GENETIC DIVERSITY ANALYSIS IN SORGHUM GERMPLASM COLLECTIONS FROM EASTERN AFRICA AS ESTIMATED BY MORPHO-AGRONOMICAL AND SSR MARKERS

By

SHADIA ABDALLAH SALIH

A thesis submitted in accordance with the requirements for the degree Philosophiae Doctor

in the Department of Plant Sciences (Plant Breeding)
Faculty of Natural and Agricultural Sciences
University of the Free State
Bloemfontein, South Africa

May 2011

Promotor: Prof. Liezel Herselman (PhD)
Co-Promotors: Prof. Maryke T. Labuschagne (PhD)
Dr. Dan Kiambi (PhD)

Declaration

I, Shadia Abdallah Salih, do hereby declare that the thesis hereby submitted for qualification for the degree Philosophiae Doctor in Agriculture at the University of the Free State represents my own original, independent work and that I have not previously submitted the same work for a qualification at another university.

I further cede copy right of the thesis in favour of the University of the Free State

Shadia ..... 19 May 2011

Shadia Abdallah Salih Date

Dedication

Dedicated

to my father and my son

Ahmed

Acknowledgements

Firstly I give glory to my creator, the Almighty God, who chose to reveal to me all the knowledge generated through my studies. He strengthened me and has been my source of courage throughout my studies. I hope in some way this work will glorify His name.

During my study several organizations and institutions collaborated directly and indirectly to my work. Without their support it would have been impossible for me to finish this research; that is why I would like to dedicate this section to recognize their helpful support and love. I am grateful to the management of Biosciences eastern and central Africa (BecA) for awarding me a PhD graduate fellowship under the project "Tapping crop biodiversity for the resource poor in east and central Africa", originated by the Generation Challenge Programme (GCP) with funds from the Canadian International Development Agency (CIDA) through BecA and the International Crops Research Institute for Semi-Arid Tropics (ICRISAT). Special thanks to the Agricultural Research Corporation (ARC) for conceding me a leave to undertake the programme. Special thanks to ICRISAT located within the International Livestock Research Institute (ILRI), Nairobi, for granting me a healthy environment for data analysis and writing of the thesis.

Without the intellectual stimulation and emotional support of supervisors, family, friends and colleagues I would never have reached this stage. It is not possible to acknowledge by name all those who influenced or help me, therefore, I would like to mention a few of those who help me so much and due of their support I was able to complete this thesis. I extend my honest appreciation to Profs. Liezel Herselman (promoter) and Maryke T. Labuschagne (co-promoter) for their enthusiastic support, scientific guidance and well appreciated effort in appraising the drafts of this thesis. Incredibly exceptional mention goes to the project principle investigator and thesis co-promoter, Dr. Dan Kiambi formerly from ICRISAT.

This work would not have been complete without the generosity of BecA coordinators, and BecA Masters students from Kenya, Uganda, Ethiopia, Eritrea, Tanzania, Sudan, Burundi, and Rwanda for collecting data from their countries, Thank you very much.

Thanks are also extended to the technical staff of the ICRISAT/ BecA located at ILRI, Nairobi for their excellent cooperation, assistance, and friendship during the genotyping analysis period in Nairobi. I am grateful to three wonderful ladies, Rosemary Mutegi, Mercy Karichi and Maggie Mwathi who introduced me to practical molecular techniques: DNA extraction, PCR amplification and genotyping. Words are not enough to express my appreciation. You mean a lot to me.

I would like to thank all members who attended the phenotypic and genotypic data analysis workshop for their advice and comments, special thanks to Professor Jorge Franco for his valuable comments and advice.

I would also like to thank Dr Evans Mutegi for his useful assistance with all computer packages I used to analyze the data. He took me up in faith and I have been honoured to reciprocate to a man who in more ways than one has been a good brother to me during my stay in Nairobi.

My sincere thanks and love are addressed to my dear father, mother, brothers, and sisters whose care and help encouraged me through the study period. Special thank to my sister Lemya, who looked after my baby during my study.

Table of contents

List of tables

List of figures

List of abbreviations

A^p	Number of private alleles
A^r	Number of rare alleles
A_s	Allelic richness
A^t	Total number of alleles
AFLP	Amplified fragment length polymorphism
AMOVA	Analysis of molecular variance
ANOVA	Analysis of variance
ARC	Agricultural Research Corporation
BC	Before Christ
BecA	Biosciences Eastern and Central Africa
bp	Base pair(s)
CIDA	Canadian International Development Agency
cpDNA	Chloroplast DNA
CBSU	Computational Biology Service Unit
CTAB	Hexadecyltrimethyl ammoniumbromide
DArT	Diversity array technology
DF	Days to 50% flowering
DNA	Deoxyribonucleic acid
dNTP	2'-deoxynucleoside 5'-triphosphate
EA	East Africa
ESIP	Ethiopian Sorghum Improvement Project
EDTA	Ethylenediaminetetraacetic acid
ET	Endosperm texture
f	Coefficient of co-ancestry
F_{IS}	Fixation index of individuals relative to the sub-population
F_{IT}	Fixation index of individuals relative to the total population
F_{ST}	Fixation index of sub-population relative to the total population/total fixation index
FAO	Food and Agriculture Organisation of the United Nations.
GC	Grain covering
GCP	Generation Challenge Programme
GD	Genetic distance

GLC	Glume colour
GM	Gaussian Model
GRC	Grain colour
GRF	Grain form
GRP	Grain plumpness
GS	Genetic similarity
H'	Shannon-Weaver diversity index
H_e	Expected heterozygosity/gene diversity
H_o	Observed heterozygosity
HCl	Hydrochloric acid
ICRISAT	International Crops Research Institute for Semi-Arid Tropics
IE	Inflorescence exsertion
ILRI	International Livestock Research Institute
JF	Juice flavour
K	Number of unknown populations/genetic clusters
KCl	Potassium chloride
MCMC	Markov Chain Monte Carlo
MDS	Multidimensional scaling
$MgSO_4$	Magnesium sulphate
mtDNA	Mitochondrial DNA
NaCl	Sodium chloride
NARS	National Agricultural Research Systems
NJ	Neighbour-joining
NPGS	National Plant Germplasm System
P(X|K)	Probability of X given K
PC	Plant colour
PCA	Principal component analysis
PCoA	Principle coordinate analysis
PCR	Polymerase chain reaction
PCS	Panicle compactness and shape
PHt	Plant height
PIC	Polymorphic information content
PL	Panicle length

PW	Panicle width
RAPD	Random amplified polymorphic DNA
RFLP	Restriction fragment length polymorphism
RNAse	Ribonuclease
SE	Senescence
SJ	Stalk juiciness
SNP	Single nucleotide polymorphism
ssp	Subspecies
SSR	Simple sequence repeat
Taq	*Thermus aquaticus*
TE	Tris/EDTA buffer
Tris-HCl	Tris (hydroxymethyl) aminomethane hydrochloride
TSWt	1000-seed weight
UPGMA	Unweighted pair-group method using arithmetic averages
USA	United States of America
USDA	United States Department of Agriculture
UV	Ultraviolet

List of SI units

°C	Degrees centigrade
cm	Centimetre(s)
g	Gram(s)
h	Hour(s)
ha	Hectare(s)
kg	Kilogram(s)
km	Kilometre(s)
m	Metre(s)
M	Molar(s)
mg	Milligram(s)
min	Minute(s)
ml	Millilitre(s)
mm	Millimetre(s)
mM	Millimolar(s)
ng	Nanogram(s)
pH	Measure of acidity/basicity
rpm	Revolutions per minute
s	Second(s)
U	Unit(s)
V	Volt(s)
v/v	Volume/volume
w/v	Weight/volume
µg	Microgram(s)
µl	Microlitre(s)
µM	Micromolar(s)
%	Percentage(s)

List of presentations and posters

S. Salih, L. Herselman, M. Labuschagne and D. Kiambi. 2011. Genetic diversity analysis in sorghum germplasm collections from eastern Africa using microsatellites. A presentation made at the ARC. Hussien Idis Hall. 2011. ARC, Wad Medani, 11 - 12 May 2011.

S. Salih, L. Herselman, M. Labuschagne and D. Kiambi. 2011. Analysis of genetic diversity of Sorghum germplasm collections from eastern Africa as estimated by SSR and morph-agronomic marker. Poster presented at the BecA - ILRI Hub Conference: "Mobilizing biosciences for Africa's development" BecA - ILRI Hub, Nairobi, 4 - 5 November 2010.

Chapter 1

General introduction

Genetic diversity among and within genera, species, subspecies, populations, and elite breeding materials is of interest in plant genetics. By combining genetic variation, high levels of diversity will in many cases provide robustness to natural ecosystems and maximize further diversification. In order to maintain the existing genetic diversity, humans increasingly manage natural ecosystems. Diversity provides insurance against catastrophic damage and act as a resource for future human use. Management by the farmer is the key determinant of genetic diversity in agricultural ecosystems (Wenzl *et al*., 2004). Sorghum [*Sorghum bicolor* (L.) Moench, 2n=20] is fifth in importance among the world's cereals (Doggett, 1988) and the major crop in warm, low-rainfall areas of the world. It is a crop with extreme genetic diversity (Subudhi *et al*., 2002) and is predominantly self-pollinating, with varying levels of outcrossing. The highest level of variability is found in the northeast quadrant of Africa, which includes Ethiopia, Eritrea, and Sudan, and most evidence points to this area as the likely principal area of its domestication (Vavilov, 1951; House, 1985; Doggett, 1988).

Sorghum is Africa's second most important cereal based on both area harvested and annual production. According to global statistics (FAO, 2008), Africa contributes over 60% to the total land area dedicated to cultivation of sorghum. Sorghum thus plays an important role as dietary staple for millions of people, especially in arid and semi-arid countries of Africa and Asia (Bantilan *et al*., 2001).

Eastern and central Africa is affected by civil strife and recurrent drought. As a result, many people are at risk in terms of food insecurity and malnutrition because of a decrease in crop production in both rainfed and irrigation areas. Sorghum, after millet, is superior in drought tolerance and adaptability to poor soils and therefore holds great potential in providing food security in the region. However, many valuable landraces of sorghum either have been lost or are under serious risk. Consequences of these losses are a high risk for genetic erosion. New germplasm that act as a source of favourable genes and/or gene complexes are needed to develop high yielding and stable varieties. Landraces,

introductions, and weedy and wild relatives of crop plants act as primary sources of these needed genes. Comprehensive knowledge of genetic diversity of cultivated and wild germplasm, the source of novel genomic regions, alleles and traits, is therefore important (Xiao *et al*., 1998; Li *et al*., 2003).

Evaluation of genetic diversity can indicate which landraces are genetically novel and most suitable for rescue and possible future use in crop improvement. Furthermore, to improve and stabilize production and utilization of sorghum in specific areas, new sorghum lines should yield equal or better than existing landraces familiar to farmers. Evaluation of genetic diversity levels among adapted, elite germplasm can provide predictive estimates of genetic variation among segregating progeny for pure line development (Manjarrez-Sandoval *et al*., 1997). The use of germplasm for cultivar improvement developed within the same region aims to reduce the risk of loosing essential adaptive characteristics through recombination (Allard, 1996). In order to improve yield and other consumer preferred traits through the use of landraces, complete information on the genetic diversity of sorghum available in the region is therefore a priority.

The accurate, fast, reliable, and cost-effective identification of plant populations and varieties is essential in agriculture as well as in pure and applied plant research (Morell *et al*., 1995). Traditionally, taxonomists classified genetic resources in sorghum based on morphological traits (Stemler *et al.*, 1977). This usually involves description of variation for morphological traits, particularly morpho-agronomical characteristics of direct interest to users. While these methods are effective for many purposes, morphological comparisons may have limitations, including subjectivity in the analysis of the character, influence of environmental or management practices on the character, limited diversity among cultivars with highly similar pedigrees, and confinement of expression of some diagnostic characters to a particular stage of development, such as flowering or seed maturity (Morell *et al*., 1995). Menkir *et al*. (1997) indicated that important traits, which are related to habitat adaptation and particular end use of the crop, exhibit enormous variability among sorghum germplasm. Hence, classifying germplasm accessions based solely on morphological characters may not provide an accurate indication of the genetic divergence among cultivated genotypes of sorghum (Ejeta *et al*., 1999).

These considerations have led to the exploration or adoption of other techniques for genetic diversity estimation and cultivar identification, including cytogenetic analysis, isozyme analysis, and molecular techniques that directly analyze polymorphism at DNA level. Molecular markers are nowadays widely used as tools to assess the soundness of morphological classification in crop plants. Microsatellites or simple sequence repeat (SSR) DNA markers have proved to be efficient and reliable in supporting conventional plant breeding programmes (Paterson *et al*., 1991; Morell *et al*., 1995; Kumar, 1999).

The objectives of this study were to:

i. Study the genetic population structure of sorghum collections held by National Agricultural Research Systems (NARS) in eastern Africa through quantifying and understanding the partitioning of genetic diversity within and between populations and within and between countries,

ii. Quantify diversity through the combined use of SSR profiles, and highly reliable morpho-agronomical characters,

iii. Develop a database of about 1720 accessions held by the East Africa (EA) NARS including passport, phenotypic, and genotypic data.

References

Allard RW. 1996. Genetic basis of the evolution of adaptedness in plants. Euphytica 92:1-11.

Bantilan MCS, Parthasarathy Rao P, Padmaja R. 2001. Future of agriculture in the semi-arid tropics: Proceedings of an International Symposium on Future of Agriculture in Semi-Arid Tropics, 14 November 2000, ICRISAT, Patancheru, India. Patancheru 502 324, Andhra Pradesh, India: International Crops Research Institute for the Semi-Arid Tropics. pp. 1-98.

Doggett H. 1988. Sorghum. Longmans (Second edition). Green and Co. Ltd., London. pp. 512.

Ejeta G, Goldsbrough PB, Tuinstra MR, Grote EM, Menkir A, Ibrahim Y, Cisse N, Weerasuriya Y, Melakeberhan A, Shaner CA. 1999. Molecular marker applications in sorghum: Paper presented at workshop on application of molecular markers, Ibadan, Nigeria, 16-20 August 1999.

FAO. 2008. World development report 2008: Agriculture for development. World Bank publisher. pp. 365.

House LR. 1985. A guide to sorghum breeding (Second edition). Patancheru AP, ICRISAT, India. pp. 206.

Kumar LS. 1999. DNA markers in plant improvement: An overview. Biotechnology Advances 17:143-182.

Li ZK, Fu BY, Gao YM, Xu JL, Vijayakumar CHM, Ali J, Lafitte R, Ismail A, Yanagihara S, Zhao MF, Dimingo J, Maghirang R, Hu FY, Zhao XQ. 2003. Discovery and exploitation of hidden genetic diversity in germplasm collections for genetic improvement of abiotic stress tolerances in rice. XIX International Congress of genetics, Melbourne, 6-11 July 2003.

Manjarrez-Sandoval P, Carter TE, Webb DM, Burton JW. 1997. RFLP genetic similarity estimates and coefficient of parentage as genetic variance predictors for soybean yield. Crop Science 37:698-703.

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

Morell MK, Peakall R, Appels R, Preston LR, Lloyd HL. 1995. DNA profiling techniques for plant variety identification. Australian Journal of Experimental Agriculture 35:807-819.

Paterson AH, Tanksley SD, Sorrells, SM. 1991. DNA markers in plant improvement. Advances in Agronomy 46:39-90.

Stemler AB, Harlan JR, de Wet JMT. 1977. The sorghums of Ethiopia. Economic Botany 31:446-460.

Subudhi PK, Nguyen HT, Gilbert ML, Rosenow DT. 2002. Sorghum improvement: past achievements and future prospects. In: Crop improvement. Edited by Kang MS. Challenges in the twenty-first century. The Haworth Press, Inc., NY, pp. 109-158.

Vavilov NI. 1951. The origin, variation, immunity and breeding of cultivated plants. Chronica Botanica 13:1-366.

Wenzl P, Caig V, Carling J, Cayla C, Evers M, Jaccoud D, Peng K, Patarapuwandol S, Uszynski G, Xia L, Yang S, Huttner E, Kilian A. 2004. Diversity array technology, a novel tool for harnessing crop genetic diversity. Proceedings of the National Academy of Science USA 101:9915-9920.

Xiao J, Li L, Yuan L, McCouch SR, Tanksley SD. 1998. Genetic diversity and its relationship to hybrid performance in heterosis in rice as revealed by PCR-based markers. Theoretical and Applied Genetics 92:637-643.

Chapter 2

Literature review

2.1 Introduction

Sorghum is the fifth most important cereal crop worldwide after wheat (*Triticum* species), rice (*Oryza* species*),* maize (*Zea mays* L.) and barley (*Hordeum vulgare* L.) with an annual average production of 61 million ton over the past decade (FAO, 1995; Folkertsma *et al*., 2005). Sorghum, together with pearl millet (*Pennisetum americanum* (L.) Leeke) and finger millet (*Eleusine coracana* (L.) Geartn), represent Africa's main contribution to the world's food supply (De Vries and Toenniessen, 2000; Folkertsma *et al*., 2005). It is an annual grass of the family Gramineae that varies between 0.5-5.0 m in height and is closely related to maize. Sorghum produces one or several tillers that emerge initially from the base and later from the stem nodes. The flower is a panicle, usually erect, but sometimes curved in a goose neck (Doggett, 1988) and the crop is predominantly self-pollinating. Cultivated sorghum has been classified into five major and ten intermediate races on the basis of grain and glume morphology (Harlan and De Wet, 1972; Folkertsma *et al*., 2005).

Grain sorghum has the ability to tolerate conditions of limited moisture and produce during periods of extended drought, in conditions that would prevent growth of other cereal crops. Several drought resistance mechanisms in sorghum make it more drought resistant compared to other grains. Therefore, sorghum can be grown under a wide range of soil and climatic conditions. Sorghum is an important major cereal in western Africa, and worldwide, due to its capacity to tolerate harsh growing conditions. Thus, the crop plays a major role under drought, heat, and poor soil conditions in the semi-arid regions of the world where other cereal crops tend to fail (Doggett, 1988; House *et al.,* 1995).

Sorghum is indigenous to Africa and is one of the oldest cultivated crops of the warm regions of Africa and Asia, especially India and China. The crop is cultivated in 100 countries around the world in the Americas, Africa, Asia, and the Pacific. Fifty nine percent of the world land area planted with sorghum is in Africa. Asian countries occupy 25% of the world sorghum area, North and Central America 11% and South America 4%. Developing countries in Asia and Africa contribute towards more than 70% of the total

sorghum production in the world. Asia alone contributes 45% of world sorghum production, North and Central America 21% and South America 6% (Bantilan *et al*., 2004). Eighty percent of the area devoted to sorghum is located within Africa and Asia, with average yields of 810 and 1150 kg/ha, respectively. The bulk of African sorghum production is centred in the savanna zone of east, west and central Africa, where grain of this crop is a major component of the daily menu for millions of people. Twelve of the 20 largest sorghum producing countries in the world are in Africa, with Nigeria being one of the leading world producers of the crop. The main producers of sorghum are the USA, India, Nigeria, China, Mexico, Sudan, and Argentina (Bantilan *et al*., 2004).

Sorghum is used to make unleavened bread, porridge, and malted beverages, including beer. The straw of traditional tall sorghums is used to make shelters in villages or around homesteads. Sorghum is a principal feed ingredient for both cattle and poultry (De Vries and Toenniessen, 2000) and in recent years it has become an important source of biofuels (Laopaiboon *et al*., 2007).

2.2 Sorghum taxonomy, origin, and domestication

Sorghum is a heterogeneous genus belonging to the botanical family Gramineae under the Andropogoneae tribe that includes the following sections: *Spitosorghum*, *Parasorghum*, *Heterosorghum*, *Chaetosorghum*, and *Eusorghum* (Garber, 1950). *Spitosorghum* and *Parasorghum* are characterized by distinct rings of hairs at each culm node and the awns of *Spitosorghum* are longer than those of *Parasorghum*. *Heterosorghum*, *Chaetosorghum*, and *Eusorghum* are characterized by hairy or globrous culm nodes with no hairs in the nodal ring. The pedicellate spikelets are reduced to subequal glumes in *Heterosorghum* and unequal glumes in *Chaetosorghum*. The section *Eusorghum* is characterized by better developed pedicellate spikelets (Snowden, 1936).

The section *Eusorghum* includes cultivated grain sorghum, a complex of closely related annual taxa from Africa and a complex of perennial taxa from southern Europe and Asia. The section *Eusorghum* is divided into two groups, the *Halapensia* and *Arundinacea* complex (Snowden, 1955). The *Halapensia* complex includes four rhizomatous species *S. controversum* (Steud.) Snowden, *S. halepense* (L.) Pers., *S. miliaceum* (Roxb.) Snowden and *S. propinquum* (Kunth) Hitch-cock. The *Arundinacea* complex, as recognized by

Snowden (1936; 1955), includes seven weedy species, 13 wild species, and 28 species of cultivated grain sorghum.

Cultivars and their wild and weedy relatives form part of the primary and secondary gene pools of sorghum (Harlan and De Wet, 1972) within the section *Eusorghum*. Three species are documented within this section: (i) *S. halepense*, a member of the secondary gene pool, is a perennial with creeping rhizomes and a native of southern Eurasia to east India, but now introduced in warm temperate regions of the world. In America, it has introgressed with grain sorghum to generate the widely dispersed Johnson grass (Celarier, 1958), (ii) *S. propinquum*, a member of the primary gene pool, is a perennial with stout rhizomes. It is a weedy species and occurs in Ceylon and southern India with distribution mainly in south east Asia and (iii) *S. bicolor*, the most important member of the primary gene pool, is described as an annual, with thick culms up to 5 m in height, often branched with many tillers. It is indigenous to Africa and comprises all cultivars of sorghum, their wild progenitors as well as weedy forms that are derivatives of crop-to-wild introgression (De Wet, 1978). Cultivated sorghum and its wild progenitors were classified under a single species, *S. bicolor*, within which three sub-specific categories are recognized: ssp. *bicolor*, ssp. *verticilliflorum* (Steud.) and ssp. *drummondii* (Steud.) (Harlan and De Wet, 1972; Doggett, 1988). All genotypes within *S. bicolor* ssp. *bicolor* have 2n=2x=20 chromosomes.

Harlan and De Wet (1972) classified sorghum cultivars into five basic races on the basis of spikelet and panicle morphology, namely *bicolor*, *kafir*, *caudatum*, *guinea*, and *durra*, with ten intermediate races representing all possible combinations between the five main races. Smith and Frederiksen (2000) reported that these 15 races of cultivated sorghum can be linked back to their specific environments and the nomadic people that first cultivated them.

Early studies on the evolution of sorghum were carried out by Snowden (1936), Harlan and De Wet (1972), Harlan *et al.* (1976), Stemler *et al*. (1977), Doggett (1988), and Doggett and Prasada Roa (1995). These studies focussed on locating the origin of sorghum in Africa and identified the region of domestication as a band stretching from southwest Ethiopia to Lake Chad. Harlan (1975) concluded that the initial domestication of sorghum occurred in a long belt across central Africa, perhaps through Ethiopia, Sudan

and Chad. De Wet and Huckabay (1967) postulated that sorghum was domesticated independently from local wild relatives of the crop in three regions: Ethiopia, tropical west Africa and southeast Africa. Doggett (1988) suggested that sorghum was domesticated about 3000 BC in the region of northeast Africa. Ethiopia in particular, is considered a centre of probable origin (Doggett and Prasada Rao, 1995). Doggett (1988) reported that the greatest genetic diversity of cultivated and wild sorghum is present in the northeast quadrant of Africa comprising Ethiopia, Sudan and east Africa.

Smith and Frederiksen (2000) reported that arthropological data indicated that hunters/gatherers consumed sorghum as early as 8000 BC. Sorghum originated in Ethiopia and surrounding countries, commencing around 4000-3000 BC. This confirmed an earlier hypothesis by Murdock (1959) that sorghum was independently domesticated in west Africa by Mande people around 4500 BC and was then introduced from west Africa to Sudan round about 4000 BC from the Lake Chad region. Moreover, sorghum occurred in archaeological sites in India, millennia before confirmed dates in Africa (Fuller, 2003). Blench (2006) suggested that wild sorghum was cultivated in the Chad-Ethiopia belt from 6000 BC onwards but that domestication took place outside Africa, perhaps in India.

