GENOTYPING SORGHUM GERMPLASM IN TANZANIA USING MICROSATELLITE MARKERS.

BY

RAPHAEL SAMWEL SALLU

A THESIS SUBMITTED IN FULFILLMENT OF REQUIREMENT FOR THE DEGREE OF MSc. OF SOKOINE UNIVERSITY OF AGRICULTURE, MOROGORO, TANZANIA.

2008

ABSTRACT

Microsatellite markers are increasingly being used in crop plants to discriminate among genotypes and as tools in marker-assisted selection. In this study microsatellite markers were used to quantify the genetic diversity within as well as among 200 accessions sampled from sorghum germplasm collection at Tanzania National gene bank germplasm collection of sorghum. Although all methods did not provide similar description of relationships between accessions, there existed some consistency in discriminating accessions which are closely related and the ones which were distantly related. But, considerable variation was found at the 39 microsatellite markers analysed, with an average number of alleles per locus equal to 9.49 within accessions, the lowest was 2.0 from X*txp*114, Xcup61 and the highest number of allele was 25 as for X*gap*206 marker. The collection of sorghum appeared moderately structured genetically with about 59% of the average gene diversity occurring among accessions. The SSR markers were moderately polymorphic, with diversity indices ranging from 0.07 to 0.91 with mean of 0.55.The UPGMA dendrogram based on SSR marker data clearly discriminated among clusters, even though some consistency in classification was observed among clusters. However, differentiation among morphologically accessions of sorghum, or among geographic origins, accounted for less than 35% of the total genetic diversity. Data in this study demonstrated that accessions of Tanzania sorghum contain a great deal of genetic diversity as indicated by the observed number of alleles. These results are in global agreement with those obtained previously with allozyme markers. It was also possible to show that microsatellite data are useful in

identifying individual accessions with a high relative contribution to the overall allelic diversity of the collection. Therefore, from the result outcome the inventory was compiled that will be used in future to characterize the rest of the sorghum germplasm and make use of the identified potential parental genotypes for mapping populations and marker assisted selection programs.

DECLARATION

I, RAPHAEL SAMWEL SALLU do hereby declare to the senate of Sokoine University of Agriculture, that this dissertation is my own work and has not been submitted for a higher degree award in any other University.

_____ _____

Raphael Samwel Sallu Date
(MSc Crop Science and Production Candidate)

The above declaration is confirmed

_____ _____

Prof. S. Nchimbi-Msolla Date

(Supervisor)

COPYRIGHT

ACKNOWLEDGEMENTS

I wish to express my appreciation to my employer Ministry of Agriculture, Food Security and Cooperatives for providing me with a scholarship, financial support and a study leave during my studies. Thanks are extended to the ICRISAT/BecA for taking care of my research cost and supervision in Tanzania and in Kenya.

I wish to express my deepest appreciation to my major Supervisor Prof. Susan Nchimbi-Msolla and to Dr Luiz Kihupi, Dr Dan Kiambi and Dr A. Mbwaga for their continuous constructive guidance, suggestions, corrections and encouragement. I would also like to thank Mr. Samuel Mndolwa the sorghum breeder at Ilonga Research station for his guidance during the phenotyping work and also Mercy Naitore Kairichi, Mr. Michael Kimani, Eric Magembe for their technical assistance during the genotyping work at ICRISAT laboratory Nairobi.

My special thanks are to Dr Jeremiah Haki the Director of Research and Training in the Ministry of Agriculture, Food Security and Cooperatives for nominating me to pursue this study. Thanks are also to Dr Alois Kullaya the Officer Incharge of Mikocheni Agricultural Research Institute, Dr Zuberi Seguni, Dr Anatolia Mpunami, Dr Grace Chipungahelo, Dr Cornel Rweyemamu and Dr E. Mneney who have contributed valuable suggestions and assistance. The author indebted to them and may God bless them. Last but not least the author's family, the beloved wife, Jane and dearest daughters, Mary Lisa, Mary Linda and Mary Lina, who deserve the greatest appreciation for their painstaking during long session of my absence from

home. But who were always ready to understand and therefore continuously encourage me in my work.

DEDICATION

This piece of work is dedicated to my parents Dr Samuel Sallu and my late mother Mary Chambo, Mrs. Agatha Sambali, Mr Mark Sallu, late brothers Gerald, Rogers and Alban Pancras Sallu.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS AND SYMBOLS

BecA	Bioscience East and Central Africa
BME	β-mercaptoethanol
BPB	bromophenol blue
CA	cytosine adenine
CTAB	mixed alkyltrimethyl-ammonium bromide
dATP	deoxyadenosine 5'-triphosphate
dCTP	deoxycytidine 5'-triphosphate
ddH_2O	double-distilled water
dH_2O	distilled water
DNA	deoxyribose nucleic acid
dNTPs	deoxynucleoside 5'-triphosphates
EDTA	ethylenediaminetetraacetate
EtBr	ethidium bromide
EtOH	ethanol
GT	guanine thymine
FAO	Food and Agriculture Organization of the United Nations
FAM	6-carboxy-fluorescine (Blue)
g	gram(s)
GCP	Generation Challenge Program
h	hour(s)
HEX	hexachloro-6-carboxy-flourescent (Green)

ICRISAT	International Crop Research for Semi-Arid Tropics
kb	Kilobases
KOAc	potassium acetate
mA	milli Amperes
MAS	marker assisted selection
min	minute(s)
ml	millilitre(s)
MW	molecular weight
NaOAc	sodium acetate
NED	6-carboxy-X-rhodamine (Yellow/Black)
ng	nanogram(s) = 10^{-9} gram
nm	nanometer(s) = 10^{-9} meter
NPGS	National Plant Germplasm System
Nacl	Sodium Chloride
OD	optical density
ODx	optical density at x nm
PAGE	Polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
pmol	Picomole
PSeGOli	Sequencing, Genotyping and Oligosynthesis Unit
RAPDs	Random Amplified Polymorphic DNAs
RFLPs	restriction fragment length polymorphisms
RNA	ribonucleic acid
Rpm	rounds per minute
QTL	Quantitative trait loci

SDS	sodium dodecyl sulphate
Sec	second(s)
SSRs	microsatellite or single sequence repeats
STE	sodium Tris-EDTA (also TEN)
TAE	Tris-acetate EDTA (buffer)
TBE	Tris-borate EDTA
TE	Tris-EDTA (buffer)
TNE	Tris Sodium (Na) EDTA (buffer)
Tris	Tris (hydroxymethyl) amino-methane
U	unit(s) of enzyme
UV	ultraviolet
U.S.A	United States of America
V	volts
VIC	tetrachloro-6-carboxy-fluorescine (green)
µg	microgram(s) = 10^{-6} gram
µl	microlitre(s) = 10^{-6} litre

CHAPTER ONE

1.0 INTRODUCTION

1.1 Background information

Sorghum [*Sorghum bicolor* (L) Moench] is the fifth important cereal grown worldwide in terms of both production and area planted (FAO, 2004). It originated from the north-east quadrant of Africa and spread from there throughout the world. Like millet and fanio, sorghum is genetically suited to hot and dry agro-ecological zones where it is difficult to grow most food grains. Three *S. bicolor* subspecies are recognized, including cultivated types (i.e., landraces) (ssp. *Bicolor*), a wild complex that is widely distributed and ecologically diverse (spp. *Verticilliflorum*), and weedy types from hybridizations between domesticated and wild sorghum (spp. *Drummondii*) (De Wet, 1978). Furthermore, five races (i.e., bicolor, caudatum, durra, guinea and kafir) and ten hybrid races that combine characteristics of two or more races have been described within spp. *Bicolor* (Harlan and De Wet, 1972). Sorghum, therefore, is a pillar of food security in the semi-arid areas of western and Central Africa.

The productivity of Sorghum [*Sorghum bicolor* (L) Moench] per unit area has been rather low due to a number of production constraints, which include presence of insect pests damage and diseases, environmental stress and inadequate supply of improved varieties. One of the serious problems observed on sorghum in Tanzania is

lack of improved varieties resistant to the *Striga* species. Striga species of economic importance are *Striga hermonthica*, *S. asiatica* and *S. forbesii* (Mbwaga and Obilana, 1993, Mbwaga *et al.*, 2000). The yield loss caused by Striga range from 40-90% and under high infestation on farmer's fields, yield loss of up to 100% has been reported.

The traditional methods used for breeding this crop have made a significant contribution to crop improvement, but they have been slow in targeting complex traits like grain yield, grain quality, and drought- or *striga* resistance. Although *Sorghum bicolor* is a species of main economic importance among the cereals crops cultivated in countries of tropical climates information regarding its genome organization and mapping is rather limited. In recent years, significant progress has been made in the use of molecular approaches for plant breeding. These approaches are applied through the use of two main, but entirely different strategies; each of them exploits the process of plant breeding. In the first strategy, a variety of transgenic crop have been produced that carry genes, which a plant breeder hitherto could not introduce by available conventional methods. In the second strategy, in several crops, molecular markers closely linked to numerous traits of economic importance have been developed (Caetano-Anolles & Gresshoff, 1997), which allows indirect selection for desirable traits in early segregating generations at seedling stages. This will save time, resources and energy that are needed not only for raising large segregating generations for several generations, but also estimating the parameters used for direct selection. Study of genetic diversity is the process by which variation among individuals or groups of individuals populations are analyzed by a specific method or a combination of methods. The molecular

diversity data can potentially bridge conservation and use when employed as a tool for mining germplasm collections for genomic regions associated with adaptive or agronomically-important traits (i.e., genes that have been important in adaptation of local environment or are associated with phenotypes selected by farmers or breeders). Methods of measuring genetic diversity have an important role within conservation programmes for genetic resources of crop plants (Newbury and Ford-Lloyd, 1997). However, in recent years, following the introduction of molecular markers in plant genetic research, considerable effort has been made to gain a better understanding of sorghum genetics and evolution, and important data have been gathered.

Conversely, DNA-based molecular markers are tools that enable plant breeders to directly evaluate genetic variation between related individuals without any concern for environmental factors and effect on gene expression levels. In addition, DNA techniques allow for the assessment of a theoretically unlimited number of polymorphic markers loci. Molecular markers, reveals differences in the DNA sequence of chromosomes derived from different progenitors. Gene tagging and QTL mapping in turn permit marker-assisted selection (MAS) in back cross, pedigree and population improvement programs (Mohan *et al*., 1997). This provides information on global genetic structure of species-typically by using twenty to fifty neutral markers which are the basis for association studies. They also determine entry points into vast germplasm collections. In the context of ex situ conservation of germplasm collections, methods based on genetic markers are used to address questions of identity, for instance to identify putative duplication accessions, and questions of relationship and structure, for instance to determine how variation is distributed

between individuals, accessions and races (Westman and Kresovich, 1997). In this study the GCP (Generation Challenge Program) has identified a set of high quality microsatellite markers that are being used for survey of a global composite of a set of germplasm.

1.2 Problem Statement and Justification

Climatologically, conditions in eastern and central Africa are changing, with the timing of rainy seasons becoming less reliable and rains resulting in lower quantities of water. This in combination with food security issues prompts the Ministry of Agriculture and Cooperative to put more emphasis on growing this crop. Sorghum, after pearl millet, is superior in drought tolerance and adaptability to poor soils and therefore has great potential in providing food security in Africa.

In Tanzania, the crop is almost grown in the most of the country but grown in large scale in semi arid areas such place like in central, western, southern and some parts of northern part of the country. The popular varieties include the local and improved ones such as Tegemeo, Macia, Weijita, Pato, Kayuma and Segaulane. In the National gene bank there are 818 sorghum accessions, their characterization and identification is only based on phenotypic expression and sometimes the locality where they were collected. Therefore, the challenges of identifying accession redundancies and uniqueness provides an opportunity for using molecular markers technology to improve collection quality and operational efficacy. Duplication or redundancies of accessions in gene bank is one of the existing problems which hinder the breeding program strategies. In order to enhance genetic potential of this crop,

there must be a comprehensive understanding of the amount and pattern of genetic variation that exist within and between the available cultivated accessions. There is a very urgent need to effectively apply molecular marker techniques for identifying and characterizing sorghum accessions that are in the National gene bank.

1.3 Objectives

In this study the main objective was to compile inventory of sorghum varieties and analyze the genetic variations among the sorghum genotypes within Tanzanian germplasm in association with phenotype for crop cultivar development.

1.3.1 Specific Objectives

The specific objectives were as follows,

i) To compile an inventory of sorghum germplasm currently available in National collections and breeding programs, together with farmer's knowledge when appropriate.

ii) To assess and characterize sorghum genetic resources available in the National Gene Bank using both molecular markers and morphological characteristics.

iii) To identify potential parental genotypes for mapping populations and marker-assisted selection program.

CHAPTER TWO

2.0 LITERATURE REVIEW

2.1 Significance of crop

Sorghum [*Sorghum bicolor* (L) Moench] represents Africa's main contribution to the world food supply. (Purseglove, 1987). It is also an important cereal crop globally, primarily because of its productivity and its tolerance to drought and heat stress (Doggett, 1988). On the basis of intense selection pressure imposed by both environment and cultivators, sorghum embodies an extensive array of morphological forms and agricultural products (Harlan and de Wet, 1972, Morden *et al*., 1990, Menkir *et al*., 1997). Sorghum has a diverse genus belonging to the tribe Andropogoneae; it consists of cultivated and wild species, many of which are interfertile. *Sorghum bicolor* spp. (2n=20) is the most important toxin agronomically in that it includes the cultivated grain races. It is a diploid, is highly self-pollinated, and possesses considerable diversity, both morphologically and in agronomic traits, such as adaptive pest resistance. It consists of five basic races (bicolor, caudatum, durra, guinea, and kafir) and several hybrid races (Doggett, 1988). With an annual average production of 61 million tones over the past decade, sorghum is the fifth important grain crop worldwide. The bulk of African sorghum production is centered in the savannah zone of western; eastern and central Africa where grain of this crop is a major component of the daily menu for millions of people. It is the most important staple crop for food security as an important source of carbohydrates and as a cash crop. It is also used for brewing, feed for livestock and raw material for

many industrial products. Economically it gives one of the highest returns per man-hour of labour spent on it (Purseglove, 1987).

As a large collection of sorghum (*Sorghum bicolor* (L.) Moench) landrace collection held at the National gene Bank represent a challenge for the maintenance of both the accessions of and the information documented for the germplasm collection. The accessibility and knowledge of the landrace collection are the essential factors for an efficient utilization of the genetic resources by both breeders and farmers. Therefore, due to its socio-economic importance, there has been substantial interest in characterizing the levels of genetic diversity present within sorghum using both phenotypic and molecular makers (Dean *et al.,* 1999, Dje *et al.,* 2000, Ghebru *et al.,* 2002, Grenier *et al*., 2000 (a&b), Uptmoor *et al*., 2003) and these analyses have provided a foundation of genetic data for making informed decisions regarding the management and utilization of genetic resources.

2.2 Germplasm collection and conservation

Plant genetic diversity is the key component of any agricultural production systems and indeed, of any ecosystem. Without it, no natural, evolutionary adjustment of any system (agriculture or natural) to changing environment and biotic conditions would be possible (Frankel and Bennet, 1970). Central to the study of genetic diversity in wild crop relatives is the concept of the gene pool (Harlan and de Wet, 1971). Despite considerable phenotypic diversity, studies of pedigree records and comparative molecular assays suggest that genetic diversity in cultivated sorghum is limited (Duncan *et al*., 1991, Ahnert *et al*., 1996).

The conservation, management and the use of germplasm maintained in gene banks poses a number of challenges to the researchers dedicated to the plant genetic resources. (Dean *et al*., 1999). Common problems include, for example, the development of strategies for sampling representative individuals in natural populations, the improvement of tools and technology for long-term conservation or high throughput genetic analysis of an increasingly high number of stored accessions (Hamon *et al.,* 2004). Central to sustainable conservation is the knowledge of genetic diversity present in gene banks. This is also key to the potential exploitation of gene banks by breeding programmes. Therefore, the characterization of the accessions maintained in the collection and the examination of the genetic relationship between them is important for the sustainable conservation and increased use of crop genetic resources.

