TAXONOMIC IDENTIFICATION OF *ELEUSINE SPP.* USING PLASTID GENES

BY:

WANYONYI SOLOMON NATEMBEA

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN THE SCHOOL OF AGRICULTURE AND BIOTECHNOLOGY

UNIVERSITY OF ELDORET,

KENYA.

2016

DECLARATION

Declaration by the Candidate

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

Wanyonyi, Solomon Natembea: Signature: _____ Date: _____.

AGR/PGB/05/11

Declaration by Supervisors

This thesis has been submitted with our approval as University supervisors.

Dr. Kiplagat K. Oliver Signature: _____ Date:_____.

University of Eldoret

Dr. Odeny A. Damaris Signature: _____ Date:_____.

ICRISAT- Nairobi

Dr. Oduori O. A. Chrispus Signature: _____ Date: _____.

KALRO - Kisii

DEDICATION

I dedicate this work to God Almighty, Who has not only given me the gift of life and wisdom but also the strength and the will to push through the challenges that surrounded this work.

To my children; Ian, Eileen, Alvin and Wayne, you are not just a challenge but a reason to want to live another day. Almighty God is with you always.

ABSTRACT

Finger millet (*Eleusine coracana* subsp. *Coracana*) is a very promising crop to alleviate the problems of food scarcity and malnutrition in the semi-arid tropics and beyond because of its known wide ecological adaptability and nutritional standards yet its production is much less than the demand. This necessitates intensive breeding programmes to improve its production and quality. Analysis and detection of the existing variability and relatedness among cultivated crop species and their immediate wild relatives are important initial steps in breeding. The wild relatives are a source of novel genes that could be exploited in improving the cultivars through introgression among other breeding methods. This necessitated this phylogenetic study. The research involved extracting DNA from two to three weeks old seedlings of 97 wild accessions of genus *Eleusine* from eastern Africa regions of Ethiopia, Uganda, Kenya and Tanzania (believed to be the primary centre of diversity) and 3 cultivated accessions and amplifying the regions of chloroplast DNA using both forward and reverse primers for the three chloroplast barcodes *rpl32-trnL* intergenic-spacer, *ndhF* gene and *rps3* gene. The amplicons were sequenced and the nucleotide sequences of the regions used to discriminate among the genotypes and to construct phylogenetic trees (Phylograms). Evaluation of the three chloroplast barcodes revealed some relatedness among them and further elucidated their overall relatedness to the cultivated finger millet. The study also mainly supported but occasionally refuted previous taxonomic classification (based on phenotypic – cytological and morphological – characters) of some genotypes e.g. accession AAU-ELU-22 was most likely *multiflora* with *97%* bootstrap support and not *intermedia* as previously classified. The results of the primers ndhF and rpl32-trnL generally showed *E. floccifolia, E. jaegeri* and *E. multiflora* being closely related while *E. coracana, E. indica. E. africana* and *E. kigeziensis* formed another clade. This generally forms two clades, the *E. floccifolia* and *E. indica* clades. The *E. indica* clade was more related to the cultivated accessions than the *E. floccifolia* clade. The rps3 primer was less informative.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF PLATES

LIST OF ABBREVIATIONS

BSA – Biological Sequence Analysis

CGIAR – Consultative Group on International Agricultural Research

EDTA – EthyleneDiamineTetraAcetate(acetic acid).

EST – Expressed Sequence Tag

ICRAF – International Centre for Research in AgroForestry

ICRISAT – International Crops Research Institute for the Semi-Arid Tropics

KSC – Kenya Seed Company

LBA – Long Branch Attraction

LCA – Last Common Ancestor

MSA – Multiple Sequence Alignment

MUM – Maximum Unique Match

NRC – National Research Council

OTUs – Operational Taxonomic Units

PCR – Polymerase Chain Reaction

SAGE – Serial Analysis of Gene Expression

SBS – Sequencing By Synthesis

SNP – Single Nucleotide Polymorphism

SSS – Sequence Search Services

TBE – Tris Borate EDTA

TE – Tris EDTA

ACKNOWLEDGEMENTS

First and foremost I thank ICRISAT/Bioinnovate through Dr. Damaris Odeny for accepting me to be attached to their project and funding this work. Secondly, I thank the University of Eldoret for accepting me to undertake my Master's degree and thirdly, to Inqaba Biotech Industries for timely and expert sequencing of the DNA fragments and above all, the sponsorship for one week Phylogenetics Workshop facilitated by Dr Jane Wright that saw me analyse and present the results with ease.

During this time, several people helped me to realize this dream at ICRAF-Nairobi Biotechnology laboratory. I only wish I could thank them individually. Actually, there is no better way to thank you but always know I am grateful.

Ibrahim Shiundu (ICRISAT-Nairobi) for linking me to ICRISAT, Vincent Njunge (ICRISAT – Nairobi) ensuring availability and my access to reagents and apparatus in the Molecular Laboratory, Dr. Paliwal Rajneesh and Annis Saiyori (both of ICRISAT – Nairobi) and Josphene Sarah, for your invaluable inputs at various points.

To all, God Almighty grants you His favours in your lives.

CHAPTER ONE

INTRODUCTION

1.1 Overview

Finger Millet [*Eleusine coracana* (L) Geartn] is a grain cereal grown in Africa and South Asia; ranking fourth in production among millets of the world after Pearl millet (*Pennisetum sp.*), Foxtail millet (*Setaria italica*) and Proso millet (*Panicum mileaceum*). Finger millet has six races (Prasada Rao *et al*.,1993) of which four are cultivated. The cultivated races consist of *Vulgaris, Plana, Compacta* and *Elongata* while the wild races include *Africana* and *Spontanea*. It is a staple grain for much of the world's population, particularly in South Asia and East Africa and it is the sixth most important grain in the world (Bisht and Mukai 2002). It is native to Africa and likely originated from the highlands of Ethiopia or Uganda (Bisht and Mukai 2002).

The current study seeks to understand the phylogeny of the genus *Eleusine* and the relatedness of cultivated finger millet (*Eleusine coracana*) to the wild relatives from eastern Africa in an attempt to trace its origin and possibly unravel the unknown genome donor. The study will also help identify the accessions in their correct taxa in cases where they maybe wrongly placed based on morphological characteristics. These will facilitate improvement of finger millet and assessment and monitoring of local finger millet biodiversity. The study will involve collecting seeds of wild accessions of the genus *Eleusine* from the East African countries of Kenya, Uganda, Tanzania and Ethiopia. The seeds will be planted and DNA will be extracted from 2-3 week seedlings and regions corresponding to ndhF, rpl32-trnL and rps3 amplified so that the sequences could be compared among the accessions and with some cultivated accessions in order to determine their degree of relatedness.

The cost and efforts of the study are worthy since the generated knowledge will be very useful to breeders in various institutions including but not limited to ICRISAT, Kenya Seed Company, Kenya Agricultural Research Institute etc. who may venture into intensive finger millet breeding. The focus is on finger millet because it plays an important role in both the dietary needs (being a staple food, used as a beverage, for cultural purposes) and income generation (fetching over double the price of sorghum

and maize (Oduori 2000)) of many rural households in semi-arid tropics. Nutritionally, it contains high amounts of amino acids deficient in most cereals namely, tryptophan, cysteine, methionine, phenylalanine and tyrosine (Malleshi and Klopfenstein., 1998). It is also a rich source of calcium (Gopalan *et al*., 2002), phosphorus and iron. Therefore finger millet can be a good remedy to malnutrition cases witnessed in many rural communities. The malting qualities of finger millet are second only to barley. Because of this, it could provide cheap and nutritious foods for preventing malnutrition in babies (NRC,1996). The malted (germinated) grain can be used to liquefy any starchy food thus it can be used to make cheap, easily digestible liquid foods for children. Though it was a predominant crop in Africa until recent decades, the crop's production has declined significantly. Ecologically, it has wide adaptations and can withstand adverse soil and weather conditions hence it is an ideal crop for dry areas since its seed can lie dormant in the soil for weeks until soil moisture is high enough (Dida *et al*., 2008). It has good storability – keeping for several years without insect damage because it is small-seeded (Iyengar *et al*., 1945) – hence suitable for food security.

More so finger millet has other uses; for instance the grain is rich in polyphenols such as phenolic acid and its derivatives, flavonoids and tannins. It is also rich in phytic acid – an antinutrient that binds minerals (Hemamaline *et al*., 1980). All these compounds can serve as antioxidants since they have been reported to scavenge on radicals. Antioxidant compounds are gaining importance due to their role in the food industry as lipid stabilisers and in preventive medicine as suppressors of excessive oxidation that causes cancer and ageing (Narmikii 1990). The free radical quenching property of finger millet was studied using ESR as a tool. The results indicate that finger millet is a potent source of antioxidant compounds (Chandra T.S. *et al*., 1996). There are three alpha (α) amylases that can be extracted from finger millet. The amylases cause starch depolymerisation which is the basis for several industrial processes such as the preparation of glucose syrups, bread making and brewing (Nirmala M. *et al*., 2003). The straw also makes good fodder that has up to 61% total digestible nutrients (NRC, 1996). The grain can as well be used as poultry feed. It is only surpassed by barley in "saccharifying" power. Finger millet is considered a delicacy in the communities that produce it. It has a high social value and therefore served to visitors to be impressed and also in festivities.

1.2 Statement of the Problem

Currently the demand for finger millet outstrips production and this is projected to continue unless productivity-enhancing options are developed and adopted. Though most parts of Kenya especially western Kenya are suitable for finger millet growing, its production levels remain low and farmers continue to achieve much lower yields than the potential. Major finger millet production constraint is lack of high yielding well-adapted varieties in dynamic climate and environmental conditions. The challenge to breeders is to develop varieties with desirable traits and adapted over a wide range of environments. When intensive breeding is undertaken, more often accessions may be identified into wrong taxonomic units. This necessitates phylogenetic study to resolve the problem.

1.3 Justification

The intensity of recurrent droughts in the East and Central Africa region has increased the urgency with which national policy makers are considering drought-tolerant crops. Finger millet offers viable options in harsh environment where other crops do poorly since it has wide adaptations (Upadhyaya *et al*., 2007b) and withstands adverse soil and weather conditions (Dida *et al*., 2008). Considering the increasing deficit in finger millet produced, improved high yielding; stable and adaptable varieties must be developed through plant breeding programmes.

However, successful breeding relies on variations among the genotypes and therefore the need to understand the genome and the origin of the crop at hand and its relationship with wild relatives from its centre of diversity which may be a source of novel genes that may be absent in the genome (Dida and Devos 2006). More so, other breeding alternatives including wide hybridization, mutation breeding and genetic engineering may be sought when genes for particular traits of economic and nutritional importance are found missing in the available germplasm including close wild relatives.

1.4 Objectives

1.4.1 Broad objective

The broad objective of this study is to contribute to enhancing breeding of finger millet for improved yield and quality.

1.4.2 Specific Objectives

1. To determine the phylogeny of wild finger millet accessions from the eastern African region.
2. To determine the relatedness of cultivated finger millet with the wild accessions using chloroplast genes.

1.5 Hypotheses

1. There is close relationship among the wild accessions from the east African region.
2. The collected wild accessions are closely related to the cultivated finger millet.

CHAPTER TWO

LITERATURE REVIEW

2.1 Finger Millet

2.1.1 Plant Type

Finger millet is in the family Poaceae/ Gramineae and sub-family Chloridoideae. The most widely accepted scientific name for finger millet is *Eleusine coracana* Gaertn. Others include: *Cynosurus coracanum* Linn; *Eleusine cerealis* Salisb.; *E. spherosperma* Stokes; *E. stricta* Roxb.; and *E. tocussa,* Fresen (Rachie & LeRoy, 1977). The term *Eleusine* is derived from *Eleusis,* an old epic city sacred to Demeter, the Greek deity presiding over agriculture while the term *coracana* is derived from *Kurukkan,* the Singhah name for this grain (Rachie & Leroy, 1977). Initially, different forms of finger millet were identified differently. Those with finely striated grains were railed *E. tocussa*; those with curved spikes *E. coracana*, whereas those with straight spikes as *E. stricta*. Later Hooker (1897) brought them all under *Eleusine coracana* Gaertn., which he considered to be the cultivated form of *E. indica.*

Some common/ local names of finger millet include *bulo* (Luhya), *kal* (Luo), *akuma* (Ateso), *wimbi* (Swahili) and bird's foot, coracana or African millet (English). Others are *Ragi, Nagli, Mandua,* or *Kurukkan* (India); *Dagussa* or *Wimbe* (East Africa); and *Vlil Rouge* or *Coracan* (French).

Finger millet has a C4 photosynthetic pathway just like maize, sorghum and sugarcane. It is a tufted annual crop, 40 – 130cm tall and it takes 2.5 – 6 months to mature. The ear head has a group of digitately arranged (finger-like, hence the name) spikes/panicles. Ayyangar (1932) described three types of ear heads: top-curved with fingers curved in at the tips only hence retaining the central hollow, in-curved with fingers curving in closing up the central hollow and the opens whose long fingers are gaped out giving a funnel shape. Grain colours are dark-brown, light-brown, reddish-brown and white. Grain shapes are round, reniform and ovoid. Three growth habits are noted; the most frequent one being erect, then decumbent and rarely prostrate. Plant colours are green (most common), purple and violet (rare). Ear size can be categorised into small, intermediate and large sizes.

The inflorescence is a terminal whorl with 2 to 10 (average of 5 or 6) spikes (fingers). The lowest spike called 'thumb' is separated from the other fingers by 2 to 5 cm length. In each finger there are about 70 spikelets each with 5 to 7 complete flowers. The spikelets increase in density i.e. number per centimetre length towards the top (Coleman, 1920; Ayyangar, 1932). Plate 2.1 is a picture of a mature inflorescence.

The flower is very small and therefore extremely difficult to manipulate and may require some magnification in order to emasculate the tiny floret. The order of flowering is thus: in spikelet it is from bottom to top, from bigger to smaller flower while in a finger, from top spikelet to bottom. The flowering process takes 6 or 7 to 10 days, most flowers opening on the third day (Ayyangar *et al*., 1932). Flowering is at its maximum when humidity is high between 95 – 99% and temperature range of between 21 – 23^{o}C minimum and 24 – 30.5^{o}C maximum (Chavan *et al*., 1955). This is usually between 1 and 8 a.m. However, high humidity prevents anther dehiscence (Coleman *et al*., 1921) hence maintaining high humidity in the vicinity of the florets for example by keeping a moist chamber around the plant or bagging the head in a light wet wrapper covered with a plastic bag until all heads have exerted their anthers which can then be wiped off with a solution of water and detergent prior to introduction of desired pollen. The optimum photoperiod is 11hrs 30min and a mean daily temperature of 26 – 30^{o}C (Porteres, 1947).

Plate 2.1: Picture of mature inflorescence of finger millet. Yield potential is influenced by the number of grains per spikelet, number of fingers per head, ear size and number of productive tillers per plant. The number of grains per spikelet ranges from 4 to 8. (source: Author, 2014)

2.1.2 Origin of Finger Millet

Early botanists such as DeCandolle (1886) and Watt (1908), and some later geneticists like Narayanaswami (1952) and Mann (1946; 1950) suggest a probable Indian origin of this millet. In India, finger millet is mentioned by Sanskrit writers and referred to as *ragi* or *rajika*. Burkhill (1935) suggested that *E. coracana* is cultivated relative of the wild species, *E. indica* (L.) Gaertn., and that its early selection by man appears to have taken place in India since it has been cultivated for a very long time there; it has a Sanskrit name, *ragi*; it was probably in India when the Aryans reached there; and its decrease in Africa from east to west suggests its introduction from the East (Rachie & LeRoy, 1977). Worth (1937) was of the opinion that it originated in India from where it spread through Arabia, Abyssinia and to the rest of Africa. Mann (1946, 1950) suggested that owing to the widespread cultivation and importance of *ragi* in India, it most probably developed there.

Later, particularly in recent times, botanists favoured the African origin theory or believed the crop originated both in Africa and in India independently for instance

Vavilov (1951) proposed Abyssinian origin; while Chandola (1959) suggested simultaneously development both in Africa and in India.Hitchcock (1950) suggested two forms of *E. coracana;* one form arising by doubling of chromosomes by *E. indica* and another form from selectionfrom wild tetraploid *E. africana* and stated that a wild tetraploid has never been found in Asia. However, Anderson (1960) called attention to the fact that there was the possibility of much earlier movement of primitive agriculture products from Africa through Yemen and the seaward edge of the Arabian Peninsula and into southern India. That this possibility had not been realised earlier is attributed to the fact that these crops bear Sanskrit names and that these millets are of minor importance and had escaped earlier attention (Rachie & LeRoy, 1977).

In more recent times, taxonomists and geneticists using more advanced techniques have studied the species *Eleusinecoracana* and its wild relatives. Kennedy-O'Byrne (1957) and Mehra (1962-63) studied the broad range of diversity in East Africa and a wild tetraploid form, *Eleusine africana*. From these evidences they suggest the origin of *Eleusine coracana* in Ethiopia, or further south, and its later transport to India in pre-Aryan times.

2.1.3 Relationship with Wild Relatives

The genus *Eleusine* is comprised of annual or perennial grasses that commonly occur in the warmer regions, particularly in southern Asia and Africa. However, one species, *E. indica*, is widely distributed in Europe and also in the New World. The species *E. Africana* (AABB), *E. coracana* (AABB) and *E. kigeziensis* (AADD) are tetraploids while *E. floccifolia* (BB), *E. indica* (AA), *E. tristachya* (AA) and *E. intermedia* (AB) are diploids. With the exception of *E. coracana,* all the other species are wild types. Hitchcock (1950) has described *Eleusine indica,* commonly referred to as goosegrass, as branching at the base; ascending to prostrate; very smooth; culms compressed, usually less than 50cm long; blades flat or folded, 38mm wide; 2 to 6 spikes. He further proposes that doubling of chromosomes of *E. indica* resulted in *E. coracana* with *E. africana* as intermediate. An example of perennial *Eleusine* species *E. compressa* has been described by Bor (LC) as stem creeping and rooting at the nodes, glabrous, basal sheaths woolly at base.