2.3 Sorghum genetic resources

Germplasm collection and conservation has become an integral component of crop improvement programmes at both national and international levels in order to prevent extinction of landraces and wild relatives of cultivated sorghum (Rosenow and Dahlberg, 2000). Many centres have been established around the world to conserve sorghum genetic resources. At global level, sorghum germplasm consists of approximately 168500 accessions, which comprises 18% landraces, 21% breeding lines, and 60% mixed categories of unknown material, with only a few wild relatives being conserved (Chandel and Paroda, 2000). One of the major organizations and countries that maintain sorghum genetic resources is the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), with the largest collection (21% of the global total). ICRISAT maintains about 36774 accessions from 90 countries, representing approximately 80% of the variability present in the crop (Gopal *et al.,* 2006). Landraces comprise 84% of the total collection compared to wild species that comprise only 1%. The United States Department of Agriculture (USDA), established around 1905, has a total of 42221

germplasm accessions that are currently being maintained at the National Plant Germplasm System (NPGS) (Dahlberg and Spinks, 1995). The Ethiopian Sorghum Improvement Project (ESIP) started with collection, evaluation, documentation, and conservation of germplasm in the early 1970s. Rosenow and Dahlberg (2000) estimated that roughly 8000 germplasm collections are being maintained and that the types of sorghum in Ethiopia are zera-zera, durra, and durra-bicolor derivatives. Zera-zera is useful in providing germplasm for improvement of food-type sorghum. The Sudanese landrace collection was established at the Tozi Research Station in 1950, with the *caudatum* race being dominant. Sudanese sorghums have been useful as sources of drought tolerance (Rosenow *et al.*, 1999). An extensive collection of sorghum genotypes has been undertaken in China, with 12836 germplasm accessions being conserved in the National Germplasm Resource Bank. About 10414 of these accessions are registered as genetic resources (Qingshan and Dahlberg, 2001).

2.4 Methods for assessing genetic variation

Information concerning germplasm diversity and genetic relationships among breeding materials could be an essential tool in crop improvement strategies. A number of methods for analysis of genetic diversity in germplasm accessions, breeding lines, and populations are currently available. Diverse data sets have been used by researchers to analyze genetic diversity in crop plants and most important among such data sets are: passport and morphological data (Smith and Smith, 1992; Bar-Hen *et al*., 1995), pedigree data (Messmer *et al*., 1993; Mohammadi and Prasanna, 2003), biochemical data obtained by analysis of isozymes (Hamrick and Godt, 1997), and storage proteins (Smith *et al*., 1987). Recently, DNA-based marker data that allows more reliable differentiation of genotypes (Mohammadi and Prasanna, 2003) have been used. Sequencing of genomic DNA is a straightforward approach for identifying variants at a locus because genes are the cause of phenotypic variation. Many studies have aimed at assessing the genetic diversity in germplasm collections of crops using allozyme markers, morphological characters, storage proteins, isozymes or molecular markers (Morden *et al*., 1989; Maquet *et al*, 1997). Karp *et al*. (1997) reported that the choice of the analytical method to be used depends on the aim of the experiment, level of resolution required, available resources and technological infrastructure, and operational and time constraints.

Accurate measurement of the level and pattern of genetic diversity can be useful in crop breeding for diverse applications including (i) analysis of genetic variability in cultivars (Smith, 1984; Zeb *et al*., 2009), (ii) identifying diverse parental combinations to create segregating progenies with maximum genetic variability for further selection (Tucak *et al*., 2010), and (iii) introgression of desirable genes from diverse germplasm into the available genetic base (Ali *et al*., 2010).

2.5 Morphological characteristics and pedigree data

Morphological and phenological methods were among the earliest genetic markers used in germplasm management (Stanton *et al*., 1994) as they rely on discriminating between individuals based on physical characteristics, e.g. maturity cycle, growth habit, leaf shape, hairiness, nature of corolla, and panicle/pod/fruit size (Van der Maesen, 1990). Morphological characters used in taxonomical classifications are easy to observe and it is possible to screen and categorize large amounts of germplasm at a low cost, which is a great advantage when managing large germplasm collections (FAO, 1995). These methods, however, have many limitations. For example, these characters may not be significantly distinct, hence require that plants grow to full maturity prior to identification (Ratnaparkhe *et al*., 1995). In addition, these characters are often influenced by environmental factors, resulting in differences in expression that complicate interpretation of results. Because different genes are expressed at different developmental stages or in different tissues, the same type of material must be used for all experiments. Furthermore, there may be a limited number of detected polymorphisms in cultivated germplasm if these methods are used (Matus and Hayes, 2002). Nevertheless, morphological and phenological characteristics are still important measures of genetic variation.

Pedigrees of varieties are defined as a complete documentation of relationships traced back to landraces and wild relatives. Malecot (1948) presented the coefficient of co-ancestry (*f*) as a well recognized kinship coefficient or coefficient of ancestry to measure the relationship based on pedigree information. This measure estimates the likelihood that two randomly drawn, homologous genes (alleles) from each of two individuals are indistinguishable by descent. Melchinger (1993) reported that assessment based on Mendelian inheritance and probability is calculated under a number of assumptions: (i) the absence of selection, mutation, migration, and drift, (ii)

regular diploid meiosis, and (iii) no relationship between individuals without a confirmed common ancestor. Pedigrees have some confines as well such as (i) strong selection, (ii) drift due to small sample size, and (iii) strange or erroneous pedigree records (Messmer *et al*., 1993). Despite these drawbacks, morphological characterization has been extensively used in self-pollinated crop species such as barley, wheat, soybean, and groundnut to study the level of genetic diversity and recognize major groupings of related cultivars (Martin *et al*., 1991). Precise inference of genetic similarity by co-ancestry requires reliable and full pedigree records.

2.6 DNA-based marker systems

Molecular genetic markers, based on DNA sequence polymorphism, offer a powerful tool to accelerate and refine assessment of genetic diversity; therefore they are increasingly being used to complement phenotypic and protein-based markers. Jones *et al*. (1997) defined a molecular marker as a DNA or protein variants which can be detected on marker level and whose inheritance can be monitored reliably. Since markers detect variations among genotypes at DNA level they provide a more direct, reliable, and efficient tool for germplasm conservation and management (Geleta *et al*., 2006). Many types of DNA-based marker systems are available for assessing genetic diversity. They differ in principle, application, amount of polymorphism detected, and cost and time required. DNA-based marker systems have several advantages over other marker types. They can be detected in all tissues at all stages of development and are not affected by the environment (Sorriano *et al*., 2005). DNA-based technologies allow not only the assessment of genetic variability but also individual DNA typing (Bling, 2000). Different marker systems such as restriction fragment length polymorphism (RFLP) (Cui *et al*., 1995; Dubreuil and Charcosset, 1998), random amplified polymorphic DNA (RAPD) (Welsh and McClelland, 1990; Williams *et al*., 1990), amplified fragment length polymorphism (AFLP) (Vos *et al*., 1995), SSRs or microsatellites (Tautz, 1989; Morgante and Olivieri, 1993; Powell *et al*., 1996a), single nucleotide polymorphism (SNP) (Weising *et al*., 2005), and others have been developed and applied.

In general, two different marker systems can be applied; (i) those based on hybridization between a probe and homologous DNA segments inside the genome, and (ii) those that use the polymerase chain reaction (PCR) technique to amplify genome segments between

arbitrary or specific oligonucleotide primer sites (Karp *et al*., 1996; Jones *et al*., 1997; Kumar, 1999).

2.6.1 Restriction fragment length polymorphism (RFLP)

The RFLP assay was the first DNA profiling technique to be widely applied to study plant variation. RFLP analysis involves digestion of genomic DNA with restriction enzymes followed by separation of the resulting fragments using gel electrophoresis and blotting onto nitrocellulose membranes (Southern, 1975). If two individuals differ in distance between sites of cleavage of a particular restriction endonuclease, the length of the fragments produced will differ when the DNA is digested with a restriction enzyme. Specific banding patterns are then visualized by hybridization with a labelled probe, which in most cases is a single copy locus probe that is species specific. To efficiently use RFLP analysis, it is necessary to test many enzymes before polymorphisms can be identified (Beckman and Soller, 1983; Karp *et al*., 1997).

The evolution of chromosomal organization, taxonomic characterization, and the measurement of genetic diversity are some areas of study that have been greatly enhanced by the use of RFLPs (Yang *et al*., 1996). RFLP analysis has been applied in sorghum as well as to other crops to study the level of genetic diversity and the phylogenetic relationships among and between populations, accessions and species (Song *et al*., 1988; 1990; Miller and Tanksley, 1990; Lubbers *et al*., 1991; Aldrich and Doebley, 1992; Demissie *et al*., 1998). RFLP analysis, using genomic single copy probes, has amongst others been used to characterize the variation among wild and cultivated species of *Oryza* (Jena and Kochert, 1991), *Lycopersicon* (Miller and Tanksley, 1990), *Musa* (Gawel *et al*., 1992), sweet potato (Jarret *et al*., 1992) and soybean (Akkaya *et al*., 1992).

In sorghum, RFLP diversity studies on 27 genotypes detected low frequencies of polymorphism (Tao *et al*., 1993). This diversity, however, was higher when maize probes were used during RFLP analysis compared to using isozymes when a set of 56 geographically and racially diverse sorghum accessions were compared (Aldrich and Doebley, 1992). RFLP analysis showed concordance between genetic differentiation and racial classification in cultivated sorghum (Deu *et al*., 1994; 1995; 2006). Cui *et al*. (1995) reported that there was greater nuclear diversity in the wild subspecies of sorghum compared to domestic accessions. Though exceptions were common, especially for the

race *bicolor*, accessions classified as the same morphological race tended to group together on the basis of RFLP similarities (Cui *et al*., 1995).

A large number of DNA probes are available for maize, wheat, sorghum, and soybean, and extensive DNA profiling with RFLP analyses is feasible (Morell *et al*., 1995). RFLP analysis is co-dominant, being able to distinguish homozygous from heterozygous individuals (Helentjaris *et al*., 1985). However, RFLP analysis is expensive, time consuming, technically demanding to assay and require a large amount of high quality DNA (10 µg) (Holton *et al*., 2000). These conditions make RFLP a technique of lower priority. As a result, other marker techniques based on PCR such as RAPDs, AFLPs and SSRs (Jones *et al*., 1997) have been discovered and are preferred.

2.6.2 Polymerase chain reaction (PCR)-based techniques

PCR was invented by Kary B. Mullis in 1985 (Saiki *et al*., 1985) and has revolutionized many areas of biological science. PCR uses the DNA polymerase enzyme which all living cells posses and use to copy their own DNA. The development of thermocyclers that have the ability to change cycling temperatures quickly and accurately, combined with the use of heat-stable DNA polymerases that stay active even after prolonged exposure to high temperatures, have facilitated the automation of this process (Pusterla *et al*., 2006).

The main advantage of PCR-based techniques over RFLP analysis is its inherent simplistic analysis and the ability to conduct PCR tests with extremely small quantities of tissue available for DNA extraction (Edwards *et al*., 1991). Currently, PCR is used worldwide in many areas of biology, agriculture, and medicine (Laurence *et al*., 2009).

2.6.3 Random amplified polymorphic DNA (RAPD)

RAPD analysis uses arbitrary primers, designed without prior knowledge of the designated target DNA sequence, that randomly amplify different regions of the genome (Welsh and McClelland, 1990; Williams *et al*., 1990; Hardys *et al*., 1992). RAPD markers require small amounts of relatively high quality DNA and are cheap and easy to use (Marsan *et al*., 1998). RAPD analysis is dominant and cannot identify heterozygous individuals and therefore has a limitation for intra-population genetic analysis (Holton *et al*., 2000). Futhermore, RAPD analysis is not reproducible and reliable (Marsan *et al*., 1998) and lack allelism (alleles cannot be attributed to loci) (Jarne and Theron, 2001).

However, this method has an advantage of detecting higher levels of polymorphism compared to isoenzymes (Fernandez *et al*., 2002). This method has been used to study diversity among wild species of *Hordeum* (Gonzalez and Ferrer, 1993), Indian mustard (*Brassica juncea* L.) (Jain *et al*., 1994), and rice (Mackill, 1995).

2.6.4 Amplified fragment length polymorphism (AFLP)

AFLP analysis is a multi-locus marker technique developed by Vos *et al.* (1995) based on the selective PCR amplification of restriction fragments from a total digest of genomic DNA. Antonio *et al*. (2004) reported that the AFLP technology has the ability to detect a large number of polymorphic fragments in a single lane rather than high levels of polymorphism at each locus such as in the case of the SSR method. AFLP analysis has a higher efficiency in detecting polymorphism than either RAPD or RFLP markers (Garcia-Mas *et al*., 2000) and has greater reproducibility than RAPD analysis (Powell *et al*., 1996b; Russell *et al*., 1997), which has led to its increased use in DNA profiling (Maughan *et al*., 1996; Powell *et al*., 1996a; Maheswaran *et al*., 1997). There are many applications of AFLP markers, genetic relationship studies being an important one (Schut *et al*., 1997; Aggarwal *et al*., 1999; Breyne *et al*., 1999; Singh *et al*., 1999; Incirli and Akkaya, 2001; Negash *et al*., 2002). The AFLP technique has been used to estimate genetic diversity in both cultivated and natural/rare populations (Hill *et al*., 1996; Lu *et al*., 1996; Sharma *et al*., 1996; Travis *et al*., 1996; Karp *et al*., 1997; Paul *et al*., 1997; Kiambi *et al*., 2005). AFLP analysis has also been used in genome mapping (Zimnoch-Guzowska *et al*., 2000), DNA fingerprinting (Powell *et al*., 1996b; Fleischer *et al*., 2004), and parentage analysis (Lima *et al*., 2002).

The suitability of AFLP analysis for cultivar identification is demonstrated by the large number of reports published on the use of the technique for genotype identification in a variety of plant species, such as Brassica, sunflower, pepper, soybean, sugar beet, lettuce, tomato (Perkin-Elmer, 1996), wheat (Donini *et al*., 1997), and barley (Pakniyat *et al*., 1997). However, Gerber *et al*. (2000) reported that AFLP analysis has lower sensitivity in detecting informative genotypic classes which might be associated with the inability to distinguish heterozygotes from homozygotes because of the dominant nature of AFLP analysis.

2.6.5 Microsatellites or simple sequence repeats (SSRs)

Microsatellites are also known as SSRs and are DNA sequences with repeat lengths of a few base pairs (2-6 bp). Mahalakshmi *et al*. (2002) reported that they are ubiquitously distributed throughout the genome of eukaryotes and abundant in genomes of plants where they are thought to be a source of genetic variation. SSRs tend to occur in non-coding regions of DNA and are flanked on each side of the repeat unit by "unordered" DNA. The flanking sequences at each of these sites are often unique. Specific primers can be designed according to the flanking sequences, which then result in single locus identification. Variation in the number of repeats can be detected with PCR and alleles that differ in length can be resolved using agarose gels or sequencing gels where single repeat differences can be resolved and all possible alleles detected (Saghai-Maroof *et al*., 1994).

Poulsen *et al*. (1993) and Kresovich *et al*. (1995) did initial research on the isolation and characterization of SSRs in cultivated *Brassica* species. SSR markers have been developed in pigeon pea and have been used to assess the degree and distribution of genetic diversity in landraces from Andhra Pradesh (Bramel *et al*., 2004; Buhariwalla and Crouch, 2004; Newbury *et al*., 2004). In sorghum, Tunstall *et al.* (2001) assessed the degree and distribution of genetic diversity in landraces from north Shewa and south Welo, Ethiopia. SSR markers have proved to be a valuable asset for breeding programmes and have been used for a wide range of applications, mostly in measuring genetic diversity (Xiao *et al*., 1996), and assigning lines to heterotic groups (Senior *et al*., 1998). SSRs have been used in genetic distance analysis (Chen *et al*., 1997), genetic analysis of breeding schemes (Kejun *et al*., 2003), estimation of genome size (Smith *et al*., 1997), population genetics (Zhang and Hewitt, 2003; Ellis and Burke, 2007), fingerprinting for legal protection of cultivars and parental lines (Kumar, 1999), and in establishing genome relationships in species with putative inter-specific parents (Dweikat, 2005). Polymorphisms have been observed with this kind of marker in loquat (Sorriano *et al*., 2005), groundnut (Krishna *et al*., 2004), perennial ryegrass (Kubik *et al*., 2001), rice (Liu *et al*., 2000) and maize (Senior and Heun, 1993; Senior *et al*., 1998). SSRs have also been found to occur in other plant genomes including soybean (Akkaya *et al*., 1992), barley (Saghai-Maroof *et al*., 1994), sorghum, and pearl millet (Taramino *et al*., 1997).

Fregene *et al.* (2003) concluded that variation in allele frequencies at many unlinked loci is the preferred method of assessing genetic diversity and differentiation and estimation of the strengths of the various forces shaping them. SSR markers are particularly attractive for studying genetic differentiation because they are co-dominant and abundant in plant and animal genomes (Folkertsma *et al*., 2005).

2.6.5.1 Advantages and limitations of microsatellites as genetic markers

SSR analysis is relatively simple and can be automated (Kresovich *et al.*, 1995; Mitchell *et al*., 1997). Most SSR markers are locus-specific (in contrast to multi-locus markers such as minisatellites or RAPDs) and show Mendelian inheritance (Saghai-Maroof *et al*., 1994). Rafalski and Tingey (1993) reported that SSRs are highly informative and PCR-based, implying that only tiny amounts of tissue are needed and even highly degraded or "ancient" DNA can be used. Due to the co-dominant nature of microsatellites, heterozygotes can be distinguished from homozygotes, in contrast to RAPD and AFLP markers which are mainly dominant markers. In addition, SSRs are highly polymorphic (Weber, 1990; Doldi *et al*., 1997; Schug *et al*., 1998) and thus the level of polymorphism in plant species studied has been greater than that found with other markers.

In sorghum, numerous SSR markers have been developed and mapped (Brown *et al*., 1996; Taramino *et al*., 1997; Bhattramakki *et al*., 2000; Kong *et al*., 2000; Schloss *et al*., 2002). However, a low number of public domain markers have been employed to analyze the genetic diversity in subsets constituted from ICRISAT (Grenier *et al*., 2000b), and USDA sorghum collections (Dean *et al*., 1999), and collections originating from single countries (Dje *et al*., 1999; Ghebru *et al*., 2002).

Despite their efficiency, SSRs have some limitations. SSR markers are time consuming and costly to develop in that the genomic regions carrying them must be identified and sequenced. They are probably rarely useful for higher-level systematics due to a too high mutation rate. Across highly divergent taxa, two problems arise. Firstly, the SSR primer sites may not be conserved (the primers used for species A may not even amplify in species B). Secondly, the high mutation rate means that homoplasy becomes much more likely; one can no longer safely assume that the two alleles identical in state are identical by origin (Spooner *et al*., 2005).

2.6.6 Diversity array technology (DArT)

Diversity array technology (DArT) is a new genotyping method that offers the highest throughput genotyping available to date. DArT is a complexity reduction, DNA hybridization-based method that simultaneously assays hundreds to thousands of markers across a genome. DArT preferentially targets low-copy genomic regions, allows automation of data acquisition and is cost competitive. Although developed some years ago, this marker technology has recently gained increasing attention (Wenzl *et al*., 2004; Tinker *et al*., 2009). However, DArT loci, due to being treated as dominant markers, limit the genetic information provided by a given locus. Huttner *et al*. (2005) documented that DArT fingerprints are useful for accelerating plant breeding, and for characterization and management of genetic diversity in domesticated species as well as in their wild relatives. The DArT genotyping method was originally developed for rice (Jaccoud *et al*., 2001) and applied to many other plant species, including barley (Wenzl *et al*., 2006), cassava (Xia *et al*., 2005), Arabidopsis (Wittenberg *et al.*, 2005), pigeon pea (Yang *et al*., 2006), wheat (Akbari *et al*., 2006), and sorghum (Mace *et al*., 2008). Mace *et al.* (2008) conducted a study to analyze a diverse set of sorghum genotypes using more than 500 markers which detected variation among 90 accessions used in the diversity analysis, and cluster analysis discriminated well among all 90 genotypes. Consequently they effectively developed DArT markers for *S. bicolor* and demonstrated that DArT provides high quality markers that can be used for diversity analyses and to create medium-density genetic linkage maps. The high number of DArT markers generated in a single assay not only provides an accurate estimate of genetic relationships among genotypes, but their even allocation over the genome also offers actual advantages for a range of molecular breeding and genomic applications (Akbari *et al*., 2006).

2.7 Molecular markers applied in sorghum germplasm

Previously reported methods based on molecular markers that have been used to study genetic diversity in sorghum germplasm include allozymes (Morden *et al.*, 1989; Ollitrault *et al*., 1989; Aldrich *et al*., 1992), mitochondrial DNA (Deu *et al*., 1995), nuclear RFLP (Deu *et al*., 1994; 1995; 2006; Cui *et al*., 1995), chloroplast DNA (Aldrich and Doebley, 1992), RAPD markers (De Oliveira *et al*., 1996; Menkir *et al*., 1997; Ayana *et al*., 2000a; 2000b), AFLP (Uptmoor *et al.*, 2003, Menz *et al*., 2004, Perumal *et al.*, 2007), and SSR analysis (Brown *et al*., 1996; Taramino *et al*., 1997; Dean *et al*., 1999;

Dje *et al*., 1999; Bhattramakki *et al*., 2000; Grenier *et al*., 2000b; Kong *et al*., 2000; Ghebru *et al*., 2002; Schloss *et al*., 2002).

Based on allozyme variation Morden *et al*. (1990) failed to offer any understandable taxonomic differentiation among species of the sub-generic section *Eusorghum* as proposed by De Wet (1978). Morden *et al.* (1990) studied the variation among 90 genebank accessions of wild congeners of cultivated sorghum in this section originating from Africa, India, and Thailand. Results might be attributed to a combination of low levels of marker polymorphism and insufficient sampling of *S. halepense* and *S. x almum* (Morden *et al.*, 1990). Their work further revealed higher levels of diversity in the wild gene pool compared to cultivated sorghum based on a comparison of the allozymic variation of S. *bicolor* ssp. *verticilliflorum* with that of cultivated *S. bicolor* spp. *bicolor* (Morden *et al.*, 1989)

Aldrich and Doebley (1992) performed a similar study to evaluate 56 accessions focusing on the geographical and racial diversity represented in cultivated sorghum (ssp. *bicolor*) and it's proposed wild progenitor (ssp. *verticilliflorum*) using nuclear and chloroplast DNA (cpDNA) RFLP analysis. They detected higher levels of nuclear diversity within wild sorghum compared to cultivated sorghum, as well as an obvious genetic variation between the two. In addition, the nuclear diversity of cultivated sorghum was found to be well encompassed within the wild sorghum gene pool. They moreover observed that nuclear diversity of the wild sorghum gene pool from north-eastern Africa was comparatively closer to cultivated sorghum.

Cui *et al.* (1995) confirmed the earlier hypothesis that central-north eastern Africa is the most likely principal area of domestication of sorghum based on observations in their RFLP analysis study on cultivated and wild genebank accessions originating from Africa, Asia, and the USA. Their results indicated that morphological races were only slightly differentiated from each other (only about 10% of genetic variation among races), while considerable genetic diversity was observed among accessions within races, and among geographical groups. On the other hand, only genebank accessions from worldwide origins were used in these studies, and the *in situ* pattern of genetic diversity at a regional scale remains unknown (Dje *et al*., 1999).

A comparative genetic diversity study of cultivated and wild sorghum using mitochondrial DNA (mtDNA) markers by Deu *et al.* (1995) indicated that domestication occurred from *S. bicolor* ssp. *verticilliflorum*, followed by diversification in cultivated sorghum in different geographic areas under different environmental and human selection pressures. These conclusions were in line with the hypothesis by Harlan *et al.* (1976).

Nkongolo and Nsapato (2003) used 35 RAPD primers to study the genetic variation within and among several sorghum populations from different agro-ecological zones in Malawi. Results indicated that sorghum accessions were genetically closely related despite considerable phenotypic diversity within and among accessions. Furthermore, Ayana *et al*. (2000a) assessed the extent of genetic variation among 80 sorghum accessions from Ethiopia and Eritrea using 20 RAPD primers and detected limited variation among accessions. Ayana *et al*. (2000b) performed RAPD analysis on wild sorghum germplasm collected *in situ* from five regions of Ethiopia, using nine decamer primers and detected low to moderate genetic variation among populations. Dahlberg *et al*. (2002) investigated variation among sorghum germplasm using seed morphology and RAPD analysis and grouped 94 accessions into four major races. Agrama and Tuinstra (2003) compared the phylogenetic relationship among 22 sorghum accessions using 32 RAPD primers and detected low levels of polymorphism among them compared to using SSR analysis. In India, Prakash *et al*. (2006) assessed the genetic diversity among 32 sorghum lines from local and exotic sorghum germplasm using 64 RAPD primers. They found that most primers were polymorphic, informative and differentiated accessions. Cluster analysis grouped the 32 sorghum accessions into two major clusters.

