and *matK* markers.

1.5.2 Specific objectives

i. To determine the taxonomic classification of yam accessions cultivated in Kenya using *rbcL* and *matK* markers.

ii. To determine the inter- and intra-species genetic variation of yam accessions grown in Kenya using *rbcL* and *matK* markers.

CHAPTER TWO

LITERATURE REVIEW

2.1 Yam plant

2.1.1 Yam habitation

Yam is an edible tuber crop that is herbaceous, twining, climbing monocot belonging to the family *Dioscoreaceae* and genius *Dioscorea (*Washaya *et al*., 2016). It is mainly grown in the tropical regions of the world where the weather conditions are favorable for their cultivation. The favorable ecological conditions for cultivation of the yam plant are warm climate with temperatures ranging from 25^{o}C to 30^{o}C and sunny weather (Barton, 2014). It is important that the plant is grown at the start of the rainy season as it requires an annual rainfall of about 1500 mm. The soil conditions that is compatible for the growth of yam plants is loamy, deep and loose in nature (Mignouna *et al*., 2009).

The crop can either be grown underground or above the soil surface/aerial (bulbil), however, majority of those that are grown are underground tubers and being dioecious plants, they may produce flowers (Gruère *et al*., 2006). Depending on the species, yam vines which grows either clockwise or anticlockwise may grow up to thirty meters long. The size of the tuber length growing in the soil can be up to 1.5 meters long while its weight can be over 70 kilograms. Their leaves can grow approximately 26 centimeters wide and long while its petiole may be shorter than leaf blade. The identifications and classification of yam species is majorly dependent on their variations in tubers and bulbils (Salawu *et al*., 2014).

2.1.2 Cultivation of yams

Both traditional and modern propagation methods are employed in the cultivation of yam plants (Aighewi *et al*., 2015). Traditionally, the yam seed is a tuber cutting/portion whose main component is a starch deposit that lacks dormant buds which are prominent among other tuber crops. Yam farmers ordinarily use tubers from the previous harvest for their next cultivation. This is occasioned by inadequate

supply of quality seeds which is a predisposing factor to poor yields as some of the seed tubers used by the farmers could be infested by parasites (nematodes) and/or infected by bacteria, fungi and viruses (Amusa *et al*., 2003; Osunde, 2008). Loss in yield is majorly contributed by poor quality of seeds that have been compromised by parasites and microorganisms. The traditional yam tuber propagation is characterized by low multiplication ratio, prolonged sprouting period and conspicuously low production. Some of the propagation methods used include: double harvesting/'milking', Anambra system (mostly in West Africa), cuttings/cut setts and sorted small tubers from the ware crop. Some of these modes of propagation provide an opportunity for pathogen dissemination and somewhat reduces the chances of emergence of new varieties as well as adaptation of the plant to environment (Dansi *et al*., 2013). In spite of the challenges that affect this cluster of methods, it still sustains the production of yams as they are able to provide seeds and some of the methods can be altered to suit different yam plants. Some of the modern methods include: mini-sets and micro-tubers technique, organ and tissue culture, vine cuttings, aeroponics system, botanic seeds and the use of temporary immersion bioreactor system for seed production (Eyitayo *et al*., 2010; Etoo *et al*., 2016). The modern methods have high multiplication ratio but they are limited by the high costs required, specialized equipment and skilled personnel needed for their development and effective application (Ondo *et al*., 2016).

2.1.3 Benefits of yams

Yam crop has benefits ranging from nutritional, social and economic purposes in areas where it is grown. In terms of nutritional value, yams contain vitamins (predominantly vitamin C), rich in carbohydrates (starch), proteins and minerals (Dabonne *et al*., 2011). It is estimated that the carbohydrate content ranges from 18.30% to 26.84%, while the moisture content ranges from 62.61% to 73.15%. Both parameters do not have significant variations among species. The yam starch which accounts for 60-80% of the tuber's dry matter is significantly responsible for rheological, physiochemical and textural characteristics of food products that are

derived from starch (Otegbayo *et al*., 2014). The fat content of yams ranges from 0.18% to 0.06% whereas crude proteins levels range from 1.30% to 1.91% (Kulasinghe *et al*., 2018). Some of the minerals present include magnesium (Mg), sodium (Na), calcium (Ca), iron (Fe), manganese (Mn) and cobalt (Co). These minerals play an important role in supplementing the nutritional needs of the consumers and to the plant itself. The β-carotene is also found in yams and it is a precursor to pro-vitamin A. Other carotenoids with important health and nutritional effects include zeaxanthin and lutein compounds which are useful in eye health and antioxidant activities (Price *et al*., 2018).

Yams also contain other important compounds such as riboflavin, phosphorus, thiamine and niacin (Udensi *et al*., 2010). Traces of tannin are also found in yam species *D. dumetorum* and are responsible for anti-rotting properties (Okwu and Ndu, 2006). Yams also contain alkaloids and saponins that are beneficial to the plant due to their toxicity properties. These compounds are responsible for the bitter taste (Adebowale *et al*., 2018). The crop contains compounds that have pharmacological properties which include saponins, dioscorine and sapogenins (Mulualem *et al*., 2018). Saponin characteristically acts as antibiotics as it is able to protect against invasion by microbes. Other medicinal potentialities of yam plant have been exploited across the globe. In China, *D. althenoides* is administered orally for the alleviation of indigestion, treat injuries and as a curative to chronic rheumatism while *D. bulbifera* matured fruits decoction are used to relieve gastric bleeding and their leaves act as antiseptic for open wounds (Ji *et al*., 2004). In Malasyia, leaves of *D. hispida* are sometimes used to treat skin conditions while mucilage from *D. laurifolia* tubers is used to cure warts. The treatments of fever and asthmatic conditions have been undertaken using boiled tubers of *D. membranacea (*Maneenoon *et al*., 2008). Yam presents an economic value to farmers that grow it as they are able to sell the crop and earn a living while socially it denotes a status, prestige and association in the society where the crop is treasured (Obidiegwu and Akpabio, 2017).

2.1.4 Challenges facing yam cultivation

Despite the benefits that accrue to yam farmers through their domestication, cultivation of the crop is faced by challenges. The challenges vary from propagation to harvesting and storage (Etoo, 2017). Soil infertility, breeding systems, pest attack in the field and perpetuated infections (fungal, viral and bacterial), are some of the main constraints to yam domestication (Frossard *et al*., 2017). Fertility of the soil is fundamental in the overall production of the yam plant that includes the nutritional status of the tuber. Breeding is presented as the main challenge that affects yam production in most of the regions that domesticate the crop (Arnau *et al*., 2010). This is because breeding is limited by germplasm and their characterization. Poor and /or lack of flowering of yams has constrained its rapid multiplication and development. However, application of molecular integrated technologies have helped improve breeding in spite of the constraints realized (Arnau *et al*., 2016). Use of microsatellite markers, AFLPs, RAPDs and SCAR markers have been employed in the evaluation of genetic diversity of yams with a view of improving its breeding however, optimization of the breeding system is yet to be achieved.

Whilst most of the challenges facing the cultivation of yam crop are within its growth period, post harvesting loses are also present (Wumbei *et al*., 2019). It is estimated that microbial tuber rot causes about 15 - 40% of the crop loss after harvesting. Some of the species (like “laribako” yam variety domesticated in Ghana) are vulnerable to pathogens thus forcing farmers to sell at a loss because of its short shelf life (Kusi *et al*., 2013). Other challenges include inadequate labor, weed pressure and price fluctuations. The use of herbicides also presents a challenge to farmers as they lack adequate knowledge on how to handle the chemicals properly thus causing undue exposure that can be of higher levels to that recommended. Unsafe handling of the pesticides leads to contamination of the tubers with residues that pose a high risk to human health upon consumption. Incorrect use of the pesticides can also cause direct exposure to the farmers which in turn pose other health related issues such as irritation and respiratory infections.

2.2 Global distribution of yam production

Yams are grown mainly for consumption and as a cash crop. It is fundementally ideal for food security as it is consumed by over 155 million people in the world (Sangakkara *et al*., 2014). West Africa is the leading producer of the crop with a total production of over sixty two (62) tonnes representing approximately 90% of the global yam production. Its production increased from 900,000 hectares in 1961 to 7 million hectares in 2014 with the main species grown being *D. alata* and *D. rotundata* (FAOSTAT, 2016). Yam production is also done in Latin America and South East Asia countries but at a relatively small scale (<10%). Presumably, yam production is confined to Africa (Fig. 2.1). There are about ten yam species that are widely domesticated in the world but the species that are majorly cultivated in West Africa include *D. rotundata*, *D. cayenensis* and *D. alata* which account for the over 90% of the total yam production in the world (Lopez-Montes *et al*., 2012). Nonetheless, this has prompted global agricultural research institutes such as IITA to focus on the development of new yam varieties that have desired qualities and agronomic traits which can improve the global yam production.

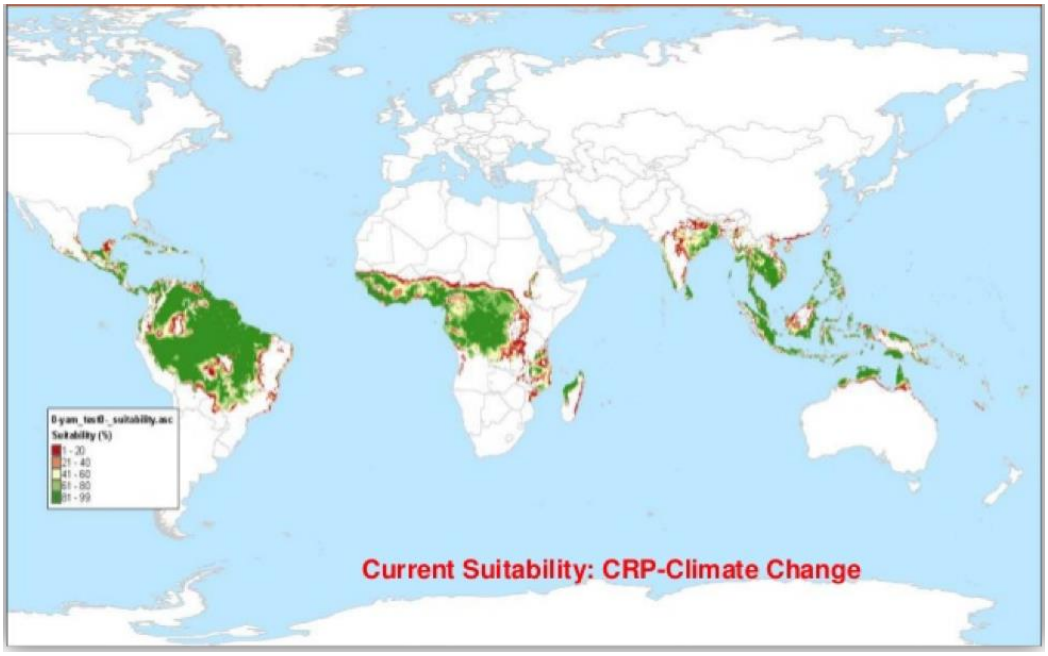


Figure 2.1: Global yam production zones
(Adapted from: International Institute of Tropical Agriculture, 2016)
Key: The Green shaded sections of the map are Yam producing regions in the world. The zones are tropical regions whose climatic conditions are favorable for the growing of yam plant.

2.3 Yams cultivation in Kenya

Available information indicates that yams species grown in Kenya are *D. rotundata*, *D. minutiflora*, *D. bulbifera*, and *D. dumetorum* (FAO, 2009). It is mainly grown for food in the Central, Eastern, Coastal and Western regions by the elderly farmers. The crop is locally referred as Ikwa or Viazi Vikuu in the dominant growing regions. Over thirteen (13) yam cultivars have been distinguished by the farmers based on their morphological characteristics which include leaf (shape, color), stem (width, texture and color), tuber (flesh color, size, shape and taste) and maturity period. Other regions like Meru South and Embu have reported the growing of a unique aerial tuber yam type. Beside the domestic yam species in Kenya, the country's wild yam species is represented by *Dioscorea quadrata* (Wambugu, 2009).

In the most recent times, the country's yam genetic diversity has been affected by genetic loss/erosion caused by drought, gross destruction, changes in the natural

habitat and neglect by farmers to continue growing the crop (Bressan *et al*., 2014). *D. minutiflora* was the mostly cultivated yam species but it has been abandoned for other crops and only grown in small scale by the elderly. *D. quadrata* is the only species whose genetic diversity is known and conserved at the National Genebank of Kenya with other yams whose species identity have not been elucidated. As a result of genetic loss, it is imperative to identify the existing yam species grown in farmers' field for their efficient utilization and conservation (Wambugu, 2009)

2.4 Genetic diversity of yams

Genetic variability in yams is high at intra- and inter-species level (Rivera-Jiménez *et al*., 2011a). This is responsible for the high genetic diversity among the yam species (more than 600 species). The genetic variability is responsible for the morphological difference in domestic and wild yams, as well as difference in isozymes and taxonomy (Sangakkara *et al*., 2014). Important information about genetic diversity can be obtained through phylogenetic analysis of the yam species. However, evaluating the yams diversity is challenging because of taxonomic complexity and ploidy variation across species (Mignouna *et al*., 2002). Moreover, ploidy variability biasness is difficult to ascertain due to poor knowledge on chromosome segregation pattern.

Molecular markers have helped detect the differences in yam varieties that are otherwise considered similar in terms of morphology and in isozyme markers (Asahina *et al*., 2010). The markers are also able to detect polymorphism pattern among species. Similarly, they are also able to identify species and or region that are at a risk of extinction and provide congruence on species concept that is important for standard biodiversity (Nixon and Wheeler, 1990). This is important in the conservation of yam biodiversity given that there are a number of factors which include pest attack, propagation methods and climate changes that are most likely to affect their diversity leading to species loss or extinction (Moritz and Potter, 2013). Phylogenetic analysis based on the chloroplast DNA markers can help distinguish

and enumerate the available *Dioscorea* species (Scarcelli *et al*., 2011). Moreover, Single nucleotide polymorphisms (SNPs) have been widely used for genotyping because of their abundance in the genome as they are able to provide quantity of ancestry in their relative proportions for the currently domesticated genotypes (Deschamps *et al*., 2012).

2.5 Chloroplast Genome

The yam chloroplast genome (Fig. 2.2), encodes one hundred and twelve (112) unique genes with which seventy nine (79) of these genes codes for proteins (Zhao *et al*., 2018). There are four ribosomal RNA (rRNA) genes and twenty nine (29) genes that code for transfer RNA (tRNA). The large single copy (LSC) region is composed of twenty two (22) tRNA genes and sixty two (62) genes that codes for proteins while the small single copy (SSC) region comprises twelve (12) protein coding genes and one (1) gene that codes for tRNA. Some of the genes that are duplicated in the inverted regions (IR) (A and B) are six protein coding genes, four rRNA genes and eight tRNAs. There are also seventeen (17) genes that contain introns,twelve of which code for proteins while five (5) code for tRNA. Both *ycf3* and *clpP* have two introns each whereas the rest of protein coding regions together with tRNA coding region have one intron. The trans-splicing gene *rps12* has the first exon in the large single copy region and the second and third exons in the inverted regions.

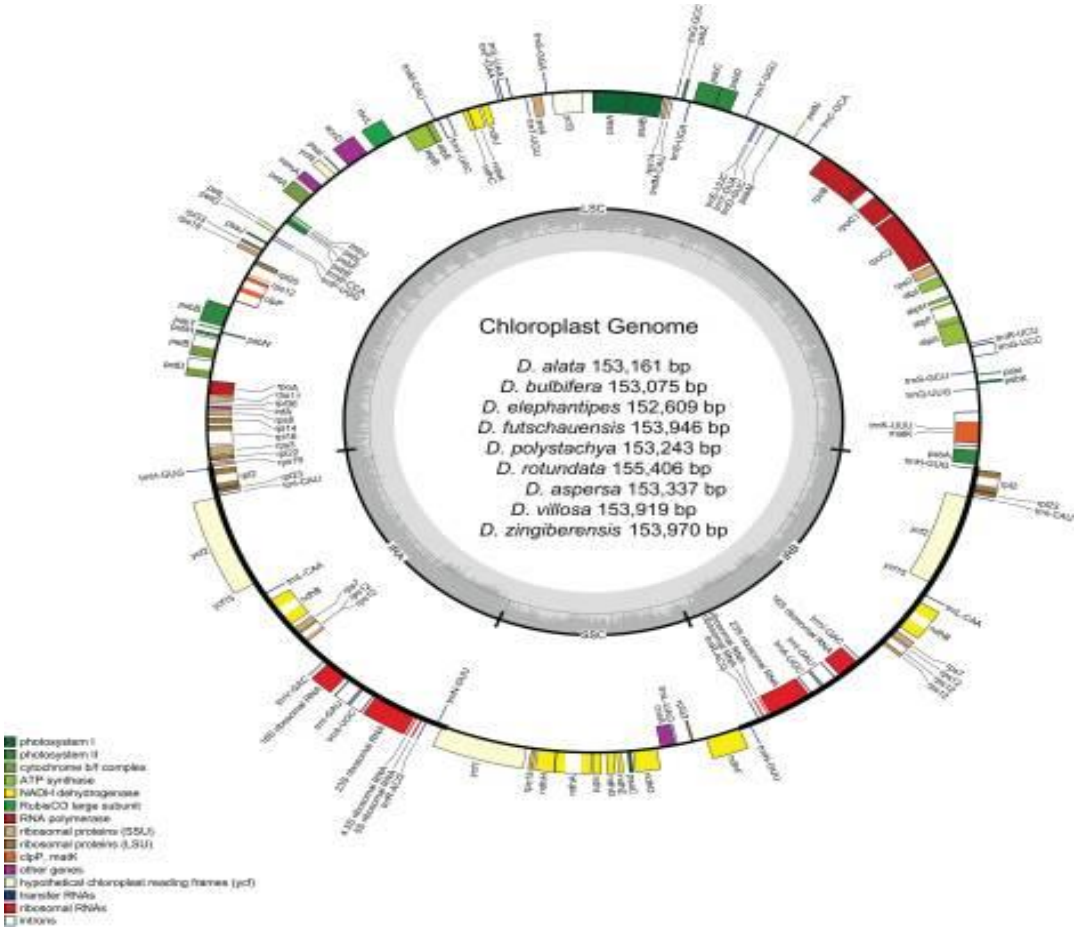


Figure 2.2: *Dioscorea* chloroplast genome
The genes are color coded with reference to their function. Genes on the outside of larger circle are transcribed in anti-clockwise manner while those inside the larger circle are transcribed in clockwise manner. GC composition is represented by the dashed region.
Adapted from Zhao *et al.* (2018). For better resolution, visit: https://www.ncbi.nlm.nih.gov/core/lw/2.0/html/tileshop_pmc/tileshop_pmc_inline.html?title=Click%20on%20image%20to%20zoom&p=PMC3&id=6284424_peerj-06-6032-g001.jpg

The transcription of the genes is based on their orientation in the larger circle. Genes oriented inside the larger circle are transcribed in a clockwise manner while those that are outside the larger circle are transcribed in an anticlockwise manner. Genes are colored with reference to their functions.

The plastome length of fourteen (14) African *Dioscorea* species varies from 151,908 base pairs (*D. burkilliana*) to 155,155 base pairs (*D. baya*) while its genetic structure is typically quadripartite (Magwé-Tindo *et al.*, 2019). Two inverted regions

(IR) separate their LSC and SSC regions. Its average GC content is 37%, however, inverted region has its GC at 43% while the SSC regions GC content is 31%. The genetic structure that is exhibited has one hundred and thirty (130) genes that correspond to eighty four (84) protein-coding genes, eight (8) rRNA genes and thirty eight (38) tRNA genes. Eighteen (18) of these genes are duplicated in the inverted region and the distribution is as follows: Eight tRNA, six protein-coding genes and four rRNA. *D. quartiniana* is an exception as it missed the *matK* gene as a result of mutation of its second codon to a stop codon. The *rps*16 gene is lost in all the fourteen species.

2.6 Ribulose 1, 5-bisphosphate carboxylase/oxygenase (*rbcL*) and Maturase K (*matK*) markers

Plant phylogenetic analysis and genetic diversity discrimination are based on chloroplast DNA (cpDNA) (Amit-Roy, 2014). Chloroplast DNA is assumed to be conserved during its evolutionary period in reference to its minimum nucleotide substitution. Currently, chloroplast DNA genes include *rbcL, ndhF, rpl16, matK* and a*tpB.* The *rbcL* and *matK* are the plastid DNA genes that have been recommended as the standard barcode markers by the Consortium for the Barcode of Life (CBOL) of the Plant Working group. This is due to their availability, and their universal primers and their high level of taxonomic resolution (Herrmann and Wink, 2014). It is estimated that nearly 92% of the plant species can be distinguished using the standard *rbcL* and *matK* barcode with the cognizance that they are of moderate dispersion with regard to their phylogeny (Burgess *et al*., 2011). The overall species discrimination success for *rbcL*+*matK*+trnH-psbA is 66.7% (Laiou *et al*., 2013). Whilst there is 100% amplification and sequencing success in *rbcL*+ *trnH-psbA* for sixty eight (68) trees from twenty four (24) taxa, *matK* didn't amplify in some plants thus recording inconsistency. In spite of the inconsistency in *matK* marker amplification, CBOL still recommends it for use in species discrimination because of its variability. Therefore, *rbcL* + *matK* markers can provide a reliable resolution of

species identity that is acceptable and which is cost effective and with minimal laboratory effort.

Ribulose 1, 5-bisphosphate carboxylase/oxygenase (rubisco) is the first enzyme of C3 cycle in plants and due to its abundance, it is the most important protein in the plant cycle (Kress and Erickson, 2007). *RbcL* gene is a single copy gene of 1428 base pairs long located in the chloroplast genome. It is universal in all plants and hence easy to align and convenient to study. Maturase K (*matK*) is positioned within the intron of lysine tRNA (trnK) chloroplast gene and has about 1500 base pairs. It is responsible for encoding maturase enzyme involved in splicing type II introns from RNA transcripts (Sun *et al*., 2012). In flowering plants, it is useful in the resolution of their intergeneric relatedness, therefore can be used for phylogenetic discrimination of plant species (Men *et al*., 2017). *RbcL* and *matK* are therefore the most commonly used DNA barcoding markers (Asahina *et al*., 2010) as they can provide discrimination of high level among species.

