PHENOTYPIC AND GENETIC DIVERSITY OF SWEET SORGHUM (*Sorghum bicolor* L.) VARIETIES AND THEIR SUITABILITY FOR ETHANOL PRODUCTION IN TANZANIA

BY

SEPERATUS PASCHAL KAMUNTU

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN CROP SCIENCE OF SOKOINE UNIVERSITY OF AGRICULTURE.

MOROGORO, TANZANIA.

2010

ABSTRACT

Sweet sorghum (*Sorghum bicolor* L.) has been recognised widely as a potential alternative source of biofuel because of its high fermentable sugar content in the stalk. With international focus on alternative fuel sources, investors are looking at sweet sorghum as a feedstock source for bioethanol production without compromising sorghum's primary use for food, fodder, and animal feed. However, for researchers to maximize its feedstock potential and to be used in future there is a need to understand the genetic diversity and relationships among sweet sorghum landraces existing in Tanzania. The objective of this study was to assess the genetic diversity and relationships among sweet sorghum accessions using SSR markers, to examine brix level variability within accessions and to establish their genetic relationships for the informed use of these accessions in breeding programs. Forty three accessions [32 accessions from farmers' fields, 10 accessions from ICRISAT] were genotyped with 42 SSR markers that generated 220 alleles with an average of 5.07 alleles per locus. Polymorphic SSRs revealed a genetic diversity of 0.48 with a range of 0.07 to 0.87 and a total of 220 alleles. Cluster analysis using dissimilarity matrices and hierarchical clustering with un-weighted pair group method with arithmetic averages was employed to analyse the genetic relationships among accessions, thus 43 accessions were grouped into 3 distinct clusters. A sufficiently large degree of genetic diversity was observed in the collected germplasm and ICRISAT Materials. (>0.75 dissimilarity between *ICSV_93046* and *WEG_Bu*; between *ICSV_93046* and *NJG* and between *S_Ind_35* and *KULY_Mch*). However, further evaluation tagging gene for sugar content is important. The information obtained from this study coupled with phenotypic characterization can be used by plant breeders to select parents for improving the sugar content in the adapted sorghum varieties and to develop segregating populations to map genes controlling sugar content in sweet sorghum.

DECLARATION

I **SEPERATUS PASCHAL KAMUNTU** do hereby declare to Senate of Sokoine University of Agriculture that the work presented here is my original work and that it has neither been submitted nor being concurrently submitted for degree award in any other Institution.

_____ _____

Seperatus Paschal Kamuntu Date

(MSc Crop Science Candidate)

The above declaration is confirmed

_____ _____

Prof. S. Nchimbi-Msolla Date

(Supervisor)

COPYRIGHT

ACKNOWLEDGEMENT

I give much thanks to my supervisor Prof. Susan Nchimbi-Msolla whose constructive criticism, encouragement and tireless assistance made this study a success. My sincere thanks are due to, Dr. Mary A. Mgonja and Dr. C. Tom Hash of International Crop Research Institute for the Semi Arid Tropics (ICRISAT) for their support, guidance, assistance, and extensive comments throughout the entire MSc. study and dissertation writing.

At the same level, I wish to acknowledge with appreciation the sponsors; the Government of Tanzania through the Ministry of Agriculture, Food Security and Cooperatives and International Crop Research Institute for the Semi Arid Tropics (ICRISAT) for financing my studies at Sokoine University of Agriculture (SUA) and at MS Swaminathan Applied Genomics Laboratory (AGL), ICRISAT-India.

I would also like to extend my sincere thanks to Zonal Director of Agricultural Research Institute Ukiriguru for supporting and allowing me to pursue a Masters degree in Crop Sciences. More appreciation is due to the extension staff and farmers of Shinyanga and Mara regions who made the materials of my study availability to me.

Thanks to all staff at the Crop Science and Production Department-SUA, ICRISAT-BecA and –ICRISAT-AGL especially Dr. Rweyemamu, Dr. Santosh, Maggie, Usha, and all friends who were extremely helpful during the research and writing of this dissertation.

Last but not least, a big thank you to my wife Dynes, my parents and family members for your assistance, inspiration and encouragement during my study. You will always be remembered.

I give all the glory to God.

DEDICATION

To my wife; Dynes, my children; Jack, Lucius, Deric and Abigel for your support, understanding and patience during this study. I thank God for your love and kindness.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS AND SYMBOLS

A P	Andhra Pradesh
A	Adenine
ABI	Applied Biosystems
AFLPs	Amplified Fragment Length Polymorphisms
AGL	Applied Genomics Laboratory
ANOVA	Analysis of Variance
BecA	Bioscience East and Central Africa
bp	base pairs
Brix	*Refractometer brix* based on refractive indices at 20°C and the percentage by mass of total soluble solids of a pure aqueous sucrose solution
°C	degree Celsius
C	Cytosine
CRBD	Randomized Complete Block Design
CTAB	mixed alkyltrimethyl-ammonium bromide
CV %	Coefficient of Variation
DArT	Diversity Array Technology
dATP	deoxyadenosine 5’-triphosphate
dCTP	deoxycytidine 5’-triphosphate
DNA	Deoxyribonucleic acid
E	East
EtOH	Ethanol
FELISA	Farming for Energy for better Livelihoods in Southern Africa
G	Guanine
HEX	hexachloro-6-carboxy-fluorescine

ICRISAT	International crops Research Institute for the Semi Arid Tropics
InDel	Insertion-Deletion
IS	International Standard Unit
KAKUTE	Kampuni ya Kusambaza Teknolojia Limited
kb	Kilobases
litres t^{-1}	Liters per metric ton (1 metric ton equal to 1265 liters of ethanol)
LSD	Least Significant Difference
LZARDI	Lake Zone Agricultural Research and Development Institute
m.a.s.l.	Metera above sea level
MDS	Multi-Dimensional Scaling
ml	Millilitre (s)
NED	6-carbon-X-rhodamine (Yellow/Black)
ng	Nanogram = 10^{-9} gram
NJ	Neighbor-Joining
nm	Nanometer(s) = 10^{-9} meter
NPGS	National Plant Germplasm System
OD	Optical Density
PCs	Principal Components
PCA	Principal Component Analysis
PCoA	Principal Coordinate Analysis
PCR	Polymerize Chain Reaction
PIC	Polymorphic Information Content
RFLPs	Restriction Fragment Length Polymorphisms
r.p.m.	Revolution per minute
S	South
SNPs	Single Nucleotide Polymorphisms

SSR	Simple Sequence Repeats
STS	Sequence Tagged Sites
T	Thymine
t	Ton
Taq	"Taq Pol"- *Thermus aquaticus*
TaTEDO	Tanzania Traditional Energy Development and Environment Organisation
TBE	Tris-borate EDTA
TE	Tris-EDTA (buffer)
TTP	Thymidine 5’-triphosphate
UPGMA	Un-weighted Paired Group Method using Arithmetic Average
UPGMC	Un-weighted Paired Group Method using Centroid
UV	Ultraviolet
V	Volts
VETA	Vocation and Training Agency
µg	Microgram(s) = 10^{-6} gram
µl	Microliter(s) = 10^{-6} liter

CHAPTER ONE

1.0 INTRODUCTION

1.1 Background Information

Sweet sorghum is a saccharinae C4 crop which is often considered to be one of the most drought resistant agricultural crops as it reduces evapotranspiration. Thus, the crop has a potential of conserving water and the capability of remaining dormant under conditions of drought and high temperatures (Prashant, 2007). The crop comprises of biochemical and morphological specialization that increase net assimilation at high temperatures (Paterson *et al*., 2009). Some varieties posses the “stay green” genes that enable them to perform photosynthesis permanently (Mamoudou *et al*., 2006) and it is well established that, sorghum crop makes more efficient use of solar radiation than does sugarcane, with about one percent of photosynthetic efficiency for the latter and two percent for the former (Channappagoudar, 2007).

Sweet sorghum is the only crop that provides grain and stem that can be used for sugar, alcohol, syrup, jaggery, fodder, fuel, bedding, roofing, fencing, paper and chewing. Sweet sorghum has 10-25% of sugar in stalk juice at physiological maturity (Reddy *et al.*, 2007). The stem sugar concentration may be quantitatively measured by high performance liquid chromatography (HPLC) or as brix, a measure of soluble solids which in sorghums is sucrose (Murray *et al.*, 2009).

Like other sorghum types, sweet sorghum originated in East Africa and spread to other parts of African, Southern Asia, Europe, Australia and the U.S. (Tawanda, 2004). Although a native to the tropics, sweet sorghum is well adapted to temperate climates (Prashant, 2007). The plant grows to a height of from 120 to over 400 cm (Gnansounou *et al.*, 2004)

having a rapid growth, high sugar accumulation, biomass production potential and wide adaptability (Reddy and Sanjana, 2003).

Sweet sorghum sugar accumulation levels can be similar to that in sugarcane (*Saccharum* spp.), a close relative, though studies on enzymatic control and carbon transport suggest that the mechanism of accumulation in the two crops is different (Lingle, 1987; Tarpley and Vietor, 2007). The stems of sweet sorghum are desired for food-grade syrup (stalks are pressed and juice is subsequently boiled). They are also desired for fresh chewing and alcohol production in places like Brazil and India (House *et al.*, 2000).

Another research has shown that the sweet sorghum varieties and hybrids have the ability to produce high stalk yields of up to 50 t ha^{-1}, with juice brix reading between 18% and 22%, and 1.5 to 2.5 t ha^{-1} grains in comparison with Durra-bicolor (15-21%) (Reddy *et al.*, 2008).

The comparative advantages of sweet sorghum (Reddy *et al*., 2005) is that sweet it is best suited for ethanol production because of its higher reducing sugar content and thus it has poorer sugar content than is a case with sugar cane. The presence of reducing sugars in sweet sorghum prevents crystallization and thus making the crops cultivars to have a fermentation efficiency of 90% (Ratanavathi, 2005).

1.2 Problem Statement and Justification

There is a growing need for petroleum throughout the world. As the world economies grow so does the need for energy. Oil has been the major source of energy in many countries. The availability of oil at low prices has for many years made researchers have had little interest in alternative fuel sources (Prashant, 2007). A climatic concern on the use of fossil

fuels in the transportation sector has been a major reason of an increasing attention of scientists in embarking on biofuel research as an alternative source of energy.

The energy crisis has put enormous pressure on the price of oil due to uncertainties in the supply. This is coupled with a sharp increase in prices because of geopolitical tensions in the oil producing countries and the rapid economic growth in China and India both of which demand huge amounts of fossil oil to run their economies (Gitay *et al.*, 2002). All these have stimulated renewed interest in exploring how energy crops could play part in diversifying energy sources to complement fossil fuel (Jumbe *et al*., 2009).

According to (UN-Energy, 2007), the global production of biofuels doubled over the past 5 years and is anticipated to double again in the next 4 years. On the other hand, the demand of biofuels as noted in OECD/FAO (2007) report is expected to grow by 170% between 2006 and 2010 and will account for 25% for the world energy needs in the next 15-20years. This demand for biofuel is partially a result of the targets different countries have set in order to reduce the consumption of fossil fuel (UN-Energy, 2007).

In Sub Sahara Africa, Malawi is among a few countries that have started producing bioethanol from sugarcane molasses since the early 1980s. Other countries which have started similar programs include Kenya, Zimbabwe, Sudan, and Uganda (Jumbe *et al*., 2009). In Tanzania, a number of multinational companies, non governmental organizations (NGOs) and smallholder farmers are implementing a number of projects aimed at increasing the supply of liquid biofuels. More than ten companies have acquired land for establishing farms for the growing of Jatropha, and other energy crops for biofuels (UN-Energy, 2007; OECD/FAO 2007 and Jumbe *et al.*, 2009). Currently, Tanzania is producing 254 megalitres of ethanol from molasses (Hagan, 2007; Karekezi, 2007).

According to TaTEDO report (2008), in Tanzania, biofuel industry is still at the infant stage, only about 650 000 hectares are allocated for biofuel production from potential 88 million hectares of arable land. Some of the key actors (Table 1) include a number of multinational companies, non governmental organizations (NGOs) and smallholder farmers who are on the ground are at various stages of developing/promoting biofuels projects aimed at increasing the supply of liquid biofuels(UN-Energy, 2007; OECD/FAO 2007; Jumbe *et al.*, 2009). Many African countries are found within the tropics with vast arable land, fertile soils and favourable climate for growing tropical crops including energy crops. As pointed out by Jumbe *et al*. (2009), the increasing of biofuels production demands is creating demand for land in Africa for growing energy crops for biofuels especially first generation biofuels. 'First-generation biofuels' are biofuels made from sugar, starch, vegetable oil, or animal fats. The basic feedstock for the production of first generation biofuels are often seeds or grains which yield starch that is fermented into bioethanol.

In view of this, it is important that the research community focuses on biofuels feedstock and approaches that do not compete with food production but rather produce food as well as fuel, and enhance food production by stimulating increased input use and crop management intensity. For that matter, sweet sorghum stalks is falling in the ‘second generation biofuels’; lingo-cellulose feed stalk which will not compromise food security as only stalk juice is used for ethanol production (Reddy *et al*., 2005). ‘Second-generation biofuels’ do not compete with food or fresh water resources or cause deforestation, on the contrary they provide socioeconomic value to local communities.

Table 1: List of biofuels producer and other stakeholders

Name of the company	Started	Purpose	Location
Diligent Tanzania Limited	2004	Production of jatropha oil	Arusha, Coast, Tanga and Singida
PROKON	2005	Production of plant oil and biodiesel from jatropha and rapeseed	Mpanda
D1 Oils Tanzania Limited	2003	Biodiesel production from plant oil particularly *Jatropha curcus* and *Moringa oleifera*	-
Sun Biofuels Tanzania Limited	2007	Planning to plant Jatropha for biodiesel production	Kisarawe
SEKAB BioEnergy Tanzania Ltd	2007	To produce ethanol from sugarcane	Bagamoyo, Rufiji and Kilwa
Kikuletwa Farm	2004	Production of oil from jatropha and Aloe vera	TPC Moshi
Africa Biofuel and Emission Reduction Company (Tanzania) Ltd	2007	Production of biodiesel from *croton megalocarpus* trees	Biharamulo
Donesta Ltd and Savannah Biofuels Ltd	2007	Biodiesel production from sunflower and jatropha	Dodoma
FELISA		Production of biodiesel from oil palm and edible oil.	Kigoma
Bioshape	2006	Production of biodiesel from Jatropha	Kilwa
InfEnergy	2005	Production of palm oil and food crops only	Mngeta - Mvomero
BioMassive	2006	Biodiesel production from jatropha and Pongamia	Lindi
KAKUTE	2000	Promotion Jatropha for women economic activities	Monduli and Arumeru.
Jatropha Products Tanzania Limited	2005	Develop and Sensitizing the potential use of Jatropha	Have projects in 5 regions
TaTEDO	1990	Sensitizing the potential use of Jatropha	Have projects in more than 10 regions

Source: Songela, (2008).

In Tanzania, sweet sorghum as a biofuel feedstock will be a potential source for ethanol production due to the fact that the crop is adapted to the environment and is grown in most parts of the semi arid areas like the central zone, the western and in some of the northern parts of the country (Mbwaga and Obilana, 1993). However, there is a need for breeding program to understand the genetic diversity, and the source material available in the country in developing improved sweet sorghum varieties and hybrid parents. If the genetic base is too narrow there may be difficulties in breeding from the available material to develop the targeted required materials (Murray *et al.*, 2009).

In the National gene bank there are 818 sorghum accessions, whose characterization and identification are based on phenotypic expression and sometimes on the locality where they were collected (Sallu, 2008). There is no genetic information which distinguishes sweet sorghum from grain sorghum germplasm. Seetharama *et al.* (1987) and Ritter *et al*. (2007) suggest that sweet sorghums are polyphyletic origin with relatives of *S. bicolor* ssp *bicolor.* Molecular markers or sugar concentration level will be appropriate tools to set the criteria in distinguishing the two (Murray *et al.*, 2009). Understanding the genetic variability and the relationship among the germplasm is important in identifying the promising combinations for the exploitation of heterosis and establishment of heterotic groups for use as source materials in breeding program.

This study aims at determining morphological, sugar content, and genetic diversity of sweet sorghum accessions in the study area especially in the lake zone of Tanzania. Simple Sequence Repeat (SSR) technique, therefore, was adopted so as to provide a quick evaluation of genetic diversity and in understanding the complexity of accessions relationships. The information would assist sorghum breeders to maximize utilization of sweet sorghum for breeding varieties or hybrids potential for biofuels as a key aspect for ethanol production in Tanzania.

1.3 Objectives

1.3.1 Overall objective

To assess the genetic diversity of sweet sorghum germplasm as a potential source of feedstock in ethanol production in Tanzania.

1.3.2 Specific objectives

(i) To determine morphological characteristics of sweet sorghum landraces grown by farmers in the study area.

(ii) To determine the genetic diversity of sweet sorghum by SSR molecular makers.

CHAPTER TWO

2.0 LITERATURE REVIEW

2.1 Sorghum Production

Sorghum is ranked the fifth most important cereal crop in the world after wheat (*Triticum species*), rice (*Oryza species)*, maize (*Zea mays*) and barley (*Hordeum vulgadre*) in both total area planted and production (Agrama and Tuinstra, 2004). Sorghum is a C4 crop hence it is able to reduce evapotranspiration thus conserving water under conditions of drought and high temperatures. Some varieties posses the “stay green” genes that enable them to perform photosynthesis permanently (Mamoudou *et al.*, 2006).

Sorghum is well adapted to hot, semi-arid tropical environments with 400-600mm rainfall; as a result, it is a very stable source of nutrition (Smith, 2008). Due to its ability to adapt to different climatic conditions, sorghum is able to grow at an altitude ranging from sea level to 1,000 m above sea level. It is also found in temperate regions and at altitudes of up to 2300 m.a.s.l. in the tropics (Mamoudou *et al*., 2006). Its value in hot, arid or semi-arid areas is due to its ability to withstand dry conditions (Tawanda, 2004). The rainfall requirements for sorghum vary within the range of 350-700 mm per growing season depending on the length of the growing cycle, short growing cycle is 90 days and a long growing cycle is more than 130 days. In many semi-arid areas where sorghum is mainly grown for food, the annual rainfall ranges from 300-750 mm. Sorghum is deep rooted and grows on soil pH ranging from pH 5 to 8.5. It can withstand temperatures above 38 ^{0}C, but the best yields are realized at the temperature range of 24-27^{0}C. The crop is adapted to a wide range of soils, especially the heavy soils commonly found in the tropics (Mamoudou *et al*., 2006).