Perumal *et al.* (2007) examined 46 converted exotic sorghum lines. Nine intermediate races of sorghum were fingerprinted using AFLP analysis in order to calculate genetic similarities between lines. They found that *caudatum* and intermediates involving *caudatum* showed a close genetic relationship with *durra* and *durra* intermediates. Morphological classification of races based on panicle traits was mostly reflected by similarity in DNA-based polymorphisms. The molecular diversity of *bicolor* and associated intermediate races was not reflective of their common morphological classification, since this race and its intermediates are quite heterogeneous.

Studies were carried out by Menz *et al*. (2004) to determine the genetic diversity of public inbreds of sorghum using mapped AFLPs. Their efforts failed to give a clear separation between B- and R-lines, suggesting that B- and R-lines did not represent well-defined heterotic groups in this set of public lines. On the other hand, cluster analysis of genetic similarity estimates revealed that classification of sorghum inbreds was based on the sorghum working groups, *zera-zera*, *kafir*, *kafir-milo*, *durra*, and *feterita*. To get an overview on the genetic relatedness of sorghum landraces and cultivars grown in low-input conditions of small-scale farming systems from southern Africa, Uptmoor *et al.* (2003) examined 46 sorghum accessions using AFLPs. UPGMA (unweighted pair-group method using arithmetic averages) clustering divided accessions into main clusters comprising landraces on the one hand and newly developed varieties on the other hand. Further sub-groupings were not unequivocal. Genetic diversity was estimated on a similar level within landraces and breeding varieties.

SSRs were also used to study genetic diversity in sorghum (Brown *et al*., 1996; Dean *et al*., 1999; Dje *et al*., 1999; 2000; Grenier *et al*., 2000b; Smith *et al*., 2000; Ghebru *et al*., 2002; Abu Assar *et al*., 2005). Results from these studies suggested that SSR markers were suitable for applications relevant to conservation and use of sorghum germplasm. Recently SSR markers have been used to study sorghum diversity in *in situ* collections and to investigate the evolutionary process that influences patterns of genetic diversity at regional, national, and local spatial scale (Barnaud *et al.*, 2007; Deu *et al.*, 2008; Sagnard *et al.*, 2008). Deu *et al.* (2008) used 28 SSR markers to perform a genetic diversity survey on 484 sorghum samples collected from 79 villages across Niger in order to understand the geographical, environmental, and social patterns of genetic diversity on different spatial scales. They detected high levels of genetic diversity that was differentiated along sorghum botanical races, geographical distribution and ethnic groupings of farmers, but low along climatic zones.

In northern Cameroon, Barnaud *et al.* (2007) used 14 SSR markers to characterize 21 sorghum landraces collected at village level among the Duupa farmers. Their results revealed significant genetic differentiation between landraces, probably due to (i) some form of barrier to inter-landrace gene flow and seed selection by farmers, (ii) existence of different mating systems among landraces, and (iii) historical factors and farmers' practices that affected patterns of genetic variation. Concerning farmers' practices,

Barnaud *et al.* (2008) concluded that selection exerted by farmers was a key factor for determining the possibility of new genetic combinations from outcrossing events and accordingly in patterns of genetic differentiation among landraces. Other studies have shown variable, though extensive outcrossing rates among landraces. In a recent study, multiple spatial scale analysis carried out by Sagnard *et al.* (2008) to characterize the evolutionary forces that shaped genetic diversity of cultivated sorghum collected *in situ* from Burkina Faso, Mali, Niger, and in the same village in Cameroon sampled by Barnaud *et al.* (2007; 2008) demonstrated no evidence of spatial genetic structure among villages separated by more than 30 km. This indicated that conventional seed exchange systems in West Africa is separate at local scale. Regarding genetic diversity between countries, Niger was found to be genetically richer than Mali, despite the fact that Mali grew sorghum in a larger agro-climatic range than Niger. These findings confirmed that the diversity of human groups acted together with the agro-ecological factors to form the composition of sorghum genetic diversity (Sagnard *et al.,* 2008).

SSR markers were used to measure and characterize diversity in a collection of gene bank accessions of cultivated and wild sorghum in a comparative genetic study by Casa *et al.* (2005) and indicated that landraces retained up to 86% of the diversity observed in wild sorghums. Genetic variation between cultivated and wild populations was found to be reasonable while little evidence was obtainable for racial differentiation in wild forms (Casa *et al*., 2005).

Because of its comprehensive socio-economic value, there has been an increase in characterizing levels of genetic diversity within sorghum using both phenotypic and molecular markers (Dean *et al*., 1999; Dje *et al*., 2000; Grenier *et al*., 2000a; 2000b; Ghebru *et al*., 2002; Uptmoor *et al*., 2003; Menz *et al*., 2004; Shehzad *et al*., 2009). These analyses have provided the establishment of genetic data for making informed decisions concerning management and utilization of genetic resources.

2.8 Comparisons based on morpho-agronomical and molecular markers

Studying the diversity of pre-breeding and breeding germplasm, and determining the uniqueness and distinctness of the phenotypic and genetic constitution of genotypes, is important to protect the plant breeder's intellectual property rights (Franco *et al*., 2001).

For conservation, evaluation, and utilization of genetic resources, different types of characters are frequently measured in each genotype: (i) quantitative characters (morpho-agronomical), (ii) qualitative characters (these are usually multi-state variables), and (iii) discrete genetic marker characteristics using for example SSR, RFLP and/or AFLP data (Franco *et al*., 1997).

When morpho-agronomical and genetic marker data are available for a set of genotypes, two types of hierarchical classifications are performed separately, namely hierarchical clustering based on the morpho-agronomical traits in which a standard metric distance (such as the squared Euclidean) is computed and a clustering strategy, such as Ward or UPGMA, is applied. The second one is obtained based on the genetic marker attributes when genetic similarities (or dissimilarities) of individuals are determined with molecular markers. Through applying any clustering strategy (such as single or complete linkage, UPGMA, centroid method, Ward method, etc.), genotypes can be clustered into groups that are as homogeneous as possible and heterogeneous among groups. Franco *et al*. (2001) reported that, in general results showed that groups formed based on both continuous and categorical classifications had low to medium consensus.

Many researchers proposed different models for combining the two data sets. Wolfe (1970) used non-hierarchical statistical methods for classifying individuals including mixture models, such as the Gaussian Model (GM) (which only deals with continuous variables). On the other hand, Franco *et al*. (1997) proposed the use of hierarchical methods such as Ward (or UPGMA), using Gower's distance.

2.9 Correlation between phenotypic and molecular marker distance

Relationships between phenotypic and molecular marker data of genotypes may provide useful information in order to determine the most promising entries for future breeding programmes. A few studies on the relationship between phenotypic and molecular data performance have been published in sorghum (Anas and Yoshida, 2004; Geleta *et al*., 2006) in which low correlations between molecular marker and phenotypic data in sorghum were reported.

Molecular markers have the ability to cover the entire genome (coding as well as non-coding regions), while phenotypic differences are connected to specific genes or coding

regions. Therefore, differences revealed by molecular markers are not necessarily associated with phenotypic variation. Hence, to express accurately the relationships among genotypes, a combination of morphological and molecular information is required.

Burstin and Charcosset (1997) investigated the relationship between molecular and phenotypic distances computed from quantitative traits. Their results revealed that the relationship is most likely triangular, which means close genetic distances are associated with close phenotypic distances, whereas distant genetic relationships can correspond with both close and distant phenotypic relationships. In many cases, the relationship between distances based on morphology and molecular markers are not easy to understand. As a result, a combination of morphological and molecular analyses may be the most useful to understand all aspects of genetic variation within a species or populations.

2.10 Measures of genetic variation

In general, measures of genetic variation of a sample depend mainly on the number of individuals sampled per population, the number of loci sampled, genotypic and allelic compositions of the population, mating system, and effective population size (Weir, 1990). Nei (1987) demonstrated that a large portion of the sampling variance of diversity estimates is due to variation of diversity levels among loci across the genome. Frankel *et al.* (1995) recommended that two different models of genetic variation are applicable at population level: (i) "richness" of any population or sample from it, related to the total number of genotypes or alleles present in the population, and (ii) "evenness" or the frequency of different types or alleles in the population or samples analyzed.

Sjogren and Wyone (1994) reported that allele richness is estimated by the mean number of alleles per locus and percentage polymorphic loci. This is sensitive to the presence or absence of distinct or rare alleles (5% or lower in frequency) in a population, as a high level of sampling error could be connected with detection of such alleles. Therefore, in addition to the total number, it would be useful to observe the number of alleles in the sample above a frequency threshold (say 5%). The percentage of polymorphic loci in a population is a basic measurement of genetic variation, as it is subjected to a large genomic sampling error; this estimate is reliable only when a large number of loci are sampled (Brown and Weir, 1983).

The evenness of allele or genotype frequencies is accounted by the measures of average observed heterozygosity, expected heterozygosity, and effective number of alleles. Heterozygosity is the most widespread measure of genetic variation within a population. Expected heterozygosity (*He*), also known as gene diversity, is a concept introduced by Nei (1978) to explain the probability by which two alleles that were arbitrarily selected are different, and is particularly useful because it is applicable to genes of different ploidy levels and in organisms of different reproductive systems (Hedrick, 2005). None of these measures are sensitive to the sampling error associated with rare alleles.

2.11 Types of distance measures

Various genetic distance measures have been proposed for analysis of molecular marker and morphological data for the purpose of genetic diversity analysis. Genetic distances can be calculated by different statistical measures depending on the data set. Dissimilarity coefficients estimate the distance or difference of two individuals and the bigger the values, the more diverse the two individuals, while similarity indices measure the similarity between two individuals. The bigger the value the more related the two individuals are (Kosman and Leonard, 2005). Kaufman and Rousseeuw (1990) reported that Euclidean distance, and square Euclidean distance are the most commonly used measures for morphological data to estimate genetic distance (GD) between individuals, whereas Gower's distance (Gower, 1971) can be used to measure genetic distance between individuals on the basis of different types of characters, such as qualitative and quantitative.

For molecular marker data, one useful similarity index is that of Nei and Li (1979):

$GD = 1-[2N_{xy}/N_x+N_y]$,

Here $2N_{xy}$ is the number of shared bands, and N_x and N_y are the number of bands observed in individual x and individual y, respectively. Other similarity indices such as the commonly used measures of genetic distance or genetic similarity (GS) using binary data namely (i) Jaccard's (1908) coefficient, (ii) simple matching coefficient (Sokal and Michener, 1958), (iii) modified Rogers' distance (Rogers, 1972), and (iv) Gower's similarity coefficients (Gower, 1971) have been extensively used in genetic distance determination (Barrett and Kidwell, 1998).

Mantel’s test is used to examine the relationship between matrices derived from different distance measures. Mantel’s test is a regression that can be applied on dissimilarity or similarity matrixes of different types of variables (Mantel, 1967). Mantel's test of significance is evaluated via permutation procedures to overcome the problem of dependent elements (Manly, 1991).

Mantel’s test is important in analysis of genetic diversity, where various data sets may be used to assess the relationships among different individuals or populations. Mantel’s test has been used in analysis of genetic diversity in many different crop plants (Kumar *et al*., 2008; Sarıkamış *et al*., 2009; Priolli *et al*., 2010).

2.12 Multivariate analysis methods

The pattern of genetic relationship among accessions can be conveniently shown by multivariate techniques such as cluster analysis, principal component analysis (PCA), principal coordinate analysis (PCoA), and multidimensional scaling (MDS).

2.12.1 Cluster analysis

Clustering is a useful tool for studying relationships among closely related cultivars or accessions. Kaufman and Rousseeuw (1990) reported that cultivars or accessions in cluster analysis are arranged in hierarchy by agglomerative algorithms according to the structure of a complex pairwise genetic proximity measure. Hierarchies emerging from cluster analysis are highly dependent on the proximity measures and clustering algorithm used. Hair *et al*. (1995) defined cluster analysis as a group of multivariate techniques whose main point is to group individuals based on the characteristics they have. That means individuals with similar descriptions are grouped into the same cluster. The resulting clusters of individuals in that case reveal high internal (within cluster) homogeneity and high external (between clusters) heterogeneity.

Generally, there are two types of clustering methods: (i) distance-based methods, in which a pair-wise distance matrix is used as an input for study by a specific clustering algorithm (Johnson and Wichern, 1992), and (ii) model-based methods, in which analysis from each cluster is assumed to be unsystematic, draws from some parametric model, and inferences about parameters related to each cluster and cluster association of each individual are

performed equally using standard statistical methods such as maximum-likelihood or Bayesian methods (Pritchard *et al*., 2000).

Distance-based clustering methods are classified into two groups: hierarchical and nonhierarchical. Hierarchical clustering methods are more frequently used in analysis of genetic diversity in crop species. These methods are performed on either a sequence of consecutive mergers or a series of successive divisions of group of individuals. The most related individuals are firstly grouped and these first groups are pooled according to their similarities. Amongst different agglomerative hierarchical methods, UPGMA (Sneath and Sokal, 1973; Panchen, 1992) is the most commonly used, followed by the Ward's minimum variance method (Ward, 1963). The non-hierarchical methods referred to as K-means clustering measures do not occupy the structure of dendrograms or trees and are based on chronological threshold, similar threshold or optimizing approaches for conveying individuals to specific clusters (Everitt, 1980).

2.12.2 Principal component analysis (PCA)

PCA, as defined by Wiley (1981), is a technique of statistical decrease to describe relations among two or more characters and to split the total difference of the novel characters into a partial number of uncorrelated new variables. The decrease is created by linear conversion of the original variables into a new set of uncorrelated variables known as principal components (PCs). The initial step in PCA is to estimate eigen values, which explain the amount of total dissimilarity that is displayed on the PC axes. The first PC summarizes most of the unpredictability present in the original data relative to all residual PCs. The second PC describes most of the variability not summarized by the first PC and uncorrelated with the first, and so on (Jolliffe, 1986).

Wiley (1981) reported that PCA can be applied to two forms of data matrices: (i) a variance-covariance matrix, and (ii) a correlation matrix. In the use of the variance-covariance matrix, absolute changes among individuals can be studied. However, with the association matrix, only differences comparative to the consistent data can be interpreted.

2.12.3 Principal coordinate analysis (PCoA)

PCoA is an ordination technique that begins with a matrix of similarities or dissimilarities between a set of individuals and aims to create a low-dimensional graphical plot of the

statistics in such an approach that distances between points in the plot are close to novel dissimilarities. Rohlf (1972) recognized that in PCoA, the treatment of missing information is more reasonable than that in PCA.

2.12.4 Multidimensional scaling (MDS)

MDS is a procedure that represents a set of individuals or genotypes (n) in a few dimensions (m) using a similarity/distance matrix between them (Johnson and Wichern, 1992). There are two types of MDS: (i) non-metric MDS, which is used when the inter-individual proximities in the map nearly match the original similarities/distances, and (ii) metric MDS, helpful when the real scales of original similarities/distances are used to get an arithmetical representation in m dimensions (Johnson and Wichern, 1992). The closeness between original similarities-distances and inter-individual proximities in the map can be tested by different methods. The most commonly used test is a numerical measure of closeness called "stress", which shows the percentage of the variance of the disparities not accounted for by the MDS model.

Rohlf (1972) reported that the actual arrangement of individuals consequential from PCA, PCoA, and MDS are typically related. On the contrary, results based on MDS contrast with PCA and PCoA since (i) differences among close individuals are, in common, reflected better by MDS, and (ii) the smaller or greater distances among individuals are not essentially represented by MDS to the equivalent scale. MDS is preferable over PCA and PCoA when the number of individuals is large (Rohlf, 1972). Simply, if there are no missing data or many more individuals than characters, PCA should be employed.

2.13 Diversity and differentiation

Diversity and differentiation are estimators of the genetic variation that are broadly used in population genetics studies. Natural selection and genetic drift favour genetic differentiation between populations which is associated with high levels of endogamy and low genetic diversity. On the other hand, gene flux enhances homogenization of genetic diversity between populations, which is associated with low differentiation and high genetic diversity. Several approaches have been used to estimate the level of differentiation in sub-division of a population (Holsinger and Weir, 2009)

2.13.1 F-statistics

F-statistics (also known as fixation indices) elucidate the level of heterozygosity in a population; in particular the degree of a reduction in heterozygosity compared to the Hardy-Weinberg expectation. F-statistics can also be thought of as a measure of the correlation between genes drawn at different levels of a (hierarchically) subdivided population. This correlation is influenced by several evolutionary processes, such as mutation, migration, inbreeding, and natural selection (Wright 1951; 1965). These three coefficients, F_{ST}, F_{IT}, and F_{IS} are interrelated so that

$$F_{ST} = F_{IT}\text{-}F_{IS}/(1\text{-}F_{IS})$$

F_{ST} is a measure of the genetic differentiation over subpopulations and is always positive. F_{IS} and F_{IT} are measures of the deviation from Hardy-Weinberg equilibrium within sub-populations and in the total population, respectively. To interpret the F_{ST} values, Wright (1978) suggested a qualitative guide in which F_{ST} values show a fluctuation: F_{ST} values of 0 to 0.05 indicate low genetic differentiations, F_{ST} values between 0.05 and 0.15 indicate moderate genetic differentiation among groups, whereas F_{ST} values between 0.15 to 0.25 and above 0.25 indicate high and very high genetic differentiation, respectively (Hartl and Clark 1997).

These statistical estimators partition the genetic variability, measured by levels of heterozygosity, into components of variability among and within populations. Furthermore, they provide a way to estimate gene flux patterns among populations (Kourti, 2002).

2.13.2 Analysis of molecular variance (AMOVA)

In order to identify variation at molecular level an analysis of molecular variance (AMOVA) is used to detect statistical differences among pre-defined phenotypic groups (Excoffier *et al*., 1992). This procedure uses the square distance between molecular marker profiles as data to perform a permutation analysis of variance, which is used to test the significance of the variance components and Φ-statistics, consequently eliminating the normality assumption necessary under conventional analysis of variance (ANOVA) analysis but unsuitable for molecular data (Excoffier *et al*., 1992).

AMOVA (Excoffier *et al*., 1992) has been widely used for estimation of variance components among and within groups (Warburton *et al*., 2002; Reif *et al*., 2003). AMOVA is based on squared Euclidean distances among individuals, and assumes that the studied populations are in Hardy-Weinberg equilibrium (Excoffier *et al*., 1992).

2.14 Conclusions

Estimating genetic diversity and defining the relationships between germplasm collections helps guarantee that germplasm is adequately collected and managed. Natural ecosystems are progressively better managed by humans with the rationale of maintaining the existing genetic diversity. This diversity is considered an assurance against disastrous damage and a resource for future human use. It is well recognized that biological diversity contributes to the robustness and sustainability of agricultural production systems, mainly in developing countries where public support to farmers in times of emergency is limited or absent (Wenzl *et al*., 2004). Comprehensive knowledge of genetic diversity in cultivated and wild sorghum germplasm, the source of novel genomic regions, novel alleles and novel traits, is important for plant genetic resources conservation and crop improvement programmes (Xiao *et al*., 1996).

Determining the genetic diversity of different accessions at DNA level holds many advantages for the plant breeder, because it may increase the effectiveness of breeding efforts to improve crop species (Barrett and Kidwell, 1998; Geleta *et al.,* 2006). This may elucidate the motivation for the development of different marker techniques. Information that is obtained with these new tools can be used to contribute to enhanced food production throughout the world. Information of genetic diversity of a germplasm collection can increase the efficiency of efforts to improve a species (Geleta *et al.,* 2006).

SSRs have been successfully applied to estimate the genetic diversity in sorghum germplasm (Smith *et al*., 2000; Uptmoor *et al*., 2003; Anas, 2004; Casa *et al*., 2005; Folkertsma *et al*., 2005; Ali *et al*., 2008; Shehzad *et al*., 2009). Morphological traits are important estimates of genetic diversity in crops but have some limitations (Alamnza-Pinzon *et al*., 2003; Fufa *et al.,* 2005). However, morphological characterization is the first step in the description and classification of germplasm (Smith and Smith, 1992). A combination of morphological and molecular analyses may be useful to understand all aspects of genetic variation within a species or population.

Sorghum was domesticated in Africa and is a significant factor of food security for more than 250 million people on the continent today (AHBFI, 2007). It is an attractive crop for plant breeders due to its adaptation to harsh environments, diverse germplasm collections, and relatively small genome size (Menz *et al*., 2002).

Eastern and central Africa is still affected by frequent drought. As a result, numerous nations are at risk because of a reduction in sorghum production in both rainfed and irrigation areas. Sorghum is cultivated under equally diverse agro-climatic conditions and practices by subsistence farmers in different communities in Africa and therefore, has great potential in providing food security in the region. Conversely, many valuable landraces of sorghum moreover have been lost or are under serious risk. Consequences of these losses are a high risk of genetic erosion. The expansion of high yielding and stable varieties necessitate an incessant supply of new germplasm as a resource of desirable genes and/or gene complexes. The most important sources of such genes are landraces, introductions, and weedy and wild relatives of the crop plants. Hence, it is necessary to study the genetic diversity and genetic relationships in germplasm accessions among east African countries using morphological and molecular markers in order to (i) understand the distribution of genetic variation in different countries, (ii) better conserve the genetic variation contained in them, and (iii) facilitate their use in new, dedicated breeding programmes for sorghum accessions.

2.15 References

Abu Assar AH, Uptmoor R, Abdelmula AA, Salih M, Ordon F, Friedt W. 2005. Genetic variation in sorghum germplasm from Sudan, ICRISAT, and USA assessed by simple sequence repeats (SSRs). Crop Science 45:1636-1644.

Aggarwal RK, Brar DS, Nandi S, Huang N, Khush GS. 1999. Phylogenetic relationships among *Oryza* species revealed by AFLP markers. Theoretical and Applied Genetics 98:1320-1328.

Agrama HA, Tuinstra MR. 2003. Phylogenetic diversity and relationships among sorghum lines using SSRs and RAPDs. African Journal of Biotechnology 2:334-340.

AHBFI. Africa Harvest Biotech Foundation International. 2007. A global vision with an African focus to fight poor nutrition with nutrient-rich crops. The Africa Biofortified Sorghum Project: Mid-Term Report, December 2007. Nairobi, Kenya; Johannesburg, South Africa; Washington DC, USA, pp. 40.

Akbari M, Wenzl P, Caig V, Carling J, Xia L, Yang S, Uszynski G, Mohler V, Lehmensiek A, Kuchel H, Hayden MJ, Howes N, Sharp P, Vaughan P, Rathnell B, Huttner E, Kilian A. 2006. Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. Theoretical and Applied Genetics 113:1409-1420.

Akkaya MS, Bhagwat AA, Cregan PB. 1992. Length polymorphisms of simple sequence repeat DNA in soyabean. Genetics 132:1131-1139.

Alamnza-Pinzon MI, Khairullah M, Fox PN, Warburton ML. 2003. Comparison of molecular markers and coefficients of parentage for the analysis of genetic diversity among spring bread wheat accessions. Euphytica 130:77-86.

Aldrich PR, Doebley J. 1992. Restriction fragment variation in the nuclear and chloroplast genomes of cultivated and wild *Sorghum bicolor*. Theoretical and Applied Genetics 85:293-302.

Aldrich PR, Doebley J, Schertz KF, Stec A. 1992. Patterns of allozyme variation in cultivated and wild *Sorghum bicolor*. Theoretical and Applied Genetics 85:451-460.

Ali ML, Rajewski JF, Baenziger PS, Gill KS, Eskridge KM, Dweikat L. 2008. Assessment of genetic diversity and relationship among a collection of US sweet sorghum germplasm by SSR markers. Molecular Breeding 21:497-509.

Ali ML, Sanchez PL, Yu S, Lorieux M, Eizenga, GC. 2010. Chromosome segment substitution lines: A powerful tool for the introgression of valuable genes from *Oryza* wild species into cultivated rice (*O. sativa*). Rice 3:218-234.

Anas YT. 2004. Genetic diversity among Japanese cultivated sorghum assessed with simple sequence repeat markers. Plant Production Science 7:217-223.

Anas YT, Yoshida T. 2004. Genetic diversity among Japanese cultivated sorghum assessed with simple sequence repeat markers. Plant Production Science 7:217-223.

Antonio AF, Garcia LL, Benchimol AM, Barbosa IO, Geraldi CL, de Souza CL Jr, de Souza, AP. 2004. Comparison of RAPD, RFLP, AFLP and SSR markers for diversity studies in tropical maize inbred lines. Genetic and Molecular Biology 27:579-588.

Ayana A, Bekele E, Bryngelsson T. 2000a. Genetic variation in wild sorghum (*Sorghum bicolor* ssp. *verticilliflorum* (L.) Moench) germplasm from Ethiopia assessed by random amplified polymorphic DNA (RAPD). Hereditas 132:249-254.

Ayana A, Bryngelsson T, Bekele E. 2000b. Genetic variation of Ethiopian and Eritrean sorghum (*Sorghum bicolor* (L.) Moench) germplasm assessed by random amplified polymorphic DNA (RAPD). Genetic Resources and Crop Evolution 47:471-482.