In response to this recognition of a limited genetic base, scientists associated with sorghum improvement have amassed extremely large ex situ collections. As of 1996, the collections at the International Crops Research Institute for semi-Arid Tropics (ICRISAT) and the U.S. National Plant Germplasm System (NPGS) numbered more than 35,000 and 40,000 accessions, respectively (Dahlberg *et al*., 1997, Eberhart *et al*., 1997). For effective conservation and use of crop genetic resources, *ex situ* collections of a taxon should contain genetic diversity that is both extensive and useful. Ideally, this diversity should be represented by a manageable number of accessions, given the constraints of available funding and staff expertise (Kresovich and McFerson, 1992). However, many plant genetic resources collections have been growing in at rates faster than that of operational support increases. The use of

molecular diversity data can potentially bridge conservation and use when employed as a tool for mining germplasm collections for genomic regions associated with adaptive or agronomically-important traits (i.e. genes that have been important in adaptation to local environment or are associated with phenotypes selected by farmers or breeders).

For development to be sustainable, conservation and use of genetic diversity must be at its core. Plant breeding is a continuous search for new sources of diversity (i.e. genes) which can be incorporated into advanced material in an effort to ease the pressure of finding new source of genetic variation. Information provided by the molecular markers is then used to design optimal procedures to manage extensive germplasm collections, in particular to highlight priorities in further sampling missions, to design germplasm regeneration programmes and to construct core collections. Molecular markers are recognized as significant tools to orient plant genetic resource conservation management, providing a means to accurately estimate the genetic diversity and genetic structure for species of interest (Hamrick and Godt, 1997). Many studies have been devoted to assessing patterns of sorghum genetic variation based on morphology (Appa-Rao *et al*., 1996, Dje *et al*., 1998) or pedigree (Jordan *et al*., 1998).

Therefore, having the knowledge of the phenotypic and genotypic diversity present in sorghum gene bank in Tanzania and dissemination of that material to sorghum breeders will be the starting point in developing improved varieties.

2.3 Methods of identification and characterization

Germplasm characterization of plant accessions deposited in gene banks has been limited and is probably a major cause for the limited use of accessions in breeding programmes (Mace *et al*., 2005). Germplasm characterization refers to the observation, measurement and documentation of highly heritable plants traits. Traits which are highly heritable can be easily seen by the eye and are expressed in all environments in a germplasm collection. The resulting data allows for identifying accessions, and building a catalogue of descriptors with embedded biological information that is essential for collection management or for direct use in agriculture (Hamon *et al.*, 2004). The characterization of plant germplasm, therefore, aims at describing and understanding the genetic diversity of the organism under study. Today, germplasm characterization has been developed based mostly on morphological descriptors, agronomic descriptors (traits) and molecular marker technology (Mace *et al*., 2005).

The current characterization of plant germplasm collections relies strongly on morphological descriptors. Morphological descriptors are reliable, easy to study and relatively low cost to evaluate. However, the use of morphological descriptors presents some limitations.

(i) Limited polymorphism, lowering the potential success of an extended classification approach, which would require a high number of descriptors in order to compensate for the small number of morphotypes;

(ii) Potential environment influence on the phenotype, making the process of evaluation and information exchange even more complex. In such cases, one should be careful with false positives when the environment affects specific morphotypes;

(iii) The impact of morphological descriptors in the viability of the individual.

Germplasm characterization based on agronomic traits, on the other hand, is particularly useful in crops of economic importance. The amount of data related to agronomic traits that is available by crop germplasm evaluation is limited (Parida *et al.,* 2005). Due to reasons that include relatively high costs and the difficulties of large-scale experimental trials, the use of agronomic evaluation to characterize germplasm collections is far from the actual need of uncovering the phenotypes of agronomic interest in accessions of a collection. More should definitely be done in this area to stimulate a higher use of stored germplasm in breeding programmes. A complete agronomic trait evaluation of crop germplasm in the next few years seems to be practically impossible to achieve. Analysis of phenotypic performance in the field in combination with molecular analysis provides useful information to increase the efficiency in plant breeding programs (House, 1985). Measurement of phenotypic and genotypic variance in field trials is a common and traditional approach to examine the genetic differences among genotypes. A field experiment is simpler and cheaper than DNA-based technology. However, unavoidable environmental variation in the field trial may mask actual genetic potency of a genotype as in the case of Hombolo field trial.

Therefore, the effective use of crop genetic resources stored in gene banks by breeding programmes is limited. The number of accessions deposited in gene banks, however, is continuously growing. Slow germplasm characterization has been pointed out as a major cause of the discrepancy (Gebhardt *et al*., 2004). According to Rafalski and Tingey (1993), many of these problems associated with breeding programs based on the phenotypic estimation of important agronomic traits can be overcome by the use of DNA-based techniques. More recently, DNA-based techniques have been used successfully in DNA finger printing of plant genome (Hongtrakul *et al*., 1997, Cervera *et al*., 1998) and in genetic diversity studies (Paul *et al*., 1997, Sonnate *et al*., 1997, Barrett and Kidwell, 1998, Chowdari *et al*., 1998b, Zhu *et al*., 1998, De-Buston *et al*., 1999). Molecular genetic markers such as RFLPs (Restriction Fragment Length Polymorphism), RAPDs (Random Amplified Polymorphic DNAs) and SSRs (single sequence repeats or microsatellites) have been used to characterize genetic diversity represented by elite inbred genotypes and cultivated races of sorghum (Brown et al., 1996, Ahnert *et al*., 1996, Menkir *et al.*,1997). The recent development and application of semi automated, robust, cost-effective molecular genetic markers, specially SSRs, now enhance the opportunity to establish and evaluate measures of quality for plant genetic resources collections (Mitchell *et al*., 1997). DNA-based assays are attractive because they reveal and their capacity to perform with great speed and efficiency, at all stages during the life cycle and independently of the environment. Molecular markers allow plant breeders to locate and follow the numerous interacting genes that determine a complex trait-QTL (quantitative trait loci), and are rapidly being adopted by crop improvement researchers globally as an effective and appropriate tool for basic and applied studies

addressing biological components in agricultural production systems (Jones *et al*., 1997, Mohan *et al*., 1997, Prioul *et al*., 1997). Molecular markers offer specific advantages in assessment of genetic diversity and in trait-specific crop improvement. Use of markers in applied breeding programs can range from facilitating appropriate choice of parents for crosses, to mapping/tagging of genes blocks associated with economically important traits. In this study genotyping was performed based on molecular marker (DNA markers) using high-throughput capillary based genotyping with 39 M_{13} labeled microsatellite primers for bulk samples and 20 selected individuals. Microsatellites are tandem repeats of short sequence motifs that occur ubiquitously in Eukaryotic genomes. One of the main features of this class of repetitive DNA is high level of variation among taxa, mainly expressed as a variable copy number of tandem repeats. Length variation of individual microsatellite loci is analyzed by PCR with a pair of locus specific flanking primers (Haussmann *et al.*, 2000). The DNA sequences flanking microsatellites are generally conserved within individuals of the same species, allowing the selection of PCR primers that will amplify the intervening SSR in all genotypes. Variation in the number of tandem repeats will result in different PCR product lengths. These repeats are highly polymorphic even among closely related cultivars, due to mutations causing variations in the number of repeating units. Unlike the other PCR-based marker techniques SSRs are inherited in a co-dominant fashion.

This allows one to discriminate between homo- and heterozygous state, and increases the efficiency of genetic mapping and population genetic studies. The SSRs have many features which are informative in revealing genetic differences among the

populations (Gupta *et al*., 1996, Powell *et al*., 1996a) The near-isogenic products of marker-assisted back-crossing programs provide genetic tools for use in improving our understanding of the mechanisms of various biotic stress tolerances (Jones *et al*., 1997, Prioul *et al*., 1997) and resistances to biotic production constraints. The value of SSRs for germplasm characterization has now been widely demonstrated in many crops (Maughan *et al.*, 1995, Provan *et al*., 1996, Taramino and Tingey 1996, Taramino *et al.*, 1997, Russel *et al*., 1997, Guilford *et al*., 1997). The high reproducibility of microsatellite in network activities has recently been demonstrated (Jones *et al*., 1997) and there are increasing numbers of examples of their use for identification of cultivars and for the construction of database (Thomas *et al.*, 1994, Guilford *et al*., 1997, Russel *et al*., 1997, Bowers and Meredith, 1997)**.** This makes them easily automated, highly polymorphic, have good analytical resolution, can be also amplified by PCR, codominant, highly informative and finally, radioisotopes are not required in the detection of SSR markers because sequence polymorphism usually can be detected by separation in agarose gels (Burr, 1994) thus making a preferred choice as markers. These systems also expose the worker to less toxic chemicals during handling and disposal. In addition, the feasibility of automated data analysis and the reproducibility of accurate sizing of microsatellite alleles to within +/-0.3 nucleotides after repeated injections are important for large scale germplasm genotyping projects (Mitchell *et al*., 1997). The genotyping with microsatellite markers was first described by Lity and Luty, and is now a widely used method in molecular genotyping. It has a firm place in forensic, diagnostic and scientific applications where haplotyping is employed in linkage analysis. Most genotyping is performed by polymerase chain reaction (PCR) with defined oligonucleotide

primers. For population analysis, accurate sizing of the microsatellite allele is crucial and as a rule, the separation capacity of the gel should be more than half the size of the repeat unit. Therefore, with dinucleotide repeat motifs analysed in this study, the detection system needed to be able to differentiate a 1bp difference. Fluorescent-based capillary detection systems have many advantages over the conventional gel based systems; automated filling of capillaries with polymer separation matrix, automated sample loading and rapid electrophoretic analysis of samples.

CHAPTER THREE

3.0 MATERIALS AND METHODS

3.1Plant materials

Two hundred accessions of sorghum [*Sorghum bicolor* (L) Moench] were obtained from the National gene bank located at Tropical Pesticide Research Institute (TPRI) Arusha. The materials for the genotyping were planted in a screen house in a well labeled pot of 350g by volume filled with sterile forest soil. The seeds were planted in bulk which consisted of more than 30 seedlings and each accession was replicated three times. For intra-population diversity 20 accessions were selected as a true representative from different geographical areas. An ammonium sulphate fertilizer was added 5 g per pot for promotion of fast vegetative growth. Irrigation of the seedlings was done as per moisture requirements. For field experiment 196 selected accessions were used for morphological characterization. The list and passport information of the accessions that were used in this study are summarized in Appendix 1.

3.2 Molecular sequencing and genotyping

Sequencing is the process of determining the exact order of bases in a segment of a DNA. In typical sequencing reactions, DNA from 96 or 384 well plates is added to a master mix buffer containing big dye, terminator reagent, primer, and the enzyme. The volume of a big dye used was 8-9.25 µl and 0.75-2.0 µl for DNA template. Sequencing reactions are then cleaned and processed on the DNA sequencer analyzers.

3.2.1 DNA extraction

For the bulk sampling 30 plants from each accession were harvested from 1-2 week old seedlings and divided in triplicate of 10 seedlings each bulk with 30 mg. These were stored in eppendorf tubes filled with 100% ethanol and kept in a fridge. The bulk samples which consisted of 3 eppendorf tubes containing 10 plants each were kept in one plastic bag and well labeled as well for intra-population diversity samples and each then transported to ICRISAT/BecA laboratory in Nairobi. The analysis of DNA bulks is the analysis of pooled DNA samples ("bulks"). Each bulk is supposed to combine equal amounts of DNA from each individual constituting the bulk. The use of DNA bulk necessitates a quantitative analysis of the obtained genotypes, in order to estimate allele frequencies. This method is thus faster and cheaper than traditional individual genotyping, and permits the realization of large population genetic studies, by using, for example, SSR markers. In this study the bulk preparation was done by pooling 10 individuals from each accession. For intra-population diversity study, 20 accessions representing different geographic origins were randomly selected (Table). Three seedlings harvest from each accessions as a separate samples and each kept in the eppendorf tube filled with 100% ethanol. For individual samples DNA extraction was done separate for each individual representing specific accession. The CTAB DNA extraction method was done according to the procedure described by Mace *et al.* (2003). The leaf samples of about 20-30 mg were cut into small pieces and loaded into 96 wells titer plate of size 15 cm. long x 10 cm. wide x 5 cm. high. Three stainless steel grinding balls (3 mm) were deposited in per titer well using a ball dispenser. The 450 µl of the pre-heated

CTAB extraction buffer (with 0.2% freshly added mercaptoethanol) poured into each titer well. Finally, the titer plate was sealed using clear adhesive films MicroAmp® (part No. 4306311) this was done to avoid one well from contaminating a sample in the adjacent well. The titer plate was clamped into the 2000 Geno/Grinder™ with the stroke rate of 500-2000/minute and adjusted using two spacers and locked down. The time set was 10 minutes, and after the lid was unlocked then the titer plate was unclamped. Then, titer plate containing macerated suspension was incubated in a water bath for 10 minutes at 65°C with occasional stirring. The incubated macerated suspensions were then transferred to a fresh eppendorf tube. The solvent extraction was done as follows, about 450 µl chloroform: isoamylalcohol (24:1) was added to each sample and inverted twice to mix and centrifuged at 12 000 rpm for 10 minutes then transferred to a fresh eppendorf tube. The clear supernatant was mixed with 0.8 volume of isopropanol and left overnight. Nucleic acids were pelleted by centrifugation at 12 000 rpm for 15 minutes. The supernatant was discarded and the pellet washed with 70% ethanol (approximately 500 µl) followed by another spin for 5-10 minutes. The ethanol was discarded and the pellet left to dry briefly at 37 ºC or at room temperature. Then the pellet was dissolved in 200 µl of TE and then 3 µl of RnaseA (10 mg/ml) were added and incubated for 30 min at 37ºC. The solvent extraction was done by as follows, 200µl phenol: chlorophorm:isoamylalcohol (25:24:1) added to each sample and invert twice to mix and centrifuged at 12 000 rpm for 10 minutes. Then the materials were transferred to a fresh eppendorf tube. 200 µl chlorophorm: isoamylalcohol (24:1) added to each sample and inverted twice to mix and centrifuged at 12 000 rpm for 10 minutes. The aqueous layer (approximately 180 µl) transferred to a fresh eppendorf tube.

The purification of the DNA was done as follows: 315 µl ethanol: sodium acetate solution was added to each sample and placed in -20ºC for 5 min.

Then centrifuged at 12 000 rpm for 5 minutes and the supernatant was decanted from each sample and pellet washed with 200 µl of 70% ethanol then centrifuged at 12 000 rpm for 5 minutes. The washing with 70% Ethanol was repeated and the supernatant decanted from each sample and the pellet was air-dried for approximately 1 hour then-resuspended in 100 µl low-salt TE buffer and stored at 4ºC.

Table 1: A list of twenty selected accessions used in the intra-population diversity study from different locations.