2.2 Sorghum Improvement in East African Countries

The Sorghum improvement programme in Tanzania Kenya and Uganda started in 1930 and ended in 1950 (Obilana, 2004). The programme started by using local popular selection Dobbs (from western Kenya) and L28 from Uganda (Doggett, 1988). These efforts resulted into development of early maturing white and brown popular resistance to bird attack. The brown grain SERENA variety was derived from the cross (Swazi P1270 x Dobbs) through pedigree breeding in 1956/57, and in 1958-1978 another three varieties were developed, two of which include -SEREDO (Serena x CK60) with brown grains and Lulu-D (SB77 x Seredo) with white grain which are still popular in east African countries (Reddy *et al*., 2008). ICRISAT came to the region in 1978 to assist in sorghum improvement; the focus being on using selected landraces as parents and adaptive testing of crossbreds. More than 500 improved sorghum varieties were identified, less than 10% of which are being grown by farmers in African countries including Tanzania (Obilana, 2004).

2.3 Sorghum Utilization

Normally, sorghum grain is ground or pounded after removing the pigmented pericarp. The flour is used to make porridge, bread or beer. Alternative uses of sorghum are increasingly becoming important while its uses as food are slowly declining. For example, there is an increasing demand for forage-sorghum genotypes (Rao *et al.,* 2004). Sorghum grain and stalk too have vast potential for industrial utilization. However, typically, sweet sorghum varieties have low grain yield and the grain is harvested for human or animal consumption (Channappaoudar *et al*., 2007). There is now a technology of producing sugar, alcohol, starch, semolina and malt products from sorghum grain.

According to Ratnavathi, (2008) sweet sorghum with a potential of 30-50 t ha^{-1} of juice biomass and 40-60% of juice recovery can serve as a potential alternative feedstock for

fuel grade ethanol. Similarly, grain sorghum can be used to replace as much as 50% of the corn in the animal feed rations without affecting animal performance (Ratnavathi, 2008). Grain sorghum is very similar to shelled corn in chemical composition except that the former is slightly higher in protein and contains little, if any carotene.

Sorghum landraces with sweet stalks are sparingly distributed across sorghum growing areas for instance in Africa and India. The green tender stalks are chewed like sugarcane. In Ethiopia, sweet-stalks sorghums are also used to make confectionary and alcohol (Prasado Rao, 1989).

The increased interest in the utilization of sweet sorghum for ethanol production in developed world has caused a re-evaluation of sweet sorghums as a source of energy (Rooney *et al.*, 2007 and Vermerris *et al.*, 2007). Other researches indicate the following: First sweet sorghum takes about 4 months to mature. Secondly water requirement of 8000 m^3 over two crops which is 4 times lower than of sugarcane (12-16 months and 3600 m^3 per crop respectively). Third the and the cost of production of ethanol from sweet sorghum per litre is slightly lower than that of from sugarcane molasses (Reddy *et al.*, 2005); also the cost of cultivation of sweet sorghum is three times lower than that of sugarcane (Dayakar *et al.*, 2004). Further the cultivation of sweet sorghum is economical in rain-fed areas where growing sugarcane is not economical. Up to 13.2 t/ha of total sugars, equivalent to 7682 litres of ethanol per hectare can be produced by sweet sorghum under favorable conditions (Kresovich and Henderlong, 1984).

2.4 Sorghum Production Constraints

The causes of low sorghum yields include losses due to biotic stresses such as diseases, insect pests (shootfly and stem boar) and *Striga*, *and* abiotic factors such as drought, high temperatures and low soil fertility (Ngugi *et al.*, 2002). The success in genetic enhancement of unimproved cultivars of sorghum can be affected by a combination of several factors, such as availability of genetic resources, inheritance and stability of the desired traits, simplicity and effectiveness of screening techniques, access to test environments, availability of technical manpower, financial and material resources (Rai *et al*., 1999).

Sorghum production has also been affected by policy biases towards certain favoured staple crops and lack of good seed industries and networking among various agencies. All these have limited the use of promising varieties (Obilana, 2004). This makes it difficult for farmers to adopt new varieties and the use of complimentary management practices for improving productivity and reducing risks. The trend also leads to underdeveloped marketing services for this crop. In Tanzania, sorghum research has been conducted by government institutions in collaboration with regional and international research organizations, in an effort to develop appropriate improved sorghum technologies for smallholder farmers. Despite such achievements, sorghum farmers still grow local varieties, while the adoption of improved varieties is generally low. However, as Mafuru *et al.* (2005) observed sorghum research activities have not adequately addressed quality attributes demanded by sorghum consumers in the market place. Therefore, most of the sorghum produced is consumed by farm households themselves.

2.5 Genetic Diversity

Understanding the distribution of genetic diversity among individuals, populations and gene pools is crucial for the efficient management of germplasm collections and breeding programs.

Diversity is the variation in living organisms within a given ecosystem (Yang *et al*., 2006). Genetic variation is recognized as one of the three fundamental levels of biodiversity, the other two being ecological diversity and species diversity. Genetic diversity is the variety of genes found among the individuals within a species. It is the raw material available to plant breeders (Peter *et al*., 2004). Plant genetic resources comprise the diversity of genetic material contained in traditional varieties and modern cultivars, as well as crop wild relatives and other wild plant species. These genetic resources can be used, now or in the future, for food and agriculture. They include resources which contribute to people's livelihoods by providing food, medicine, feed for domestic animals, fibre, clothing, shelter, energy and a multiple of other products and services.

Crop diversity is one of the most fundamentally important resources for human life on earth. It provides the natural, biological basis of our ability to grow the food required today, as well as to meet the challenges of population growth, changing climates and constantly evolving pests and diseases. Modern cultivars developed by scientific plant breeding for modern intensive agriculture have high genetic uniformity. Genetic uniformity makes the crops vulnerable to disease and pest epidemics as well as natural catastrophes (Liu *et al*., 2003). An example is the production of hybrid maize with limited diversity in the Corn Belt of the USA which led to the epidemic of southern leaf blight in 1970. All the hybrid maize cultivars adopted and produced in the 1960s had the cytoplasmic male sterility gene. This led to the loss of about 15% of the US maize crop in the early 1970s because the

cultivars were susceptible to the new race of *Helminthosporium maydis* (Liu *et al*., 2003). Therefore, the conservation and sustainable use of plant genetic resources are essential to the sustainable development of agricultural production hence the knowledge of crop genetic diversity in the cultivated crops is very important. Diversity analysis is routinely carried out using sequencing of selected gene(s) or molecular marker technologies (Yang *et al.*, 2006).

2.6 DNA Molecular Markers

Markers fall into one of the three broad classes: those based on visually assessable traits (morphological and agronomic traits), those based on gene product (biochemical markers), and those relying on a Deoxyribonucleic acid (DNA) assay (molecular markers).

Molecular markers should not be considered as normal genes, as they usually do not have any biological effect, and instead can be thought of as constant landmarks in the genome. They are identifiable DNA sequences, found at specific locations of the genome, and transmitted by the standard laws of inheritance from one generation to the next. The existence of various molecular techniques and differences in their principles and methodologies require careful consideration in choosing one or more of such marker types (Semagn *et al*., 2006).

More recently, DNA-based marker systems have been used successfully in DNA fingerprinting of plant genome and in genetic diversity studies (Agrama and Tuinstra, 2004). These are identifiable DNA sequences found at specific regions of the genome and transmitted by the standard laws of inheritance from one generation to the next. Molecular markers rely on a DNA assay while morphological markers rely on visible traits (Tawanda, 2004).

There are many types of molecular markers which have been developed and applied. This includes restriction fragment length polymorphisms (RFLPs), Amplified Fragment Length Polymorphisms (AFLPs), Random Amplified Polymorphic DNA (RAPD), Sequence Tagged Sites (STS), microsatellites also known as Simple Sequence Repeats (SSR) and Single Nucleotide Polymorphisms (SNPs) (Tawanda, 2004; Yang *et al.,* 2006). The application of any type of marker in the assessment of diversity among accessions will depend on the crop species, technical expertise, lab equipment and cost, suitability for the specific study and the desired results (Agrama and Tuinstra, 2004).

Although some of these marker types are very similar, some synonymous, and some identical (Reddy *et al*., 2002), there are still a wide range of techniques for researchers to choose upon. The various molecular markers can be classified into different groups based on: (a) Mode of transmission (bi-parental nuclear inheritance, maternal nuclear inheritance, maternal organelle inheritance, or paternal organelle inheritance), (b) Mode of gene action (dominant or co-dominant markers) and, (c) Method of analysis; hybridization-based or Polymarised Chain reaction (PCR) based markers).

Microsatellites or SSRs PCR based markers are short repeated DNA sequences in the genome, 2 to 4 nucleotides in length (Tawanda, 2004). These repeat motifs are flanked by conserved nucleotide sequences from which forward and reverse primers can be designed to amplify the DNA section by PCR containing the SSR. Simple sequence repeats require very small amounts of DNA (Daniel *et al*., 1999). SSR alleles can be separated by gel electrophoresis and visualized by silver-staining, autoradiography, or other staining with Ethidium bromide under UV-light. On one hand the automatic PCR product separation using fluorescent detection systems on DNA sequencers, such as ABI Prism, one of the selective primers must be labelled with different colour dyes (fluorophore) at the 5` end

such as 6-carboxyfluorescine (6-FAM), hexachloro-6-carboxy-fluorescine (HEX) or tetrachloro-6-carboxy-fluorescine (TET). On the other hand it is only fragments containing a priming site complementary to the fluorophore labelled primer will be detected by the sequencers. There are four essential elements of fluorescence detection system: (a) An excitation source, (b) A fluorophore (c) Wavelength filters to isolate emission photons from excitation photons and (d) A detector that registers emission photons and produces a recordable output, usually as an electrical signal or a photographic image.

Regardless of the application, compatibility of these four elements is essential for optimizing fluorescence detection. For high throughput analysis, three to nine different reactions labelled with different dyes can be multiplexed and loaded in a single lane (e.g., ABI Prism® 377 DNA Sequencer) or in a single injection (e.g. ABI Prism 3730 DNA Analyzer). An internal size standard labelled in a different colour need to be loaded to estimate the size of SSR amplification fragments using computer programs e.g. GeneScan and GeneMapper software's from Applied Biosystems.

SSR analysis is amenable to multiplexing and allows genotyping to be performed on large numbers of lines. They are reproducible, co-dominant and highly polymorphic markers. They also exhibit uniform genome coverage, and are transferable between mapping populations (Agrama and Tuinstra, 2004). SSRs are highly informative and abundant, occurring on average every 6 - 7 kb (Daniel *et al.*, 1999). These features have made them useful molecular markers.

Agrama and Tuinstra (2004) applied SSR and RAPD markers in sorghum germplasm analysis to compare suitability for quantifying genetic diversity and their results indicate that SSR markers were highly polymorphic with an average of 4.5 alleles per primer.

The RAPD primers were less polymorphic with nearly 40% of the fragments being monomorphic. The results also indicate that the genetic distances calculated from SSR data were highly correlated with the distances based on the geographic origin and race classifications.

The SSR markers had become powerful ideal tool for genotype assignment, marker assisted breeding, genetic mapping and diversity assessment (Gupta and Varshney, 2000). Comparative studies in crop plants have shown that microsatellite markers are more variable than most other molecular markers. Recent studies have shown that SSRs are conserved in related species, and may allow for the analysis of different species by the same microsatellite loci.

2.7 Analyses for Genetic Diversity

The use of established multivariate statistical algorithms is an important strategy for classifying germplasm, ordering variability for large number of accessions, or analysing genetic relationships among breeding materials. According to (Mohammadi and Prasanna, 2003), multivariate analytical techniques, which simultaneously analyse multiple measurements on each individual under investigation, are widely used in the analysis of genetic diversity irrespective of the dataset (morphological, biochemical, or molecular marker data). Among these algorithms, cluster analysis, principal component analysis (PCA), principal coordinate analysis (PCoA), and multidimensional scaling (MDS) are, at present, most commonly employed and appear particularly useful (Brown-Guedira *et al.*, 2000).

2.7.1 Cluster analysis

Cluster analysis according to Hair *et al*, (1995) is a group of multivariate techniques whose primary purpose is to group individuals or objects based on the characteristics they possess, so that individuals with similar descriptions are mathematically gathered into the same cluster. The resulting clusters of individuals should then exhibit high internal (within cluster) homogeneity and the high external (between clusters) heterogeneity (Mohammadi and Prasanna, 2003). Thus, if the classification is successful, individuals within a cluster shall be closer when plotted geometrically and different clusters shall be farther apart (Hair *et al*., 1995).

There are broadly two types of clustering methods: (i) distance-based methods, in which a pair-wise distance matrix is used as an input for analysis by the specific clustering algorithm (Mohammadi and Prasanna, 2003), leading to a geographical representation (such as a tree or dendogram) in which clusters may be visually identified; and (ii) model-based method, in which observations from each cluster are assumed to be random draws from some parametric model, and inference about parameters corresponding to each cluster and cluster membership of each individual are performed jointly using standard statistical methods such as maximum-likelihood or Bayesian methods (Pritchard *et al*., 2000).

Cluster analysis based on algorithms such as Un-weighted Paired Group Method using Arithmetic Average UPGMA, Un-weighted paired Group Method using Centroid UPGMC, Ward's, Single linkage and Complete Linkage has drawbacks. These techniques do not provide an objective definition of what constitutes an optimal tree or dendogram, and systematic errors are likely to be introduced during cluster analysis and tree reconstruction (Mohammadi and Prasanna, 2003). Such constraints may possibly be overcome by employing alternative methods such as neighbor joining or Fitch-Margoliash, which

remove the assumption that the data are ultra-metric (Swofford *et al.*, 1996). Neighbor joining has been more commonly used for phylogenetic studies; but very few researchers (Liu *et al.*, 2000) have applied this method for intra-specific differentiation in crop plants. Whatever algorithm is used for generating the dendogram, it is useful to carry out bootstrapping of the allelic frequencies (followed by calculation of genetic distances, etc.) to assess the reliability of the nodes.

2.7.2 Principle coordinate analysis (PCoA)

PCoA is a scaling or ordination method that starts with a matrix of similarities or dissimilarities between a set of individuals and aims to produce a low dimensional graphical plot of the data in such a way that distances between points in the plot are close to original dissimilarities. It allows visualization of differences among individuals and identifies possible groups. The reduction is achieved by linear transformation of the original variables into a new set of uncorrelated variables known as principal components PCs. The first step in PCA is to calculate eigen values, which define the amount of total variation that is displayed on the PC axes. The first PC summarizes most of the variability present in the original data relative to all remaining PCs. The second PC explains most of the variability not summarized by the first PC and uncorrelated with the first and so on (Mohammadi and Prasanna, 2003). The proportion of each variation accounted for by each PC is expressed as the eigen value divided by the sum of the eigen values. The eigenvector defines the relation of the PC axes to the original data axes.

CHAPTER THREE

3.0 MATERIAL AND METHODS

3.1 Plant Material

Sweet sorghum seeds used in this study were collected from farmers in Shinyanga and Mara Regions of Lake Zone in Tanzania in November 2008. The area lies within 33º 49' 8.9" E and 34º26'51.1" S with an altitude ranging from 1135 to 1369 m.a.s.l. Thirty sweet sorghum accessions were collected on the basis of sweetness of their stalks. These sorghum accessions were provided by farmers with their traditional knowledge of sweetness of sorghum stalks as opposed to grain sorghum stalks. Three to four panicles were randomly picked from each accession, and were bulked at Lake Zone Agricultural Research Institute Ukiriguru Mwanza (LZARDI).

3.2 Phenotypic Diversity Study

3.2.1 Experimental design

A trial was conducted at the Lake Zone Agricultural Research and Development Institute-Ukiriguru (LZARDI) in Mwanza and planted on November 2008. Enough seeds to sow a trial were sampled from 30 different accessions collected from farmers and 2 accessions from ICRISAT for phenotypic characterization (Table 2). A randomized complete block design (RCBD) single-row plots with 32 rows each with 10 m-long at a spacing of 75 cm between the rows and 15 cm within the row. The trial was replicated three times.

Table 2: Sweet sorghum accessions collected from Shinyanga, Mwanza and Mara-Tanzania and ICRISAT-Nairobi-Kenya

Sno	Accession name	Abbreviation	Origin	ID	Elevation (m)	South	East
1.	*Mwanagudungu*	*MWGD_Mnc*	Bariadi	134	1300	3º09'36.6"	33º51'39.1"
2.	*Mnanso*	*Mna*	Meatu	133	1220	3º32'30.0"	34º26'51.1"
3.	*Unknown*	*Unk_Ise*	Kishapu	125	1141	3º38'59.0"	33º50'58.7"
4.	*Wegita*	*WEG_Sa*	Bariadi	141	1271	2º48'10.3"	33º59'13.2"
5.	*Ngudungu*	*NGD_Ki*	Bariadi	137	1336	2º46'40.9"	34º09'29.9"
6.	*Gudungu*	*GDG_Chi*	Bariadi	140	1369	2º53'47.5"	34º10'22.4"
7.	*Kinyenche*	*KINY*	Bariadi	141	1271	2º48'10.3"	33º59'13.2"
8.	*Limuche*	*LIM*	Kishapu	127	1152	3º38'20.9"	33º53'18.9"
9.	*Wegita-Bunda*	*WEG_Bu*	Bunda	142	1238	2º01'25.5"	33º52'24.3"
10.	*Kenya Redi (mchanganyiko)*	*KENR*	Maswa	135	1291	3º08'27.9"	33º52'11.5"
11.	*Samumba Serena*	*SA_Se*	Meatu	133	1220	3º32'30.0"	34º26'51.1"
12.	*Tegemea/selemani (Mchanganyiko)*	*TEG-Sel*	Maswa	135	1291	3º08'27.9"	33º52'11.5"
13.	*Tegemea*	*TEG_Bu*	Bariadi	137	1336	2º46'40.9"	34º09'29.9"
14.	*Mkungu nyamatu*	*MKG_At*	Bariadi	141	1271	2º48'10.3"	33º59'13.2"
15.	*Ngudungu*	*NDH_Chi*	Bariadi	139	1368	2º53'47.5"	34º10'22.4"
16.	*Ng'holongo Tembe*	*NGH_Te*	Meatu	133	1220	3º32'30.0"	34º26'51.1"
17.	*KSV 700*	*KSV_700*	ICRISAT				
18.	*Mwanagudungu*	*MWA_Ma*	Bariadi	141	1271	2º48'10.3"	33º59'13.2"
19.	*Lulu*	*LUL*	Bariadi	141	1271	2º48'10.3"	33º59'13.2"
20.	*Mwanagudungu*	*MWA-Pj*	Meatu	133	1220	3º32'30.0"	34º26'51.1"
21.	*Nkombituna*	*NKO_Ko*	Kishapu	127	1152	3º38'20.9"	33º53'18.9"
22.	*Wegita Nyepe*	*WEG_Ny*	Bunda	142	1238	2º01'25.5"	33º52'24.3"
23.	*Unknown*	*UNK_Ki*	Kishapu	122	1135	3º37'36.1"	33º 49' 8.9"
24.	*Serena&Selemani Mchanganyiko*	*SEL_Mch*	Kishapu	129	1192	3º30'55.7"	34º18'13.2"
25.	*IS23/31*	*IS23/31*	ICRISAT				
26.	*Redi*	*RMCH*	Maswa	135	1291	3º08'27.9"	33º52'11.5"
27.	*Ng'wanagudungu -Sangija*	*NGW_Sa*	Bariadi	140	1369	2º53'47.5"	34º10'22.4"
28.	*Ng'holongo*	*NGH_Chi*	Baradi	140	1369	2º53'47.5"	34º10'22.4"
29.	*Mkombituna*	*MKO*	Meatu	133	1220	3º32'30.0"	34º26'51.1"
30.	*Kulya mchanganyiko*	*KULY_Mch*	Bariadi	140	1369	2º53'47.5"	34º10'22.4"
31.	*Ng'wajinila*	*NGJ*	Meatu	133	1220	3º32'30.0"	34º26'51.1"
32.	*Nkulya Nyeupe*	*NKLY_Ny*	Bariadi	140	1369	2º53'47.5"	34º10'22.4"

3.2.2 Data collection

The data were collected using the International Board of Plant Genetic Resources IBPGR and ICRISAT IBPGR/ICRISAT, (1993) procedure. Data was recorded according to Ali, (2008) whereby three randomly sampled plants at the middle row of each accession were used and the mean was obtained. Plants that were used for scoring were tagged prior flowering to avoid biasness. The following variables in Table 4 were recorded: Days to 50% flowering (days), number of nodes at maturity, plant height at maturity (cm), panicle

length (cm), peduncle length (cm), 100-seed weight (g) average number of seeds per panicle, and grain yield (t ha^{-1}). At physiological maturity, according to Murray (2009), six sweet sorghum stalks samples were randomly harvested, weighed and juice extracted on accession basis. Sorghum grains were shelled, dried and weighed. The extraction of juice was done by a motorized machine made by vocation educational and training agency (VETA) Mwanza. The brix content was estimated by hand refractometer according to Ali *et al.* (2008). The data were recorded using the excel data-sheet.