Bantilan MCS, Deb UK, Gowda CLL, Reddy BVS, Obilana AB, Evenson RE. 2004. Sorghum genetic enhancement, research process, dissemination and impacts. Patancheru 502 324, Andhra Pradesh, India. International Crops Research Institute for the Semi-Arid Tropics, pp. 320.

Bar-Hen A, Charcosset A, Bourgoin M, Guiard J. 1995. Relationships between genetic markers and morphological traits in a maize inbred lines collection. Euphytica 84: 145-154.

Barnaud A, Deu M, Garine E, Mckey D, Joly HI. 2007. Local genetic diversity of sorghum in a village in northern Cameroon: structure and dynamics of landraces. Theoretical and Applied Genetics 114:237-248.

Barnaud A, Trigueros G, Mckey D, Joly HI. 2008. High outcrossing rates in fields with mixed sorghum landraces: how are landraces maintained? Heredity 101:445-452.

Barrett BA, Kidwell KK. 1998. AFLP-based genetic diversity assessment among wheat cultivars from the Pacific Northwest. Crop Science 38:1261-1271.

Beckman JS, Soller M. 1983. Restriction fragment length polymorphisms in genetic improvements: methodologies, mapping and costs. Theoretical and Applied Genetics 67:35-43.

Bhattramakki D, Dong JM, Chabra AK, Hart GE. 2000. An intergrated SSR and RFLP linkage map of *Sorghum bicolor* (L) Moench. Genome 43:988-1002.

Blench R. 2006. Language, archaeology and the African past. African Archaeology Series. Volume (10). Altamira Press, Lanham, MD, pp. 361.

Bling HFJ. 2000. Detection of adulteration of basmati rice with non-premium long grain rice. International Journal of Food Science and Technology 35:257-256.

Bramel PJ, Ford-Lloyd BV, Chandra S, Newbury HJ. 2004. Assessment of degree and distribution of genetic diversity in pigeonpea landraces using SSR markers. In: Assessing the risk of losses of biodiversitry in traditional cropping systems. Edited by Bramel PJ. A case study of pigeonpea in Andhra Pradesh. International Crops Research Institute for Semi-arid tropics Pantacheru, 502 324, Andhra Pradesh, India, pp. 75.

Breyne P, Rombaut D, Van Gysel A, Van Montagu M, Gerats M. 1999. AFLP analysis of genetic diversity within and between *Arabidopsis thaliana* ecotypes. Molecular and General Genetics 261:627-634.

Brown AHD, Weir BS. 1983. Measuring genetic variability in plant populations. In: Isozymes in plant genetics and breeding. Edited by Tanskley SD, Orton TJ. Part A. Elsevier, Amsterdam. pp. 219-229.

Brown SM, Hopkins MS, Mitchell SE, Senior ML, Wang TY, Duncan RR, Gonzalez-Candelas F, Kresovich S. 1996. Multiple methods for the identification of polymorphic simple sequence repeats (SSRs) in sorghum [*Sorghum bicolor* (L.) Moench]. Theoretical and Applied Genetics 93:190-198.

Buhariwalla HK, Crouch JH. 2004. Optimization of marker screening protocol to assess the degree and distribution of genetic diversity in landrecas of pigeonpea. In: Assesing the risk of losses of biodiversity in traditional cropping systems. Edited by Bramel PJ. A case study of pigeonpea in Andhara Pradesh. International Crops Institute for the semi-Arid tropics Pantacheru, 502 324, Andhra Pradesh, India, pp. 168.

Burstin J, Charcosset A. 1997. Relationships between phenotypic and marker distance: Theoretical and experimental investigation. Heredity 78:477-483.

Casa AM, Mitchell SE, Hamblin MT, Sun H, Bowers JE, Paterson AH, Aquadro CF, Kresovich S. 2005. Diversity and selection in sorghum: simultaneous analyses using simple sequence repeats. Theoretical and Applied Genetics 111:23-30.

Celarier RP. 1958. Cytotaxonomic notes on the subsection *halepensia* of the genus *Sorghum*. Bulletin of the Torrey Botany Club 85:49-62.

Chandel S, Paroda R. 2000. Status of plant genetic resources conservation and utilization in Asia-Pacific region. Regional synthesis report. Asia-Pacific Association of Agricultural Research Institutions, FAO, Bangkok, pp. 224-229.

Chen X, Temnykh S, Xu Y, Cho YG, McCouch SR. 1997. Develpoment of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L.). Theoretical and Applied Genetics 95:533-567.

Cui YX, Xu GW, Magill CW, Schertz KF, Hart GE. 1995. RFLP-based assay of *Sorghum bicolor* (L) Moench. genetic diversity. Theoretical and Applied Genetics 90:787-796.

Dahlberg JA, Spinks MS. 1995. Current status of the US sorghum germplasm collection. International Sorghum and Millets Newsletter 36:4-12.

Dahlberg JA, Zhang X, Hart GE, Mullet JE. 2002. Comparative assessment of variation among sorghum germplasm accessions using seed morphology and RAPD measurements. Crop Science 42:291-296.

De Oliveira AC, Richter T, Bennetzen JL. 1996. Regional and racial specificities in sorghum germplasm assessed with DNA markers. Genome 39:579-587.

De Vries J, Toenniessen G. 2000. Securing the harvest: biotechnology, breeding and seed systems for African crops. CABI Publication, Wallingford, pp. 207.

De Wet JMJ. 1978. Systematics and evolution of sorghum sect. Sorghum (Gramineae). American Journal of Botany 65:477-484.

De Wet JMJ, Huckabay JP. 1967. The origin of *Sorghum bicolor*. II. Distribution and domestication. Evolution 21:787-802.

Dean RE, Dahlberg JA, Hopkins MS, Mitchell SE, Kresovich S. 1999. Genetic redundancy and diversity among 'Orange' accessions in the U.S. national sorghum collection as assessed with simple sequence repeat (SSR) markers. Crop Science 39:1215-1221.

Demissie A, Bjørnstad Å, Kleinhofs A. 1998. Restriction fragment length polymorphisms in landrace barleys from Ethiopia in relation to geographic, altitude, and agro-ecological factors. Crop Science 38:237-243.

Deu M, Gonzalezdeleon D, Glaszmann JC, Degremont I, Chantereau J, Lanaud C, Hamon P. 1994. RFLP diversity in cultivated sorghum in relation to racial differentiation. Theoretical and Applied Genetics 88:838-844.

Deu M, Hamon P, Chantereau J, Dufour P, Dhont A, Lanaud C. 1995. Mitochondrial-DNA diversity in wild and cultivated sorghum. Genome 38:635-645.

Deu M, Rattunde HFW, Chantereau J. 2006. A global view of genetic diversity in cultivated sorghum using a core collection. Genome 49:168-180.

Deu M, Sagnard F, Chantereau J, Calatayud C, Herault D, Mariac C, Pham JL, Vigouroux Y, Kapran I, Traore PS, Mamadou A, Gerard B, Ndjeunga J, Bezancon G. 2008. Niger-wide assessment of *in situ* sorghum genetic diversity with microsatellite markers. Theoretical and Applied Genetics 116:903-913.

Dje Y, Forcioli D, Ater M, Lefebvre C, Vekemans X. 1999. Assessing population genetic structure of sorghum landraces from North-western Morocco using allozyme and microsatellite markers. Theoretical and Applied Genetics 99:157-163.

Dje Y, Heuertz M, Lefebvre C, Vekemans X. 2000. Assessment of genetic diversity within and among germplasm accessions in cultivated sorghum using microsatellite markers. Theoretical and Applied Genetics 100:918-925.

Doggett H. 1988. Sorghum. Longman Scientific and Technical Essex, England, pp. 512.

Doggett H, Prasada Rao KE. 1995. Sorghum. In: Evolution of crop plants (second edition). Edited by Smartt J, Simmonds NW. Longman Group, Burnt Mill, pp. 173-180.

Doldi ML, Vollmann J, Lelly T. 1997. Genetic diversity in soybean as determined by RAPD and microsatellite analysis. Plant Breeding 116:331-335.

Donini P, Elias ML, Bougourd SM, Koebner RMD. 1997. AFLP fingerprinting reveals pattern differences between DNA extracted from different plant organs. Genome 40:521-526.

Dubreuil P, Charcosset A. 1998. Genetic diversity within and among maize populations: a comparisons between isoenzyme and nuclear RFLP loci. Theoretical and Applied Genetics 96:577-587.

Dweikat I. 2005. A diploid, interspecific, fertile hybrid from cultivated sorghum, *Sorghum bicolor*, and the common Johnsongrass weed *Sorghum halepense*. Molecular Breeding 16:93-101.

Edwards K, Johnstone C, Thompson C. 1991. A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucleic Acids Research 19:1349-1458.

Ellis JR, Burke JM. 2007. EST-SSRs as a resource for population genetic analyses. Heredity 99:125-132.

Everitt B. 1980. Cluster analysis. 2th edition, Halstead Press, New York, pp. 136.

Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics 131:479-491.

FAO. 1995. Production year book. Vol. 49. FAO, Roe, Italy.

Fernandez ME, Figueiras AM, Beinito C. 2002. The use of ISSR and RAPD markers for detecting DNA polymorhism, genotype identification and genetic diversity among barley cultivars with unknown origin. Theoretical and Applied Genetics 104:845-851.

Fleischer R, Horlemann C, Schwekendiek A, Kling C, Weber G. 2004. AFLP fingerprinting in hop: analysis of the genetic variability of the Tettnang variety. Genetic Resources and Crop Evolution 51:211-220.

Folkertsma RT, Frederick H, Rattunde HFW, Chandra S, Raju GS, Hash CT. 2005. The pattern of genetic diversity of Guinea-race *Sorghum bicolor* (L.) Moench landraces as revealed with SSR markers. Theoretical and Applied Genetics 111:399-409.

Franco JE, Crossa JM, Díaz J, Taba T, Villaseñor J, Eberhart SA. 1997. A sequential clustering strategy for classifying gene bank accessions. Crop Science 37:1656-1662.

Franco JE, Crossa JM, Ribaut JM, Betran J. 2001. A method for combining molecular markers and phenotypic attributes for classifying plant genotypes. Theoretical and Applied Genetics 103:944-952.

Frankel OH, Brown AHD, Burdon JJ. 1995. The conservation of plant biodiversity. Cambridge University Press, Cambridge, England. pp. 299.

Fregene MA, Suarez M, Mkumbira J, Kulembeka H, Ndedya E, Kulaya A, Mitchel S, Gullberg U, Rosling H, Dixon AGO, Dean R, Kresovich S. 2003. Simple sequence repeats marker diversity in cassava landraces; genetic diversity and differentiation in an asexually propagated crop. Theoretical and Applied Genetics 107:1083-1093.

Fufa H, Baenziger PS, Beecher BS, Dweikat I, Graybosch RA, Eskridge KM. 2005. Comparison of phenotypic and molecular marker-based classifications of hard red winter wheat cultivars. Euphytica 145:133-146.

Fuller DQ. 2003. African crops in prehistoric South Asia: a critical review. In: Food, fuel and fields. Edited by Neumann K, Butler A, Kahlheber S. Progress in Africa Archaeobotany, Africa Praehistorica 15. Heinrich-Barth-Institute, Colonge, pp. 239-271.

Garber ED. 1950. Cytotaxonomic studies in the genus sorghum. University of California Publications in Botany 23:283-361.

Garcia-Mas J, Oliver M, Gómez-Paniagua H, De Vicente MC. 2000. Comparing AFLP, RAPD and RFLP markers for measuring genetic diversity in melon. Theoretical and Applied Genetics 101:860-864.

Gawel NJ, Jarret RL, Whittermore A. 1992. Restriction fragment length polymorphism (RFLP) - based phylogenetic analysis of *Musa*. Theoretical and Applied Genetics 84:286-290.

Geleta N, Labuschagne MT, Viljoen CD. 2006. Genetic diversity analysis in sorghum germplasm as estimated by AFLP, SSR and morpho-agronomical markers. Biodiversity and Conservation 15:3251-3265.

Gerber S, Mariette S, Streiff R, Bodénès C, Kremer A. 2000. Comparison of microsatellites and amplified fragment length polymorphism markers for parentage analysis. Molecular Ecology 9:1037-1048.

Ghebru B, Schmidt RJ, Bennetzen JL. 2002. Genetic diversity of Eritrean sorghum landraces assessed with simple sequence repeat (SSR) markers. Theoretical and Applied Genetics 105:229-236.

Gonzalez JM, Ferrer E. 1993. Random amplified polymorphic DNA analysis in *Hordeum* species. Genome 36:1029-1031.

Gopal RV, Upadhyaya HD, Gowda CLL. 2006. Current status of sorghum genetic resources at ICRISAT: their sharing and impacts. International Sorghum and Millets Newsletter 47:9-13.

Gower JC. 1971. A general coefficient of similarity and some of its properties. Biometrics 27:857-871.

Grenier C, Bramel-Cox PJ, Noirot M, Prasada Rao KA, Hamon P. 2000a. Assessment of genetic diversity in three subsets constituted from ICRISAT sorghum collection using random vs. non-random sampling procedures. A. Using morpho-agronomical and passport data. Theoretical and Applied Genetics 101:190-196.

Grenier C, Deu M, Kresovich S, Bramel-Cox PJ, Hamon P. 2000b. Assessment of genetic diversity in three subsets constituted from ICRISAT sorghum collection using random vs. non-random sampling procedures. B. Using molecular markers. Theoretical and Applied Genetics 101:197-202.

Hair JR, Anderson RE, Tatham RL, Black WC. 1995. Multivariate data analysis with readings. 4th edition, Prentice-Hall, Englewood Cliffs, NJ.

Hamrick JL, Godt MJW. 1997. Allozyme diversity in cultivated crops. Crop Science 37:26-30.

Hardys H, Ballick M, Schiewater B. 1992. Amplification of random amplified polymorphic DNA (RAPD) in molecular ecology. Molecular Ecology 1:55-63.

Harlan JR. 1975. Crops and man. American Society of Agronomy Madison, pp. 284.

Harlan JR, De Wet JMJ. 1972. A simplified classification of cultivated sorghum. Crop Science 12:172-177.

Harlan JR, De Wet JMJ, Stemler ABL. 1976. The origins of African plant domestication. Mouton Press, The Hague, Netherlands, pp. 498.

Hartl DL, Clark AG. 1997. Principles of population genetics. 3rd edition. Sunderland (MA): Sinauer Associates. pp. 542.

Hedrick PW. 2005. Genetics of populations, 3rd edition. Jones and Bartlett, Sudbury, MA, pp. 725.

Helentjaris T, King G, Slocum M, Siedenstrang C, Wegman S. 1985. Restriction fragment polymorphosim as probes for plant diversity and their development as tools for applied plant breeding. Plant Molecular Biology 5:109-118.

Hill M, Witsenboer H, Zabeau M, Vos P, Kesseli R, Michelmore R. 1996. PCR-based fingerprinting using AFLPs as a tool for studying genetic relationships in *Lactuca* spp. Theoretical and Applied Genetics 93:1202-1210.

Holsinger KE, Weir BS. 2009. Fundamental concepts in genetics. Genetics 10:639-650.

Holton TA, Cross M, Karakousis A, Henry RJ. 2000. Genetic screening of Australian barley varieties using microsatellite markers. Paper presented to the Plant and Animal Genome VIII Conference, San Diego, California, USA, 9-12 January.

House LR, Osmanzai M, Gomez MI, Monyo ES, Gupta SC. 1995. Agronomic principles. In: Sorghum and millets: Chemistry and technology. Edited by Dendy DAV. American Association of Cereal Chemists, Saint Paul Minnesota, USA, pp. 27-67.

Huttner E, Wenzl P, Akbari M, Caig V, Carling J, Cayla C, Evers M, Jaccoud D, Peng K, Patarapuwandol S, Uszynski G, Xia L, Yang S, Kilian A. 2005. Diversity array technology: A novel tool for harnessing the genetic potential of orphan crops In: Discovery to delivery: BioVision Alexandria 2004. Edited by Serageldin I, Persley GJ. Proceedings of the 2004 Conference of the World Biological Forum. CABI Publishing: UK, pp. 145-155.

Incirli A, Akkaya MS. 2001. Assessment of genetic relationships in durum wheat cultivars using AFLP markers. Genetic Resources and Crop Evolution 48:233-238.

Jaccard P. 1908. Nouvelles researchers sur la distribution florale. Bulletin de la Society Vaudoise des Sciences Naturelles 44:223-270.

Jaccoud D, Peng K, Feinstein D, Kilian A. 2001. Diversity arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Research 29:e25.

Jain A, Bhatia S, Bhanga SS, Prakash S, Lakshmikumara MN. 1994. Potential use of random amplified polymorphic DNA (RAPD) technique to study the genetic diversity in Indian mustard (*Brassica juncea*) and its relationships to heterosis. Theoretical and Applied Genetics 8:116-122.

Jarne P, Theron A. 2001. Genetic structure in natural populations of flukes and snails: a practical approach and review. Parasitology 123:S27-S40.

Jarret RL, Gawel N, Whittermore A. 1992. Phylogenetic relationships of the sweet potato *Ipomea batatas* (L). Lam. Journal of American Society of Horticultural Science 117:633-637.

Jena KK, Kochert GC. 1991. Restriction fragment length polymorphism analysis of CCDD genome species of the genus *Oryza* L. Plant Molecular Biology 16:633-637.

Johnson AR, Wichern DW. 1992. Applied multivariate statistical analysis. (3rd edition), Prentice-Hall, Englewood Cliffs, NJ, pp. 767.

Jolliffe IT. 1986. Principal component analysis. Springer-Verlag, Berlin. pp. 522.

Jones CJ, Edwards KJ, Castaglione S, Winfield MO, Sala F, Van De Wiel C, Bredemeijer G, Vosman B, Mathes M, Daly A, Brettschneider R, Bettini P, Buiatti M, Maestri E, Malcevschi A, Mamiroli N, Aert R, Volckaert G, Rueda J, Linacero R, Vaazquez A, Karp A. 1997. Reproducibilty testing of RAPD, AFLP, and SSR markers in plants by a network of European laboratories. Molecular Breeding 3:381-390.

Karp A, Kresovich S, Bhat KV, Ayad WG, Hodgkin T. 1997. Molecular tools in plant genetic resources conservation: a guide to the technologies. IPGRI Technical Bulletin No. 2. International Plant Genetic Resources Institute, Rome, Italy.

Karp A, Seberg O, Buiatti M. 1996. Molecular techniques in the assessment of botanical diversity. Annals of Botany 78:143-149.

Kaufman L, Rousseeuw PJ. 1990. Finding groups in data: An introduction to cluster analysis. A Wiley-Interscience Publication, NY, pp. 335.

Kejun L, Major G, Spencer M, Smith JS, Buckler ED, Doebley J. 2003. Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. Genetics 165:2117-2128.

Kiambi DK, Newbury HJ, Ford-Lloyd BV, Dawson I. 2005. Contrasting genetic diversity among *Oryza longistaminata* from different geographic origins using AFLP. African Journal of Biotechnology 4:308-317.

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

Kosman E, Leonard KJ. 2005. Similarity coefficients for molecular markers in studies of genetic relationships between individuals for haploid, diploid, and polyploidy species. Molecular Ecology 14: 415-424.

Kourti A. 2002. Estimates of heterozygosity and patterns of geographic differentiation in natural populations of the medfly (*Ceratitis capitata*). Hereditas 137:173-179.

Kresovich S, Szewe-Mcfadden AK, Bliek SM, McFerson JR. 1995. Abundance and characterization of simple-sequence repeats (SSRs) isolated from size-fractioned genomic library of *Brassica napus* L. (rapeseed). Theoretical and Applied Genetics 19:206-211.

Krishna KG, Zhang J, Burow M, Pittman NR, Delikostadinov SG, Lu Y, Puppala N. 2004. Genetic diversity analysis in Valencia peanut (*Arachis hypogaea* L.) using microsatellite markers. Cellular and Molecular Biology Letters 9:685-697.

Kubik C, Sawkins M, Meyer AW, Gaut SB. 2001. Genetic diversity in seven perennial ryegrass (*Lolium perenne* L.) cultivars based on SSR markers. Crop Science 41: 1565-1572.

Kumar LS. 1999. DNA markers in plant improvement: An overview. Biotechnology Advances 17:143-182.

Kumar V, Sharma S, Kero S, Sharma S, Sharma AK, Kumar M, Bhat KV. 2008. Assessment of genetic diversity in common bean (*Phaseolus vulgaris* L.) germplasm using amplified fragment length polymorphism (AFLP). Scientia Horticulturae 116:138-143.

Laopaiboon L, Thanonkeo P, Jaisil P, Laopaiboon P. 2007. Ethanol production from sweet sorghum juice in batch and fed-batch fermentations by *Saccharomyces cerevisiae*. World Journal of Microbiology and Biotechnology 23:1497-1501.

Laurence G, Gaudin JC, Bensadoun P, Rebillat I, Morel Y. 2009. Real-time PCR assay for detection of a new stimulant for poxyvirus biothreat agents. Applied and Environmental Microbiology 75:1614-1620.

Lima MLA, Garcia AAF, Oliveira KM, Matsuoka S, Arozono H, de Souza CL Jr, de Souza AP. 2002. Analysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugar cane (*Saccharum* spp). Theoretical and Applied Genetics 104:30-38.

Liu CD, Rossnagel BG, Scoles GJ. 2000. The development of oat microsatellite markers and their use in identifying relationships among *Avena* species and oat cultivars. Theoretical and Applied Genetics 101:1259-1268.

Lu J, Knox MR, Ambrose MJ, Brown JKM, Ellis THN. 1996. Comparative analysis of genetic diversity in pea assessed by RFLP and PCR-based methods. Theoretical and Applied Genetics 93:1103-1111.

Lubbers EL, Gill KS, Cox TS, Gill BS. 1991. Variation of molecular markers among geographically diverse accessions of *Triticum tauschii*. Genome 34:354-361.

Mace ES, Xia L, Jordan DR, Halloran K, Parh DK, Huttner E, Wenzl P, Kilian A. 2008. DArT markers: diversity analyses and mapping in *Sorghum bicolor*. BMC Genomics 9:26.

Mackill DJ. 1995. Classifying Japonica rice cultivars with RAPD markers. Crop Science 35:889-894.

Mahalakshmi V, Aparna P, Ramadevi S, Rodomiro O. 2002. Genomic sequence derived simple sequence repeats markers. A case study with *Medicago* spp. Electronic Journal of Biotechnology 5:233-242.

Maheswaran M, Subudhi PK, Nandi S, Xu JC, Parco A, Yang DC, Huang N. 1997. Polymorphism, distribution, and segregation of AFLP markers in a double haploid rice population. Theoretical and Applied Genetics 94:39-45.

Malecot G. 1948. Les mathematiques de 1'heredite. Masson and Cie Paris. pp. 63.

Manly BFJ. 1991. Randomization and Monte Carlo methods in biology. Chapman and Hall, London, pp. 120.

Mantel N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Research 27:209-220.

Maquet A, Zoro Bi I, Delvaux M, Wathelet B, Baudoin JP. 1997. Genetic structure of a lima bean base collection using alloenzyme markers. Theoretical and Applied Genetics 95:980-991.

Marsan AP, Castiglioni P, Fusari F, Kuiper M, Motto M. 1998. Genetic diversity and its relationship to hybrid performance in maize as revealed by RFLP and AFLP markers. Theoretical and Applied Genetics 96:219-227.

Martin JM, Blake TK, Hockett EA. 1991. Diversity among North American spring barely clutivars based on coefficient of parentage. Crop Science 31:1131-1137.

Matus IA, Hayes PM. 2002. Genetic diversity in three groups of barley germplasm assessed by simple sequence repeats. Genome 45:1095-1106.

Maughan PJ, Saghai-Maroof MA, Buss GR, Huestis GM. 1996. Amplified fragment length polymorphism (AFLP) in soyabean: species diversity, inheritance, and near-isogenic line analysis. Theoretical and Applied Genetics 93:392-401.

Melchinger AE. 1993. Use of RFLP markers for analysis of genetic relationships among breeding materials and prediction of hybrid performance. In: First international crop science congress. Edited by Buxton DR. CSSA, Madison, WI, pp. 621-627.

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

Menz MA, Klein RR, Mullet JE, Obert JA, Unruh NC, Klein PE. 2002. A high-density genetic map of *Sorghum bicolor* (L.) Moench based on 2926 AFLP, RFLP and SSR markers. Plant Molecular Biology 48:483-499.

Menz MA, Klein RR, Unhruh NC, Rooney Wl, Klein PE, Mullet JE. 2004. Genetic diversity of public inbreds of sorghum determined by mapped AFLP and SSR markers. Crop Science 44:1236-1244.

Messmer MM, Melchinger AE, Boppenmaier J, Herrmann RG. 1993. Relationships among early European maize (*Zea mays* L.) inbred lines: II. Comparison of pedigree and RFLP data. Crop Science 33:944-950.