Girma *et al*. (2015) reported that the combination of both *rbcL* and *matK* had an optimal efficiency of 76.2% as DNA barcode as it identified sixteen species out of the twenty one yam plants under study. *RbcL* has a good record with regards to amplification and the quality of sequence, but has poor species discrimination power at 47.6% identification ability. *MatK* identifies species to 63.2% level and additionally detects inter-specific variations. The use of combined *rbcL* + *matK* markers as standard DNA barcode for yam species identification and classifications is reliable and acceptable. However, Li *et al*. (2014) explains that *rbcL* and *matK* markers are highly conserved and both universal but fails to discriminate species. These markers registered a poor resolution of 6.25% which is attributable to their low sequence variation. Although the markers are good for species identification, they failed in distinguishing the *Calligonum* species with shrubs found in arid and semi-arid regions of Europe, Africa and Asia.

Three barcode markers *rbcL*, *matK* and Internal Transcribed Spacer 2 (ITS2) are able to discriminate species at ~44%, ~81% and ~72% respectively of vascular plants in

Canada (Braukmann *et al*., 2017). At the genus level, the markers are able to correctly identify the plants as follows 91% (*rbcL*), 97% (ITS2) and 98% (*matK).* These markers, however, have the least discrimination ability in assigning the plant taxa to their biogeographic location. Their resolution ranges from 36% for *rbcL* to 69% in *matK*. The resolution in the woodland region plants is 52% for *rbcL* and 87% in *matK*. From this information on genus identification and species discrimination, it can be inferred that barcode markers are effective in these two functions. The difference in the species resolution by the markers is due to selection, phylogeography structure, hybridization, lineage sorting and species demography (Naciri and Linder, 2015).

RbcL has a strong purifying selection in autotrophic plants which reduces its evolution rate and ability to distinguish related species (Wicke *et al*., 2011). Conversely, *matK* is under relaxed purifying selection and has a high rate of nucleotide substitution therefore better ability in species discrimination (Duffy *et al*., 2009). However, *matK* lack conserved priming sites which largely affects its sequences recovery. Moreover, nuclear markers can be considered as barcode markers because of their high rate of nucleotide substitution which helps them evolve rapidly (Drouin *et al*., 2008). Additionally, multiple alleles among the nuclear markers make them useful for identification of hybrids. It is only ITS2 marker that has met the criterion for inclusion as a barcode marker among the nuclear markers (Naciri *et al*., 2012).

MatK gene evolution is always with low homoplasy which explains its significance of being close to neutrality in regard to selection (Müller *et al*., 2006). *RbcL* and other barcode markers have relatively higher homoplasy compared to *matK* which exhibits negative maximum sites to homoplasy. Higher homoplasy can be correlated to lower performance of a marker in the distinguishing of species. This is because high homoplastic sites translate to more non-informative sites, high number of randomized sites in the informative sites and low phylogenetic structure.

The identification of three divergent species; *Solanum nigrum*, *Euphorbia helioscopia* and *Dalbergia sissoo* using *rbcL* and *matK* markers revealed conserved genome information (Wattoo *et al*., 2016). This information is useful in profiling the medicinal potential of some of the plants of these species as each of them represents their family. Provision of such information on species studied confirms that barcode primers can accurately identify and discriminate plants into their respective genus and species level. Moreover, the authentication of species by the markers can provide an opportunity to discover new traits among the species.

Whilst discriminating *Dipterocarpaceae* family, the sequence success rate for the barcode markers is 83.7%, 81% and 54% for *rbcL*, *matK* and *trnL-F*, respectively (Carneiro *et al*., 2019). *MatK* proved to be the most efficient in the authentication at species level. The genetic distances of the respective marker sequences are 0.020 (*matK)*, 0.026 (*trnL-F*) and 0.017 (*rbcL*). Therefore, these markers are able to provide resolution to the identification of plants thus reliable tools for DNA barcoding and evolutionary inventories.

Techen *et al*. (2017) reported that there is a possibility that LFY genomic region is a useful as a molecular marker for distinguishing *Dioscorea* species. This region in the gene is about 300 base pairs with sequence repeats ACCCT that can be inserted or deleted in either position of the gene. This deletion/insertion repeats provides variation in the gene which is useful in species discrimination. The nuclear marker also exhibits 100% sequence success thus translating to 100% species authentication and discrimination.

The success of species identification is dependent on the database exhaustiveness that is in use. This is because missing species cannot be identified (Liu *et al*., 2011). DNA barcoding being a bioinformatics and molecular tool is used to identify species and discover new ones (DeSalle *et al*., 2005). This makes it an invaluable tool for discriminating species and detecting errors in the process of species authentication as they are able to provide credible information on intra- and inter-specific regions (Fatima *et al*., 2019). Moreover, the combination of these barcode markers can

produce an optimal evolutionary tree that can provide insights on the species co-existence and relatedness (Kang *et al*., 2017).

2.7 Polymorphism and DNA divergence in yam species

Genetic variability among yam species can be linked to breeding efforts which ultimately determines their population structure as easily demonstrated by gene flow (Siadjeu *et al*, 2018). In order to explicitly determine the breeding system of yams, its genetic diversity must be ascertained correctly. This information on diversity informs breeding which eventually guides the improvement, utilization and conservation of yam genetic resources.

RbcL in yams is the most conserved chloroplast gene with reference to the number of conserved sites and its sequence length, while *matK* has the greatest nucleotide variations based on length of its sequence and the number of variable sites present (Sun *et al*., 2012). The numbers of nucleotides that are conserved in *rbcL* are 522 of 553 nucleotides while the variations in *matK* are 110 of 752. *M*at*K* has 81 of 752 parsimony informative sites while *rbcL* has 31 of 553 sites. This implies that *matK* is the best DNA barcode for the reconstruction of a phylogenetic tree for the *Dioscorea* species (Lahaye *et al*., 2008). *Mat*K (0.0095±0.0167) has the maximum intra-specific variation while *rbc*L (0.0019±0.0045) has the minimum. Similarly, Wilcoxon signed rank tests shows that the inter-specific divergence of *mat*K (0.0295±0.0249) is higher than that of *rbcL* (0.0125±0.0076) (Sun *et al*., 2008).

The relationships of guinea yam germplasm in Benin based on forty one (41) simple sequence repeats (SSRs) indicated that it has thirteen (13) polymorphic sites and one hundred and thirteen (113) polymorphic alleles, while the number of alleles in each loci ranges from 4 to 13 with an average of 8.69 alleles per loci (Loko *et al*., 2016). The genetic distances between the yam cultivars has a range of between 0.45 and 0.04 while the polymorphic information content is 0.76. This indicates that yam cultivars have high genetic diversity as depicted by 0.78 heterozygote value. The yam landraces has a 96% variation within its population and 4% between its

populations. This information on genetic diversity can inform the process of yam plant genetic improvement.

On using the AFLP molecular technique, Rivera-Jiménez *et al*. (2011b) while determining the genetic variation distribution at inter specific and intra-specific level, confirmed that there is a correspondence between morphological characteristics and molecular basis of a yam species. The study indicates that *D. alata* is genetically closer to *D. rotundata* than to *D. esculenta*. The genetic distance between *D. alata* and *D. rotundata* is at 41.84% while that of *D. esculenta* and *D. trifida* is at 33.51%. With the availability of such information, the origin of each species can be traced while the formulation of breeding strategies can be inferred and executed accordingly.

The genetic diversity of forty African yam accessions had 95.2% polymorphism upon evaluation using AFLPs (Ojuederie *et al*., 2014). The average number of amplified polymorphic bands in each primer is 411 and ranges from 360 to 520. The polymorphism information content ranges 0.9447 to 0.9626 while their similarity was between 0.66 and 0.911. Pencil yam (*Vigna lanceolata*) has 138 alleles with an average of 7.67 alleles per locus. Its overall genetic diversity was 0.62 and the polymorphism information content per marker varies between 0.06 and 0.90 with a mean of 0.61 (Nubankoh *et al*., 2015). These findings confirm the robustness of AFLPs in genetic variation evaluations. The province of such incredible information can be a lead to selection of progeny with desirable phenotypic characteristics.

Muthamia *et al*. (2013) obtained a total of one hundred and thirty one (131) alleles amplified with a minimum of two (2) alleles and thirteen (13) alleles in Kenyan yam accessions using simple sequence repeats (SSR). These alleles have a minimum of 64 base pairs and a maximum of 368 base pairs. The accessions from the Eastern Kenya are unique while those from the coastal region have no unique alleles. Polymorphic percentage of the accessions from the eastern Kenya is 76.3% while that of West Africa is 32.8%. Shannon index is 0.1444 for West African accessions while that of Central Kenya cultivars is 0.2366. Variation of the cultivars is

significant at 88.0% within population and distinctively clustered into four groups regardless of their geographical pattern thus indicating their relatedness.

Akakpo *et al.* (2017) opines that the diversity in wild type yam was approximately 30% more important than that in cultivated African yam, *D. rotundata*. Evidence of selection was only found in two percent of the genes studied. Photosynthesis and phototropism are affected by selection due to changes in adaptation from wild plant to a domesticated plant. Cultivated yam Tajima's D is skewed to a positive value of 0.77 (mean). Tajima's D distribution of two wild type yams is centered on zero reflecting a global equilibrium relationship between SNP occurrence and their frequencies. The average differentiation between *D. rotundata* and *D. praehensilis* is higher than between *D. rotundata* and *D. abyssinica,* (F_{ST} = 0.21 and 0.16, respectively, *p*-value <0.001). There are at least two SNPs that are putatively under selection in two hundred and thirty eight (238) contigs which are retained as candidates.

The characterization of yam species grown in Nigeria using morphological traits and isozyme markers shows that there are two clusters that are domesticated by farmers (Efisue, 2015). Cluster I comprise of species *D. abyssica*, *D. schimperana*, *D. alata* and *D. dumetorum* while cluster II consists of *D. cayenesis* and *D. rotundata*. The three isozymes that ably discriminates species are SKDH, 6-PGD and AAT. The information provided by the isozymes showed that *D. alata* is distantly related to both *D. cayenesis* and *D. rotundata* while *D. cayenesis* is closely related to *D. rotundata*. This indicates a high genetic diversity and high genetic similarity, respectively, of the species used in the study. The two mega-clusters have a coefficient of similarity of 0.32 and 0.50 with coefficient above 0.60 having decreased species number. In spite of the two clusters present, there are sub-clusters that existed within each cluster which have no significant differences. AAT isozyme marker is the best discriminant with a polymorphism of 50.0% – 96.4%

The genetic diversity of the dominant yam species in Ethiopia is 0.36 and has a corresponding Shannon's index of 0.53 (Ousmael *et al*., 2019). These yam species

assessed includes *D. alata, D. bulbifera*, and *D. cayenensis/rotundata* complex. Their polymorphism at genus level is at 97.4%. Individual species polymorphism is rated at 92.2% for *D. cayenensis/ rotundata)* and 71.4% for *D. bulbifera*. The genetic diversity of *D. cayenensis/ rotundata* complex is 0.33 whereas its Shannon's index is 0.49 while *D. bulbifera* has a genetic diversity of 0.24 and Shannon's index of 0.47. The variation between populations is high at 63.9% while that among population is at 36.1%. Correspondingly, there is high variation within population than between populations in *D. cayenensis/ rotundata* complex at 53.6% and 46.4% respectively. From the findings it can be inferred that the genetic diversity of yam species in Ethiopia is high which can be fundamentally important for studying their genetic breeding system with the view of its improvement.

The nucleotide diversity of nine yams species with 800 base pairs is estimated to range from 0 to 0.0175 in clade A (*D. zingiberensis*, *D. villosa*, and *D. futschauensis*) while clade B (*D. rotundata*, *D. bulbifera*, *D. aspersa*, *D. alata*, *D. elephantipesn* and *D. polystachya)* has a range of 0 to 0.02533 (Zhao *et al*., 2018). The mean nucleotide diversity in clade A is 0.00334 while that of clade B has a mean diversity of 0.00926. Whilst the nucleotide diversity informs the yam variations, eight regions in the genome that could be used as potential markers for the identification of the plant genus are found in both the LSC and SSC segments. Five of these potential markers are located at the LSC while three of them are located in the SSC region. Six of the potential markers are located in the intergenic region which include trnK-trnQ, trnS-trnG, trnC-petN, trnE-trnT, petG-trnW-trnP, and trnL-rpl32 while the other two are located in the coding regions and include ycf1 and ndhF (Rogalski *et al*., 2015).
The Shannon indexes of the predominant yam species domesticated in Brazil have nearly similar values to each other at 0.23 and 0.27. The *D. rotundata* has an index of 0.23 while *D. cayenensis* has a value of 0.27 (Silva *et al*., 2017). The variation between the yam growing regions of Brazil is significant at 50.9%, despite a significant total variation that is attributable to variation within regions that is at 49.1%. Mantel test reveals that these yam species have a high and significant association between geographic and genetic distances for morphological (r=0.68;

P=< 0.01) and molecular markers (r=0.64; P =<0.01). There is also a high and significant association between molecular genetic distances and morphological characteristics (r=0.56; P=<0.01). Polymorphism averages at 56.8% while their information content ranges from 0.34 to 0.70. The genetic variation is significantly in similar proportion between species (49.4%; P =< 0.01) and within species (50.6%; P=< 0.01).

The diversity of twenty one (21) yam landraces cultivated in China has a varied polymorphism of 95.3% when evaluated using inter simple sequence repeat (ISSR) and a corresponding 93.5% on sequence related amplified polymorphism (SRAP) markers (Wu *et al*., 2014). There is significant variation among yam population compared to within the populations, as depicted by 40.39% and 23.83% respectively. Consequently, their mean gene flow is 0.1081 that is a derivative of high genetic differentiation of 0.8222, suggesting that the gene flow is restricted within the population. The markers used, ably discriminates the species into respective clusters, therefore, providing useful information on the existent genetic diversity of yams in China.

The genetic coefficient of greater yam (*D. alata*) that is grown by farmers in China averages at 0.70 and has a range of between 0.42 and 0.91 (Wu *et al*., 2019). The population of this particular species is however clustered into two groups with a genetic coefficient of 0.62. Cluster I has 125 genotypes while Cluster II has 17 genotypes. The variation within the Chinese *D. alata* population is at 99.0% while that among population is a paltry 1%. In terms of genetic differentiation, there is no much significance among the yams in various geographical locations as it rated at 0.013. The ploidy of this population is 87.3% diploid and generally polyploid with numerous ploidy levels.

Cao *et al* (2018) observes that the comparison of six genotypes of *Dioscorea polystachya* chloroplast genome originating from different geographical regions in China have one hundred and forty one (141) SNPs. There are seventy (70) simple sequence repeats of which twenty four (24) are polymorphic in the six genotypes.

Complete genome study of this species reveals important and reliable molecular information that is fundamental for the elucidation of its biogeographic structure. It also reveals information on the variation within and between its populations. There are eighty eight (88) transitions compared to fifty three (53) transversions bringing the ratio to 1:0.6 (5:3) revealing a bias to transition mutation. With this variation, the frequency of SNP mutation from Cytosine to Thymine and from Guanine to Adenosine is high while that from Cytosine to Guanine and from guanine to cytosine is lowest. Moreover, this species has only five small inversions of between 2 to 51 base pairs in its chloroplast genome.

Elephant foot yam (*Amorphophallus paeoniifolius* (Dennst.)) that is domesticated in India has a polymorphism information content that ranges from 0.620 to 0.937, and is averaging at 0.803 (Santosa *et al*., 2017). Most of the allelic loci of the species are highly heterozygous (H_o>0.600). Each locus has an average of F_{IT} (0.249), F_{IS} (0.032) and F_{ST} (0.279) respectively. Majority of the loci produced very high F_{ST} values of between 0.201 and 0.476. The variation exhibited by the species is as follows: among populations (11%), among individuals within population (8%), among regions (16%) and among individuals across all populations (65%). Phylogenetic assessment indicates that species are clustered with respect to their cultivation regions/origin.

2.8 Phylogeography divergence

Phylogeography elucidates the spatial distribution of species in a given region, which includes species variation, selection and habitat and adaptation (Schaal *et al*., 1998). It may also present the species geographic latitude and time of divergence. With the above concepts on a species, its genetic structure and diversity within population can be understood. Information on species variation, evolution and distribution in a given habitat is important to its breeding improvement and safeguard against attack from pests and diseases (Seal *et al*., 2014).

Domestication of yam species in the regions' growing the crop can be traced to different sources (Pickersgill, 2007). These include wild plant, hybridization and seeds dispersal from one region to other. Yam domestication process signifies the end of wild plants and beginning of cultivation of the adopted plant that has to undergo several stages before it is fully domesticated. The process ends during selection of the crop to be domesticated by farmers which inform the next stage of formulating a criterion for differentiating the wild plants from the domestic ones as well as the intra-species and inter species distinction (Clement, 1999).

Yams have been domesticated in three different continents at least at three different times. These continents include Asia that domesticated *D. alata*, Africa that domesticated *D. rotundata* and America that domesticated D. *trifida* as their inception yam plant (Harlan, 1992). In Africa, the origin of yam domestication can be traced to River Niger belt, a region between western part of Nigeria and eastern region of Ghana. It is also important to recognize that the five crops which are sorghum, yam, cowpea, pearl millet and African rice, domesticated in Africa have their origin in Niger belt (www.fao.org/faostat). On the yam plant, *D. rotundata* domestication in Africa can be attributed to be have been contributed by either or both savannah *D. abyssinica* and the species from the forest *D. praehensilis* (Fig. 2.3). Other yam species that have been domesticated are likely to be derivatives of either of the two yam species or a product of hybridization between *D. abyssinica* and *D. praehensilis (*Magwé-Tindo *et al.*, 2018). Scarcelli *et al* (2019) while using principal component analysis alludes that, the best traced origin of yam domestication is northern Benin with latitude 11.1° ± 4.4°N and longitude 2.8° ± 8.9°E (Fig. 2.4).

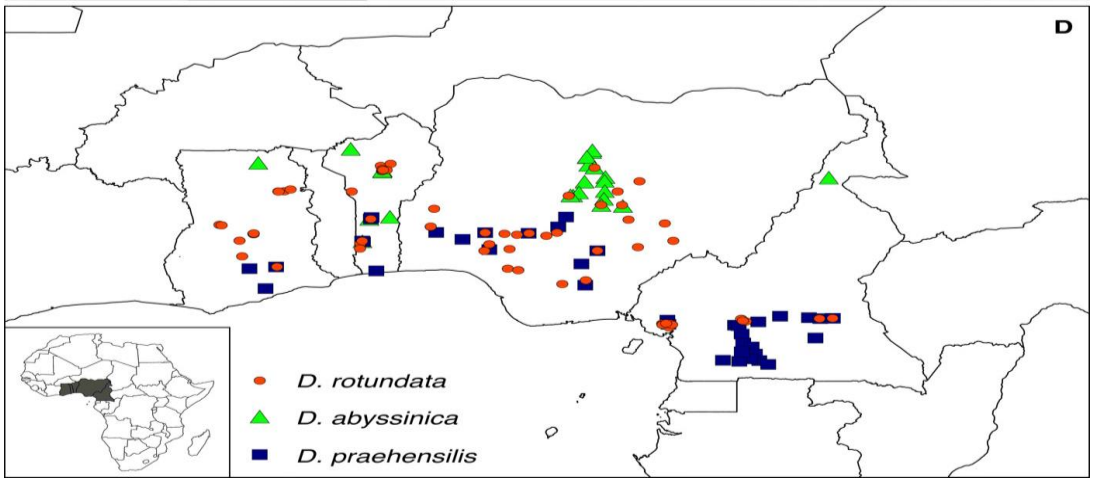


Figure 2.3: The origin of Africa's yam plant
Map representing the origin and domestication of yam species in Africa. Orange color represents *D. rotundata*, Green color represents *D. abyssinica* and Blue represents *D. praehensilis.* Domestication of Africa *D. rotundata* can be linked to both *D. abyssinica and D. praehensilis* (Scarcelli *et al*., 2019**).**

The divergence of the yam species from its cradle of domestication is that the forest progeny, *D. praehensilis*, split into two main clusters that were geographically distinct. The split led to their domestication in Cameroon and another one in West Africa while another cluster was *D. abyssinica* while the fourth cluster is *D. rotundata.*

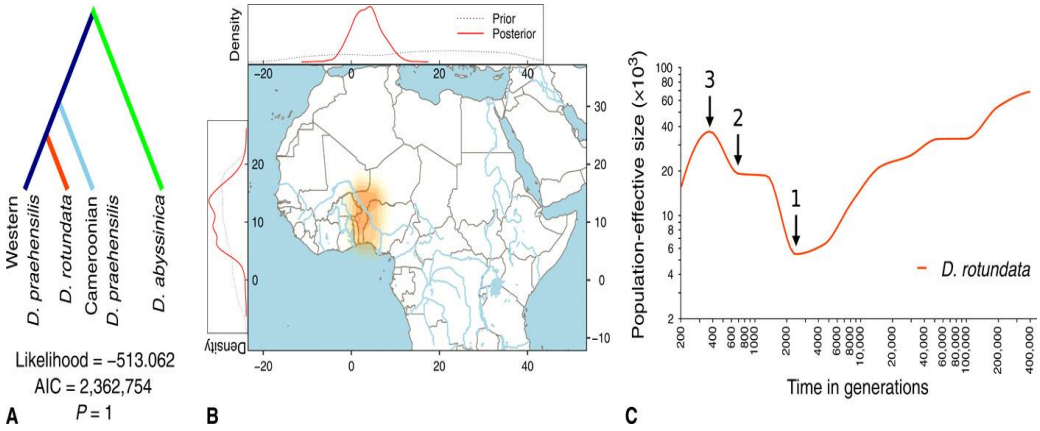


Figure 2.4: History of Yam Domestication in Africa
A - Represent the relationship of yam species inferred through coalescent-based analysis. **B** - Represents the geographical origin of cultivated yam species based on an approximate Bayesian spatial model. **C** - Represent the demographic history of cultivated yam populations. The arrows (1, 2 and 3) represents the generations of expansion and decrease in yam cultivation. Arrow 1 is shows the expansion in period ca. 2500, arrow 2 indicates the expansion in period ca. 700 while ca. 400 denotes the latest decrease in yam cultivation (Scarcelli *et al*., 2019)

Africa's history of yam domestication indicates that about fifteen thousand (15,000) generations ago, the cultivated and the wild populations declined (Terhorst *et al*, 2017). At Ca 2,500 generations ago, the cultivated population was at its minimum whilst their size was divided into four in comparison to their ancestral population size. Thereafter, in two phases, the cultivated population increased its size to a maximum level at Ca 400 generation ago. In the first phase of population increase, a robust increase at fourfold level occurred between Ca 2,500 and Ca 1,500 generation among the cultivated populations. The second phase of population increase occurred between Ca. 700 and Ca. 400 generations ago and the increase doubled the populations. It is notable that in the last Ca 400 generations ago, the cultivated population declined by half. In the last Ca. 300 generations, the western *D. praehensilis* exhibited similar population trend as the cultivated populations which was declining.