3.2.3 Data analysis

Grain yield and brix value are of primary importance in this study, therefore all evaluation were made including yield data and juice content. The data were analyzed using MSTAT C package version 2.1. The analysis of variance (ANOVA) was calculated using the one way (in randomized blocks) method in order to come up with F-statistics (Gomez and Gomez, 1984). Means were separated by Duncan Multiple Range Test at 5%.

Statistical model:

$$Y_{ijk} = \mu + B_i + T_{ij} + \epsilon_{ijk}$$

Where: Y_{ijk} = Response.

μ = General effect.

B_i = Block effect.

T_{ij} = Treatment effect.

ϵ_{ijk} = Random effect.

3.3 Genetic Diversity Study

Forty two accessions were used in this study during DNA extraction at BecA-ICRISAT-Nairobi centre (Appendix 1); 32 accessions from Tanzania and ten accessions from ICRISAT Nairobi. These ten accessions were known as sweet sorghum and were included purposely as a control for comparison.

3.3.1 DNA extraction

About 20 seeds of each of 42 accessions were sterilized by 70% EtOH to get rid of fungal infections. These seeds were grown in well-labelled petri-dishes and incubated at 27^{o}C in the laboratory at BecA Nairobi. A blotting moist paper was placed in each petri-dish to initiate seed germination. After seven days, 1.0-1.2 g roots and leaf tissues were harvested and bulked from each accession for DNA extraction as per Mace *et al*. (2004) with minor modifications.

Using this procedure (96-PCR tube rack and PCR-tube stripes each containing 8 tubes) were labelled for orientation. Two stainless still balls were placed in each tube stripe and put in liquid Nitrogen to chill the tubes and dry the roots and leaf tissues. The tubes were closed using strip caps after filling each 8 tubes to avoid mixing.

The extraction buffer was prepared by adding 170 µl ß-mecaptoethanol to 100ml CTAB and then putting in 65°C water bath. The CTAB was heated to soften the tissue for grinding and to increase the number of cells released after grinding. The strip caps were removed and 450 µl of pre heated (65°C) extraction buffer (CTAB) was added into each tube using a multi-channelled pipette and the tubes covered. When the extract mixes with CTAB, the cells rapture and lyses for DNA to come into the solution.

The chilled plate with roots and leaf tissues were weighed then set on the genogrinder machine. The genogrinder machine was run at 1000 strokes per minute for 10 minutes to grind the roots and leaf tissues. The macerated substance was incubated for 30 minutes at 65°C in the water bath with occasional mixing. The plate with macerated substance was weighed and then centrifuged for 2 minutes at 3500 revolutions per minute (r.p.m.) to make the ground material to settle at the bottom of the strips so that contamination will not take place when opening the caps.

Solvent extraction was done by adding 450 µl chloroform: isoamylalcohol (24:1) to each tube using the multi-channelled pipette and inverted twice to have it fully mixed. The chloroform: isoamylalcohol at this stage also removes the proteins from the cells. The plates were centrifuged at 5000 r.p.m. for 15 minutes at low temperature (+4°C). Approximately 450 µl of the supernatant was transferred into fresh strip tubes.

Equal volume (450 µl) of isopropanol (stored at -20°C) was added and the tubes inverted once to have it fully mixed. The plates were centrifuged at 5000 r.p.m. for 15 minutes. This was necessary to precipitate the crude DNA pellet. The supernatant was decanted and the DNA pellet air-dried for 30minutes. 200 µl low salt TE buffer was added to each sample to dissolve the DNA and 3 µl RNase-A (10 mg/ml) added to each of the sample to remove RNAs. The DNA samples were incubated for 30 minutes at 37°C overnight at room temperature (in the dark).

A second solvent extraction was done by adding 200 µl phenol: chloroform: isoamylalcohol (25:24:1) to each sample and inverting twice to have it mixed (phenol

removes proteins), then centrifuged at 5000 r.p.m. for 5 minutes. A fixed volume of 180 µl of the top layer was transferred to fresh strip tubes and chloroform: isoamylalcohol (24:1) added to each tube and then inverted twice to have it mixed (chloroform: isoamylalcohol removes any traces of phenol and proteins). The plates were centrifuged at 5000 r.p.m. for 5 minutes. A fixed volume of top layer (approximately 180 µl) was transferred to fresh strip tubes.

To purify the DNA, 315 µl ethanol: sodium acetate solution was added to each sample and placed in -20^0C for 5 minutes then centrifuged at 5500 r.p.m. for 5minutes in $+4^0$C. The supernatant from each sample was decanted and the pellets washed by adding 150 µl of 70% ethanol then centrifuging at 5500 r.p.m. for 5 minutes. The supernatant from each sample was decanted and the pellet air-dried for approximately 30 minutes. The pellet was re-suspended 100 µl low salt TE buffer.

3.3.2 DNA quality

After isolation of DNA, quantification and analysis of quality are necessary to ascertain the approximate quantity of DNA obtained and the suitability of DNA sample for further analysis. This is important for PCR amplification of target DNA.

The genomic DNA was checked for quality and quantified by gel electrophoresis. 1.5 µl DNA sample was mixed with 1.5 µl 5x loading bromo-phenol blue dye and 2 µl double distilled water were loaded into 0.8% agarose gel wells stained with ethidium bromide (10mg/ml) submerged in an electrophoresis unit containing 1 x TBE buffer. An electric current of 100V was applied for 60 minutes. This method is based on the ethidium bromide fluorescent staining of DNA. Ethidium bromide is a fluorescent dye, which intercalates

between the stacked bases. After the run of gel it was removed and photographed under UV light using a video capture system (Flowgen IS 1000) (Fig. 1), UV was absorbed by the DNA and transmitted to the dye and the bound dye itself absorbed radiation. This energy was retransmitted to the region of the visible spectrum. The quantity of DNA was estimated by comparing the fluorescent yield of the samples with standards, lambda (λ) DNA at known concentrations of 25 ng, 50 ng and 100 ng filled at the outer wells at the left edges of the gel. This provided a very rapid and sensitive means of estimating the nucleic acid concentration. The gel indicated the DNA, which migrates as a tight band indicating high molecular weight and/or presence of RNA. The DNA samples 16, 29, and 41 were visually identified on the gel as not having good quality and they could possibly therefore be degraded/sheared (Fig. 1). However, the nano-drop spectrophotometer was used to measure the absorbance to get the concentration and purity. According to Hoisington *et al.* (1994), Nano-drop Spectrophotometric determination provides a simple and accurate estimation of the concentration of nucleic acids in a sample. Purines and pyrmidines in nucleic acid show maximum absorption of around 260nm (e.g. dATP: 259 nm; dCTP: 272 nm; dTTP: 247 nm) if the DNA sample is pure without significant contamination from proteins or organic solvents. The ratio of OD_{260}/OD_{280} was determined to assess the purity of the sample. The quantified genomics DNA (Appendix 5) was transported to MS Swaminathan Applied Genomics Laboratory (AGL) at ICRISAT, Patancheru, India.

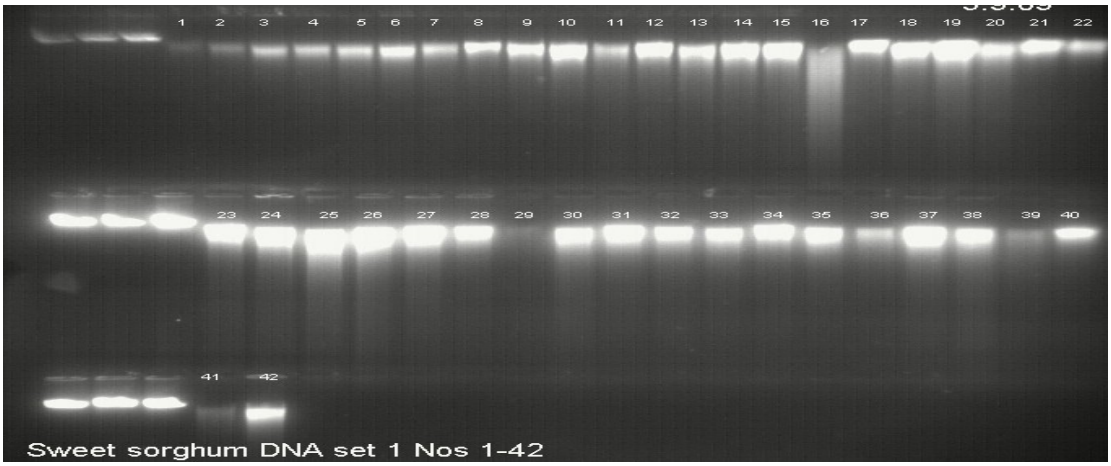

Figure 1: Agarose gel images showing the bulk DNA extracted from 42 sweet sorghum accessions.

3.3.3 DNA normalization

The quality and quantity of the DNA was further standardized by comparing with known concentration of uncut-λ DNA standards) using agarose gel electrophoresis at AGL to ensure uniformity in the results. Sample DNA was subsequently diluted to a concentration of 2.5 ng/µl.

3.3.4 Optimization of primers and DNA amplification (PCR)

The objective of conducting this procedure is to select the best candidate markers and then to optimize conditions for their amplification. Optimization of microsatellite systems involves a more or less comprehensive survey of PCR conditions for amplification of candidate loci. The aim here is to focus on signal strength and purity in order to obtain products from various loci with non-overlapping ranges of allele sizes, which can be amplified with similar efficiency under a standard set of conditions to enable multiplexing for high throughput analysis (Schlotterer, 1998).

However, another six DNA (*ICSL73200*, *ICSL73221*, *ICSV111*, *S35*, *SSV84* and *BTx632*) samples from ICRISAT-Patancheru India were included to make a total of 48 DNA samples (Appendix 1). Six DNA samples were purposely included to make an economical use of a 96 DNA well plates which are normally used in the polymerase chain reaction (PCR) machine during DNA amplification. In this case, 48 DNA samples were doubled to make a complete 96 DNA well plate.

3.3.5 SSR amplification

Simple sequence repeat (SSR) markers were used for molecular genotyping of the sweet sorghum accessions in this study. A set of 48 sorghum SSR markers, well distributed across sorghum genome, were used for genotyping (Table 3).

This included a set of 39 sorghum SSR markers that were used for fingerprinting a core collection of 3365 sorghum accessions in a Generation Challenge Program project. Primer pairs for the SSR markers used were previously defined by Brown *et al*. (1996) (3 *Xgap* markers), Taramino *et al*. (1997) (*SbAGB*02), Kong *et al*. (2000) (15 out of 17 *Xtxp* markers), Bhattramakki *et al*. (2000) (*Xtxp*114 and *Xtxp*320), Schloss *et al*. (2002) (7 *Xcup* markers), Ramu *et al*. (2009) (*Xisep*0310 and *Xisep*0107), and CIRAD (6 *gpsb* and 12 *mSbCIR* markers). Further genotyping analysis was accomplished as per Ramu *et al*. (2009).

Table 3: Characteristics of each locus: repeat type and number of alleles

Marker	Repeat motif	No. Allele.
Xtxp040	(GGA)7	1
mSbCIR223	(AC)6	2
mSbCIR276	(AC9	2
Xcup53	(TTTA)5	2
Xisep0310	CCAAT(4)	3
gpsb148	CA	5
Xisep0107	TGG(4)	3
Xtxp278	(TTG)12	3
Xtxp321	(GT)4 +(AT)6	5
Xtxp273	(TTG)20	7
Xcup11	(GCTA)4	3
Xcup62	(GAA)6	3
Xtxp136	(GCA)5	2
Xtxp012	(CT)22	2
mSbCIR246	(CT)7.5	4
mSbCIR262	(CATG)3.25	4
Xcup63	(GGATGC)4	3
Xcup14	(AG)10	4
mSbCIR240	(TG)9	5
mSbCIR300	(GT)9	3
gpsb067	GT	4
Sb4_72	(AG)16	4
mSbCIR248	(GT)7.5	5
gpsb123	AC	6
mSbCIR238	(AC)26	7
mSbCIR306	(GT)7	4
Xtxp021	(AG)18	4
Xcup61	(CAG)7	4
gpsb089	TG	6
SbAG_B02	(AG35)	7
Xtxp145	(AG)22	5
mSbCIR286	(AC)9	5
mSbCIR329	(AC)8.5	5
gpsb151	CT	7
Xtxp015	(TC)16	6
Sb6_84	(AG)14	7
Xtxp010	(CT)10	6
Xtxp320	(AAG)20	8
Xtxp057	(GT)21	9
Xtxp295	(TC)19	10
gpsb069	TC	13
Xtxp265	(GAA)19	15

3.3.6 Capillary electrophoresis and fragment analysis

Primer pairs for each locus, where one primer was labelled at the 5' end with FAM, VIC, PET or NED fluorescent dyes were obtained from Applied Biosystems (MS Swaminathan Genomic Service lab GSL). Multiplex PCRs were performed according to Matsuoka *et al.* (2002) with two to four SSR loci were assayed simultaneously in each reaction. The PCRs were performed in 25-µl volumes containing 2.5 ng of template DNA, 10 pmol of each of forward and reverse primers for each locus, 10X PCR buffer (Promega, Madison, WI), 0.2 mM dNTPs, 50 mM MgCl2 and 1.0 U/µl *Taq* polymerase (Promega).

The reactions were carried out using the Gene Amp PCR systems 9600 thermocycler (Applied Biosystems) and touch-down PCR amplification. The PCR temperature profile consisted denaturisation step at 94°C for 4 minutes initial temperature, followed by 94°C for 1 minute, 51°C for 1minute and 72°C for 1 minute. The next 40 cycles were at 94°C, 51°C and 72°C for 1 minute each. In the last cycle, extension time at 72°C was increased to 20 minutes and then it drops down to 10°C which is the final temperature.

Samples containing 2 µl of PCR products, 6.3 µl of formamide, 0.2 µl GenescanTM -5-LIZTM size standard (Applied Biosystems) and 2.5 µl of distilled water were denatured at 94°C for 5 min and assayed on a capillary DNA sequencer (Applied Biosystems, model 3730). After denaturation, the reaction plates were taken to Applied Biosystems (ABI) for analysis.

3.3.7 Genetic diversity data analysis

Allele-calling was done using Genemapper 4.0 ABI. Allele sizes were exported as an Excel file and the datawere analyzed using PowerMarker V3.25 (Liu, 2005). Allelobin

software was used to check the quality of markers. The parameters for allele scoring and heterozygosity include: allele height, the number of base pairs (bp)/allele size and the number of alleles per loci.

The software generated a number of alleles and polymorphic information content (PIC) for each of the SSR loci. The formula used for calculating plolymorphic information content is:

$$PIC = 1 - \sum n (P_i)^2$$

Where n = Total number of alleles present in a primer

P_i = Proportion of number of alleles present in accessions

A dissimilarity matrix was generated from the data and subsequently grouped by weighted Neighbor-Joining method using DARwin v5 (Perrier *et al.*, 2003). The dissimilarity coefficient was used to perform principal coordinates of analyses (PCoA) and constructed weighted Neighbor-Joining Tree (Saitou and Nei 1987) with a bootstrapping value of 10000 using DARwin v5.0.

CHAPTER FOUR

4.0 RESULTS AND DISCUSSION

4.1 Phenotypic and Sugar Content Analysis

The evaluated traits under field condition were growth traits, agronomic performance and yield response using visual observations and measurements on standard scale. The results from ANOVA in Table 4 on variables of yield components characteristics of accession show high significant differences among wet stalk weight, 100 grain weight, the number of seeds per panicle, brix value, and days to 50% flowering. On the other hand, grain yield, plant height, panicle width, panicle length, and the number of internodes were not significant ($P \leq 0.05$).

The significant variation of wet stalk, grain yield and brix value indicates that it is possible to make a selection of the best accessions by using these traits. Such variability between accessions shows that there is an opportunity for improvement if the right material is selected (Edmeades *et al.*, 1997a).

Table 4: ANOVA extract and probability levels for variables scored

Source of variance	DF	Wet stalk weight (kg)	100 grain weight (gm)	Brix % value	50 % flowering (days)	Grain yield (t ha^{-1})	Plant height (cm)	Panicle width (cm)	Panicle length (cm)	No. of internodes
Replication	2	1.009	0.978	44.648	116.010	40.598	4316.715	11.729	22.582	26.042
Accessions	31	1.497***	0.376***	27.015***	202.440***	4.680	861.067	3.291	20.769	1.407
EMS	62	0.195	0.095	5.511	23.150	2.548	868.016	3.561	17.200	1.472
Total	95									

Note: *= Probability significant at 0.05 level; ** = significant at 0.01 and ***= significant at 0.001 level, EMS = Error mean squares

4.1.1 Wet stalk weight

The wet stalk weight differed significantly among accessions as indicated in Table 4. The top three accessions; *NG_Te* (4.5 Kg), *NGJ* (2.7 Kg) and *KSV_700* (1.9 Kg) recorded a significant higher wet stalk weight (Table 5), and this indicates that these three accessions are potential to produce high biomass. As Reddy *et al*. (2007) report there is a strong positive correlation among wet stalk weight, plant height, juice volume and sugar yield. Breeders would like to develop tall varieties and/or hybrids which will give high biomass for folder as well as juice volume for ethanol production. In this study, the juice volume was not recorded.