Miller JC, Tanksley SD. 1990. RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. Theoretical and Applied Genetics 80:437-448.

Mitchell SE, Kresovich S, Jester CA, Javier Hernandez C, SzewcMcFadden AK. 1997. Application of multiplex PCR and fluorescence-based, semi automated allele sizing technology for genotyping plant genetic resources. Crop Science 37:617-624.

Mohammadi SA, Prasanna BM. 2003. Analysis of genetic diversity in crop plants - salient statistical tools and considerations. Crop Science 43:1235-1248.

Morden CW, Doebley JF, Schertz KF. 1989. Allozyme variation in old world races of *Sorghum bicolor (*Poaceae). American Journal of Botany 76:247-255.

Morden CW, Doebley K, Schertz KF. 1990. Allozyme variation among the spontaneous species of *Sorghum* section Sorghum (Poaceae). Theoretical and Applied Genetics 80:296-304.

Morell MK, Peakall R, Appels R, Preston LR, Lloyd HL. 1995. DNA profiling techniques for plant variety identification. Australian Journal of Experimental Agriculture 35:807-819.

Morgante CW, Olivieri AM. 1993. PCR-amplified microsatellites as markers in plant genetics. Plant Journal 3:175-182.

Murdock GP. 1959. Staple subsistence crops of Africa. Geographical Review 50:521-540.

Negash A, Tsegaye A, Treuren R, Visser B. 2002. AFLP analysis of *Enset* clonal diversity in south and southwestern Ethiopia for conservation. Crop Science 42:1105-1111.

Nei M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583-590.

Nei M. 1987. Molecular evolutionary genetics. Columbia University Press, New York. pp. 526.

Nei M, Li WH. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proceedings of the National Academy of Science USA 79:5269-5273.

Newbury HJ, Burns MJ, Baggot C, Ford-Lloyd BV. 2004. Development of SSR molecular markers to assess the degree and distribution of genetic diversity in landraces of pigeonpea. In: Assessing the risk of losses of biodiversity in traditional cropping systems. Edited by Bramel PJ. A case study of pigeonpea in Andhra Pradesh. International Crops Research Institute for the Semi-arid tropics Pantacheru. Andhra Pradesh India, pp. 133-141.

Nkongolo KK, Nsapato L. 2003. Genetic diversity in *Sorghum bicolor* (L.) Moench accessions from different ecogeographical regions in Malawi assessed with RAPDs. Genetic Resources and Crop Evolution 50:149-156.

Ollitrault P, Arnaud M, Chantereau J. 1989. Polymorphisme enzymatique des sorghos. II. Organisation génétique et évolutive des sorghos cultivés. Agronomie of Tropicale 44:211-222.

Pakniyat H, Powell W, Baird E, Handley LL, Robinson D, Scrimgeour CM, Nevo E, Hackett CA, Caligara PDS, Forster BP. 1997. AFLP variation in wild barley with reference to salt tolerance and associated ecogeography. Genome 40:332-341.

Panchen AL. 1992. Classification, evolution and the nature of biology. Cambridge University Press, Cambridge, England. pp. 403.

Paul S, Wachira FN, Powell W, Waught R. 1997. Diversity and genetic differentiation among populations of Indian and Kenyan tea (*Camellia sinensis* (L.) O. Kuntze) revealed by AFLP markers. Theoretical and Applied Genetics 94:255-263.

Perkin-Elmer. 1996. AFLP Plant Mapping Kit – the PCR marker of choice for plant mapping.

Perumal R, Renganayaki K, Menz MA, Katilé S, Dahlberg J, Magill CW, Rooney WL. 2007. Genetic diversity among sorghum races and working groups based on AFLP and SSR. Crop Science 47:1375-1383.

Poulsen G, Hahl B, Saghai-Maroof G, Weissing K. 1993. Abundance and polymorphism of simple repetitive DNA sequences in *Brassica napus* L. Theoretical and Applied Genetics 85:994-1000.

Powell W, Machary GC, Provan J. 1996a. Polymorphism revealed by single sequence repeats. Trends in Plant Science 1:215-222.

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey C, Rafalski A. 1996b. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Molecular Breeding 2:225-238.

Prakash SPJ, Biji SP, Michael KRS, Ganesa K, Chandra R. 2006. Genetic diversity analysis of sorghum (*Sorghum bicolor* L. Moench) accessions using RAPD markers. Crop Science 1:109-112.

Priolli RHG, Pinheiro JB, Zucchi MI, Bajay MM, Vello NA. 2010. Genetic diversity among Brazilian soybean cultivars based on SSR loci and pedigree data. Brazilian Archives of Biology and Technology 53:519-531.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945-959.

Pusterla N, Madigan JE, Leutenegger CM. 2006. Real-time polymerase chain reaction: A novel molecular diagnostic tool for equine infectious diseases. Journal of Veterinary Internal Medicine 20:3-12.

Qingshan L, Dahlberg JA. 2001. Chinese sorghum genetic resources. Economic Botany 55:401-425.

Rafalski A, Tingey SV. 1993. Genetic diagnosis in plant breeding: RAPDs, microsatellites and machines. Trends in Genetics 9:275-279.

Ratnaparkhe MB, Gupta VS, Ven Murthy MR, Ranjekar PK. 1995. Genetic fingerprinting of pigeon pea (*Canjnus cajan*) (L.) Millsp. and its wild relatives using RAPD markers. Theoretical and Applied Genetics 91:893-898.

Reif JC, Melchinger AE, Xia XC, Warburton ML, Hoisington DA, Vasal SK, Beck D, Bohan M, Frisch M. 2003. Use of simple sequence repeats (SSRs) for establishing heterotic groups in subtropical maize. Theoretical and Applied Genetics 107:947-957.

Rogers JS. 1972. Measures of genetic similarity and genetic distance. Studies in genetics. VII. University Texas Publication 2713:145-153.

Rohlf FJ. 1972. An empirical comparison of three ordination techniques in numerical taxonomy. Systematic Zoology 21:271-280.

Rosenow DT, Dahlberg JA. 2000. Collections, conversion and utilization of sorghum. In: Sorghum: origin, history, technology and production. Edited by Smith CW, Frederiksen RA. New York, USA: John Wiley and Sons, Inc, pp. 309-328.

Rosenow DT, Woodfin CA, Clark LE, Sij JW. 1999. Drought resistance in exotic sorghums. In: 1999 Annual Meetings Abstracts. ASA, CSSA, and SSSA, Madison, WI, pp. 166.

Russell JR, Fuller JD, Macaulay M, Hatz BG, Jahoor A, Powell W, Waugh R. 1997. Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. Theoretical and Applied Genetics 95:714-722.

Saghai-Maroof MA, Biyashev RM, Yang GP, Zhang Q, Allard RW. 1994. Extraordinarily polymorphic microsatellite DNA in barley: Species diversity, chromosomal locations and population dynamics. Proceedings of the National Academy of Science USA 91:5466-5470.

Sagnard F, Barnaud A, Deu M, Barro C, Luce C, Billot C, Rami JF, Bouchet S, Dembele D, Pomies V, Calatayud C, Rivallan R, Joly H, Brocke KV, Toure A, Chantereau J, Bezancon G, Vaksmann M. 2008. Multi-scale analysis of sorghum genetic diversity: Understanding the evolutionary processes for *in situ* conservation. Cahiers Agricultures 17:114-121.

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N. 1985. Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230:1350-1354.

Sarıkamış G, Yaşar F, Bakır M, Kazan K, Ergül A. 2009. Genetic characterization of green bean (*Phaseolus vulgaris*) genotypes from eastern Turkey. Genetic and Molecular Research 8:880-887.

Schloss SJ, Mitchell SE, White GM, Kukatla R, Bowers JE, Paterson AH, Kresovich S. 2002. Characterization of RFLP probe sequences for gene discovery and SSR development in *Sorghum bicolor* (L.) Moench. Theoretical and Applied Genetics 105:912-920.

Schug MD, Hutter CM, Wetterstrand KA, Gaudette MS, Mackay TFC, Aquadro CF. 1998. The mutation rates of di-, tri-, and tetra-nucleotide repeats in *Drosophila melanogaster*. Molecular Biology Evolution 15:1751-1760.

Schut JW, Qi X, Stam P. 1997. Association between relationship measures based on AFLP markers, pedigree data and morphological traits in barley. Theoretical and Applied Genetics 95:1161-1168.

Senior ML, Heun M. 1993. Mapping maize microsatellites and polymerase chain reaction confirmation of targeted repeats using a CT primer. Genome 36:884-889.

Senior ML, Murphy JP, Goodman MM, Stuber CW. 1998. Utility of SSRs for determining genetic similarities and relationships in maize using agarose gel system. Crop Science 38:1088-1098.

Sharma SK, Knox MR, Ellis THN. 1996. AFLP analysis of the diversity and phylogeny of Lens and its comparison with RAPD analysis. Theoretical and Applied Genetics 93:751-758.

Shehzad T, Hisato O, Makoto K, Kazutoshi O. 2009. Development of SSR-based sorghum (*Sorghum bicolor* (L.) Moench) diversity research set of germplasm and its evaluation by morphological traits. Genetic Resources and Crop Evolution 56:809-827.

Singh A, Negi MS, Rajagopal J, Bhatia S, Tomar UK, Srivastava PS, Lakshmikumaran M. 1999. Assessment of genetic diversity in *Azadirachta indica* using AFLP markers. Theoretical and Applied Genetics 99:272-279.

Sjogren P, Wyone PI. 1994. Conservation genetics and detection of rare alleles in finite populations. Conservation Biology 8:267-270.

Smith CW, Frederiksen RA. 2000. Sorghum: origin, history, technology, and production New York, NY: John Wiley and Sons, pp. 824.

Smith JSC. 1984. Genetic variability within U.S. hybrid maize: Multivariate analysis of isozyme data. Crop Science 24:1041-1046.

Smith JSC, Chin ECL, Shu H, Smith OS, Wall SJ, Senior ML, Mitchell SE, Kresovich S, Ziegle J. 1997. An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.): comparisons with data from RFLPs and pedigree. Theoretical and Applied Genetics 95:163-173.

Smith JSC, Kresovich S, Hopkins MS, Mitchell SE, Dean RE, Woodman WL, Lee M, Porter K. 2000. Genetic diversity among elite sorghum inbred lines assessed with simple sequence repeats. Crop Science 40:226-232.

Smith JSC, Paszkiewics S, Smith OS, Schaeffer J. 1987. Electrophoretic, chromatograhic and genetic techniques for identifying associations and measuring genetic diversity among corn hybrids. In: Proceedings of the 42nd annual corn sorghum research conference, Chicago, IL. American Seed Trade Assocation, Washington, DC, pp. 187-203.

Smith JSC, Smith OS. 1992. Fingerprinting crop varieties. Advances in Agronomy 47:85-141.

Sneath PHA, Sokal RR. 1973. Numerical taxonomy; the principles and practice of numerical classification. San Francisco: Freeman, pp. 573.

Snowden JD. 1936. The cultivated race of sorghum. Allard and Son, London, pp. 274.

Snowden JD. 1955. The wild fodder sorghums of the genus *Eu-sorghum*. Journal of the Linnean Society of London 55:191-260.

Sokal RR, Michener CD. 1958. A statistical method for evaluating systematic relationships. University of Kansas Science Bulletin 38:1409-1438.

Song KM, Osborn TC, William PH. 1988. *Brassica* taxonomy based on nuclear restriction fragment polymorphisms (RFLP). Theoretical and Applied Genetics 75:784-794.

Song KM, Osborn TC, William PH. 1990. Brassica taxonomy based on nuclear restriction fragment polymorphisms (RFLP). 3 Genome relationships in Brassica and related genera and the origin of *B. oleracea* and *B. rapa* (syn. *campestris*). Theoretical and Applied Genetics 79:497-506.

Sorriano MJ, Romero C, Vilanova S, Lla`cer G, Badenes ML. 2005. Genetic diversity of loquat germplasm (*Eriobotrya japonica* [Thunb] Lindl) assessed by SSR markers. Genome 48:108-114.

Southern EM. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. Journal of Molecular Biology 98:503-517.

Spooner D, van Treuren R, de Vicente MC. 2005. Molecular markers for genebank management. IPGRI Technical Bulletin No 10, pp. 136.

Stanton MA, Stewart JM, Percival AE, Wendel JF. 1994. Morphological diversity and relationships in A-genome cottons, *Gossypium arboretum* and *G. herbaceum.* Crop Science 34:519-527.

Stemler ABL, Harlan JR, de Wet JMJ. 1977. The sorghums of Ethiopia. Economic Botany 31:446-460.

Tao Y, Manners JM, Ludlow MM, Henzell RG. 1993. DNA polymorphisms in grain sorghum [*Sorghum bicolor* (L.) Moench]. Theoretical and Applied Genetics 86:679-688.

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

Tautz D. 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Research 17:6463-6471.

Tinker NA, Kilian A, Wight CP, Heller-Uszynska K, Wenzl P, Rines HW, Bjornstad A, Howarth CJ, Jannink JL, Anderson JM. 2009. New DArT markers for oat provide enhanced map coverage and global germplasm characterization. BMC Genomics 10:39.

Travis SE, Maschinski J, Keim P. 1996. An analysis of genetic variation in *Astragalus cremnophylax* var. *cremnophylax*, a critically endangered plant, using AFLP markers. Molecular Ecology 5:735-745.

Tucak M, Popović S, Čupić T, Grljušić S, Meglič V, Jurković Z. 2010. Efficiency of phenotypic and DNA markers for a genetic diversity study of alfalfa. Russian Journal of Genetics 46:1314-1319.

Tunstall V, Teshome A, Torrance JK. 2001. Distribution, abundance and risk of loss of sorghum landraces in four communities in North Shewa and South Welo, Ethiopia. Genetic Resources and Crop Evolution 48:131-142.

Uptmoor R, Wenzel W, Friedt W, Donaldson G, Ayisi K, Ordon F. 2003. Comparative analysis on the genetic relatedness of *Sorghum bicolor* accessions from Sourthern Africa by RAPDs, AFLPs and SSRs. Theoretical and Applied Genetics 106: 1316-1325.

Van der Maesen LJG. 1990. Pigeonpea: origin, history, evolution, and taxonomy. In: The pigeon pea. Edited by NeneYH, Hall SD, Sheila VK. C.A.B. International UK, pp. 15-46.

Vos P, Hogers R, Bleeker M, Reijans M, Van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research 23:4407-4414.

Warburton ML, Zianchun X, Crossa J, Franco J, Melchinger AE, Frisch M, Bohn M, Hoisington D. 2002. Genetic characterization of CIMMYT inbred maize lines on open pollinated populations using large scale fingerprinting methods. Crop Science 42:1832-1840.

Ward JH. 1963. Hierarchical grouping to optimize on objective function. American Statistical Association Journal 56:236-244.

Weber JL. 1990. Informativeness of human $(dC\text{-}dA)_n$ $(dG\text{-}dT)_n$ polymorphisms. Genomics 7:24-530.

Weir BS. 1990. Genetic data analysis. Sinauer Associates, Sunderland, pp. 574.

Weising K, Nybom H, Wolff K, Kahl G. 2005. DNA Fingerprinting. In: Plants: Principles, methods and applications. 2nd edition. London: CRC Press, pp. 444.

Welsh J, McClelland M. 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Research 18:7213-7218.

Wenzl P, Caig V, Carling J, Cayla C, Evers M, Jaccoud D, Peng K, Patarapuwandol S, Uszynski G, Xia L, Yang S, Huttner E, Kilian A. 2004. Diversity array technology, a novel tool for harnessing crop genetic diversity. Proceedings of the National Academy of Science USA 101:9915-9920.

Wenzl P, Li H, Carling J, Zhou M, Raman H, Paul E, Hearnden P, Maier C, Xia L, Caig V, Jaroslava O, Cakir M, Poulsen D, Wang J, Raman R, Smith KP, Muehlbauer GJ, Chalmers KJ, Kleinhofs A, Huttner E, Kilian A. 2006. A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits. BMC Genomics 7:206-228.

Wiley EO. 1981. Phylogenetics: The theory and practice of phylogenetics and systematic. John Wiley, New York, pp. 456.

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey S. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research 18:6531-6535.

Wittenberg AHJ, Van der Lee T, Cayla C, Kilian A, Visser RGF, Schouten HJ. 2005. Validation of the high-throughput marker technology DArT using the model plant *Arabidopsis thaliana.* Molecular Genetic and Genomics 274:30-39.

Wolfe JH. 1970. Pattern clustering by multivariate mixture analysis. Multivariate Behaviour Researcher 5:329-350.

Wright S. 1951. The genetical structure of populations. Annals of Eugenics 15:323-354.

Wright S. 1965. The interpretation of population structure by *F*-statistics with special regard to systems of mating. Evolution 19:395-420.

Wright S. 1978. Variability within and among natural populations. Evolution and the genetics of populations. Chicago: University of Chicago Press, pp. 465.

Xia L, Peng K, Yang S, Wenzl P, Carmen de Vicente M, Fregene M, Kilian A. 2005. DArT for high-throughput genotyping of cassava (*Manihot esculenta*) and its wild relatives. Theoretical and Applied Genetics 110:1092-1098.

Xiao J, Li J, Yuan L, McCouch SR, Tanksley SD. 1996. Genetic diversity and its relationship to hybrid performance in heterosis in rice as revealed by PCR-based markers. Theoretical and Applied Genetics 92:637-643.

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

Yang W, de Oliveira AC, Godwin I, Schertz K, Bennetzen JL. 1996. Comparison of DNA marker technologies in characterizing plant genome diversity: variability in Chinese sorghums. Crop Science 36:1669-1676.

Zeb B, Khan IA, Ali S, Bacha S, Mumtaz S, Swati ZA. 2009. Study on genetic diversity in Pakistani wheat varieties using simple sequence repeat (SSR) markers. African Journal of Biotechnology 8:4016-4019.

Zhang DX, Hewitt GM. 2003. Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. Molecular Ecology 12:563-584.

Zimnoch-Guzowska E, Marczewski W, Lebecka R, Flis B, Schäfer-Pregl R, Salamini F, Gebhardt C. 2000. QTL analysis of new sources of resistence to *Erwinia carotovora* ssp *atroseptica* in potato done by AFLP, RFLP, and resistence-gene-like markers. Crop Science 40:1156-1167.

Chapter 3

Phenotypic diversity in sorghum accessions based on morphological and agronomical traits

3.1 Abstract

A total of 1013 sorghum accessions from Sudan, Eritrea, Ethiopia, Rwanda, Burundi, Tanzania, and Uganda were evaluated for 13 qualitative traits. Phenotypic frequencies between accessions in seven countries were calculated. Phenotypic diversity index, H', was analysed and the result indicated the between countries component of diversity to be relatively smaller than the variation in H' among characters within countries. A total of 920 sorghum accessions from Sudan, Ethiopia, Rwanda, Burundi, Tanzania, and Uganda were evaluated for five quantitative characters to determine the extent of morphological variation. Analysis of variance revealed highly significant differences between accessions pooled over countries and between the six countries. Results suggested the occurrence of significant regional differentiation and existence of significant phenotypic variation between accessions as a whole. Multivariate methods, including cluster analysis, were used on quantitative and combined traits data to estimate the patterns and distribution of phenotypic variation. Cluster analysis grouped accessions into six groups, mainly based on geographical origin. Results in qualitative and quantitative traits data showed that there existed a wide morpho-agronomical diversity among accessions studied. Therefore, future germplasm collection should take all levels of variation into consideration.

3.2 Introduction

Information on the genetic diversity within and among closely related crop species is essential for effective use and management of genetic resources. It is particularly useful in characterizing individual accessions and cultivars, in detecting genetic material with novel genes and thereby rescuing them from erosion, and as a general guide in selecting parents for crossing in breeding programmes (Karp, 2002).

Categorizing germplasm accessions into morphologically similar and presumably genetically similar groups are most useful when the population structure in a collection is unknown (Marshall and Brown, 1975). Genetic relationships among a large number of accessions can be summarized using cluster analysis by placing similar accessions into

groups. Phenotypic diversity indexes of morphological characters and/or multivariate analysis of quantitative characters have been used previously to measure genetic relationships within cereal crop species. Examples include tef (Assefa *et al*., 1999), barley (Tolbert *et al*., 1979; Bekele, 1984; Negassa, 1985; Demissie and Bjornstad, 1996), tetraploid wheat (Tesfaye *et al*., 1991; Bechere *et al*., 1996) and Ethiopian wheat (Negassa, 1986).

Morphological traits, for which the variant allelic phenotypes are adequately discrete to allow their segregation to be followed, are the easiest and generally most economical of all markers to assay. However, discrete morphological traits, though they have high heritability, are limited in number, each being conditioned by a few genes (Karp *et al*., 1996; 1997). Thus, only a small portion of the genome could be covered. They are usually characterized by epistasis, pleiotropy and dominant-recessive relationships, further limiting their value as an ideal genetic marker (Smith and Smith, 1992).

Statistical analysis of quantitative morpho-agronomical traits along with eco-geographic information (De Wet *et al*., 1976) was one of the earliest methods used for estimating genetic diversity in sorghum. It is still widely used to quantify the level and distribution of variation in large sorghum germplasm collections (Prasada Rao and Ramanatha Rao, 1995; Teshome *et al*., 1997; Ayana and Bekele, 1999). Using multivariate analysis procedures, Ayana and Bekele (1999) revealed that the morphological variation in sorghum germplasm from Ethiopia and Eritrea was structured by environmental factors.

In sorghum, as is true for other crop plants, the earliest methods for estimating genetic diversity included Mendelian analysis of discrete morphological traits (Doggett, 1988). Earlier studies of morphological traits have shown that eastern Ethiopian sorghum is believed to be predominantly of the race *durra* (Brooke, 1958; Stemler *et al*., 1977; Doggett, 1988). In a previous study, Geleta (1997) investigated the level of morpho-agronomic trait variability in sorghum landraces from the eastern highland regions of Ethiopia based on quantitative trait data. Using *ex situ* conserved sorghum accessions from Ethiopia and Eritrea, Ayana and Bekele (1998) reported that high and comparable levels of phenotypic variation existed between the regions of origin. Nevertheless, *in situ* patterns of genetic diversity at country as well as regional scale have not been investigated and remain less understood.

In this study, both qualitative and quantitative traits were used to estimate the levels of variation among sorghum accessions grown in eastern Africa. The main objectives of the study were to: (i) estimate the extent of genotypic diversity among sorghum accessions based on 13 qualitative and five quantitative traits, and (ii) assess the regional patterns of phenotypic diversity using qualitative and quantitative traits.

3.3 Materials and methods

3.3.1 Plant material

About 1568 accessions comprising of landraces and farmer varieties from different geographic origins, races and breeding lines from Sudan, Eritrea, Ethiopia, Rwanda, Burundi, Tanzania, and Uganda were selected as the core material for the study. Kenya was excluded from analyses because of failure to describe all the data due to excessive rainfall.

3.3.2 Experimental plot design

Each batch of 196 accessions (Appendix 1) was planted in each country using a different randomization (Table 3.1). In each country three replicates per accession were used to satisfy statistical analysis requirements. A balanced lattice design of 14x14 with 196 genotypes was used. A smaller design was used when and where samples were less than 196. To avoid border effects, three short rows of 3 m each were used per accession. Scoring of phenotypic characters was done on the middle row to counter possible border effects. Inter- and intra-row spacing was decided by the breeder, following common practice in each country.

3.3.3 Qualitative traits

3.3.3.1 Plant materials

A total of 1013 accessions comprising landraces and farmer varieties from different geographic origins, races and breeding lines from Sudan, Eritrea, Ethiopia, Rwanda, Burundi, Tanzania, and Uganda were selected as the core material. Kenya was excluded from the analyses because of failure to describe all data due to excessive rainfall. The number of accessions per country varied due to failure to germinate as well as due to environmental (temperature and water stress) factors. For Eritrea excessive rainfall caused the reduction of accessions used.

Table 3.1 Countries, number of accessions, phenotypic site and site characteristic

Country	No of accessions	Phenotypic site	Site characteristic
Burundi	163	Mahwa	Midlands
Ethiopia	196	Arsingele	Highlands
Rwanda	107	Rwerere	Highlands
Sudan	138	Wad Medani	Lowlands
Tanzania	196	Hambolo	Lowlands
Uganda	120	Serere	Lowland
Eritrea	93	Hahale	Midlands

3.3.3.2 Methods

Quantitative trait evaluation was conducted in the field during 2007. Three plants of each accession were replicated three times to avoid any bias, and randomly selected and tagged just before flowering. To categorize each accession morphologically, published sorghum descriptors (IBPGR/ICRISAT, 1993) were used. Table 3.2 lists the qualitative traits, their descriptors and codes used in analyses. A total of 13 of the presented 21 characters were used in this study due to missing characters in many countries (failure to describe all characters).