Arnau *et al*. (2017) indicates that the genetic diversity of *D. alata* is 0.66 which demonstrates a moderate to high polymorphism level. It similarly demonstrates that there are two genepools that are divergent in India and Vanuatu which provides a clear genetic differentiation between yam species in Asia and South Pacific which both have secondary diversification. The tetraploid cultivars were from numerous geographic locations demonstrating that *D. alata* grown in Africa and America, is genetically similar to that grown in South Pacific and Asia where it was domesticated in about 6000 years ago.

Since Eocene, Neotropical *Discorea* has four main origins as indicated by its calibrated tree (Courto *et al*., 2018). The greatest divergent lineage of yams originated in the period between Eocene and Oligocene in Southern Andes region. This is also present in the areas of Central America, Atlantic forest and Northern Andes. It is estimated that at the boundary of Cretaceous–Palaeocene which is 63.7Mya of period 52.6 –77.0 Mya, is the origin of the most diverse genus of the family *Dioscorea*. Similarly, it is at Palaeocene that nearly all the divergence at gene level occurred while the two Neotropical clades originated at period Eocene and Oligocene. The crown age of NWI is 31.2 Mya (23.6–39.2 Mya) and NWII of *c.* 28.2 Mya (19.5–37.2 Mya) originated during this period while NWIII that comprise of four species, originated at *c.* 17.7 Mya (11.53–24.15 Mya) which is Oligocene period. Finally, at Miocene, 14.3 Mya (5.0–25.2 Mya), the fourth cluster of yam that include Neotropical *D. mollis* and Malagasy *D. antaly,* diverged.

The Brazilian yam cultivars of *D. cayenensis* and *D. rotundata* are spatially grown in two distinct regions in the country (Silva *et al*., 2017). The species are spatially separated into two distinct groups which are cluster I and II. Cluster I cultivars originated from the southeast region and they belong to *D. cayenensis* species. Cluster II cultivars occur predominantly in the Northeast region and they belong to *D. rotundata*. The cultivars from the south region belonged to either species. The Brazilian yam spatial distribution, exhibits low geographic mobility as a result of the

farmers living in the same locality for more than twenty years. It is also possible that these species were distinctively introduced to the farmers in their respective regions.

2.9 Haplotypes

The genus *Dioscorea* is divided into several sections into which the species that belong therein are clustered (Oli, 2006). Enantiophyllum is one of the largest sections of this genus that is comprised of *D. cayenensis, D. opposita, D. alata, D. transversa, D. rotundata* and *D. japonica*. When viewed (physically) upwards from the ground, the members of this section are characterized by vines that twine in a clockwise manner. Whilst members of *Enantiophyllum* twine clockwise, those that belong to section *Lasiophyton* twine to the left. Some of the *Lasiophyton* species include *D. hispida* and *D. dumetorum*. The other sections include *Combilium* (*D. esculenta*), *Opsophyton* (*D. bulbifera*) and *Macrogynodium* (*D. trifida*). These sections though exist as distinct, evolutionary changes occur as some of the species are domesticated which may result to genetic alteration, therefore producing diversity in the original species. Most of the yam haplotypes present have originated in accordance to evolutionary changes that have occurred in these sections, with some being shared haplotypes.

The evolution of yam species genotypes by Chair *et al* (2005) characterizes the *Discorea* species into four distinct groups with six haplotypes present. Haplotype one (H1) consists of cultivars *D. abyssinica* or *D. praehensilis* which together with cultivars in Haplotype two (H2), belong to group *Enantiophyllum.* The cultivars that belong to H2 includes *D. minutiflora, D. smilacifolia* and *D. burkilliana*. These species can be considered to be of the same genetic group because they diverged later than other species in this section of *Enantiophyllum.* Only *D. togoensis* belonged to the Haplotype three (H3) and was the last to diverge in the section *Enantiophyllum. D. abyssinica* of subset haplotype five (H5), distinctively diverged from haplotype one but had its alleles shared with *D. alata* of Haplotype six (H6). It can be inferred that it is the same ancestor that gave rise to Asian and South Pacific *D. alata* and *D. abyssinica* in Africa.

There is a close genetic resemblance between Florida air-potato (*D. bulbifera*) to specimens examined from China, than to those from Africa (Croxton *et al*., 2011). At least two introductions of invasive population that are established in Florida explains their low intraspecific genetic diversity. The haplotypes of invasive population of *D. bulbifera* identified in Florida are distinct but genetically similar. Therefore, this data informs that these populations have a minimum of two introductions. The *D. bulbifera* haplotypes in Africa are found in both West and East Africa; however, they are genetically distant to those from China and Florida. The African haplotypes are not geographically structured. This fact about haplotype genetic distances in the three continents is confirmed by analysis using both parsimonious network analysis and neighbor-joining analysis.

The African *D. rotundata* and *D. cayenensis* share haplotypes while morpho-type *abyssinica* has three haplotypes (Djedatin *et al*., 2006). One of the haplotypes of D. *abyssinica* is shared with the *D. cayenensis-rotundata* species complex and with *D. praehensilis* morphotypes. The sharing of haplotypes indicates that these cultivars may belong to the same species. Other *Dioscorea* species such as *D. smilacifolia, D. minutiflora* and *D. burkilliana* that are the resultant of variation in *Enantiophyllum* section, suggests that they belong to the same genetic group.

The minor allele frequency of *D. alata* has an average of 0.25 and varies from 0.05 to 0.5 (Agre *et al*., 2019). The expectation of its heterozygosity is that it averages at 0.34 and varies from 0.10 to 0.50; however, upon observation the heterozygosity varies from 0.03 to 0.69 and averages at 0.38 on evaluation of the SNP markers. The polymorphism content is 0.36 and the markers are unequally distributed in the 21 chromosomes with the highest density being in chromosome 5 and lowest in chromosome 11. The order of descending SNPs density is as follows; chromosome(s) 5, 4 and 8.

Guinea yams together with their wild relatives have an average of 8.6 alleles in each loci among the southern Ethiopian cultivars (Mengesha, *et al*., 2012). The genetic diversity among these yams is high as indicated by its expected heterozygosity that is

averaging at 64%. The wild yams are allele rich when compared to domesticated yams. Both cultivated and wild yams have significantly low mean F_{ST} which indicates that the diversity is within the population rather than between the populations. The high diversity, heterozygosity and low mean F_{ST} observed in this population indicates that there is likelihood that genetic drift has had a small impact on the population.

There is an existent rich diversity of yam plant in India with approximately fifty species that are grown in the east, west and northeastern regions of the country (Sheela *et al*., 2014). Of the nearly fifty yam species grown, *D. alata,* is the predominant cultivar. Whilst there is diversity of these species, genotype divergence is observed between the northeastern states and the southern states that can be defined as rich in biodiversity. Sixty four alleles with an average of nine alleles per locus are identified among the seven loci of the *D. alata* species. The identified alleles range from 103 to 317 base pairs. The smallest difference between the highest allele length and lowest allele length is 39. However, thirty eight (38) alleles are identified as unique and there are present in less than ten percent of the population with eleven and twenty seven found in the northeastern and southern populations respectively. The total frequency of the allelic population ranges from 0.016 to 0.492 while that of unique alleles ranges from 0.016 to 0.143.

CHAPTER THREE
MATERIALS AND METHODS

3.1 Plant materials

A total of 22 samples of morphologically distinct yam varieties were collected from major growing regions in Kenya as well as from the National Genebank of Kenya (Table 3.1). Thirteen yam tubers were obtained from central (Muranga, Kirinyaga and Nyeri) and upper eastern (Meru) regions, two young leaf samples from Rift valley (Molo-Nakuru) region and seven young leaf samples from the National Genebank of Kenya. The yam varieties collected from the field were highly cultivated by farmers but their species were unknown. The selected yam varieties were packaged in boxes and transported at room temperatures to the laboratory within eight hours of sampling then stored at -80^{o}C freezer. Selected leaf samples were packaged in zip lock bags, transported in cooler box and stored at -80^{o}C freezer.

Table 3.1: Samples of yam varieties used in the study

No	Variety Code	Variety Name	Locality
1	A_N	Amola	National Genebank of Kenya
2	B_N	Obiotungi	National Genebank of Kenya
3	25_N	TDr2579	National Genebank of Kenya
4	19_N	TDr0097	National Genebank of Kenya
5	06_N	TDr0060	National Genebank of Kenya
6	X1_N	TDr2436	National Genebank of Kenya
7	M_N	Makakkwa	National Genebank of Kenya
8	00E_Meru	00E_Meru/1	Meru
9	1E_Meru	1E_Meru/1	Meru
10	2E_Meru	2E_Meru/1	Meru
11	3E_Meru	3E_Meru/1	Meru
12	4E_Meru	4E_Meru/1	Meru
13	5E_Muranga	5E_Muranga/1	Muranga
14	6E_Muranga	6E_Muranga/1	Muranga
15	7E_Kirinyaga	7E_Kirinyaga/1	Kirinyaga
16	8E_Nyeri	8E_Nyeri/1	Nyeri
17	9E_Muranga	9E_Muranga/1	Muranga
18	10E_Kirinyaga	10E_Kirinyaga/1	Kirinyaga
19	11E_Nyeri	11E_Nyeri/1	Nyeri
20	12E_Nyeri	12E_Nyeri/1	Nyeri
21	M2_Molo	M2_Molo/1	Rift Valley (Molo)
22	M3_Molo	M3_Molo/1	Rift Valley (Molo)

3.2 Sample preparation

Yam tubers from the field were washed, peeled, cut into small pieces and ground in liquid nitrogen before storing at -80ºC. Leaf samples were thoroughly washed with distilled water and stored at -80ºC ready for genomic DNA extraction. To minimize cross contamination and infections, the preparations was carried out in a sterile environment.

3.2.1 DNA extraction

Genomic DNA extraction for all the yam cultivars was carried using the Cetyl Trimethylammonium Bromide (CTAB) method (Datukishvili *et al*., 2010) with some modifications. The DNA extraction protocol was modified accordingly as need

dictated (Abdel-Latif and Osman, 2017). The CTAB extraction buffer (100 ml) was prepared from the following components: 2 grams CTAB, 0.5 M EDTA, 10% SDS, 5 M NaCl, 1 M Trisaminomethane (Tris), 1% Polyvinylpyrrolidone and double distilled water. The modifications included grinding the leaf/tuber sample in 600 µl of CTAB buffer with 150 µl of 10% Sodium Dodecyl Sulfate (SDS) and centrifuging at 18659 G-force/14000 rpm (MIKRO 200R Microliter Centrifuge- Hettich, Tuttlingen, Germany) for ten minutes. Yam leaf/tuber of approximately 5 grams was allowed to thaw on ice for five minutes and ground in 600 µl of CTAB buffer with 150 µl of 10% SDS using sterilized mortar and pestle. The sample mixture was incubated at 65ºC for one hour with vigorous vortexing at quarter hour intervals. The contents were incubated at room temperature for 10 minutes, 1 µl of RNAse added to each sample and incubated at 37ºC for ten minutes. Centrifugation was done at 18659 G-force for ten minutes and supernatant extracted into a sterile Eppendorf tube. Two times extraction of the supernatant with 500 µl of Chloroform-Isoamyl alcohol (24:1) was undertaken at 18659 G-force for ten minutes at each step. The upper phase was then transferred to a sterile Eppendorf tube and 300 µl of ice cold isopropanol added then incubated overnight. The sample mixture was then centrifuged at 18659 G-force for ten minutes and the supernatant discarded while the pellet preserved. Each of the sample pellets was washed twice with 400 µl of 70% ethanol and centrifuged at 18659 G-force for five minutes. The pellet was air dried at room temperature for thirty minutes before each was suspended in 50 µl of nuclease free water. The suspended DNA was stored at -20ºC. The quality of the isolated DNA was ascertained using gel electrophoresis method (Kumari, 2012).

3.2.2 Polymerase chain reaction (PCR) analysis

Polymerase chain reaction (PCR) amplifications were performed with the use of an Applied Biosystems 96-Well Veriti Thermal Cycler (ThermoFisher Scientific, USA). Two primers targeting *matK* (www.barcoding.si.edu) and *rbcL* (Cuénoud *et al.*, 2002) genes were used. The sequences for each of the primers are:

*matK*_F: 5’CCTATCCATCTGGAAATCTT3’,

*matK*_R: 5’GTTCTAGCACAAGAAAGTCG3’,

*rbcL*_1_F: 5’ATGTCACCACAAACAGAAAC3’ and

*rbcL*_74R: 5’TCGCATGTACCTGCAGTAGC3’.

Primer optimizations were performed on gradient PCR at six different annealing temperatures. The annealing temperature that produced the band with desired quality and expected size was selected for use in both markers. The amplification was done in a final volume of 20 µl using Ampliqon Taq DNA Polymerase 2x Master Mix RED with 2 mM $MgCl_2$ final concentration PCR Mastemix (Ampliqon, Stenhuggervej 22, Denmark) that contained: Tris-HCl pH 8.5, $(NH4)_2S0_4$, 4 mM $MgCl_2$, 0.2% Tween® 20, 0.4 mM of each dNTP, Ampliqon Taq DNA polymerase and Inert red dye and stabilizer. Amplification was carried out under the following conditions: initial denaturation at 94°C for 5 minutes, then 35 cycles at 94°C denaturation for 30 seconds, 58°C annealing for 45 seconds and 72°C extension for 30 seconds, and the final extension at 72°C for 7 minutes (Maloukh *et al*., 2017) for *rbcL* gene while *matK* gene was amplified with same protocol as *rbcL* gene but with the exception of annealing temperature which was set at 48°C. As part of quality check and ascertaining contamination, nuclease free water (non-template negative control) was subjected to the aforementioned PCR protocol.

A 5 µl aliquot of the PCR product mixed with 3 µl of the bromophenol blue loading dye, was subjected to 1% (w/v) agarose gel electrophoresis stained with ethidium bromide (EtBr) (0.5 µg/ml). The agarose gel was prepared by dissolving 0.4 grams of agarose in 40 ml of 1× Tris-Borate/EDTA (TBE) buffer. The mixture was then heated on a hot plate until boiling, before cooled and then stained with 1 µl of 0.5 µg/ml EtBr. The gel mixture was then casted on an electrophoresis plate with combs inserted to create wells and allowed to solidify for thirty minutes. An aliquot of 5 µl of the amplicon and 3 µl of the bromophenol blue loading dye were mixed and loaded into wells of the prepared gel. Electrophoresis was run at 80 Volts for one hour and the bands visualized under UV Tran illuminator (BIORAD). This was done

by placing the gel onto trans-illuminator above the viewing window before putting down the hinged safety lid and switching on the illuminator for visualization. The results were evaluated for the presence of bands of the expected sizes - 710 base pairs for *rbcL* and approximately 830 base pairs for *matK*.

3.2.3 Sequencing and sequence assembly

Amplification products with the expected band sizes were purified using the GeneScript QuickClean DNA Gel extraction kit (Piscataway NJ, USA) as per the manufacturer's instructions. The cleaning procedure removes non-nucleic acid contaminants, unused dNTP's and inhibitors. This kit was chosen because it is able to achieve ninety percent (90%) recovery rate for DNA fragments ranging 100 - 3000 base pairs. Bidirectional Sanger sequencing was done by Macrogen Lab Standard-seq single using ABI 3730xl System-Big Dye v.3.1. The sequences generated were edited and assembled in Geneious Prime 2019.2 (https://www.geneious.com/) and viewed on Jalview version 2.11.0. Both *de novo* assembly and assembling by mapping to reference sequences was done. For *de novo* assembly, the border of each marker was selected based on the quality of the conservation sites set at 50% and above. Mapping to the respective reference genome sequences for both markers was done as the sequences were required for phylogeography, DNA polymorphism and divergence and haplotype analysis. *MatK* sequences were mapped to NC_039708.1:c3223-1664, NC_039835.1:c3205-1646, NC_024170.1:c6127-4568 and NC_039836.1:c3214-1655 references sequences while *rbcL* generated sequences were mapped to NC_039707.1:54667-56100, NC_039708.1:54453-55886, NC_039835.1:54325-55758 and NC_039836.1:54484-55917 sequences.

3.3 Sequence alignment and phylogenetic analysis

Sequence similarity of *de novo* assembled sequences was searched by BLASTn and sequences aligned using MUSCLE Version 3.8 (Edgar, 2004) implemented in Geneious Prime 2019.2. Phylogenetic evaluations were executed using Maximum

Likelihood (ML) technique for tree drawing with TN93 model (Tamura *et al*., 2013) for both markers. Because the two markers had different evolutionary rates, their sequence alignment was conducted separately before the tree was built. The sequences aligned on Muscle version 3.8 were exported and viewed on Jalview version 2.11.0.; then trimmed to a requisite conservatory quality of >50%. The resultant aligned sequences were then exported to Geneious Prime with which PhyML package had been installed. Phylogenetic tree reconstruction was undertaken at 1000 bootstrap value in TN93 model. The tree was visualized in an inbuilt geneious prime visualization system.

3.4 Phylogeographic divergence of yam lineages

Chloroplast DNA (cpDNA) *rbcL* and *matK* markers were used to examine the differentiation and phylogeographic history of the *Dioscorea* species grown in Kenya. Phylogeography patterns of *Dioscorea* species was done using Bayesian model implemented in Bayesian Evolutionary Analysis Sampling Tree (BEAST; Qin *et al*., 2013). BEAST version 1.10 which uses the posterior probability to estimate the divergence/differeniation of species over time was used for phylogeography analysis. Analysis was performed through ten million steps of Markov Chain Monte Carlo (MCMC) to build the tree posteriorly. Maximum clade credibility tree was annotated using Tree Annotator available in BEAST with 10% burn-in of all the trees bult to remove the non significant trees. Subsequently, the resultant annotated tree was visualized on FigTree version 1.4.4.

3.5 DNA polymorphisms

Sequence polymorphism of cpDNA markers (*rbcL* and *matK*) were screened in all the amplified samples as described by Besnard *et al*. (2007). DNA Sequence Polymorphism (DnaSP) version 6.12.03 was used to analyze polymorphism in the *rbcL* and *matK* genes of the sequenced samples. The sequences were first mapped to reference sequence, aligned on MUSCLE version 3.8 (Edgar, 2004), viewed and then trimmed on Jalview version 2.11.0 to a uniform sequence length. A permutation

approach was used to estimate the significance of sequence differences between and within the yam species. This involved the estimation of statistical pairwise nucleotide differences. The resultant single nucleotide polymorphisms (SNP) were categorized to be singleton and Parsimony informative sites while positions of the SNPs noted together with the nucleobase altered.

3.6 DNA divergence between populations

DNA divergence between yam populations was measured by computing their variance Pi, Dxy and Da as outlined by Jukes and Cantor algorithm implemented in DnaSP version 6.12.03 described by Kartavtsev (2011). The DNA divergence between yam populations was measured using Analysis of Molecular Variance (AMOVA) and significance tested at 1000 permutations (Excoffier *et al*., 2005).

3.7 Tajimas neutrality test

Neutrality test was conducted to ascertain the frequency of mutation among the yam species and to establish if there was any selection taking course using Tajima D, Fu &Li's D* and Fu & Li's F* indices. The statistical significance was calculated with a thousand permutations and pairwise differences used to estimate the molecular distances among sequences (Zhao and Gong, 2015). Its graph was computed using sliding window analysis with a length of 100 sites and step size of 25 bases.

3.8 Haplotype analysis

Haplotype analysis of yam genotypes was conducted to ascertain the network of the cpDNA haplotypes (Guan, 2014). This was done by constructing a cpDNA network using DnaSP version 6.12.03 which generated the haplotype data files. Arlequin 3.5.2.2 was used for haplotype inference by estimating allele frequencies at each loci. Analysis of Molecular Variance was also undertaken at 1000 permutations (Beaumont and Nichols, 1996; Excoffier *et al*., 2009). Significance for all the analysis was evaluated at 5% confidence level.

CHAPTER FOUR

RESULTS

4.1 Polymerase chain reaction amplification of *rbcL* and *matK* genes in yam species

Amplifications of *rbcL* and *matK* regions were performed on samples of yam varieties followed by sequencing of resultant amplicons. For all the samples, single amplicons of ~710 and 830 base pairs were obtained with the markers *rbcL* and *matK*, respectively. The PCR products were sequenced to confirm the identity/species of the yam varieties, define their differentiation times and establish their haplotype inference. A total of 31 good quality sequences were obtained from bidirectional sequencing, with 20 and 11 being for *rbcL* and *matK* regions, respectively. Samples of varieties 25MatK_N, 06MatK_N, MMatK_N and 10MatK_Kirinyaga were not considered in the analysis because of their quality which was below the set sensitivity on sequence similarity.

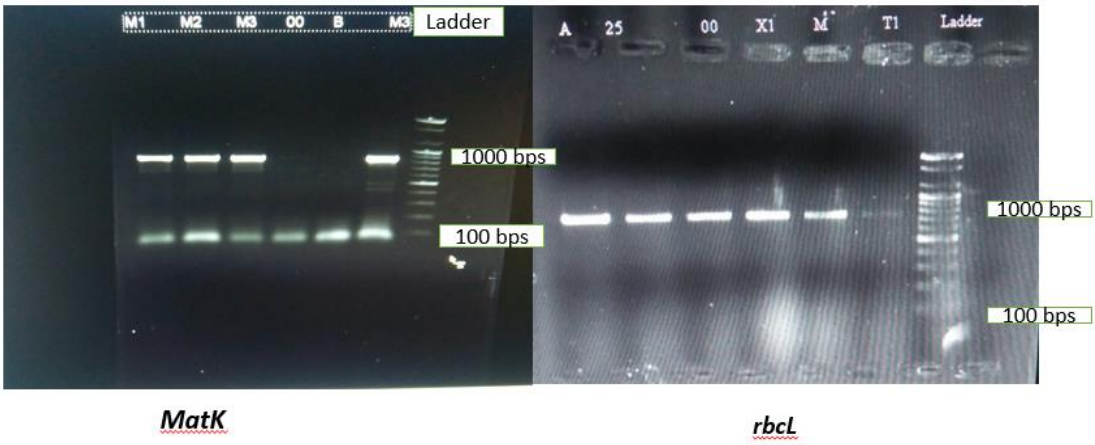


Figure 4.1: Gel image for *matK* and *rbcL* amplified DNA products represented on 100 bp molecular ladder (Quick-Load® Purple 100 bp DNA Ladder, New England Biolabs, Ipswich, MA, United States). The *matK* image represent the unpurified DNA products while the *rbcL* image respresent the purified DNA products. Expected *matK* and *rbcL* gene products were 830bps and 710 basepairs respectively. T1 indicates a purified product with low DNA concentration.