Number	Serial Number	Accession number	Collectors number	Place collected	Altitude (m)	Latitude	Longitude
1	151	TZA 3493	KEL 324	Nzega	1100 m	04 03' 12" S	33 12' 28" E
2	152	TZA 4166	CCM 125	Bukoba	1140 m	03 06' 04" S	31 07' 45" E
3	153	TZA 4247	MN 28	Bukoba	1111 m	01 12' 45" S	03 24' 26" E
4	154	TZA 3418	KEL 245	Igunga	1075 m	04 35' 48" S	33 50' 40" E
5	155	TZA 3436	KEL 263	Igunga	1050 m	04 14' 11" S	33 54' 00" E
6	156	TZA 3494	KEL 325	Nzega	1110 m	04 03' 12" S	33 12' 28" E
7	157	TZA 4001	NAM 94	Ukerewe	1080 m	02 07' 02" S	33 09' 07" E
8	162	TZA 4044	NAM 137	Ukerewe	1100 m	01 56' 14" S	32 52' 40" E
9	164	TZA 4155	CCM 114	Biharamulo	1100 m	03 06' 02" S	31 07' 44" E
10	166	TZA 3904	LNA 323	Nachingwea	220 m	10 31' 36" S	32 26' 28" E
11	170	TZA 3864	LNA 283	Nachingwea	190 m	10 07' 12" S	38 28' 51" E
12	173	TZA 3616	LNA 33	Mtwara	200 m	10 29' 51" S	39 50' 57" E
13	174	TZA 3882	LNA 301	Nachingwea	190 m	10 07' 12" S	38 28' 46" E
14	175	TZA 3943	NAM 36	Serengeti	1100 m	01 31' 44" S	34 31' 15" E
15	179	TZA 3835	LNA 254	Newala	660 m	40 30' 58" S	39 13' 27" E
16	181	TZA 3992	NAM 85	Musoma	1100 m	01 37' 29" S	34 13' 12" E
17	182	TZA 3994	NAM 87	Musoma	1100 m	01 35' 53" S	34 11' 14" E
18	186	TZA 3965	NAM 58	Serengeti	1500 m	01 41' 21" S	34 33' 44" E
19	189	TZA 4162	CCM 121	Biharamulo	1140 m	03 06' 24" S	31 08' 11" E
20	193	TZA 3147	HPL 63	Kasulu	1150 m	04 37' 28" S	30 15' 32" E

3.2.2 DNA Quality check and quantification.

T**o check the quality and concentration of extracted DNA 2 µl of the sample was run on 0.8% agarose gel.** Standard 0.8% (w/v) agarose gel was prepared by dissolving 2.4 g agarose in 300 ml 1XTBE electrophoresis buffer (0.04M Tris-Borate, 0.001M EDTA, pH 8.0). The mixture was heated in a hot plate or microwave to allow the agarose to dissolve and form a gel. The gel was allowed to cool to about 20^{o}C before adding 5 µl of ethidium bromide (10 mg/ml). The gel was then poured into a horizontal gel tray fitted with appropriate combs. After about 40 minutes of gel polymerization, the combs were carefully removed and the tray immersed in an electrophoresis tank containing electrophoresis buffer (1% x TBE). Then 2 µl of extracted DNA products of each individual sorghum accession was mixed with 2µl of bromophenol blue dye diluted 6x (3:1 water: 6 x dye, 50Mm EDTA, 50mM Nacl, 50% Glycerol) and then loaded into separate lanes (slots) of the submersed agarose gel. The samples were run alongside 1.0 µl (5ng/µl) lambda DNA digested with Hind III enzyme alone or Hind III and EcoRI enzymes at 120 volts for 45 minutes. After the run the gel was removed and photographed under UV light using a video capture system (Flowgen IS 1000) Figure 1. The quantity of each DNA sample was analyzed by using a Nano drop® ND-1000 spectrophotometer V3**.**1 (Nanodrop Technologies. INC.) After quantification, each DNA sample was diluted further at the ratio of 1:10 using double distilled H_20 (ddH_20) to a final working concentration of approximately 5ng/µl of DNA. All samples were put into PCR-384 well plates.

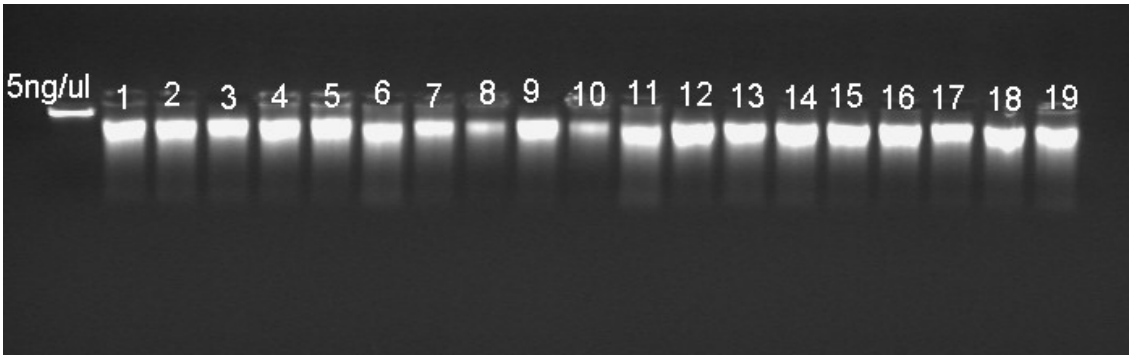

Figure 1: The quality of DNA samples that were run alongside 1.0 µl (5ng/µl) lambda DNA digested.

3.2.3 Polymerase Chain Reaction (PCR)

In this study a set of 39 sorghum SSR primers for bulk samples and 20 for individual samples were used for genotyping (Table 2). SSR markers were chosen based on three criteria: genome position, repeat size (ranging from dinucleotide to hexa-nucleotide repeats) and the number of previously reported alleles (Ranging from two to six). In order to analyse the length of the PCR products by electrophoresis and a leaser detection system, the forward primers were to carry a fluorescent dye label (M_{13} labelled).

These fluorescent dyes were 6-carboxy-fluorescine FAM (Blue), hexachloro-6-carboxy-fluorescent HEX (Green), 6-carboxy-X-rhodamine NED (Yellow/Black), tetrachloro-6-carboxy-fluorescine VIC (Green), or PET (Red) (PE-Applied Biosystems, Foster City, Calif.), allowing post-PCR pooling of the 69 primer products into seventeen groups of four primer products each, with each primer product given group being labelled with a different dye.

Table 2: Detailed information of the microsattelite markers used in the study

No (	Marker (SSR locus)	GCP registry allele size for BTx623	Repeat motif	Chromosome	Allele size range (bp)	Set for coloading
1	mSbCIR238	75	(AG)26	2	69-129	set 1
2	Xtxp21	185	(AG)18	4	145-227	set 1
3	Xtxp114	214	(AGG)8	3	196-245	set 1
4	mSbCIR262	214 and 222	(CATG) 3.25	7	208-446	set 1
5	SbAGB02	118	(AG)35	5	92-176	set 2
6	Xcup53	194	(TTTA)5	1	182-202	set 2
7	Xcup14	211	(AG)10	3	209-251	set 2
8	Xtxp320	290	(AAG)20	1	250-329	set 2
9	mSbCIR329	111 and 117	(AC)8.5	10	73-121	set 3
10	Xcup63	145	(GGATGC)4	2	127-163	set 3
11	Xgap72	191	(AG)16	9	168-229	set 3
12	Xtxp136	243	(GCA)5	10	238-246	set 3
13	mSbCIR300	110	(GT)9	5	74-118	set 4
14	Xtxp10	145	(CT)14	6	119-155	set 4
15	Xisep0310	204	(CCAAT)4	2	158-219	set 4
16	Xtxp278	249	(TTG)12	5	225-319	set 4
17	mSbCIR246	100	(CA)7.5	5	86-114	set 5
18	mSbCIR306	122	(GT)7	1	118-126	set 5
19	gpsb067	180	(GT)10	8	160-190	set 5
20	Xtxp57	251	(GT)21	9	213-285	set 5
21	mSbCIR248	89	(GT)7.5	10	79-111	set 6
22	mSbCIR283	143	(CT)8(GT)8.5	7	11-157	set 6
23	Xtxp265	213	(GAA)19	9	163-246	set 6
24	gpsb123	296	(CA)7+(GT)5	8	284-304	set 6
25	mSbCIR223	116	(AC)6	2	101-124	set 7
26	Xtxp141	163	(GA)23	7	126-175	set 7
27	Xcup61	198	(CAG)7	3	189-204	set 7
28	Xtxp145	236 and 238	(AG)22	9	204-278	set 7
29	Xtxp40	138	(GGA)7	5	108-144	set 8
30	Xtxp12	193	(CT)22	4	143-215	set 8
31	mSbCIR276	230	(AC)9	3	221-252	set 8
32	Xgap206	126	(AC)13/(AG)20	6	86-164	set 9
33	Xtxp15	215	(TC)16	10	197-273	set 9
34	Xtxp321	206	(GT)4+(AT)6+(CT)21	8	180-252	set 9
35	Xtxp273	196	(TTG)20	8	148-243	set 9
36	Xgap84	183	(AG)14	2	171-235	set 10
37	Xcup02	198	(GCA)6	6	186-216	set 10
38	mSbCIR240	112	(TG)9	8	102-180	set 10
39	mSbCIR286	128	(AC)9	1	104-150	set 10

PCR conditions for each of the 39 SSR markers for bulk samples and 20 for individuals' samples were optimized and PCR reactions were set in 10-µl volumes in

384-well PCR plates (ABGene, Rochester, N.Y.) manually. Each PCR reaction contained 0.2pmol of primer, 1-2 m*M* $mgcl_2$, 0.1-0.2 m*M* dNTP, 0.2U Amplitaq Gold polymerase (PE-Applied Biosystems) and 1x PCR buffer (PE-Applied Biosystems). Temperature cycling was carried out using the GeneAmp PCR systems 9600 (PE-Applied Biosystems) and touch-down PCR amplification: One 15-min denaturation cycle, followed first by ten cycles of 94°C for 10s, 61°C for 20 s (ramp of 1 per cycle) and 72°C for 30 s, then by 31cycles of 94°C for 10s, 54°C for 20 s and 72°C for 30 s. After completion of the 31 cycles, a final extension of 20 minutes at 72°C was included to minimize the +A overhang.

3.2.4 Detection of PCR products.

There are two common methods for manual analysis of fragments based on the sizes separation matrix used either agarose or denaturing/non-denaturing polyacrylamide gel electrophoresis (PAGE). Separated fragments are detected either by ethidium bromide or silver staining of the gels. Accurate sizing is difficult with both agarose and polyacryalamide gels and these matrices do not allow resolution to within a single base pair unit. In addition, mobility is affected by sequence composition so the repeat unit confounds migration of complementary strands in a gel based system. For instance CA strands moves faster on denaturing polyacrylamide gels than GT strands and this can result in two bands instead of one for the same allele (Ghebru *et al*., 2002). In this study the separated fragment were detected by a gel stained with ethidium bromide.

After PCR, 16 **samples were chosen randomly from the 384 well plates for electrophoresis. To check on the amplification of each SSR 3µl of the PCR**

products were run on 2% agarose gel. Standard 2% (w/v) agarose gel was prepared by dissolving 2g agarose in 100ml 1XTBE electrophoresis buffer (0.04M Tris-Borate, 0.001M EDTA, pH 8.0). The mixture was heated in a hot plate or microwave to allow the agarose to dissolve and form a gel. The gel was allowed to cool to about 20°C before adding 2ul of ethidium bromide (10mg/ml). The gel was then poured into a horizontal gel tray fitted with appropriate combs. After about 40 minutes of gel polymerization, the combs were carefully removed and the tray immersed in an electrophoresis tank containing electrophoresis buffer (1% x TBE). About 3µl of each of the randomly chosen PCR product was mixed with 2µl of loading dye and then loaded into separate lanes (slots) of the submersed agarose gel. The samples were run alongside 1.0µl 50 bp DNA step ladder at 100 volts for 45 minutes. After the run the gel was photographed under UV light. Each amplified SSR fragment was visualized as a distinct band.

3.2.5 Capillary electrophoresis

Genotyping was carried out by capillary electrophoresis using the ABI PRISM 3730 ABI **and Prism 3130** (Applied Biosystems), a fluorescent based capillary detection system that uses polymer as the separation matrix. This facilitated the accurate sizing of the microsatellite allele to within ± 0.3 base pairs (Buhariwalla and Crouch 2004). PCR products were co-loaded post-PCR based on dye label, fragment size and fluorescence to reduce the unit cost of high throughput genotyping. The amount of labelled PCR products was 0.75-1µl (depending on the intensity of the bands on agarose gel) were loaded mixed with HiDi genetic analysis grade formamide (PE-

Applied Biosystems) and ROX-labelled Gene Scan™ 500 LIZ-3730® size standard (PE-Applied Biosystems). For 1 ml of HiDi 12 µl Gene Scan were added and mixed it vigorously and 9.25µl of the mixture was combined with the PCR product with 3-4 primers labelled with different dyes. The mixtures were denatured for 5 minutes at 72ºC the plate was removed from the PCR machine and covered with ice for few minutes before sent to PSeGOli (Sequencing, Genotyping and Oligosynthesis Unit) covered with aluminium foil to avoid light and air. Before the sample was sent to PSeGOli the sample sheet was prepared based on plate loading sequence. T**he amplified DNA fragments were screened by capillary electrophoresis on the ABI Prism 3130 and ABI 3730 genetic analyzer (Applied Biosystems).** The peaks were sized and the alleles called using genotyper, genemapper softwares and the internal ROX GS500LIZ-3730 size standard. This system has the advantages of automated filling of capillaries, automated sample loading and rapid electrophoresis (Buhariwalla and Crouch, 2004). To verify the repeatability of each PCR and each capillary electrophoresis run, a control sample (accession BTx623, A, B, and C) were included during the PCR of each SSR marker and during each capillary electrophoresis run. Allelobin software was used for checking the quality of all markers.

3.2.6 Fragment analysis

The fragments **were analysed using the Genemapper V3.7 software (Applied Biosystems). The Genemapper software performs size calling which includes peak detection and fragment size matching. Bins which represent a fragment size or base pair range and dye colour that define an allele, are constructed from reference data. Algorithms determine if peaks represents alleles. When a peak**

from a data sample matches the location of a bin, the software makes an allele call. Alleles are automatically assigned allele calls based on the bin definitions. Once complete the genemapper software represents the results including the quality values .The results were stored in the genemapper data base.

3.3 Statistical data analysis

All SSR markers showed high reproducibility, with high consistency in the amplified product between the PCR and ABI runs of the controls, BTx623, A, B, and C. Therefore, all 39 markers were included in the analysis.

The in**formativeness of each primer was deduced using the polymorphic information content (PIC) as described by Weir (1996)**

$$PIC=1- \Sigma Pi^2$$

Where Pi is the frequency of the i^{th} allele.

PIC values give the information that each marker impacts to the study, which is the measure of the usefulness of each marker in distinguishing one individual from another.

PIC values are affected by the number and frequency of alleles in the population under study. Genetic diversity was measured in terms of number of alleles per locus and Nei's unbiased estimate of gene diversity (H) (Nei, 1987). The differences in value between H and observed heterozygosity (direct count) can provide an indication of

deviations from random mating in relation to Hardy-Weinberg (H-W) equilibrium (Weir, 1996). Distance matrices were calculated on the basis of Rogers modified distance (Wright, 1978) between regions and between countries. The distance matrices were subjected to Unweighted Paired Group Method using Arithmetic averages (UPGMA) cluster analysis. Distance matrices and clustering were determined using 'Tools for population Genetic Analysis (TFGPA software, Version 1.3 Miller, 1997 (htt//bioweb.usu.edu/mpmbio/ index.html). Bootstrap resampling (n=1000) was performed to test the robustness of the dendrogram topology.

3.4 Agro-morphological characterization

3.4.1The protocol for characterization by phenotyping

Measurement of phenotyping variance in field trials is a common and traditional approach to examine the genetic differences among genotypes. In this study the germplasm phenotyping was based only on characterization rather than evaluation. 196 accessions comprising landraces and farmers varieties from different geographic origins of Tanzania were collected from the National gene bank where are conserved for future use.

3.4.2 Experimental and plot design

Each batch of 196 accessions was replicated in two different sites representing different agro-climatic zones in this case Ilonga and Hombolo. In each site 3 replicates per accession were used. The experimental design was a 14x14 balanced lattice with 196 genotypes. To avoid border effect, three short rows of 3 m each were used per accession. The plot size was 3m x 1.5m with 30 cm spacing between plants and 75 cm

spacing between rows. Fertilizer N, P_2O_5 and K_2O were applied at the rate of 20g per plant. Hand weeding and watering were practiced when necessary.

3.4.3 Selection of plants and scoring procedures

Scoring of phenotypic characters was done in the middle row and the two outer rows were considered as borders (net plot size was 3m x 0.75m). The Bioversity International (formally IPGRI)/ICRISAT descriptor lists (1993) with some amendments were used score morphological characters (Appendix 2). Characters denoted by (0) and (+) were substituted with numerical descriptors (1) and (2) respectively. Three plants per accession were used for scoring and the mean was obtained (for quantitative characters). For qualitative characters, the numerical codes provided in the descriptor list were used instead of the actual characters. For intra-accession variation, codes of different characters observed were recorded as well as the visual assessment of the dominant character noted as well. The three plants scored per row were selected randomly and tagged just before flowering to avoid any bias. A blank space was used instead of 0 when scoring of character is not appropriate or applicable. Descriptors used in the study contained the qualitative and quantitative traits. Seed yields (the weight of 100 seeds measured in grams per accession) were recorded per accession for all the replications and the means were calculated. 100 seed weight was done at the physiological maturity when the seed contained 12% moisture content. Other yield components consisted of the ear length (from the base of head to tip) and harvest index (ratio of grain weight to total above- ground weight), grain sub-coat taken by scratching the outer seed cover and scored as (1) absent and (2) for present. The number of grain per panicle where three heads were used for scoring the

character and the mean was calculated. The magnifying glass was used to see clearly and score the features of grain plumpness and the grain form. For endosperm texture, colour and type the seed was dissected longitudinally and the magnified glass was used. The primary agronomic traits scored were plant height (taken from the ground to the tip of panicle of the main stalk), days of flowering (50% of plants in the plot were heading), dry weight, stalk juiciness (done at the fourth internode at maturity), plant colour (done at the flowering stage), juice flavor, leaf mid-rib colour (assessed at the plant flowering stage on any of the fully bloomed leaves and colour chart used to obtain the right colour), waxy bloom (the character was assessed on the on the stalk at the flowering). The descriptors that are associated with the inflorescence were inflorescence compactness and shape (the score was based on the shapes from descriptors), inflorescence exsertion (measured in cm from the base of the flag leaf of head), inflorescence length (made in the middle of the panicle with the ruler or tape measure in cm underneath the head) and width.