Table 5: Performance of 32 sweet sorghum accessions evaluated at (LZARDI) Ukiriguru Mwanza in 2009

S/no	Accession Name	Wet stalk weight (kg)	100 grain weight (gm)	Brix %	Day to 50% Flowering (days)	Grain yield t ha^{-1}	Plant height (cm)	Panicle length (cm)	Mean peduncle length (cm)	No. of nodes
1.	MWGD_Mnc	1.1	2.5	8.6	74	5.1	156.9	22.6	27.9	9
2.	Mna	1.6	2.3	14.0	87	3.6	187.8	20.1	38.4	10
3.	Unk_Ise	0.8	2.1	14.3	77	6.5	188.4	16.0	23.9	9
4.	WEG_Sa	0.9	1.9	9.8	77	5.1	158.0	15.9	31.9	9
5.	NGD_Ki	1.2	2.7	8.6	74	6.0	153.9	22.3	33.0	9
6.	GDG_Chi	0.8	2.4	11.5	74	3.3	154.8	19.6	28.3	10
7.	KINY	1.1	1.8	13.7	93	5.3	159.5	16.6	27.1	9
8.	LIM	1.3	2.3	13.2	87	5.1	204.5	20.4	31.3	11
9.	WEG_Bu	1.5	2.3	6.1	80	5.9	161.2	17.0	31.8	9
10.	KENR	1.6	1.9	12.5	84	3.3	179.7	15.6	31.2	10
11.	SA_Se	0.9	2.4	7.3	76	5.4	142.3	17.5	27.6	9
12.	TEG-Sel	1.1	2.5	11.7	84	3.3	163.1	13.3	19.3	9
13.	TEG_Bu	1.3	2.1	14.5	83	4.3	157.6	19.8	35.8	10
14.	MKG_At	1.0	1.7	7.3	77	3.7	164.1	19.3	33.8	10
15.	NDH_Chi	1.6	2.7	4.9	80	7.1	155.8	17.1	24.7	10
16.	NGH_Te	4.5	1.9	13.5	103	3.7	160.0	23.9	32.0	8
17.	KSV_700	1.9	2.5	13.7	88	2.7	125.8	14.8	27.5	9
18.	MWA_Ma	1.2	2.3	10.9	77	4.2	186.0	20.7	32.0	9
19.	LUL	1.0	2.9	11.7	77	3.2	175.3	17.6	31.8	11
20.	MWA-Pj	1.0	2.6	5.5	75	5.4	184.5	14.9	20.9	10
21.	NKO_Ko	1.0	2.4	8.2	75	4.0	181.9	15.9	32.7	10
22.	WEG_Ny	1.2	1.9	7.1	79	5.7	163.6	18.3	26.0	10
23.	UNK_Ki	0.8	2.2	10.4	83	3.4	192.1	18.8	30.5	9
24.	SEL_Mch	0.7	2.4	8.3	73	4.0	158.1	23.7	37.3	9
25.	IS23/31	0.9	2.7	15.0	73	2.3	154.8	17.9	31.7	9
26.	RMCH	1.0	2.3	12.3	83	4.6	172.2	18.9	24.1	11
27.	NGW_Sa	1.2	2.7	10.2	76	6.2	152.9	18.2	30.3	10
28.	NGH_Chi	0.8	1.9	5.9	79	2.9	176.2	17.9	31.6	9
29.	MKO	1.2	2.3	12.0	74	3.8	168.3	16.7	27.1	9
30.	KULY_Mch	1.1	1.8	8.8	80	5.3	196.8	19.6	33.2	10
31.	NGJ	2.7	1.5	14.2	108	2.7	176.9	18.6	36.2	9
32.	NKLY_Ny	1.0	1.6	11.1	84	4.4	181.2	14.5	23.3	11
	Mean	1.3	2.2	10.5	81	4.4	168.6	18.2	29.8	10
	SE±	0.3	0.2	1.4	3	0.9	17	2.4	3.7	1
	CV (%)	34.5	13.8	22.3	6	36.1	17.5	22.8	21.6	13
	LSD (5%)	0.7	0.5	3.8	8	2.6	48.1	6.8	10.5	2

4.1.2 Days to 50 % flowering

The difference among accessions in flowering time was highly significant and ranged from 72 (*IS23/31*) to 108 (*NGJ*) days with mean flowering of 81 days. Physiologically, under normal condition, plants which are early maturing would reach flowering stage very early. But again early flowering has been sited as one of the survival mechanisms in plant especially under stress condition (Nilsen and Orcutt, 1996). Generally according to Reddy (2005) and IBPGR/ICRISAT (1993), sorghum varieties which have days to 50% flowering are less than 80 days, fall under early or medium maturing varieties. Sixty three percent of 32 accessions tested (Table 5) reached days to 50% flowering almost in 80 days or less. This implies that these accessions are suitable regarding the maturity. For the case of ethanol production, the shorter time the crop will take to mature is an advantage to the crop because in a year, one crop can be produced as compared to sugarcane which takes more than one year to produce the first crop. Therefore this implies that sweet sorghum will be effectively and efficiently in crop production and land utilization.

4.1.3 Grain yield

In this study, grain yield did not differ significantly among the accessions. Higher grain yield (t ha^{-1}) was recorded in accession *NDH_Chi* (7.1 ton), *Unk_Ise* (6.5 ton) and *NGW_Sa* (6.2 ton) (Table 5). Accessions *NGJ* (2.7 ton) recorded the lowest grain yield. It has been reported that stress especially moisture content normally affect plant growth by accelerating senescence of plant leaves (Bilaro, 2008) minimising the proportion of green leaves as a result; hence the amount of assimilates available for the sink is reduced. In this study, it is possible that the prolonged drought that started at initiation to booting stage at LZARDI Ukiriguru as indicated in Appendix 1 could also have affected the grain yield. On the other hand sorghum is a plant known as tolerant to water stress, but the extent of individual effects varies depending on the intensity of the stress as well as genetic

difference among the accessions. It is expected that grain yield will not be similar if the trial is tested for more than once and in different environment to observe their significance.

4.1.4 Plant height

The analysis of variance show that accessions did not differ significantly on plant height. The highest plant height was recorded in accession *LIM*; 204 cm while *KSV_700* recorded 125 cm as the lowest plant height with the mean average of 168.6 cm. However, there was a significant variation ($P \leq 0.001$) among accessions in the number of internodes. At maturity, the number of internodes ranged from 8 to 11 with the mean average of 9.6. Accession *NKLY_Ny* and *LIM* had the highest number of 11 internodes. The number and length of internodes relates to plant height. Sorghum plant height is affected by environmental conditions. However, the findings obtained in this study was inconsistence with those reported by Reddy *et al.* (2007) and Ali *et al.* (2008), that sweet sorghum varieties had a height ranging from 130-240 cm and 104 to 374 cm respectively. The plant height and the number of internodes apart from environmental effects can also be influenced by the genotype by itself, that is, some accessions being genetically taller with more number of internodes than others. A taller plant has an advantage due to fact that height may be correlated with brix and cane yield as noted by Reddy *et al.* (2007). A variation of the number of internodes among the accession could be utilized during the selection for improvement

4.1.5 Brix percentage

There was a significant variation ($P \leq 0.001$) among accessions in brix percentage (Table 6). According to Weighted Neighbor Joining by simple matching, 32 accessions were mainly grouped into three levels of brix values; Low brix (%) <10, medium brix (%) 10-13 and high brix (%) >13. Forty one percent of 32 accessions recorded low brix value, 31 percent

recorded medium brix value and 28 percent recorded high brix value (Table 5; Fig. 2). The high brix values were recorded from accessions *IS23/31*(15.0), *TEG_Bu* (14.5), *Unk_Ise* (14.3) (Table 5). The *NGJ* accession seems to be potential for ethanol production although in terms of grain yield it was not good unlike other accessions. These results are different from those in the study by Ali *et al.* (2008) whereby the highest brix value ranged from 18.0 to 19.3 and Reddy *et al.* (2007) recorded the highest brix value of 16.0 to 20.0.

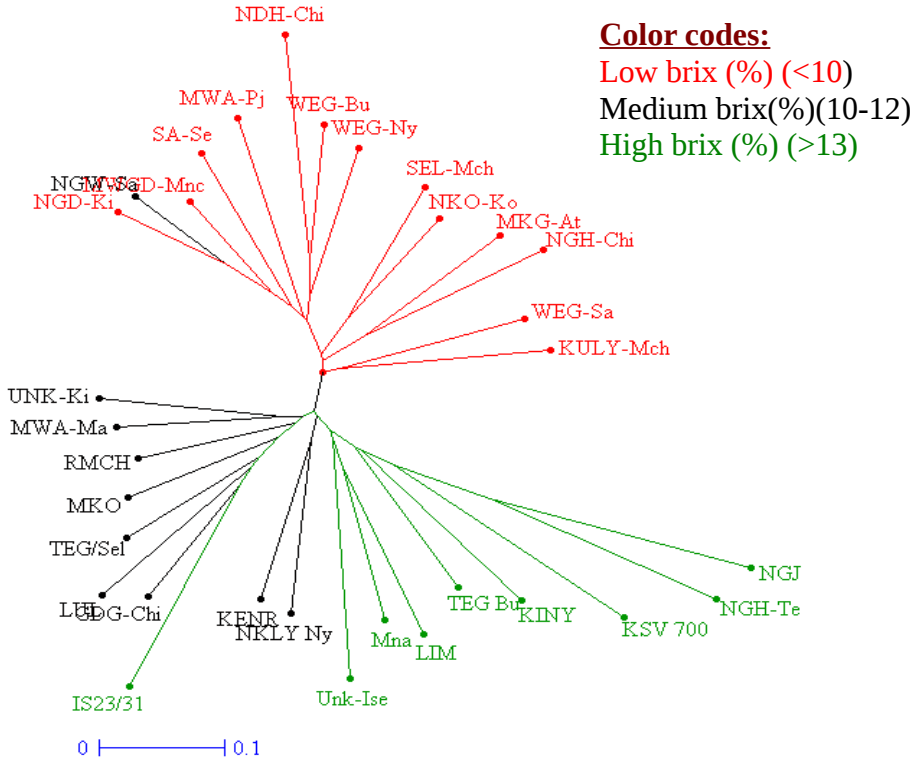

Figure 2: Radial dendogram of 32 sweet sorghum brix value dissimilarity index by simple matching based on NJ clustering-dendogram (Darwin V.5.25)

The data in this study were collected at one location and for one season and possibly the variability for brix levels could have been studied well under different environments (GxE) which in fact was not part of this study. However, the variability in the 32 accessions is an indication that possibly there is a high potential for genetic improvement to produce sorghum for high sweet-stalk yield coupled with high sucrose content. It has been reported from other research that 112 t of sorghum stalk (25 t ha^{-1}) has 23.47% juice with 8.5% total reducing sugar, brix of 12 and pH 5 produce alcohol yield of about 16.38 litres t^{-1} of stalk (Ratnavathi *et al*., 2003).

It is already noted from the study that sweet sorghum feedstalk with a brix level above 12 are suitable for ethanol production (Ratnavathi *et al*., 2003, Reddy *et al*, 2003, Reddy *et al.*, 2005 and Rao *et al.*, 2008).

In this study, twelve accessions namely; *MKO, RMCH, KENR, LIM, NGH_Te, KINY, KSV_700, Mna, NGJ, Unk_Ise, TEG_Bu* and *IS23/31* had brix levels above 12 and grain yield between 2.3 and 6.5 t ha^{-1} (Table 5). This indicates that these accessions can be used to produce ethanol. Other findings (Reddy *et al.,* 2008) shown sweet sorghum varieties and hybrids as having the ability to produce extremely high stalk yields of up to 50 t ha^{-1}, with juice brix reading ranging from 18 to 22% and 1.5 to 2.5 t ha^{-1} grain.

The dissimilarity distance for brix levels indictes that accession *NGD_Chi* and *NGW_Sa* measured the lowest distance matrix (0.19). This means that they were comparable in terms of brix contents. The highest dissimilarity distance (0.70) was noted among; *KSV_700 and NGJ; and NDH_Chi and NGJ* and therefore they were diverse in terms of brix values (Table 6.). A pair-wise distinction of accessions in this study indicates, only six pairs which had ≥ 0.65 dissimilarity distances. The six pairs are *NGH_Te/NDH_Chi* (0.69), *NGH_Te/KSV_700* (0.68), *NGJ/NDH_Chi*, (0.70), *NGJ/IS23/31* (0.65), *NGJ/KSV_700* (0.70), *KSV_700/NDH_Chi* (0.66). Out of six pairs, two pairs (*NGH_Te/NDH_Chi)* and (*NGJ/NDH_Chi*) were from farmers' fields (Table 6). Accessions which had high brix levels and diverse from their counterparts of low brix levels, could possibly be possessing genes for increasing the amount of brix in their stem juices. These pairs could be used to select accessions with high brix levels. Further evaluation of these accessions could assist in selecting parent materials for tagging such potential gene.

Table 6: Distinct relationships of of brix levels by NJ simple matching dissimilarity matrix (Darwin V.5.25)

S.no	Names of accessions pair		Distance matrix
1.	*NGJ*	*SEL_Mch*	0.60
2.	*NGJ*	*MKG_At*	0.60
3.	*NGH_Te*	*NGH_Chi*	0.61
4.	*NGH_Te*	*WEG-Bu*	0.61
5.	*NGJ*	*LUL*	0.61
6.	*IS23/31*	*NDH_Chi*	0.61
7.	*NGH_Te*	*NGD_Ki*	0.62
8.	*NGJ*	*WEG_Bu*	0.62
9.	*NGH_Te*	*MWA_Pj*	0.63
10.	*NGH_Te*	*IS23/31*	0.63
11.	*NGJ*	*NGH_Chi*	0.63
12.	*NGJ*	*SA-Se*	0.63
13.	*NGJ*	*MWA_Pj*	0.64
14.	*NGJ*	*NGD_Ki*	0.64
15.	*NGJ*	*IS23/31*	0.65
16.	*KSV_700*	*NDH_Chi*	0.66
17.	*NGH_Te*	*KSV_700*	0.68
18.	*NGH_Te*	*NDH_Chi*	0.69
19.	*NGJ*	*NDH_Chi*	0.70
20.	*NGJ*	*KSV_700*	0.70

4.1.6 Relationships among yield components

Correlation analysis (Table 7) indicates that days to 50% flowering and days to maturity show a high and positive significant correlation with wet stalk weight (r = 0.569*** and 0.741***) respectively. This implies that the longer the time accession takes to flower the more the accumulation of stalk weight and therefore the longer the time for harvesting (like *NGJ*). However, brix value was significantly negatively correlated with grain yield (r= -0.342**) while positively and highly significantly correlated with days to 50% flowering (r= 0.362***). This could possibly be that more assimilates accumulated by plant are converted and stored in the form of sugar in stalks instead of grain yield. This is why most of sweet sorghum varieties have low grain weight per spike as well as grain yield. Accessions which measured a high brix value (>13) as well as grain yield of above 5.0 t^{-1}ha were *Unk_Ise*, *KINY and LIM* (Table 5). This implies that these accessions can be used as sweet sorghum varieties by farmers and/or in hybrid development of sweet sorghum which can optimally give good grain yield as well as high brix value.

The plant height was positively and highly significantly correlated with the number of internodes (r=0.588***). This implies that plant height relates to number of nodes and the length of internodes. As the number of nodes increases plant height tends to increase as well. This implies that the more the number of nodes are, the long the growing time. Taller plants usually reach flowering stage at a later stage than the shorter plants. However, as Reddy *et al*. (2007) note the length of internodes is sensitive to temperature and water supply. Therefore, environmental interaction generally has an influence on plant height as well as grain yield. On the other hand, plant height did not show any significance correlation with stalk weight and brix levels (Table 7). This could be attributed to environmental effect. Repeating the trial could possibly give an understanding on the influence of environmental effect.

Table 7: Correlation coefficiencies of sweet sorghum accessions among the traits

Trait	Wet stalk weight (kg)	Brix (%)	Grain yield t/ha	50% flowering (days)	Days to maturity (days)	Plant height (cm)	No of Internodes
Wet stalk weight (kg)	1.000						
Brix (%)	0.148	1.000					
Grain yield t/ha	0.106	-0.342**	1.000				
50% flowering (days)	0.569***	0.362***	-0.151	1.000			
Harvesting (days)	0.741***	0.252*	-0.184	0.751***	1.000		
Plant height (cm)	-0.023	-0.013	0.150	-0.006	-0.028	1.000	
No of Internodes	0.004	-0.178	0.316*	-0.105	-0.207*	0.588***	1.000

Note: *= Probability significant at 0.05 level; ** = significant at 0.01 and ***= significant at 0.001 level

Sorghum plant have the ability to cease growing for some time if there is water stress and resumes later when moisture is adequate. Lack of positive significant correlation between plant height and the time to 50% flowering (r = -0.006) (Table 7) contradicts the findings reported by Reddy *et al*. (2007) and this could probably be attributed to bad weather condition. The drought was noted at panicle initiation to booting stage as indicated (Fig. 3; Appendix 1.). During this period the rainfall record in February (37cm) and March (79cm) that might have reduced the potential seed numbers. According to Vanderlip, (1993), it is considered that the panicle initiation to booting stage is the most critical period for grain production since the seed number per plant accounts for 70 percent of the grain yield.