4.2 Sequence assembly and alignment

Multiple alignments of cleaned sequences produced 692 and 795 base pairs for *rbcL* and *matK*, respectively (Appendix 1 and 2). Upon alignment and subsequent viewing on ESPript 3, single nucleotide polymorphic sites were identified. The polymorphic sites were recognized to be both synonymous and non-synonymous mutations on annotation. These mutations were found in all the four *Dioscorea* species identified during the study. The codon position of the respective SNPs was also depicted.

4.3 Phylogenetic reconstruction

To ascertain the identity of the resultant consensus sequences, blasting against the NCBI database sequences was done and yam species identified based on percentage similarity and respective E values. All the blasted sequences had similarity percentage of 99.0% and above while their corresponding E-value was 0.0.

For *rbcL* marker, a well–resolved phylogeny supported by bootstrap values was obtained and revealed four strongly supported species-specific clades (Clade A, B, C and D) (Fig. 4.2). The species identified included *Dioscorea bulbifera, Dioscorea alata, D. minutiflora* and *D. cayenensis.* The tree was classified in an ascending order based on bootstrap values in each of the species, ancestral node. Clade A which represented *D. minutiflora* had a support value of 62.5 and the accessions in this clade were 11ErbcL_Nyeri, 10ErbcL_Kirinyaga, 8ErbcL_Nyeri, 7ErbcL_Kirinyaga, 5ErbcL_Muranga, 4ErbcL_Meru, 3ErbcL_Meru, 1ErbcL_Meru and 0ErbcL_Meru, all from central and eastern regions of Kenya. *D. minutiflora* identified represented 45.0% of the total yam accessions cultivated in Kenya. The relationship in this section was generally monophyletic. *D. cayenensis* (Clade B) was strongly supported by a bootstrap value of 80.2 and its accessions included 6ErbcL_Muranga, BrbcL_N, 25rbcL_N, ArbcL_N, X1rbcL_N, 19rbcL_N and MrbcL_N. *D. cayenensis* represented 35.0% of the yam species identified with only 5.0% being cultivated by farmers in the field. The relationship in this section was largely paraphyletic. *Dioscorea alata* (Clade C) had a support value of 88.5 with accessions being 06rbcL_N and 2ErbcL_Meru while its relationship was monophyletic. *D. alata*

represented 10.0% of the total yam species identified with 5.0% being that cultivated in the field. Clade D which was *Dioscorea bulbifera* had the highest support value of 98.5, a monophyletic relationship and its accessions were M3rbcL_Molo and M2rbcL_Molo. This represented a 10.0% of the yam species identified (Fig. 4.2).

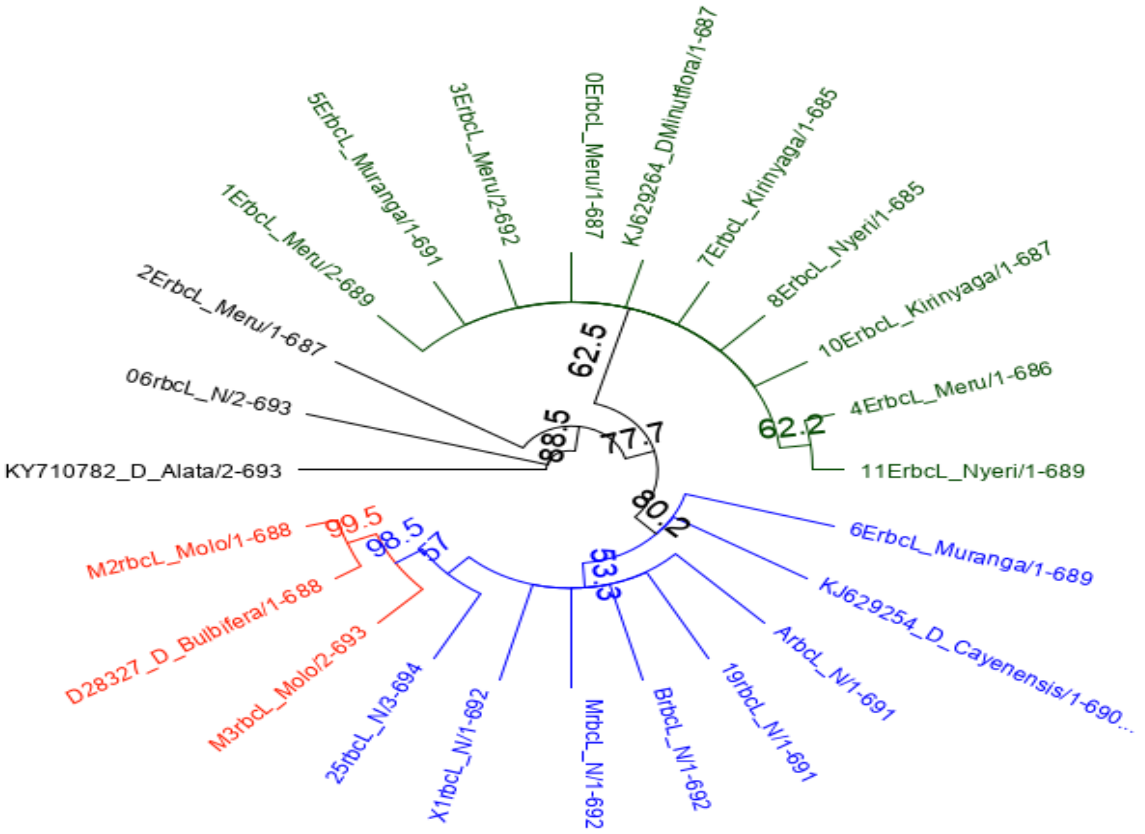


Figure 4.2: Phylogenetic tree reconstruction based on *rbcL* marker.
Each color indicates a different clade whereas a clade represents a distinct species. Clade A, green color represents *D. minutiflora;* Clade B, blue color is *D. cayenensis;* Clade C, black color represents *D. alata* and Clade D, red color represents *D. bulbifera* species.

Three clades were obtained by Maturase K (*matK*) marker on phylogenetic reconstruction, where each clade/subclade represented different *Dioscorea species* (Fig. 4.3). One subclade with more than two accessions was observed to segregate from clade A. The species identification included *D. Minutiflora*, *D. cayenensis subspecies rotundata (cayenensis-rotundata* Complex) and *Dioscorea bulbifera.* Clade A contained *D. minutiflora* with a support value of 55.4, a polyphyletic

relation and its cultivars included 1EMatK_Meru, 5EMatK_Muranga and 0EMatK_Meru. Clade B consisted of *D. cayenensis subspecies rotundata* which was supported by a value of 62.6 with a monophyletic relationship and its accessions included 7ErMatK_Kirinyaga, BMatK_N, AMatK_N, X1MatK_N and 19MatK_N. *Dioscorea bulbifera* of Clade C was strongly supported by a value of 100 and its accessions were M3MatK_Molo and M2MatK_Molo and it had a monophyletic relationship. Accession 8EMatK_Nyeri had ambiguous identity despite scoring 99.85% similarity for *D. minutiflora.*

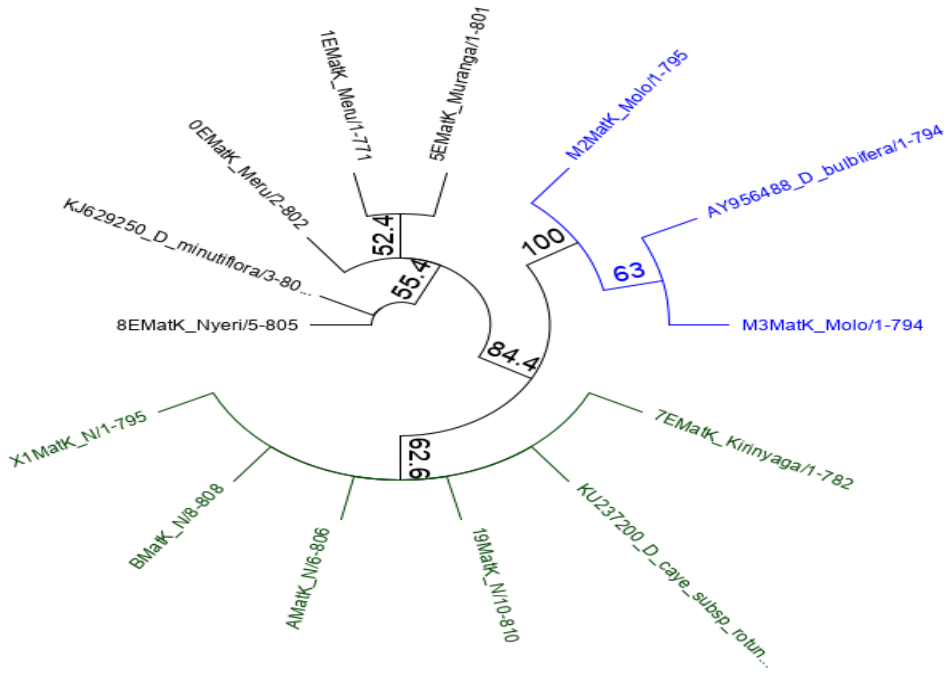


Figure 4.3: Phylogenetic tree reconstruction based on *matK* marker.
Different colors are used to represent the respective clades which represent different species. Clade A, black color represents *D. minutiflora;* Clade B, green color is *D. cayenensis-rotundata complex;* Clade C, blue color represents *D. bulbifera* species.

The use of combined *rbcL* and *matK* markers to build a phylogenetic tree was also explored (Fig. 4.4). The discriminatory power by individual markers upon combination was clearly notable. *MatK* was able to discriminate the species *D. cayenensis* by *rbcL* to be a *D. cayenensis rotundata* complex. Only X1MatK had

ambiguous identity while 8EMatK_Nyeri was correctly classified to belong to *D. minutiflora* species.

The cultivars which belong to the species *bulbifera* were correctly classified by the two markers and their relationship resolved to be monophyletic. The accessions of the species *alata* were correctly classified by *rbcL* marker and its relationship was monophyletic. All the accessions for *D. minutiflora* were correctly resolved by the two markers except 7EMatK_ Kirinyaga. Accession 7EMatK_ Kirinyaga was identified by *matK* to be a *D. cayenensis subspecies rotundata*.

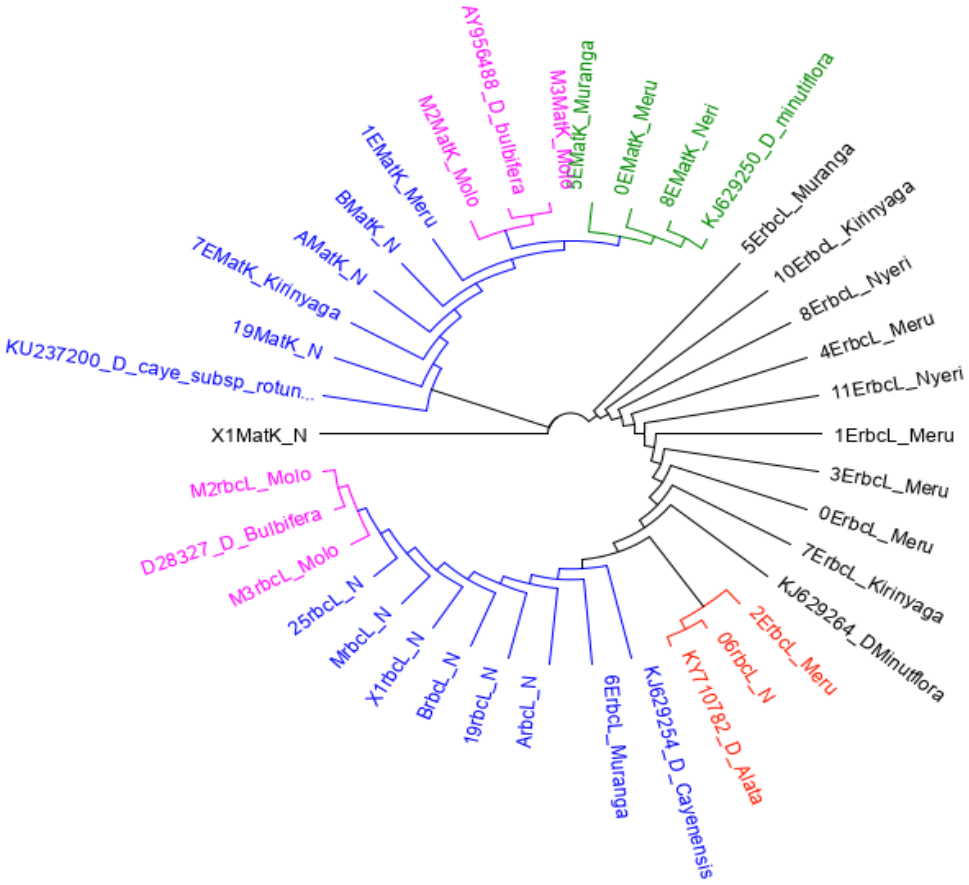


Figure 4.4: Phylogenetic tree reconstruction based on both *rbcL* and *matK* markers. Different colors are used to represent the respective clades which represent different species. Clade A, blue color represents *cayenensis-rotundata* complex; Clade B, purple color is *D. bulbifera*; Clade C, red color represents *D. alata* species. Colors green and black represents Clade D and E respectively which both represent *D. minutiflora* species. The two colors were chosen to represent the species, *D. minutiflora*, because of inconsistency in some of its accessions that included ambiguity in 8EMatK_Nyeri and reclassification by *matK* marker of 7EMatK_ Kirinyaga cultivar.

4.4 Phylogeography analysis

The yam accessions maximally differentiated into two distinct clades (Clade I and Clade II) with reference to *rbcL* sequences and Clades were differentiated in 1991 (Fig. 4.5). The differentiation covered a period of 20 years (1991-2011). Clade I contained subclades which were classified into A, B, C, D, E and F based on the node age that indicated their divergence time from their common origin (ancestor). These included A_1993 (0ErbcL_Meru) B_1995 (2ErbcL_Meru, 3ErbcL_Meru and 4ErbcL_Meru), C_1998 (06rbcL_N, 8ErbcL_Nyeri, 5ErbcL_Muranga and 6ErbcL_Muranga), D_2000 (7ErbcL_Kirinyaga), E_2002 (11ErbcL_Nyeri, 10ErbcL_Kirinyaga) and F_2011 (25rbcL_N and 19rbcL_N). The accessions in this clade were from upper eastern (Meru) and central (Kirinyaga, Muranga and Nyeri) regions of Kenya. Accessions from the Gene bank in this clade are 06rbcL_N, 25rbcL_N and 19rbcL_N which were originally from Nigeria then cultivated in Kenya and preserved at the National Gene Bank of Kenya.

Clade II had only two instant time of divergence (1994 and 1995). The accessions that diverged in the years 1994 include 1ErbcL_Meru, BrbcL_N, ArbcL_N, X1rbcL_N and MrbcL_N while those for year 1995 include M3rbcL_Molo and M2rbcL_Molo. The cultivars included in this clade were from Rift Valley and also those originally obtained from Nigeria and cultivated in Kenya. Only one cultivar from Meru was included in clade II.

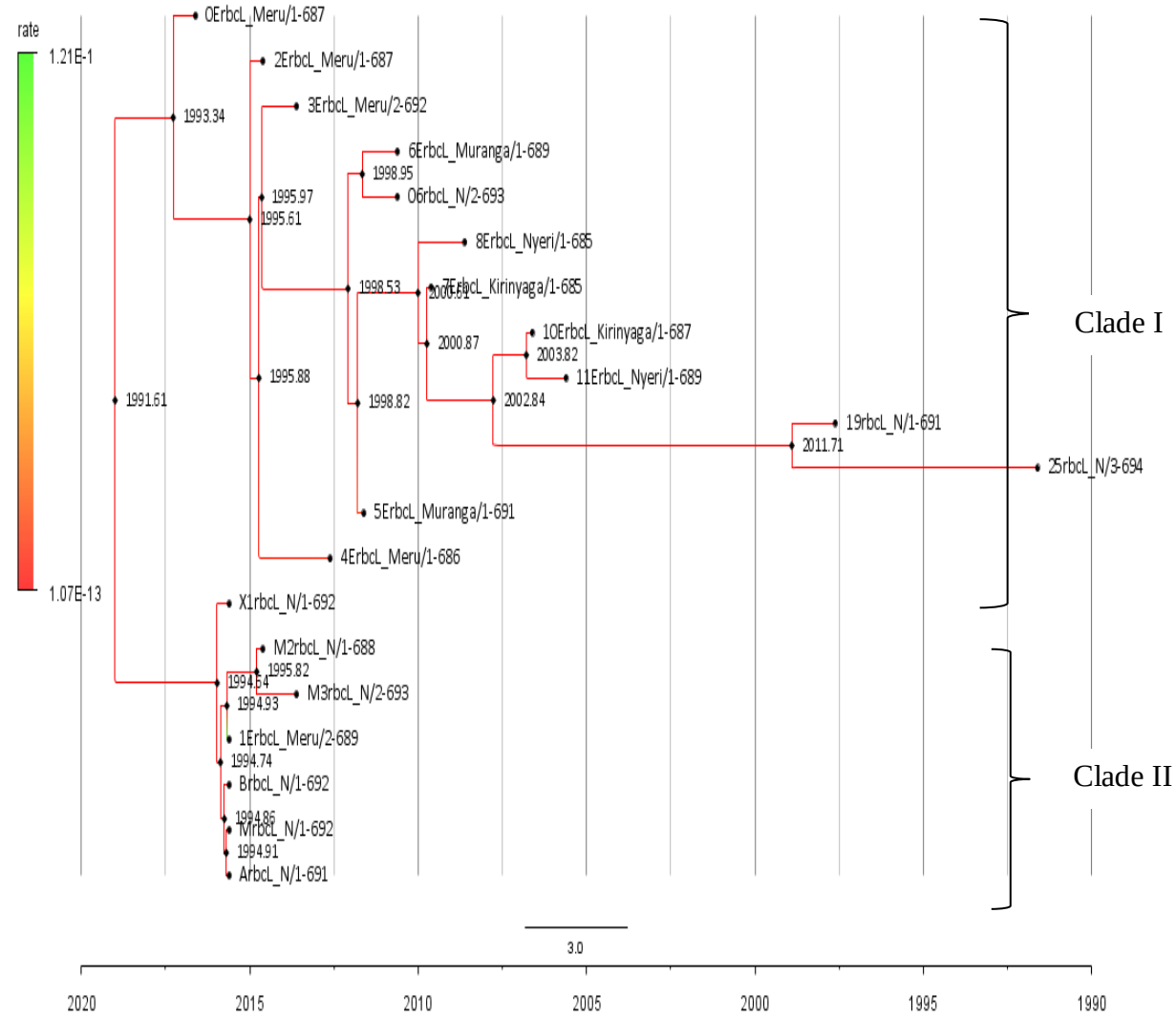


Figure 4.5: Phylogeography representation based on *rbcL* marker.
Two distinct clades resulting from the species differentiation were obtained. Clade I and Clade II generally represented species from different geographical region. The differentiation time indicated inform the relatedness of the cultivars in each clade.

Yam accessions were distinctively categorized into two clades (Clade I and Clade II) that differentiated in 1999 based on *matK* region (Fig. 4.6). The differentiation covered a period of 8 years (1999-2007). Clade I contained accessions 1EMatK_Meru and 0EMatK_Meru that diverged in the year 1999. All the accessions in clade I were from upper eastern (Meru) region of Kenya. Clade II contained subclades that differentiated in the year 2000 (M2MatK_Molo and X1MatK_N), 2001(M3MatK_Molo), 2002 (5EMatK_Muranga), 2004 (8EMatK_Nyeri and

AMatK_N), 2006 (7ErMatK_Kirinyaga) and 2007 (BMatK_N and 19MatK_N). The accessions in clade II were from Rift Valley, Central (Muranga, Kirinyaga and Nyeri), upper Eastern (Meru) regions of Kenya and National Genebank of Kenya.

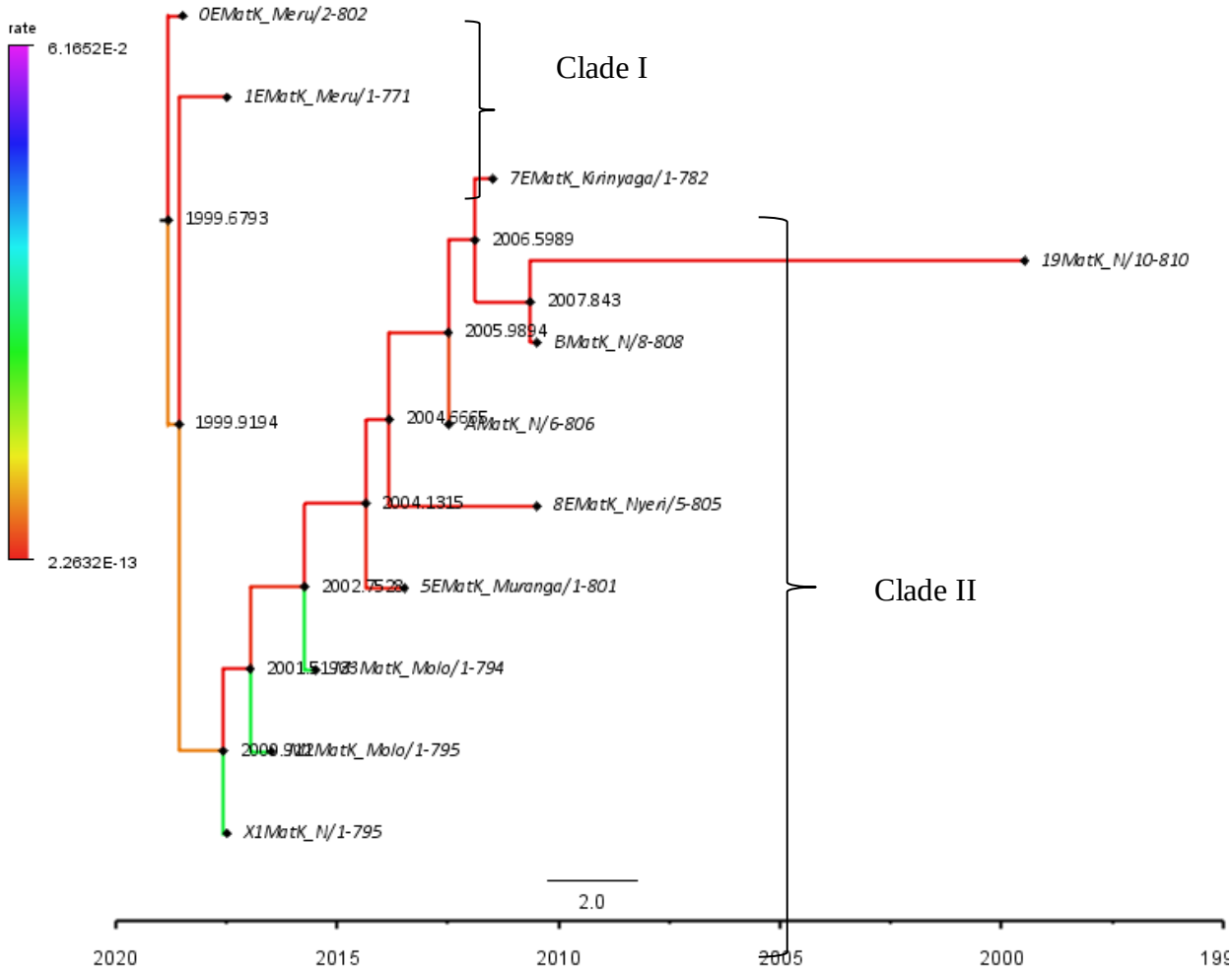


Figure 4.6: Phylogeography representation based on *matK* marker.
Two distinct clades resulting from the species differentiation were obtained. Clade I and Clade II generally represented species from respective geographical regions. The differentiation time indicated inform the relatedness of the cultivars in each clade.