3.5 Data analysis

The data were analyzed using the GenStat (Discovery edition 2 product of VSN international) based on the means from the ANOVA output or variable that are significant different. Data were converted from the excel sheet to tab delimited. The Euclidean distance method (Kaufman and Rousseau, 1990) was used to construct distance matrices. Cluster analysis was performed using the genetic matrices generated by the Euclidean distance method to reveal patterns of genetic relationships among genotypes. Dendrograms were constructed using UPGMA clustering (Hintze, 2000).

The co-phetic correlation (Kaufman and Rousseau, 1990, Hintze, 2000) for each dendrogram was computed as a measure of goodness of fit' for each dendrogram.

CHAPTER FOUR

4.0 RESULTS AND DISCUSSION

4.1Genetic diversity and heterozygosity estimates from bulk samples

The diversity present in Tanzanian sorghum accessions was initially suggested by their morphological variation, their different growth properties, their varied uses and geographical positions. From this study it is indicated that SSR markers were highly polymorphic with an average of nine alleles per primers (Table 3). An analysis of genetic diversity among sorghum accessions indicated that the genetic distances calculated from SSR data were highly correlated with distances based on geographic origin. As reported in a number of genetic-diversity studies on other species and populations, SSR loci were able to uniquely identify each of our Tanzanian accessions. The cases observed in this study revealed that the numbers of alleles were much higher and usually across a wider size range (2-25) than those previously reported in studies of the same loci in other sorghum varieties (Brown *et al*., 1996, Dean *et al*., 1999, Djé *et al.*, 2000). Although allelic information may be inconclusive because same size PCR products can be seen for different alleles (Schlotterer, 1998), the number of observed alleles for most of the loci was also greater in this study (Table 3). The 39 SSR evaluated in this study provided a fair coverage of the genome and, were inherited in a Mendelian fashion. Each of the 39 primers that were employed in this study gave amplification products in all of the 200 accessions. The final data set therefore, was based on results from 39 SSR. As an indication of polymorphism, the number of

alleles and their frequency was analyzed. No null alleles were observed, a summary of the number of alleles, diversity indices, and average heterozygosity value of each locus is presented in (Table 3). The number of alleles per locus ranged from two (X*cup* 61, X*txp*114 and X*txp* 136) to 25 (*Xgap* 206) and the average number of alleles per locus was 9. The overall gene diversity in the entire sample was moderate (H_T=0.59) with two thirds of the diversity due to differentiation among accessions, but the highest gene diversity H_T=0.92 (X*txp* 265) and the lowest 0.078 (X*isep*0310). Generally, all the SSR markers were polymorphic, averaging nine allele per locus with diversity indices of 0.0892 or higher for most of the loci. The observed heterozygosity among the Tanzania accessions ranged from 0.1 % (*mSbCIR 238*) to 17 % (*Xxtp273*) with an average of 10.9% per locus (Table 2). The polymorphism information content (PIC) values for SSR loci ranged from 0.077 (Xisep0310) to 0.91(Xtxp265) with very high discriminative and a mean value of 0.55. The PIC provides estimates of the discrimatory power of a locus, or loci by taking into account, not only the number of allele that are expressed, but also relative frequencies of those alleles. PIC value ranges from zero (monomorphic) to one (very highly discriminative, with many alleles each equal and low frequency). As expected, the dinucleotide repeat-containing SSRs were, generally, more variable than those with longer motifs. Moreover, X*gap* and X*txp* SSRs usually had more alleles than X*cup* loci. This finding was probably due to differing SSR origins. The X*gap* and X*txp* markers were isolated from either small-insert genomic libraries (Brown *et al.*, 1996, Kong *et al*., 2000) or bacterial artificial chromosome

(BAC)-end sequences (Bhattramakki *et al.*, 2000). These loci, therefore, were more likely to include noncoding regions than the *Xcup* SSRs that were developed from low-copy RFLP probe sequences located primarily near or in genes (Schloss *et al.*, 2002).

Table 3: A summary of the number of alleles, diversity indices, major allele frequency and average heterozygosity value of each locus

Marker	Major Allele Frequency	Availability	Allele No.	Gene Diversity	Hetero-zygosity	PIC	Quality Index	Allelic Drift
gpsb067	0.4747	0.9851	7	0.6586	0.1465	0.6021	0.2391	0.01
gpsb123	0.7186	0.9900	6	0.4613	0.1206	0.4379	0.1956	0.01
mSbCIR223	0.4625	0.9950	6	0.6206	0.1300	0.5437	0.3748	0.12
mSbCIR238	0.2382	0.9502	18	0.8852	0.0157	0.8759	0.3014	-0.01
mSbCIR240	0.7577	0.9751	7	0.3875	0.0867	0.3433	0.1074	0.02
mSbCIR246	0.8945	0.9900	5	0.1951	0.0854	0.1873	0.1375	0.01
mSbCIR248	0.7632	0.9453	3	0.3855	0.1053	0.3479	0.1101	0.00
mSbCIR262	0.6154	0.9701	5	0.5600	0.1128	0.5123	0.4249	-0.14
mSbCIR276	0.6103	0.9701	4	0.4963	0.1179	0.3984	0.1043	0.00
mSbCIR283	0.2437	0.9900	13	0.8402	0.1558	0.8213	0.5144	-0.02
mSbCIR286	0.6742	0.9851	10	0.5117	0.0758	0.4806	0.1626	0.01
mSbCIR300	0.5455	0.9851	4	0.6003	0.1061	0.5343	0.1611	0.03
mSbCIR306	0.6005	0.9900	4	0.5265	0.1005	0.4440	0.1741	0.02
mSbCIR329	0.5536	0.9751	4	0.5892	0.0969	0.5223	0.092	-0.01
SbAGB02	0.5075	0.9900	9	0.6730	0.0804	0.6346	0.1669	0.01
Xcup02	0.4797	0.9801	5	0.6416	0.1472	0.5764	0.1117	-0.06
Xcup14	0.7386	0.9801	11	0.4128	0.0863	0.3673	0.1904	-0.03
Xcup53	0.8864	0.9851	3	0.2052	0.0657	0.1909	0.1646	0.08
Xcup61	0.5306	0.9751	2	0.4981	0.1735	0.3741	0.0638	-0.10
Xcup63	0.8401	0.9801	3	0.2809	0.0558	0.2620	0.0597	-0.01
Xgap206	0.1231	0.9900	25	0.9300	0.0553	0.9257	0.3547	0.01
Xgap72	0.4949	0.9851	8	0.6724	0.1414	0.6286	0.157	0.00
Xgap84	0.3325	0.9950	19	0.7738	0.1800	0.7438	0.1175	0.00
Xisep0310	0.9598	0.9900	6	0.0782	0.0201	0.0771	0.0915	0.00
Xtxp10	0.3781	1.0000	11	0.7395	0.1592	0.6992	0.2866	0.00
Xtxp114	0.5459	0.9751	2	0.4958	0.1429	0.3729	0.0602	-0.05
Xtxp12	0.2247	0.9851	23	0.8943	0.1313	0.8862	0.122	0.00
Xtxp136	0.5178	0.9801	2	0.4994	0.1523	0.3747	0.0941	-0.04
Xtxp141	0.2667	0.9701	15	0.8634	0.1128	0.8508	0.0892	0.00
Xtxp145	0.2969	0.9552	16	0.8653	0.0417	0.8552	0.376	-0.02
Xtxp15	0.3538	0.9701	12	0.7840	0.1282	0.7558	0.1281	0.00
Xtxp21	0.3665	0.9502	18	0.7442	0.1099	0.7071	0.2802	-0.03
Xtxp265	0.1392	0.9652	19	0.9197	0.1031	0.9141	0.2931	-0.01
Xtxp273	0.5457	0.9801	12	0.6527	0.1777	0.6222	0.1293	0.00
Xtxp278	0.9436	0.9701	5	0.1086	0.0308	0.1067	0.0639	0.00
Xtxp320	0.2436	0.9701	12	0.8445	0.2051	0.8259	0.1783	-0.01
Xtxp321	0.2236	0.9900	20	0.8781	0.1608	0.8673	0.1484	0.00
Xtxp40	0.9385	0.9701	5	0.1179	0.0359	0.1154	0.0941	0.00
Xtxp57	0.3975	0.9950	11	0.7697	0.1100	0.7431	0.2968	0.00

Mean	**0.5238**	**0.9788**	**9**	**0.5913**	**0.1093**	**0.5520**

4.1.1 Characterization of microsatellite loci

As in Figure 2, it demonstrated a typical electrophoregram picks obtained with multiple samples loading. Optimal separation between the five loci was obtained by combining differences in fragment sizes and in fluorescent dyes. The total number of putative alleles at each locus and the size ranges of these alleles as observed in this study are given in (Table 3). It appears that all 39 loci scored in this study are highly polymorphic with many alleles (from 2–25) putative alleles) and a wide range of product sizes (Table 2). For most loci the size range of PCR products obtained here is substantially wider.

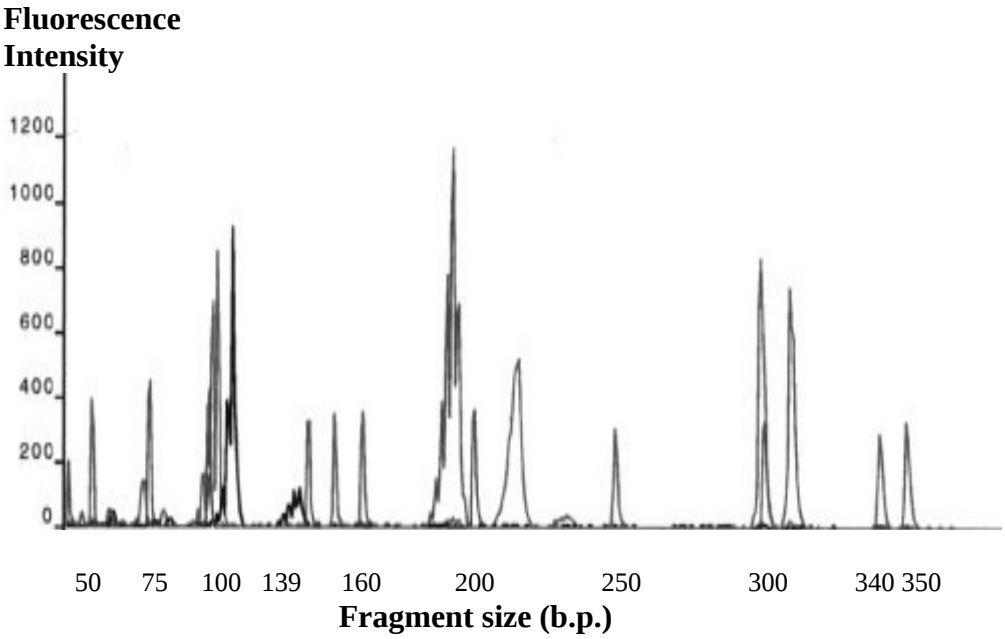

Figure 2: Typical electrophoregram produced by the automated DNA sequencer after multiple samples loading.

4.1.2 Genetic relationships among accessions

The neighbor-joining phenogram depicting genetic relation among *Sorghum bicolor* accessions shows two major groups and a minor group. The minor group consists of four accessions which suggested that these accessions originated from the same geographical area.

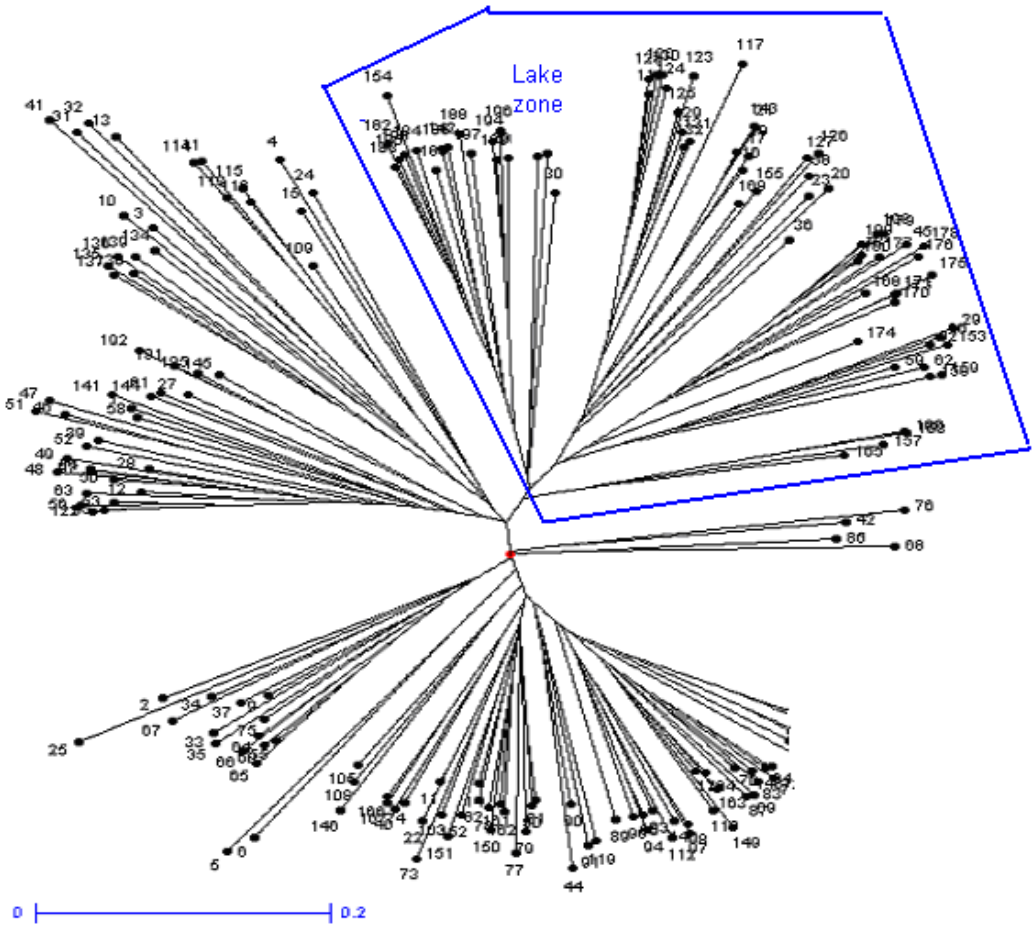

Figure 3: Neighbor-joining phenogram depicting genetic relationship among 200 *Sorghum bicolor* accessions.

Neighbor-joining analysis indicated that sorghums accessions from the lake zone generally formed a coherent group that branched further to small clusters Figure 3. Moreover, most of the accessions are from the same geographic region (Lake Zone)

tended to be genetically more similar to each other than those from more distant locations (accessions from Lindi, 170, 171 and 174 etc) indicating the presence of population structures that are much very much similar in that region. Results from independent diversity studies can also lend to support to the selection hypothesis. Race bicolor resembles spontaneous weedy sorghums but lacks the ability to disperse seeds naturally (i.e., seeds do not shatter) (de Wet, 1978). Because it's long, clasping glumes, elongated seed, and open panicles are considered to be primitive characters; bicolor is thought to be the race most closely related to wild sorghums (Harlan and de Wet, 1972) and Figure 8(1)**.** It should be emphasized that branch support for groupings of cultivated accessions was low (most bootstrap values ≤ 20), indicating a history of gene flow among the various races and/or recent common ancestry. Dendrogram generated from distance matrices of these sorghums accessions (Fig.4) showed three major clusters and two minor clusters, with one major cluster largely composed of accessions from the Lake Zone and central Tanzania. Beyond this observation, major clustering was based by region of collection/origin. The accessions from other regions such as Lindi, Nachingwea and Newala clustered in the same clusters as Lake Zone found group in all trees. Genetic-distance matrix data from all the comparisons of the 200 total accessions were used to visualize the relationships among the accessions. Distinct clustering of the Lake zone sorghum accessions are seen on the top two major clusters and those from the other areas mixed in the bottom cluster. Mixed groups of both the Lake zone accessions and other sorghums can also be seen in the middle clusters.