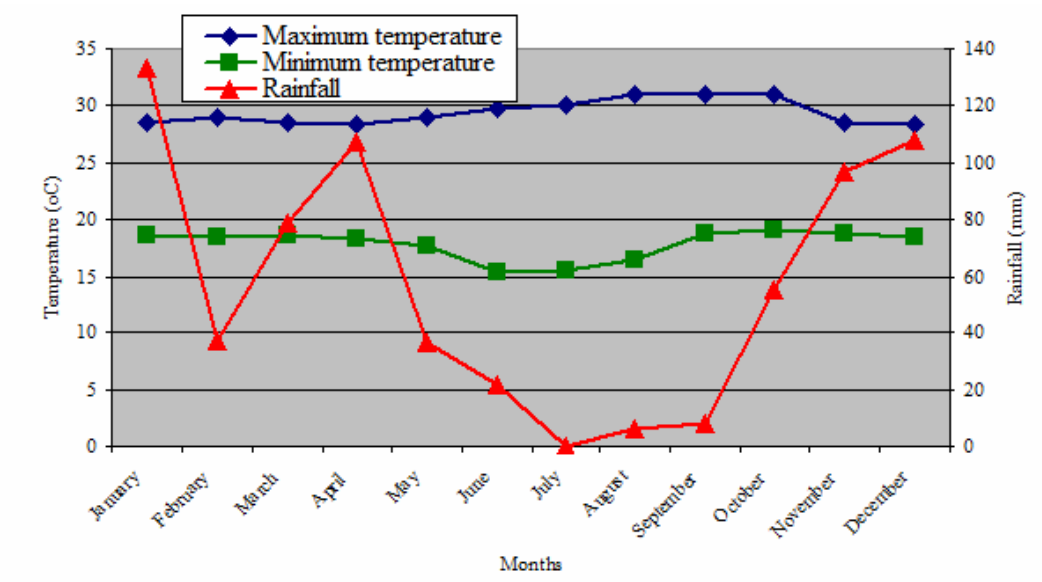

Figure 3: Ranfall and Temperature year 2009 at LZARDI Ukiriguru

4.1.7 Genetic diversity of sweet sorghum accession

Forty eight sorghum SSRs, which were distributed across the genome, *Xtxp*339 failed to amplify. The PCR amplification was successful for other 47 SSRs across the genotypes /accessions screened and the best markers with good yield amplification with low proportional of heterozygozity were retained for further analysis. However, 5 SSR loci (*mSbCIR*283, *Xcup*02, *Sb5*-206, *Xtxp*114 and *Xtxp*141) had very high proportion of heterozygocity, and hence they were discarded from further analysis. The remaining, 42 SSRs (Table 8) that had a good performance were taken for further analyses. The observed allele sizes ranged from 103 bp (*msbCIR*238) to 311 bp (*gpsb*123). Of the 42 SSR markers, one of the genic SSR loci (*Xtxp*040) was monomorphic in this diverse set of sorghum accessions. Heterozygosity values of polymorphic SSR markers ranged from 0.00 (*Xtxp*278, *Xcup*53) to 0.33 (*gpsb*067), with a mean of 0.13. These 42 SSRs revealed a total of 220 alleles with a range of 2 to 15 alleles and with an average of 5.07 alleles per primer pair lower than 7.6 from another study of sweet sorghum (Wang *et al*., 2009). This is possibly due to polymorphism of SSR markers and diversity of germplasm accessions used in this study (Table 8) and which confirms their usefulness for genetic analysis.

Table 8: Details of 42 SSRs markers used across 43 sweet sorghum accessions

Marker	Major Allele Frequency	Genotype No	Sample Size	No. of obs.	Allele No	Gene Diversity	Heterozy gosity	PIC
Xtxp040	1.00	1	43	43	1	0.00	0.00	0.00
mSbCIR223	0.97	3	43	43	2	0.07	0.02	0.07
mSbCIR276	0.97	3	43	43	2	0.07	0.02	0.07
Xcup53	0.95	2	43	43	2	0.09	0.00	0.08
Xisep0310	0.95	3	43	43	3	0.09	0.00	0.09
gpsb148	0.91	5	43	43	5	0.17	0.09	0.17
Xisep0107	0.87	4	43	43	3	0.23	0.02	0.21
Xtxp278	0.84	3	43	43	3	0.28	0.00	0.26
Xtxp321	0.83	5	43	43	5	0.31	0.16	0.30
Xtxp273	0.80	8	43	43	7	0.35	0.07	0.34
Xcup11	0.78	4	43	43	3	0.37	0.02	0.34
Xcup62	0.65	4	43	43	3	0.46	0.23	0.37
Xtxp136	0.56	3	43	43	2	0.49	0.09	0.37
Xtxp012	0.53	3	43	43	2	0.50	0.19	0.37
mSbCIR246	0.74	5	43	43	4	0.42	0.05	0.38
mSbCIR262	0.52	4	43	43	4	0.52	0.14	0.41
Xcup63	0.64	4	43	43	3	0.51	0.12	0.44
Xcup14	0.63	5	43	43	4	0.51	0.05	0.44
mSbCIR240	0.65	6	43	43	5	0.52	0.09	0.48
mSbCIR300	0.63	6	43	43	3	0.53	0.16	0.48
gpsb067	0.57	6	43	43	4	0.57	0.33	0.51
Sb4_72	0.55	9	43	43	4	0.59	0.26	0.53
mSbCIR248	0.58	8	43	43	5	0.61	0.21	0.57
gpsb123	0.58	10	43	43	6	0.62	0.19	0.59
mSbCIR238	0.58	7	43	43	7	0.62	0.00	0.59
mSbCIR306	0.44	4	43	43	4	0.66	0.00	0.60
Xtxp021	0.49	9	43	43	4	0.66	0.26	0.61
Xcup61	0.40	6	43	43	4	0.67	0.14	0.61
gpsb089	0.45	8	43	43	6	0.68	0.14	0.62
SbAG_B02	0.45	11	43	43	7	0.68	0.16	0.63
Xtxp145	0.47	9	43	43	5	0.68	0.12	0.64
mSbCIR286	0.44	7	43	43	5	0.69	0.09	0.65
mSbCIR329	0.38	8	43	43	5	0.71	0.23	0.66
gpsb151	0.35	10	43	43	7	0.73	0.16	0.68
Xtxp015	0.43	12	43	43	6	0.73	0.23	0.69
Sb6_84	0.41	12	43	43	7	0.74	0.19	0.71
Xtxp010	0.38	12	43	43	6	0.76	0.33	0.72
Xtxp320	0.36	15	43	43	8	0.77	0.26	0.74
Xtxp057	0.30	14	43	43	9	0.83	0.19	0.81
Xtxp295	0.27	15	43	43	10	0.83	0.19	0.81
gpsb069	0.19	21	43	43	13	0.88	0.26	0.87
Xtxp265	0.23	23	43	43	15	0.88	0.21	0.87
Mean	**0.59**	**7.55**	**43**	**43**	**5.07**	**0.53**	**0.13**	**0.48**

The polymorphism information content (PIC) values for the SSR loci ranged from 0.07 (*msbCIR*223, *msbCIR*276) to 0.87 (*Xtxp*265, *gpsb*069) with an average of 0.48 similar to 0.46 reported by Schloss *et al*. (2002) but lower than diversity value of 0.62, 0.58 reported by Agrama and Tuinstra (2003) and Smith (2000) respectively. SSR loci which produced a higher number of alleles (7-15) also had a higher number of gene diversity ranging from 0.0 to 0.88 per locus similar to the results of Ali *et al*. (2008). Sixteen SSR loci (*Xtxp*021,

*Xcup*61, *gpsb*089, *SbAG_B*02, *Xtxp*145, *SbCIR*286, *SbCIR*329, *gpsb*151, *Xtxp*015, *Sb6*_84, *Xtxp*010, *Xtxp*320, *Xtxp*057, *Xtxp*295, *gpsb*069, *Xtxp*265) manifested PIC value of more than 0.6 indicating their potential in-formativeness to detect differences among the accessions based on their genetic relationships. Dinucleotide repeat containing SSRs were generally, more variable than those with longer motifs as noted in other studies by Enoki, *et al.* (2002). However, other investigators often prefer to work with loci containing tri- and tetra-nucleotide repeat arrays rather than dinucleotide arrays. Thus, allele sizing is less error prone using tri- and tetra-nucleotide repeats than di-nucleotide repeats (Diwan and Cregan, 1997). In this study, 29 out of 43 SSRs were di-nucleotide and this could be the reason why there was higher number of allelic diversity in the materials used. In general, SSRs with dinucleotide repeats displayed a higher number of alleles than tri- and tetra nucleotide repeats (Tables 3 and 8.). Another possible reason for a relatively high gene diversity in this study could be the different sources of materials used that is those from ICRISAT being improved and those obtained from farmers' fields. In the case of 48 accessions that were used in the analysis, five of them; (*NGH_Te, ICSL73220, ICSL73221, ICSV111, BTx623*) were excluded in the analysis due to the fact that they had more than 20% missing markers (Table 8). This could be the absence of amplification due to the mutations in franking regions (Stachel *et al.,* 2000). Another possible reason for obtaining such results could be the contamination or shearing of DNA which finally end up with amplification failure. In another study, Semagn *et al.* (2006) reported to have a generally low recovery rate for useful SSR primer due to different reasons: (a) the primer may not amplify any PCR product; (b) the primer may produce very complex, weak or non-specific amplification patterns and (c) the amplification product may not be polymorphic. Regarding this study, it might be possible that there were low levels of repetitive DNA sequences in their genome of these accessions used and that is why they failed to show alleles at the loci. This problem could be resolved by DArT technique which allows

simultaneous scoring of hundreds of restriction site based polymorphisms between accessions and which do not require DNA sequence information or site specific oligonucleotides.

4.1.8 Genetic relationship among accessions

The genetic relationship between accessions from farmers' fields and ICRISAT materials was assessed via cluster analysis using Weighted Neighbor-Joining clustering-based (Saitou and Nei 1987) and principal component of analysis (PCoA) of similarity matrix of raw SSR and data based similarity index (Nei, 1978). This is a distance-based method, in which a pair-wise dissimilarity index by simple matching was used as an input for analysis. Therefore, Neighbor-Joining tree analysis clustered the accessions into branches by fixing each accession in a branch. The resulting dendogram (Fig. 4) indicates that almost all accessions could be distinguished and clustered into three main *Clusters, A, B,* and *C.*

Cluster A was the largest group which was subdivided into *Cluster A1&A2.* It included 30 accessions dominated by ICRISAT materials. *Cluster B* included 10 accessions and *Cluster C* had 4 accessions. All of the accessions in *Cluster B* and *C* were accessions from farmers' fields. Fourteen accessions were in *Cluster A1* namely; *SPV 422, IESV 92001 DL, SDSL 90167, S_Ind_35, IESV 92008 DL, IESV 92028 DL, IESV 93042 SH, S_Nair_35, LUL, Mna, IS_2331, IS23/31, SSV_84,* and *LIM.* Out of these 14, three accessions (*LUL, Mna, LIM)* were from farmer's fields and the rest were from ICRISAT. This suggests that these accessions could be closely related and might have the same genetic background with ICRISAT materials. Another possibility is that these accessions might have been introduced to farmers from research centres and given different names to suit the interest of the farmers. For example, the presence of *LUL,* which is possibly related to 'LULU' that was developed by ICRISAT through pedigree breeding (Obilana, 2004), could be the

reason why these were clustered together. Sixteen accessions were clustered together in *ClusterA2* whereby three of them (*ICSV_93046, KSV_700, IESV_91018LT*) being ICRISAT materials and 13; *SEL_Mch, SA_Se, UNK_Ki, Unk_Ise, NKLY_Ny, KULY_Mch, MKG_At, ICSV_93046, KSV_700, TEG_Bu,* and *TEG_Sel* being from farmers' fields. When closely examined cluster *A2* in Figure 4 branches further into smaller clusters in which some accessions which were collected from farmers formed a coherent group while the ICRISAT material forming another small branch indicating the distinction within the *cluster A.* This increases the chances that the accession from farmers' fields shares some genes with ICRISAT materials. It can also be due to the recurrent recombination of the ICRISAT materials and farmers' materials and this confirms the assertions that the farmer materials are among the parents used in the development of improved varieties.

The clustering of *B and C* contained only accessions from farmers' fields. Therefore, dendogram has clearly distinguished ICRISAT materials from accession from farmers' fields and this could be the evidence that they have different origins. These results are in agreement with comparable results on barley genotypes published by Ordon *et al.* (1997).

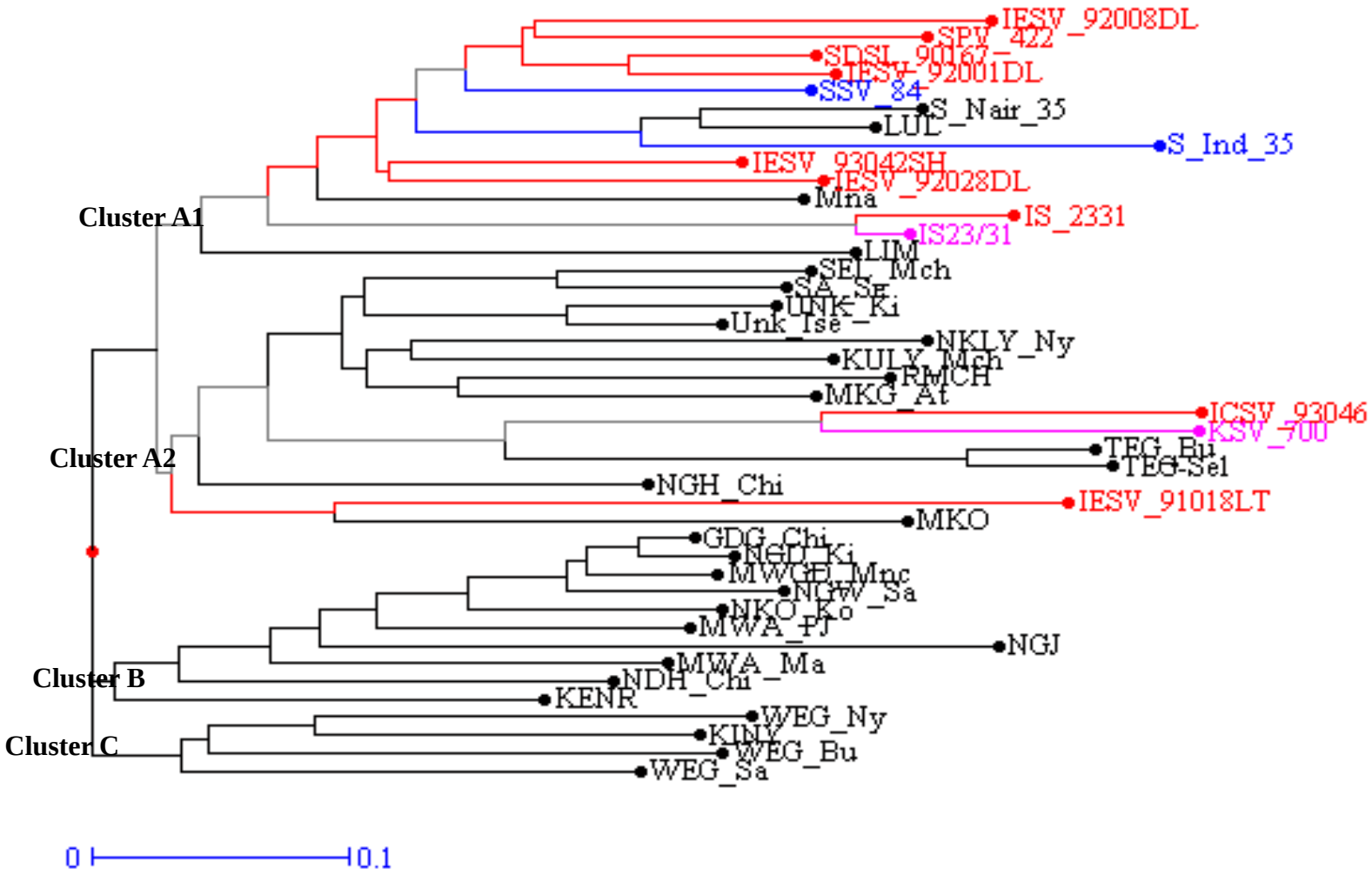

Figure 4: Tree diagram based on 42 SSRs and 43 accessions by simple matching dissimilarity indices NJ clustering. (Darwin V.5.25)

Among the 43 accessions studied 31 (72.1%) were genetically distinct with a dissimilarity matrixes of ≥ 0.70 (Table 9). The highest genetic distance was observed in (*S_Ind_35 and KULY_Mch)* (0.79) followed by (*NGJ* and *ICSV_93046)*, *(ICSV_93046* and *WEG_Bu)* and *(S_Ind_35* and *IS23/31)* with 0.77 each. Accessions *S_Ind_35*, *IS23/31* and *ICSV_93046* originated from ICRISAT, while *KULY_Mch*, *WEG_Bu* and *GNJ* were collected from farmers' fields. This indicates a sufficiently large degree of genetic diversity in the collected germplasm and improved cultivars (Table 9). On the other hand, 12 (30.2%) had a genetic diversity ≤ 0.20 and the lowest genetic diversity (0.06) being recorded between (*GDG_Chi* and *NGD_Ki*) (Appendix 2). The implication is that the smaller the value of the distance matrixes is, the less distinct the accessions are. Therefore, according to the results, these accessions were closely related. The other reason for close relatedness is that all of the accessions were collected from farmers' fields and farmers have been using and exchanging these accessions among themselves for quite some time. This tendency could have contributed much to making these accessions less distinction among them. The situation has been clearly shown by a dendogram (Fig. 4) in which *Cluster B* accessions came from farmers' field.

Table 9: The dissimilarity of sweet sorghum accessions based on Neighbor-Joining

S.no.	Names of accession pair		Dissimilarity matrix	S.no.	Names of accession pair		Dissimilarity matrix
1.	*NGJ*	*NKLY_Ny*	0.70	17.	*TEG_Bu*	*NGW_Sa*	0.71
2.	*NGJ*	*IESV_93042SH*	0.70	18.	*WEG_Bu*	*KSV_700*	0.71
3.	*S_Ind_35*	*Unk_Ise*	0.70	19.	*ICSV_93046*	*KINY*	0.73
4.	*S_Ind_35*	*LIM*	0.70	20.	*NGJ*	*SPV_422*	0.73
5.	*TEG_Sel*	*MWGD_Mnc*	0.70	21.	*S_Ind_35*	*KSV_700*	0.73
6.	*TEG_Sel*	*NGW_Sa*	0.70	22.	*S_Ind_35*	*MWA_Pj*	0.73
7.	*WEG_Bu*	*TEG_Sel*	0.70	23.	*WEG_Bu*	*TEG_Bu*	0.73
8.	*ICSV_93046*	*NGW_Sa*	0.71	24.	*NGJ*	*IESV_91018L T*	0.74
9.	*NGJ*	*MKG_At*	0.71	25.	*TEG_Sel*	*NGJ*	0.74
10.	*NGJ*	*IESV_92008DL*	0.71	26.	*S_Ind_35*	*RMCH*	0.75
11.	*S_Ind_35*	*GDG_Chi*	0.71	27.	*TEG_Bu*	*NGJ*	0.75
12.	*S_Ind_35*	*MKG_At*	0.71	28.	*NGJ*	*ICSV_93046*	0.77
13.	*S_Ind_35*	*GDG_Chi*	0.71	29.	*S_Ind_35*	*IS23/31*	0.77
14.	*S_Ind_35*	*NKO_Ko*	0.71	30.	*WEG_Bu*	*ICSV_93046*	0.77
15.	*TEG_Bu*	*MWGD_Mnc*	0.71	31.	*S_Ind_35*	*KULY_Mch*	0.79
16.	*TEG_Bu*	*LIM*	0.71				

The PCoA was performed on the basis of Weighted Neighbor-Joining Dissimilarity index by simple matching of 43 accessions. A scatter plot of the first and second axes of non-metric multi-dimensional scaling (MDS) (Fig. 5) reveals three inter-relationship groupings among accessions. These groupings were *Clusters I, II* and *III*. *Cluster 1* had eleven accessions, *Cluster II* seventeen accessions and *Cluster III* three accessions.

Although PCoA is based on different algorithms using Eigen vector, the grouping of accessions was similar to that of dendogram (Fig. 4), very few accessions clustered differently. For instance, accession *MKO* and *LIM* were not clustered with ICRISAT materials as they did in the dendogram. The inter and intra-clusters variations of accessions provide useful information to sorghum breeders and especially sweet sorghum in terms of the selection of related or unrelated parental materials. Therefore, the results can enable the breeders to maximize their variability in the breeding programs which probably may produce good segregating progenies.