4.5 DNA polymorphism

4.5.1 Genetic diversity of *rbcL* polymorphism

Twelve (12) segregation sites (S) or variable sites were identified within the *rbcL* gene of the accessions. The nucleotide diversity π was 0.00392 while the average number of nucleotide differences was 2.668 (Table 4.1). The *rbcL* gene had a total of 669 monomorphic sites and a sequence length of 692 base pairs while twelve polymorphic sites identified comprised of 3 singleton mutations and 9 parsimony informative bases (Table 4.1). The corresponding codons in the *rbcL* gene were mutants and belonged to both synonymous and non-synonymous mutations. These mutations were considered to be both singleton and parsimony informative sites.

Table 4.1: DNA polymorphism based on *rbcL* marker

Polymorphic sites/ Segregation sites (S)	12	Position in the Gene	Reference Codon	Amino Acid	Variant Codon	Resulting Amino Acid
Singleton	3	5	TGT	Cys	TTT	Phe
		7	TGG	Trp	GGG	Gly
		682	GGG	Gly	AGG	Arg
Parsimony informative sites	9	12	ATT	Ile	ATC	Ile
		58	CTA	Leu	ATA	Ile
		163	TGG	Trp	CGG	Arg
		280	CTA	Leu	TTA	Leu
		430	ACA	Thr	GCA	Ala
		433	CGG	Arg	TGG	Trp
		469A	TGG	Trp	CGG	Arg
		551A	TCT	Ser	TTT	Phe
		670	GGG	Gly	AGG	Arg
Nucleotide diversity π (Pi)	0.00392					
k(Average number of nucleotide difference)	2.668					
Sequence Length (base pairs)	692					
No of Sequences	20					

The polymorphic sites in the respective accessions occurred at codon positions **5** and **7** in 4ErbcL_Meru, **12** in both 4ErbcL_Meru and 11ErbcL_Nyeri and they were all present in *D. Minutiflora* species. *D. minutiflora* acccessions had variations (Parsimony informative sites) in codon positions **58** and **438**. All species except *D. bulbifera* had polymorphic sites at codon positions **163**, **280**, **430** and **670**. Cultivars that belonged to *D. alata* varied at codon positions **438**, **469** and **551**. Synonymous mutations occurred in positions 12, 58, 280 and 430. Mutations considered non synonymous occurred in positions 5, 7, 163, 433 469, 551 670 and 682. These non-synonymous mutations were attributed to genetic diversity of the yam accessions.

4.5.2 DNA divergence between populations

Results on DNA divergence between yam populations was measured by computing their variance Pi, Dxy and Da as outlined by Jukes and Cantor algorithm implemented in DnaSP, are presented in Table 4.2.

Table 4.2: DNA divergence between *Dioscorea* species populations on *rbcL* marker

	DB	*DM*	*DB)*	*DA*	*DB*	*DC*	*DM*	*DA*	*DM*	*DC*	*DA*	*DC*
Population	*(P1)*	*(P2)*	*(P1)*	*(P2)*	*(P2)*	*(P2)*	*(P1)*	*(P2)*	*(P1)*	*(P2)*	*(P1)*	*(P2)*
Polymorphic sites in each population	0	4	0	0	0	0	4	0	4	0	0	0
Total number of polymorphic sites	10		7		4		7		6		3	
Average number of nucleotide difference (kt)	2.836		4.667		1.556		1.855		1.658		1.167	
Nucleotide diversity(pt)	0.00416		0.00679		0.00227		0.00272		0.00244		0.0017	
Number of fixed differences	6		7		4		4		2		3	
Mutations polymorphic in P1 but Monomorphic in P2	0		0		0		4		4		0	
Mutations polymorphic in P2, but monomorphic in P1	4		0		0		0		0		0	
Shared Mutations	0		0		0		0		0		0	
Average number of nucleotide differences between populations	6.556		7.000		4.000		3.556		2.556		3.000	
Average nucleotide substitution per site between population	0.00961		0.01019		0.00584		0.00521		0.00375		0.00438	
Number of net nuc. subs. per site between populations	0.00884		0.01019		0.00584		0.00444		0.00298		0.00438	

KEY: DB = *Dioscorea bulbifera, DM = Dioscorea minutiflora, DA = Dioscorea alata, DC= Dioscorea cayenensis.*

The accessions elucidated by *rbcL* marker had no shared mutations between the yam species population. The average number of nucleotide substitution per site between populations ranged from 0.00375 to 0.01019 which corresponded to its number of net nucleotide substitution per site between populations that ranged from 0.00298 to 0.01019. Only mutations in *D. minutiflora* species were polymorphic but monomorphic in the other yam species populations. The total number of fixed differences between populations was: *D. bulbifera* and *D. minutiflora* = 6, *D. bulbifera* and *D. alata* = 7, *D. bulbifera* and *D. cayenensis* = 4, *D.alata* and *D. minutiflora* = 4, *D. cayenensis* and *D. minutiflora* = 3, *D. alata* and *D. cayenensis* = 3. The number of fixed differences between populations was inferred from the total polymorphic sites between the populations. The fixed differences number alluded to the diversity and relatedness of the accessions used in this study. Species *D. minutiflora* had the highest number of polymorphic sites between populations implying a diversity of its accessions.

4.5.3 Tajima neutrality

Neutrality test was conducted on the twenty sequences of *rbcL* gene to establish if there was any selection taking course using Tajima D, Fu &Li's D* and Fu & Li's F* indices. The computed Tajima D value was -0.755777 while the Fu and Li's D* value was 0.22299 and Fu and Li's F*= -0.0593 and the nucleotide diversity was 0.00392 less than the conventional value of 0.05. The cultivars exhibited negative selection.

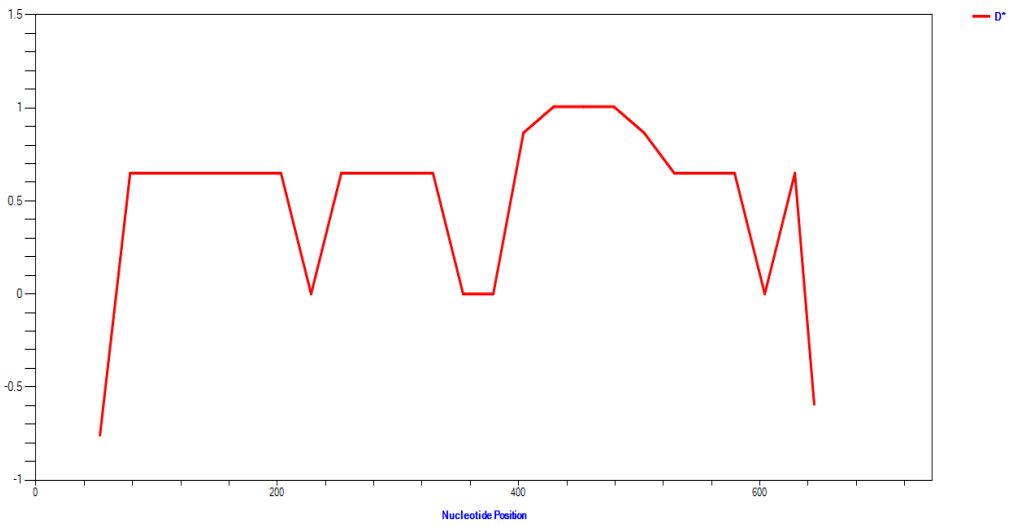


Figure 4.7: Tajima neutralityy graph for *rbcL* marker.
The graph was computed using sliding window analysis with a length of 100 sites and step size of 25 bases.

4.5.4 Haplotype analysis

Haplotype analysis of yam species was conducted to ascertain the network of the cpDNA haplotypes. Twelve polymorhic sites within the *rbcL* sequence were involved in the detemination of haplotypes of the accessions. There were 7 haplotypes within the *rbcL* sequence with a diversity index of 0.800 and variation of 0.00374 (Table 4.3).

Table 4.3: Haplotype analysis based on *rbcL* marker

Number of Haplotypes	Haplotype Diversity	Variation of Haplotype Diversity
7	0.8	0.004

Haplotype Number	*Dioscorea species*	Haplotype	Accessions
Hap_1: 2	*D. bulbifera*	GTTCTCACTCGG	[M2rbcL_Molo and M3rbcL_Molo]
Hap_2: 1	*D. minutiflora*	TGCACTGTTCAG	[4ErbcL_Meru]
Hap_3: 2	*D. alata*	GTTCCTGTCTAG	[2ErbcL_Meru and 06rbcL_N]
Hap_4: 7	*D. cayenensis*	GTTCCTGCTCAG	[6ErbcL_Muranga, ArbcL_N, 25rbcL_N, 19rbcL_N, BrbcL_N, MrbcL_N and X1rbcL_N]
Hap_5: 1	*D. minutiflora*	GTCACTGTTCAG	[11ErbcL_Nyeri]
Hap_6: 1	*D. minutiflora*	GTTACTGTTCAA	[1ErbcL_Meru]
Hap_7: 1	*D. minutiflora*	GTTACTGTTCAG	[5ErbcL_Muranga, 3ErbcL_Meru, 0ErbcL_Meru, 7ErbcL_Kirinyaga, 8ErbcL_Nyeri and 10ErbcL_Kirinyaga]

Haplotypes 4 and 7 had the highest number of accessions with each having seven and six accesions respectively. Haplotypes 1 and 3 had two accessions each while each of the haplotypes 2, 5 and 6 had only one accession (Table 4.3). The distribution of haplotypes was in accordance to their geographical location and species type. Haplotypes 1 contained accessions from Rift valley (Molo) region which belonged to *D. bulbifera* species. Haplotype 2 and 6 had accessions from upper eastern (Meru) that belonged to *D. minutiflora* species. Haplotype 3 had accessions from both upper eastern (Meru) and National Genebank of Kenya that belonged to species *D. alata*. Haplotype 4 comprised of accessions from National Genebank of Kenya and one from central (Muranga) region of Kenya which were *D. cayenensis* species. Haplotype 5 had one yam accession from central (Nyeri) region of Kenya which was *D. minutiflora* species. Haplotype 7 comprised of accessions of *D. minutiflora* from

central (Nyeri, Muranga, and Kirinyaga) and upper eastern (Meru) regions of Kenya. Haplotype 3 and 4 comprised of yam accessions from different geographical location. No accession was placed in multiple haplotypes.

4.5 6 Genetic diversity of *matK* polymorphism

Fourteen (14) segregation sites (S) were identified within the *matK* sequences of the yam accessions. The nucleotide diversity π in the eleven (11) sequences was 0.00632 while the average number of nucleotide differences was 4.87273. A total of 757 monomorphic sites /invariant base pairs and 14 polymorphic sites were elucidated. The polymorphic sites comprised of both singleton mutations (2) and twelve parsimony informative bases (Table 4.4). The corresponding codons in the *matK* sequence were mutants and belonged to both synonymous and non-synonymous mutations. These mutations were both singleton and parsimony informative sites.

Table 4.4: DNA polymorphism based on *matK* marker

Polymorphic sites/Segregation Sites (S)	14	Position in the Gene	Reference		Variant Codon	Resulting Amino Acid
Singleton	2	758	CAC	Hist	CTC	Leu
		784	TTG	Leu	CTG	Leu
Parsimony informative sites	12					
		83	CAG*	Gln	CCG	Pro
		257	GGA*	Gly	GAA	Glu
		403	GGA*	Gly	AGA	Arg
		412	CTT*	Leu	TTT	Phe
		468	CAT	Hist	CAC	Hist
		495	TTC	Phe	TTT	Phe
		532	CCT	Pro	TTT	Phe
		533	CCT*	Pro	TTT	Phe
		545	AGG	Arg	AAG	Lys
		561	GAA	Glu	GAG	Glu
		658	ATA	Ile	GTA	Val
		672	CCC	Pro	CCT	Pro
Nucleotide Diversity π (Pi)	0.00632					
k(Average number of nucleotide difference)	4.87273					
Sequence Length (base pairs)	795					
No of Sequences	11					

The singleton sites in the accession 8EMatK_Nyeri occurred at codon positions **758** and **784** which was a *D. minutiflora* species. For other accessions in this species, parsimony informative sites occurred at codon positions **257** and **468.** All species had variations in codon positions **83, 403, 495, 532, 533, 545, 561, 658, 672** except *D. bulbifera*. *D. cayenensis-rotundata* varied at codon position **412.** Synonymous mutation occurred in positions 468, 495, 545, 561, 658, 672 and 784. The synonymous mutations were attributed to selection that occurred within the respective yam species population. Mutations considered non synonymous occurred in positions 83, 257, 403, 412, 532, 533 and 758. These non-synonymous mutations were attributable to genetic diversity of the yam accessions.

4.5.7 DNA divergence between populations

DNA divergence between yam populations based on *matK* gene was measured by computing their variance Pi, Dxy and Da as outlined by Jukes and Cantor algorithm implemented in DnaSP.

Table 4.5: DNA divergence between yam species populations based on *matK* marker

Population	*D. bulbifera (P1)*	*D minutiflora (P2)*	*D. bulbifera (P1)*	*D. caye_rotundata*	*D. minutiflora (P1)*	*D. caye_rotundata (P2)*
Polymorphic sites in Each population	0	4	0	0	2	0
Total Number of Polymorphic sites	13		10		5	
Average Number of nucleotide difference (kt)	6.533		5.333		2.214	
Nucleotide diversity(pt)	0.00847		0.00682		0.00287	
Number of fixed differences	11		10		3	
Mutations polymorphic in P1 but monomorphic in P2	0		0		2	
Mutations polymorphic in population 2, but monomorphic in population 1	2		0		0	
Shared mutations	0		0		0	
Average number of nucleotide differences between populations	11.500		10.000		3.50	
Average nucleotide substitution per site between population	0.01492		0.01279		0.00454	
Number of net nuc. subs. per site between populations	0.01427		0.01279		0.00389	

There were no shared mutations between the yam species population based on *matK* marker (Table 4.5). The average number of nucleotide substitution per site between populations ranged from 0.00454 to 0.01492 which corresponded to its number of net nucleotide substitution per site between populations that ranged from 0.00389 to 0.01427. Only mutations in *D. minutiflora* species were polymorphic but

monomorphic in the other yam species populations. The total number of fixed differences between populations was: *D. bulbifera* and *D. minutiflora* = 11, *D. bulbifera* and *D. cayenensis-rotundata* = 10, *D. cayenensis-rotundata* and *D. minutiflora* = 3. The number of fixed differences was inferred from the total polymorphic sites between the populations. Relatedness among the yam species can be deduced from their polymorphic site differences/ number of fixed differences. The relatedness of *D. minutiflora* and *D. cayenensis-rotundata* was 3 base differences while that of *D. cayenensis-rotundata* and *D. bulbifera* was 10 base difference. *D. bulbifera* and *D. minutiflora* was 11 bases genetically distant.

4.5.8 Tajima neutrality

Neutrality test was conducted on the eleven sequences of *matK* sequence to establish if selection existed among the populations using Tajima D, Fu &Li's D* and Fu & Li's F* indices . The computed Tajima D value was 0.08564 while the Fu and Li's D* value was 0.93064 and Fu and Li's F* = 0.73729 and the nucleotide diversity was 0.00632, which was less than the conventional value of 0.05. According to *matK* sequence these yam accessions exhibited balanced selection as there were few numbers of singleton mutation as referred from the positive Tajima D.

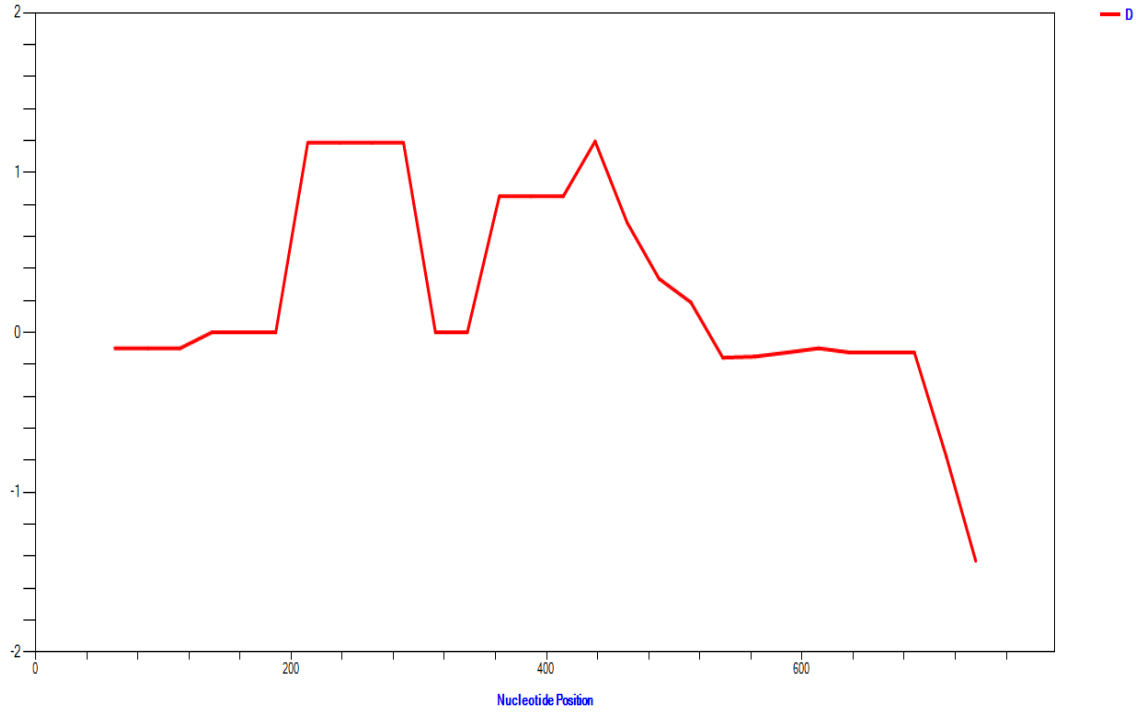


Figure 4.8: Tajima neutrality graph for *matK* marker
The graph was computed using sliding window analysis with a length of 100 sites and step size of 25 bases.

4.5.9 Haplotype analysis

Haplotype analysis of yam species was conducted to ascertain the network of the cpDNA haplotypes. There were 4 haplotypes within the *matK* gene with a diversity index of 0.745 and variation of 0.00956. The haplotypes were generated from the 14 polymorphic sites within this gene (Table 4.6).

Table 4.6: Haplotype analysis based on *matK* marker

Number of Haplotypes	Haplotype Diversity	Variation of Haplotype Diversity
4	0.745	0.00956

Haplotype Number	*Dioscorea species*	Haplotype	Accessions
Hap_1: 2	*D. bulbifera*	AGGCTCCCGAACAT	[M2MatK_Molo and M3MatK_Molo]
Hap_2: 1	*D. minutiflora*	CAACCTTTAGGTTC	[8EMatK_Nyeri]
Hap_3: 5	*D. cayenensis rotundata*	CGATTTTTAGGTAT	[7EMatK_Kirinyaga, 19MatK_N, AMatK_N, BMatK_N and X1MatK_N]
Hap_4: 3	*D. minutiflora*	CAACCTTTAGGTAT	[0EMatK_Meru, 5EMatK_Muranga and 1EMatK_Meru]

Majority of the accessions were classified to belong to haplotypes 3 and 4. Each of the haplotypes had five and three accessionsrespectively (Table 4.6b). Haplotype 1 had two accessions while haplotype 2 had only one accession. The haplotype distribution elucidated the species and their respective geographical locations. Haplotype 1 accessions were derived from Rift valley region and belonged to species, *D. bulbifera*. Haplotype 2 had an accession derived from Nyeri Kenya which belonged to *D. minutiflora* species. Haplotype 3 comprised of accessions from both Kirinyaga and National Genebank of Kenya which belonged to *D. cayenenis rotundata*. Haplotype 4 comprised of accessions from Muranga and Meru which belong to species *D. minutiflora*. Haplotype 3 comprised of accessions from different geographical locations. None of the accessions belonged to multiple haplotypes.

CHAPTER FIVE

DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS

5.1 Discussion

Species identification of yam accessions cultivated in Kenya was undertaken using *rbcL* and *matK* barcoding markers. Based on *rbcL* and *matK* sequences, DNA polymorphism, haplotypes, DNA divergence, phylogeography and phylogenetic relations of the yam species were evaluated. Both markers were able to discriminate the accessions into their respective yam species. *MatK* was able to discriminate up to subspecies level making it good barcoding marker candidate for inter-specific divergence resolution among the yam species. This was evidenced by its discrimination of cultivars identified as *D. cayenensis* species by *rbcL* to species *D. cayenensis subspecies rotundata*. It was also able to resolve the identity of accession 7ErbcL_Kirinyaga identified as *D. minutiflora* by *rbcL* to D. *cayenensis subspecies rotundata.* This agrees with Carneiro *et al*. (2019) that *matK* proved to be the most efficient in plant authentication at species level.

Similarly, *rbcL* is also a good candidate for yam plant barcoding as it can universally amplify and produce quality amplicons resulting to quality sequences. However, its discriminatory ability is fairly efficient because of its highly conserved sequences. Hollingsworth *et al.* (2011) observed that *rbcL* marker is an efficient barcoding marker but lowly efficient in inter-specific divergence discrimination. This was contrary to Li *et al.* (2014) that *rbcL* and *matK* failed to discriminate *Calligonum* species. By combining the two (*rbcL+ matK*) markers, barcoding and discrimination of yam species can be achieved. Girma *et al.* (2016) determined that combining *rbcL+ matK* markers was 76.2% optimum for yam barcoding and *matK* marker detected high interspecific variation and identified 63.2% of yam species. Sun *et al.* (2012) demonstrated that a combination of the two markers achieved a higher success rate of species discrimination than a single marker. Overall, the two markers were efficient with reference to yam species discrimination. This concurs with

previous reports from studies in West Africa and China (Wu *et al*., 2019; Efisue, 2015).