Mehra (1962) studied a collection of 35 plants from a farmer's field in Uganda in order to unravel the relationships and evolution of the cultivated species of *E. coracana* in

Africa. He utilized six morphological characters, including width of rachis, width of stem, length of spikelet, length of glume, length of lemma, and width of raceme. From the study we get the following phenotypic descriptions: *Eleusine indica:*2n=18 chromosomes; it has a smaller plant; narrow rachis; thin stems, relatively short glumes and lemma, and spikelets; shattering spikelets, small seeds are enclosed in glumes and thin racemes.*E. africana:* 2n = 36 chromosomes; it has a larger plant but generally similar to E. indica; has wider rachis, thicker stems and longer spikelets, glumes, and lemmas; the spikelets shatter and it has shattering seeds as well.*E. coracana* (African-highland type): 2n -36 chromosomes; longer lemmas, glumes, and spikelets; spikelets are non-shattering and it has plump grains; the seeds are enclosed inside the glumes. *E. coracana* (Afro-Asiatic type): 2n — 36 chromosomes; it has shorter glumes, lemmas and spikelets; has non-shattering spikelets and plump seed; and seeds thresh free from the glumes. Glume length is controlled by G-1, G-2, and G-3 factors, one of which in the dominant condition gives short glumes; any two give medium glumes; and all three produce long glumes.

The eco-geographic adaptation of the two types of *E. coracana* have been described as Afro-Asiatic type growing in drier areas below 4,000 ft. and its distribution in Africa extending from Ethiopia and the Sudan to Kenya, Uganda and Tanzania; whereas the African highland type being present in the same areas, but in damp sites above 4,000 ft. with an extended distribution south through Malawi, Rhodesia, and into the Union of South Africa (Mehra 1962, 63).

2.1.4 The Allotetraploid Origin of Finger Millet

Eleusine coracana is an allotetraploid with 36 chromosomes i.e. x = 9, 2n = 4x = 36 (A Shakoor). Allopolyploids usually arise as a result of hybridization leading to new species (Ellstrand *et al*., 1996; Seehausen, 2004; Paun *et al*., 2009). Approximately 21.8% of grass species have arisen as a result of hybridization (Knobloch, 1968, 1972). With recent advances in molecular phylogeny, the evolutionary consequences of merging parental genomes into polyploidy nuclei have attracted much attention (Ge at al., 1999; Petersen *et al*., 2006; Mason-Gamer *et al*., 2010). Consequently, several phylogenetic relationships of *Eleusine* have been done in order to identify its allotetraploid origin e.g. by Hilu KW *et al*. (2005); Qing Liu *et al*. (2011) etc. Using two low-copy nuclear markers *pepc4* (gene for phosphoenol pyruvate carboxylase) and *EF-1α* (gene for eukaryotic elongation factor 1-alpha), Qing Liu *et al*. (2011)

found that three tetraploid *Eleusine* species *E. africana, E. coracana* and *E. kigeziensis* share a common ancestor with the two diploid species *E. indica* and *E. tristachya* which are considered maternal parents for the allotetraploids.

The *pepc4* data support independent allotetraploid origins for *E. kigeziensis* and the *E. Africana – E. coracana* clade, but both have *E. indica – E. tristachya* clade as the maternal parents. However, the paternal parents involved in the original hybridization events remain unknown (Liu *et al*., 2011). Bisht and Mukai (2000, 2001a, 2002) had indicated that *Eleusine floccifolia* (2n = 2x = 18) could be the BB donor species of *E. coracana*. This has, however, been refuted by others like Neves *et al*. (2005); Devarumath *et al*. (2010) and Liu *et al*. (2011), who still maintain that the BB genome donor species remains unidentified and may possibly be extinct. *Eleusine kigeziensis* (2n = 4x = 36 or 38) (AADD) is the third tetraploid species of the genus. *Eleusine indica* (2n = 2x = 18) (AA) and *E. jaegeri* (2n = 2x = 20) (DD) are proposed to be the wild progenitors of *E. kigeziensis* (Bisht and Mukai 2002; Devarumath *et al*. 2010). On the contrary, Neves *et al*. (2005) proposed *E. kigeziensis* to be an autotetraploid, and *E. indica* as being closely related to *E. kigeziensis* but not the direct genome donor to it.

Genomic *in situ* hybridization (GISH) studies identified *E. indica* (AA genome) and *E. floccifolia* (Forssk.) Spreng. (BB genome) as candidate progenitors for two tetraploids, *E. africana* and *E. coracana* (Bisht and Mukai, 2001*a*, *b*). Phylogenetic analyses of nuclear ribosomal DNA (nrDNA) internal transcribed spacer (ITS) and plastid *trnT-trnF* sequences (Neves *et al.*, 2005) revealed the following: In the ITS locus A clade, *E. indica* was confirmed to be closely related to three allotetraploids (*E. africana*, *E. coracana* and *E. kigeziensis*), and this group, in turn, was sister to *E. tristachya*. However, no diploid was detected as the second genome donor for the B locus. The GISH results showed genetic similarity of chromosomes, but were not useful in elucidating interspecific phylogenetic relationships. Bisht and Mukai (2002) reported similar hybridization signals between *E. coracana* and two hybrid pairs, *E. indica–E. floccifolia* and *E. tristachya–E. floccifolia*, but, so far, the degree of relationship between *E. indica* and *E. tristachya* remains unresolved. Plastid sequence similarity between *E. indica* and *E. tristachya* suggests that *E. tristachya* became separated from its ancestor very recently (Neves *et al.*, 2005).

Three hypotheses are proposed to explain why the paternal parents remain unidentified (1) the paternal parent of *E. kigeziensis* may come from outside of *Eleusine;* (2)

genomes derived from paternal parent may have undergone a rapid evolutionary divergence (or degradation) after allotetraploid speciation and are therefore untraceable or (3) the paternal parents are extinct in the wild (Liu *et al*., 2011). For *E. africana* and *E. coracana*, the paternal parents should be members of the genus, but diverged earliest, so greatly that they are rare to find or are possibly extinct (Wendel, 2000; Petersen *et al*., 2006).

Herbarium specimens at Kew were studied by Mehra (1962, 63) using the metroglyph analyses devised by Anderson (1960) to evaluate the variation pattern amongst the different specimens based on cytological and morphological policies. As a result of his findings, Mehra (1963) suggested a working hypothesis on the origin of cultivated *E. coracana* (L.) Gaertn., based on both cytotaxonomic and morphological analyses. The hypothesis is thus: *E. indica* (2n = 18) crossed with an unknown species of *Eleusine* (also 2n = 18 chromosomes) and produced a wild hybrid, a diploid which doubled its chromosomes and produced a wild tetraploid with 2n = 36 chromosomes. This evolved into *E. africana*, also a tetraploid, from which was selected *E. coracana* with 2n = 36. The stage between *E. africana* and *E. coracana* is characterized by repeated hybridization and continuing introgression of the two types and was known as the African highland type. Other genotypes may have evolved from this African highland type in possessing the glume inhibitory' mechanism, thereby becoming the Afro-Asiatic type. Alternatively, this second Afro-Asiatic type may have evolved directly from *E. indica* by another *Eleusine* species. It appears highly likely that selection by man has played an important role in the final evolutionary stages of either cultivated types (Richie and LeRoy, 1977).

2.1.5 Elucidating the Allotetraploid Nature

From his study of chromosomal complement in *E. coracana* (and several millet species and other crops) Rau (1929) found that *E. coracana* probably had 36 chromosomes. Krishnaswami and Ayyangar (1935) reported on haploid chromosome numbers in metaphase plants endogenous as follows: *E. coracana* = 18; *E. indica* = 9; *E. brevifolia* = 18; and *E. aegyptica* = 17. Secondary pairing was noted in *E. coracana*, *E. brevifolia* and *E. aegyptica*. They concluded that *E. indica* appeared to be a diploid; *E. coracana* and *E. brevifolia* tetraploids, and *E. aegyptica* probably a tetraploid with one pair lost (4 x — 2). The basic number was presumed to be 9.

Later, embryological studies of finger millet at Tamil Nadu (Madras-UA, 1954) saw the chromosome numbers of some additional wild Eleusines being worked out and reported. These are *E. lagopoides* 2n = 36; *E. compressa* 2n = 45; and *E. verticillata* 2n = 36. Another genus which closely resembles finger millet, *Dactyloctenium*, was also investigated. The basic numbers of the caryotypes were different, however, *D. aegypticum* had 2n = 48 and *D. scindicum* also had 2n = 48 chromosomes. Sharma and De (1956) studied morphology of somatic chromosomes and found similar results, *E. indica* had 2n = 18; *E. coracana* had 2n = 36 and *E. aegyptica* had 2n = 45 chromosomes. They noted prominent similarities in karyotypes within this genus and noted few meiotic irregularities in *E. aegyptica*. Meiotic and mitotic observations were made and reported on several varieties of *E. coracana* which had 2n = 36; *E. indica* and *E. verticillata* both with 2n =18; and *Dactyloctenium aegypticum* with 2n = 36 chromosomes (Chandola, 1959). Meiosis was found regular in *Eleusine, Setaria, Pennisetum*, and *D. aegyptium*. Stoma and pollen-grain size were considered to be the criteria of polyploidy in Eleusine and the other millets. Secondary association suggested that *Eleusine* and *Dactyloctenium* had X = 7 while in *Setaria* it was X = 5. It was suggested that *E. coracana* may have originated as an allotetraploid from crosses between parents each with n = 9 chromosomes. Cytological studies revealed regular pairing and formation of eighteen bivalents during meiosis in the sporocytes (micro sporocytes) in *E. coracana* and *E. africana*. Regular bivalent formation during meiosis, the presence of duplicate factors and polymeric factors suggest *E.coracana* to be an allotetraploid. The cultivated and wild types at the tetraploid level cross readily with each other producing a variety of hybrid types (Richie & LeRoy, 1977).

Other confirmatory evidence on the chromosome number in *E. indica* was reported by Singh (1965a, 1965b). Both pollen grain mitosis and root tip cells were used in these studies. Recommendations were made on the use of emergent ear heads to study pollen cell mitosis. The evidence from these genetical and cytological studies suggests the cultivated species of *Eleusine* to be an allotetraploid by the presence of duplicate and polymeric factors. It is quite evident that *E. indica* is an immediate ancestor of the cultivated species; however, there may be intermediate forms such as *E. africana* (Richie & LeRoy, 1977).

2.1.6 Ecological Conditions for Growth

Finger millet has wide adaptations (Upadhyaya *et al*., 2007b) and can withstand adverse soil and weather conditions (Dida *et al*., 2008). It is an ideal crop for dry areas since its seed can lie dormant in the soil for weeks until soil moisture is high enough. The crop takes only 45 days to mature enabling it to escape drought. It grows at an altitude of 0- 2,400 feet above sea level. On average finger millet requires well-distributed moderate rainfall of up to 1000 mm per annum without prolonged droughts. It flourishes in well-drained sandy loam soils that are relatively fertile and a temperature range of 18^0 C minimum and 27^0 C maximum. Its thick adventitious root system makes it ideally suited to these conditions (McMaster 1962).

2.1.7 Major Production Regions

Approximately three quarters of the world's finger millet production is in the regions immediately surrounding Lake Victoria in East Africa and southern India in the south-eastern part of Karnataka. India produces about 40-45% of the total world production and most of the rest is produced in central and eastern Africa.

The most important regions of finger millet cultivation in Africa are in Uganda in the vicinity of Lake Victoria and Lake Kyoga, and between Lake Tanganyika and Lake Victoria in the western, eastern and southern part of the continent. There are also limited areas in the Central African Republic, southern Chad, north-eastern Nigeria, northern and eastern Zaire, the Sudan, Ethiopia, Somaliland, Uganda, Kenya, Rwanda, Burundi, Tanganyika, Malawi, Zambia, Rhodesia, Mozambique and Madagascar (Johnson & Raymond, 1964; Mnyenyembe & Gupta, 1998; Obilana *et al*., 2002). It is the most important cereal crop in Uganda.

2.1.8 Production Constraints

Many factors have constrained finger millet production especially in east African regions. This explains why it production has been very low in spite of its nutritional, economic and social importance. Finger millet is considered a poor peoples' crop hence most growers are low income earners who can't afford inputs; severe lack of government support and an enabling policy environment. Because of this there is limited access to the government extension services and rare or no demonstrations by state departments of Agriculture. More so, demonstrations if any are conducted in more favourable agricultural environments.

Finger millet has been a neglected crop compared to other cereals such as wheat, maize and rice in terms of research etc. Therefore research and extension must be increased. This necessitates a policy change to increase emphasis on the crop; unavailability of seeds of improved varieties. The seeds are found in research institutions and gene banks and less known to farmers. As a result of this, most farmers use their own seeds of traditional varieties (primitive cultivars/ landraces). About only 10% of farmers use seeds from research institutions (especially KSC & KARI), church organizations and NGOs; 10% from unregistered seed growers and only 5% purchase commercial seed; lack of high yielding, well adapted varieties. This is due to poor technology adoption.

There is need to raise awareness of new varieties and technology through existing networks; abiotic stress factors i.e. diseases, weeds and pests in which case there is a notorious one for each factor e.g. for diseases, there is blast caused by a fungus (*Pyricularia grisea)*; for weeds there is wild finger millet (*Eleusine indica)* which looks very similar during the vegetative stage and for pests we have bird predators especially the notorious *Quelea quelea;* competition from other cereal crops such as maize, sorghum and pearl millet as food and barley for brewing. This may be one reason why it has been given low priority; small size of seeds makes it difficult to handle at all stages i.e. planting, threshing, and packaging, marketing and milling. At planting, a very fine seedbed is required hence farmers must work hard and long to prepare the land; scarcity of processing equipment. It is difficult to mill finger millet. Some of the reasons include the following (1) Very small seeds and therefore they need fine screens (2) The seed coat is tightly bound to the endosperm (3) The grain is so soft and friable, hence not easy to remove the outside without crushing the inside.

2.2 Finger Millet Breeding

2.2.1 The Importance of Breeding Finger Millet

Several investigators have noted the importance of breeding finger millet. It had been found to give very high yields under intensive cultivation and the grain also possesses high nutritive value, thereby justifying intensive efforts to evolve improved strains (Gokhale *et al*., 1931). For many years, much varietal improvement had been carried out in India (Iyer, 1954; Chavan & Shendge, 1957), most of which to 1957, resulted from pure line selection of local varieties and releasing several strains.

Although the early efforts on the improvement of finger millet have largely concentrated on the development of pure breeding strains and varieties, Ayyangar (1932) proposed considering the development and improvement of composite varieties. He proposed this technique in the attempt to develop varieties with a much greater range of adaptation to weather and climates prevailing under erratic rain-fed conditions. He further theorized that such populations would contribute at least a small amount of heterosis by virtue of some natural crosses amongst the different genotypes comprising it. Nevertheless, all improved commercial strains that have been released are, insofar as is known, pure line strains, selected from local or other material and/or a combination amongst specific types (Rachie & Leroy, 1977).

2.2.2 Traits to be tapped in Breeding Finger Millet

With the potential that exists with finger millet, plant breeders need to breed for finger millet with desired characteristics/traits (Salasya *et al*., 2009). These traits include the following: High grain and fodder yield; early maturity with robust growth; resistance to diseases especially blast; grain colour – many people prefer dark brown colour to white colour; plant height – medium height is preferred as the plants can resist lodging yet appropriate for harvesting; ear size – medium sized compact ears are preferred; high finger number and high tillering; drought tolerance; uniformity in height – this facilitates harvesting; wide adaptation to local conditions; grain quality traits e.g. high density grains, good palatability, good brewing qualities etc; good storability and viability; adaptation to poor soil and large panicle/ head size.

2.2.3 Some Varieties of Finger Millet in Kenya

Several finger millet genotypes are grown in Kenya. Among them are advanced cultivars and farmers' varieties (landraces). Improved varieties that have been released include U-15 which originated from Uganda, it has a purple pigmentation, short and is high yielding, Gulu E which also originated from Uganda, it is green with no pigmentation, its resistant to blast and lodging and reputed for high yields, Okhale-1 which originated from Nepal, it has purple pigmentation, its resistant to *Striga*, blast and lodging. It is also high yielding (Riley, 1997), P-224 which originated from Uganda, it is green with no plant pigmentation, it is high yielding but susceptible to *Striga*, blast and lodging (Von Brook, 1990), P-283 which originated from Uganda, it has no pigmentation, it has moderate yield but resistant to lodging, Seremi 1, KNE 648, IE 4115 and Ikhule. Landraces usually have local names e.g. Nanjala Brown

which is tall with purple pigmentation, it is susceptible to *Striga,* blast and lodging and has moderate yield (Oduori *et al*., 2008).

2.2.4 Characterization of Finger Millet Germplasm Resources

To facilitate utilization of germplasm by plant breeders, it must be characterised for both qualitative and quantitative traits of agronomic importance. Characterization should be done in standard and commonly understood language so that researchers in different institutes and countries can use the information easily and effectively. An expert committee sponsored by the then International Board for Plant Genetic Resources (IBPGR) formulated a list of descriptors for finger millet (IBPGR 1985). This includes 30 passport data, 45 morphological data and resistance to 32 stress factors (5 abiotic, 27 biotic) for describing finger millet germplasm accessions.