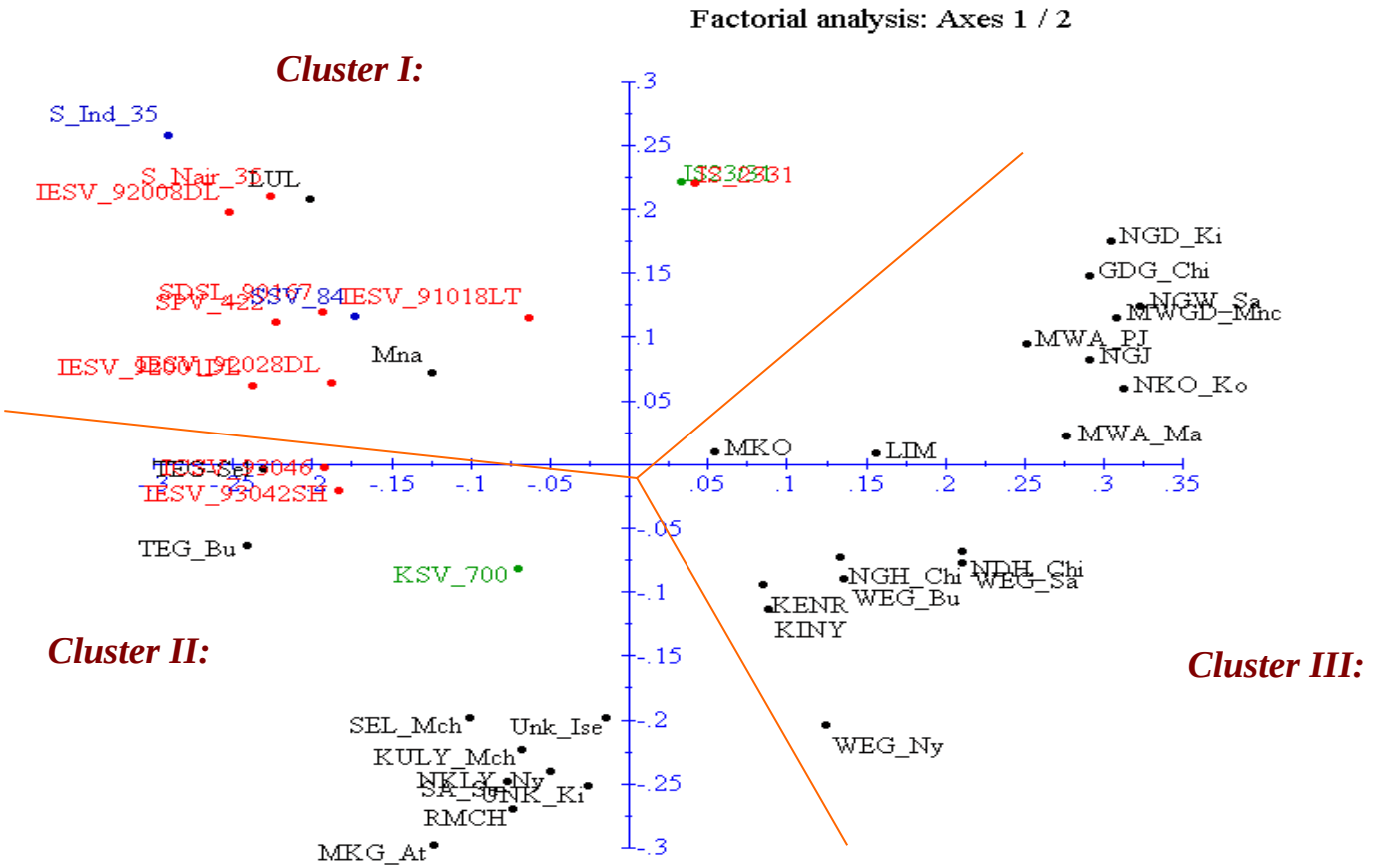

Figure 5: PCoA scatter plot by simple matching dissimilarity indices for 42 sorghum SSR loci and 43 sorghum accessions (Darwin V.5.25)

CHAPTER FIVE

5.0 CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

The objective of this study was to assess the genetic diversity and relationships among sweet sorghum accessions using SSR markers, to examine brix level variability within the colleted accessions and to establish their genetic relationships.

The study has revealed valuable information on the relationship among sweet sorghum accessions from farmers' fields and ICRISAT materials, in that there is a distinct relationship between them. The accessions collected from near-by locations were similar genetically, but still had variability within. The accessions from farmers fields were diverse than the materials from ICRISAT.

(i) The significant variation ($P \leq 0.001$) was observed among accessions in brix percentage. Twelve accessions namely; *MKO, RMCH, KENR, LIM, NGH_Te, KINY, KSV_700, Mna, NGJ, Unk_Ise, TEG_Bu* and *IS23/31* had brix levels above 12 and grain yield between 2.3 and 6.5 t ha^{-1} This indicates that these accessions can be used to produce ethanol. However, *NGJ* accession seems to be potential for ethanol production although in terms of grain yield it was not good unlike other accessions

(ii) In this study, grain yield did not differ significantly among the accessions. Higher grain yield (t ha^{-1}) was recorded in accession *NDH_Chi* (7.1 ton), *Unk_Ise* (6.5 ton) and *NGW_Sa* (6.2 ton). The high brix values were recorded from accessions *IS23/31*(15.0), *TEG_Bu* (14.5), *Unk_Ise* (14.3).

(iii) The highest dissimilarity distance (0.70) in terms of brix values was noted among; *KSV_700 and NGJ; and NDH_Chi and NGJ* and therefore they were diverse. These pairs could be used to select accessions with high brix levels.

(iv) The dendogram distinguishes three main *Clusters*, *A, B,* and *C* of accessions *whereby* all of the accessions in *Cluster B* and *C* were dominated by accessions from farmers' fields and Clusters, *A,* dominated by ICRISAT materials .The highest genetic distance was observed in (*S_Ind_35 and KULY_Mch)* (0.79) followed by (*NGJ* and *ICSV_93046), (ICSV_93046* and *WEG_Bu)* and *(S_Ind_35* and *IS23/31)* with 0.77 each. This indicates a sufficiently large degree of genetic diversity and therefore evidence that they have different origins.

5.2 Recommendations

(i) The phenotypic information show the variability in-terms of grain yield as well as brix content. It is recommended to do further evaluation on accessions which had:

- High brix levels (≥13), grain yield (≥ 5.0 t^{-1}ha) and diverse from their counterparts of low brix levels.
- Dissimilarity distance ≥ 0.70 for brix levels that is (*KSV_700 and NGJ; and NDH_Chi and NGJ*).

These accessions could possibly be possessing genes for increasing the amount of brix in their stem juices and grain yield. Their high brix value could assist in selecting parent materials for tagging such potential gene.

(ii) The diversity has been observed in the studied accession especially among pairs; (*S_Ind_35 and KULY_Mch)* (*NGJ* and *ICSV_93046), (ICSV_93046* and *WEG_Bu)* and *(S_Ind_35* and *IS23/31)*, It is recommended to be confirmed by DArT technique to ascertain results obtained with SSR markers in this study. DArT technique will be able to distinguish whether the gene constitution of accessions from farmers' fields and accession from ICRISAT is different for sweet sorghum traits.

(iii) To broaden the genetic base of sweet sorghum in Tanzania, it is recommended to carry out the germplasm collection of sweet sorghum from sorghum growing areas in the country. The existing level of genetic diversity in farmers' materials is likely to represent significance advancement in genetic improvement for sweet sorghum program in the country in future

REFERENCES

Agrama, H. A. and Tuinstra, M. R. (2003). Phylogenetic diversity and relationships among sorghum accessions using SSRs and RAPDs. *African Journal of Biotechnology* 2(10): 334-340.

Ali, M. L. Rajewski, J. F., Baenzieger, P. S., Gill, K. S. and Dweikat, E. I. (2008). Assessment of genetic diversity and relationship among a collection of US sweet sorghum germplasm by SSR markers. *Molecular Breeding* 21: 497-509.

Bhattramakki, D., Dong, J., Chhabra, A. K. and Hart, G. (2000). An integrated SSR and RFLP linkage map of *Sorghum bicolor* (L.) Moench. *Genome* 43: 988-1002.

Bilaro, A. L. (2008). Evaluation of quality protein maize synthetic germplasm for drought tolerance. A dissertation for award of MSc Degree at Sokoine University of Agriculture, Morogoro, Tanzania. 104 pp.

Brown, S. M., Hopkins, M. S. and Mitchell, S. E. (1996). Multiple methods for the identification of polymorphic simple sequence repeats (SSRs) in sorghum [*Sorghum bicolor* (L.) Moench]. *Theoretical and Applied Genetics* 93: 190-198.

Brown-Guedira, G. L., Thompson, J. A., Nelson, R. L. and Warburton, M. L. (2000). Evaluation of genetic diversity of soybean introductions and North American ancestors using RAPD and SSR markers. *Crop Science* 40: 815-823.

Channappagoudar, B. B., Biradar, N. R., Patil, J. B. and Hiremath, S. M. (2007). Assessment of sweet sorghum genotypes for cane yield, juice characters and sugar levels. *Journal of Agricultural Science* 20(2):294-296.

Daniel, J. P., Nathalie, I. and Jean, B. (1999). Sequence-tagged-site (STS) markers of arbitrary genes: the amount and nature of variation revealed in Norway spruce. *Heredity* 83: 239-248.

Dayakar, R. B., Ratnavathi, C. V., Karthikeyan, K., Biswas, P. K., Rao, S. S., Vijay Kumar, B. S. and Seetharama, N. (2004). *Sweet sorghum cane for bio-fuel production: A SWOT analysis in Indian context.* National Research Centre for Sorghum, Rajendranagar, Hyderabad, AP 500 030. India. 20 pp.

Diwan, N. and Cregan, P. B. (1997). Automated sizing of fluorescent-labelled simple sequence repeat (SSR) markers to assay genetic variation in soybean. *Theoretical and Applied Genetics* 95: 723-733.

Doggett, H. (1988). *Sorghum - second edition. Tropical agriculture series.* UK, IDRC, Canada: Longman Scientific and Technical Publishers. 512pp.

Edemeades, G. O., Banzinger, M., Pnadey, S., Bolanos, J., Chapman, S. C., Ortega, A., Lafitte, H. R. and Fischer, K. S. (1997 a). Recurrent selection unde managed drought stress improve grain yield in tropical maize. In: *Proceeding of symposium on developing drought and low-N tolerant maize.* (Edited by Edmeades, G. O. *et al.),* 25-29 March 1996, El Batan, Mexico. 415-425pp.

Enoki, H., Sato, H. and Koinuma, K. (2002). SSR analysis of genetic diversity among maize inbred lines adapted to cold regions of Japan. *Theoretical and Applied Genetics* 104: 1270-1278.

Gitay, H., Suárez, A., Watson, R. T. and Dokken, D. J. (2002). *Intergovernmental Panel on Climatic Change: Climate Change and Biodiversity* (IPCC Technical Paper V) IPCC. Geneva, Switzerland. 85pp [http://www.ipcc.ch/publications_and_data/ publications_and_data_technical_papers.htm]. Site visited on 30/9/2009.

Gnansounou, E., Duriant, A. and Wyman, C. E. (2004). Refining sweet sorghum to ethanol and sugar: economic trade-offs in the context of North China. *Bioresource Technology* 96(9): 985-1002.

Gomez, K. A. and Gomez, A. A. (1984). *Statistical Procedures for Agricultural Research* (2nd edition). John Willey and Sons, Inc. Singapore. 240 pp.

Gupta, P. K. and Varshney, R. K. (2000). The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica* 113: 163-185.

Hagan, E. B. (2007). Biofuels assessment report-ECOWAS Sub-Region. A paper presented at the AU/Brazil/UNIDO Biofuels Seminar in Africa, Addis Ababa, Ethiopia. [www.unido.org/fileadmin/ext_media/Services/Energy_and_Climate_ Change/ …/70710_Biofuels__ECOWAS__Dr._Ben_Hagan.ppt]. Site visited on 12/03/2010.

Hair, J. R., Anderson, R. E., Tatham, R. T., and Black, W. C. (1995). *Multivariate data analysis with readings. 4th edition.* Prentice-Hall, Englewood Cliffs, NJ. 287 pp.

Hoisington, D., Khairallah, M. and Gonzalez-de-Leon, D. (1994). Laboratory Protocols: CIMMYT Applied Biotechnology Center. Second Edition, Mexico, D.F. CIMMYT.

House, H. R., Gomez, M., Murty, O.S., Sun, Y., and Verma, B. N. (2000). *Development of some agricultural industries in several African and Asian countries. In: Sorghum Origin, history, technology, and production* (Edited by Smith, C.W. and Frederiksen, A. R.), John Wiley & Sons, New York. pp. 131-190.

IBPGR/ICRISAT (International Board for Plant Genetic Resources/International Crops Research Institute for the Semi-Arid Tropics) (1993*). Descriptors for sorghum* [*Sorghum bicolor* (L.) Moench]. IBPGR, Rome, Italy. 38pp.

Jumbe, C. B. L., Msiska, F. B. M. and Madjera, M. (2009). Biofuels development in Sub-Saharan Africa: Are the policies conducive? *Energy Policy*. doi:10.1016/j.enpol. 2009.06.064 [www. elsevier.com/locate/enpol]. Site visited on 24/11/2009.

Karekezi, S. (2007). Biofuels in Eastern and Southern Africa. A paper presented at the AU/Brazil/UNIDO Biofuels Seminar in Africa, Addis Ababa, Ethiopia. [www.complete-bioafrica.net/…Report_on_National_Policcies_and_Strategies. pdf]. Site visited on 26/11/2009.

Kong, L., Dong, J. and Hart, G. E. (2000). Characteristics, linkage-map positions, and allelic differentiation of *Sorghum bicolor* (L.) Moench DNA simple-sequence repeats (SSRs). *Theoretical and Applied Genetics* 101: 438-448.

Kresovich, S. and Henderlong, P. R. (1984). Agronomic potential of sorghum as a raw material for ethanol production in Central Ohio. *Energy in Agriculture* 3: 145-153

Lingle, S. E. (1987). Sucrose metabolism in the primary culm of sweet sorghum during development. *Crop Science* 27: 1214-1219.

Liu, K. and Muse, S.V. (2005). PowerMarker: Integrated analysis environment for genetic molecular data. *Bioinformatics* 21: 2128-2129.

Liu, K. J., Goodman, M. Muse, S. Smith, J. S. Buckler, E. and Doebley, J. (2003). Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. *Genetics* 165: 2117-2128.

Liu, S., Cantrell, R. G., McCarty, J. C., and Stewart, M. D. (2000). Simple sequence repeat-based assessment of genetic diversity in cotton race accessions. *Crop Science* 40: 1459-1469.

Mace, E. S., Buhariwalla, H. K. and Crouch, J. H. (2004). A high-throughput DNA extraction protocol for tropical molecular breeding programs. *Plant Molecular Biology Reporter* 21: 459-459.

Mafuru, J. M., Kamuntu, S. P. and Norman, D. W. (2005). Strengthening market-orientation of agricultural research in low-income countries using sorghum as an example. In: *Proceedings of Farming Systems and Poverty: Making a difference* (Edited by Dixon, J. *et al*.), 31st October - 3rd November 2005, Rome, Italy.

Mamoudou, H. D., Hurry, G., Alfred, S., Alphons, G. J. and Van, B. (2006). Sorghum grain as human food in Africa: relevance of content of starch and amylase activities. *African Journal of Biotechnology* 5(5): 384-395.

Matsuoka,Y., Mitchell, S.E., Kresovich, S., Goodman, M. and Doebley, J. (2002) Microsatellites in Zea-variability, patterns of mutations, and use for evolutionary studies. *Theoretical and Applied Genetics* 104: 436-450.

Mbwaga, A. M. and Obilana, A. T. (1993). Distribution and host specificity of striga asiatica and striga hermonthica on cereals in Tanzania. *Interantional Jornal of Pest Management* 39:449-451.

Mohammadi, S. A. and Prasanna, B. M. (2003). Analysis of genetic diversity in crop plants- Salient statistical tools and considerations. *Crop Science* 43: 1235-1248.

Murray, S. C., Hamblin, M. T., Mitchell, S. E. and Kresovich, S. (2009). Sweet Sorghum Genetic Diversity and Association Mapping for Brix and Height. *Plant Genome* 2: 48-62.

Nei, M. (1978). Estimation of average heterozygosity and genetic distances from a small number of individuals. *Genetics* 89: 583-590.

Ngugi, H. K., King, S. B., Abayo, G. O. and Reddy, Y. V. R. (2002). Prevalence, incidence and severity of sorghum diseases in western Kenya. *Plant Disease.* 86: 65-70.

Nilsen, E. T. and Orcut, D.M. (1996). *The physiology of plant under stress. Abiotic factors*. John Wiley and Sons Inc. New York. 688pp.

Obilana, A. B. (2004). Sorghum breeding research in Africa. In: *Sorghum genetic enhancement: research process, dissemination and impacts, Dissemination and impacts* (Edited by Bantilan, M.C.S. *et al.*). Patancheru 502 324, International Crop Research Institute for the Semi-Arid Tropics. Andhra Pradesh, India. pp 105-136.

Ordon, F., A. Schiemann, and Friedt, W. (1997). Assessment of the genetic relatedness of barley accessions (*Hordeum vulgare* s.l.) resistant to soil-borne Mosaic inducing viruses (Ba MMV, BaYMV-2) using RAPDs. *Theoretical and Applied Genetics* 94:325-330.

Organization for economic cooperation and development/food and agriculture organization. (OECD/FAO) (2007). OECD-FAO Agricultural Outlook 2007-2016. OECD/FAO, Paris. [http://www.unep.fr/scp/rpanel/pdf/Assessing_Biofuels_Full_ Report.pdf]. Site visited on 26/02/2010.

Paterson, A. H., Bowers, J. E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., Schmutz, J., Spannagl, M., Tang,

H., Wang, X., Wicker, T., Bharti, A. K., Chapman, J., Feltus, F. A., Gowik, U., Grigoriev, I. V., Lyons, E., Maher, C. A., Martis, M., Narechania, A., Otillar, R. P., Penning, B. W., Salamov, A. A., Wang, Y., Zhang, L., Carpita, N. C., Freeling, M., Gingle, A. R., Hash, C. T., Keller, B., Klein, P., Kresovich, S., McCann, M. C., Ming, R., Peterson, D. G., Rahman, M., Ware, D., Westhoff, P., Mayer, K. F. X., Messing, J. and Rokhsar, D. S. (2009). The *Sorghum bicolor* genome and the diversification of grasses. *Nature*. Vol 457|29 January 2009| doi: 10.1038/nature07723. [www.nature.com/reprints]. Site visited on 19/11/2009.