Both *rbcL* and *matK* markers were able to discriminate four yam species from farmers' fields in Kenya and two species from the National Genebank of Kenya (GBK). The identified yam species from farmers' fields included *D. Minutiflora, D. cayenensis subspecies rotundata (cayenensis-rotundata* Complex), *D. alata* and *Dioscorea bulbifera.* The yam species from the GBK included *D. cayenensis subspecies rotundata (Cayenensis-Rotundata* Complex) and *D. alata*. *D. minutiflora* was the most cultivated yam species by farmers in Kenya. The results agree with the available information which indicated that yams species grown in Kenya are *D. rotundata, D. minutiflora, D. bulbifera,* and *D. dumetorum* (FAO, 2009). Similarly, Kariuki (2012) indicated that taxonomic position of Kenyan cultivated yams belonged to seven major species represented by *Dioscorea cayanensis, D. rotundata, D. mangenotiana, D. alata, D. sehimperiana, D. dumetorum* and *D. bulbifera.*

Phylogeography establishes the congruence of geographical and phylogenetic relationships of species in order to explain the process that defines their populations' genetic diversity in space and time. The yam accessions were maximally differentiated into two distinct clades based on both *rbcL* and *matK* sequences. Clade I cultivars using *rbcL* gene were from Meru, Kirinyaga, Muranga, Nyeri and GBK. Yam accessions in this clade belonged to *D. minutiflora, D. alata* and *D. cayanensis* species and were obtained from farmers' fields while accessions from GBK included in this clade, belonged to *D. cayanensis and D. alata* species. Yam accessions in Clade II based on *rbcL* sequence were from farmers' fields and GBK. The accessions in this clade belonged to *D. bulbifera* and *D. cayenensis* species.

For *matK* sequences, clade I had cultivars from Meru which belonged to *D. minutiflora* species, while Clade II comprised of cultivars from all the sampled regions and belonged to species *D. bulbifera, D. minutiflora* and *D. cayenensis subspecies rotundata.* Each of the species in this clade was a derivative of a sub clade that fulfilled the phylogeography affirmation of congruence that exists between

species geographic location and its phylogeny. It is evident from the results that the clade that an accession belonged to was informed by its geographic location and its species. The yam species clades originated from their time of divergence. It is important to note that phylogeography information can be used to predict the gene flow within population in a given region. The results agree with Arnau *et al*. (2017) that there are two genepools that are divergent in India and Vanuatu which provides a clear genetic differentiation between yam species in Asia and South Pacific which both have secondary diversification.

Yams were independently domesticated in Asia, America and Africa in three different times with the species being *Dioscorea alata*, *Dioscorea trifida* and *Dioscorea rotundata*, respectively (Scarcelli *et al.* 2019). *D. rotundata* was mainly grown in West Africa belt. *D. cayenensis subspecies rotundata* predominantly in Nigeria differentiated in 1994 and 2002 for *rbcL* marker and in the year 2000, 2005 and 2007 for *matK*. This confirmed that *D. rotundata* that was originally domesticated in Nigeria, is the *D. cayenensis subspecies rotundata* that differentiated in the 1990s and 2000s for both markers. The presence of D. *cayenensis subspecies rotundata* cultivated in Kenya in a different clade but together with *D. alata* from GBK and originally obtained from Nigeria confirmed that the species origin is Nigeria but through divergence, is domesticated in Kenya.

The domestication of *D. bulbifera* was mainly in the rift valley Kenya. The species is native to Asia but introduced to tropical Africa and mainly grows in areas with high humid, temperatures of 25-35 °C and annual rainfall of above 1000 mm (Wunderlin and Hansen, 2008). These environmental conditions are found in the Rift valley (Molo) region where the species is domesticated in Kenya. The variability of this accession is small as its monophyletic relationship was supported by a bootstrap value of >98% similarity in both markers. Being domesticated in Rift Valley (Molo) could mean that the environmental conditions in the region supports the species and its genetic variation is small as the divergence time was 1994 for *rbcL* and 2004 and 2005 for *matK* and present in the same clade for the two genes.

The presence of both *D. minutiflora* and *D. alata* in the same clade for *rbcL* marker demonstrated their closeness with regard to genetic diversity. *D. alata* origin is not well documented although it is believed to have originated from South East Asia in the 16th century and dispersed to other regions of the world including Western Africa (Arnau, 2017). The differentiation of *D. alata* in 1995 and 1998 in both farmers' fields yam accessions and GBK (originally from Nigeria) indicated that they were related as affirmed by their existence in the same clade. This also implied that their genetic variation could be low.

Yam accessions for *D. minutiflora* were distributed in both clades in the two markers with their divergence time being 1994 to 1995 and 2002 to 2006 for *rbcL and matK*, respectively. These accessions closely clustered with *D. alata* and *D. cayenensis subspecies rotundata* species implying that their differentiation time was near each other and therefore their genetic variation could be small. It also confirmed that the *D. cayenensis* is a product of hybridization of male *D. minutiflora* and female *D. rotundata (*Terauchi *et al.*, 1992). The origin of the cultivar could also be traced to tropical regions of Eastern Africa which is why it is a dominant yam accession in Kenya.

The nucleotide diversity among the yam species was revealed to be 0.00392 (for *rbcL*) and 0.00632 (for *matK*). This demonstrated the genetic variation among the yam species as it depicts polymorphisms. Polymorphism in the *matK* sequence was slightly higher than that in *rbcL* sequences therefore *rbcL* is slightly conserved and serves a good candidate for barcoding while *matK* was good for interspecific discrimination. The haplotype diversity was shown to be 0.800 for *rbcL* and 0.745 for *matK* sequences among the yam accessions. Haplotype diversity was slightly higher in the *rbcL* gene sequence as compared to *matK* gene of the yam accessions. The slightly higher haplotype diversity in the *rbcL* sequence can be linked to the total number of haplotypes identified (seven) as compared to four haplotypes determined by *matK* sequence. *RbcL* gene sequence of yam accessions had a lower nucleotide diversity and higher haplotype diversity while *matK* sequence had a higher

nucleotide diversity and lower haplotype diversity. Mulualem *et al.* (2018) found that the genetic diversity of yam landraces in Ethiopia ranged from 0.00 to 0.80 with a polymorphism of 0.30 information content. Arnau (2017) using SSRs demonstrated that the diversity among *D. alata* ranged from 0.20 to 0.86 which is close to the haplotype diversity of the two markers used in the current study.

The polymorphism among the yam accessions resulted in both synonymous and non-synonymous mutations. The synonymous mutations can be attributed to selection that occurred within the respective species population. In the *rbcL* sequences, synonymous mutations occurred at positions 12, 58, 280 and 430 while non-synonymous mutations occurred at positions 5, 7, 163, 433 469, 551 670 and 682. Ude *et al*. (2019) noted that transversional mutation of G/T occurred at a consensus position of 335 then transitions at 362 (A/G), 368 (A/G), 371 (C/T) and 391 (C/T) within the yam accessions in Nigeria. Synonymous mutations in the *matK* gene sequences occurred at positions 468, 495, 545, 561, 658, 672 and 784 while non-synonymous mutations occurred at positions 83, 257, 403, 412, 532, 533 and 758. The nature of the substitutions that occurred were both trans-versions and transition. Transition mutations can be attributed to selection while transversion mutation can be linked to genetic diversity.

The divergence between yam populations was also evaluated based on the two markers. The average number of nucleotide substitution per site in *rbcL* sequences between populations ranged from 0.00375 to 0.01019 while its nucleotide diversity ranged from 0.0017 to 0.00416. The nucleotide substitution per site in *matK* gene sequences between populations ranged from 0.00454 to 0.01492 while nucleotide diversity ranged from 0.00287 to 0.00847. The DNA divergence between the yam populations was similar in the two markers as there were no shared mutations. Similarly, the divergence in a descending order was that the diversity was highest between *D. minutiflora* and *D. bulbifera* followed by *D. bulbifera* and *D. cayenensis subspecies rotundata*, while *D. minutiflora* and *D. cayenensis subspecies rotundata* was the lowest. This implies that differentiation between the yam accessions is as

demonstrated by their divergence. This is consistent with Mulualem *et al.* (2018) who demonstrated using simple sequence repeats (SSR) that the genetic distances of yam species found in Ethiopia had a range of 0.00 to 1.0.

Tajima D result by *rbcL* sequences indicated that yam accessions were under sweep selection while *matK* gene sequences demonstrated that yam populations were under balanced selection. The Tajima D for *rbcL* was -0.75777 while that for *matK* was 0.08564. Positive Tajima value indicates a balanced selection where alleles with intermediate frequency thrive. It also denotes a population that is formed recently from two different populations which *Dioscorea* populations can exhibit. Negative Tajima value denotes purifying selection where excessive polymorphisms of low frequency occur. It also implies that population growth is being exhibited. Both markers have pointed out that the yam plant is under selection pressure and growth. There were 3 singleton sites in *rbcL* gene sequences and 2 singleton sites in *matK* sequence which corresponded to the nature of selection among the yam species. These findings concur with Akakpo *et al.* (2017) that cultivated yam showed a skewed distribution to positive Tajima D value. Similarly, Wicke *et al*. (2011) observed that *rbcL* has a strong purifying selection in autotrophic plants which reduces its evolution rate and ability to distinguish related species. Conversely, *matK* is under relaxed purifying selection and has a high rate of nucleotide substitution therefore better ability in species discrimination (Duffy *et al*., 2009).

5.2 Conclusions

Both *matK* and *rbcL* markers have proved to be good candidate barcodes for yam species discrimination and identification. *MatK* marker is efficient for subspecies resolution while *rbcL* is good for barcoding. The two markers were able to identify four species (*D. bulbifera, D. alata, D. minutiflora* and *D. cayenensis*) that are domesticated in Kenya. The cultivation of *D. minutiflora* is seemingly dominant in both central and upper eastern Kenya. The other three species are cultivated in the rift valley, eastern and central Kenya. It is evident that the yam status in the country

stands at four species. The distribution of the species in the Kenya is clustered into two main groups with regard to their phylogeography. Polymorphism in the yam species exhibits both synonymous and non-synonymous mutations. *RbcL* gene sequence of yam accessions had a lower nucleotide diversity and higher haplotype diversity while *matK* gene sequence recorded a higher nucleotide diversity and lower haplotype diversity. The DNA divergence between the yam species populations was similar based on the two markers and with no shared mutations.

5.3 Recommendations

There is need to establish a reference database for yam species cultivated in Kenya. Further studies should be conducted to establish whether species grown in the respective geographical region is due to their origin or favorable climatic conditions. The identified four yam species should be conserved at the National Genebank of Kenya. Field trials on the adaptability of each of the four yam species in the different growing regions should be conducted.

REFERENCES

Adebowale, A. R. A., Wahab, A. B., Sobukola, P. O., Obadina, A. O., Kajihausa, E. O., Adegunwa, O. M., & Tomlins, K. (2018). The antinutritional and vitamin composition of high-quality yam flour as affected by yam specie, pretreatment, and drying method. *Food Science and Nutrition*, *6*(8), 1985-1990.

Agre, P., Asibe, F., Darkwa, K., Edemodu, A., Bauchet, G., Asiedu, R., & Asfaw, A. (2019). Phenotypic and molecular assessment of genetic structure and diversity in a panel of winged yam (*Dioscorea alata*) clones and cultivars. *Scientific Reports*, *9*(1), 1-11.

Aighewi, B. A., Asiedu, R., Maroya, N., & Balogun, M. (2015). Improved propagation methods to raise the productivity of yam (*Dioscorea rotundata Poir.). Food Security*, *7*(4), 823-834.

Akakpo, R., Scarcelli, N., Dansi, A., Djedatin, G., Thuillet, A. C., Rhoné, B., & Vigouroux, Y. (2017). Molecular basis of African yam domestication: analyses of selection point to root development, starch biosynthesis, and photosynthesis related genes. *BMC Genomics*, *18*(1), 782.

Amit Roy, S. R. (2014). Molecular Markers in Phylogenetic Studies-A Review. *Journal of Phylogenetics and Evolutionary Biology*, *02*(02). https://doi.org/10.4172/2329-9002.1000131.

Amusa, N. A., Adigbite, A. A., Muhammed, S., & Baiyewu, R. A. (2003). Yam diseases and its management in Nigeria. *African Journal of Biotechnology*, *2*(12), 497-502.

Andres, C., AdeOluwa, O. O., & Bhullar, G. S. (2016). Yam (*Dioscorea* spp.). *Encyclopedia of Applied Plant Sciences*, *3*(October), 435–441. https://doi.org/10.1016/B978-0-12-394807-6.00177-5.

Andriamparany, J. N., Brinkmann, K., Jeannoda, V., & Buerkert, A. (2014). Effects

of socio-economic household characteristics on traditional knowledge and usage of wild yams and medicinal plants in the Mahafaly region of south-western Madagascar. *Journal of Ethnobiology and Ethnomedicine*, *10*(1). https://doi.org/10.1186/1746-4269-10-82.

Anokye, M., Tetteh, J. P., & Otoo, E. (2014). Morphological Characterization of Some Water Yam (*Dioscorea alata* L.) Germplasm in Ghana. *Journal of Agricultural Science and Technology B Journal of Agricultural Science and Technology*, *4*(March), 518–532.

Arnau, G., Abraham, K., Sheela, M. N., Sartie, A., & Asiedu, R. (2010). Root and Tuber Crops. *Handbook of Plant Breeding*.

Arnau, G., Bhattacharjee, R., Sheela, M. N., Malapa, R., Lebot, V., Abraham, K., & Pavis, C. (2017). Understanding the genetic diversity and population structure of yam (*Dioscorea alata* L.) using microsatellite markers. *PLoS One*, *12*(3), e0174150.

Arnau, G., Maledon, E., Nudol, E., & Gravillon, M. C. (2016). Progress and challenges in genetic improvement of yam (*Dioscorea alata* L.).[S13. 06]. WCRTC.

Asahina, H., Shinozaki, J., Masuda, K., Morimitsu, Y., & Satake, M. (2010). Identification of medicinal *Dendrobium* species by phylogenetic analyses using *matK* and *rbcL* sequences. *Journal of Natural Medicines*, *64*(2), 133–138. https://doi.org/10.1007/s11418-009-0379-8.

Asala, S. W., & Ebukiba, E. S. (2016). Profitability of yam production in Southern Guinea Savanna zone of Nigeria. *Net Journal of Agricultural Science*, *4*(1), 9-14.

Balogun, M. O., & Gueye, B. (2013). Status and prospects of biotechnology applications to conservation, propagation and genetic improvement of yam. *Bulbous Plants: Biotechnology*, 92-112.

Banson, K., & Danso, K. (2013). Improving the size and market value of an underutilized yam (*Dioscorea esculenta*) in Ghana: implications for crop breeding and production choices. *Journal of Life Sciences*, *7*(7), 732.

Barlagne, C., Cornet, D., Blazy, J. M., Diman, J. L., & Ozier-Lafontaine, H. (2017). Consumers' preferences for fresh yam: a focus group study. *Food Science and Nutrition*, *5*(1), 54-66.

Barton, H. (2014). Yams: origins and development. *Encyclopedia of Global Archaeology*, 7943-7947.

Besnard, G., Rubio De Casas, R., & Vargas, P. (2007). Plastid and nuclear DNA polymorphism reveals historical processes of isolation and reticulation in the olive tree complex (Olea europaea). *Journal of Biogeography*, *34*(4), 736–752. https://doi.org/10.1111/j.1365-2699.2006.01653.x.

Braukmann, T. W., Kuzmina, M. L., Sills, J., Zakharov, E. V., & Hebert, P. D. (2017). Testing the efficacy of DNA barcodes for identifying the vascular plants of Canada. *PLoS One*, *12*(1).

Bressan, E. A., Briner Neto, T., Zucchi, M. I., Rabello, R. J., & Veasey, E. A. (2014). Genetic structure and diversity in the *Dioscorea cayenensis/D. rotundata* complex revealed by morphological and isozyme markers. *Genetics and Molecular Research*, *13*(1), 425–437. https://doi.org/10.4238/2014.January.21.10.

Burgess, K. S., Fazekas, A. J., Kesanakurti, P. R., Graham, S. W., Husband, B. C., Newmaster, S. G., & Barrett, S. C. H. (2011). Discriminating plant species in a local temperate flora using the *rbcL+matK* DNA barcode. *Methods in Ecology and Evolution*, *2*(4), 333–340. https://doi.org/10.1111/j.2041-210X.2011.00092.x.

Cao, J., Jiang, D., Zhao, Z., Yuan, S., Zhang, Y., Zhang, T., & Huang, L. (2018). Development of chloroplast genomic resources in Chinese yam (*Dioscorea polystachya*). *BioMed Research International*, *2018*.

Carneiro, C. D. M. M., Brambach, F., Jair, K. H. B., Krutovsky, K. V., Kreft, H., Tjitrosoedirdjo, S. S., & Gailing, O. (2019). Integrating DNA Barcoding and Traditional Taxonomy for the Identification of Dipterocarps in Remnant Lowland Forests of Sumatra. *Plants (Basel, Switzerland)*, *8*(11).

Chaïr, H., Perrier, X., Agbangla, C., Marchand, J. L., Dainou, O., & Noyer, J. L. (2005). Use of cpSSRs for the characterization of yam phylogeny in Benin. *Genome*, *48*(4), 674-684.Clement, C. R. (1999). 1492 and the loss of Amazonian crop genetic resources. I. The relation between domestication and human population decline. *Economic Botany*, *53*(2), 188.

Cornago, D. F., Rumbaoa, R. G. O., & Geronimo, I. M. (2011). Philippine Yam (*Dioscorea* spp.) tubers phenolic content and antioxidant capacity. *Philippine Journal of Science*, *140*(2), 145-152.

Couto, R. S., Martins, A. C., Bolson, M., Lopes, R. C., Smidt, E. C., & Braga, J. M. A. (2018). Time calibrated tree of *Dioscorea* (*Dioscoreaceae*) indicates four origins of yams in the Neotropics since the Eocene. *Botanical Journal of the Linnean Society*, *188*(2), 144-160.

Croxton, M. D., Andreu, M. A., Williams, D. A., Overholt, W. A., & Smith, J. A. (2011). Geographic origins and genetic diversity of air-potato (*Dioscorea bulbifera*) in Florida. *Invasive Plant Science and Management*, *4*(1), 22-30.

Cuénoud, P., Savolainen, V., Chatrou, L. W., Powell, M., Grayer, R. J., & Chase, M. W. (2002). Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid *rbcL*, *atpB*, and *matK* DNA sequences. *American Journal of Botany*, *89*(1), 132–144. https://doi.org/10.3732/ajb.89.1.132.

Dabonne, S., Kouakou, D. J. E., Martin, K., Hubert, K., & Kouame, L. P. (2011). Enzymatic profile of different parts of yam (*Dioscorea*) tuber during storage. *Agriculture and Biology Journal of North America*, *2*(4), 591-597.

Dansi, A., Orobiyi, A., Dansi, M., Assogba, P., Sanni, A., & Akpagana, K. (2013). Sélection de sites pour la conservation in situ des ignames sauvages

apparentées aux ignames Cultivées: Cas de *Dioscorea praehensilis* Au Bénin. *International Journal of Biological and Chemical Sciences*, *7*(1), 60-74.

DeSalle, R., Egan, M. G., & Siddall, M. (2005). The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1462), 1905-1916.

Deschamps, S., Llaca, V., & May, G. D. (2012). Genotyping-by-sequencing in plants. *Biology*, *1*(3), 460-483.

Djedatin, G., Perrier, X., Agbangla, C., & Noyer, J. L. (2006). *Relationship between wild and cultivated yams:* Case study of yam domestication in Benin.

Drouin, G., Daoud, H., & Xia, J. (2008). Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Molecular Phylogenetics and Evolution*, *49*(3), 827-831.

Duffy, A. M., Kelchner, S. A., & Wolf, P. G. (2009). Conservation of selection on *matK* following an ancient loss of its flanking intron. *Gene*, *438*(1-2), 17-25.

Efisue, A. A. (2015). Genetic diversity Study of *Dioscoreas* Using Morphological Traits and Isozyme Markers Analyses. *Nigerian Journal of Biotechnology*, *30*(1), 7-17.

Excoffier, L., Laval, G., & Schneider, S. (2005). Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics*, *1*, 117693430500100. https://doi.org/10.1177/117693430500100003.

Eyitayo, O. A., Anthony, T. O., & Theresas, I. (2010). Economics of seed yam production using minisett technique in Oyo State, Nigeria. *Field Actions Science Reports. The Journal of Field Actions*, *4*.

Fatima, T., Srivastava, A., Somashekar, P. V., Hanur, V. S., & Rao, M. S. (2019). Development of DNA-based species identification and barcoding of three important timbers. *Bulletin of the National Research Centre*, *43*(1), 76.

Food and Agricultural Organization. (2009). Country Repot on the State of Plant Genetic Resources for Food and Agricuture. *Kenya Agricultural Research Institute, National Genebank of Kenya*. Retrieved from http://www.fao.org

Frossard, E., Aighewi, B. A., Aké, S., Barjolle, D., Baumann, P., Bernet, T., & Ilboudo, L. J. (2017). The challenge of improving soil fertility in yam cropping systems of West Africa. *Frontiers in Plant Science*, *8*, 1953.

Fu, R. H. Y., Kikuno, H., & Maruyama, M. (2011). Research on yam production, marketing and consumption of Nupe farmers of Niger State, central Nigeria. *African Journal of Agricultural Research*, *6*(23), 5301–5313. https://doi.org/10.5897/AJAR11.586.

Girma, G., Spillane, C., & Gedil, M. (2016). DNA barcoding of the main cultivated yams and selected wild species in the genus *Dioscorea*. *Journal of Systematics and Evolution*, *54*(3), 228-237.

Gruère, G., Giuliani, A., & Smale, M. (2006). *Marketing underutilized plant species for the benefit of the poor: a conceptual framework*. Intl Food Policy Res Inst.

Guan, Y. (2014). Detecting structure of haplotypes and local ancestry. *Genetics*, *196*(3), 625–642. https://doi.org/10.1534/genetics.113.160697.

Harlan, J. R. (1992). *Crops and Man* (No. Ed. 2). American Society of Agronomy.

Herrmann, F., & Wink, M. (2014). Use of *rbcL* sequences for DNA barcoding and authentication of plant drugs used in Traditional Chinese Medicine, (January). https://doi.org/10.7287/peerj.preprints.196v1.