3.3.3.3 Data analysis

Phenotypic frequency distributions of the characters were calculated for all accessions. The Shannon-Weaver diversity index (H′) was computed using the phenotypic frequencies to assess the phenotypic diversity for each character for all accessions. The Shannon-Weaver diversity index as described by Perry and McIntosh (1991) is given as:

$$H' = 1-\sum_{i=1}^{n} p_i \log_e p_i$$

where p_i is the proportion of accessions in the i^{th} class of an n-class character and n is the number of phenotypic classes of traits. Each H′ value was divided by its maximum value ($\log_e n$) and normalized in order to keep the values between 0 and 1. By pooling various characters across collection sites, the additive properties of H' were used to evaluate diversity of localities and characters within the population.

Table 3.2 Character, descriptor and codes used for characterization of qualitative traits in sorghum accessions used in the study

Character	Descriptor and code
Plant colour (PC)	Pigmented (1) and tan (2)
Stalk juiciness (SJ)	Not juicy (1) and juicy (2)
Juice flavour (JF)	Sweet (1) and insipid (2)
Panicle compactness and shape (PCS)	Very lax panicle (1), very loose erect primary branches (2), very loose drooping primary branches (3), loose erect primary branches (4), loose drooping primary branches (5), semi-loose erect primary branches (6), semi-loose drooping primary branches (7), semi-compact elliptic (8), compact elliptic (9), compact oval (10), half broom corn (11), and broom corn (12)
Inflorescence exsertion (IE)	Slightly exserted (1), exserted (2), well-exserted (3), and panicle recurved (4)
Glume colour (at maturity) (GLC)	White (1), sienna (2), mahogany (3), red (4), purple (5), black (6), and grey (7)
Grain covering (GC)	25% grain covered (1), 50% grain covered (3), 75% grain covered (5), grain fully covered (7), and glume longer than grain (9)
Awns (at maturity)	Absent (1) and present (2)
Grain colour (GRC)	White (1), yellow (2), red (3), brown (4), and buff (5)
Grain plumpness (GRP)	Dimpled (3) and plump (7)
Grain form (GRF)	Single (1) and twin (2)
Endosperm texture (ET)	Completely corneous (1), mostly corneous (3), intermediate (5), mostly starchy (7), and completely starchy (9)
Senescence (SE)	Very slightly senescent (1), slightly senescent (3), intermediate (5), mostly senescent (7), and completely senescent (9)

3.3.4 Quantitative traits

3.3.4.1 Plant material

About 920 accessions comprising of landraces and farmer varieties from different geographic origins, races and breeding lines from Sudan, Ethiopia, Rwanda, Burundi, Tanzania, and Uganda were selected as the core material (Table 3.1). Kenya and Eritrea were excluded from the analyses because of failure to describe all the data due to excessive rainfall.

3.3.4.2 Parameters measured

Data was collected for five morpho-agronomical characters (Table 3.3). For every accession, data was recorded from the middle row on three randomly selected individual plants, except for days to 50% flowering, which was recorded on plot basis (Table 3.3).

3.3.4.3 Statistical analysis

Pair-wise genetic distance estimates for the mean value of each accession for the five quantitative traits as well as combined data (both qualitative and quantitative) were obtained based on the Gower's distance matrix. Clustering was done using UPGMA and relationships displayed as a phenogram. Analyses were done using the R software programme, version 2.8.1 (R Development Core Team, 2008). In order to draw the un-rooted trees, data was imported into DARwin5 version 5.0.155 software (Perrier *et al.*, 2003). Dissimilarities were estimated based on Euclidean distance matrix and hierarchical clustering analyses were performed. ANOVAs were done using an unbalanced design instead of the square lattice used in planting due to missing entries (failure to germinate) using Genstat software (VSN International Ltd. 2007).

3.4 Results

3.4.1 Qualitative traits

3.4.1.1 Estimates and analysis of diversity

The level of phenotypic diversity based on Shannon-Weaver diversity index (H′) and its partitioning within and between countries are given in Table 3.4. The 13 characters differed in their distribution as well as the amount of variation. The overall average phenotypic diversity (H′) among accessions was 1.00, varying from 0.01 (grain form) to 1.77 (panicle compactness and shape). Grain form was highly monomorphic, while plant colour, stalk juiciness, juice flavour, awns, and grain plumpness were intermediate (H′ of

Table 3.3 Character code and description of the quantitative characters recorded in the study

Character	Code	Description
Plant height (cm)	PHt	Height of the main stalk from the ground to the tip of the panicle
Days to 50% flowering (days)	DF	From emergence to when 50% of plants have started flowering
Panicle length (cm)	PL	Length of panicle from its base to tip
Panicle width (cm)	PW	Width of panicle in natural position at the widest part
1000-seed weight (g)	TSWt	Weight of 1000 seed counts

Table 3.4 Estimates of diversity (H′) and its partitioning into within and between countries for 13 qualitative characters in 1013 sorghum accessions

Character	H′	H_{cl}	H_{cl}/H'	$(H'-H_{cl})/H'$
Plant colour (PC)	0.55	0.23	0.42	0.58
Stalk juiciness (SJ)	0.69	0.51	0.74	0.26
Juice flavour (JF)	0.68	0.44	0.65	0.35
Panicle compactness and shape (PCS)	1.77	0.00	0.00	1.00
Inflorescence exsertion (IE)	1.36	0.92	0.68	0.32
Glume colour (GLC)	1.56	0.39	0.25	0.75
Grain covering (GC)	1.23	0.19	0.15	0.85
Awns	0.58	0.31	0.53	0.47
Grain colour (GRC)	1.45	0.00	0.00	1.00
Grain plumpness (GRP)	0.69	0.26	0.34	0.62
Grain form (GRF)	0.01	0.01	1.00	0.00
Endosperm texture (ET)	1.03	0.19	0.18	0.82
Senescence (SE)	1.45	0.00	0.00	1.00
Average	1.00	0.27	0.38	0.62

H′ = Diversity index for each character calculated from entire data set; **H_{cl} =** Average diversity index of each character for the seven countries; **H_{cl}/H' =** Proportion of diversity within countries; **$(H'-H_{cl})/H'$** = Proportion of diversity between countries in relation to the total variation.

0.55-0.69), and panicle compactness and shape, inflorescence exsertion, glume colour, grain covering, grain colour, endosperm texture, and senescence were highly polymorphic (H′ of 1.03-1.77). The overall mean of the proportion of diversity within countries (H_{cl}/H') was 0.38 ranging from 1.00 for grain form to 0.00 for the panicle compactness and shape, grain colour, and senescence. The overall average of the proportion of diversity between countries in relation to the total variation was 0.62 with the highest value (1.00) for panicle compactness and shape, grain colour and senescence and lowest for grain form (0.00) (Table 3.4).

The H′ pooled across characters by country ranged from 0.08 to 0.51 (Table 3.5). The countries that had the highest H′ were Sudan (0.51), Tanzania (0.37), and Ethiopia (0.30). The lowest mean values of H′ were from Rwanda (0.08) and Burundi (0.18). The highest diversity index was obtained for inflorescence exsertion (IE) (0.92), while an intermediate diversity index was obtained for stalk juiciness (SJ) (0.51), juice flavour (JF) (0.44), and glume colour (GLC) at maturity (0.39) and the lowest values (0.00) for panicle compactness and shape (PCS), grain colour (GRC), grain form (GRF), and senescence (SE).

3.4.1.2 Character distribution

The frequency distribution of the different classes for the 13 qualitative characters is shown in Table 3.6. Seventy six percent of the 1013 germplasm samples were found to have pigmented colour across all countries. However, samples from Burundi and Tanzania showed a high frequency of tan coloured plants. Juicy stalks were present in 44% of samples. Most sorghum accessions with juicy stalks were grown in Tanzania, Burundi, and Eritrea, while non-juicy stalks were mainly grown in Ethiopia, Sudan, and Rwanda. The distribution of juice flavour followed the same trend with 42% of samples having sweet juice.

Table 3.5 Estimates of the Shannon-Weaver diversity index (H′) for 13 qualitative characters in sorghum accessions by country

Countries	Characters													
	PC‡	SJ	JF	PCS	IE	GLC	GC	Awn	GRC	GRP	GRF	ET	SE	Mean
Burundi	0.54	0.68	0.69	0.00	0.00	0.00	0.00	0.44	0.00	0.00	0.04	0.00	0.00	0.18
Eritrea	0.00	0.45	0.31	0.00	1.21	0.00	0.00	0.69	0.00	0.00	0.00	0.00	0.00	0.21
Rwanda	0.00	0.47	0.44	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.00	0.00	0.00	0.08
Sudan	0.00	0.56	0.54	0.00	1.37	1.64	1.34	0.54	0.00	0.66	0.00	0.00	0.00	0.51
Tanzania	0.69	0.46	0.38	0.00	1.27	0.00	0.00	0.06	0.00	0.67	0.00	1.31	0.00	0.37
Uganda	0.15	0.60	0.54	0.00	1.22	0.00	0.00	0.00	0.00	0.34	0.00	0.00	0.00	0.22
Ethiopia	0.23	0.37	0.15	0.00	1.33	1.09	0.00	0.44	0.00	0.06	0.00	0.00	0.00	0.30
Mean	0.23	0.51	0.44	0.00	0.92	0.39	0.19	0.31	0.00	0.26	0.01	0.19	0.00	

‡ Character abbreviations as defined in Table 3.2.

Table 3.6 Percentage frequency distribution of different phenotypic classes for 13 qualitative characters in sorghum for seven countries*

Country	Plant colour		Stalk juiciness		Juice flavour		Panicle compactness and shape											
Descriptor code	1	2	1	2	1	2	1	2	3	4	5	6	7	8	9	10	11	12
Burundi	4	12	7	9	9	7	0	0	0	0	0	1	0	9	4	1	1	0
Eritrea	9	0	2	8	8	1	0	0	0	0	1	0	0	5	0	2	0	0
Ethiopia	18	1	17	2	1	19	0	0	0	0	1	0	6	6	3	1	1	2
Rwanda	11	0	9	2	2	9	0	0	0	0	0	2	0	5	4	0	0	0
Sudan	14	0	10	3	3	10	0	0	0	0	0	0	0	4	5	3	0	0
Tanzania	9	10	3	16	17	2	0	0	6	0	0	1	2	6	2	1	2	0
Uganda	11	0	8	3	3	9	0	0	0	0	2	2	0	3	3	0	0	0
All	76	24	56	44	42	58	0	0	6	0	5	6	9	38	21	8	3	2

Country	Inflorescence exsertion				Glume colour							Grain covering				
Descriptor code	1	2	3	4	1	2	3	4	5	6	7	1	3	5	7	9
Burundi	0	1	13	2	3	0	0	2	0	11	0	11	3	1	0	0
Eritrea	5	1	1	2	0	0	0	0	0	9	0	0	5	3	2	0
Ethiopia	4	7	3	5	0	0	1	0	1	3	13	6	7	4	1	2
Rwanda	3	5	2	0	0	0	0	0	0	11	0	10	1	0	0	0
Sudan	3	3	3	4	3	0	2	1	0	3	3	2	7	3	1	0
Tanzania	7	6	4	2	0	5	4	6	0	4	0	8	9	1	1	0
Uganda	4	5	3	1	0	0	0	6	0	5	0	12	0	0	0	0
All	26	28	30	16	6	6	8	16	2	46	16	48	32	12	5	2

* Code descriptors as given in Table 3.2.

Table 3.6 Percentage frequency distribution of different phenotypic classes for 13 qualitative characters in sorghum for seven countries* (continued)

Country	Awn		Grain colour					Grain plumpness		Grain form	
Descriptor code	1	2	1	2	3	4	5	3	7	1	2
Burundi	13	3	1	0	14	1	0	16	0	16	0
Eritrea	5	4	2	2	1	3	1	9	0	9	0
Ethiopia	3	16	6	2	4	3	4	0	19	20	0
Rwanda	11	0	0	0	10	0	0	0	10	11	0
Sudan	10	3	4	6	2	2	0	8	5	14	0
Tanzania	19	0	9	2	3	5	0	12	8	19	0
Uganda	12	0	2	0	6	4	0	11	1	12	0
All	73	27	24	13	39	19	5	56	44	100	0

Country	Endosperm texture					Senescence				
Descriptor code	1	3	5	7	9	1	3	5	7	9
Burundi	0	0	0	0	16	0	13	3	0	0
Eritrea	0	0	0	8	1	0	0	9	0	0
Ethiopia	0	0	0	7	12	0	2	8	10	0
Rwanda	0	0	0	3	7	0	1	6	0	3
Sudan	0	0	3	5	5	0	1	3	8	1
Tanzania	0	2	5	2	9	3	9	6	1	0
Uganda	0	0	0	5	7	7	4	1	0	0
All	0	3	10	30	57	10	30	36	18	4

* Code descriptors as given in Table 3.2.

Of the 12 phenotypic classes for panicle compactness and shape, 38% accessions were semi-compact elliptic type samples, followed by compact elliptic (21%). Samples from Ethiopia and Tanzania showed semi-loose drooping primary branches, while compact oval branches were seen in Sudan and Eritrea. There were four phenotypic classes observed for inflorescence exsertion and the well-exserted character was dominant, while slightly exserted and exserted mainly occurred in Tanzania and Ethiopia, respectively.

Samples from Ethiopia showed a high frequency of grey glume colour, which agreed with Ayana and Bekele (1998) who reported that the most predominant glume colour in sorghum samples from Hararge was grey. The most frequent glume colour in this study was black. The class of 25% grain covered by glumes was the most frequent (48%) for all countries, while Tanzania mainly showed a 50% covering. Most of the germplasm samples (73%) were awnless across all countries, while the majority of samples from Ethiopia had awns.

The most abundant grain colour for most countries was red (39%) and white (24%). All samples from Rwanda were red. The dominant grain colour in Ethiopia and Sudan was white and yellow, respectively. The occurrence of dimpled grain was more frequent (56%) than plumped grain (44%) for all samples. All samples from Burundi and Eritrea were dimpled, while all samples from Ethiopia and Rwanda were plumped. All samples had the single grain form, although Ayana and Bekele (1998) and Abdi *et al.* (2002) found twin seeded grains in their study.

Endosperm texture was mostly completely starchy (57%) and only a few were of the mostly corneous type (3%) with total absence of completely corneous endosperm texture. The intermediate (36%) and slightly senescent (30%) types of senescence, out of the five phenotypic classes, were most frequent in this study.

3.4.2 Quantitative traits

3.4.2.1 Clustering based on quantitative data

The dendrograms drawn based on Gower's distance (Figures 3.1-3.3) show the clustering of sorghum accessions based on quantitative data. Figure 3.1 represents a hierarchical clustering of the 920 sorghum accessions based on the five quantitative traits and using Gower's distance. The same data was used to draw Figure 3.2, but in order to obtain a

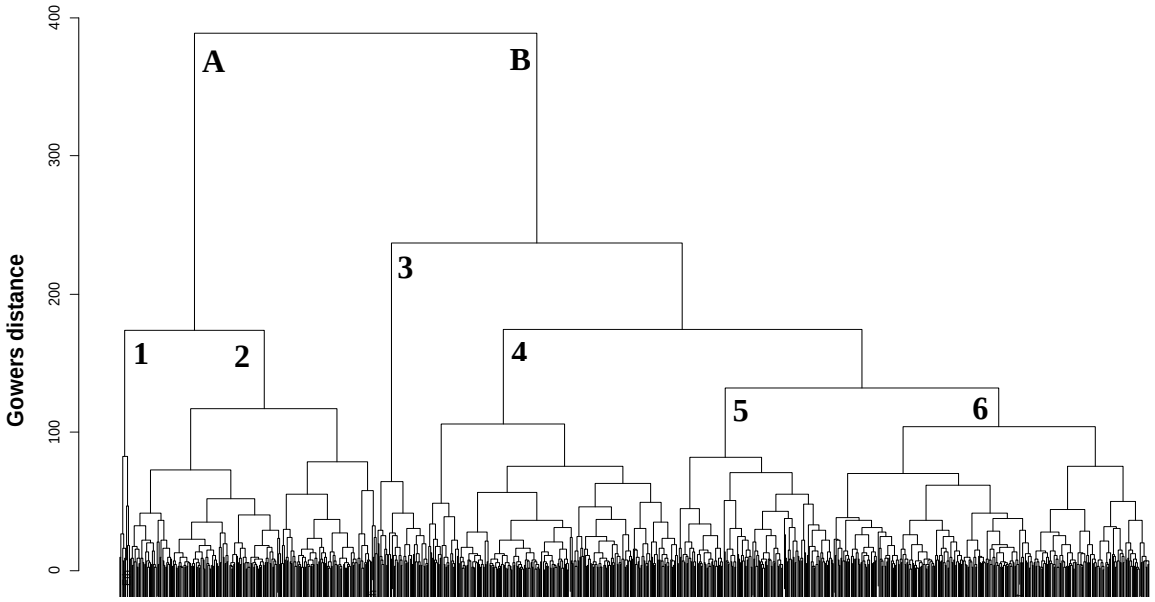

Figure 3.1 Hierarchical dendrogram, based on Gower's distance and UPGMA clustering, showing cluster groups among the 920 sorghum accessions based on five quantitative traits.

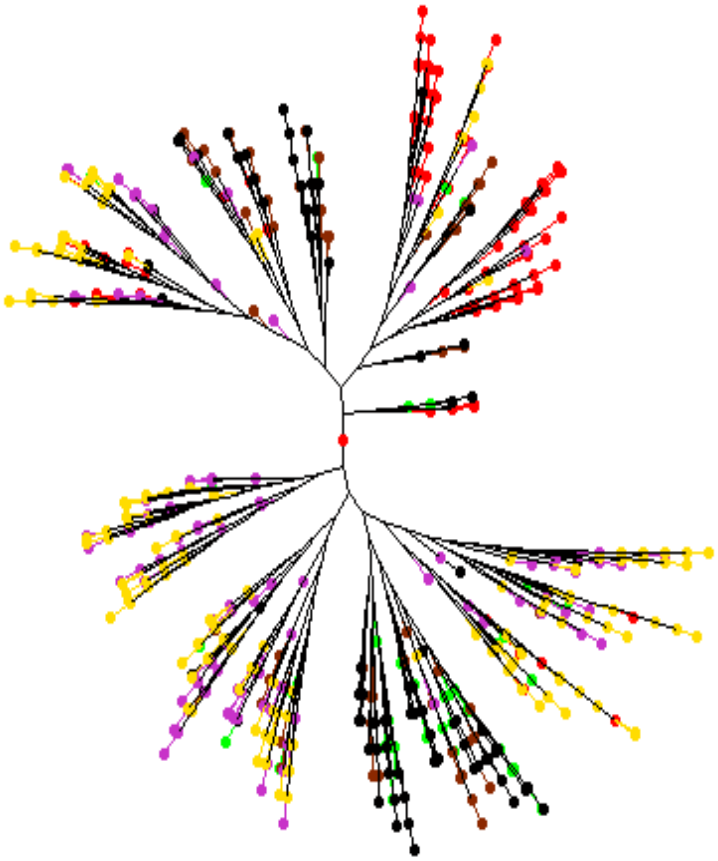

Figure 3.2 **Unrooted tree drawn using Euclidian distances and hierarchical clustering in DARwin5 software. The tree shows cluster groups among the 920 sorghum accessions based on five quantitative traits.**

Burundi, Rwanda, Uganda, Sudan, Tanzania, and Ethiopia.

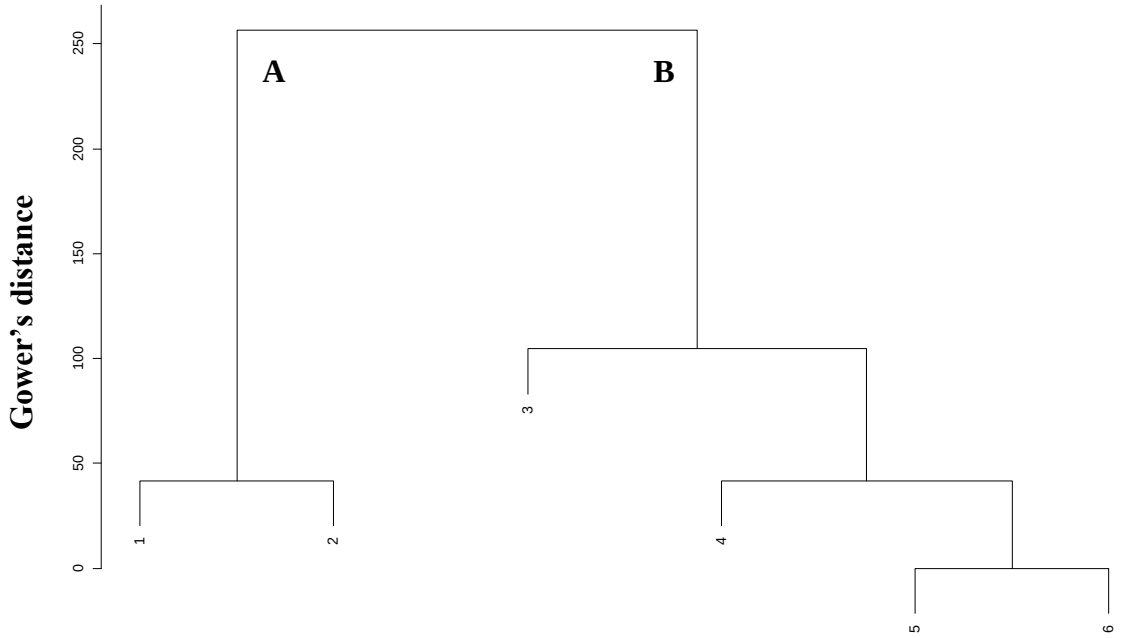

Figure 3.3 **Simplified dendrogram showing the main cluster groups among the 920 sorghum accessions based on five quantitative traits, Gower's distance and UPGMA clustering.**

better visual representation of the 920 accessions based on Euclidean distance, clustering was done using an unrooted tree. In order to simplify the data further, a dendrogram showing only the six main groups (Figure 3.3) was drawn. Two major clusters (A and B, Figure 3.3) were formed at a genetic distance of about 2.5. The first major group (A) contained about 25% of the accessions and was divided into two subgroups at a genetic distance of 0.4 which included groups 1 and 2 [group 1 (1% of accessions) included accessions from Burundi and Ethiopia while group 2 (24% of accessions) included accessions from all countries except Rwanda] (Table 3.7). The second major group (B) was further divided into two subgroups at a genetic distance of 1 (Figure 3.3). The first of these subgroups (group 3) mainly included accessions from Tanzania. The second of these subgroups was further divided into two subgroups, the first of these subgroups (group 4) formed at a genetic distance of about 0.4 (with 24% of the accessions), and the second group consisted of the rest of the accessions. In total, six clusters were formed (Table 3.7). The second main cluster (B) contained about 75% of the accessions, and wide intra-cluster variation was observed.

Table 3.7 Distribution of 920 sorghum accessions by country into six clusters using average values of quantitative characters

Country	Cluster						Country total
	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	
Burundi	6	126	0	28	1	2	163
Ethiopia	4	68	0	65	9	50	196
Rwanda	0	0	0	19	3	85	107
Sudan	0	10	1	35	39	53	138
Tanzania	0	5	45	12	80	54	196
Uganda	0	12	0	62	7	36	120
Total	10	221	46	224	139	280	920

About 77% of Burundi's accessions were accumulated in group 2, while 79% of Rwanda's accessions were found in group 6, and 52% of Uganda accessions were found in group 4. Ethiopian accessions were spread throughout all groups except group 3, while Sudanese and Tanzanian materials were absent from group 1.

3.4.2.2 Morphological and agronomic variability

Analysis of variance (ANOVA) revealed highly significant ($p<0.001$) differences among accessions for all quantitative traits studied (Table 3.8). There was a wide range of expression across accessions for all traits, including a 105 day range in 50% days to flowering, 385.5 cm range in plant height, 35 cm range in panicle length, 22 cm in panicle width and 8.8 g range in 1000-seed weight.

Highly significant differences ($p<0.001$) were found between accessions pooled over countries for the five characters studied (Table 3.8) and between the seven countries of origin for the 920 sorghum accessions for the five characters (Table 3.9). Results suggested the occurrence of significant regional differentiation and existence of significant phenotypic variation between accessions as a whole.

Table 3.8 Statistical analysis of five quantitative characters

Character	Min	Max	Mean	St Dev	SE±	Variance	CV%
PHt[†]	50.0	435.5	234	72.1	2.4	5198.5**	30.9
DF	42.0	147.7	86.25	19.6	0.6	382.6**	22.7
PL	5.0	40.0	21.0	5.72	0.19	32.8**	27.3
PW	1.0	23.0	7.00	2.44	0.08	5.95**	35.2
TSWt	0.8	9.60	2.10	0.99	0.03	0.97**	40.1

[†] Character abbreviations as defined in Table 3.3. **P<0.001
Min = Minimum; Max = Maximum; St Dev = Standard deviation; SE = Standard error; CV = Coefficient of variation.