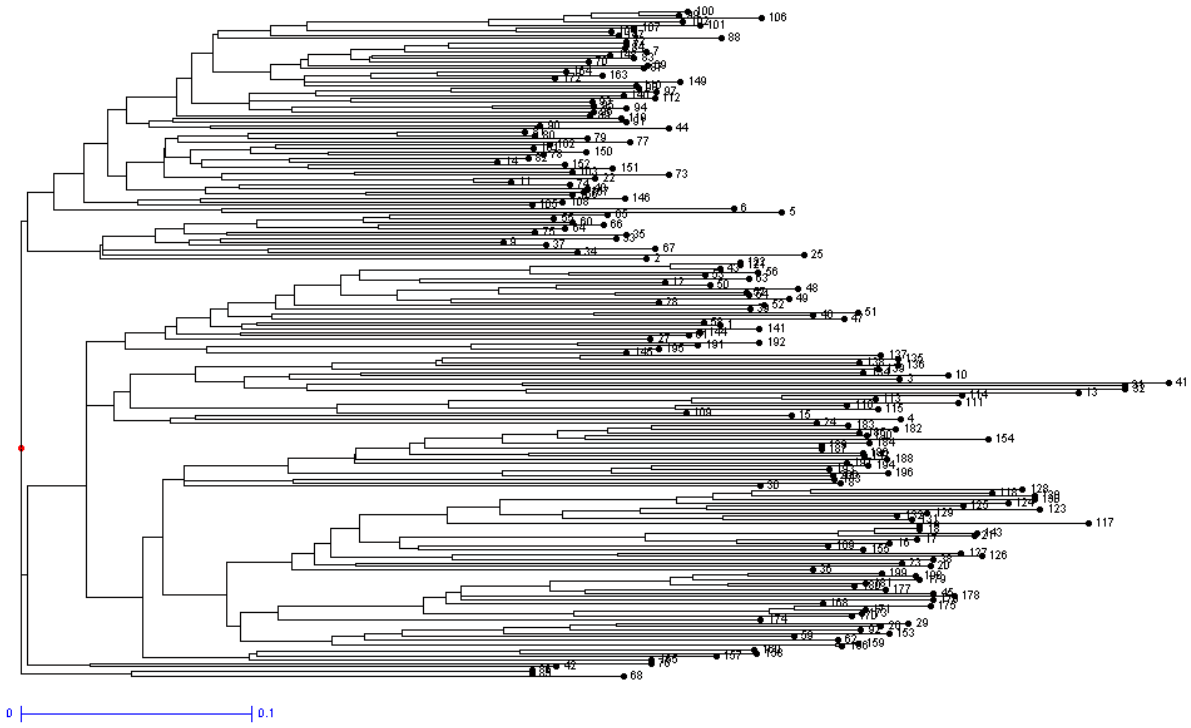

Figure 4: Dendrogram showing relationship among 200 Tanzanian *Sorghum bicolor* accessions

When closely examined, the smaller clusters observed consist of a group of lake zone and central Tanzania. These accessions can only be found within the clusters for example clusters bearing accessions (76, 42, 86 and 68). Most of the accessions given the same name or similar identification characters by farmers were grouped together, or more or less agree with their classification. The other prominent clustering of Tanzanian sorghum are group from Southern Tanzania Figures 3, 4, 5 and a diverse group of central Tanzania accessions. This suggests that genotypes from heterogeneous accessions always clustered together, either with their cohorts or with individuals from closely related accessions.

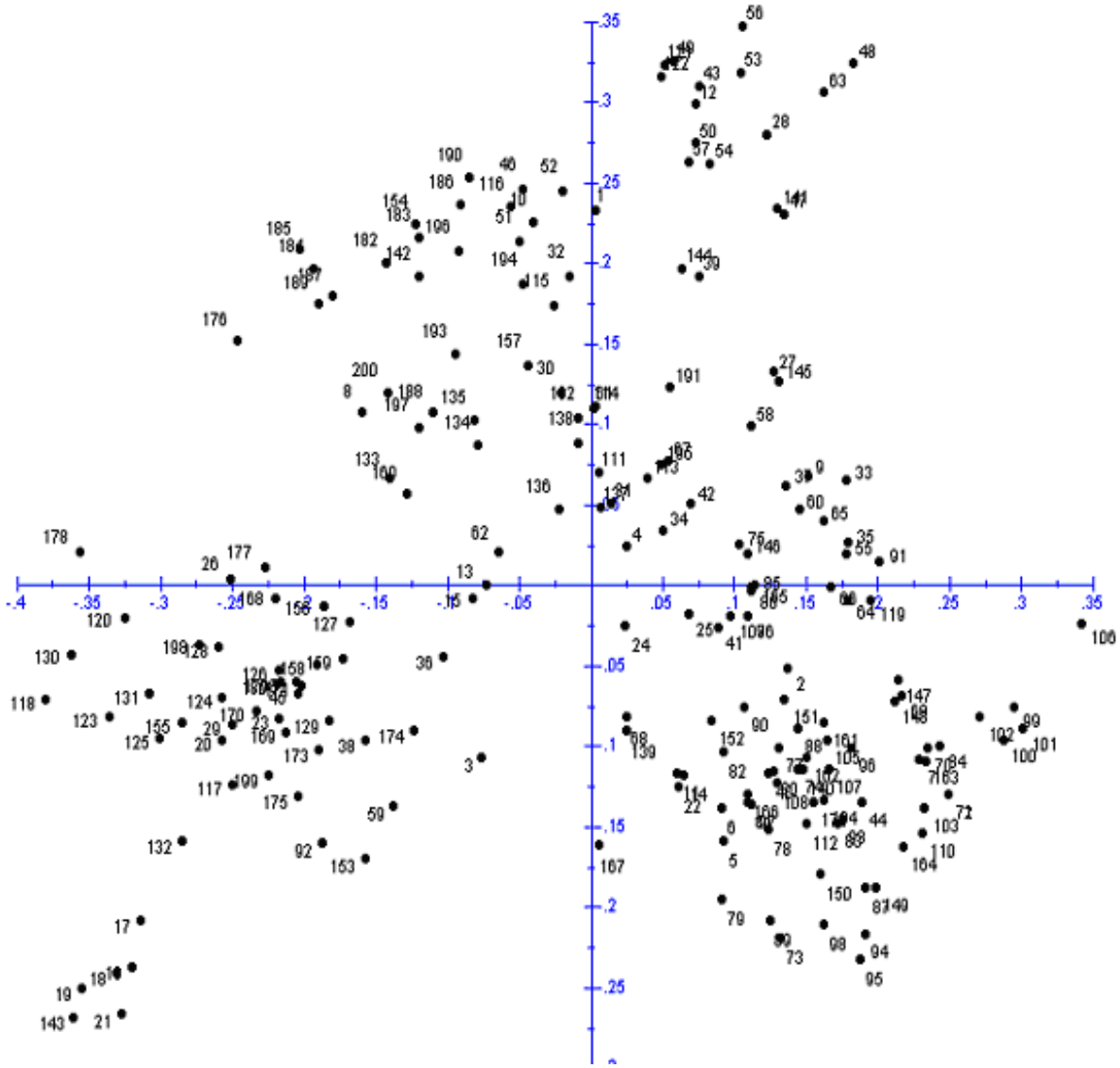

Figure 5: Scatter plot of sorghum accessions against the two first principal coordinates computed from the genetic distance matrix

4.1.3 Genetic diversity and heterozygosity estimates from individual samples

For this study 20 accessions were selected within the material used in this study from different areas for the intra-population diversity study as in Table 1. Three individual plants were selected per accession a total number of sixty. The areas with low altitude are those areas along the southern coastal part of Tanzania (Mtwara and Nachingwea).The 18 primers were used to amplify the DNA from 60 sorghum individual plants derived from 20 selected accessions. These accessions were selected to represent area where sorghum is grown in large scale and used as a staple crop. Statistics of genetic diversity within and among accessions are given in Table 4. High genetic polymorphism for microsatellite loci was observed within accessions as, on average, more than two-thirds of the loci were polymorphic for a given accession (mean percentage of polymorphic loci, p=67.0%), more than two alleles per locus per accession were found on average (A= 2.0) and the probability that two randomly sampled alleles in a given accession are different is higher than a quarter (He = 0.29). It was shown that accessions with either low or very high polymorphism were rather rare within the sample, and none were completely fixed at some loci. Despite this high level of genetic variation within accessions of cultivated sorghum, the average proportion of heterozygote individuals was found to be low (Ho=0.31), in agreement with the high level of inbreeding noted. The overall gene diversity in the entire sample was found to be very high (Ht=0.90), with two-thirds of the diversity due to differentiation among accessions.

Table 4: Descriptive statistics (by population) of the genetic diversity within and among 20 accessions of sorghum used in the intra-population study classified by geographic origin

Number	Access. Number	*n*	*p*	*A*	*Ap*	*He*	*Ho*	*f*
1	TZA 3493	2.89	0.72	2.44	3.00	0.49	0.47	0.047
2	TZA 4166	2.89	0.61	1.77	2.27	0.37	0.46	-0.315
3	TZA 4247	2.89	0.94	2.83	2.94	0.66	0.33	0.561
4	TZA 3418	2.94	0.55	1.77	2.40	0.29	0.33	0.212
5	TZA 3436	3.00	0.83	2.27	2.53	0.50	0.24	0.466
6	TZA 3494	3.00	0.50	1.77	2.55	0.29	0.29	-0.266
7	TZA 4001	2.94	0.61	1.88	2.45	0.36	0.35	0.485
8	TZA 4044	2.83	0.83	2.33	2.60	0.52	0.21	0.664
9	TZA 4155	2.88	0.72	1.94	2.30	0.42	0.20	0.282
10	TZA 3904	2.94	0.55	1.72	2.30	0.32	0.32	0.166
11	TZA 3864	2.94	0.5	1.61	2.22	0.26	0.27	0.239
12	TZA 3616	2.88	0.55	1.77	2.40	0.33	0.21	0.110
13	TZA 3882	2.94	0.66	2.11	2.66	0.43	0.30	0.333
14	TZA 3943	2.83	0.72	1.94	2.30	0.45	0.31	-0.072
15	TZA 3835	2.94	0.72	2.11	2.53	0.44	0.47	0.518
16	TZA 3992	2.94	0.77	2.33	2.71	0.48	0.24	0.351
17	TZA 3994	2.94	0.66	1.77	2.16	0.37	0.33	0.099
18	TZA 3965	2.94	0.66	1.94	2.41	0.37	0.34	0.110
19	TZA 4162	2.83	0.66	1.94	2.41	0.39	0.34	0.235
20	TZA 3147	2.94	0.66	1.94	2.00	0.31	0.31	0.058
	Mean	**2.91**	**0.67**	**2.00**	**2.46**	**0.4**	**0.31**	**0.262**

Key: ***n***= Mean sample size over all the loci; ***P***= proportion of polymorphic loci; **A**= mean number of alleles per locus; ***Ap***= alleles per polymorphic locus; ***He***= expected heterozygosity; ***Ho***= observed heterozygosity; ***f***= inbreeding co-efficient

The *F*-statistics over the 20 accessions are given for each microsatellite locus as well as their averages over loci (Table 5). Altogether very little variation was observed among loci with respect to the values of these *F*-statistics. The overall inbreeding

coefficient, F_{it}, has a mean of 0.52, showing that the sorghum germplasm collection is highly inbred overall. This inbreeding is due in part to inbreeding within individual accessions, with a mean value F_{is} = 0.55 and in part to strong genetic differentiation among accessions as shown by a mean fixation index of F_{st} = 0.49.

Table 5: Descriptive *F*-statistics (by locus) over the 20 accessions of sorghum for each microsatellite locus; F_{it}, overall inbreeding coefficient; F_{st}, fixation index; F_{is}, average inbreeding coefficient within accessions

Locus	*p*	*A*	*Ap*	*He*	*Ho*	*f*	F_{is}	F_{it}	F_{st}	*Nm*
gpsb123	1	4.0	4.0	0.48	0.10	0.79	0.34	0.77	0.65	0.13
mSbCIR223	1	5.0	5.0	0.54	0.10	0.81	0.51	0.81	0.62	0.15
mSbCIR238	1	14	14	0.9	0.16	0.81	0.43	0.81	0.67	0.12
mSbCIR240	1	5.0	5.0	0.62	0.80	-0.27	-0.61	-0.28	0.20	0.95
mSbCIR283	1	8.0	8.0	0.74	0.00	0.10	1.00	1.00	0.75	0.07
mSbCIR306	1	5.0	5.0	0.43	0.11	0.73	0.49	0.73	0.46	0.28
SbAGB02	1	11	11	0.74	0.27	0.64	0.20	0.63	0.54	0.21
Xcup61	1	6.0	6.0	0.73	0.30	0.59	0.06	0.58	0.55	0.19
Xtxp12	1	10	10	0.86	0.25	0.71	0.42	0.70	0.48	0.26
Xtxp136	1	3.0	3.0	0.52	0.55	-0.06	-0.55	-0.06	0.31	0.54
Xtxp141	1	9.0	9.0	0.69	0.08	0.87	0.79	0.87	0.40	0.36
Xtxp265	1	17	17	0.91	0.08	0.38	-0.14	0.37	0.45	0.29
Xtxp278	1	3.0	3.0	0.45	0.56	-0.45	-0.67	-0.48	0.11	1.88
Xtxp321	1	11	11	0.80	0.65	0.66	0.41	0.69	0.48	0.26
Xtxp40	1	4.0	4.0	0.20	0.26	0.42	-0.44	0.41	0.59	0.16
Xtxp57	1	12	12	0.83	0.11	0.44	0.04	0.43	0.41	0.35
Sb4-72	1	8.0	8.0	0.85	0.46	0.06	-0.51	0.05	0.37	0.42
Sb6-84	1	7.0	7.0	0.78	0.80	0.84	0.49	0.84	0.70	0.10
Mean	**1**	**7.88**	**7.9**	**0.067**	**0.31**	**0.53**	**0.055**	**0.52**	**0.49**	**0.25**

4.1.4 Genetic relationships among 20 accessions and 60 individuals

From Figure 6, twenty neighbor-joining dendrograms were generated; the clustering of individuals was not similar. The majority of the individuals analyzed were genetically more similar but not identical as shown in Figure 6.