Perrier, X., Flori, A. and Bonnot, F. (2003). Data analysis methods. In: *Genetic diversity of cultivated tropical plants*. (Edited by Hamon, P., Seguin, M., Perrier, X., Glaszmann, J. C.), Enfield, Science Publishers, Montpellier. pp. 43-76.

Peter, W., Vanessa, C., Jason, C., Cypril, C., Margaret, E., Damian, J., Kaiman, P., Sujin, P., Grzegorz, U., Ling, X., Shiying, Y., Eric, H., and Andrzej, K. (2004). Diversity Array technology, a novel tool for harnessing crop genetic diversity. *Plant Molecular. Biology* 21: 459-459.

Prasado Rao, K. E., Mengesha, M. H. and Gopa Red, V. (1989). International use of sorghum germplasm. In *The Use of plant genetic resources*. (Edited by Brown A. H. D.), Cambridge University Press. 382pp.

Prashant, V. B. (2007). Economic of on-farm ethanol production using sweet sorghum. Dissertation for award of MSc. degree at Oklahoma State University Stillwater, Oklahoma, USA. 59pp.

Pritchard, J. K., Stephens, M. and Donnelly, P. (2000). Instance of population structure using multilocus, genotype data. *Genetics* 155: 945-959.

Rai, K. N., Murty, D. S., Andrews, D. J. and Bramel-Cox P. J. (1999). Genetic enhancement of pearl millet and sorghum for the semi-arid tropics of Asia and Africa. *Genome* 42 :617-628.

Ramu, P., Kassahun, B., Senthilvel, S., Kumar, C. A., Jayashree, B., Folkertsma, R. T., Reddy, L. A., Kuruvinashetti, M. S., Haussmann, B. I. G. and Hash, C. T. (2009). Exploiting rice-sorghum synteny for targeted development of EST-SSRs to enrich the sorghum genetic linkage map. *Theoretical and Applied Genetics* 119:1193-1204.

Rao, N. K., Bramel, P. J., Reddy, V. P. and Deb, U. K. (2004). Conservation, utilization and distribution of sorghum germplasm. In: *Sorghum genetic enhancement: Research process, dissemination and impacts*. (Edited by Baantilan, M. C. S. *et al*.), International Crops Research Institute for the Semi-Arid Tropics. Patancheru 502 324, Andhra Pradesh, India. pp. 43-62.

Rao, S. S., Seetharama, N., Dayakar, B., Ratnavathi, C. V. and Reddy, C. S. (2008). Sweet sorghum-a potential energy crop for biofuels production in India. In: *Sorghum improvement in the new millennium*. (Edited by Reddy, B. V. S. *et al.*), International Crops Research Institute for the Semi-Arid Tropics. Patancheru 502 324, Andhra Pradesh, India. pp. 281-288.

Ratnavathi C. V. (2008). Alternative use of sorghum grain and other products. In: *Sorghum improvement in the new millennium*. (Edited by Reddy, B. V. S. *et al.*), International

Crops Research Institute for the Semi-Arid Tropics. Patancheru 502 324, Andhra Pradesh, India. pp. 275-280.

Ratnavathi, C. V., Dayakar, R. B. and Seetharama, N. (2003). Sweet sorghum stalk: *A suitable raw material for fuel alcohol production. National research center for sorghum (NRCS)*, NRCS Report Number 12/2003, NATP (NRCS) series No. 1. 8 pp.

Ratnavathi, C.V., Dayakar, R. B., Padmaja, P. G., Ravi-Kumar, S., Reddy, C. S., Vijay-Kumar, B. S., Pallavi, M., Komala, V. V., Gopala, K. D., and Seetharama, N. (2005). *Sweet sorghum the world wonder crop for biofuel production*. NRCS Technical report Number 27. Rajendranagar, Hyderabad, Andhra Pradesh 500 030, India. 24pp.

Reddy, B. V. S., Kumar, A. A. and Reddy, P. S. (2008). Genetic improvement of sorghum in the semi-arid tropics. In: *Sorghum improvement in the new millennium.* (Edited by Reddy, B. V. S. *et al.*), International Crops Research Institute for the Semi-Arid Tropics. Patancheru 502 324, Andhra Pradesh, India. pp. 105-123.

Reddy, B. V. S., Ramiah, B., Ashok, K. A. and Sanjan, R. P. (2007). Selection of restorers and varieties for stalk sugar traits in sorghum. *Journal of SAT Agricultural Research* 5(1): 1-3.

Reddy, B. V. S. and Sanjana R. P. (2003). Sweet sorghum: characteristics and potential. *International sorghum and millets Newsletter* 44: 26-28.

Reddy, B.V.S., Ramesh, S., Reddy, P. S., Ramaiah, B., Salimath, P. M. and Kachapur, R. (2005). Sweet sorghum a potential alternative raw material for bio-ethanol and bio-energy. *International sorghum and Millets Newsletter* 46: 79-86.

Reddy, M. P, Sarla, N. and Siddiq, E. A. (2002). Inter simple sequence repeat (ISSR) polymorphism and its application in plant breeding. *Euphytica* 128: 9-17.

Ritter, K. B., McIntyre, C. L., Godwin, I. D., Jordan, D. R. and Chapman, S. C. (2007). An assessment of the genetic relationship between sweet and grain sorghums, within *Sorghum bicolor* ssp. *bicolor* (L.) Moench, using AFLP markers. *Euphytica* 157: 161-176.

Rooney, W. L., Blumenthal, J., Bean, B. and Mullet, J. E. (2007). Designing sorghum as a dedicated bioenergy feedstock. *Biofuels Bioproducts Biorefining.* 1: 147-157.

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

Sallu, R. S. (2008). Genotyping sorghum germplasm in Tanzania using microsatellite markers. Dissertation for award of MSc. Degree at Sokoine University of Agriculture. Morogoro, Tanzania. 75pp.

Schloss, S. J., Mitchell ,S. E., White, G. M., Kukatla, R., Bowers, J. E., Paterson, A. H. and Kresovich, S. (2002). Characterization of RFLP clone sequences for gene discovery and SSR development in *Sorghum bicolor* (L). Moench. *Theoretical and Applied Genetics* 105: 912-920.

Schlotterer, C. (1998). Microsatellites In: *Molecular genetic analysis of populations: A practical approach.* (Edited by Hoelzel A. R.), IRL Press, Oxford, pp. 237-261.

Seetharama, N., Rao, K., Subramanian, V. and Murty, D. (1987). Screening for sweet stalk sorghums and environmental effect on stalk sugar concentrations. In: *Technology and applications for alternative uses of sorghum: Proceedings of the national seminar*. Parbhani, India. pp. 169-179.

Semagn, K. Bjørnsta, Å. and Ndjiondjop, M. N. (2006). An overview of molecular marker methods for plants. *African Journal of Biotechnology*. 5 (25): 2540-2568 [http://www.academicjournals.org/AJB]. Site visited on 30/01/2010.

Smith, J. S. C., Kresovich, S., Hopkins, M. S., Mitchell, S. E., Dean, R. E., Woodman, W. L., Lee, M. and Porter, K. (2000). Genetic diversity among elite sorghum inbred lines assessed with simple sequence repeats. *Crop Science* 40: 226-232.

Smith, S. E. (2008). What is Sorghum? [http://www.wisegeek.com/what is sorghu.htm]. Site visited on 07/01/2008.

Songela F. A. (2008). Status of Biofuels Industry in Tanzania. Biofuels Stakeholders Workshop at Oasis Hotel, Morogoro.9th - 11th June 2008. [www.tnrf.org/.../E-INFO_Biofuels_workshop_presentation_on_Status_of_Biofuels_Industry_in_Tanzania_Songela_2008.pdf]. Site visited on 07/10/2010

Stachael, M., Lelly, T., Grausgruber, H. and Vollmann, J. (2000). Application of microsatellites in wheat (*Triticum aestivum* L.) for studying genetic differentiation caused by selection for adaptation and use. *Theoretical and Applied Genetics* 100: 242-248.

Swofford, D. L., Olsen, G. J., Waddell, P. J. & Hillis, D. M. (1996). Phylogenetic inference. In: *Molecular Systematics, 2nd Ed*. (Edited by Hillis, D. M. Moritz, C. & Mable, B. K.). Sinauer, Sunderland, Massachusetts. pp. 407-514.

Taramino, G, Tarchini, R, Ferrario, S, Lee, M., and Pe´, M. E. (1997) Characterization and mapping of simple sequence repeats (SSRs) in *Sorghum bicolor*. *Theoretical and Applied Genetics* 95: 66-72.

Tarpley, L., and Vietor, D.M. (2007). Compartmentation of sucrose during radial transfer in mature sorghum culm. *BioMed Central Plant Biology* 7: 33-47.

TaTEDO, (2008). Competence Platform on Energy Crop and Agroforestry Systems for Arid and Semi-arid Ecosystems - Africa In: *National Policies and Strategies on Bioenergy in Africa Case Study: Tanzania (*Edited by Sawe E.). 8pp [www.compete-bioafrica.net/.../COMPETE-032448-NationalBioenergyPolicy-Tanzania-0801.pdf]. Site visited on 23/05/2010.

Tawanda, Z. (2004). DNA-Based methods in sorghum diversity studies and improvement. *Plant Biotechnology Center,* Ohio State University. [www.biotec.or.th/Biosafety/ web/db/attach/radDO1F1.pdf]. Site visited on 14/02/2010.

UN-Energy. (2007). Sustainable Energy: A framework for decision-makers. [http://esa.un.org/un-energy/pdf/susdev.Biofuels.FAO.pdf]. Site visited on 15/02/2010.

Vanderlip, R. L. (1993). How a Grain Sorghum Plant Develops, Kansas State University. [www.uaex.edu/Other_Areas/publications/PDF/MP297/1_growth...]. Site visited on 05/10/2010.

Vermerris, W., Saballos, A., Ejeta, G., Mosier, N. S., Ladisch, M. R. and Carpita N. C. (2007). Molecular breeding to enhance ethanol production from corn and sorghum stover. *Crop Science* 47(3): 142-153.

Wang, M. L., Zhu, C., Barkley, N. A., Chen, Z., Erpelding, J. E., Murray, S. C., Tuinstra, M. R., Tesso, T., Pederson, G. A. and Yu, J. (2009). Genetic diversity and population structure analysis of accessions in the US historic sweet sorghum collection. *Theoretical and Applied Genetics* 120(1): 13-23.

Yang, S., Pang, W., Ash, G., Harper, J. Carling, J., Wenzl, P., Huttner, E., Zong, X., and Kilian, A. (2006). Low level of genetic diversity in cultivated Pigeon pea compared to its wild relatives is revealed by diversity arrays technology. *Theoretical and Applied Genetics* 113: 585-595.

APPENDICES

Appendix 1: Rainfall data at Ukiriguru station from January 2008 to December 2009

	2008			2009		
Months	Maximum temperature (°C)	Minimum temperature (°C)	Rainfall (cm)	Maximum temperature (°C)	Minimum temperature (°C)	Rainfall (cm)
January	28.4	18.2	61.8	28.6	18.6	133.1
February	27.7	18.1	135.7	29.0	18.5	37.0
March	28.2	18.0	153.0	28.5	18.7	79.2
April	28.2	17.5	96.8	28.3	18.4	107.4
May	29.2	17.5	13.2	29.0	17.8	36.6
June	28.9	15.7	0.0	29.8	15.4	21.5
July	29.3	15.9	1.3	30.0	15.5	0.0
August	30.0	17.1	3.5	31.0	16.5	6.2
September	30.3	18.6	30.2	31.0	18.8	8.2
October	28.9	18.9	117.7	31.0	19.1	55.2
November	27.9	19.5	162.1	28.5	18.8	96.7
December	29.0	18.7	98.5	28.3	18.5	107.8

Appendix 2: Less distinct of brix levels based on NJ simple matching dissimilarity matrix (Darwin V.5.25)

S.no.	Names of accession pair		Dissimilarity matrix
1.	*GDG_Chi*	*NGD_Ki*	0.06
2.	*NGD_Ki*	*MWGD_Mnc*	0.10
3.	*GDG_Chi*	*MWGD_Mnc*	0.11
4.	*TEG_Bu*	*TEG-Sel*	0.11
5.	*NGW_Sa*	*NGD_Ki*	0.13
6.	*NGW_Sa*	*GDG_Chi*	0.13
7.	*UNK_Ki*	*Unk_Ise*	0.14
8.	*NGW_Sa*	*MWGD_Mnc*	0.15
9.	*NKO_Ko*	*GDG_Chi*	0.15
10.	*S_Ind_35*	*LUL*	0.15
11.	*SDSL_90167*	*IESV_92001DL*	0.15
12.	*NKO_Ko*	*NGD_Ki*	0.17
13.	*SEL_Mch*	*SA_Se*	0.19

Appendix 3: Brix levels relationship of sweet sorghum accessions based on NJ (Darwin V.5.25

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1 RMCH																															
2 TEG/Sel	0.30																														
3 KENR	0.32	0.37																													
4 MWGD-M	0.37	0.42	0.40																												
5 NKLY N	0.31	0.36	0.30	0.39																											
6 TEG Bu	0.34	0.39	0.37	0.42	0.36																										
7 WEG-Ny	0.38	0.43	0.41	0.30	0.41	0.43																									
8 NGH-Te	0.50	0.55	0.53	0.59	0.53	0.46	0.60																								
9 MKO	0.28	0.29	0.35	0.40	0.34	0.36	0.41	0.53																							
10 Mna	0.33	0.38	0.36	0.41	0.35	0.35	0.42	0.51	0.35																						
11 MWA-Pj	0.41	0.46	0.44	0.32	0.44	0.46	0.34	0.63	0.44	0.45																					
12 LIM	0.35	0.40	0.38	0.43	0.37	0.37	0.44	0.53	0.37	0.30	0.47																				
13 GDG-Chi	0.32	0.28	0.39	0.44	0.38	0.41	0.45	0.57	0.31	0.40	0.48	0.42																			
14 KULY-M	0.36	0.41	0.39	0.37	0.38	0.41	0.38	0.57	0.39	0.40	0.41	0.42	0.43																		
15 NGH-Chi	0.40	0.45	0.43	0.38	0.42	0.45	0.39	0.61	0.42	0.44	0.42	0.46	0.47	0.39																	
16 LUL	0.35	0.30	0.41	0.46	0.40	0.43	0.47	0.59	0.33	0.42	0.50	0.44	0.32	0.45	0.49																
17 NGW-Sa	0.41	0.46	0.44	0.24	0.43	0.46	0.34	0.62	0.43	0.45	0.35	0.47	0.48	0.40	0.42	0.50															
18 WEG-Bu	0.39	0.44	0.42	0.31	0.42	0.44	0.29	0.61	0.42	0.44	0.35	0.46	0.46	0.39	0.40	0.48	0.35														
19 NDH-Chi	0.47	0.52	0.50	0.39	0.50	0.52	0.37	0.69	0.50	0.52	0.43	0.53	0.54	0.47	0.48	0.56	0.43	0.32													
20 NGJ	0.51	0.56	0.55	0.60	0.54	0.48	0.61	0.42	0.54	0.53	0.64	0.55	0.58	0.59	0.63	0.61	0.64	0.62	0.70												
21 NGD-Ki	0.41	0.46	0.44	0.25	0.43	0.46	0.34	0.62	0.44	0.45	0.36	0.47	0.48	0.41	0.42	0.50	0.19	0.35	0.43	0.64											
22 Unk-Ise	0.37	0.42	0.41	0.46	0.40	0.39	0.47	0.55	0.40	0.38	0.50	0.40	0.44	0.45	0.48	0.46	0.49	0.48	0.56	0.57	0.50										
23 NKO-Ko	0.35	0.40	0.38	0.33	0.37	0.40	0.34	0.56	0.37	0.39	0.37	0.41	0.42	0.35	0.36	0.44	0.36	0.35	0.43	0.58	0.36	0.43									
24 SEL-Mch	0.37	0.42	0.40	0.35	0.39	0.42	0.36	0.58	0.40	0.41	0.39	0.43	0.44	0.37	0.38	0.46	0.38	0.37	0.45	0.60	0.39	0.46	0.25								
25 SA-Se	0.40	0.45	0.43	0.28	0.43	0.45	0.33	0.62	0.43	0.45	0.35	0.46	0.47	0.40	0.41	0.49	0.31	0.34	0.42	0.63	0.32	0.49	0.36	0.38							
26 UNK-Ki	0.30	0.35	0.36	0.41	0.35	0.37	0.42	0.53	0.33	0.36	0.45	0.38	0.37	0.40	0.43	0.39	0.44	0.43	0.51	0.55	0.45	0.41	0.38	0.41	0.44						
27 MKG-At	0.37	0.42	0.40	0.36	0.40	0.42	0.37	0.59	0.40	0.41	0.40	0.43	0.44	0.37	0.30	0.46	0.39	0.38	0.46	0.60	0.40	0.46	0.33	0.36	0.39	0.41					
28 MWA-M	0.28	0.33	0.34	0.39	0.33	0.36	0.40	0.52	0.31	0.35	0.43	0.37	0.35	0.38	0.42	0.37	0.43	0.41	0.49	0.54	0.43	0.39	0.37	0.39	0.42	0.28	0.39				
29 WEG-Sa	0.35	0.40	0.37	0.35	0.37	0.40	0.37	0.56	0.37	0.39	0.40	0.41	0.42	0.32	0.38	0.44	0.39	0.38	0.46	0.58	0.39	0.43	0.33	0.35	0.39	0.38	0.36	0.37			
30 KINY	0.38	0.43	0.41	0.46	0.41	0.34	0.48	0.46	0.40	0.39	0.51	0.41	0.45	0.45	0.49	0.47	0.50	0.49	0.57	0.48	0.50	0.43	0.44	0.46	0.50	0.41	0.47	0.40	0.44		
31 IS23/31	0.39	0.35	0.45	0.51	0.45	0.47	0.52	0.63	0.37	0.46	0.55	0.48	0.34	0.49	0.53	0.39	0.54	0.53	0.61	0.65	0.54	0.51	0.48	0.51	0.54	0.43	0.51	0.42	0.48	0.51	
32 KSV 700	0.44	0.40	0.50	0.55	0.49	0.52	0.56	0.68	0.42	0.51	0.59	0.53	0.38	0.54	0.58	0.44	0.59	0.58	0.66	0.70	0.59	0.55	0.53	0.55	0.59	0.48	0.55	0.47	0.53	0.56	0.45

Appendix 4: Genetic relationship of 43 sweet sorghum accessions based on NJ (Darwin V.5.25