ICRISAT's large collection of finger millet accessions were characterised in batches over the years at the ICRISAT research farm at Patancheru, India, during the rainy seasons. The accessions were also classified into six botanical races following Prasada Rao *et al*. (1993). The races are *Africana, Spontanea, Compacta, Elongata, Plana* and *Vulgaris.Africana* and *Spontanea* are wild races while the others are cultivated. The formation of races is based primarily on spike type, for instance race *Compacta* has high proportion of fisty, compact and incurved spikes; race *Elongata* has high proportion of long-open, pendulous and top-curved spikes; race *Plana* has mostly top-curved spikes while races *Spontanea* and *Vulgaris* have incurved or top-curved spikes. Glume status is the distinguishing trait between cultivated and wild races. The two wild races have mostly prominent glume while cultivated races have all three classes – prominent, non-prominent and medium.

2.3 DNA Barcoding

DNA barcoding is a taxonomic method that uses a short genetic marker in an organism's DNA to identify it as belonging to a particular species. DNA barcoding, based on highly conserved sequence information, provides new tools for systematics (Hebert and Barrett, 2005) and phylogeny (Wyman *et al*., 2004; Leebens-Mack *et al*., 2005; Jansen *et al*., 2006; Hansen *et al*., 2006). DNA barcodes consist of short sequences of DNA between 400 and 800 base pairs that can be routinely amplified by PCR (polymerase chain reaction) and sequenced. When sequences of the barcodes are similar or almost similar then the organisms may be considered to belong to the same species. Deep intra-specific divergences indicate possible new species. A standard

sequence threshold percentage for a new species is set for example at 2.7% divergence – i.e. as percentage of the number of nucleotides that differ in the sequences of that particular segment (barcode).

Morphologically distinguishable taxa may not require barcoding; however, subspecies (ssp.), cultivars (cv.), eco- and morphotypes, mutants, species complex and clones can be diagnosed with molecular barcoding. Barcode of a specimen can be compared with sequences derived from other taxa, and in the case of dissimilarities, species identity can be determined by molecular phylogenetic analyses based on molecular operational taxonomic units – MOTU (Floyd *et al*., 2002). Unlike molecular phylogeny which determines the pattern of relationship, MOTU identifies unknown sample in terms of a pre-existing classification.

The use of DNA or protein sequences to identify organisms was proposed as a more efficient approach than traditional taxonomic practices (Tautz *et al*., 2003; Blaxter *et al*., 2004). For animals, a segment of about 600bp of mitochondrial gene cytochrome oxidase subunit I (CO1) is a widely used barcode in a range of animal groups (Hebert *et al*., 2003), but this locus is unsuitable for use in plants due to its low mutation rate (Kress *et al*., 2005; Cowen *et al*., 2006; Fazekas *et al*., 2008). Moreover, hybridization and polyploidy, common in plants, make it difficult to define species boundaries (Rieseberg *et al*., 2006; Fazekas *et al*., 2009). A chloroplast gene such as matK (maturase K) or a nuclear sequence such as ITS (internal transcribed spacer) may be an effective target for barcoding in plants (Kress *et al*., 2005). Chloroplast genes rbcL and matK were proposed in 2009 as barcodes but recently, other regions are used. The number and identity of DNA sequences that should be used for barcoding in a particular project or endeavour is a matter of debate (Pennisi, 2007; Ledford, 2008) and so is an ideal DNA barcode.

Generally, an ideal DNA barcode should be chosen on the following criteria: Sufficiently variable among species but conserved enough to be less variable within species; standardised with the same region of DNA genome; should contain enough phylogenetic information; extremely robust, with highly conserved priming sites (to design primers) and highly reliable DNA amplifications and sequencing and short enough to allow amplification of degraded DNA (Blaxter, 2004; Kress *et al*., 2005).

Of the multitude of nuclear regions employed for phylogenetic analysis (Grob *et al*., 2004; Small *et al*., 2004), one potential barcode candidate is the nuclear ribosomal DNA internal transcribed spacer region (ITS) which often successfully discriminates species (Stoeckle, 2003; Chase *et al*., 2005; Kress *et al*., 2005). However, plastid genome along with nuclear ITS are most promising loci for plant barcoding (Chase *et al*., 2005; Kress *et al*., 2005). Non-coding regions are more appropriate due to more variations (maximally variable locus) than coding regions because of presumed reduction of functional constraints (Gielly and Taberlet, 1994; Shaw *et al*., 2005).

Even so, unlike the nuclear genome, the plastid genome is uniparentally inherited. In groups of plants that exhibit extensive hybridization and introgression, reliance on a single (usually) maternally inherited gene would be problematic, hence incorporation of multiple nuclear regions is necessary to make confident identifications (Chase *et al*., 2005). The suggestion to include multiple loci (Kress *et al*., 2005) in barcoding systems was welcomed by critics of barcoding (Rubinoff *et al*., 2006). However, a tiered approach is suggested so that instead of simply adding another locus, a first tier coding region common across the land plants provide resolution at a certain rank (e.g. family or genus) and a more variable (coding or non-coding) region at species level. This minimises difficulty in aligning non-coding regions from highly divergent genera (NewMaster *et al*., 2006).

2.4 The Plastid DNA

All plastids are derived from proplastids (formerly ‘eoplastids’) present in the meristematic regions of the plant. There are several types of plastids which perform different function e.g. chloroplasts for photosynthesis (derived from etioplasts); chromoplasts for pigment synthesis and storage (hence coloured depending on the pigment); gerontoplasts that control dismantling of the photosynthetic apparatus during senescence and the colourless leucoplasts for monoterpene synthesis, for storage of starch (amyloplasts), oil (elaioplasts) and protein (proteinoplasts). To evolutionists, plastids are thought to have originated from endosymbiotic cyanobacteria. Their inheritance is generally maternal in angiosperms and paternal in gymnosperms. In normal intraspecific crossing to give normal hybrids of the same species, it is 100% uniparental but in interspecific hybridization, it is more erratic so that in angiosperms, about 20% may show biparental inheritance of plastids.

Plastid genes encode rRNAs and tRNAs and proteins involved in photosynthesis and plastid genes transcription and translation.However, nuclear genes encode the vast majority of plastid proteins and the expression of plastid genes and nuclear genes is highly co-regulated to coordinate proper development of plastids in relation to cell differentiation.

The plastid DNA exists as large protein-DNA complexes called plastid nucleoids. Each nucleoid particle has more than ten copies of the plastid DNA. Each copy called a plastome is a circular 75-250kb molecule, mostly 120–217 kb, with a unique exception of green alga *Floydiella terrestris* with huge cpDNA of 521.168 kb (Gyulai *et al*., 2012).A plastome has about 100 genes. A proplastid has a single nucleoid located in the centre of the plastid but developing plastid has many nucleoids localised at the periphery of the plastid, bound to the inner envelope membrane. Nucleoids differ in morphology, size and location within the plastid depending on the type of plastid. Several plastids may be interconnected by stromules within which proteins and presumably small molecules move.

2.5 To Distinguish ptDNA from nuDNA

Plastid DNA may be present in the nuclear genome. However, it is instable in nuclear genome probably due to change in the neo gene (deletion or sequence decay) or by silencing of the gene through epigenetic mechanisms as commonly occurs in plant transgenes (Shepard & Tennis, 2009). The nuclear genomes of many higher plants are extensively methylated at cytosine residues (Doerfler *et al*. 1990; Jeddoloh & Richards 1996) while plastid DNA is generally unmethylated (Scott & Possingham 1980) and in rare cases, very few sites are (Ngernprasirtsiri *et al*. 1989). Hence ptDNA and nDNA can be differentiated by use of differential digestion with methylation-sensitive restriction enzymes e.g. Hpa II that recognises CCGG but does not cleave if the internal cytosine base is methylated. Combination of Hpa II and EcoRI restriction enzymes result in ptDNA forming small Hpa II fragments while nDNA forms large EcoR1 fragments. The fragments are separated by agarose gel electrophoresis.

2.6 The Use of cpDNA Barcodes

Chloroplast DNA (cpDNA) has been used very frequently in plant systematic and phylogenetic studies. The plastid genome has some advantages in relation to the nuclear genome for traceability purposes which include the DNA being uniparentally

inherited (maternally in angiosperms and paternally in gymnosperms, in general, with exceptions), which facilitates determination of the maternal parent in hybrids and allopolyploids (Ackerfield and Wen, 2003); relatively abundant (generally, about 50-100 chloroplasts per cell) compared to nuclear DNA (generally 2n) and being circular (not linear like nuclear DNA), and therefore resistant to exonucleases hence highly conserved, i.e., relatively free of large deletions, insertions, transpositions, inversions and SNPs (single nucleotide polymorphism), which make it advantageous for phylogenetic studies. Chloroplasts also have a double membrane that makes their DNA more resistance to degradation, (which is also a significant advantage for forensics); and a lower mutation rate than nuclear genomic sequences. The DNA stability may be an advantage for traceability analyses, despite of a low polymorphism level.

There are about 100 functional genes in the chloroplast genome. It contains, with few exceptions (IRL – IRless), two duplicate regions in reverse orientation, known as the inverted repeats (IR) of 10–76 kb, which divides the chloroplast genome into large (LSC) and small single-copy (SSC) regions. Some chloroplast regions like psbA-trnH intergenic spacer and rps16 intron gene evolve relatively rapidly. There are a number of non-coding cpDNA regions which are useful target of study such as the intergenic spacer of atpB-rbcL (Manen & Natali, 1995;Baker *et al*., 1999; Asmussen & Chase, 2001; Manen *et al*., 2002), the rps16 intron, matK, ndhF, ycf6-psbM, and psbM-trnD (Oxelman *et al*., 1997; Andersson & Rova, 1999; Downie & Katz-Downie, 1999; Wallander & Albert, 2000; S˘ torchova´ & Olson, 2007), rpL16 intron (Jordan *et al*., 1996; Baum *et al*., 1998), trnL-F (Wallander & Albert, 2000), and psbA-trnH spacer, trnH-psbA (Kress *et al*., 2005; Chase *et al*., 2005), by using universal primers. For phylogenetic investigations cpDNA has been more readily exploited than the nuclear genome for barcoding, similar to mitochondrial genomes of animals. Kress *et al*. (2005) compared plastid genomes of *Atropa* and *Nicotiana*, and recorded that nine intergenic spacers trnK-rps16, trnH-psbA, rp136-rps8, atpB-rbcL, ycf6-psbM, trnV-atpE, trnC-ycf6, psbM-trnD, and trnL-F fulfill the barcode criteria. For comparison, ITS had a much higher divergence value (13.6%) than any plastid regions, especially rbcL, which is far the lowest in divergence (0.83%).

Chloroplast DNA can be used in characterization of plastid haplotypes for thorough studies of population genetics, phylogenetics and taxonomic background of grass

species and also to investigate nucleo-cytoplasmic effects – plastid signals controlling nuclear gene expression can be positive or negative. Chloroplast DNA sequences can be obtained in two ways, either by sequencing chloroplast DNA clones found in genomic libraries as "contamination" or by sequencing high purity extracted chloroplast DNA cloned into sequencing vectors.

2.7 Phylogenetics

Phylogenetic inference is the best estimate of the evolutionary history of a group of organisms based on inheritance of ancestral characteristics (Wright J., 2013). From molecular phylogenetics, phylogenetic (evolutionary) trees are constructed, which show evolutionary relationships among groups of organisms and also used to understand biological processes e.g. functions of genes or proteins etc. Specific characters among related organisms are believed to have appeared by mutation (Wright J., 2013). There are two basic types of mutations in DNA (1) substitution of a nucleotide for another. This can either be (a) transitions i.e. a purine for a purine (A↔G) or a pyrimidine for a pyrimidine (C↔T) or (b) transversions i.e. a purine for a pyrimidine or vice vasa. Transversions are less frequent than transitions. Substitutions that result in synonymous codons are 'silent' i.e. no change in amino acid sequence due to degeneracy of the genetic code. These mainly occur at the third codon position. Substitutions that result in non-synonymous codons cause a change of the amino acid residue hence change in phenotype. Non-sense substitutions result in a stop codon. (2) Insertion or deletion of one or more nucleotides. These cause shifts in reading frames of protein-coding genes (frameshift mutations). Gaps in multiple-aligned sequences may be due to either insertion or deletion and are therefore grouped as indels. Indels are more frequent in non-coding regions of the genome (Wright J., 2013).

Dobzhansky (1973) stated that nothing in biology makes sense except in the light of evolution. Phylogeny is in the midst of a renaissance, heralded by the widespread application of new analytical approaches and molecular techniques. Phylogenetic analyses provide insights into relationships at all levels of evolution. The phylogenetic trees now available at all levels of the taxonomic hierarchy for animals and plants play a pivotal role in comparative studies in diverse fields from ecology to molecular evolution and comparative genetics (Soltis & Soltis, 2000). Taxonomy is a synthetic science, drawing upon data from such diverse fields as morphology, anatomy, embryology, cytology, and chemistry. Traditional taxonomists use multiple

morphological traits to delineate species. Today, such traits are increasingly being supplemented with DNA based information. In recent years, development of techniques in molecular biology including those for molecular hybridization, cloning, restriction endonuclease digestions and nucleic acid sequencing have provided many new tools for the investigation of phylogenetic relationships. The reconstructions of angiosperm phylogeny have relied largely on plastid and mitochondrial genes (Chase *et al*., 1993; Nandi *et al*., 1998; Savolainen *et al*., 2000; Hilu *et al*., 2003; Zhu *et al*., 2007; Qiu *et al*., 2010) and sometimes entire plastid genomes (Jansen *et al*., 2007; Moore *et al*., 2007, 2010) and nuclear genes (Doyle *et al*., 1994; Soltis *et al*., 1997; Mathews & Donoghue, 1999; Finet *et al*., 2010; Lee *et al*., 2011).

A diverse array of molecular techniques are available for studying genetic variability, including restriction site analysis, analysis of DNA rearrangements, gene and intron loss, and the dominantly used PCR based techniques followed by DNA sequencing and cladistics analyses of the nuclear genome (nuDNA) and both organelle genomes of mitochondria (mtDNA) and chloroplast (cpDNA) (Martins & Hellwig, 2005; Mitchell & Wen, 2005). Multiple sequence alignments software programs of BioEdit Sequence Alignment Editor (North Carolina State University, USA) (Hall, 1999), CLUSTAL W (Thompson *et al*., 1997), MAFFT, BLAST analysis of the NCBI databases (Altschul *et al*., 1997) and MEGA6 (Tamura *et al*., 2011) are available for inferring Phylogeny.

The three commonly used methods for phylogenetic analysis are MP (maximum parsimony), ML (maximum likelihood), and (BI) Bayesian inference. Of them ML (maximum likelihood) was found to be the most discriminative (Hillis *et al*., 1994). The maximum parsimony algorithm (Farris, 1970; Swofford *et al*., 1996) searches for the minimum number of genetic events (e.g. nucleotide substitutions) to infer the shortest possible tree (i.e., the maximally parsimonious tree). Often the analysis generates multiple equally most parsimonious trees.

When evolutionary rates are drastically different among the species analysed, results from parsimony analysis can be misleading e.g. due to long-branch attraction (Felsenstein, 1978). Parsimony analysis is most often performed with the computer program PAUP* 4.0 (Swofford, 2002), and MEGA (Tamura *et al*., 2007, 2011).

The maximum likelihood (ML) method (Felsenstein, 1985; Hillis *et al*., 1994) evaluates an evolutionary hypothesis in terms of the probability that the proposed model and the hypothesized history would give rise to the observed data set properly. The topology with the highest maximum probability or likelihood is then chosen. This method may have lower variance than other methods and is thus least affected by sampling error and differential rates of evolution. It can statistically evaluate different tree topologies and use all of the sequence information. The Bayesian phylogenetic inference is model-based method and was proposed as an alternative to maximum likelihood (Rannala & Yang, 1996).

2.8 The Genebank Sequence Database

The GeneBank sequence database is an open access, annotated collection of all publicly available nucleotide sequences and their protein translations. This database is produced at National Centre for Biotechnology Information (NCBI) as part of the International Nucleotide Sequence Database Collaboration. GeneBank and its collaborators receive sequences produced in laboratories throughout the world from distinct organisms. GeneBank is built by direct submissions from individual laboratories, as well as from bulk submissions from large-scale sequencing centres. The Entrez Nucleotide and BLAST (Basic Local Alignment Search Tool) are the two main ways to search and retrieve data from GenBank at http://www.ncbi.nlm.nih.gov/genbank/

CHAPTER THREE

MATERIALS AND METHODS

3.1 Experimental Materials

Genotypes included 97 wild accessions of *Eleusines* collected by ICRISAT/Bioinnovate from Ethiopia, Tanzania, Uganda and Kenya (eastern African region believed to be the primary centre of diversity) and 3 varieties of cultivated finger millet (presented in appendix I). Three primers (both forward and reverse) for the chloroplast DNA barcodes *rpl32-trnL* intergenic spacer, *ndhF* and *rps3* genes were used.

3.2 Methodology

3.2.1 Sowing of Accessions

Seeds of accessions were planted in well labelled trays in a glasshouse at ICRAF Research Nursery (World Agroforestry Centre, Nairobi). Ugandan, Kenyan and Tanzanian accessions were sown on 22/10/2013 in a 6×11-hole tray whose layout is shown in table 3.1. The accessions that failed to germinate or had poor germination were re-sown on 4/11/2013. The accessions and their layout in the tray are shown in table 3.2. Ethiopian accessions were sown on 15/11/2013 in three, 4×7-hole trays as shown in tables 3.3a - c.