Hollingsworth, P. M., Graham, S. W., & Little, D. P. (2011). Choosing and using a plant DNA barcode. *PLoS One*, *6*(5), e19254.

Ji, H., Shengji, P., & Chunlin, L. (2004). An ethnobotanical study of medicinal plants used by the Lisu people in Nujiang, northwest Yunnan, China. *Economic Botany*, *58*(1), S253-S264.

Kang, Y., Deng, Z., Zang, R., & Long, W. (2017). DNA barcoding analysis and phylogenetic relationships of tree species in tropical cloud forests. *Scientific Reports*, *7*(1), 1-9.

Kariuki, S. M. (2012). *Molecular phylogeny of Kenya Dioscorea species (Yams) and the quantification of their Dioscin levels* (Doctoral dissertation, University of Nairobi, Kenya).

Kartavtsev, Y. P. (2011). Divergence at Cyt-b and Co-1 mtDNA genes on different taxonomic levels and genetics of speciation in animals. *Mitochondrial DNA*, *22*(3), 55–65. https://doi.org/10.3109/19401736.2011.588215.

Kress, W. J., & Erickson, D. L. (2007). A Two-Locus Global DNA Barcode for Land Plants: The Coding *rbcL* Gene Complements the Non-Coding trnH-psbA Spacer Region. *PLoS One*, *2*(6). https://doi.org/10.1371/journal.pone.0000508.

Kulasinghe, W. M. A. A., Wimalasiri, K. M. S., Samarasinghe, G., Silva, R., & Madhujith, T. (2018). Macronutrient composition of selected traditional yams grown in Sri Lanka. *Tropical Agricultural Research*, *29*(2), 113-122.

Kusi, F., Asante, S. K., Adjebeng-Danquah, J., Nutsugah, S. K., Buah, S. S. J., Owusu, R. K., Wiredu, A. N., Sugri, I. & Zakaria, M. (2013). Participatory integrated pest management strategy for improving shelf-life of yam (*Dioscorea* spp.). *International Journal of Advance Agriculture Research*, *1*, 124-132.

Lahaye, R., Van der Bank, M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G., & Savolainen, V. (2008). DNA barcoding the floras of biodiversity hotspots. *Proceedings of the National Academy of Sciences*, *105*(8), 2923-2928.

Laiou, A., Mandolini, L. A., Piredda, R., Bellarosa, R., & Simeone, M. C. (2013). DNA barcoding as a complementary tool for conservation and valorization of forest resources. *ZooKeys*, (365), 197.

Li, Y., Feng, Y., Wang, X. Y., Liu, B., & Lv, G. H. (2014). Failure of DNA barcoding in discriminating *Calligonum* species. *Nordic Journal of Botany*, *32*(4), 511-517.

Liu, J. I. E., Moeller, M., Gao, L. M., Zhang, D. Q., & LI, D. Z. (2011). DNA barcoding for the discrimination of Eurasian yews (*Taxus* L., *Taxaceae*) and the discovery of cryptic species. *Molecular Ecology Resources*, *11*(1), 89-100.

Loko, Y. L., Bhattacharjee, R., Agre, A. P., Dossou-Aminon, I., Orobiyi, A., Djedatin, G. L., & Dansi, A. (2017). Genetic diversity and relationship of Guinea yam (*Dioscorea cayenensis* Lam.–*D. rotundata* Poir. complex) germplasm in Benin (West Africa) using microsatellite markers. *Genetic Resources and Crop Evolution*, *64*(6), 1205-1219.

Lopez-Montes, A., Bhattacharjee, R., & Girma Tessema, G. (2012). Yam breeding at IITA: achievements, challenges, and prospects.

Magwé-Tindo, J., Wieringa, J. J., Sonké, B., Zapfack, L., Vigouroux, Y., Couvreur, T. L., & Scarcelli, N. (2019). Complete plastome sequences of 14 African yam species (*Dioscorea* spp.). *Mitochondrial DNA Part B*, *4*(1), 74-76.

Magwé-Tindo, J., Wieringa, J. J., Sonké, B., Zapfack, L., Vigouroux, Y., Couvreur, T. L., & Scarcelli, N. (2018). Guinea yam (*Dioscorea* spp., *Dioscoreaceae*) wild relatives identified using whole plastome phylogenetic analyses. *Taxon*, *67*(5), 905-915.

Maloukh, L., Kumarappan, A., Jarrar, M., Salehi, J., El-wakil, H., & Rajya Lakshmi, T. V. (2017). Discriminatory power of *rbcL* barcode locus for authentication of some of United Arab Emirates (UAE) native plants. *3 Biotech*, *7*(2), 1–7. https://doi.org/10.1007/s13205-017-0746-1.

Maneenoon, K., Sirirugsa, P., & Sridith, K. (2008). Ethnobotany of *Dioscorea* L. (*Dioscoreaceae*), a major food plant of the Sakai tribe at Banthad Range, Peninsular Thailand. *Ethnobotany Research and Applications*, *6*, 385-394.

Men, Q., Xue, G., Mu, D., Hu, Q., & Huang, M. (2017). Mitochondrial DNA markers reveal high genetic diversity and strong genetic differentiation in populations of *Dendrolimus kikuchii* Matsumura (*Lepidoptera: Lasiocampidae*). *PLoS One*, *12*(6), e0179706. https://doi.org/10.1371/journal.pone.0179706.

Mengesha, W. A., Demissew, S., Fay, M. F., Smith, R. J., Nordal, I., & Wilkin, P. (2013). Genetic diversity and population structure of Guinea yams and their wild relatives in South and South West Ethiopia as revealed by microsatellite markers. *Genetic Resources and Crop Evolution*, *60*(2), 529-541.

Mignouna, H. D., Abang, M. M., Asiedu, R., & Geeta, R. (2009). Yam (*Dioscorea*) husbandry: Cultivating yams in the field or greenhouse. *Cold Spring Harbor Protocols*, *2009*(11), pdb-prot5324.

Mignouna, H. D., Dansi, A., & Zok, S. (2002). Morphological and isozymic diversity of the cultivated yams (*Dioscorea cayenensis/Dioscorea rotundata complex*) of Cameroon. *Genetic Resources and Crop Evolution*, *49*(1), 21–29. https://doi.org/10.1023/A:1013805813522.

Mignouna, H., Abang, M., & Asiedu, R. (2008). Genomics of yams, a common source of food and medicine in the tropics. *Genomics of Tropical Crop Plants*, 549–570. https://doi.org/10.1007/978-0-387-71219-2_23.

Moritz, C. C., & Potter, S. (2013). The importance of an evolutionary perspective in conservation policy planning. *Molecular Ecology*, *22*(24), 5969–5971. https://doi.org/10.1111/mec.12565.

Müller, K. F., Borsch, T., & Hilu, K. W. (2006). Phylogenetic utility of rapidly evolving DNA at high taxonomical levels: contrasting *matK*, *trnT-F*, and *rbcL* in basal angiosperms. *Molecular Phylogenetics and Evolution*, *41*(1), 99-117.

Mulualem, T., Mekbib, F., Hussein, S., & Gebre, E. (2018). Analysis of Biochemical Composition of Yams (*Dioscorea* spp.) Landraces from Southwest Ethiopia.

Agrotechnology, *7*, 177.

Mulualem, T., Mekbib, F., Shimelis, H., Gebre, E., & Amelework, B. (2018). Genetic diversity of yam (*Dioscorea* spp.) landrace collections from Ethiopia using simple sequence repeat markers. *Australian Journal of Crop Science*, *12*(8), 1222.

Muthamia, Z. K., Morag, F. E., Nyende, A. B., Mamati, E. G., & Wanjala, B. W. (2013). Estimation of genetic diversity of the Kenyan yam (*Dioscorea* spp.) using microsatellite markers. *African Journal of Biotechnology*, *12*(40).

Naciri, Y., & Linder, H. P. (2015). Species delimitation and relationships: the dance of the seven veils. *Taxon*, *64*(1), 3-16.

Naciri, Y., Caetano, S., & Salamin, N. (2012). Plant DNA barcodes and the influence of gene flow. *Molecular Ecology Resources*, *12*(4), 575-580.

Ngwe, M. F. S. N., Omokolo, D. N., & Joly, S. (2015). Evolution and Phylogenetic Diversity of Yam Species (*Dioscorea* spp.): Implication for Conservation and Agricultural Practices. *PLoS One*, *10*(12), 1–13. https://doi.org/10.1371/journal.pone.0145364.

Nixon, K. C., & Wheeler, Q. (1990). An amplification of the phylogenetic species concept. *Cladistics* 6: 211–223. and. 1992. Extinction and the origin of species.

Nubankoh, P., Pimtong, S., Somta, P., Dachapak, S., & Srinives, P. (2015). Genetic diversity and population structure of pencil yam (*Vigna lanceolata*) (*Phaseoleae, Fabaceae*), a wild herbaceous legume endemic to Australia, revealed by microsatellite markers. *Botany*, *93*(3), 183-191.

Obidiegwu, J. E., & Akpabio, E. M. (2017). The geography of yam cultivation in southern Nigeria: Exploring its social meanings and cultural functions. *Journal of Ethnic Foods*, *4*(1), 28-35.

Obidiegwu, J. E., Lyons, J. B., & Chilaka, C. A. (2020). The *Dioscorea* Genus (Yam)—an Appraisal of Nutritional and Therapeutic Potentials. *Foods*, *9*(9),

1304.

Ojuederie, O. B., Balogun, M. O., Fawole, I., Igwe, D. O., & Olowolafe, M. O. (2014). Assessment of the genetic diversity of African yam bean (*Sphenostylis stenocarpa* Hochst ex. A Rich. Harms) accessions using amplified fragment length polymorphism (AFLP) markers. *African Journal of Biotechnology*, *13*(18).

Okwu, D. E., & Ndu, C. U. (2006). Evaluation of the phytonutrients, mineral and vitamin contents of some varieties of yam (*Dioscorea* spp.). *International Journal of Molecular Medicine and Advance Science*, *2*(2), 199-203.

Oli, M. T. (2006). *Assessing diversity in yams (Dioscorea* spp.*) from Ethiopia based on morphology, AFLP markers and tuber quality, and farmers' management of landraces*. Cuvillier Verlag.

Osunde, Z. D. (2008). Minimizing postharvest losses in Yam (*Dioscorea* spp.): Treatments and techniques. *Using food science and technology to improve nutrition and promote national development. International Union of Food Science and Technology*, 66-88.

Otegbayo, B., Oguniyan, D., & Akinwumi, O. (2014). Physicochemical and functional characterization of yam starch for potential industrial applications. *Starch-Stärke*, *66*(3-4), 235-250.

Otoo, E. (2017). Yam breeding in Ghana. *Journal of Agricultural Science*, *9*(10), 122-136.

Otoo, E., Anyakanmi, T. G., Kikuno, H., & Asiedu, R. (2016). In vivo yam (*Dioscorea* spp.) vine multiplication technique: the plausible solution to seed yam generation menace. *Journal of Agricultural Science*, *8*(2), 88-97.

Ousmael, K. M., Tesfaye, K., & Hailesilassie, T. (2019). Genetic diversity assessment of yams (*Dioscorea* spp.) from Ethiopia using inter simple sequence repeat (ISSR) markers. *African Journal of Biotechnology*, *18*(30), 970-977.

Paschoalin, V. M. F. (2015). DNA barcoding assessment of the genetic diversity of varieties of taro , Colocasia esculenta (L .) Schott in Brazil. *Breeding and Genetic Engineering - The Biology and Biotechnology Research*, (September), 33.

Ondo O. P., Kevers, C., & Dommes, J. (2016). Effects of planting methods and tuber weights on growth and yield of yam cultivars (*Dioscorea rotundata Poir*.) in Gabon. *International Research Journal of Agricultural Science and Soil Science*, *6*, 32-42.

Pickersgill, B. (2007). Domestication of plants in the Americas: insights from Mendelian and molecular genetics. *Annals of Botany*, *100*(5), 925-940.

Price, E. J., Bhattacharjee, R., Lopez-Montes, A., & Fraser, P. D. (2018). Carotenoid profiling of yams: Clarity, comparisons and diversity. *Food Chemistry*, *259*, 130-138.

Qin, A., Wang, M., Cun, Y., Yang, F., Wang, S., Ran, J., & Wang, Q. (2013). Phylogeographic Evidence for a Link of Species Divergence of Ephedra in the Qinghai-Tibetan Plateau and Adjacent Regions to the Miocene Asian Aridification. *PLoS One*, *8*(2). https://doi.org/10.1371/journal.pone.0056243.

Rivera-Jiménez, H. J., Álvarez-Soto, A., Palacio-Mejía, J. D., Barrios-Leal, D. Y., & López-Álvarez, D. (2011a). Genetic diversity intra and inter-specific yam (*Dioscorea* spp.) from the Colombian Caribbean region by AFLP markers. *Acta Agronómica*, *60*(4), 328-338.

Rivera-Jiménez, H. J., Álvarez-Soto, A., Palacio-Mejía, J. D., Barrios-Leal, D. Y., & López-Álvarez, D. C. (2011b). Intra and interspecific genetic diversity of yam (*Dioscorea* spp.) from the Colombian Caribbean region by AFLP markers. *Acta Agronómica*, *60*(4), 325–333.

Rogalski, M., do Nascimento Vieira, L., Fraga, H. P., & Guerra, M. P. (2015). Plastid genomics in horticultural species: importance and applications for plant population genetics, evolution, and biotechnology. *Frontiers in Plant*

Science, *6*, 586.

Salawu, M. B., Ibrahim, A. G., Lamidi, L. O., Salau, M. A., & Ogunleye, B. T. (2014). Economic analysis of yam processing in Oyo State. *Economic Analysis*, *6*(39).

Sangakkara, Ravi; Frossard, E. (2014). Home gardens and *Dioscorea* species – A case study from the climatic zones of Sri Lanka. *Journal of Agriculture and Rural Development in the Tropics and Subtropics*, *115*(1), 55–65. Retrieved from http://www.jarts.info/index.php/jarts/article/view/2014020344898/807

Santosa, E., Lian, C. L., Sugiyama, N., Misra, R. S., Boonkorkaew, P., & Thanomchit, K. (2017). Population structure of elephant foot yams (*Amorphophallus paeoniifolius* (Dennst.) Nicolson) in Asia. *PLoS One*, *12*(6).

Scarcelli, N., Barnaud, A., Eiserhardt, W., Treier, U. A., Seveno, M., d'Anfray, A., … Pintaud, J. C. (2011). A set of 100 chloroplast DNA primer pairs to study population genetics and phylogeny in monocotyledons. *PLoS One*, *6*(5), 1–8. https://doi.org/10.1371/journal.pone.0019954.

Scarcelli, N., Cubry, P., Akakpo, R., Thuillet, A. C., Obidiegwu, J., Baco, M. N., & Mariac, C. (2019). Yam genomics supports West Africa as a major cradle of crop domestication. *Science Advances*, *5*(5), eaaw1947.

Schaal, B. A., Hayworth, D. A., Olsen, K. M., Rauscher, J. T., & Smith, W. A. (1998). Phylogeographic studies in plants: problems and prospects. *Molecular Ecology*, *7*(4), 465-474.

Seal, S., Turaki, A., Muller, E., Kumar, P. L., Kenyon, L., Filloux, D., & Iskra-Caruana, M. L. (2014). The prevalence of badnaviruses in West African yams (*Dioscorea cayenensis-rotundata*) and evidence of endogenous pararetrovirus sequences in their genomes. *Virus Research*, *186*, 144-154.

Sheela, M. N., Abhilash, P. V., Asha, K. I., & Arnau, G. (2014, August). Genetic diversity analysis in greater yam (*Dioscorea alata* L.) native to India using

morphological and molecular markers. In *XXIX International Horticultural Congress on Horticulture: Sustaining Lives, Livelihoods and Landscapes (IHC2014): 1118* (pp. 51-58).

Siadjeu, C., Mayland-Quellhorst, E., & Albach, D. C. (2018). Genetic diversity and population structure of trifoliate yam (*Dioscorea dumetorum Kunth*) in Cameroon Revealed by Genotyping-By-Sequencing (GBS). *BMC Plant Biology*, *18*(1), 359.

Sonibare, M. A., Asiedu, R., & Albach, D. C. (2010). Genetic diversity of *Dioscorea dumetorum* (Kunth) Pax using amplified fragment length polymorphisms (AFLP) and cpDNA. *Biochemical Systematics and Ecology*, *38*(3), 320-334.

Silva, L. R., Mezette, T. F., Nascimento, W. F. D., Silva, E. F., & Veasey, E. A. (2017). Spatially structured morphological and molecular diversity among D*ioscorea cayenensis* and *D. rotundata* yam accessions. *Plant Genetic Resources*, *15*(4), 296-309.

Sun, X. Q., Zhu, Y. J., Guo, J. L., Peng, B., Bai, M. M., & Hang, Y. Y. (2012). DNA barcoding the *Dioscorea* in China, a vital group in the evolution of monocotyledon: Use of *matK* gene for species discrimination. *PLoS One*, *7*(2), e32057.

Tamiru, M., Natsume, S., Takagi, H., White, B., Yaegashi, H., Shimizu, M., & Urasaki, N. (2017). Genome sequencing of the staple food crop white Guinea yam enables the development of a molecular marker for sex determination. *BMC Biology*, *15*(1), 86.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, *30*(12), 2725–2729. https://doi.org/10.1093/molbev/mst197.

Techen, N., Parveen, I., & Khan, I. A. (2017). A single molecular marker to distinguish between species of *Dioscorea Genome*, *60*(3), 201-207.

Terauchi, R., Chikaleke, V. A., Thottappilly, G., & Hahn, S. K. (1992). Origin and phylogeny of Guinea yams as revealed by RFLP analysis of chloroplast DNA and nuclear ribosomal DNA. *Theoretical and Applied Genetics*, *83*(6-7), 743-751.

Terhorst, J., Kamm, J. A., & Song, Y. S. (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nature Genetics*, *49*(2), 303.

Ude, G. N., Igwe, D. O., McCormick, J., Ozokonkwo, O., Harper, J., Ballah, D., & Chiezey, V. O. (2019). Genetic Diversity and DNA Barcoding of Yam Accessions from Southern Nigeria. *American Journal of Plant Sciences*, *10*, 179-207.

Udensi, E. A., Oselebe, H. O., & Onuoha, A. U. (2010). Antinutritional assessment of D. alata varieties. *Pakistan Journal of Nutrition*, *9*(2), 179-181.

Verter, N., & Bečvářová, V. (2015). An analysis of yam production in Nigeria. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, *63*(2), 659–665. https://doi.org/10.11118/actaun201563020659.

Kumari, V. (2012). Simplified extraction of good quality genomic DNA from a variety of plant materials. *African Journal of Biotechnology*, *11*(24), 6420–6427. https://doi.org/10.5897/AJB11.2366.

Wambugu, P. W. (2009). The State of Plant Genetic Resources for Food and Agriculture in Kenia. *Submitted to FAO Commission on Plant Genetic Resources for Food and Agriculture*.

Washaya, S., Mupangwa, J. F., & Muranda, E. (2016). Nutritional Value and Utilization of Yams (*Dioscorea steriscus*) by Residents of Bindura Town High Density Suburbs, Zimbabwe. *Advances in Agriculture*, *2016*.

Wattoo, J. I., Saleem, M. Z., Shahzad, M. S., Arif, A., Hameed, A., & Saleem, M. A. (2016). DNA Barcoding: Amplification and sequence analysis of *rbcL* and

matK genome regions in three divergent plant species. *Advancements in Life Sciences*, *4*(1), 03-07.

Wicke, S., Schneeweiss, G. M., Depamphilis, C. W., Müller, K. F., & Quandt, D. (2011). The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Molecular Biology*, *76*(3-5), 273-297.

Wilkin, P., Muasya, A. M., Banks, H., Furness, C. A., Vollesen, K., Weber, O., & Demissew, S. (2018). A New Species of Yam from Kenya , *Dioscorea kituiensis* : Pollen Morphology , Conservation Status , and Speciation Author (s): Paul Wilkin , A . Muthama Muasya , Hannah Banks , Carol A . Furness , Kai Vollesen , Odile Weber and Sebsebe Demissew Published by : American Society of Plant Taxonomists Stable URL : https://www.jstor.org/stable/40540496 REFERENCES Linked references are available on JSTOR for this article : You may need to log in to JSTOR to access the linked references . A New Species of Yam from Kenya , *Dioscorea kituiensis* : Pollen Morphology , Conservation Status , and Speciation, *34*(4), 652–659.

Wu, W., Chen, C., Zhang, Q., Ahmed, J. Z., Xu, Y., Huang, X., & Huang, D. (2019). A comparative assessment of diversity of greater yam (*Dioscorea alata*) in China. *Scientia Horticulturae*, *243*, 116-124.

Wu, Z. G., Li, X. X., Lin, X. C., Jiang, W., Tao, Z. M., Mantri, N., & Bao, X. Q. (2014). Genetic diversity analysis of yams (*Dioscorea* spp.) cultivated in China using ISSR and SRAP markers. *Genetic Resources and Crop Evolution*, *61*(3), 639-650.

Wumbei, A., Bawa, J. K. A., Akudugu, M. A., & Spanoghe, P. (2019). Absence of Effects of Herbicides Use on Yam Rots: A Case Study in Wulensi, Ghana. *Agriculture*, *9*(5), 95.