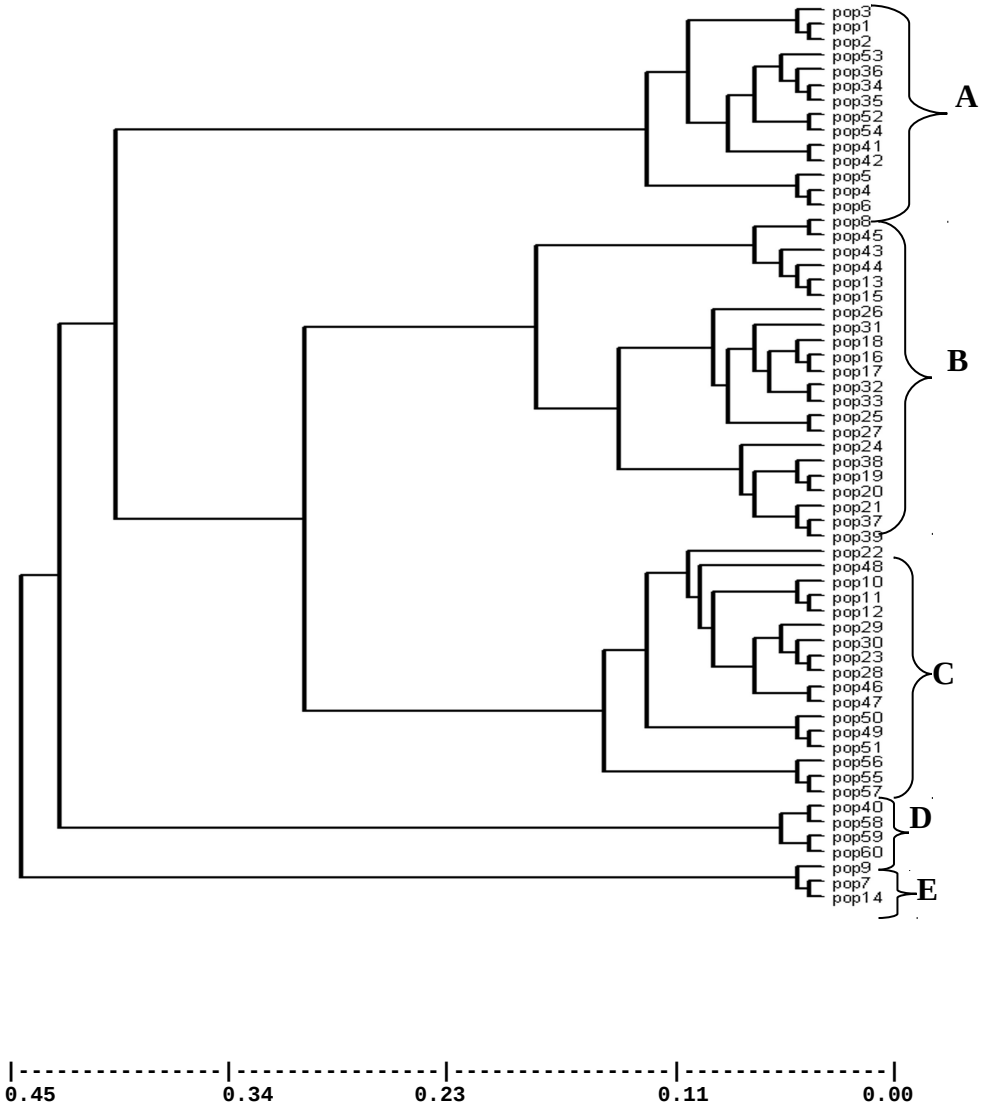

Figure 6: Dendrogram showing intra-population relationships among 60 individuals selected from sorghum 20 accessions

There were no genetically identical individuals but most of them clustered at the same cluster as for individuals from accessions (1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 14, 16, 17, 18 and 20). In Figure 6, dendrograms generated from distance matrices of individuals alone showed three major clusters (A, B and C) with two small clusters (D and E) largely composed of accessions from number 3 and 20 collected/originated from Kasulu and Bukoba.

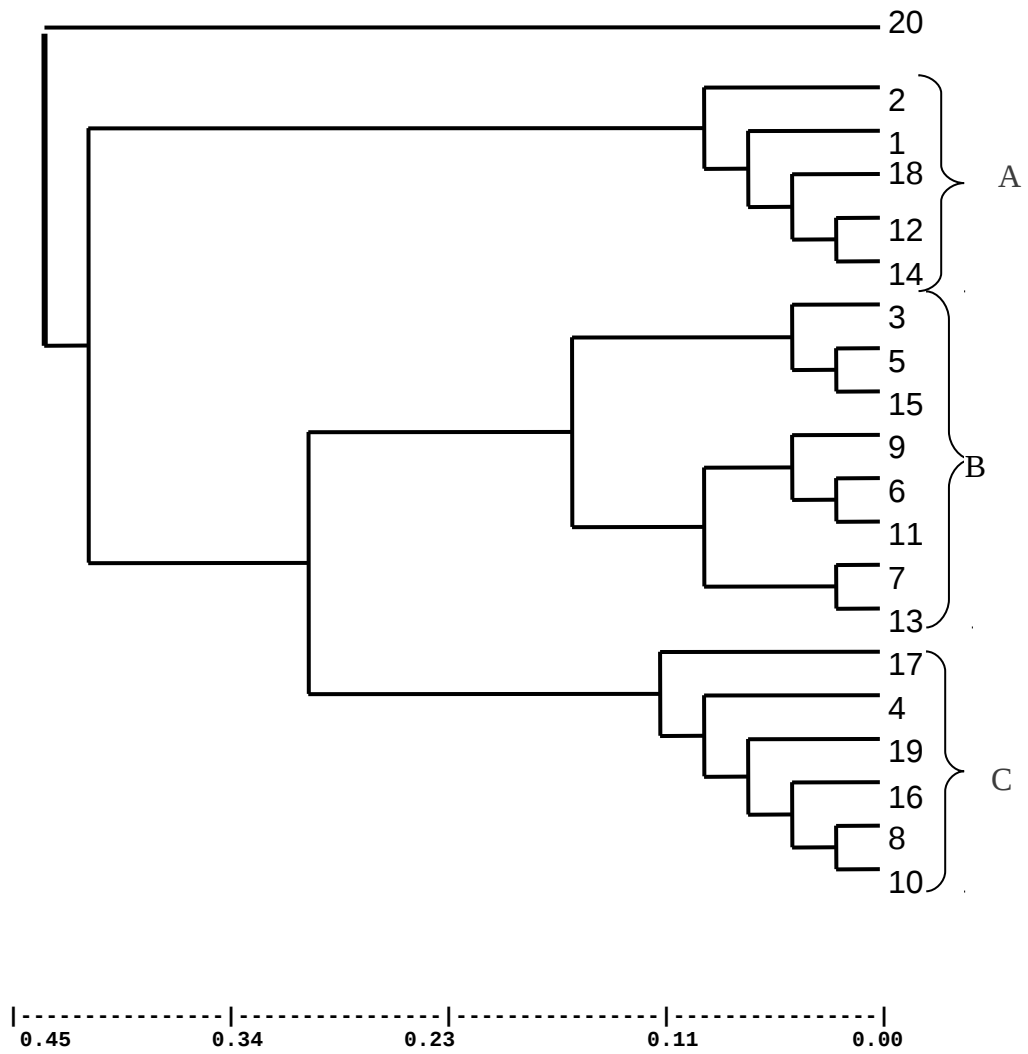

Figure 7: Dendrogram showing relationships among the 20 sorghum accessions.

Beyond this observation, major clustering by region of collection was not observed as in Figures 6 and 7. Figure 7 formed three major clusters (A, B and C), but accession number 20 from the western part of Tanzania the Kasulu district. This accession may have crossed from neighboring countries such as Rwanda, Burundi and Democratic Republic of Congo; this can be proved with the other future studies.

While the area of functional genomics is still in its infancy, we anticipate that the use of population diversity based approaches will allow the mining of germplasm collections and SSR diversity data for identifying interesting genomic regions. Certainly, molecular data from species with little population structure and intermediate levels of LD would be well suited for re-analysis. It must be stressed that while these approaches may be advantageous for identifying genomic regions that differ from the average observed in the genome, some of these departures may also result from non-equilibrium population history. Functional studies (e.g., mutant screening, genetic complementation, expression analysis, biochemical localization and characterization, etc) are still required to establish causation.

Therefore, this study provides a first detailed analysis and quantification of genetic diversity in Tanzanian sorghum accessions. The data also reaffirm the power of SSR markers to distinctly group closely related landraces. Several authors have indicated that SSR technology is highly cost-effective (Smith *et al.*, 2000) and that this technology could easily be employed in resource-poor countries like Tanzania. It could provide efficient and fast screening for both germplasm conservation and crop improvement. The efficiency of using germplasm as a genetic resource can be improved if genetic diversity information is available.

4.2 Field morphological characterization

The morphological characterization work was performed in two sites, all the quantitative and qualitative data were collected as per Bioversity International sorghum descriptor formerly International Plant Genetic Resources Institute (IPGRI). The

accessions performance at Ilonga plots was excellent; this is due to the fact that the amount of rainfall received at Ilonga was higher (1300 mm/annual) compared to Hombolo with an average rainfall of 650 mm/annual (Table 6). Also the soil fertility at Ilonga plot was much higher compared to Hombolo before the start of the experiment (Table 6). At Ilonga the average 100 seeds weight per accession was 3.7g compared to that of Hombolo 1.9g. The plants were taller at Ilonga with an average 300 cm while those from Hombolo were about 210 cm above the ground. The plants at Ilonga were able to show all plant characters unlike; at Hombolo where it was difficult to observe some of the phenotypic characters such as goose neck type of exsertion Figure 6(10). Therefore, the gene expression for the quantitative and qualitative data at Ilonga was higher than Hombolo. A higher degree of variation was observed in days to flowering for the accessions planted in Hombolo and Ilonga but shows a small difference for the accessions grown in the same area. For Ilonga the average days to flower for the accessions planted were 80 days compared to Hombolo which was 92 days this is due to the effect of genotype x environmental interaction. Theoretically, the phenotypic performance is the results of interaction between genotypic (particular assemblages of genes in each individual) and environmental factors (all non-genetic circumstance that influenced the phenotypic value (Falconer, 1970). High phenotypic variation was observed among accessions as in Figures 8, 9 and 10. These accessions were scattered along the dendrogram. Four regions (Southern Tanzania, Lake Victoria, Central Tanzania and Western Tanzania) were recognized as origins of these accessions Figure 9 A-E.

Table 6: Weather informations, physical and chemical properties of the soils at Hombolo and Ilonga agricultural research stations experimental plots

a)Ilonga					
Chemical properties	**%**	**Physical properties**	**%**	**Weather**	
pH (water)	4.56	Sand	68.0	Temperature ºC	27.5
OC	2.45	Silt	8.20	Rainfall (mm/yr)	1300
Total N	0.21	Clay	23.8		
Ca	0.89				
Mg	0.54				
K	0.78				
Na	1.62				
CEC Me/100g	2.18				
P (ppm)	1.03				
BS	25.5				
b) Hombolo					
Chemical properties	**%**	**Physical properties**	**%**	**Weather**	
pH (water)	5.81	Sand	72.0	Temperature ºC	30.5
OC	1.07	Silt	4.50	Rainfall (mm/yr)	650
Total N	0.13	Clay	23.5		
Ca	1.22				
Mg	0.41				
K	0.24				
Na	0.19				
CEC Me/100g	1.88				
P (ppm)	0.65				
BS	12.5				

4.2.1 Hierarchical cluster analysis

"Cluster analysis" refers to "a group of multivariate in 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" (Hair *et al*., 1995).

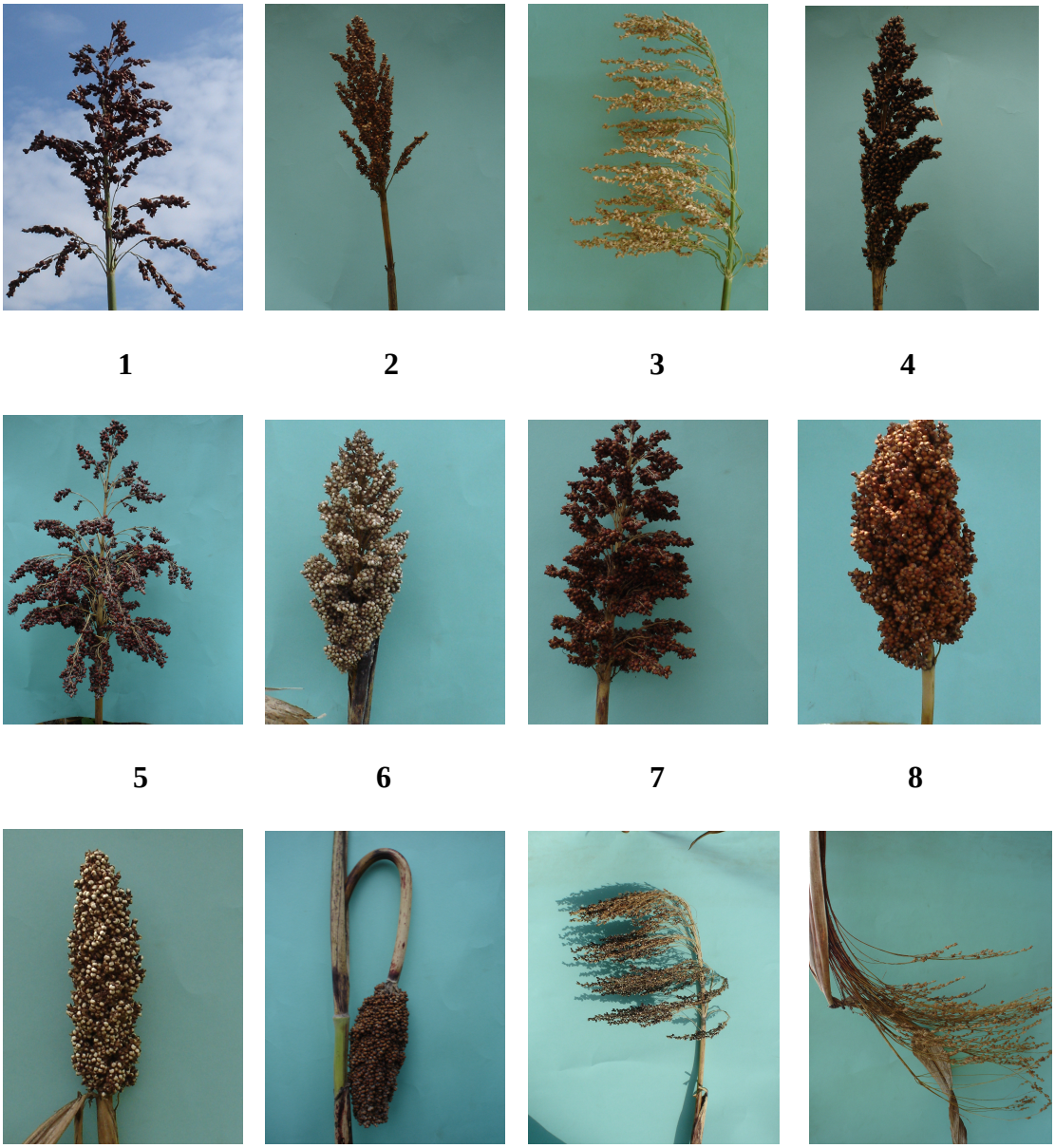

Figure 8: Inflorescence compactness and shape as example of morphological characters

Key 1. Very lax panicle (typical of wild sorghum) 2. Very loose erect primary branches 3. Very loose drooping primary branches 4. Loose erect primary branches 5. Loose drooping primary branches 6. Semi-loose erect primary branches 7. Semi-loose drooping primary branches 8. Semi-compact elliptic 9. Compact oval 10. Compact oval 11. Half broom corn 12. Broom corn

The resulting clusters of individuals should then exhibit high internal (within cluster) homogeneity and high external (between clusters) heterogeneity. Thus, if the

classification is successful, individuals within a cluster shall be closer when plotted geometrically and different clusters shall be further apart (Hair *et al*., 1995). Hierarchical clustering methods are more commonly employed in analysis of genetic diversity in crop species.

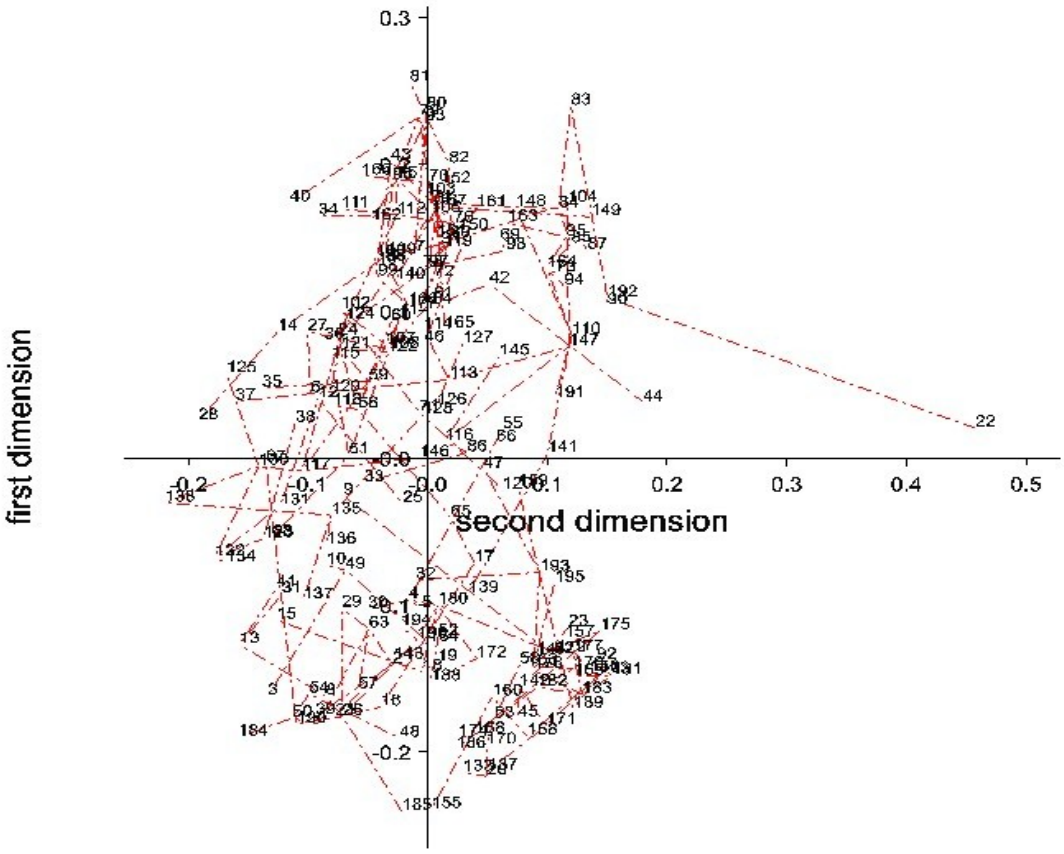

Figure 9: Minimum spanning scatter tree showing the highly loaded principal coordinates thus compares the groupings to the clusters as computed from morphological traits

The clustering among the sorghum accessions assessed using phenotypic characters showed that all accessions were distinctly placed in the respective positions in spinning

scatter tree Figure 9. In Figure 9 the accessions are classified at least into four main clusters and a small cluster.