Table 3.1 Layout for Ugandan, Kenya and Tanzanian accessions in a 6×11-hole tray

UG 1	UG 02	UG 02	UG 03	UG 8	UG 9
UG 9	UG 10	UG 11	UG 18	UG 18	UG 19
UG 19	UG 20	MS 1	MS 3	MS 4	MS 5
MS 5	MS 6	MS 7	MS 8	MS 9	MSN 10
MS 11	MS 12	MS 13	MS 9	MS 13	MS 15
MS 16	MS 17	MS 17	MS 18	MS 18	MS 19
MS 19	MS 21	MS 21	MD 48	EDL 9	EDL 10
EDL 15	EDL 16	EDL 19	EDL 22	EDL 24	EDL 25
EDL 26	EDL 27	EDL 30	EDL 31	LEN 7	LEN 21
LEN 24	LEN 27	LEN 37	LESK 10	LESK 18	EDL 34

Table 3.2 Layout for re-sown accessions that had failed to germinate

UG 03	UG 03	UG 8	UG 8	LESK 18	LESK 18
LEN 27	LEN 27	LEN 24	LEN 24	LEN 21	LEN 21
EDL 22	EDL 22	EDL 19	EDL 19	EDL 25	EDL 25
EDL 27	EDL 27	EDL 26	EDL 26	EDL 31	EDL 31

Table 3.3a Tray one layout for some Ethiopian accessions

AAU-ELU-01	AAU-ELU-02	AAU-ELU-03	AAU-ELU-04
AAU-ELU-05	AAU-ELU-06	AAU-ELU-07	AAU-ELU-08
AAU-ELU-09	AAU-ELU-10	AAU-ELU-11	AAU-ELU-12
AAU-ELU-13	AAU-ELU-14	AAU-ELU-15	AAU-ELU-16
AAU-ELU-17	AAU-ELU-18	AAU-ELU-19	AAU-ELU-20
AAU-ELU-21	AAU-ELU-22	AAU-ELU-23	AAU-ELU-24
AAU-ELU-25	AAU-ELU-26	AAU-ELU-27	AAU-ELU-28

Table 3.3b Tray two layout for some Ethiopian accessions

AAU-ELU-29	AAU-ELU-30	AAU-ELU-31	AAU-ELU-32
AAU-ELU-33	AAU-ELU-34	AAU-ELU-35	AAU-ELU-36
AAU-ELU-37	AAU-ELU-38	AAU-ELU-39	AAU-ELU-40
AAU-ELU-41	AAU-ELU-42	AAU-ELU-43	AAU-ELU-44
AAU-ELU-46	AAU-ELU-47	AAU-ELU-48	AAU-ELU-49
AAU-ELU-50	AAU-ELU-51	AAU-ELU-52	AAU-ELU-53
AAU-ELU-54	AAU-ELU-56	AAU-ELU-57	AAU-ELU-58

Table 3.3c Tray three layout for some Ethiopian accessions

AAU-ELU-59	AAU-ELU-60	AAU-ELU-61	AAU-ELU-62
AAU-ELU-63	AAU-ELU-64	AAU-ELU-65	AAU-ELU-66
AAU-ELU-67	AAU-ELU-68		

3.2.2 DNA Extraction

DNA was extracted from fresh leaves of two to three weeks old seedlings (plate 3.1) using Isolate II plant DNA kit by Bioline as per the manufacturer's protocol (appendix III). The accessions were assigned laboratory numbers in sampling order prior to DNA extraction. Three accessions UG 18, LEN 24 and EDL 25 were sampled twice, to act as accuracy check and therefore they had two laboratory numbers 14 & 21, 38 & 70 and 39 & 72 respectively. During sampling of tissues, necessary precautions were taken like sterilizing the apparatus and the surface using 70% ethanol and keeping the samples on ice. The tissues were macerated with a Geno/grinder TissueLyser.

Plate 3.1: Picture of seedlings of some accessions in a 6×11-hole tray. The layout and the labels helped identify the various accessions. (source: Author, 2013)

3.2.3 DNA integrity check

DNA integrity/quality check was done using 0.8% (w/v – 0.8g agarose in 100ml 1×TBE buffer) agarose gel electrophoresis stained with Gel Red® (Biotium Inc., USA), 5µL/100ml. The samples were run against λDNA of known concentration, in this case 50ng/µL. Degraded DNA, absence of DNA in the extract or too low DNA

concentrations necessitated re-extraction of those samples. A mixture of 2µL DNA and 1µL 3× loading buffer dye (25mg bromophenol blue (0.25%), 25mg xylene xyanol (0.25%) and 4g sucrose (40%)) were loaded into the wells and run for 45 minutes at 80 volts (Electrophoresis power supply – EPS 600, Pharmacia Biotech) in a 1×TBE buffer (0.1M Tris base, 0.1M boric acid and 0.02M EDTA; pH 8.0) (Santie M. D. *et al*., 2015). The gel was then visualized under UV light and photographed (plate 3.2).

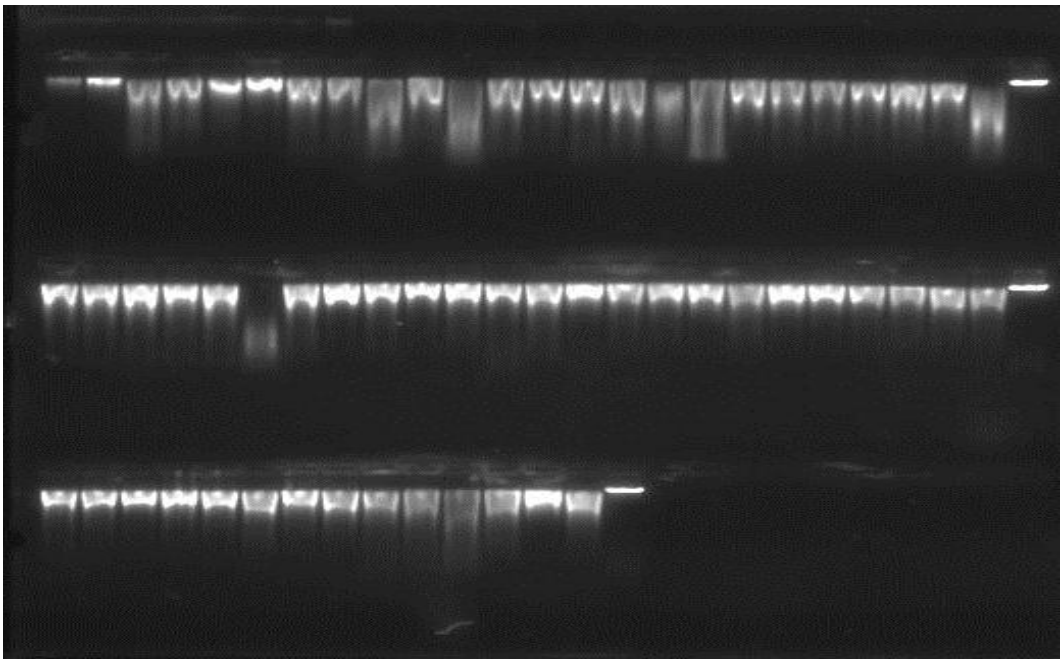

Plate 3.2: Gel picture of genomic DNA of some samples. At end of each row is 50ng/µL conc. λDNA (Source: Author, 2013).

3.2.4 DNA Quantification

Estimation of DNA concentration (DNA quantification) was done by Qubit Fluorometric Quantification as per the protocol (appendix IV), nanodrop spectrophotometry (plate 3.3) and gel electrophoresis against known concentration of Lambda DNA. Just a few samples were quantified using Qubit and their results were used to estimate the other samples on the gel. This was done to save on cost and time. Nanodrop (plate 3.3) was used to check the reasonableness of the Qubit. For any one sample, nanodrop would give a very high value compared to Qubit®. This may be attributed to the ability of Qubit® to discriminate between dsDNA and ssDNA, hence considered more accurate although this depends on accuracy when diluting the various reagents.

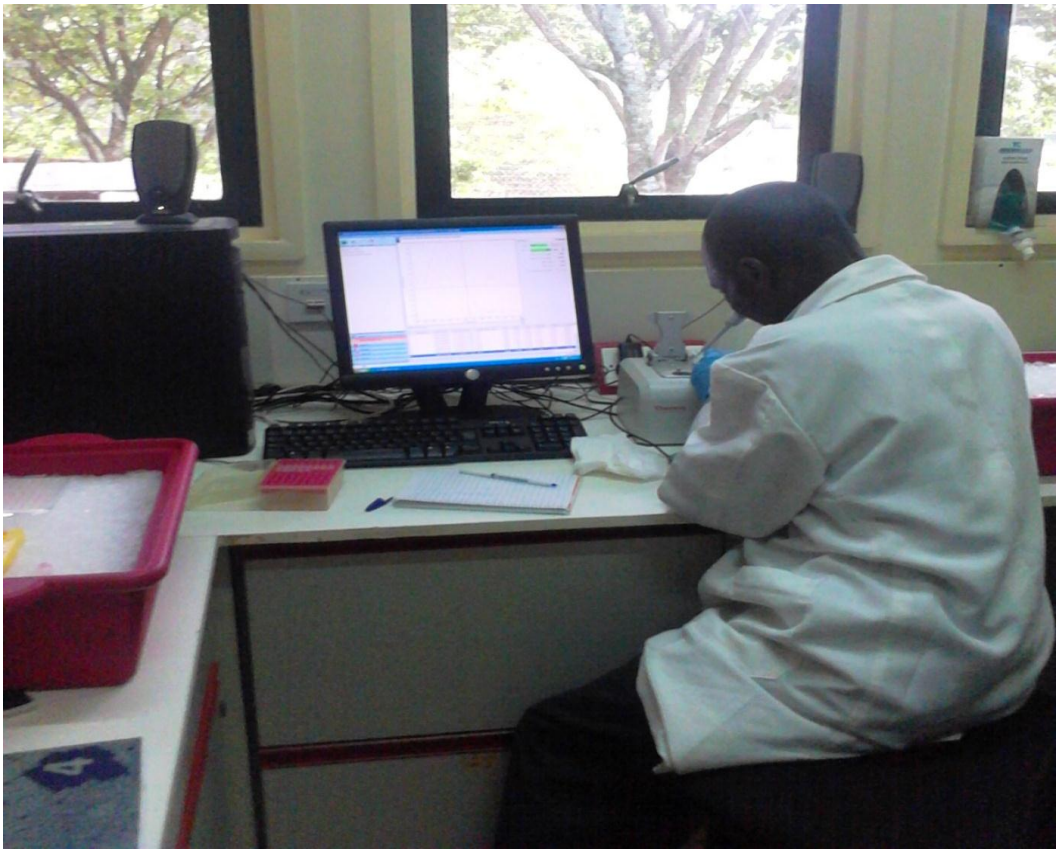

Plate 3.3: Nanodrop spectrophotometry. It was used to check whether the Qubit® readings were reliable (Source: Author).

3.2.5.DNA Dilution

Samples were then diluted to a 30ng/µL concentration using TE (1mM Tris; 0.01mM EDTA). Dilutions were based on the formula$C_1 \times V_1 = C_2 \times V_2$; Where C_1= estimated concentration of the DNA sample; V_1= volume of the DNA to pick; C_2= final concentration to constitute i.e. 30ng/µL and V_2= final volume to constitute i.e. 100 µLs. Plate 3.4 shows the diluted samples of DNA in 96-well plates on an ice bath.

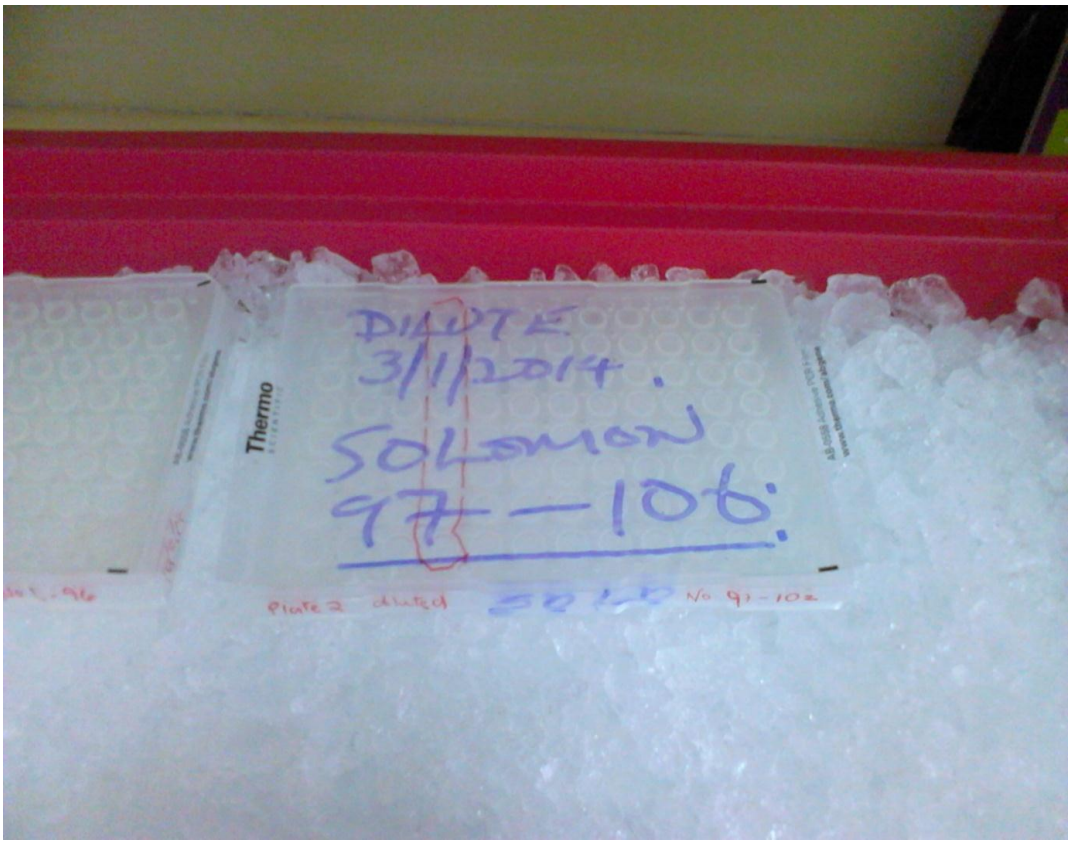

Plate 3.4: Dilute DNA in 96-well plates. Plate 1 for the first 96 samples and plate two for samples 97-106 (indicated in red). As seen, all through the process, DNA was on ice, whenever removed from the fridge at 4^{o}C. Stock DNA was kept at -20^{o}C to be used in future as and when need arises (Source: Author).

The concentrations of the diluted DNA were ascertained on 0.8% (w/v) agarose gel against 20ng/µL and 40ng/µL λ DNA. The results are shown in plate 3.5 below. Samples were loaded at 2µL DNA for 1µL 3× loading dye and run on 0.8% agarose gel at 80V for 1 hour. The samples showing absence or little DNA were noted. They were samples: 11, 17, 25, 33, 36, 49, 62, 63, 66, 70, 78, 83, 86, 89, 90, 97 & 105. Plate 3.6 is a gel picture of their new dilutions from stock. Samples 17 and 41 had degraded DNA while 86 and 89 had no DNA, and they were therefore left out.

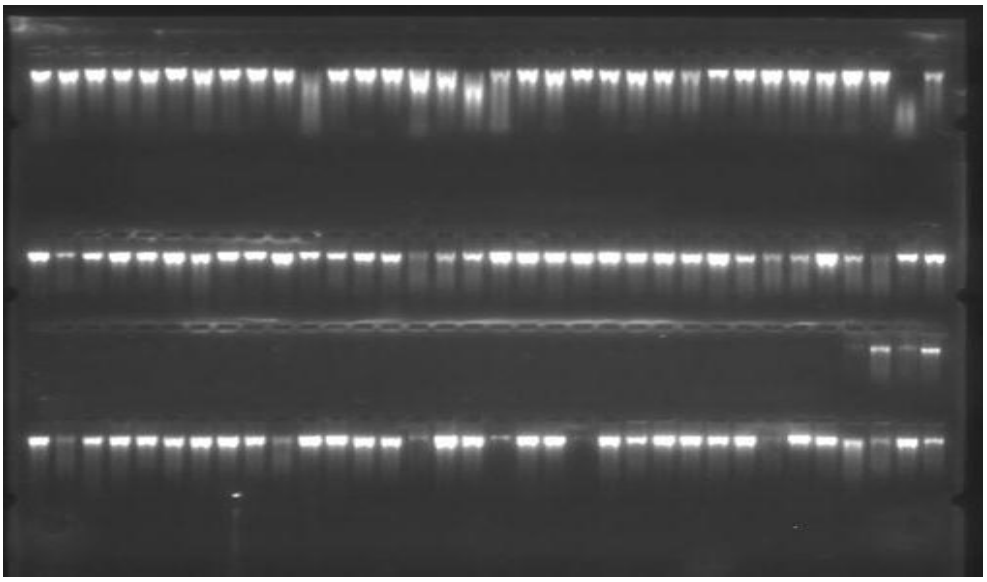

Plate 3.5: Gel picture of diluted genomic DNA for all samples. The samples are in order of the Lab. Numbers (appendix I), beginning from the first row and ending with the third row.