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	46
1	MWGD_Mnc																																										
2	Mna	0.56																																										
3	Unk_Ise	0.54	0.48																																									
4	WEG_Sa	0.43	0.55	0.45																																								
5	NGD_Ki	0.10	0.52	0.56	0.42																																						
6	GDG_Chi	0.11	0.51	0.52	0.40	0.06																																					
7	KINY	0.55	0.54	0.50	0.42	0.52	0.54																																				
8	LIM	0.39	0.55	0.50	0.50	0.45	0.40	0.44																																			
9	WEG_Bu	0.48	0.52	0.51	0.38	0.55	0.54	0.38	0.48																																		
10	KENR	0.43	0.46	0.42	0.40	0.49	0.46	0.43	0.48	0.39																																	
11	SA_Se	0.56	0.51	0.27	0.52	0.62	0.60	0.51	0.55	0.50	0.37																																
12	TEG-Sel	0.70	0.54	0.55	0.61	0.63	0.65	0.68	0.70	0.70	0.61	0.50																															
13	TEG_Bu	0.71	0.50	0.51	0.60	0.64	0.64	0.64	0.71	0.73	0.64	0.50	0.11																														
14	MKG_At	0.63	0.48	0.35	0.56	0.69	0.65	0.54	0.60	0.55	0.45	0.27	0.54	0.52																													
15	NDH_Chi	0.36	0.57	0.52	0.39	0.38	0.35	0.52	0.50	0.50	0.32	0.46	0.62	0.62	0.51																												
17	KSV_700	0.65	0.63	0.54	0.63	0.64	0.62	0.61	0.62	0.71	0.61	0.57	0.52	0.52	0.60	0.61																											
18	MWA_Ma	0.29	0.55	0.49	0.36	0.23	0.26	0.46	0.46	0.50	0.40	0.57	0.64	0.63	0.58	0.31	0.67																										
19	LUL	0.60	0.37	0.50	0.58	0.55	0.55	0.57	0.62	0.57	0.55	0.55	0.51	0.52	0.57	0.63	0.64	0.56																									
20	MWA-Pj	0.30	0.54	0.50	0.39	0.24	0.24	0.44	0.45	0.50	0.44	0.56	0.61	0.63	0.63	0.40	0.64	0.32	0.52																								
21	NKO_Ko	0.23	0.55	0.50	0.39	0.17	0.15	0.55	0.44	0.55	0.45	0.56	0.63	0.63	0.63	0.35	0.61	0.25	0.60	0.24																							
22	WEG_Ny	0.57	0.61	0.48	0.39	0.57	0.58	0.32	0.51	0.43	0.46	0.50	0.68	0.65	0.51	0.48	0.63	0.48	0.67	0.51	0.52																						
23	UNK_Ki	0.58	0.50	0.14	0.52	0.57	0.54	0.52	0.56	0.58	0.46	0.32	0.55	0.50	0.30	0.52	0.55	0.49	0.56	0.51	0.51	0.49																					
24	SEL_Mch	0.55	0.46	0.32	0.52	0.61	0.58	0.55	0.49	0.51	0.46	0.19	0.56	0.50	0.36	0.52	0.62	0.57	0.52	0.57	0.58	0.55	0.37																				
25	IS23/31	0.45	0.43	0.58	0.56	0.46	0.46	0.52	0.49	0.51	0.56	0.64	0.64	0.65	0.62	0.60	0.64	0.50	0.44	0.49	0.51	0.60	0.62	0.61																			
26	RMCH	0.62	0.57	0.42	0.57	0.68	0.67	0.48	0.51	0.54	0.49	0.40	0.61	0.57	0.31	0.56	0.63	0.57	0.58	0.60	0.63	0.44	0.38	0.44	0.63																		
27	NGW_Sa	0.15	0.57	0.51	0.44	0.13	0.13	0.60	0.48	0.58	0.49	0.58	0.70	0.71	0.67	0.38	0.64	0.31	0.60	0.31	0.23	0.58	0.56	0.61	0.50	0.67																	
28	NGH_Chi	0.37	0.48	0.31	0.40	0.39	0.35	0.51	0.43	0.49	0.40	0.44	0.55	0.52	0.40	0.42	0.60	0.35	0.56	0.36	0.36	0.54	0.35	0.46	0.46	0.44	0.40																
29	MKO	0.51	0.57	0.52	0.51	0.54	0.51	0.64	0.60	0.61	0.51	0.52	0.65	0.64	0.57	0.49	0.69	0.51	0.54	0.54	0.52	0.58	0.54	0.50	0.61	0.60	0.55	0.45														
30	KULY_Mc	0.57	0.49	0.40	0.49	0.61	0.57	0.54	0.58	0.52	0.45	0.36	0.57	0.52	0.36	0.48	0.60	0.50	0.55	0.57	0.56	0.51	0.32	0.32	0.64	0.44	0.61	0.45	0.48													
31	NGJ	0.43	0.69	0.58	0.55	0.40	0.40	0.64	0.60	0.63	0.60	0.61	0.74	0.75	0.71	0.60	0.67	0.48	0.64	0.43	0.40	0.64	0.64	0.68	0.64	0.65	0.42	0.52	0.60	0.68												
32	NKLY_Ny	0.61	0.46	0.42	0.51	0.65	0.61	0.46	0.52	0.50	0.46	0.44	0.61	0.56	0.33	0.45	0.57	0.54	0.57	0.64	0.61	0.50	0.39	0.43	0.55	0.44	0.63	0.43	0.61	0.37	0.71											
33	IS_2331	0.50	0.49	0.61	0.61	0.52	0.52	0.61	0.51	0.56	0.61	0.67	0.67	0.68	0.65	0.58	0.67	0.51	0.52	0.56	0.55	0.63	0.64	0.65	0.08	0.68	0.54	0.49	0.61	0.67	0.69	0.58										
34	ICSV_930	0.67	0.56	0.57	0.71	0.65	0.62	0.73	0.68	0.77	0.65	0.57	0.51	0.46	0.63	0.64	0.30	0.71	0.60	0.68	0.62	0.69	0.57	0.57	0.64	0.63	0.68	0.63	0.63	0.54	0.77	0.63	0.67									
35	IESV_910	0.61	0.61	0.61	0.61	0.61	0.57	0.68	0.65	0.62	0.52	0.58	0.64	0.68	0.65	0.51	0.68	0.64	0.49	0.60	0.62	0.67	0.63	0.61	0.67	0.64	0.60	0.58	0.51	0.61	0.74	0.68	0.65	0.64								
36	SPV_422	0.57	0.46	0.54	0.68	0.61	0.60	0.54	0.56	0.61	0.54	0.55	0.56	0.55	0.54	0.62	0.63	0.61	0.37	0.62	0.68	0.57	0.56	0.51	0.45	0.48	0.61	0.55	0.65	0.54	0.73	0.57	0.52	0.61	0.58							
37	IESV_920	0.58	0.40	0.48	0.61	0.62	0.60	0.50	0.52	0.61	0.46	0.49	0.45	0.43	0.51	0.60	0.60	0.60	0.36	0.57	0.62	0.60	0.51	0.46	0.48	0.44	0.63	0.51	0.58	0.49	0.68	0.54	0.56	0.51	0.61	0.25						
38	SDSL_90	0.57	0.36	0.46	0.56	0.56	0.54	0.54	0.56	0.61	0.45	0.50	0.48	0.46	0.52	0.55	0.60	0.54	0.26	0.52	0.55	0.61	0.51	0.48	0.48	0.52	0.57	0.51	0.55	0.48	0.63	0.54	0.56	0.52	0.55	0.30	0.15					
39	S_Nair_3	0.61	0.37	0.52	0.60	0.58	0.58	0.61	0.64	0.60	0.56	0.57	0.49	0.50	0.60	0.61	0.65	0.58	0.15	0.55	0.62	0.69	0.58	0.55	0.49	0.61	0.63	0.60	0.60	0.54	0.65	0.60	0.52	0.57	0.55	0.39	0.38	0.25				
40	IESV_920	0.62	0.51	0.61	0.65	0.64	0.62	0.60	0.50	0.68	0.57	0.58	0.46	0.52	0.62	0.62	0.63	0.64	0.39	0.61	0.69	0.68	0.64	0.57	0.56	0.58	0.68	0.56	0.60	0.58	0.71	0.67	0.54	0.58	0.49	0.33	0.26	0.36	0.37			
41	IESV_920	0.58	0.38	0.50	0.58	0.60	0.61	0.56	0.55	0.55	0.49	0.43	0.58	0.56	0.51	0.63	0.68	0.60	0.39	0.56	0.63	0.62	0.51	0.40	0.49	0.52	0.62	0.50	0.52	0.46	0.65	0.52	0.54	0.62	0.51	0.39	0.37	0.37	0.42	0.38		
42	IESV_930	0.55	0.43	0.42	0.55	0.57	0.55	0.50	0.54	0.48	0.43	0.33	0.51	0.49	0.43	0.52	0.57	0.52	0.32	0.56	0.56	0.56	0.43	0.37	0.56	0.49	0.60	0.49	0.57	0.43	0.70	0.45	0.62	0.52	0.54	0.37	0.36	0.33	0.32	0.40	0.31	
46	S_Ind_35	0.51	0.65	0.70	0.69	0.68	0.71	0.67	0.70	0.63	0.63	0.57	0.62	0.67	0.71	0.71	0.73	0.27	0.69	0.73	0.71	0.68	0.62	0.56	0.77	0.75	0.67	0.56	0.67	0.79	0.69	0.63	0.64	0.52	0.50	0.51	0.45	0.33	0.40	0.46	0.44	0.36
47	SSV_84	0.30	0.46	0.56	0.55	0.52	0.55	0.57	0.57	0.50	0.50	0.50	0.49	0.48	0.58	0.60	0.55	0.32	0.50	0.58	0.63	0.51	0.48	0.39	0.54	0.55	0.50	0.57	0.52	0.64	0.46	0.48	0.56	0.58	0.29	0.27	0.23	0.35	0.43	0.31	0.36	0.50	0.35

1. MWGD_Mnc. 2. Mna, 3. Unk_Ise, 4. WEG_Sa, 5. NGD_Ki, 6. GDG_Chi, 7. KINY, 8. LIM, 9. WEG_Bu, 10. KENR, 11. SA_Se, 12. TEG-Sel, 13. TEG_Bu, 14. MKG_At, 15. NDH_Chi, 17. KSV_700, 18. MWA_Ma, 19. LUL, 20. MWA-Pj, 21. NKO_Ko, 22. WEG_Ny, 23. UNK_Ki, 24. SEL_Mch, 25. IS23/31, 26. RMCH, 27. NGW_Sa, 28. NGH_Chi, 29. MKO, 30. KULY_Mch. 31. NGJ, 32. NKLY_Ny, 33. IS_2331, 34. ICSV_93046, 35. IESV_91018LT, 36. SPV_422, 37. IESV_92001DL, 38. SDSL_90167, 39. S_Nair_35, 40. IESV_92008DL, 41. IESV_92028DL, 42. IESV_93042SH, 46. S_Ind_35, and 47. SSV_84

Appendix 5: Nano readings genomics DNA of 42 accessions

S No.	Sample ID name	Genotype	Ng/µl	A260	A280	260/280	260/230
1.	set 1 swt sorg 1.1	MWGD_Mnc	43.64	0.87	0.50	1.74	1.17
2.	set 1 swt sorg 1.2	Mna	80.27	1.61	0.87	1.84	1.47
3.	set 1 swt sorg 1.3	Unk_Ise	141.42	2.83	1.49	1.90	1.89
4.	set 1 swt sorg 1.4	WEG_Sa	120.41	2.41	1.29	1.87	1.87
5.	set 1 swt sorg 1.5	NGD_Ki	103.10	2.06	1.13	1.82	1.48
6.	set 1 swt sorg 1.6	GDG_Chi	137.55	2.75	1.50	1.83	1.67
7.	set 1 swt sorg 1.7	KINY	91.06	1.82	0.98	1.85	1.56
8.	set 1 swt sorg 1.8	LIM	151.60	3.03	1.60	1.89	1.69
9.	set 1 swt sorg 1.9	WEG_Bu	81.92	1.64	0.88	1.86	1.78
10.	set 1 swt sorg 1.10	KENR	94.09	1.88	0.99	1.91	1.86
11.	set 1 swt sorg 1.11	SA_Se	49.78	1.00	0.56	1.78	1.51
12.	set 1 swt sorg 1.12	TEG-Sel	119.04	2.38	1.25	1.91	1.99
13.	set 1 swt sorg 1.13	TEG_Bu	67.65	1.35	0.73	1.86	1.69
14.	set 1 swt sorg 1.14	MKG_At	81.25	1.63	0.87	1.87	1.36
15.	set 1 swt sorg 1.15	NDH_Chi	109.57	2.19	1.15	1.91	1.88
16.	set 1 swt sorg 1.16	NGH_Te	72.61	1.45	0.76	1.90	1.74
17.	set 1 swt sorg 1.17	KSV_700	60.21	1.20	0.63	1.93	1.48
18.	set 1 swt sorg 1.18	MWA_Ma	83.63	1.67	0.89	1.89	1.50
19.	set 1 swt sorg 1.19	LUL	236.41	4.73	2.53	1.87	1.80
20.	set 1 swt sorg 1.20	MWA-Pj	79.98	1.60	0.88	1.82	1.39
21.	set 1 swt sorg 1.21	NKO_Ko	95.12	1.90	0.99	1.92	1.90
22.	set 1 swt sorg 1.22	WEG_Ny	136.66	2.73	1.45	1.88	1.89
23.	set 1 swt sorg 1.23	UNK_Ki	81.33	1.63	0.86	1.90	1.72
24.	set 1 swt sorg 1.24	SEL_Mch	101.42	2.03	1.07	1.89	1.94
25.	set 1 swt sorg 1.25	IS23/31	86.43	1.73	0.88	1.97	1.77
26.	set 1 swt sorg 1.26	RMCH	153.14	3.06	1.64	1.87	1.55
27.	set 1 swt sorg 1.27	NGW_Sa	160.86	3.22	1.73	1.86	1.48
28.	set 1 swt sorg 1.28	NGH_Chi	99.36	1.99	1.12	1.78	0.97
29.	set 1 swt sorg 1.29	MKO	59.87	1.20	0.72	1.67	0.73
30.	set 1 swt sorg 1.30	KULY_Mch	101.30	2.03	1.07	1.90	1.64
31.	set 1 swt sorg 1.31	NGJ	64.65	1.29	0.65	2.00	1.57
32.	set 1 swt sorg 1.32	NKLY_Ny	53.89	1.08	0.55	1.96	1.47
33.	set 1 swt sorg 1.33	IS 2331	92.73	1.86	0.99	1.87	1.36
34.	set 1 swt sorg 1.34	ICSV 93046	87.07	1.74	0.93	1.88	1.59
35.	set 1 swt sorg 1.35	IESV 91018LT	72.43	1.45	0.75	1.93	1.65
36.	set 1 swt sorg 1.36	SPV 422	83.26	1.67	0.89	1.87	1.77
37.	set 1 swt sorg 1.37	IESV 92001 DL	253.23	5.07	2.69	1.88	1.75
38.	set 1 swt sorg 1.38	SDSL 90167	197.96	3.96	2.14	1.85	1.74
39.	set 1 swt sorg 1.39	S 35	44.97	0.90	0.47	1.91	1.76
40.	set 1 swt sorg 1.40	IESV 92008 DL	97.55	1.95	0.91	2.16	1.60
41.	set 1 swt sorg 1.41	IESV 92028 DL	101.07	2.02	1.09	1.85	1.46
42.	set 1 swt sorg 1.42	IESV 93042 SH	123.64	2.47	1.47	1.68	0.96

Appendix 6: Forty eight sorghum accessions from ICRISAT and Tanzania

Acc no	Id abbreviation	Designated name	Origin	Acc no	Id abbreviation	Designated name	Origin
1.	*MWGD_Mnc*	*Mwanagudungu*	Tanzania	25.	*KSV_700*	*KSV 700*	ICRISAT_Nairobi
2.	*Mna*	*Mnanso*	Tanzania	26.	*RMCH*	*Redi*	Tanzania
3.	*Unk_Ise*	*Unknown Shija-Iseme*	Tanzania	27.	*NGW_Sa*	*Ng'wanagudungu-Sangija*	Tanzania
4.	*WEG_Sa*	*Wegita-Sapiwi*	Tanzania	28.	*NGH_Chi*	*Ng'holongo-Chinamile*	Tanzania
5.	*NGD_Ki*	*Ngudungu-Kihandanila*	Tanzania	29.	*MKO*	*Mkombituna*	Tanzania
6.	*GDG_Chi*	*Gudungu-chinamile*	Tanzania	30.	*KULY_Mch*	*Kulya-mchanganyiko*	Tanzania
7.	*KINY*	*Kinyenche*	Tanzania	31.	*NGJ*	*Ng'wajinila*	Tanzania
8.	*LIM*	*Limuche*	Tanzania	32.	*NKLY_Ny*	*Nkulya-Nyeupe*	ICRISAT_Nairobi
9.	*WEG_Bu*	*Wegita-Bunda*	Tanzania	33.	*IS 2331*	*IS 2331*	ICRISAT_Nairobi
10.	*KENR*	*Kenya Redi (mchanganyiko)*	Tanzania	34.	*ICSV 93046*	*ICSV 93046*	ICRISAT_Nairobi
11.	*SA_Se*	*Samumba Serena*	Tanzania	35.	*IESV 91018LT*	*IESV 91018LT*	ICRISAT_Nairobi
12.	*TEG-Sel*	*Tegemea/selemani (Mchanganyiko)*	Tanzania	36.	*SPV 422*	*SPV 422*	ICRISAT_Nairobi
13.	*TEG_Bu*	*Tegemea-Bumela*	Tanzania	37.	*IESV 92001 DL*	*IESV 92001 DL*	ICRISAT_Nairobi
14.	*MKG_At*	*Mkungunyamatu*	Tanzania	38.	*SDSL 90167*	*SDSL 90167*	ICRISAT_Nairobi
15.	*NDH_Chi*	*Ngudungu-Nsulwa Sabi*	Tanzania	39.	*S 35*	*S 35*	ICRISAT_Nairobi
16.	*NGH_Te*	*Ng'holongo-Tembe*	Tanzania	40.	*IESV 92008 DL*	*IESV 92008 DL*	ICRISAT_Nairobi
17.	*IS23/31*	*IS23/31*	Tanzania	41.	*IESV 92028 DL*	*IESV 92028 DL*	ICRISAT_Nairobi
18.	*MWA_Ma*	*Mwanagudungu-Myana*	Tanzania	42.	*IESV 93042 SH*	*IESV 93042 SH*	ICRISAT_Nairobi
19.	*LUL*	*Lulu*	Tanzania	43.	*ICSL73200*	*ICSL73200*	ICRISAT_India
20.	*MWA_Pj*	*Mwanagudungu-Paji*	Tanzania	44.	*ICSL73221*	*ICSL73221*	ICRISAT_India
21.	*NKO_Ko*	*Nkombituna*	Tanzania	45.	*ICSV111*	*ICSV111*	ICRISAT_India
22.	*WEG_Ny*	*Wegita-Nyupe*	Tanzania	46.	*S35*	*S35*	ICRISAT_India
23.	*UNK_Ki*	*Unknown-Kishapu*	Tanzania	47.	*NTJSSV84*	*NTJSSV84*	ICRISAT_India
24.	*SEL_Mch*	*Serena Mchanganyiko na Selemani*	Tanzania	48.	*BTx632*	*BTx632*	ICRISAT_India