The clusters are mainly dominated by the accessions from the lake zone as in Figure 3. In Figure 9**,** the clustering of the accessions showed a very interesting relationship with a closet within and between them that is beyond the rescaled cluster combine distance. These accessions as per passport data showing as coming from the same area that is from the lake zone.

.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

Data in this study demonstrate that accessions of Tanzanian sorghums contain a considerable genetic diversity as indicated by the observed higher number of alleles. The study provides a first detailed analysis and quantification of genetic diversity in Tanzanian sorghum accessions. The data also reaffirm the power of SSR markers to distinctly group closely related landraces. SSRs is a promising marker system especially for sorghum accessions, which are closely related, with medium level of polymorphism, and thus high level of genetic uniformity. Several authors have indicated that SSR technology is highly cost-effective (Smith *et al.*, 2000) and that this technology could easily be employed in resource-poor countries like Tanzania. It could provide efficient and fast screening for both germplasm conservation and crop improvement. It appears that microsatellite data pertaining to diversity within and among accessions of a crop species may be useful in devising strategies for managing highly structured germplasm collections. In this study it was found that molecular characterization in conjunction with phenotypic factors can provide reliable information for assessing, among other factors, the amount of genetic diversity, the structure of diversity in samples and populations, rate of genetic divergence among populations and the distribution of diversity in populations found in different locations. Therefore, the accessions inventory was compiled according to the results out come which can also, be used in future to compile the rest of the sorghum germplasm present at the National Plant Genetic Resources Centre (NPGRC). The

same method can be used to assess and characterize sorghum genetic resources available at the National Plant Genetic Resources Centre (NPGRC). In the study part of the sorghum germplasm from the National Plant Genetic Resources Centre (NPGRC) gene bank was characterized and the potential parental genotypes for mapping populations and mass assisted selection program were identified.

The efficiency of using plant genetic resources can be improved if genetic diversity information is available. The conservation activities range from collecting, compiling of an inventory and managing the germplasm through identifying genes that adds value to genetic resources. Finally, it seems that although all methods did not provide similar description of relationships between accessions, there existed some consistency in discriminating accessions which are closely related and the ones which were distantly related. All methods have advantages and disadvantages for practical applications under different circumstances. From this study, knowledge was obtained on genetic diversity using molecular markers which were not applied before in Tanzania.

5.2 Recommendations

Therefore, the study was of importance for gaining more knowledge for improving sorghum breeding programmes in Tanzania. Therefore, it is recommended that,

(i) Conventional breeding which has been used for sorghum breeding programmes needs to be supplemented with genetic mapping and transformation to identify and transfer specific genes in order to hasten progeny selection for increased yield and grain quality. Since morphological

traits do not cover entire genome, this has to be confirmed at DNA level where possible. Low levels of correlation existed between agronomical, morphological, combined agronomical and morphological and SSRs analysis encompassed the expression of genotype, environment and their interactions.

(ii) Because these materials are farmers' varieties, conservation strategies in Tanzania should focus on these resources. Many SSR loci will be significantly linked to important agronomic traits. Hence, Tanzanian sorghums deserve broader characterization at both molecular and agronomic levels, including the molecular mapping of important traits (Lin *et al.*, 1995, Peng *et al.*, 1999, Lan and Paterson, 2000) and the findings can be used in future crop improvement. Their use in genetic conservation should be encouraged, provided that successful strategies for improving the analysis throughput could be introduced (Donini *et al.*, 1998). Furthermore, molecular analysis should concentrate on the use of other DNA marker systems in order to generate more information such as the use of marker assisted selection technique.

(iii) As indicated in this study that qualitatively inherited morphological characteristics can be used to characterize varieties for collection and maintenance of germplasm and for parental selection through heterotic groups (groups with large distances between them) for improving local varieties. Since morphological traits are influenced by environment, it is recommended that experiment should be repeated for two more years at different sites and to be confirmed at a DNA level where possible.

(iv) Conventional and molecular breeding should be used in plant breeding and germplasm conservation in Tanzania. Future research should focus on comparing these methods in terms of feasibility, efficiency and accuracy by involving more tests over different environmental trials and years (for agronomical and morphological characteristics).

REFERENCES

Ahnert, D., Lee, M., Austin, D.F., Livini, C., Openshaw, S.J., Smith, J.S.C., Porter, P., and Dalton, G. (1996). Genetic diversity among elite sorghum inbred lines assessed with DNA markers and pedigree information. *Crop Sci.* 36:1385–1392.

Appa-Rao, S., Prasada-Rao, K.E., Mendesha, M.H., Gopal-Reddy, V. (1996). Morphological diversity in sorghum germplasm from India. *Gen. Res. Crop Evol.* 43: 559-567.

Barrett, B.A., Kidwell, K.K. (1998). AFLP-based genetic diversity assessment among wheat cultivars from the Pacific Northwest. *Crop Sci.* 38: 1261-1271.

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.*

Bowers, J.E. and Meredith, C.P. (1997). The percentage of a classic grape wine. Cabernet Sauvignon. *Nat. Genet.* 16: 84-87.

Brown, S.M., Hopkins M.S., Mitchell SE, Senior, M.L., Wang, T.Y., Duncan, R.R., Gonzalez-Candelas, F., Kresovich, S. (1996*)* Multiple methods for the identification of polymorphic simple repeats (SSRs) in sorghum (*Sorghum bicolor* (L) Moench). *Theor Appl Genet* 93:190–198.

Buhariwalla, H.K. and Crouch, J.H. (2004). *Optimization of marker screening protocol to assess the degree and distribution of genetic diversity in landraces of pigeon Pea*. John Wiley and Sons, Inc., New York. pp. 67 - 76

Burr, B. (1994). *Some concepts and new methods for molecular mapping in plants*. In: Philips RL, Vasil IK (Eds.) DNA-based Markers in Plants. Kluwer Academic Publishers, Dordrecht, the Netherlands, pp 1-7.

Caetano-Anollés, G. & P.M. Gresshoff, (1997). *DNA Markers: Protocols, Applications, and Overviews.* Willey-Liss, John Willey & Sons, Inc., Publication, New York. Pp. 45-78

Cervera, M.T., Cabezas, J.A., Sancha, J.C., Martínez de Toda, F., Martínez-Zapater, J.M. (1998). Application of AFLPs to the characterization of grapevine *Vitis vinifera* L. genetic resources. A case study with accessions from Rioja (Spain). *Theor. Appl. Genet.* 97: 51-59.

Chowdari, K.V., Venkatachalam S.R., Davierwala, A.P., Gupta, V.S., Ranjekar, P.K., Govila, O.P. (1998b). Hybrid performance and genetic distance as revealed by the (GATA) 4 microsatellite and RAPD markers in pearl millet. *Theor. Appl. Genet.* 97: 163-169.

Dahlberg, J.A., Hash, J., Kresovich, S., Maunder, A.B. and Gilbert, M. and Szewc-McFadden, B. (1997). *Application of multiple PCR and fluores- 1997. Sorghum and pearl millet genetic resources utilization.p 42–54. In* Proceedings of the International Conference on Genetic Improvement of Sorghum and Millet. SICNA, Lubbock, TX. pp.60-65

Dean, R.E., Dahlberg, J.A., Hopkins M.S., Mitchell S.E., Kresovich, S. (1999). Genetic redundancy and diversity among orange accessions in the US National Sorghum Collections as assessed with simple sequence repeat (SSR) markers. *Crop Sci* 39:1215-1221.

De-Bustos, A., Soler, C., Jouve, N. (1999). Analysis by PCR-based markers using designed primers to study relationships between species of *Hordeum* (Poaceae). *Genome* 42: 129-138.

De' Wet, J.M.J. (1978). Systematics and evolution of Sorghum Sect. *Sorghum* (grammineae). *AM J Bot* 65:477-484.

Djè, Y., Ater, M., Lefèbvre, C., Vekemans, X. (1998). Patterns of morphological and allozyme variation in sorghum landraces of Northwestern Morocco. *Gen. Res. Crop Evol.* 45: 541-548.

Djè Y., Heuertz, M., Lefèbvre, C., Vekemans, X. (2000). Assessment of genetic diversity within and among germplasm accessions in cultivated sorghum using microsatellite markers. *Theor. Appl. Genet.* 100 918-925.

Doggett, H. (2nd Eds.) (1988). Sorghum. John Wiley and Sons, Inc., New York. 87pp

Donini, P., Stephenson, P., Bryan, G.J., Koebner, R.M.D. (1998). The potential of microsatellites for high throughput genetic diversity assessment in wheat and barley. *Genet Res Crop Evol* 45:415–421.

Duncan, R.R., Bramel-Cox, P.J and F.R. Miller. (1991). *Contributions of introduced sorghum germplasm to hybrid development in the USA*. p. 69–102. *In* H.L. Shands and L.E. Weisner (Eds.) *Use of plant introductions in cultivar development*. CSSA Spec. Publ. 17. CSSA, Madison, WI.

Eberhart, S.A., Bramel-Cox P.J and Prasada Rao K.E. (1997). *Preserving genetic resources.* p. 25–41. In*: Proceedings of the International Allard. 1994. Extraordinarily polymorphic microsatellite DNA in Conference on Genetic Improvement of Sorghum and Millet,* SICNA, Lubbock, TX.

Falconer, D.S. (1970). *Introduction of Quantitative Genetics*. Longman Group, London. 112-134.

FAO. (2004). Production yearbook 2002 no. 56. *FAO Statistic Series* no. 176, Rome

Frankel, O.H. and E. Bennet (Eds)(1970). *Genetic Resources on plants-their exploration and conservation.* Blackwell Scientific Publication, Oxford (pp210-228).

Gebhardt, C., Ballvora, A., Walkemeier, B., Oberhagemann, P. & Schuler, K. (2004). Assessing genetic potential in germplasm collections of crop plants by marker-trait association: a case study for potatoes with quantitative variation of resistance to late blight and maturity type. *Mol. Breeding* 13:93-102.

Ghebru, B., Schmidt, R.J., Bennetzen, J.L. (2002). Genetic diversity of Eritrean sorghum landraces assessed with simple sequence repeats (SSR) markers. *Theor Appl Genet* 105:229-236.

Grenier, C., Bramel-Cox, P.J., Noirot, M., Prasada-Rao, K.A., Hamon, P. (2000a*).* Assessment of Genetic diversity in three subset constituted from the ICRISAT sorghum collections using random vs.non-random sampling procedures. A. Using morpho-agronomical and passport data. *Theo Appl Genet* 101:190-196.

Grenier, C., Deu, M., Kresovich, S., Bramel-Cox, P.J., Hamon, P. (2000b). Assessment of Genetic diversity in three subset constituted from the ICRISAT sorghum collections using random vs. on-random sampling procedures. B. Using molecular markers. *Theor Appl Genet* 101:197-202.

Guilford, P., Prakash, S., Zhu, J.M., Rikkerink, E., Gardiner, S., Basset, H. and Foster, R. (1997). Microsatellite in *Malus domestica* (apple): abundance. Polymorphism and cultivar identificatio*n. Theor. Appl. Genet*. 94:249-254.

Gupta, P.K., Balayan, H.S., Sharma, P.C., Ramesh, B. (1996). Microsatellite in plants: a new class of molecular markers. *Crr Sci* 70:45-55.

Hair, J.R., Anderson, R.E., Tatham, R.L. and Black, W.C. (1995). *Multivariate data analysis with readings*. 4th edition, Prentice-Hall, Englewood Cliffs, NJ. pp.23-29

Hamon, S., Frison, E & Navaro, L. (2004). *Connecting plant germplasm collection and genomic centers: How to better link curators, molecular biologist and genetists*? Pp. 33-42 *in M.C.* de Vicente, ed. *The evolving role banks in the fast-developing field of molecular genetics. Issues in genetic resources*, No. XI, august 2004. Rome, International Plant Genetic Resources Institute. pp. 78-80

Hamrick, J.L., Godt, M.J.W. (1997). Allozyme diversity in cultivated crops. *Crop Sci* 37:26–30.

Harlan, J.R and De Wet, J.H.J. (1971). Towards a rational classification of cultivated plants. *Taxon* 20:509-517.

Harlan, J.R and De Wet, J.H.J. (1972). A simplified classification of cultivated sorghum. *Crop Sci* 12:172-176.

Haussmann, H.H., Geiger, D.E., Hess, C.T., Hash, J and Bramel-Cox, P (Eds.)(2000). *Application of molecular markers in plant breeding*. Training manual for a seminar held at IITA, Ibadan, Nigeria, (pp 5-10).

Hintze, J. L., (2000). *Number Cruncher Statistical Systems (NCSS).* Statistical System for window. Kaysville, Utah. pp. 34-37

Hongtrakul, V., Huestis, G.M., Knapp, S.J. (1997). Amplified fragment length polymorphisms as a tool for DNA fingerprinting sunflower germplasm: Genetic diversity among oilseed inbred lines. *Theor. Appl. Genet*. 95:400-407.

House, L.R. (1985). *A guide to sorghum breeding*. (2nd Edn). Patancheru, A.P. 502324, India; International Crops Research Institute for the Semi-Arid Tropics. pp. 1-99

IBPGR/ICRISAT. (1993). *Descriptors for Sorghum* [*Sorghum bicolor* (L) Moench]. International Board of Plant Genetic Resources. Rome, Italy/International Crop Research Institute for Semi-Arid Tropics, Patancheru, India.

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

Jordan, D.R., Tao, Y.Z., Godwin, I.D., Henzel, R.G., Cooper, M., McIntyre, C.L. (1998). Loss of genetic diversity associated with selection resistance to sorghum midge in Australia sorghum. *Euphytica* 102:7.

Kaufman, L and Rousseau, P.J. (1990). *Finding groups in data. An introduction to cluster analysis* pp 342. A Wiley-Interscience Publication, New York.

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

Kresovich, S., and McFerson, J.R. (1992). Assessment and management of plant genetic diversity: considerations of intra- and interspecific variations*. Field Crops Res*. 29:185–204.

Lan, T.H., Paterson, A.H. (2000). Comparative mapping of quantitative trait loci sculpting the curd of *Brassica oleracea*. *Genetics* 155:1927–1954.

Lin, Y.R., Schertz, K.F., Paterson, A.H. (1995*)*. Comparative analysis of QTLs affecting plant height and maturity across the *poaceae,* in reference to an interspecifc sorghum population. *Genetics* 141:391–411

Mace, E.S., Mathur, P.N., Godwin, I.D., Hunter, D., Taylor, M.B., Singh, D., DeLacy, I.H. and Jackson, G.V.H**.** (2005). *Development of a regional Core Collection (Oceania) for taro, Colocasia esculenta* (L.) Schott, *based on molecular and phenotypic characterization. (in press)*. *In* V. Ramanatha Rao, P.J. Matthews & P.B. eyzaguire, eds. *The global diversity of taro: ethnobotany and conservation.* Rome, IPGRI; Osaka, Japan, Minpaku (National Museum of Ethnology, Osaka, Japan).