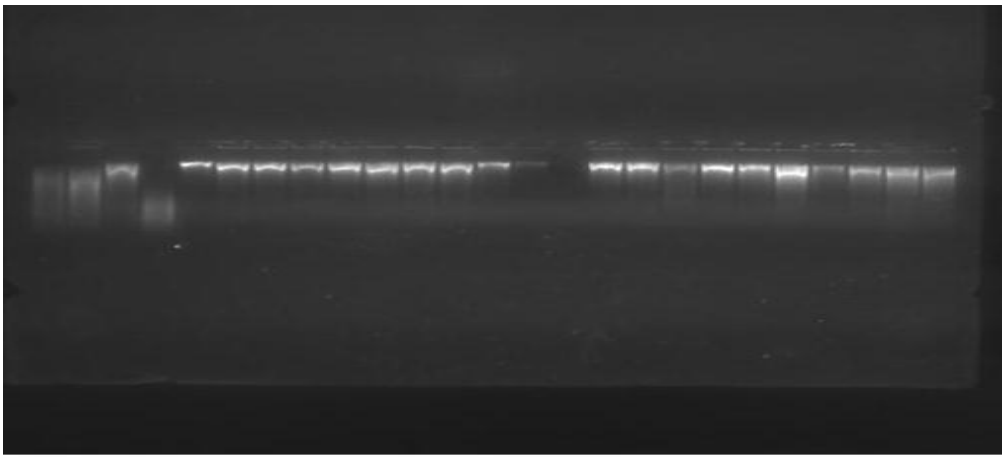

Plate 3.6: Gel picture of new dilutions for problematic samples. Dilutions were made from stock that had been kept at -20^{o}C, for the problematic samples noted in plate 6 above.

3.2.6 PCR Master Mix

Three PCRs were performed using the three primers *rpl32-trnL*(For – 5` - CAGTTCCAAAAAAACGTACTTC – 3`; Tm 50.1^{o}C; Rev – 5` - CTGCTTCCTAAGAGCAGCGT – 3`; Tm 53.8^{o}C), *ndhF* (For – 5` - ACTGCAGGATTAACTGCGTT – 3`; Tm 51.3^{o}C; Rev – 5` - GACCCACTCCATTGGTAATTC – 3`; Tm 52.1^{o}C) and *rps3*(For – 5` - TCAGACTTGGTACAACCCAA – 3`; Tm 49.9^{o}C; Rev – 5` - TCTTCGTCTACGAATATCCA – 3`; Tm 47^{o}C;) by BioNeer®. *Rpl32-trnL and ndhF* were performed on GeneAmp 96 well PCR system 9700®Applied Biosystems, USA thermocycler while *rps3* was performed on Veriti 96 well Thermal cycler, Applied Biosystems. For samples 97-106, PCR was performed in tubes. PCR master mix for each primer comprised of 2µL of 10× PCR buffer, 0.8µL of 50mM $MgCl_2$, 1.6µL of dNTP mix (2mM concentration each), 2µL each of 2pM forward and reverse primers, 0.08µL of Taq DNA polymerase (Sibenzyme Ltd, Russia), 60ng template DNA and DPC treated water to top up to a reaction volume of 20µLs (Santie M. D. *et al*., 2015). For the all 106 samples the master mixes were prepared as shown in table 3.4 below. For each sample, 18µL of the master mix was dispensed in the respective tube and 2µL of the DNA added.

Table 3.4: Master mix volume preparations for the three primers

Component	**Stock Conc.**	**React. Conc.**	**1 React. Vol.**	**110 Reactions**
PCR Buffer	10×	1×	2	220
MgCl2	50mM	2mM	0.8	88
dNTPs	2mM	0.16mM	1.6	176
F –primer	2pMoles	0.2pMoles	2	220
R – primer	2pMoles	0.2pMoles	2	220
Taq Polymerase	5U/µL	0.2U/µL	0.08	8.8
Water			9.52	1047.2
Total Volume (µL)			18	1980

3.2.7 PCR Profiles

PCR profiles were performed according to New England BioLabs inc. protocol with some modifications due to differences in melting temperatures of the primers and expected length of amplicons. The profiles were optimised to: *rpl32-trnL* – 1st denaturation at 94°C for 5min; 35 cycles of denaturation at 94°C for 2min, annealing at 60°C for 1min and elongation at 72°C for 2min and final elongation at 72°C for 10min; *ndhF* - 1st denaturation at 94°C for 5min; 35 cycles of denaturation at 94°C for 1min, annealing at 58°C for 1min and elongation at 72°C for 1min and final elongation at 72°C for 7min and *rps3* - 1st denaturation at 94°C for 5min; 35 cycles of denaturation at 94°C for 1min, annealing at 52°C for 1min and elongation at 72°C for 1min and final elongation at 72°C for 7min.

3.2.8 PCR Products Resolution

This step is necessary to ascertain whether amplification took place or the samples that were not amplified. PCR products were resolved on 1% agarose (w/v) run at 80V for 45min (Santie M. D. *et al*., 2015). Plate 3.7 is a gel picture of rps3 amplicons for the first 96 samples in the order of their laboratory numbers (appendix I). ndhF and rpl32-trnL primers also gave similar good results**.**

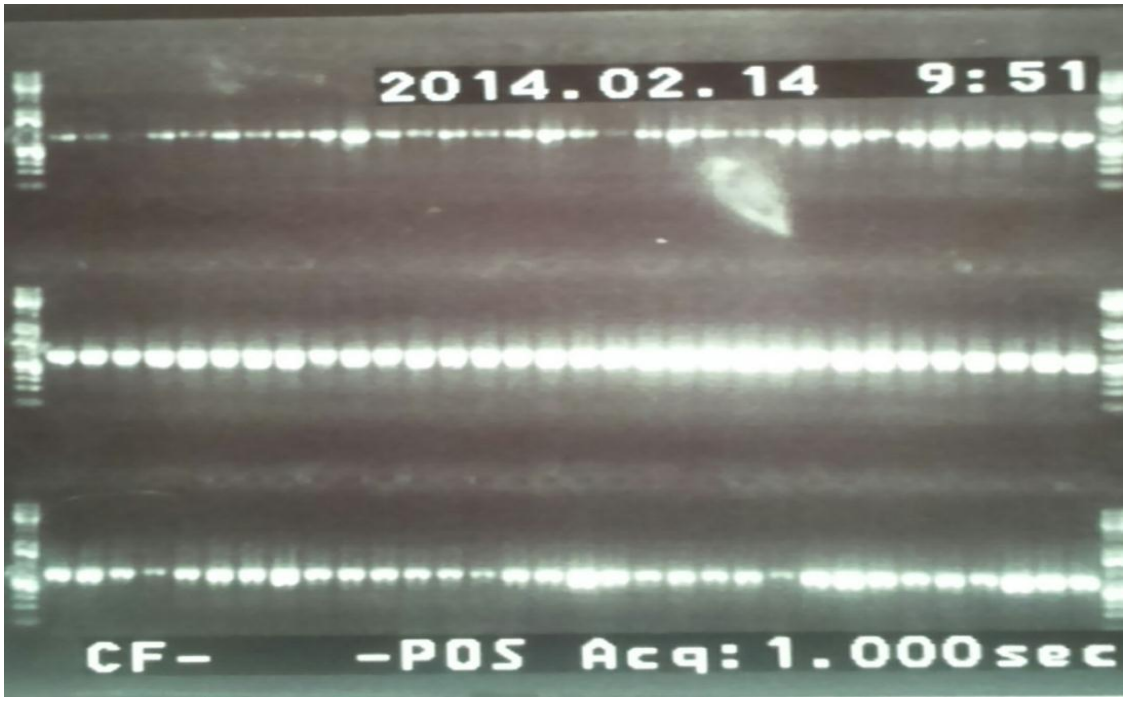

Plate 3.7: PCR results for rps3 primer. The gel picture shows the first 96 samples in order of their laboratory numbers (appendix I). At beginning and ending of each row is a ladder (O'GeneRuler 100bp Plus DNA Ladder, ready-to-use, by ThermoScientific)(Source: Author).

3.2.9 PCR Products Cleaning

This procedure is necessary in order to remove unused dNTPs and primers prior to sequencing. The products were cleaned using ExoSAP-IT® as per the manufacturer's protocol (appendix V) and then send for sequencing. Sequencing was done at Inqaba Biotechnical Industries (Pty) Ltd, Pretoria, South Africa.

3.3 Data Analysis

The AB1 chromatograms were changed to fasta format in MEGA6. The reverse sequences were reversed and complimented to the forward. For each primer, two folders were created, one for forward and the other for reverse compliment fasta files.

Base calling and sequence editing were done simultaneously using ChromasPro and BioEdit version 7.2.5 (12/11/2013) software. The forward (F) and Reverse Compliment (RC) sequence pairs for each sample were opened in BioEdit and then pairwise alignment done allowing the ends to slide. The sequences were edited in BioEdit by comparing the pairwise-aligned F and RC sequences Vis a Vis the AB1

Chromatograms opened in ChromasPro software. The unresolved ends (beginning and ending of sequences) were trimmed off. A consensus sequence was then created for each sequence pair in BioEdit and the sequences exported into a folder by splitting into individual fasta files i.e. for each sample there were forward, reverse compliment and consensus sequences. Each set was put into a folder.

For each primer, one consensus sequence was used to do blast into the NCBI and Phytozome databases in order to confirm that the sequences were the correct ones and also to retrieve similar sequences already in the databases. Sequences in the databases that scored over 99% hits were retrieved. The consensus sequences and retrieved sequences in fasta format were then combined into a single file for each primer and then multiple-aligned using online MAFFT version 7. The alignment was then copied in fasta format and opened in BioEdit. Plate 3.8 shows part of alignment of sequences for the *ndhF* gene.

The alignments were trimmed to begin and end at same point in BioEdit and then saved as a single fasta file that could be opened in Notepad. The three files of aligned, trimmed combined consensus and retrieved sequences (a file set for each primer) were then opened in MEGA6 and then converted into Mega format.

Phylogenetic reconstruction analysis was done using MEGA6. Statistical method used was neighbour Joining (NJ) (Saitou and Nei., 1987). This is a distance matrix method that produces unrooted tree that shows relationship but does not assume ancestry. It uses star decomposition method by first producing a star tree and then pulling out neighbouring pairs at a time. Bootstrap method (Efron B., 1979) was used as a test of phylogeny with 1000 Bootstrap replications. It was used to assess the reliability that the clustering of the accession is as shown in the trees. The percentage values indicate the probability of the tree coming out the same way in the various replications. A good reliability is 70% and above while anything below 50% generally is unreliable. Jukes-Cantor model of DNA evolution was used. This model assumes uniform rate of substitution of nucleotides at all sites. This is a 1-parameter model most preferred for closely related sequences.

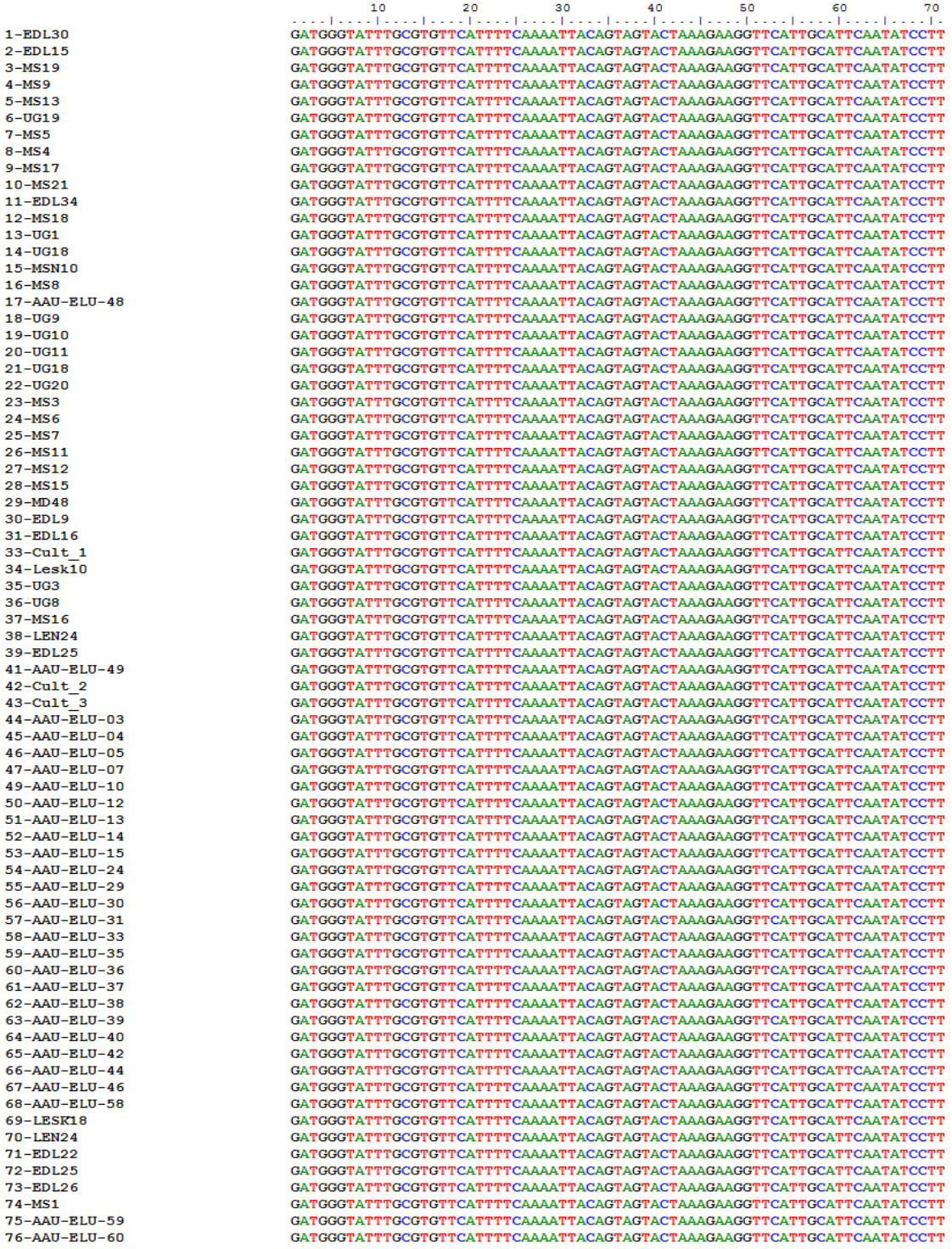

Plate 3.8: A page from BioEdit showing part of ndhF alignment. Shown above, all the consensus sequences began at the same point and (not shown) ended at the same point. There were gaps (not seen above) at some points, presumably due to indels (Insertions and Deletion mutations) and also nucleotide differences (SNPs)(source: Aouthor).

CHAPTER FOUR

RESULTS

4.1 *rpl32-trnL* primer

This primer locus was quite informative. After editing, it remained with 651 bases which were used in the construction of the tree shown in figure 4.1. All Ethiopian accessions previously grouped as *E. floccifolia* formed a distinct group apart from the *E. coracana – E.indica* and *E. africana – E. kigeziensis* groups.

MD48 was classified as it were previously, as *E. kigeziensis* with 95% bootstrap support.

Genotypes UG2, UG3, UG9, UG11, UG20, MS3, MS7 and MS15 were clustered together with retrieved *E. africana* sequence with 98% bootstrap support. These are Ugandan and Tanzanian accessions. AAU-ELU-65 of Ethiopian origin had only 10% divergence from them and was previously classified as *E. africana.*

AAU-ELU-48, AAU-ELU-02 and AAU-ELU-47 had 95% Bootstrap support that they will always cluster together. All are previously classified as *E. floccifolia.* This group was close to MS 16 and a retrieved *Eleusine multiflora* sequence. MS 16 was previously classified as *E. jaegeri* hence its close relationship with *E. floccifolia* and *E. multiflora* (see *ndhF*).

AAU-ELU-01, AAU-ELU-57 and AAU-ELU-68, all of which are classified as *E. floccifolia* had 86% supports for relatedness with AAU-ELU-64 which is classified as *E. multiflora.* Surprisingly, there were two groups of *E. floccifolia* each in close association with *E. multiflora* (Bootstrap consensus tree puts the two groups together). Both are diploids, but different in growth habit and number of chromosomes – *E. floccifolia* is perennial and x=9 whereas and *E. multiflora* is annual and x=8.

Most genotypes grouped with *E. coracana* (including the three cultivated genotypes as expected) and this group had 72% supports for relatedness with *E. indica.*

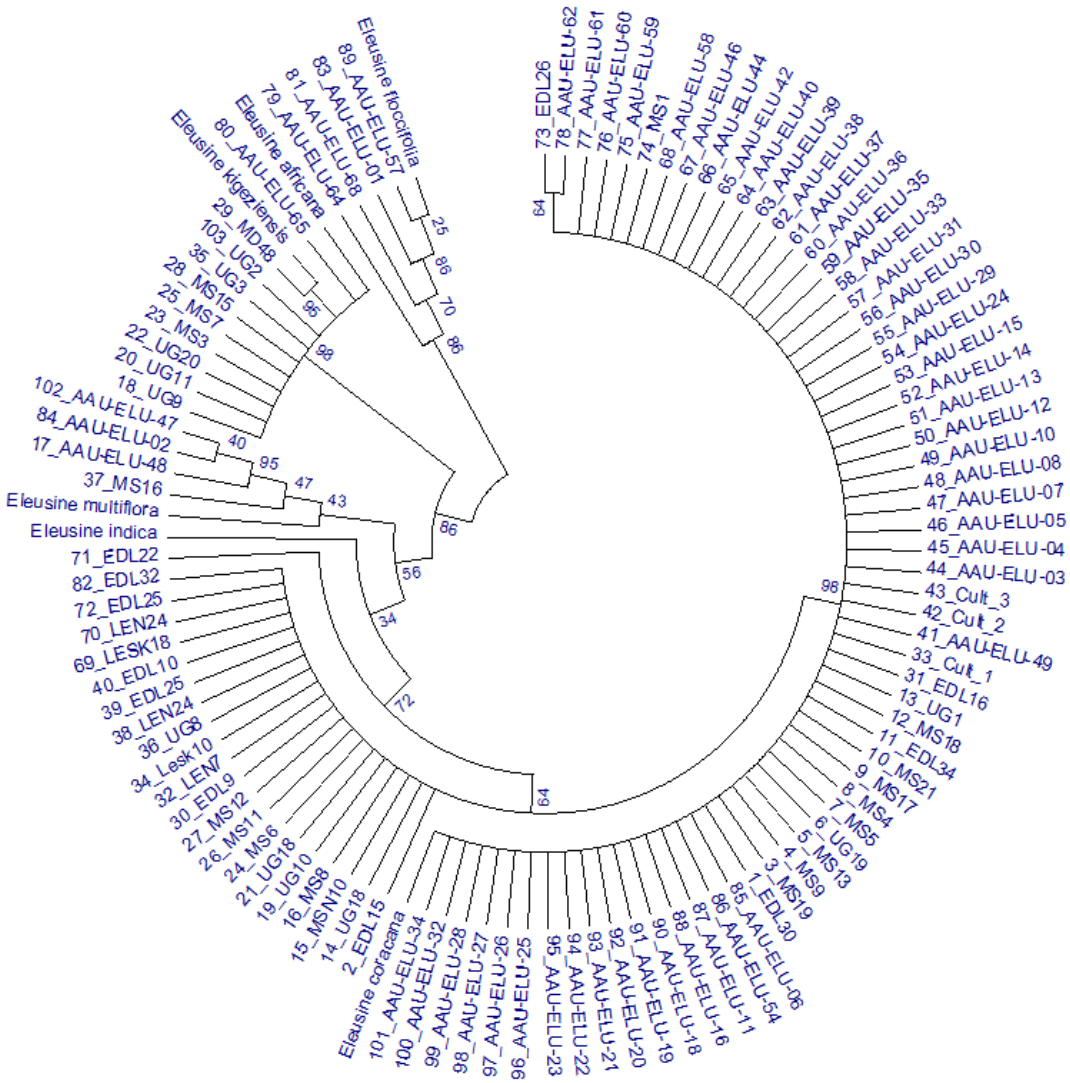

Figure 4.1: Phylogram of *rpl32-trnL* barcode.