Wunderlin, R. P., & Hansen, B. F. (2008). Atlas of Florida Vascular Plants Institute for Systematic Botany. *Florida Center for Community Design and Research.*

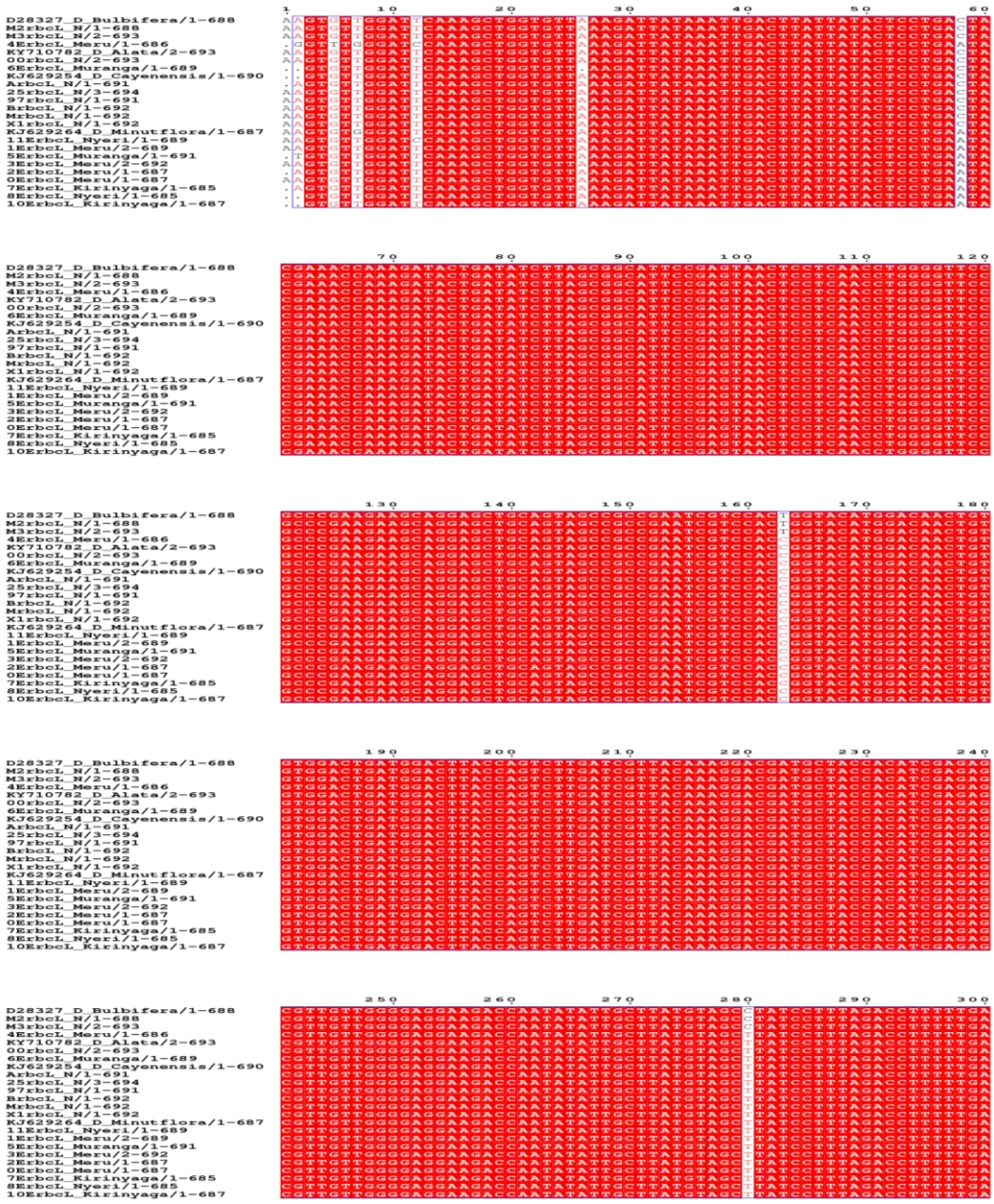
University of South Florida, Tampa, FL, USA. Available at: www. plantatlas. usf. edu/(accessed 19 January 2016).

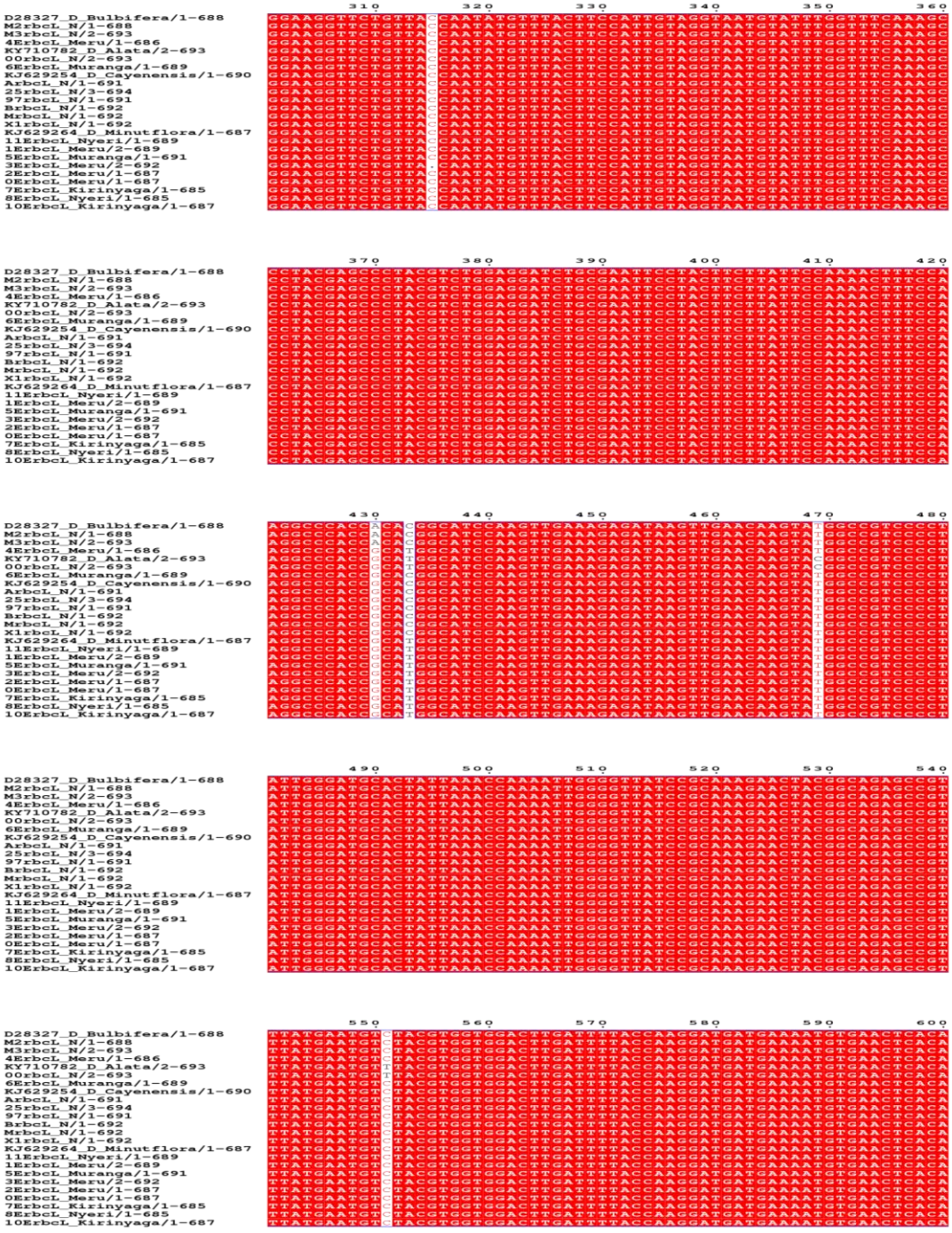
Zhao, Y. J., & Gong, X. (2015). Genetic divergence and phylogeographic history of two closely related species (Leucomeris decora and Nouelia insignis) across the "Tanaka Line" in Southwest China. *BMC Evolutionary Biology*, *15*(1), 1–13. https://doi.org/10.1186/s12862-015-0374-5

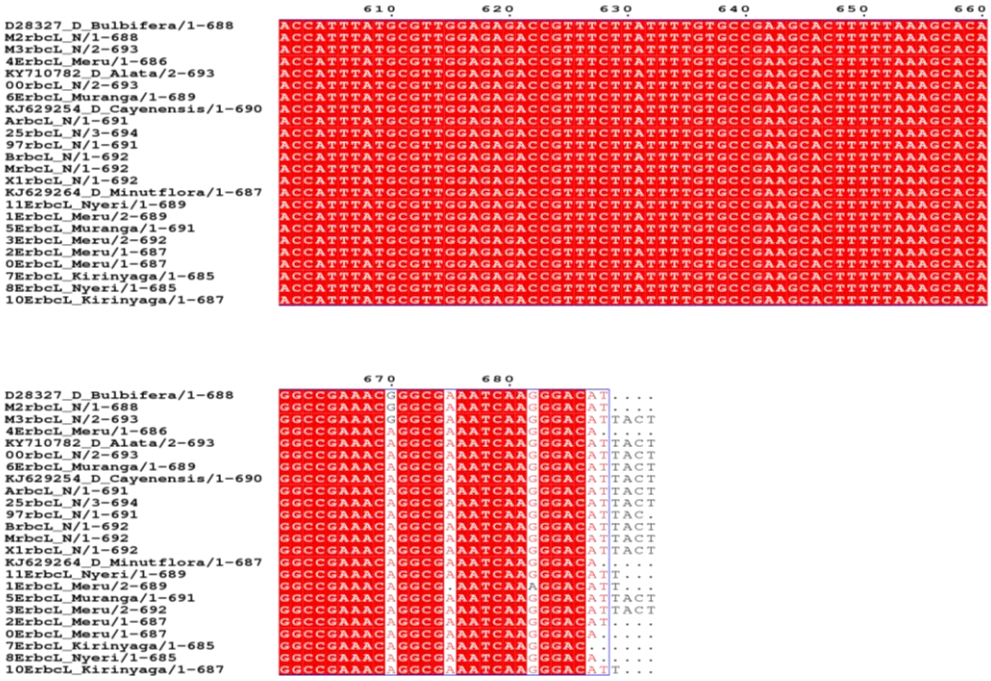
Zhao, Z., Wang, X., Yu, Y., Yuan, S., Jiang, D., Zhang, Y. & Huang, L. (2018). Complete chloroplast genome sequences of *Dioscorea*: Characterization, genomic resources, and phylogenetic analyses. *PeerJ*, *6*, e6032.

APPENDICES

Appendix 1: Multiple sequence alignment of *rbcL* sequences



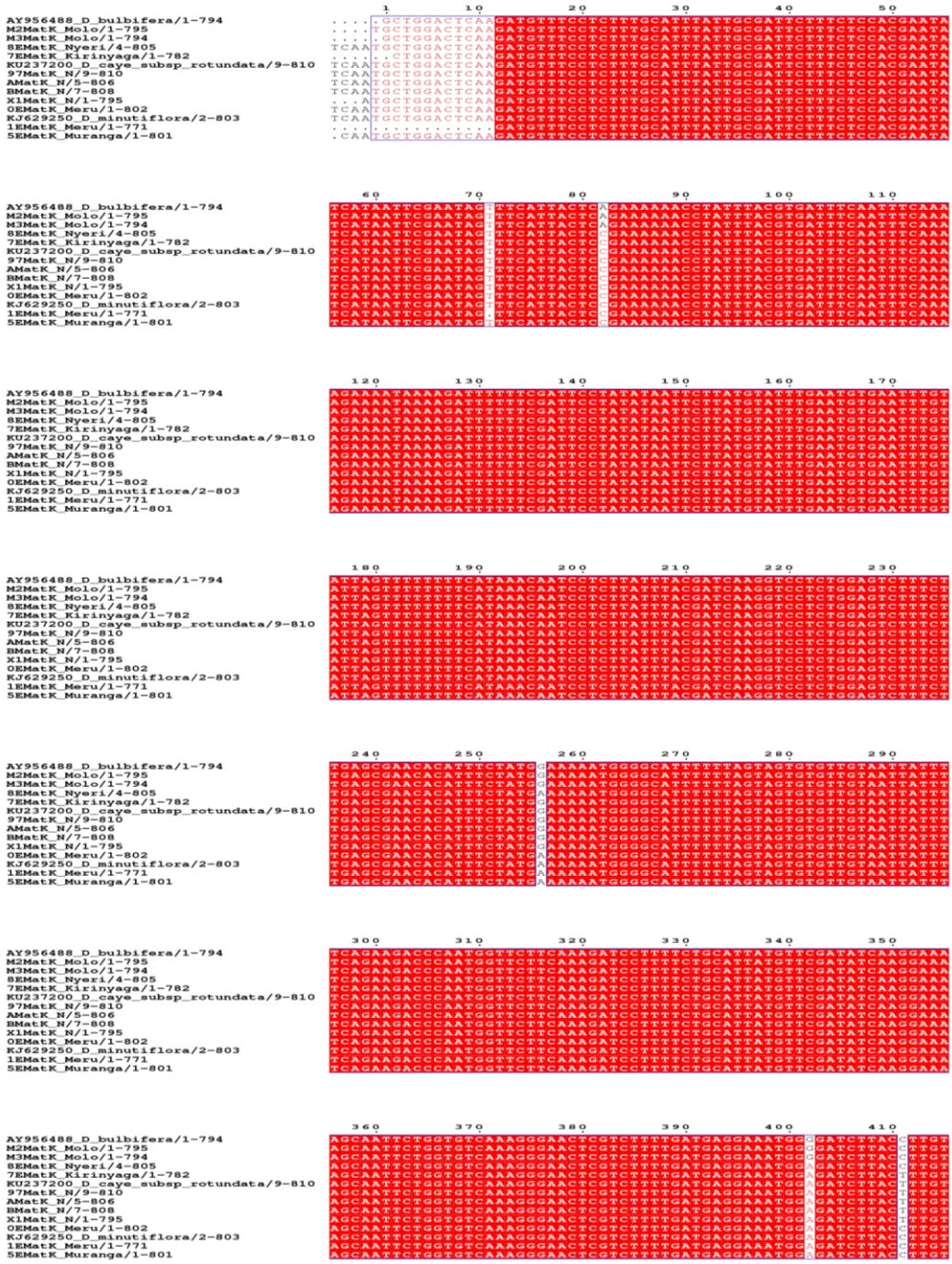

310 320 330 340 350 360
370 380 390 400 410 420
430 440 450 460 470 480
490 500 510 520 530 540
550 560 570 580 590 600
D28327_D_Bulbifera/1-688
M2rbcL_N/1-688
M3rbcL_N/2-693
4ErbcL_Meru/1-686
KY710782_D_Alata/2-693
00rbcL_N/2-693
6ErbcL_Muranga/1-689
KJ629254_D_Cayenensis/1-690
ArbcL_N/1-691
25rbcL_N/3-694
97rbcL_N/1-691
BrbcL_N/1-692
MrbcL_N/1-692
X1rbcL_N/1-692
KJ629264_D_Minutflora/1-687
11ErbcL_Nyeri/1-689
1ErbcL_Meru/2-689
5ErbcL_Muranga/1-691
3ErbcL_Meru/2-692
2ErbcL_Meru/1-687
0ErbcL_Meru/1-687
7ErbcL_Kirinyaga/1-685
8ErbcL_Nyeri/1-685
10ErbcL_Kirinyaga/1-687

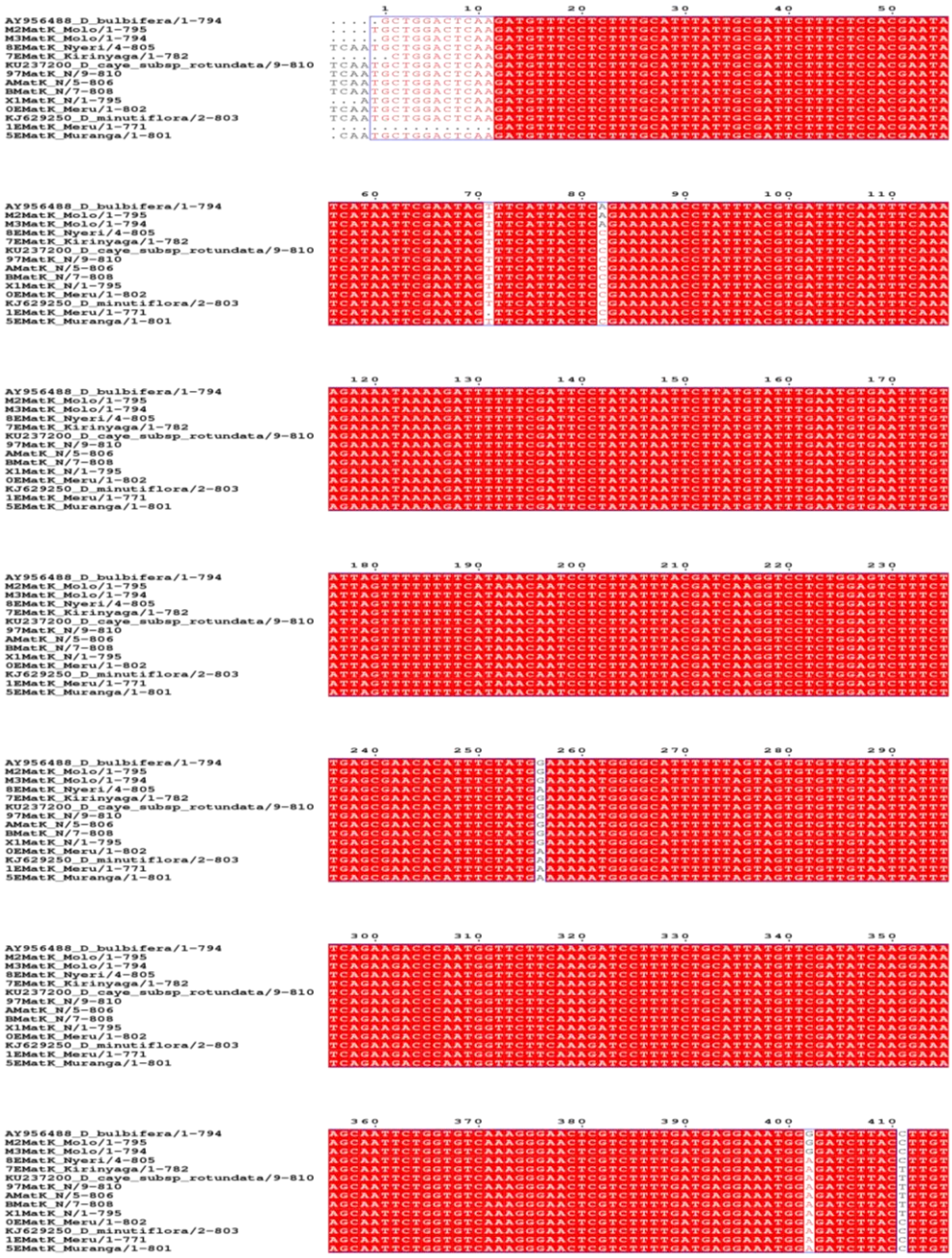


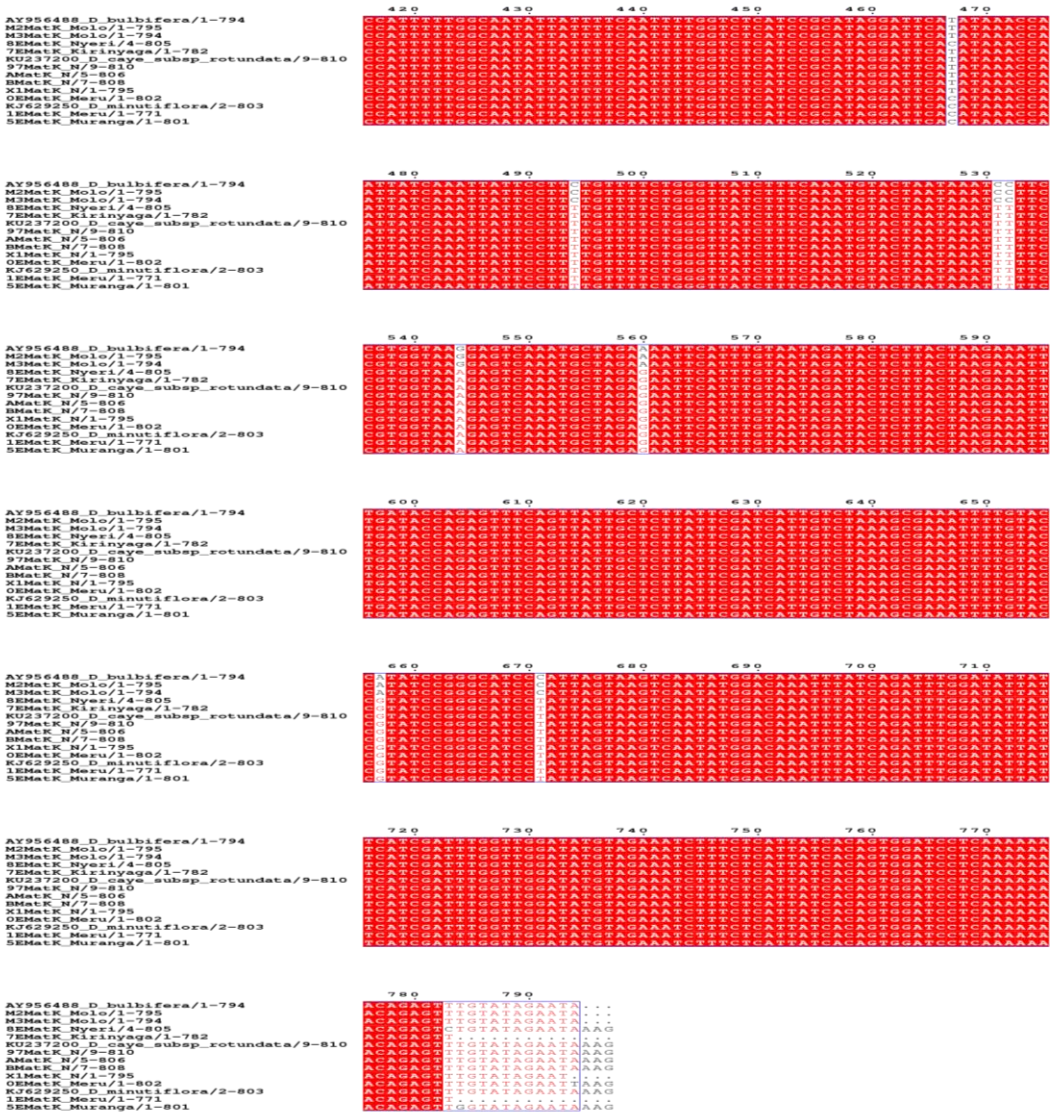
http://espript.ibcp.fr/ESPript/temp/1841381870/0-0-1578501684-esp.300.png
http://espript.ibcp.fr/ESPript/temp/1841381870/0-1-1578501684-esp.300.png
http://espript.ibcp.fr/ESPript/temp/1841381870/0-2-1578501684-esp.300.png

ESPript 3 (http://espript.ibcp.fr); (Robert and Gouet, 2014).

Appendix 2: Multiple sequence alignment of *matK* sequences




AY956488_D_bulbifera/1-794
M2MatK_Molo/1-795
M3MatK_Molo/1-794
8EMatK_Nyeri/4-805
7EMatK_Kirinyaga/1-782
KU237200_D_caye_subsp_rotundata/9-810
97MatK_N/9-810
AMatK_N/5-806
BMatK_N/7-808
X1MatK_N/1-795
0EMatK_Meru/1-802
KJ629250_D_minutiflora/2-803
1EMatK_Meru/1-771
5EMatK_Muranga/1-801



http://espript.ibcp.fr/ESPript/temp/1581429972/0-0-1578502717-esp.300.png

http://espript.ibcp.fr/ESPript/temp/1581429972/0-1-1578502717-esp.300.png

ESPript 3 (http://espript.ibcp.fr); (Robert and Gouet, 2014).