Country means for all characters are shown in Table 3.9. Significant country differentiation was observed for mean plant height, mean days to 50% flowering, mean panicle length, mean panicle width and for mean 1000-seed weight.

The mean for plant height for accessions from Burundi was higher than those for all other countries. The highest mean for number of days for 50% flowering was observed for the accessions from Rwanda followed by Ethiopia, while accessions from Sudan had the shortest number of days. Accessions from Ethiopia and Rwanda were noted for having higher means for panicle length and panicle width than those from other countries, while accessions from Sudan had the shortest panicle length and accessions from Tanzania the

shortest panicle width. The highest mean values for 1000-seed weight were observed for Sudan, while the lowest mean value was found in accessions from Rwanda. Accessions from Sudan could thus be a good source of early flowering, moderate plant height and high 1000-seed weight, for which there is an urgent need in eastern Africa.

Table 3.9 Country means for the five quantitative characters in sorghum

	Character				
Country	**PHt**[†]	**DF**	**PL**	**PW**	**TSWt**
Burundi	321.9	78.15	20.48	7.79	1.94
Ethiopia	258.5	100.19	23.30	8.77	3.06
Rwanda	206.2	122.59	23.11	9.15	1.60
Sudan	214.3	65.43	18.92	5.69	3.73
Tanzania	157.2	81.90	19.69	4.07	2.21
Uganda	243.8	72.76	20.21	6.97	1.85
Variance	233.35**	659.68**	17.15**	238.14**	208.80**
SE±	47.96	9.12	5.48	1.61	0.67
CV%	20.5	10.6	26.2	23.3	27.5

[†]Character abbreviations as defined in Table 3.3. **P<0.001

SE = Standard error; CV = Coefficient of variation.

3.4.3 Combined quantitative and qualitative traits

3.4.3.1 Clustering based on combined quantitative and qualitative traits

The dendrograms (Figures 3.4-3.6) show the clustering of sorghum accessions based on both qualitative and quantitative data. Two major clusters (A and B) were formed at a genetic distance of about 0.07 (Figure 3.6). The first major group (A) contained about 26% of the accessions. Subgroups at a genetic distance of 0.01 included groups 1 and 2 (group 1 mainly included accessions from Sudan while group 2 mainly included accessions from Ethiopia) (Table 3.10). The second major group (B) was further split into two subgroups at a genetic distance of 0.06. The first of these subgroups (group 3) included 28% of all accessions that included accessions from Burundi, Tanzania, Sudan, and Uganda. The second of these subgroups was further split into two sub-subgroups at a genetic distance of 0.03. Group 4 which contained accessions from all the countries except Burundi and Rwanda but mainly consisted of accessions from Tanzania.

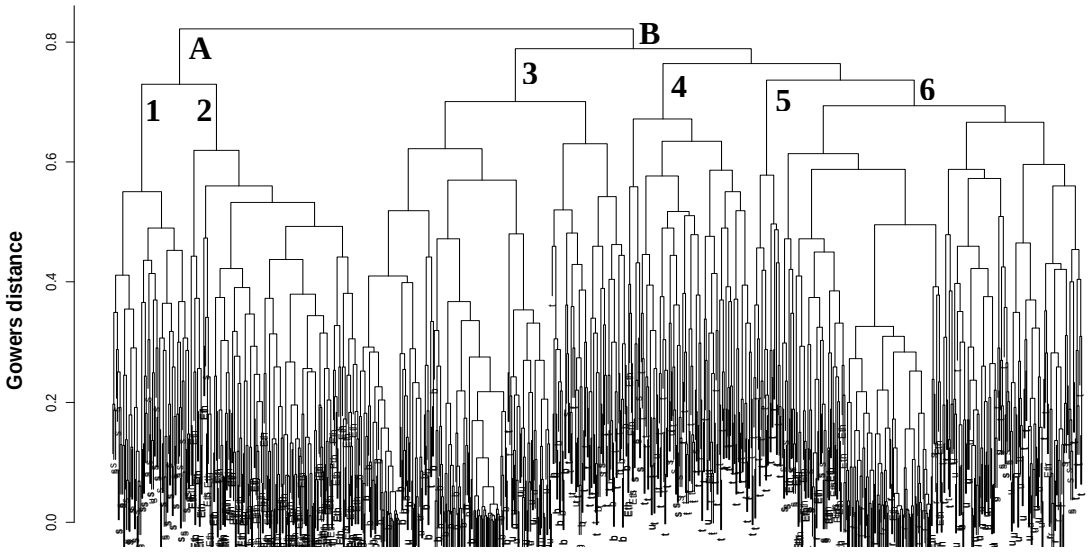

Figure 3.4 **Dendrogram showing cluster groups among the 920 sorghum accessions based on combined quantitative and qualitative traits using Gower's distance and UPGMA clustering.**

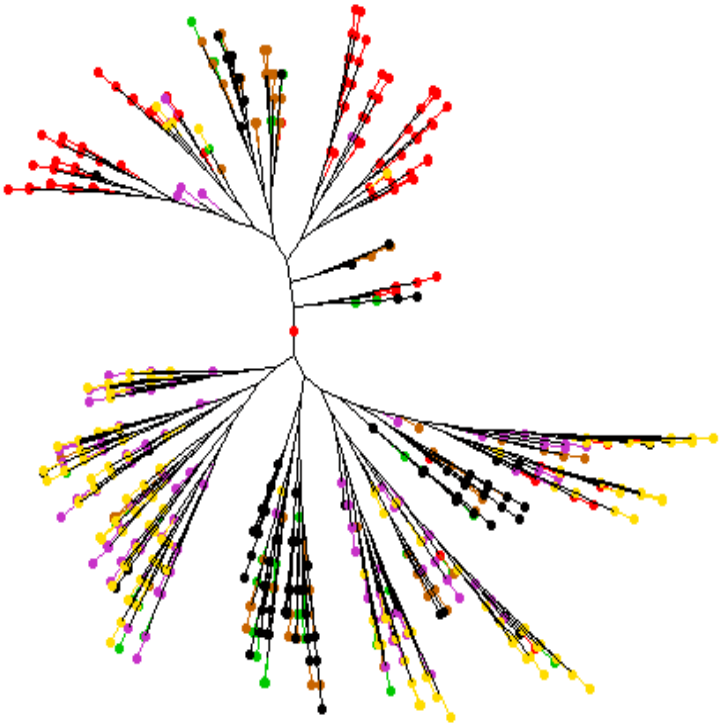

Figure 3.5 **Unrooted tree showing cluster groups among the 920 sorghum accessions based on combined traits using Euclidean distance matrix and hierarchical clustering in DARwin5 software.**

Burundi, Rwanda, Uganda, Sudan, Tanzania, and Ethiopia.

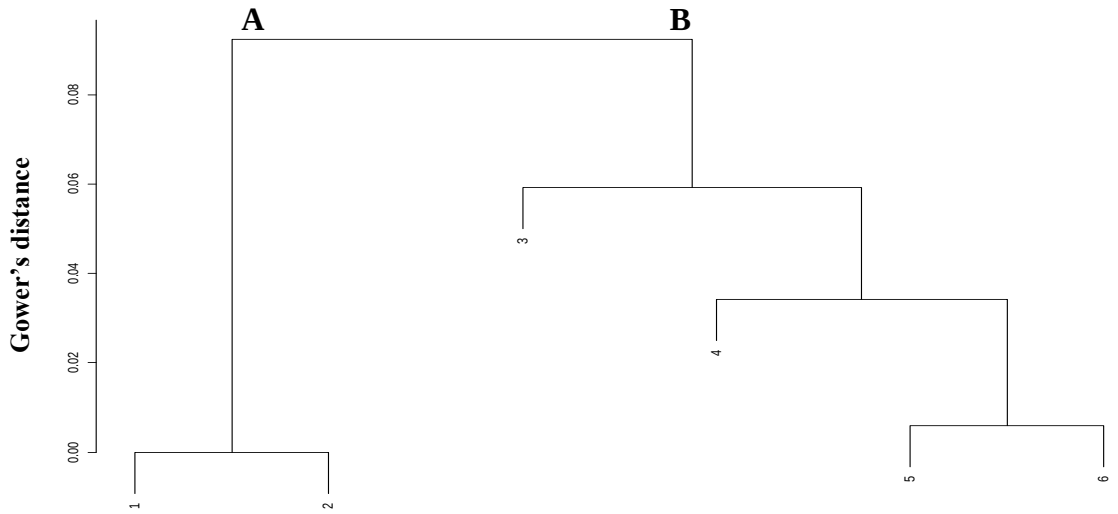

Figure 3.6 Simplified dendrogram showing the main cluster groups among the 920 sorghum accessions based on combined trait data.

The second of these sub-subgroups was further split into two subgroups at a genetic distance of 0.01 and included groups 5 and 6. Group 5 contained only Tanzanian accessions while group 6 contained accessions from all the countries with the highest percentage (38%) from Rwanda. In total, six clusters were formed (Table 3.10). The second cluster contained about 75% of the accessions, and a wide intra-cluster variation was observed.

Table 3.10 Distribution of the 920 sorghum accessions into six clusters by country using average values of combined quantitative and qualitative traits

Country	Cluster						Country total
	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	
Burundi	0	0	161	0	0	2	163
Ethiopia	0	157	0	6	0	33	196
Rwanda	0	0	0	0	0	107	107
Sudan	73	4	4	17	0	40	138
Tanzania	0	0	44	90	27	35	196
Uganda	1	0	46	7	0	66	120
Total	74	161	255	120	27	283	920

About 99% of Burundi's accessions were located in group 3, with the rest in group 6. Ethiopian accessions were mainly accumulated in group 2 (80%), while 100% of Rwanda's accessions were found in group 6. Sudanese accessions were spread throughout all groups except in group 5, with the majority in group 1 (53%). Tanzanian materials were absent in groups 1 and 2 with the majority in group 4 (46%). Accessions from Uganda were absent in groups 2 and 5 with the majority in group 6 (55%).

3.5 Discussion

3.5.1 Qualitative traits

The overall mean diversity index (H′) (0.01-1.77) obtained in the present study was higher than those of Ayana and Bekele (1998) (overall H′ for Welo = 0.07-0.75) and Abdi

et al. (2002) (overall H′ for North Shewa and South Welo = 0.04-0.77). Reasons for this high diversity of sorghum accessions in the current study might be due to the fact that these accessions were found in relatively complex and heterogeneous ecologies and the non-uniform climatic conditions. Results of the present study will help to advocate and define the need for proper strategies for future collection of germplasm and for their *in situ* and *ex situ* conservation and utilization. Such a consideration has previously been discussed by Ayana and Bekele (1998) and Dje *et al*. (1998; 1999) as one of the restrictions in developing strategies for preservation.

Panicle compactness and shape in this study showed high variation between countries. This could be explained in two ways. Firstly, the patterns of distribution of the different panicle types appeared to follow temperature, humidity, and rainfall patterns. Results detected mainly semi-compact and compact panicle types relating to the relatively hot and dry regions of Sudan and Eritrea while Tanzania and Ethiopia mainly showed a high frequency of very loose drooping primary branches, indicating relatively cool and wet regions. Ayana and Bekele (1998) reported that the compact and semi-compact panicle types were more frequent in relatively hot and dry regions of Eritrea, while the loose panicle types with drooping primary branches occurred abundantly in relatively cool and wet regions of Ethiopia. This finding corroborated suggestions of Stemler *et al.* (1975) and Prasada Rao and Mengesh (1981) that the open panicle of sorghum is an adaptive trait which facilitates quick drying of the panicle in areas of high rainfall and humidity, thereby minimizing grain weathering due to fungal diseases such as grain mould. Secondly, the distribution of the different panicle types reflected the distribution of different races of sorghum (Harlan and De Wet, 1972; Stemler *et al.*, 1977; Dogget, 1988). Race *durra* (characteristic of a compact oval panicle) is grown in Africa along the northeastern coastal region (Sudan, Eritrea, and Ethiopia). The presence of compact, semi-compact, very loose and semi-loose types in this study confirmed the occurrence of race *durra*, *caudatum*, *guinea* and *bicolor* types (Stemler *et al.*, 1977; Harlan, 1992).

Results on inflorescence exsertion in the current study indicated that the majority of accessions had slightly exserted, exserted to well exserted inflorescence, with the latter being dominant. Geleta and Labuschagne (2005) found that a well-exserted character was

dominant in germplasm from the eastern highland of Ethiopia, confirming results of this study.

Results obtained from covered grain in the current study agreed with the finding of Geleta and Labuschagne (2005) who reported that 25% grain covered by glumes was the most frequent for the eastern highland of Ethiopia, while Ayana and Bekele (1998) reported a grain covering ranging from 25-100% for samples from Hararge with 75% covering being the most frequent. The discrepancy between these results could be due to the nature of the samples considered. Of the accessions tested in the current study, 48% showed 25% grain coverage and 32% showed 50% coverage. Thus grain coverage was mainly 25-50%.

Most of the germplasm samples (73%) were awnless across all countries, while the majority of samples from Ethiopia had awns. However, in a previous study, the frequency distribution of awned sorghum landraces for all populations was reported to be fairly equal in the Welo region of Ethiopia (Ayana and Bekele, 1998), while Geleta and Labuschagne (2005) reported that most of the germplasm sampled from the eastern highland of Ethiopia had awns. The deviation is most probably due to the nature of the sample considered in the three studies.

The most abundant grain colour for most countries was red and white. The dominant grain colour in Ethiopia and Sudan was white and yellow, respectively. Results from this study agreed with that reported by Abdi *et al.* (2002) when they found that the grain colour for sorghum landraces from Ethiopia was yellow. Geleta and Labuschagne (2005) demonstrated that the most frequent grain colour in their study was white, yellow and red. Ayana and Bekele (1998) indicated that both human and natural selection could be the cause for high variation in Ethiopian grain colour.

Dimpled grain (56%) was more frequent than plumped grain (44%). This result deviated from results reported by Teshome *et al.* (1997) and Abdi *et al.* (2002), where about 69% and 83% of the grain were recorded as dimpled, respectively. Abdi *et al*. (2002) indicated that the variation may be related to natural and farmer selections prevailing in the region, which may cause changes in the genetic diversity of a specific trait through time. All samples had the single grain form. This was in contrast with Ayana and Bekele (1998) and Abdi *et al.* (2002) who found twin seeded grains in their study.

Sorghum grain is an energy-providing food because of its high starch content (Hulse *et al.*, 1980). The detection of 57% completely starchy endosperm texture in the current study could likely be related to the high milling quality of sorghum grains and its nutritional quality, as sorghum flours are fermented and used to make a wide range of pancakes in India and most parts of Africa. The high frequency of starchy endosperm in this study confirmed results from previous studies (Kebede, 1991; Ayana and Bekele, 1998; Abdi *et al.*, 2002).

In crops, geographic factors that reflect social and political differences may be as significant as ecological factors in determining the distribution of genetic diversity. In many studies, using genebank accessions, the country of origin appears to be an extremely important factor in determining the distribution of genetic diversity (Allard *et al*., 1972; Kahler and Allard, 1981; Saghai-Maroof *et al*., 1990). Two previous studies which used qualitative and quantitative data supported the absence of a clear grouping of accessions based on geographical origin (Teshome *et al*., 1997; Ayana and Bekele, 1999). However, accessions in this study mainly grouped according to geographical origin.

3.5.2 Quantitative traits

The country's variance for days to flowering (659.68) was greater than variance between accessions pooled over countries (382.6). This agreed with reports of Ayana and Bekele (2000) when they studied 415 sorghum accessions representing different regions of Ethiopia, Eritrea, and a group of introduced accessions. Their results showed significant regional variation for days to flowering. Appa Rao *et al.* (1996) studied morphological diversity in sorghum germplasm from India and reported substantially higher differences among different states than within states of India. Since significant variation was found between countries and between accessions within countries, it would be necessary in future to collect from as many countries as possible and adequately sample the variable populations from different localities in a country in order to sample the variation.

In general, accessions from Ethiopia and Rwanda were characterized by late flowering and longer panicle length and wider panicle width. In contrast, accessions from Sudan were characterized by moderate plant height and early flowering, suggesting the possibility of obtaining genes for early and short stature from accessions of Sudan. Accessions from Sudan had relatively short and narrow panicles which are attributes of

the race *durra* (Dogget, 1988; Dogget and Prasada Rao, 1995). On the other hand, long and fewer primary branches, which are characteristic features of race *bicolor* and *guinea* (Doggett, 1988; Doggett and Prasada Rao, 1995), were recorded for Ethiopia, Rwanda, Burundi and Uganda. In other words, there were similarities between accessions from Sudan and Tanzania, between Ethiopia and Rwanda and between Burundi and Uganda.

The coefficient of variation observed for most of the characters agreed with those reported by Abu-el-Gasim and Kambal (1975) for indigenous sorghum of Sudan, and by Ayana and Bekele (2000) for sorghum germplasm from Ethiopia and Eritrea. The coefficients of variation of this study were 30.9 for plant height, 22.7 for days to flowering, 27.3 for panicle length, 35.2 for panicle width, and 40.1 for 1000-seed weight that compared well with the values obtained within the ranges of study by Ayana and Bekele (2000) e.g. 11.49 to 61.81 for plant height, 6.24 to 30.59 for days to flowering, 12.63 to 29.12 for panicle length, 11.32 to 98.55 for panicle width, and 9.99 to 74.62 for 1000-seed weight.

The wide range of variation detected by the various univariate statistics for the characters studied was in agreement with previous studies on Ethiopian sorghum germplasm (Gebrekidan, 1973; Kebede, 1991), and on sorghum germplasm from Ethiopia and Eritrea (Ayana and Bekele, 2000). The different levels of regional variability of a particular character could be due to differences in forces of selection and/or differences in the intensity of a particular selection force. Similar results were reported in tetraploid and hexploid wheat (Bekele, 1984), and in tetraploid wheat (Pecetti and Damania, 1996).

With the hierarchical clustering analyses, considerable variability was observed between and within clusters. In spite of the fact that accessions have been collected from different countries of east Africa, some accessions were similar and occurred in the same clusters while some accessions originally collected in the same country were placed in different clusters. The two dendrograms based on quantitative and combined data both separated the 920 sorghum accessions into six groups. However, the two dendrograms grouped the accessions rather differently. For example, accessions from Burundi and Rwanda mainly grouped in one group each in both dendrograms, while Ethiopian accessions mainly grouped into one group in the combined dendrogram but mainly into three groups in the quantitative dendrogram. Accessions from Sudan, Tanzania, and Uganda were found in

most of the groups in both dendrograms. Overlapping of the clustering patterns of the sorghum accessions was a hint of the lack of large sorghum accession differentiation, which could indicate the presence of gene flow among accessions. On the other hand, even though some accessions were originally collected from different countries, they shared some morphological characteristics, for example accessions from Rwanda and Uganda clustered together in the quantitative dendrogram. However, all clusters were different from one another, indicating that sorghum accessions in the study were morphologically variable with some extent of non-uniformity (Abdi *et al.*, 2002).

The large variation observed in this study and previous studies (Gebrekidan, 1973; Teshome *et al.*, 1997; Ayana and Bekele, 2000) could be credited to many factors. One is the fact that sorghum is grown in different environmental conditions. These include different rainfall, temperature, altitude, and growing periods. Other factors are cultural, historical, and economic system differences among people who cultivate sorghum (Stemler *et al.*, 1977), which contribute to its variation. The various physical, biological, and human factors as well as complex interaction among such factors all seem to have contributed to the wide range of variation of the crop in each country. Another source of variation must come from gene flow between cultivated sorghum and its wild weedy relatives (Doggett and Majisu, 1968).

3.6 Conclusions

The overall diversity index indicated the existence of a wide range of genetic variability in the indigenous eastern African sorghum germplasm. Generally, there was high morphological diversity, implying that *in situ* conservation has been present and genetic erosion has not been widespread. Moreover, results of this study support the hypothesis that eastern Africa is a centre of domestication for sorghum.

The outcome from this study may help in structuring future germplasm collections in many geographical areas instead of collecting comprehensively within individual regions. Nevertheless, future collection operations of sorghum germplasm as source of diversity should also account distribution of variation. Priorities of germplasm collection should focus on areas with relatively large variation, accordingly, future germplasm collection should take into account all levels of variation, and especially plant height and maturity are important characters in sorghum adaptation and breeding.

Results of this study will help to advocate and define the need for appropriate strategies for future collection of germplasm and for *in situ* and *ex situ* conservation and utilization. The role of farmer management adds another aspect to the cause of diversity. This effort underlines the necessity of complementary *in situ* conservation strategies for sorghum genetic resources whereby preservation effects are related to country development projects that highlight the maintenance of conventional farming systems by relying on and giving awareness to the preservation of biological and genetic diversity in these systems. On-farm conservation by small scale farmers not only influences patterns of distribution of different traits but also determines the adaptive qualities of combined traits to be maintained for greater yield and dry matter composition.

3.7 References

Abdi A, Bekele E, Asfaw Z, Teshome A. 2002**.** Pattern of morphological variation of sorghum (*Sorghum bicolor* (L.) Moench) landraces in qualitative characters in North Shewa and South Welo, Ethiopia. Hereditas 137:161-172.

Abu-el-Gasim EH, Kambal AE. 1975. Variability and inter-relations among characters in indigenous grain sorghum of the Sudan. East African Agricultural and Forestry Journal 41:125-133.

Allard RW, Babbel GR, Clegg MT, Kahler AL. 1972. Evidence of coadaptation in *Avena barbata*. Proceedings of the National Academy of Science USA 69:3043-3048.

Appa Rao S, Prasada Rao KE, Mengesha MH, Gopal Reddy V. 1996. Morphological diversity in sorghum germplasm from India. Genetic Resources and Crop Evolution 43:559-567.

Assefa K, Ketema S, Tefera H, Nguyen HT, Blum A, Ayele M, Bai G, Simane B, Kefyalew T. 1999. Diversity among germplasm lines of the Ethiopian cereal tef [*Eragrostis tef* (Zucc.) Trotter]. Euphytica 106:87-97.

Ayana A, Bekele E. 1998. Geographical patterns of morphological variation in sorghum (*Sorghum bicolor* (L.) Moench) germplasm from Ethiopia and Eritrea: qualitative characters. Hereditas 129:195-205.

Ayana A, Bekele E. 1999. Multivariate analysis of morphological variation in sorghum (*Sorghum bicolor* (L.) Moench) germplasm from Ethiopia and Eritrea. Genetic Resources and Crop Evolution 46:273-284.

Ayana A, Bekele E. 2000. Geographical pattern of morphological variation in sorghum (*Sorghum bicolor* (L.) Moench) germplasm from Ethiopia and Eritrea: Quantitative characters. Euphytica 115:91-104.

Bechere E, Belay G, Mitiku D, Merker A. 1996. Phenotypic diversity of tetraploid wheat landraces from north-central regions of Ethiopia. Hereditas 124:165-172.

Bekele E. 1984. Analysis of regional patterns of phenotypic diversity in the Ethiopian tetraploid and hexaploid wheats. Hereditas 100:131-154.

Brooke C. 1958. The durra complex in the central highlands of Ethiopia. Economic Botany 2:192-204.

De Wet JMJ, Harlan JR, Price EG. 1976. Variability in sorghum bicolor. In: Origin of African plant domestication. Edited by Harlan JR, De Wet JMJ, Stemler ABL. Mouton Publishers, the Hauge, Paris, pp. 453-462.

Demissie A, Bjornstad A. 1996. Phenotypic diversity of Ethiopian barley in relation to geographical regions, altitudinal range, and agro-ecological zones: as an aid to germplasm collection and conservation strategy. Hereditas 124:17-29.

Dje Y, Ater M, Lefebvre C, Vekemans X. 1998. Patterns of morphological and allozyme variation in sorghum landraces of northwestern Morocco. Genetic Resources and Crop Evolution 45:541-548.

Dje Y, Forcioli D, Ater M, Lefebvre C, Vekemans X. 1999. Assessing population genetic structure of sorghum landraces from north-western Morocco using allozyme and microsatellite markers. Theoretical and Applied Genetics 99:57-163.

Doggett H. 1988. Sorghum. Longman Scientific and Technical Essex, England, pp. 512.

Doggett H, Majisu BN. 1968. Disruptive selection in crop improvement. Heredity 23:1-22.

Doggett H, Prasada Rao KE. 1995. Sorghum. In: Evolution of Crop Plants, 2nd edition. Edited by Smartt J, Simmonds NW. Longman Group UK limited, pp. 173-180.

Gebrekidan B. 1973. The importance of the Ethiopian sorghum germplasm in the world sorghum collection. Economic Botany 23:442-445.