4.2 *ndhF* primer

Just like *rpl32-trnL*, this primer locus was also quite informative (figure 4.2). After editing, it remained with 719 bases (being the longest of the three) that were used to construct the tree. The Ethiopian wild accessions AAU-ELU-48, AAU-ELU-49, AAU-ELU-68, AAU-ELU-01, AAU-ELU-02 and AAU-ELU-47 were classified as *E. floccifolia* as it previously was. However, AAU-ELU-16 previously classified as *E. coracana* was also included in this *E. floccifolia* group. They have 94% bootstrap support for relatedness to retrieved *E. jaegeri* sequence. Like *E. floccifolia, E. jaegeri* is also a diploid and perennial but with different chromosome number, x=10. As with *rpl32-trnL* locus, *E. multiflora* was close to *E. floccifolia.*

Ethiopian accession AAU-ELU-22 clustered with *E. multiflora* with 97% support yet previously it had been grouped as *E. intermedia.* Both species are diploids but while *E. multiflora* is annual and x=8, *E. intermedia* is perennial (and with similar seed characters with *E. floccifolia)* and x=9.

MS 16, a Kenyan accession previously classified as *E. jaegeri*, stood independent – as it were with *rpl32-trnL* primer – with 73% support for relatedness to *E. multiflora* – *E. floccifolia* group.

All other genotypes except LESK 10 (61% divergence from the group) grouped together with *E. coracana, E. kigeziensis* and *E. africana* (all tetraploids) with 40% bootstrap support for their relatedness with *E. indica* (a diploid). Within the group, MS11, MS12, MS6, UG18, UG10, MS8 and MSN10 formed a subgroup with 65% bootstrap support. These are Ugandan and Kenyan accessions. It is worth noting that UG18 had two samples laboratory numbers 14 and 21 (separate from point of sampling and DNA extraction) as a control, and they both appear in this subgroup ascertaining some degree of accuracy of the results.

LESK 10 stood independent with 64% support for relatedness to the *E. coracana* – *E. indica* group.

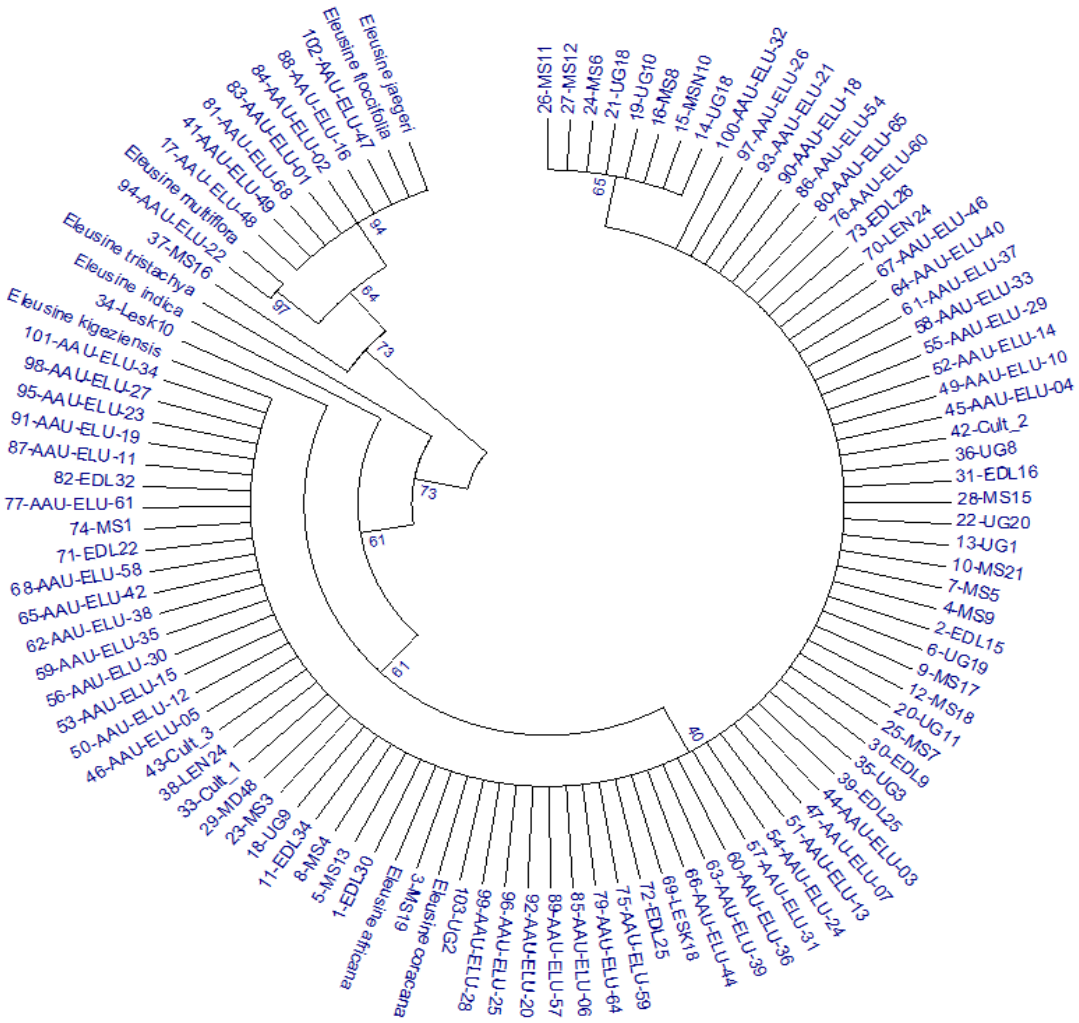

Figure 4.2: Phylogram of *ndhF* barcode.

4.3 *rps3* primer

This locus was the shortest of the three barcodes remaining with only 579 bases after trimming the sequences to make them all equal. Its phylogram is shown in figure 4.3 below. Though it was less informative, but just like the other two, it grouped accessions classified as *E. floccifolia* distinct from the other genotypes although BLAST did not retrieve any *floccifolia* sequence. However, this group clustered with BLAST-retrieved *E. jaegeri* sequence. They include AAU-ELU-48, AAU-ELU-49, AAU-ELU-68, AAU-ELU-01, AAU-ELU-02, AAU-ELU-57 and AAU-ELU-47; all Ethiopian accessions which were previously classified as *E. floccifolia.*

All the other genotypes clustered together with *E. coracana*. This group had 40% supports for relatedness with *E. indica*, a diploid wild relative considered the maternal genome donor of *E. coracana*.

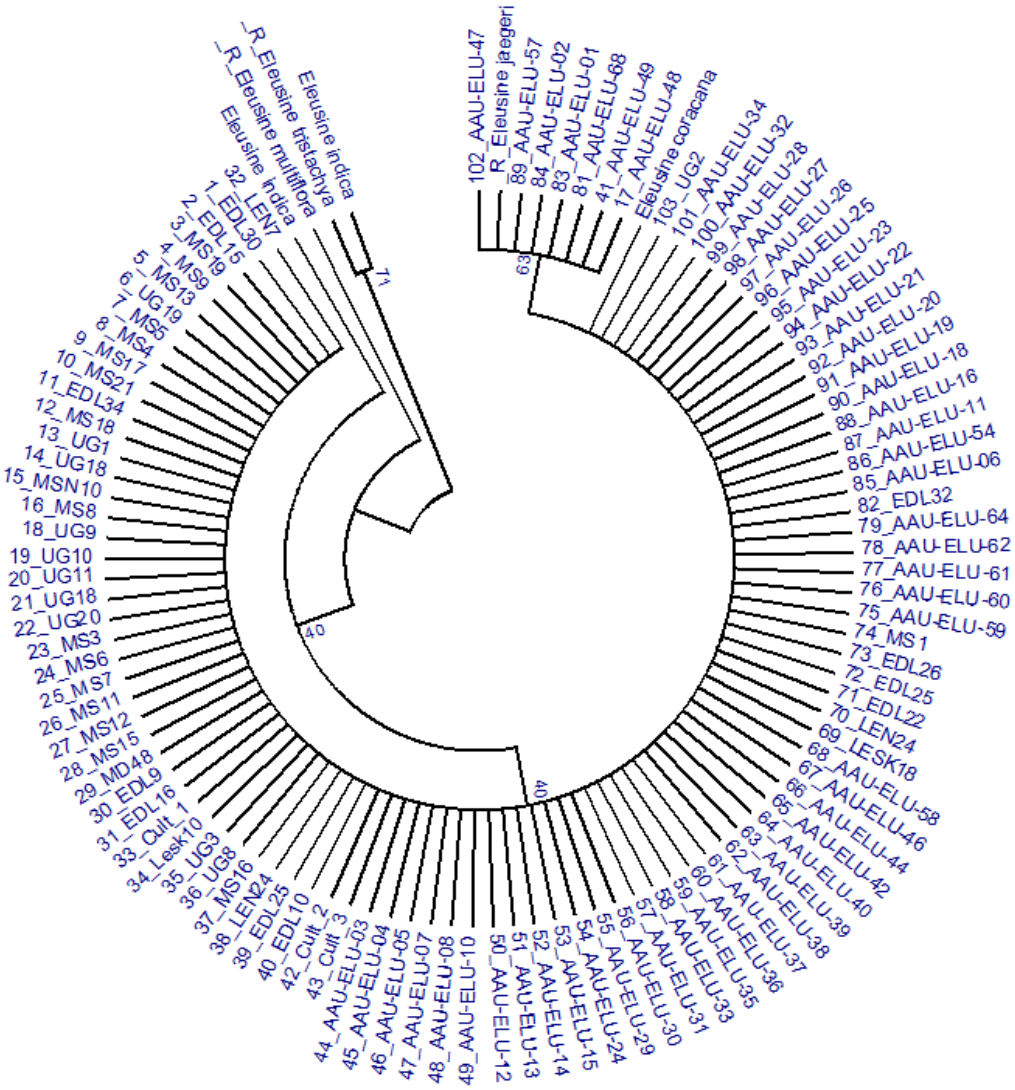

Figure 4.3: Phylogram of *rps3* barcode.

CHAPTER FIVE

DISCUSSION OF RESULTS

Generally, all the three ctDNA loci showed *E. floccifolia* as being distinct from *E. indica* and *E. coracana* but closely associated with *E. jaegeri*. *E. floccifolia* and *E. jaegeri* are both perennials but with different chromosome numbers, x=9 and x=10 respectively. *E. multiflora* with x=8 is affiliated to this group. On the other hand *E. indica* and *E. tristachya* associated with the three tetraploids *E. coracana, E. africana* and *E. kigeziensis* which is consistent with what others have found especially Hilu K. W. (1987) using restriction fragment homology of ctDNA, Bisht and Mukai (2001) using Genomic *in situ* Hybridization (GISH), Neves *et al*., (2005) using nuclear ribosomal DNA (nrDNA) internal transcribed spacer (ITS) and chloroplast *trnT-trnF* and Liu *et al*., (2011) using two low-copy nuclear markers *pepc4* and *EF-1α.* This implies that this group shares the same chloroplast genome distinct from the former *E. floccifolia* group. However, this does not refute the possibility of *E. floccifolia* being the paternal genome donor since chloroplasts are generally maternally inherited in angiosperms (**Kirk a**nd **Tilney-Bassett** 1978; **Sears** 1980, 1983; **Whatley** 1982 quoted by Hilu KW 1987).

While most of the accessions clustered as per the above clades following morphological characterization, some showed contrary results with the barcodes for instance AAU-ELU-22 previously classified as *intermedia* had 97% support for closeness with retrieved *multiflora* sequence suggesting that the accession is *multiflora* and not *intermedia*. This shows that morphological characterization is less reliable being subjective and environmentally influenced and therefore some of the accessions had been wrongly identified taxonomically basing on morphological characterization.

Bisht and Mukai (2001) had shown using genomic *in situ* hybridization that while *E. indica* and *E. tristachya* chromosomes hybridized the same set of *E. coracana* chromosomes (15-18 chromosomes – thus confirming *E. indica* as one of the genome donors and *E. tristachya* as sister); the chromosomes of *E. floccifolia* hybridized the other set of the *E. coracana* chromosomes, which led them conclude that it is the paternal genome donor. From the accessions collected, only some Ethiopian

accessions clustered with BLAST-retrieved *floccifolia* sequence and none from accessions from the other eastern African countries. This then suggests origin of finger millet being narrowed specifically to Ethiopia.

Mehra (1962, 63), building on Kennedy O'Byrne's (1957) work, did metroglyph analyses devised by Anderson on the broad range of finger millet diversity in east Africa and from the results of the cytotaxonomic and morphological data, Mehra did not only suggest the origin of finger millet in Ethiopia but also hypothesised its evolution thus: *E. indica* (2n=18) crossed with an **unknown species** (2n=18) and produced a wild hybrid (2n=18). The wild hybrid doubled its chromosomes and produced a wild tetraploid 2n=36 from which *E. coracana* was selected by man (basically because of agronomic traits). If Mehra's hypothesis holds, then, since the wild hybrid arose as a result of normal hybridization, only the nuclear genome was modified but chloroplast genome was intact, similar to the mother (*E. indica*) since chloroplasts are maternally inherited. This explains the close relationship between *E. indica* and *E. coracana* (confirming *E. indica* as the mother) when comparing the chloroplast genome but shows *E. floccifolia* as being distant thus neither proofing nor disproving its paternity. Thus a combination of GISH, SSR markers and GBS can successfully elucidate a hybridization process while Plastid DNA identifies the maternal genome donor (generally in angiosperms) or paternal genome donor (generally in gymnosperms).

CHAPTER SIX

CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

1. The study showed wild *coracana* accessions and the species *africana, indica, tristachya, and kigeziensis* to be more related to the cultivated accessions of finger millet than *floccifolia, jaegeri* and *multiflora.*

2. The primers ndhF and rpl32-trnL were more informative than rps3. The two primers had relatively more bases (719 and 651 respectively) while rps3 had only 579 bases. Therefore, longer barcodes are more reliable than shorter ones probably due to more variable base sites (SNPs).

3. The results have also shown that morphological characterisation is less reliable in identification of accessions into the various taxonomic groups and therefore molecular techniques such as DNA barcoding can be used for more reliability during breeding programmes.

6.2 Recommendations

1. The study recommends the use of molecular techniques to identify the accessions into the various taxonomic groups prior to use in breeding programmes.

2. When barcoding is the molecular technique of choice in taxonomic identification, longer barcodes are recommended for more reliable results.

3. The study recommends use of *E. kigeziensis* in breeding programmes to improve the cultivars for any noted trait of economic importance e.g. blast disease or *striga* weed resistance.

REFERENCES

Agrawal R, Agrawal N, Tandon R, Raina S. N., 2014. Chloroplast genes as genetic markers for inferring patterns of change, maternal ancestry and phylogenetic relationships among Eleusine species. *AoB PLANTS 6: plt056*; doi:10.1093/aobpla/plt056

Ansevin A. T., Vizard D. L., Brown B. W., McConathy J., 1976. "High-resolution thermal denaturation of DNA. I. Theoretical and practical considerations for the resolution of thermal subtransitions", *Biopolymers**15** (1):* 153–74, doi:10.1002/bip.1976.360150111, PMID 1244898

Ayyangar G. N. R. and Rao P. K., 1931. Inheritance of characters in ragi, Eleusine Coracana Gaertn., Part VI. Earhead shapes. *Indian L. Agr. Sci. 1*:434 – 444

Ayyangar G. N. R., Rao P. K. and Wariar U. A., 1933. Inheritance of characters in ragi, Eleusine corarana Gaertn., the finger millet; Part VII. Fist-like earheads. *Indian J, Agr. Sri. 3*: 1072 - 1079.

Ayyangar G. N. R., Wariar U. A. and Rao P. K., 1932. Inheritance of characters in ragi, Eleusine coracana Gaertn., Part VI. Earhead shapes. Indian J. Agr. Sri. **2: 254- 265.**

Barbeau W. E., Hilu K. W., 1993. Protein, calcium, iron and amino acid content of selected wild and domesticated cultivars of finger millet. Plant Foods for Human Nutrition 43:97–104.