Mace, E.S., Buhariwala, H.K., Crouch, J.H. (2003) A high-throughput DNA extraction protocol for tropical molecular breeding programs. *Plant Mol Biol Rep*21:459a-459h.

Maughan, P.J., Sarghai-Maroof, M.A. and Buss, G.R. (1995). Microsatellite and amplified sequence length polymorphism in cultivated and d soy bean. *Genome*. 38:715-723.

Mbwaga, A. M. and Obilana, A. T. (1993): Distribution and host specificity of Striga asiatica and Striga hermonthica on cereals in Tanzania *in International Journal of Pest management* 39:449-451.

Mbwaga, A. M., Kaswende, J. and Shayo, E. (2000): *A reference Manual on Striga distribution and Control in Tanzania*. FARMESA Publication 2000.

Menkir, A., Goldsbrough, P., Ejeta, G. (1997). RAPD based assessment of genetic diversity in cultivated races of sorghum. *Crop Sci*. 37: 564-569.

Mitchell, S.E., S. Kresovich, C.A., Jester, C.J., Hernandez and Szewc-Fadden, A.K. (1997). Application of multiplex PCR and fluorescence-based, semi-automated allele sizing technology for genotyping plant genetic resources. *Crop Sci* 37: 617–624.

Mohan, M., Quarries, S., Cause, M and de'Vienne, D. (1997). Dissecting complex physiological functions through the use of molecular quantitative genetics. *Journal of experimental Botany* 48:1151-1163.

Morden, C.W., Doebley. K., Schertz, K.F. (1990) Allozyme variation among the spontaneous species of *Sorghum* section Sorghum (Poaceae*). Theor Appl Genet* 80:296–304.

Nei, M. (1987) *Molecular evolutionary genetics*. Columbia University Press, New York, USA.

Newbury, H. Ford-Lloyd, B.V. (1997) *Estimation of genetic diversity*. In: Maxted N, Ford-Lloyd BV, Hawkes JG (eds) Plant genetic conservation-the in situ approach. Chapman and Hall, London, pp 192–206.

Paul, S., Wachira, F.N., Powell, W., Waugh, R. (1997). Diversity and genetic differentiation among populations of Indian and Kenyan tea (*Camellia sinensis* (L.) O. Kuntze) revealed by AFLP markers. *Theor. Appl. Genet*. 94: 255-263.

Parida, A., Parani, M., Lakshmi, M., Nivedita, R., Elango, S. and Parida, A. **(**1998). Nature and extent of genetic variation and species diversity in Indian mangroves. *IAEA Techdoc*, 1047:95-105.

Peng, Y., Schertz, K.F., Cartinhour, S., Hart, G.E. (1999) Comparative genome mapping of *Sorghum bicolor* (L.) (Moench) using an RFLP map constructed in a population of recombinant inbred lines. *Plant Breed* 118:225–235.

Powell, W., Machray, G.C., Provan, J. (1996 a). Polymorphism revealed by simple sequence repeats. *Trend Plant Sci* 1:215-222.

Purse glove, J. W. (1987). *Tropical crops, Monocotyledons*. Vol. 1 and 2.

Prioul, J.L., Quarrie, S., Causse, M and de'Vienne, D. (1997). Dissecting complex physiological functions through the use of molecular quantitative genetics. *Journal of Experimental Botany* 48:1151-1163.

Provan, J., Powell, W and Waugh, R. (1996). Microsatellite analysis of relationships within cultivated potato (*Solanum tuberosum*). *Theor. Appl. Genet*. 92: 1078-1084.

Rafalski J.A and Scott V.T (1993) Genetic diagnosis in plant breeding: RAPDs, microsatellites and machines, *TIG vol*. 9 no. 8.

Russell, J., Fuller, J., Young, G., Thomas, B., Taramino, G., Macaullay, M., Waugh, R and Powell, W. (1997). *Discriminating between barley: species diversity, chromosomal locations and population dynamics.* Proc. Natl Acad. SCI. USA.91: 5466-5470.

Schlotterer, C. Microsatellites. In: Hoelzel AR (1998)(Eds.) *Molecular genetic analysis of populations: a practical approach*, 2nd edn. Oxford University Press, New York, pp 238–261.

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

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

Sonnate, G., Marangi, A., Venora, G., Pignone, D. (1997). Using RAPD markers to investigate genetic variation in chickpea. *Genet. Breed*. 51: 303-307.

Taramino, G and Tingey, S. (1996). Simple sequence repeats for germplasm analysis and mapping in maize. *Genome*. 39: 277-287.

Taramino, G., Tarchini, R., Ferrario, S., Lee, M., Pe,' M.E. (1997). Characterization and mapping of simple repeat (SSRs) in *Sorghum bicolar*, *Theor. Appl. Genet. 95*: 66-72.

Thomas, M.R., Cain, P and Scott, N.S. (1994). DNA vines: a universal methodology and data base for describing cultivars ad evaluating genetic relatedness. *Plant Mol. Biol*. 25: 939-949.

Uptmoor, R., Wenzel, J., Friedt, W., Donaldson, G., Ayisi, K., Ordon, F. (2003). Comperative analysis on the genetic relatedness of *sorghum bicolor* accessions from Southern Africa by RAPDs, AFLPs and SSRs. *Theor Appl Genet* 106:1316-1325.

Weir, B.S. (1996). *Genetic data analysis II.* 2nd ed. Sinauer Associates. Inc., Sunderland, MA.

Westman, A.L., Kresovich, S. (1997). *Use of molecular marker techniques for description of plant genetic variation.* In: Callow JA, Ford-Lloyd BV, Newbury HJ (Eds) Plant genetic resources. Conservation and use. CAB Int, Wallingford, UK, pp 9–48.

Wright, S. (1978). The genetical structure of populations. *Ann Eugenics* 15:323–354.

Zhu, J., Gale, M.D., Quarrie, S., Jackson, M.T., Bryan, G.J. (1998). AFLP markers for the study of rice biodiversity. *Theor. Appl. Genet*. 96:602-61.

APPENDICIES

Appendix 1: A list of sorghum accessions used in the study

S.No.	Accession No.	Collectors No.	Where collected
1	TZA 1213	TLSC 375	
2	TZA 1254	TLSC 419	
3	TZA 1209	TLSC 371	
4	TZA 1286		
5	TZA 1301	PR 5588	
6	TZA 1187	TLSC 348	
7	TZA 1299	PR 5582	
8	TZA 1204	TLSC 366	
9	TZA 930		
10	TZA 1251	TLSC 416	
11	TZA 1223	TLSC 386	
12	TZA 989	CER 64	
13	TZA 981	CER 47	
14	TZA 925		
15	TZA 1185	TLSC 346	
16	TZA 937		
17	TZA 924		
18	TZA 1408	AMF 152	
19	TZA 974	CER 30	
20	TZA 1217	TLSC 379	
21	TZA 1190	TLSC 351	
22	TZA 928		
23	TZA 970	CER 24	
24	TZA 1319	PR 5630	
25	TZA 1309	PR 5612	
26	TZA 1320	PR 5631	
27	TZA 1308	PR 5607	
28	TZA 1184	TLSC 345	
29	TZA 949		
30	TZA 1316	PR 5627	
31	TZA 938		
32	TZA 1259		
33	TZA 1016	MPR 45	
34	TZA 1312	PR 5620	
35	TZA 919		
36	TZA 1192	TLSC 353	
37	TZA 1318	PR 5629	
38	TZA 965	CER 15	
39	TZA 955		
40	TZA 969	CER 21	

41	TZA 1302	PR 5591
42	TZA 1279	
43	TZA 1295	
44	TZA 1224	TLSC 387
45	TZA 1283	
46	TZA 983	CER50
47	TZA1058	TLSC 204

S.No.	Accession No.	Collectors No.	Where collected
48	TZA 971	CER 25	
49	TZA 950		
50	TZA 1010	MPR 34	
51	TZA 936		
52	TZA 1007	MPR 24	
53	TZA 1292		
54	TZA 942		
55	TZA 1191	TLSC 352	
56	TZA 1091	TLSC 244	
57	TZA 1160	TLSC 319	
58	TZA 1242	TLSC 207	
59	TZA 948		
60	TZA 985	CER 55	
61	TZA 1285		
62	TZA1287		
63	TZA 1305	PR 5595	
64	TZA1004	MPR 13	
65	TZA 934		
66	TZA 924		
67	TZA 1003	CER 96	
68	TZA 966	CER 16	
69	TZA 945		
70	TZA 953		
71	TZA 939		
72	TZA 976	CER 38	
73	TZA 946		
74	TZA 1282		
75	TZA 954		
76	TZA 1206	TLSC 368	
77	TZA 1293		
78	TZA 921		
79	TZA 1011	MPR 35	
80	TZA 1272		
81	TZA 1183	TLSC 344	
82	TZA 935		
83	TZA 984	CER 53	
84	TZA 1291		
85	TZA 979	CER 43	
86	TZA 1005	MPR 16	
87	TZA 1250	TLSC 415	
88	TZA 986	CER 57	
89	TZA 1247	TLSC 411	

S.No.	Accession No.	Collectors No.	Where collected
90	TZA 1294		
91	TZA 931		
92	TZA 1249	TLSC 413	
93	TZA 920		
94	TZA 922		
95	TZA 932		
96	TZA 941		
97	TZA 1261		
98	TZA 1210	TLSC 372	
99	TZA 992	CER 67	
100	TZA 1229	TLSC 393	
101	TZA 1274		
102	TZA 1277		
103	TZA 1220	TLSC 382	
104	TZA 1201	TLSC 362	
105	TZA 1267		
106	TZA 1182	TLSC 343	
107	TZA 1264		
108	TZA 1272		
109	TZA 1270		
110	TZA 1198	TLSC 359	
111	TZA 1186	TLSC 347	
112	TZA 1226	TLSC 389	
113	TZA 1222	TLSC 384	
114	TZA 1278		
115	TZA 1231	TLSC 395	
116	TZA 1276		
117	TZA 1239	TLSC 403	
118	TZA 1265		
119	TZA 1238	TLSC 402	
120	TZA 1189	TLSC 350	
121	TZA1246	TLSC 410	
122	TZA 1197	TLSC 358	
123	TZA 1243	TLSC 407	
124	TZA 1001	CER 89	
125	TZA 1205	TLSC 367	
126	TZA 1241	TLSC 405	
127	TZA 973	CER 28	
128	TZA 1273		
129	TZA 1200	TLSC 361	
130	TZA 1271		
131	TZA 996	CER 74	
132	TZA 1218	TLSC 380	
133	TZA 1236	TLSC 400	
134	TZA 918		
135	TZA 991	CER 66	
136	TZA 1257		
137	TZA 1000	CER 87	
138	TZA 1008	MPR 25	
139	TZA 1208	TLSC 370	

140	TZA 1230	TLSC 394	
141	TZA 1303	PR 5592	
142	TZA 998	CER 85	
143	TZA 1275		
144	TZA 1304	PR 5593	
145	TZA 929		
146	TZA 1203	TLSC 365	
147	TZA 947		
S.No.	**Accession No.**	**Collectors No.**	**Where collected**
148	TZA 994	CER 72	
149	TZA 1263		
150	TZA 1216	TLSC 378	
151	TZA 3493	KEL 324	Nzega, Tabora
152	TZA 4166	CCM 125	Bukoba Rural, Kagera
153	TZA 4247	MN 28	Bukoba Rural, Kagera
154	TZA 3418	KEL 245	Igunga, Tabora
155	TZA 3436	KEL 263	Igunga, Tabora
156	TZA 3494	KEL 325	Nzega, Tabora
157	TZA 4001	NAM 94	Ukerewe, Mwanza
158	TZA 4171	CCM 130	Biharamulo, Kagera
159	TZA 3949	NAM 42	Serengeti, Mara
160	TZA 3966	NAM 59	Serengeti, Mara
161	TZA 3921	NAM 13	Tarime, Mara
162	TZA 4044	NAM 137	Ukerewe, Mwanza
163	TZA 3426	KEL 253	Igunga, Tabora
164	TZA 4155	CCM 114	Biharamulo, Kagera
165	TZA 4031	NAM 124	Ukerewe, Mwanza
166	TZA 3904	LNA 323	Nachingwea, Lindi
167	TZA 4219	CCM 178	Mwanza
168	TZA 4009	NAM 102	Ukerewe, Mwanza
169	TZA 4198	CCM 157	Mwanza
170	TZA 3864	LNA 283	Nachingwea, Lindi
171	TZA 3955	NAM 48	Serengeti, Mara
172	TZA 4158	CCM 117	Biharamulo, Kagera
173	TZA 3616	LNA 33	Mtwara
174	TZA 3882	LNA 301	Nachingwea, Lindi
175	TZA 3943	NAM 36	Serengeti, Mara
176	TZA 4146	CCM 105	Biharamulo, Kagera
177	TZA 3970	NAM 63	Musoma Rural, Mara
178	TZA 3912	NAM 4	Tarime,Mara
179	TZA 3835	LNA 254	Newala, Mtwara
180	TZA 3920	NAM 12	Tarime, Mara
181	TZA 3992	NAM 85	Musoma Rural, Mara
182	TZA 3994	NAM 87	Musoma Rural, Mara
183	TZA 3425	KEL 252	Igunga, Tabora
184	TZA 3424	KEL 251	Igunga, Tabora
185	TZA 3993	NAM 86	Musoma Rural, Mara
186	TZA 3965	NAM 58	Serengeti, Mara
187	TZA 4189	CCM 148	Biharamulo, Kagera
188	TZA 4021	NAM 114	Ukerewe, Mwanza
189	TZA 4162	CCM 121	Biharamulo, Kagera

190	TZA 3880	LNA 299	Nachingwea, Lindi
191	TZA 3983	NAM 76	Musoma Rural,Mara
192	TZA 4027	NAM 120	Ukerewe, Mwanza
193	TZA 3147	HPL 63	Kasulu, Kigoma
194	TZA 4011	NAM 104	Ukerewe, Mwanza
195	TZA 4004	NAM 97	Ukerewe, Mwanza
196	TZA 3878	LNA 297	Nachingwea, Lindi
197	TZA 4023	NAM 116	Ukerewe, Mwanza
S.No.	**Accession No.**	**Collectors No.**	**Where collected**
198	TZA 3933	NAM 25	Ukerewe, Mwanza
199	TZA 123	WMK 70	Sumbawanga
200	TZA 125	WMK 72	Sumbawanga

Appendix 2: The parameters used for the morphological characterization

Characteristics	Descriptions
Plant height (cm)	Height above the ground measured in cm at harvest
Plant colour	done at the flowering stage using the standard colour chart
Stalk juiciness	done at the fourth internode's at maturity
Juice flavours	scored at the fourth internode of the stalk at plant maturity age
Midrib colour	assessed at plant flowering stage on any of the fully bloomed leaves.
Wax bloom	the character was assessed on the on the stalk at the flowering
Days of flower	taken when 50% of plants in the plot were heading
Inflorescence compactness and shape	the score was based on the picture from descriptors
Inflorescence exsertion (cm)	measured in cm fro the base of the flag leaf of head.
Inflorescence length (cm)	made in the middle of the panicle with the ruler in cm underneath the head
Inflorescence width (cm)	made in the middle of the panicle with the ruler or tape measure in cm underneath the head
Glume colour	measured at maturity stage, colour codes from the descriptor used
Grain covering	Taken by scratching the outer seed cover and scored as (1) absent and (2) for present.
Awns	character assessed at maturity and scored as 1 absent and 2 as present
Shattering	scored immediately after physiological maturity
Grain colour	scored at maturity stage using colour codes provided from the descriptor
Grain lustre	assessed at maturity stage scored as 1 absent and 2 present.
100 seed weight (g)	100 g grain weight per accession
Grain number per panicle	where three heads were used for scoring the character and the mean was calculated
Grain sub-coat	Taken by scratching the outer seed cover and scored as (1) absent and (2) for present.
Grain plumpness	The magnifying glass was used to see clearly and score the features of grain plumpness
Grain form	Assessed at physiological maturity when the seeds were still in the panicle
Endosperm texture	the seed was dissected longitudinally and the magnified glass was used
Endosperm colour	the seed was dissected longitudinally and the magnified glass was used
Endosperm type	the seed was dissected longitudinally and the magnified glass was used
Senescence	assessed at grain maturity for the death of leaves and stalk.