Birky C. W., 2001. "The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models". *Annu. Rev. Genet.* ***35***: 125–48. doi:10.1146/annurev.genet.35.102401.090231. PMID 11700280.

Bisht M. S., Mukai Y., 2001a. Genomic in situ hybridization identifies genome donor of finger millet (Eleusine coracana). *Theoretical and Applied Genetics 102*:825–832.

Bisht M. S., Mukai Y., 2001b. Identification of genome donors to the wild species of finger millet, Eleusine africana by genomic in situ hybridization. *Breeding Science 51:*263–269.

Bisht M. S., Mukai Y., 2002. Genome organization and polyploid evolution in the genus Eleusine (Poaceae). *Plant Systematics and Evolution 233*:243–258.

Blaxter M., 2004. The promise of a DNA taxonomy. *Philos. Trans. R. Society London. B. Biol. Sci. 359*: 669-679. PMID: 15253352.

Chase M. W., Salamin N., Wilkinson M., Dunwell J.M., Kesanakurthi R. P., Haider N. and Salolainen V., 2005. Land plants and DNA barcodes: short-term and long-term goals. *Philos. Trans. R. Society London. B. Biol. Sci. 360:* pp. 1889 – 1895. PMID: 16214746.

Dobzhansky T., 1973. Nothing makes sense except in the light of Evolution. *Am. Biol. Teach. 35*

Efron B., 1979. Bootstrap methods: another look at the Jackknife. *The annals of statistics 7(1):* 1-26. Doi:10,1214/aos/1176344552

Fazekas A.J., Burgess K.S., Kesanakurti P.R., Graham S.W., Newmaster S.G., Husband B.C., Percy D.M., Hajibabaei M., Barrett S.C., 2008. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS ONE 3 (7), e2802*.

Fazekas A.J., Kesanakurti P.R., Burgess K.S., Percy D.M., Graham S.W., Barrett S.C., Newmaster S.G., Hajibabaei M., Husband B.C., 2009. Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? *Mol. Ecol. Resour. s1*, 130–139.

Ganopoulos I., Aravanopoulos F., Madesis P., Pasentsis K., Bosmali I., et al,. 2013. Taxonomic Identification of Mediterranean Pines and Their Hybrids Based on the High Resolution Melting (HRM) and trnL Approaches: From Cytoplasmic Inheritance to Timber Tracing. *PLoS ONE 8(4): e60945*. doi:10.1371/journal.pone.0060945

Gopal Reddy V., Upadhyaya H. D., Gowda C. L. L. and Singh S., 2009. Characterisation of eastern African finger millet germplasm for qualitative and quantitative characters at ICRISAT. *Journal of SAT Agricultural Research 7*.

Gyulai G., Szabo Z., Wichmann B., Bittsanszky A., Waters Jr. L., Toth Z., Dane F., 2012. Conservation genetics – Heat Map analysis of nuSSRs of a DNA of archaeological watermelons (Cucurbitaceae, Citrullus l. lanatus) compared to current varieties.

Hall T. A., 1999. BioEdit: a user-friendly biological Sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids. Ser.41*:95-98.

Hilu K.W., Johnson J. L., 1992. Ribosomal DNA variation in finger millet and wild species of Eleusine (Poaceae). *Theoretical and Applied Genetics 83*: 895–902.

Hilu K. W., Johnson JL. 1997. Systematics of Eleusine Gaertn. (Poaceae, Chloridoideae): chloroplast DNA and total evidence. *Annals of the Missouri Botanical Garden 84*: 841–847.

Hilu K. W., 1988. Identification of the 'A' genome of finger millet using chloroplast DNA. *Genetics 118*:163–167.

Hilu K. W., 1995. Evolution of finger millet: evidence from random amplified polymorphic DNA. *Genome 38*:232–238.

Hiremath S. C., Chennaveeraiah M. S., 1982. Cytogenetical studies in wild and cultivated species of Eleusine (Gramineae). *Caryologia 35*:57–69.

Hiremath S. C., Salimath S. S., 1991. The quantitative nuclear DNA changes in Eleusine (Gramineae). *Plant Systematics and Evolution 178*:225–233.

Hiremath S.C., Salimath S. S., 1992. The 'A' genome donor of Eleusine coracana (L.)Gaertn.(Gramineae). *Theoretical and Applied Genetics 84*:747–754.

Jansen R.K., Cai Z., Raubeson L.A., Daniel, H., Depamphilis, C.W., Leebens-Mack, J., Muller, K.F., Guisinger-Bellian, M., Haberle, R.C., Hansen, A.K., 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc. Natl. Acad. Sci. USA 104*, 19369–19374.

Kennedy-O'Byrne J., 1957. Notes on African grasses: XXIX. A new species of Eleusine from tropical and South Africa. *Kew. Bull. 1*: 65-74.

Knobloch I.W., 1972. Intergeneric hybridization in flowering plants. *Taxon 21*: 97–103.

Kress W.J., Wurdack, K.J., Zimmer E. A., Weigt L. A and Jansen, D. H., 2005. Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. USA, 102*: pp. 8369-8374

Liu Q., Peterson P.M., 2010. Advances in systematics of adaptively radiated Eleusine Gaertn. (Poaceae). *Journal of Tropical and Subtropical Botany 18:*

Liu Q, Triplett JK, Wen J, Peterson M. 2011. Allotetraploid origin and divergence in Eleusine (Chloridoideae, Poaceae): evidence from low-copy nuclear gene phylogenies and a plastid gene chronogram. *Annals of Botany 108*:1287–1298.

National Research Council, 1996. Lost crops of Africa, vol. 1, Grains. *National Academy Press, Washington, DC.*

Neves, S. S., Swire-Clark, G., Hilu, K. W. and Baird, W. V. 2005. Phylogeny of Eleusine (Poaceae: Chloridoideae) based on nuclear ITS and plastid trnT-trnF sequences. *Molecular Phylogenetic Evolution 35*: 395 – 419.

Neves, S. S. 2011. Eleusine. In Kole, C. (Ed), Wild crop relatives: Genomic and breeding resources, millets and grasses. DOI 10.1007/978-3-642-14255-0_7, # *Springer-Verlag Berlin Heidelberg 2011*.

NewMaster S.G., Fazekas A.J. and Ragupathy S., 2006. DNA barcoding in land plants: evaluation of rbcL in a multigene tiered approach. *NRC Research press website at http://canjbot.nrc.ca*

Oduori C.O.A., P. Tongoona, J. Derera, O.M. Odongo. 2008. Finger millet hybridisation using ethrel chemical hybridising agent. *In Ph.D Thesis*.

Oduori C.O.A., P. Tongoona, J. Derera, O.M. Odongo. 2008. Studies of genetic components of agronomic traits and resistance to blast disease and Striga in six elite finger millet varieties of western Kenya. *In Ph.D Thesis*.

Oduori C.O.A., 2005. The importance and research status of finger millet in Africa. Nairobi: *The McKnight Foundation Collaborative Crop Research*.

Pennisi E, 2007. Taxonomy, wanted: a barcode for plants. *Science 318*, pp.190-191.

Pérez-Jiménez M., Besnard G., Dorado G., Hernandez P., 2013. Varietal Tracing of Virgin Olive Oils Based on Plastid DNA Variation Profiling. *PLoS ONE 8(8): e70507*. doi:10.1371/journal.pone.0070507

Rachie K. O and LeRoy V. P., 1977. The Eleusines: A review of the world literature. *ICRISAT – Hyderabad, India*. Pp. 2-51.

Ririe, K.M., Rasmussen, R.P., Wittwer, C.T., 1997. "Product differentiation by analysis of DNA melting curves during the polymerase chain reaction". *Anal Biochem.* ***245*** *(2):* 154–60, doi:10.1006/abio.1996.9916, PMID 9056205

Sachidanandam, Ravi; Weissman, David; Schmidt, Steven C.; Kakol, Jerzy M.; Stein, Lincoln D.; Marth, Gabor; Sherry, Steve; Mullikin, James C. et al., 2001. "A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms". *Nature* ***409*** *(6822):* 928–33. doi:10.1038/35057149. PMID 11237013

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

Santie M. De Villiers, Vincent N. Michael, Eric O. Manyasa, Annis N. Saiyiorri, Santosh Deshpande, 2015. Compilation of an informative microsatellite set for genetic characterisation of East African finger millet (Eleusine coracana). *Electronic Journal of Biotechnology vol. 18 (2):77-82,* doi:10.1016/j.ejbt.2014.12.001

Sheppard A. E., Timmis J. N., 2009. Instability of Plastid DNA in the Nuclear Genome. *PLoS Genet 5(1): e1000323*. doi:10.1371/journal.pgen.1000323

Stoeckle, M., 2003. Taxonomy, DNA and the bar code of life. Bioscience, 53: 796-797

Tamura K, Stecher G, Peterson D, Filipski A, and Kumar S, 2013. **MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0.** Molecular Biology and Evolution 30: 2725-2729.

Upadhyaya, H.D., et al., 2010. Genetic diversity for grain nutrients contents in a core collection of finger millet (Eleusine coracana (L) Gaertn.) germplasm. *Field Crops Res.* doi:10.1016/j.fcr.2010.11.017

Wendel JF. 2000. Genome evolution in polyploids. *Plant Molecular Biology 42*: 225–249.

Wise, Robert R. (2006). "The Diversity of Plastid Form and Function", *Advances in Photosynthesis and Respiration* ***23****. Springer. pp. 3–26.* doi:10.1007/978-1-4020-4061-0_1.

Wright Jane, 2013. Introductory phylogenetic workshop manual. Inqaba biotec.

APPENDICES

Appendix I: FINGER MILLET SAMPLE IDENTITIES

LAB NO.	ACCESSION NO.	SPECIES
1	EDL 30	*E. coracana*/hybrid *africana?*
2	EDL 15	*E. africana*
3	MS 19	*E. africana*
4	MS 9	*E. africana*
5	MS 13	*E. coracana*
6	UG 19	
7	MS 5	*E. coracana*
8	MS 4	*E. africana*
9	MS 17	
10	MS 21	
11	EDL 34	*E. intermedia/jaegeri* (Mt. Meru collection)
12	MS 18	
13	UG 1	*E. coracana*
14	UG 18	*E. africana*
15	MSN 10	*E. africana*
16	MS 8	*E. africana*
17	AAU-ELU-48	*E. floccifolia*
18	UG 9	*E. africana*
19	UG 10	*E. africana*
20	UG 11	*E. africana*
21	UG 18	*E. africana*
22	UG 20	*E. africana*
23	MS 3	*E. africana*
24	MS 6	*E. africana*
25	MS 7	*E. africana*
26	MS 11	*E. africana*
27	MS 12	*E. africana*
28	MS 15	

29	MD 48	*E. kigeziensis*
30	EDL 9	*E. africana*
31	EDL 16	*E. africana*
32	LEN 7	*E. africana*
33	TRAY 2 NO. 5	*E. coracana* (Cultivated)
34	LESK 10	*E. africana*
35	UG 3	*E. africana*
36	UG 8	*E. africana*
37	MS 16	
38	LEN 24	*E. africana*
39	EDL 25	*E. africana*
40	EDL 10	*E. africana*
41	AAU-ELU-49	*E. floccifolia*
42	TRAY 3 NO. 3	*E. coracana* (Cultivated)
43	TRAY 5 NO. 3	*E. coracana* (Cultivated)
44	AAU-ELU-03	*E. africana*
45	AAU-ELU-04	*E. africana*
46	AAU-ELU-05	*E. africana*
47	AAU-ELU-07	*E. coracana*
48	AAU-ELU-08	*E. africana*
49	AAU-ELU-10	*E. africana*
50	AAU-ELU-12	*E. africana*
51	AAU-ELU-13	*E. africana*
52	AAU-ELU-14	*E. africana*
53	AAU-ELU-15	*E. africana*
54	AAU-ELU-24	*E. coracana*
55	AAU-ELU-29	*E. africana*
56	AAU-ELU-30	*E. coracana*
57	AAU-ELU-31	*E. coracana* – cultivar Baruda
58	AAU-ELU-33	*E. africana*
59	AAU-ELU-35	*E. africana*
60	AAU-ELU-36	*E. africana*
61	AAU-ELU-37	*E. africana*

62	AAU-ELU-38	*E. africana*
63	AAU-ELU-39	*E. africana*
64	AAU-ELU-40	*E. coracana*
65	AAU-ELU-42	*E. africana*
66	AAU-ELU-44	*E. africana*
67	AAU-ELU-46	*E. africana*
68	AAU-ELU-58	*E. africana*
69	LESK 18	*E. africana*
70	LEN 24	*E. africana*
71	EDL 22	*E. africana*
72	EDL 25	*E. africana*
73	EDL 26	*E. coracana*/hybrid *africana?*
74	MS 1	*E. africana*
75	AAU-ELU-59	*E. africana*
76	AAU-ELU-60	*E. africana*
77	AAU-ELU-61	*E. africana*
78	AAU-ELU-62	*E. africana*
79	AAU-ELU-64	*E. multiflora*
80	AAU-ELU-65	*E. africana*
81	AAU-ELU-68	*E. floccifolia*
82	EDL 32	
83	AAU-ELU-01	*E. floccifolia*
84	AAU-ELU-02	*E. floccifolia*
85	AAU-ELU-06	*E. coracana*
86	AAU-ELU-54	*E. africana*
87	AAU-ELU-11	*E. africana*
88	AAU-ELU-16	*E. coracana*
89	AAU-ELU-57	*E. floccifolia*
90	AAU-ELU-18	*E. africana*
91	AAU-ELU-19	*E. africana*
92	AAU-ELU-20	*E. africana*
93	AAU-ELU-21	*E. africana*
94	AAU-ELU-22	*E. africana*

95	AAU-ELU-23	*E. africana*
96	AAU-ELU-25	*E. africana*
97	AAU-ELU-26	*E. africana*
98	AAU-ELU-27	*E. africana*
99	AAU-ELU-28	*E. africana*
100	AAU-ELU-32	*E. africana*
101	AAU-ELU-34	*E. africana*
102	AAU-ELU-47	*E. floccifolia*
103	UG 2	*E. africana*

N/B – Sample 14 is same as 21, 38 same as 70 and 39 same as 72 – accuracy check.

- The MS series are accessions collected from Kenya, UG series from Uganda, AAU-ELU series from Ethiopia and the rest from Tanzania.

Appendix II: THESIS RESEARCH PROJECT ACTIVITIES

Date: 22/10/2013

The Ugandan, Kenya and Tanzanian accessions were sown in a 6×11-hole tray.

Date: 4/11/2013

Re-sowing of accessions that failed to germinate

Date: 6/11/2013

DNA extraction of 16 samples using Isolate II plant DNA kit by Bioline® as per the manufacturer's protocol.

Lab. ID	Accession No
1	EDL 30
2	EDL 15
3	MS 19
4	MS 9
5	MS 13
6	UG 19
7	MS 5
8	MS 4
9	MS 17
10	MS 21
11	EDL 34
12	MS 18
13	UG 1
14	UG 18
15	MSN 10
16	MS 8

Date: 8/11/2013

DNA integrity of the extracted samples was checked by gel electrophoresis on agarose. 0.8g agarose was dissolved in 100ml of 1× TBE buffer and 5µl of gel red added to aid in visualization under UV light. The mixture was boiled in a microwave

for 3min for agarose to dissolve. Loading – 2µl DNA and 1µl 3× loading dye. The gel was run at 100V for 1 hr.

DNA extraction of the following accessions

Lab. ID	Accession No
17	UG 2
18	UG 9
19	UG 10
20	UG 11
21	UG 18
22	UG 20
23	MS 3
24	MS 6
25	MS 7
26	MS 11
27	MS 12
28	MS 15
29	MD 48
30	EDL 9
31	EDL 16
32	LEN 7
33	LEN 37
34	LESK 10.

Date: 13/11/2013

DNA extractions for the following accessions

Lab. ID	Accession No
35	UG 3
36	UG 8
37	MS 16
38	LEN 24
39	EDL 25

40	EDL 10
41	Cult 1
42	Cult 2
43	Cult 3

Sample 23 (MS 3) was re-extracted together with the above.

14/11/2013

DNA Quantification of the 43 samples using Qubit®Fluorometric Quantification as per the manufacturer's protocol. The Qubit® 2.0 Fluorometer was set as follows:

- Assay – DNA
- dsDNA broad range – since it was genomic DNA.
- The prepared standards I and II were read as prompted.
- The samples were read in succession and the data saved on a flash disc at the end.

15/11/2013

Ethiopian wild accessions were sown in three, 4×7-hole trays.

19/11/2013

DNA samples quantification with Qubit, nanodrop and agarose gel electrophoresis.

20/11/2013

DNA samples were diluted to 30ng/µl concentration with TE using the formula$C_1V_1 = C_2V_2$; thus; $V_1 = \frac{C_2V_2}{C_1}$. Where C_1= estimated concentration of the DNA sample; V_1= volume of the DNA to pick; C_2= final concentration to constitute e.g. 20ng/µL and V_2= final volume to constitute e.g. 100 µLs. Concentration of the diluted DNA was ascertained on 0.8% agarose against 20ng/µL and 40ng/µL λ DNA.

22/11/2013

DNA re-extraction was done on the following accessions which had very low, absent or degraded DNA.

Sample No.	Accession No.

1	EDL 30
3	MS 19
4	MS 9
5	MS 13
7	MS 5
8	MS 4
9	MS 17
10	MS 21
11	EDL 34
12	MS 18
14	UG 18
15	MSN 10
16	MS 8
17	UG 2
18	UG 9
19	UG 10
20	UG 11
22	UG 20
23	MS 3
24	MS 6
25	MS 7
26	MS 11
27	MS 12

28	MS 15
29	MD 48
31	EDL 16
32	LEN 7
33	LEN 37
34	LESK 10
74	MS 1

19/11/2013

DNA concentrations of a few samples were confirmed or compared with λDNA on agarose gel after Qubit reading. The samples and their Qubit® readings were;

1	17.4
14	44.2
23	30.2
30	13

Same concentrations were made as follows

Sample	Quantity taken	Water added	Final volume
Lambda DNA	1	3	4
1	2.9	1.1	4
14	1.1	2.9	4
23	1.66	2.34	4
30	3.85	0.15	4

2µL of 3× loading dye were added to make a volume of 6µLs.

20/11/2013

Samples were diluted to 30ng/µL. and the concentration confirmed on 0.8% agarose gel, ready for PCR.

The primers were also tested with a few samples. Their Tm temperatures are as follows:*rpl32-trnL* - F – 50.1, R – 53.8; *rps3* - F – 49.9, R – 47; *ndhF* - F – 51.3, R – 52.1. Stock concentration of primers was 100pM/µL. They were diluted to 2pM/µL.

PCR Master Mix for the 43 samples with rpl32-trnL primer was thus;

Component	**1× Vol.**	**For 50 samples**
PCR Buffer (10×)	1	50
MgCl2(50mM)	0.4	20
dNTPs	0.8	40
F primer	0.5	25
R primer	0.5	25
Taq polymerase	0.04	2
Water	4.76	238
DNA (40ng)	2µL	

PCR Profile

Step	Temperature	Time	
1st denaturation	94	1min	
2nd denaturation	94	3min	
Annealing	60	1min	35 cycles (loop).
Elongation	72	1min	

Final elongation	72	7min
Holding	15	∞

21/11/2013

PCR products were resolved on 2% agarose for 1hr at 100V. Loading was 2µL PCR products for 1µL 3× loading dye.

22/11/2013

DNA was re-extracted for samples 1, 3, 4, 5, 7, 8, 9, 10, 11, 12, 14, 15, 16, 17, 18, 19, 20, 22, 23, 24, 25, 26, 27, 28, 29, 31, 32, 33, 34 and 74.

26/11/2013

DNA was extracted for samples 44- 73.

27/11/2013

DNA integrity for the samples extracted on 22nd and 26th Nov. was done on 0.8% agarose gel.

DNA Quantification using Qubit® was done on some samples (listed below with their readings).

1	57.8
7	79.2
11	50.6
19	80.8
34	78.8
44	82.0
50	61.8
61	67.6

66 84.0

73 59.6

Readings of stds I – 116.31; II – 12185.60.

Fluorometer settings

- dsDNA
- broad range
- units – ng/µL
- DNA sample volume added 1µL.

29/11/2013

DNA quantification for the 74 samples (so far extracted) on 0.8% agarose. The gel picture was poor hence the activity was repeated on 2/12/2013.

2 and 3/12/2013

TE was prepared for purpose of diluting the final DNA samples for use in PCR.

1. Tris-HCl (1M)
    a. Molecular weight of Tris-base – 121.14g/mol.
    b. Add 30.285g Tris- base in 250ml water as follows. Add water up to 150ml then HCl to balance pH to 8.0 then top up.
2. EDTA (0.5M)
    a. Molecular weight – 372.24.
    b. Add 46.53g in 250 ml water, add NaOH pellets one at a time as you stir until dissolved. Or 23.265g in 125ml water.
3. TE (10mM Tris; 0.1mM EDTA)
    a. To make 100ml 1×TE: add 1ml (1000µL) Tris-HCl (1M) + 0.02ml (20µL) EDTA (0.5M) + 98.98ml sterile distilled water.

b. To prepare 100ml of 1mM Tris; 0.01mM ETA: add 10ml 1×TE + 90ml sterile distilled water.

It was autoclaved overnight.

Ethanol(100%)was diluted to 70% for use in sterilizing the surface and some apparatus.

N/B – in all dilutions the formula$C_1V_1 = C_2V_2$was used, using sterile (autoclaved) double distilled water.

TE Buffer is used to dilute DNA because it binds the DNA PROTECTING IT FROM FURTHER DEGRADATION AND SHEARING.

4/12/2013

DNA of the 74 samples was diluted with TE while primers were diluted with sterile double distilled water. Primers were diluted to 2pMoles/µL from 100pM/µL concentration.

5/12/2013

Some test PCRs were performed with some samples to test the primers and optimize the profiles. Reaction volume of 10µL (including 1.5µL DNA) was used. The profiles used were:

ndhF –1st denaturation at 94^{o}C for 5 min, 2nd denaturation at 94^{o}C for 3min, annealing at 59^{o}C for 1min, elongation at 72^{o}C for 2min and final elongation at 72^{o}C for 10min. Samples used were:

- Tube 1 – 1, 9, 17, 25, 33, 41, 49, 57.
- Tube 2 – 2, 10, 18, 26 34, 42, 50, 58.

rps3 – 1st denaturation at 94^{o}C for 5min, 2nd denaturation at 94^{o}C for 30sec, annealing at 53^{o}C for 1 min, elongation at 72^{o}C for 1 min and final elongation at 72^{o}C for 7min. samples were:

- Tube 3 – 11, 19, 27, 35, 43, 51, 59, 67.
- Tube 4 – 12, 20, 28, 36, 44, 52, 60, 68.

rpl32-trnL – 1st denaturation at 94°C for 5 min, 2nd denaturation at 94°C for 30sec, annealing at 60°C for 1 min, elongation 72°C for 2min and final elongation at 72°C for 10min. Samples were: tube 5= tube 1 and tube 6= tube 2 of *ndhF*.

9/12/2013

Gel electrophoresis of diluted DNA of the 74 samples. 3rd row last is λ20ng/µL.

10/12/2013

DNA was extracted for samples 75 – 82 and re-extraction for 3, 9, 14, 19, 26, 29 and 32.

11/12/2013

Integrity check of the above samples was done on 0.8% agarose gel run at 80V for 45min.

N/B – samples 82 (EDL 27) and 84 (LEN21) were discarded due to too low DNA concentration. Sample 83 (EDL 31) then became 82. Samples were quantified on agarose and diluted to 30ng/µL.

16 to 18/12/2013

Test PCRs for the 82 samples with the three primers were done.

PCR Master Mix for the reactions was thus:

Component	Stock Conc.	Reaction Conc.	1 rxn Volume	85 reactions
PCR Buffer	10×	1×	2	170
MgCl2	50mM	2mM	0.8	68
dNTPs	2mM	0.16mM	1.6	136
F –primer	2pMoles	0.2pMoles	2	170
R – primer	2pMoles	0.2pMoles	2	170
Taq	5U/µL	0.2U/µL	0.08	6.8

Polymerase				
Water			9.52	809.2
DNA	30ng/µL	30ng	2µL	
			20µL	

18µLs of the master mix was dispensed in each tube, and then 2µLs of sample DNA added in respective tubes.

There were too many dimers; it was established that primer concentration was too high; stock had been taken to be 10 pMol instead of 100pMoles. Very few samples amplified with *ndhF*.

19/12/2013

DNA was extracted for samples 83 – 106. Integrity was checked on agarose, they were quantified and diluted to 30ng/µL using TE.

20/12/2013

Gel electrophoresis of diluted samples – row 1 – 83-96, λ 25ng/µL; row 2 – 97-106, λ25ng/µL, λ 50ng/µL.

PCR for a *ndhF* few samples using *rps3* and primers

rps3

A – 2,10,18,26,34,42,50,58.

B – 23,31,39,47,55,63,71,79.

ndhF

C = A; D = B.

Master Mix

Component	1× volume	20 reactions

dNTPs	0.8	16
PCR Buffer	1	20
MgCl2	0.4	8
F-primer	1	20
R-primer	1	20
Taq polymerase	0.04	0.8
Water	4.26	85.2
DNA	1.5	
	10µL in each tube	

Profiles

rps3

- 94-5, (94-1, 58-1, 72-1)35, 72-7, 15-∞.

ndhF

- 94-5, (94-1, 53-1, 72-1)35, 72-10, 15-∞.

PCR products gel electrophoresis – 1st row – rps3, 2nd – *ndhF*.

Required profile for *rpl32-trnL*

94-5, (94-2, 60-1, 72-2)35, 72-10, 15-∞

After confirmation of the working of the primers and optimization of the primers, gel electrophoresis of all the diluted samples was done in preparation of the PCR for sequencing which was done in February 2014.

Appendix III: STANDARD PROTOCOL FOR PURIFYING PLANT GENOMIC DNA

The ISOLATE II Plant DNA Kit includes two different lysis buffers for optimal results with most common plant species. Please refer to section 7.3 (of product manual) for choosing the optimal lysis buffer system for your plant sample and for information on how to process more sample material than in the standard protocol.

Before you start:

- Ensure Wash Buffer PAW2 and RNase A are prepared (section 7.4 of product manual)
- Preheat Elution Buffer PG to 65^{o}C.

1. Homogenization

Homogenize up to 100mg wet weight or up to 20mg dry weight (lyophilized) plant material. Refer to homogenization methods (section 7.2 of product manual).

Proceed to cell lysis using Lysis Buffer PA1 (step 2a) or alternatively Lysis Buffer PA2 (step 2b).

2. Lysis

a. Cell lysis with Lysis Buffer PA1

Transfer resulting powder to a new tube and add 400µl Lysis Buffer PA1. Vortex mixture thoroughly.

Note: If sample does not resuspend easily e.g. due to plant powder absorbing too much buffer, add more Lysis Buffer PA1. Note that the volumes of RNase A (step 2a) and Binding Buffer PB (step 4) have to be increased proportionally.

Add 10µl RNase A solution and thoroughly mix sample. Incubate at 65^{o}C for 10 min.

Note: For certain plants, increasing incubation time to 30-60 min may be required.

Proceed to step 3.

b. Cell lysis with Lysis Buffer PA2

Transfer resulting powder to a new tube and add 300µl Lysis Buffer PA2. Vortex mixture thoroughly.

Note: If sample does not resuspend easily e.g. due to plant powder absorbing too much buffer, add more Lysis Buffer PA2. The volumes of RNase A, precipitation Buffer PL3 (step 2b), and Binding Buffer PB (step 4) however, have to be increased proportionally.

Add 10µlRNase A solution and thoroughly mix sample. Incubate at 65^{o}C for 10 min.

Note: For certain plants, increasing incubation time to 30-60 min may be required.

Add 75µl Precipitation Buffer PL3, mix thoroughly and incubate for 5 min on ice to precipitate SDS completely.

Proceed to step 3.

3. Filter crude lysate

Place an ISOLATE II Filter (violet) into a new Collection Tube (2ml) and load lysate onto column. Centrifuge for 2 min at 11,000 × g. Collect the clear flow-through and discard the ISOLATE II Filter.

Repeat the centrifugations step if not all liquid has passed through the filter.

If a pellet is visible in the flow-through, transfer the clear supernatant without disturbing the pellet to a new 1.5ml microcentrifuge tube (not supplied).

4. Adjust DNA binding conditions

Add 450µl Binding Buffer PB. Mix thoroughly by pipetting up and down 5 times or by vortexing.

5. Bind DNA

Place an ISOLATE II Plant DNA Spin Column (green) into a new Collection Tube (2ml) and load sample (max. of 700µl).

Centrifuge for 1 min at 11,000 × g and discard flow-through.

The maximum loading capacity of the ISOLATE II Plant DNA Spin Column is 700µ. For higher volumes repeat the loading and centrifugation steps.

6. Wash and dry silica membrane

Add 400µl Wash Buffer PAW1 to the ISOALTE II Plant DNA Spin Column. Centrifuge for 1 min at 11,000 × g and discard flow-through.

Add 700µl Wash Buffer PAW2 to the ISOLATE II Plant DNA Spin Column. Centrifuge for 1 min at 11,000 × g and discard flow-through.

Add another 200µl Wash Buffer PAW2 to the ISOLATE II Plant DNA Spin Column. Centrifuge for 2 min at 11,000 × g in order to remove wash buffer and to dry the silica membrane completely.

7. Elute DNA

Place the ISOLATE II Plant DNA Spin Column into a new 1.5ml microcentrifuge tube.

Pipette 50µl Elution Buffer PG (65^{o}C) onto the membrane. Incubate the ISOLATE II Plant DNA Spin Column for 5 min at 65^{o}C. Centrifuge for 1 min at 11,000 × g to elute the DNA.

Repeat this step with another 50µl Elution Buffer PG (65^{o}C) and elute into the same tube.

Appendix IV: QUBIT® PROTOCOL

Experimental Protocol

Performing the Qubit® dsDNA BR Assay

The protocol below assumes that you will be preparing standards for calibrating the Qubit® 2.0 Fluorometer. If you plan to use the last calibration performed on the instrument, you will need fewer tubes (step 1.1) and less working solution (step 1.3). More detailed instructions on the use of the Qubit® 2.0 Fluorometer (corresponding to steps 1.9–1.12 and 2.1–2.6) can be found in the user manual accompanying the instrument. For sample purity determinations, it is possible to use the Qubit® 2.0 Fluorometer to calculate the amount of dsDNA and RNA in the same sample. Simply perform each assay for your sample.

1.1 Set up the required number of 0.5-mL tubes for standards and samples. The Qubit® dsDNA BR assay requires 2 standards.

Note: Use only thin-wall, clear, 0.5-mL PCR tubes. Acceptable tubes include Qubit® assay tubes (500 tubes, Cat. no. Q32856) or Axygen® PCR-05-C tubes (VWR, part no. 10011-830).

1.2 Label the tube lids.

Note: It is important to label the lid of each standard tube correctly as calibration of the Qubit® 2.0 Fluorometer requires that the standards be introduced to the instrument in the right order.

1.3 Make the Qubit® working solution by diluting the Qubit® dsDNA BR reagent 1:200 in Qubit® dsDNA BR buffer. Use a clean plastic tube each time you make the Qubit® working solution. **Do not mix the working solution in a glass container.**

Note: The final volume in each assay tube must be 200 µL. Each standard tube requires 190 µL of Qubit® working solution, and each sample tube requires anywhere from 180–199 µL. Prepare sufficient Qubit® working solution to accommodate all standards and samples.

For example, for 8 samples, prepare enough working solution for the samples and 2 standards: ~200 µL per tube in 10 tubes yields 2 mL of working solution (10 µL of Qubit® reagent plus 1990 µL of Qubit® buffer).

1.4 Load 190 µL of Qubit® working solution into each of the tubes used for standards.

1.5 Add 10 µL of each Qubit® standard to the appropriate tube, then mix by vortexing 2–3 seconds. Be careful not to create bubbles.

Note: Careful pipetting is critical to ensure that exactly 10 µL of each Qubit® dsDNA BR standard is added to 190 µL of Qubit® working solution.

1.6 Load the Qubit® working solution into individual assay tubes so that the final volume in each tube after adding sample is 200 µL.

Note: Your sample can be anywhere from 1–20 µL, therefore, load each assay tube with a volume of Qubit® working solution anywhere from 180–199 µL.

1.7 Add each of your samples to assay tubes containing the correct volume of Qubit® working solution (prepared in step 1.6), then mix by vortexing 2–3 seconds. The final volume in each tube should be 200 µL.

1.8 Allow all tubes to incubate at room temperature for 2 minutes.

1.9 On the Home Screen of the Qubit® 2.0 Fluorometer, press **DNA**, then select **dsDNA Broad Range** as the assay type. The Standards Screen is displayed.

Note: If you have already performed a calibration for the selected assay, the Qubit® 2.0 Fluorometer prompts you to choose between reading new standards and using the previous calibration. See *Calibrating the Qubit® 2.0 Fluorometer* above for calibration guidelines.

1.10 On the Standards Screen, select to run a new calibration or to use the last calibration:

Press **Yes**to run a new calibration, then:

Insert the tube containing Standard #1 in the Qubit® 2.0 Fluorometer, close the lid, then press **Read**. The reading takes approximately 3 seconds. Remove Standard #1.

Insert the tube containing Standard #2 in the Qubit® 2.0 Fluorometer, close the lid, then press **Read**. Remove Standard #2.

OR

Press **No** to use the last calibration. The Sample Screen is displayed. Insert a sample tube into the Qubit® 2.0 Fluorometer, close the lid, then press **Read**.

After the measurement is completed, the result is displayed on the screen.

Note: The value given by the Qubit® 2.0 Fluorometer at this stage corresponds to the concentration after your sample was diluted into the assay tube. You can record this value and perform the calculation yourself to find out the concentration of your original sample (see *Calculating the concentration of your sample* below) or the Qubit® 2.0 Fluorometer performs this calculation for you (see *Dilution Calculator* on page 5).

1.11 To read the next sample, remove the sample from the Qubit® 2.0 Fluorometer, insert the next sample, then press **Read Next Sample**.

1.12 Repeat sample readings until all samples have been read.

Appendix V: USB® ExoSAP-IT® PCR Product Clean-up

BRIEF PROTOCOL (Affimetrix, Inc. 2011):

1. Remove ExoSAP-IT reagent from -20^{o}C freezer and keep on ice throughout this procedure.

2. Mix 5 µL of a post-PCR reaction product with 2 µL of ExoSAP-IT reagent for a combined 7 µL reaction volume. Note: When treating PCR product volumes greater than 5 µL, simply increase the amount of ExoSAP-IT reagent proportionally.

3. Incubate at 37^{o}C for 15 minutes to degrade remaining primers and nucleotides.

4. Incubate at 80^{o}C for 15 minutes to inactivate ExoSAP-IT reagent.

5. The PCR product is now ready for use in DNA sequencing, SNP analyses, or other primer-extension applications. Treated PCR products may be stored at -20^{o}C until required.

Note: Store ExoSAP-IT reagent in a non-frost-free freezer.