Geleta N. 1997. Variability and association of morpho-agronomic characters with reference to highland sorghum [*Sorghum bicolour* (L.) Moench] landraces of Hararghe, eastern Ethiopia. M.Sc. Dissertation, Alemaya University of Agriculture, Ethiopia.

Geleta N, Labuschagne MT. 2005. Qualitative traits variation in sorghum [*Sorghum bicolour* (L.) Moench] germplasm from eastern highland of Ethiopia. Biodiversity and Conservation 14:3055-3064.

Harlan JR. 1992. Crops and man, 2nd edition. American Society of Agronomy. Crop Science Society of American. Madison, WI., USA, pp. 284.

Harlan JR, De Wet JMJ. 1972. A simplified classification of cultivated sorghum. Crop Science 12:172-177.

Hulse JH, Laing EM, Pearson OE. 1980**.** Sorghum and the millets: their composition and nutritive value. Academic Press, London, UK, pp. 81-91.

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.

Kahler AL, Allard RW. 1981. Worldwide patterns of genetic variation among four esterase loci in barley (*Hordeum vulgare* L.) Theoretical and Applied Genetics 59:101-111.

Karp A. 2002. The new genetic era: will it help us in managing genetic diversity? In: Managing plant genetic diversity. Edited by Engels JMM, Ramanatha Rao V, Brown AHD, Jackson MT. Wallingford and Rome, CAB International and IPGRI, pp. 43-56.

Karp A, Kresovich S, Bhat KV, Ayad WG, Hodgkin T. 1997. Molecular tools in plant genetic resources conservation: a guide to the technologies. IPGRI Technical Bulletin No. 2. International Plant Genetic Resources Institute, Rome, Italy.

Karp A, Seberg O, Buiatti M. 1996. Molecular techniques in the assessment of botanical diversity. Annual Botany 78:143-149.

Kebede Y. 1991. The role of Ethiopian sorghum germplasm resources in the national breeding programme. In: Plant Genetic Resources of Ethiopia. Edited by Engels JM, Hawkes JG, Worede M. Cambridge University Press. Cambridge, pp. 315-322.

Marshall DR, Brown AHD. 1975. Optimum sampling strategies in genetic conservation. In: Crop genetic resources for today and tomorrow. Edited by Frankel OH, Hawkes JG. Cambridge University Press, London, pp. 53-80.

Negassa M. 1985. Patterns of phenotypic diversity in an Ethiopian barley collection and the Arsi-Bale highland as a centre of origin of barley. Hereditas 102:139-150.

Negassa M. 1986. Estimates of phenotypic diversity and breeding potential of Ethiopian wheats. Hereditas 104:41-48.

Pecetti L, Damania AB. 1996. Geographic variation in tetraploid wheat (*Triticum turgidum* ssp. *turgidum durum*) landraces from two provinces in Ethiopia. Genetic Resources and Crop Evolution 43:395-407.

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

Perry MC, McIntosh MS. 1991. Geographical patterns of variation in the USDA soybean germplasm collection: I. Morphological traits. Crop Science 31:1350-1355.

Prasada Rao KE, Mengesh MH. 1981. Apointed collection of Zera-zera sorghum in the Gambella area of Ethiopia. Genetic Resources Unit Progress Report No. 33. ICRISAT, Patancheru P.O., A.P., India.

Prasada Rao KE, Ramanatha Rao V. 1995. The use of characterisation data in developing a core collection of sorghum. In: Core collections of plant genetic resources. Edited by Hodgkin T. John Wiley and Sons, Chichester, UK, pp. 109-116.

R Development Core Team. 2008. R: A language and environment for statistical computing. R. Foundation for Statistical Computing, Vienna, Austria (http://www.R-project. org).

Saghai-Maroof MA, Allard RW, Qhang Q. 1990. Genetic diversity and ecogeographical differentiation among ribosomal DNA alleles in wild and cultivated barley. Proceedings of the National Academy of Science USA 87:8486-8490.

Smith JSC, Smith OS. 1992. Fingerprinting crop varieties. Advances in Agronomy 47:85-141.

Stemler ABL, Harlan JR, De Wet JMJ. 1975. Evolutionary history of cultivated sorghums (*Sorghum bicolor* (L.) Moench) of Ethiopia. Bulletin of the Torrey Botanical Club 102:322-333.

Stemler ABL, Harlan JR, de Wet JMJ. 1977. The sorghums of Ethiopia. Economic Botany 31:446-460.

Tesfaye T, Getachew B, Worede M. 1991. Morphological diversity in tetraploid wheat landrace populations from central highlands of Ethiopia. Hereditas 114:171-176.

Teshome A, Baum BR, Fahrig L, Torrance JK, Arnason TJ, Lambert JD. 1997. Sorghum [*Sorghum bicolor* (L.) Moench] landrace variation and classification in North Shewa and South Welo, Ethiopia. Euphytica 97:255-263.

Tolbert DM, Qualset CO, Jain SK, Craddock JC. 1979. A diversity analysis of a world collection of barley. Crop Science 19:789-794.

VSN International Ltd. 2007. Genstat Discovery Edition 3. VSN International Ltd. Hernel Hempstead, UK.

Chapter 4

Genetic diversity analysis in sorghum based on microsatellite (SSR) analysis

4.1 Abstract

Analysis of genetic diversity in crops has a strong impact on plant breeding and maintenance of genetic resources. For the most part, it is valuable for characterization of individuals, accessions, and cultivars in identifying duplications in germplasm collections and for selecting parents. Sorghum is one of the most important cereal crops providing food in east Africa, which is believed to be its centre of origin and diversity. Genetic diversity among sorghum accessions from Sudan, Kenya, Uganda, Ethiopia, Eritrea, Rwanda and Burundi was assessed using 39 SSR markers. The objectives of this study were to (i) assess the extent and pattern of genetic diversity among sorghum accessions from the selected countries using SSR markers and (ii) study the genetic structure and how diversity is distributed within and between countries. All 39 SSR loci were polymorphic and revealed a total of 941 alleles in 1108 sorghum genotypes. Sudan had the highest genetic diversity followed by Ethiopia. AMOVA showed all variance components to be highly significant and the bulk of variation to be partitioned within countries (68.1%) compared to among countries (31.9%). Genetic differentiation between countries based on F_{ST} was high and highly significant (F_{ST}=0.32). Neighbour-joining (NJ) analysis formed two distinct clusters according to geographic regions namely the central region (Kenya, Burundi, Uganda, and Rwanda) and the eastern region (Sudan, Ethiopia, and Eritrea). Population structure analysis revealed six distinct populations corresponding to NJ analysis and geographical origin of accessions. This study addressed the pattern of genetic diversity of sorghum landraces at a large scale. Results suggested that despite sorghum's predominantly autogamous mating system, countries clustered totally separately with no integration, thus underlining the role of farmers' practices in the maintenance of landrace identity and genetic diversity. High levels of diversity were shown in central and eastern Africa.

4.2 Introduction

Sorghum is one of the world's leading cereal crops, providing food, feed, fiber, fuel, and chemicals/biofuels across a range of environments and production systems. Based on morphology, it is divided into five races (*bicolor*, *caudatum*, *guinea*, *durra*, and *kafir*),

along with the ten intermediate races resulting from all possible inter-race crosses. (Harlan and de Wet, 1972).

Genetic diversity and characterization of cultivars integrated in world collections are essential for classification, management, utilization, and ultimately for further efficient collection of exotic germplasm. Morphological characterization was the first method used by researchers to select superior genotypes in spring wheat (Briggs, 1991), and sorghum (Beta and Corke, 2001). However, several studies demonstrated that morphological markers are not appropriate for traits with low heritability and that they are highly affected by environments (Smith and Smith, 1992; Redfearn *et al*., 1999; Cadee, 2000). Genetic variation in sorghum has been evaluated in a number of studies using these morphological markers (Appa-Rao *et al*., 1996; Dje *et al*., 1998; Dahlberg, 2000; Geleta and Labuschagne, 2005; Barnaud *et al*., 2007). Although this approach has been effective, it is time consuming and is based on only a few traits. These limitations have lead to the development of molecular markers.

Molecular marker technology also contributes towards studying genetic diversity. It has numerous advantages including the availability of large numbers of markers, coverage of the entire genome, and their expression is not affected by the environment (Gepts, 1993). Genetic diversity in sorghum has been estimated using several types of molecular markers viz., allozymes (Aldrich *et al*., 1992; Dje *et al*., 1998), RFLP (Aldrich and Doebley, 1992; Tao *et al*., 1993; Cui *et al*., 1994; Deu *et al*., 2006), RAPD (Vierling *et al*., 1994; Ayana *et al*., 2000; Agrama and Tuinstra, 2003; Nkongolo and Nsapato, 2003; Uptmoor *et al*., 2003), AFLP (Uptmoor *et al*., 2003; Menz *et al*., 2004), and SSR (Brown *et al*., 1996; Taramino *et al*., 1997; Dje *et al*., 1998; Smith *et al*., 2000; Uptmoor *et al*., 2003; Menz *et al*., 2004; Casa *et al*., 2005). In each of these studies, authors studied a precise subset of sorghum germplasm.

Microsatellites or SSRs are relatively easy to perform, can be automated (Kresovich *et al*., 1995; Mitchell *et al*., 1997), and show Mendelian inheritance (Saghai-Maroof *et al*., 1994). SSRs are highly informative and are PCR-based. In addition, SSRs are highly polymorphic (Weber, 1990; Doldi *et al*., 1997; Schug *et al*., 1998). SSR markers are particularly attractive for studying genetic differentiation because they are co-dominant (Akkaya *et al*., 1992) and abundant in the genome (Lagercrantz *et al*., 1993). However, it

needs characterization of primers in every species of interest, which is expensive and time consuming (Kubik *et al*., 2001).

The potential of semi-automated, robust, cost-effective molecular genetic markers, specifically SSRs, allows the evaluation of germplasm collections for enhanced breeding (Morell *et al*., 1995; Mitchell *et al*., 1997). A previous study based on qualitative and quantitative characters identified significant variability amongst sorghum accessions sampled from the same region as the present study (Geleta, 1997). However, there exists no molecular data for comparison of sorghum accessions from different countries. The objectives of the study were to (i) assess the extent and pattern of genetic diversity among sorghum accessions from selected countries using SSR markers and (ii) study the genetic structure and how diversity is distributed within and between countries.

4.3 Materials and methods

4.3.1 Plant material

A total of 1108 accessions from Sudan, Ethiopia, Eritrea, Kenya, Uganda, Rwanda, and Burundi were selected as the core samples (Table 4.1). Samples from Tanzania were excluded from analysis because a different genotyping platform was used, that complicated comparison of results with that of other countries.

Table 4.1 Number of accessions genotyped per country

Country	Number of accessions
Sudan	208
Ethiopia	189
Eritrea	140
Kenya	189
Uganda	118
Rwanda	99
Burundi	165
Total	1108

4.3.2 DNA extraction

Twenty five seeds from each of the collected panicles of sorghum populations were randomly selected and planted in potted plastic trays at room temperature in the laboratory. To ensure sufficient lighting throughout the germination and growth period, trays were placed on a bench next to a glass widow.

DNA isolation from sorghum accessions was carried out according to the modified protocol of Mace *et al*. (2004), with a GenoGrinder (Geno/Grinder 2000, Spex Certiprep USA) using freshly harvested leaves of two week old seedlings from each sample that were placed into the wells of a 96-well format Geno/Grinder plate together with two 4 mm stainless steel grinding balls, on liquid nitrogen. Four hundred and fifty µl pre-heated (65°C) hexadecyltrimethyl ammoniumbromide (CTAB) extraction buffer [3% (w/v) CTAB, 1.4 M NaCl, 20 mM (hydroxymethyl) aminomethane hydrochloride (Tris-HCl), 20 mM ethylenediaminetetraacetic acid (EDTA), pH 8.0, 0.17% (v/v) β-mercaptoethanol] was added to each of the samples. The sample leaf tissues were homogenized in the Geno/Grinder at 500 strokes per min for 10 min. The plate was transferred and incubated in a water bath for 10 min at 65°C with occasional mixing.

Four hundred and fifty µl of chloroform-isoamylalcohol (24:1) was added and gently inverted twice and centrifuged at 10000 rpm for 10 min. A fixed volume of 400 µl of the upper aqueous layer was transferred to another labelled tube. To precipitate the DNA pellet, 280 µl of ice-cold isopropanol was added and mixed gently. The mix was kept at -20°C for one hour and then centrifuged for 10 min at 10000 rpm. The supernatant was removed and the DNA pellet air-dried for 30 min. Then 500 µl of 70% (v/v) ethanol was added for washing, followed by a brief centrifugation of 3 min at 10000 rpm. DNA was air-dried and later re-suspended in 100 µl of low-salt TE [10 mM Tris-HCl, 0.1 mM EDTA (pH 8.0)] and 1 µl RNAse (10 mg/ml) was added before incubating in an oven 37°C for 1 h.

To check the DNA quality, 3 µl of each sample was loaded on ethidium stained 0.8% (w/v) agarose gel and electrophoresed at 100 V for 1 h and visualized under UV light. Samples were normalized to a final concentration of 10 ng/µl by adding variable volumes of distilled water to a final volume of 100 µl for PCR amplification.

4.3.3 SSR amplification

PCR reactions were performed using a Mastercycler (Eppendorf) in a total volume of 10 µl, containing 5 µl of template DNA, 0.2 U of Amplitaq Gold *Taq* DNA polymerase, 1x PCR buffer (10 mM Tris-HCl pH 8.3, 50 mM KCl, 1.5 mM $MgSO_4$, 1.5 mM $MgCl_2$), 0.16 mM of each 2'-deoxynucleotide 5'-triphosphate (dNTP), 2 µM sequence-specific reverse primer, 0.04 µM 5'-M13 tailed sequence-specific forward primer, 0.16 µM 5'-

flourescently labelled M13 universal sequence primer. Subsets of 39 SSR markers were used in the study because of their high polymorphic content and broad coverage of the sorghum genome (Table 4.2).

PCR conditions as described by Folkertsma *et al.* (2005) were used: initial denaturation at 94°C for 15 min followed by 10 cycles of 94°C for 15 s, annealing for 20 s using a touchdown strategy where temperatures declined from 61°C to 50°C, and extension at 72°C for 30 s. This was followed by 35 cycles of 94°C for 10 s, 54°C for 20 s, 72°C for 30 s and a final extension step of 20 min at 72°C. The consistency of amplification was established for each primer set by subjecting 3 µl of the PCR products to 2% (w/v) agarose gel electrophoresis at 100 V for 45 min. Then 2 µl of the PCR products was denatured at 94^{o}C for 5 min in 8 µl of Hi-Di formamide with 0.13 µl of GeneScan 500 LIZ internal size standard (Applied Biosystems). The PCR product was subjected to capillary electrophoresis for allele detection. Allelic data was scored using the software Genemapper 3.7 (Applied Biosystems).

4.4 Data analysis

4.4.1 Diversity analyses

Total number of alleles (A^t), number of rare alleles (A^r, alleles with a frequency of <5% in a group), observed heterozygosity (H_o), unbiased expected heterozygosity or gene diversity (H_e) and polymorphic information content (PIC) were computed using the software PowerMarker V3.0 (Liu and Muse, 2004) to evaluate levels of genetic diversity. In view of the fact that the observed number of alleles in a sample is highly dependent on the sample size, the software HP RARE (Kalinowski, 2005) was additionally used to calculate allelic richness (A_s) and private alleles (A^p, alleles unique to a group), using the rarefaction technique recommended by Petit *et al*. (1998). Using Wilcoxon's matched-pairs signed-rank test as implemented in the software GenStat (VSN International Ltd., 2007), significant differences between defined groups across all loci, were assessed.

4.4.2 Analysis of population structure

In order to assess the structure of genetic diversity within and among countries, four complementary approaches were used: *F*-statistics, neighbour-joining analysis, AMOVA, and a Bayesian model-based clustering method.

Table 4.2 List of microsatellite primers used in this study

Marker	Chromosome	Repeat motif	Forward primer	Reverse primer	Tm (ºC)†	Min allele (bp)	Max allele (bp)
Gpsb067	8	$(GT)_{10}$	TAGTCCATACACCTTTCA	TCTCTCACACACATTCTTC	49	160	190
Gpsb123	8	$(CA)_7$+$(GA)_5$	ATAGATGTTGACGAAGCA	GTGGTATGGGACTGGA	50	284	304
MSbCIR246	5	$(CA)_7$	TTTTGTTGCACTTTTGAGC	GATGATAGCGACCACAAATC	55	86	114
MSbCIR262	7	$(CATG)_3$	GCACCAAAATCAGCGTCT	CCATTTACCCGTGGATTAGT	57	208	446
MSbCIR300	5	$(GT)_9$	TTGAGAGCGGCGAGGTAA	AAAAGCCCAAGTCTCAGTGCTA	61	74	118
MSbCIR329	10	$(AC)_8$	GCAGAACATCACTCAAAGAA	TACCTAAGGCAGGGATTG	54	73	121
Sb5-206=XGap206	6	$(AC)_{13}$/$(AG)_{20}$	ATTCATCATCCTCATCCTCGTAGAA	AAAAACCAACCCGACCCACTC	55	86	164
Sb6-84=XGap84	2	$(AG)_{14}$	CGCTCTCGGGATGAATGA	TAACGGACCACTAACAAATGATT	55	171	235
SbAGB02	5	$(AG)_{35}$	CTCTGATATGTCGTTGTGCT	ATAGAGAGGATAGCTTATAGCTCA	55	92	176
Xcup02	6	$(GCA)_6$	GACGCAGCTTTGCTCCTATC	GTCCAACCAACCCACGTATC	54	186	216
Xcup14	3	$(AG)_{10}$	TACATCACAGCAGGGACAGG	CTGGAAAGCCGAGCAGTATG	54	209	251
Xcup53	1	$(TTTA)_5$	GCAGGAGTATAGGCAGAGGC	CGACATGACAAGCTCAAACG	54	182	202
Xcup61	3	$(CAG)_7$	TTAGCATGTCCACCACAACC	AAAGCAACTCGTCTGATCCC	54	189	204
Xcup63	2	$(GGATGC)_4$	GTAAAGGGCAAGGCAACAAG	GCCCTACAAAATCTGCAAGC	54	127	163
XtXp010	6	$(CT)_{14}$	ATACTATCAAGAGGGGAGC	AGTACTAGCCACACGTCAC	50	119	155
XtXp015	10	$(TC)_{16}$	CACAAACACTAGTGCCTTATC	CATAGACACCTAGGCCATC	55	197	273
XtXp040	5	$(GGA)_7$	CAGCAACTTGCACTTGTC	GGGAGCAATTTGGCACTAG	55	108	144
XtXp057	9	$(GT)_{21}$	GGAACTTTTGACGGGTAGTGC	CGATCGTGATGTCCCAATC	55	213	285
XtXp145	9	$(AG)_{22}$	GTTCCTCCTGCCATTACT	CTTCCGCACATCCAC	55	204	278
Xisep0310	2	$(CCAAT)_4$	TGCCTTGTGCCTTGTTTATCT	GGATCGATGCCTATCTCGTC	60	159	219
MSbCIR223	2	$(AC)_6$	CGTTCCAATGACTTTTCTTC	GCCAATGTGGTGTGATAAAT	55	104	124
MSbCIR238	2	$(AC)_{26}$	AGAAGAAAAGGGGTAAGAGC	CGAGAAACAATTACATGAACC	55	69	129
MSbCIR240	8	$(TG)_9$	GTTCTTGGCCCTACTGAAT	TCACCTGTAACCCTGTCTTC	55	104	180
MSbCIR248	10	$(GT)_7$	GTTGGTCAGTGGTGGATAAA	ACTCCCATGTGCTGAATCT	56	79	111
MSbCIR276	3	$(AC)_9$	CCCCAATCTAACTATTTGGT	GAGGCTGAGATGCTCTGT	53	222	252
MSbCIR283	7	$(CT)_8$ $(GT)_8$	TCCCTTCTGAGCTTGTAAAT	CAAGTCACTACCAAATGCAC	54	111	157
MSbCIR286	1	$(AC)_9$	GCTTCTATACTCCCCTCCAC	TTTATGGTAGGATGCTCTGC	55	110	150
MSbCIR306	1	$(GT)_7$	ATACTCTCGTACTCGGCTCA	GCCACTCTTTACTTTTCTTCTG	55	118	126

†Tm = annealing temperature

Table 4.2 Continued

Marker	Chromosome	Repeat motif	Forward primer	Reverse primer	Tm (ºC)[†]	Min allele (bp)	Max allele (bp)
Sb4-72=XGap72	9	$(AG)_{16}$	TGCCACCACTCTGGAAAAGGCTA	CTGAGGACTGCCCCAAATGTAGG	55	169	229
XtXp012	4	$(CT)_{22}$	AGATCTGGCGGCAACG	AGTCACCCATCGATCATC	55	143	215
XtXp021	4	$(AG)_{18}$	GAGCTGCCATAGATTTGGTCG	ACCTCGTCCCACCTTTGTTG	60	151	227
XtXp114	3	$(AGG)_{8}$	CGTCTTCTACCGCGTCCT	CATAATCCCACTCAACAATCC	50	196	226
XtXp136 (Kaf3)	10	$(GCA)_{5}$	GCGAATAGCATCTTACAACA	ACTGATCATTGGCAGGAC	55	240	246
XtXp141	7	$(GA)_{23}$	TGTATGGCCTAGCTTATCT	CAACAAGCCAACCTAAA	55	133	175
XtXp265	9	$(GAA)_{19}$	GTCTACAGGCGTGCAAATAAAA	TTACCATGCTACCCCTAAAAGTGG	55	168	246
XtXp273 (Pbbf)	8	$(TTG)_{20}$	GTACCCATTTAAATTGTTTGCAGTAG	CAGAGGAGGAGGAAGAGAAGG	55	148	217
XtXp278	5	$(TTG)_{12}$	GGGTTTCAACTCTAGCCTACCGAACTTCCT	ATGCCTCATCATGGTTCGTTTTGCTT	50	225	318
XtXp320 (PhyB)	1	$(AAG)_{20}$	TAAACTAGACCATATACTGCCATGATAA	GTGCAAATAAGGGCTAGAGTGTT	54	251	329
XtXp321	8	$(GT)_{4}+(AT)_{6}+(CT)_{21}$	TAACCCAAGCCTGAGCATAAGA	CCCATTCACACATGAGACGAG	55	180	252

[†]Tm = annealing temperature

Fixation indexes can be used to study population structure, to assign the degree of differentiation within a population among groups of inhabitants (F_{SG}), within groups among inhabitants (F_{GT}), and within a population among inhabitants (F_{ST}) (Hartl and Clark, 1997). It ranges from 0 (indicating no differentiation between the overall population and its subpopulations) to a maximum of 1. However, in practice the observed fixation index is much lower than 1, even in extremely differentiated populations.

Considering the seven countries for which enough individuals were sampled, F_{ST} values (Weir and Cockerham, 1984) were computed as a measure of the genetic diversity within and among countries. Distance method pairwise differentiation was used to estimate the pairwise genetic differentiation between countries. Permutation procedures (1000 permutations) were performed to test the significance of differences between values. Calculations were carried out using ARLEQUIN 3.0 (Excoffier *et al*., 2005).

To investigate the genetic relationships among and between countries, a neighbour-joining (NJ) cluster analysis algorithm implemented in the software DARwin 5.0.155 (Perrier *et al.*, 2003) was used. Dissimilarities between all pairs of individual genotypes were estimated based on a simple matching procedure. The genetic structure of sorghum accessions was additionally investigated using AMOVA using ARLEQUIN 3.0 (Excoffier *et al*., 2005). The significance of the partitioning of genetic variance among groups was tested.

To assign sorghum accessions to populations based on their genotypes, the data set was subjected to the Bayesian model-based clustering method implemented in the software STRUCTURE 2.2 (Pritchard *et al.*, 2000) using the admixture model. The Bayesian based model assumes that each individual inherited some portion of its ancestry from one of the *K* populations. The method of Evanno *et al.* (2005) was used to determine the true number of *K* populations in the dataset. All STRUCTURE analyses were performed using the high performance computing resources of the computational biology service unit (CBSU) from Cornell University (http://cbsuapps.tc.cornell.edu/structure.aspx). With the assumed number of populations (*K*) varying from 1 to 10, 20 replicate runs per *K* value, with a burn in of 50000 Markov Chain Monte Carlo (MCMC) iterations followed by 10^6 iterations of data collection. Evanno *et al*. (2005) reported that the modal value of the

distribution of Δ*K* is located at the real *K*. The modal value was illustrated graphically by plotting the Δ*K* values against successive *K* values. After the identification of the 'true K', the run showing the highest P(X|K) value was considered in drawing a barplot of the proportion of an individual's genome assigned to each of the clusters using the software STRUCTURE 2.2 (Pritchard *et al*., 2000).

4.5 Results

4.5.1 Polymorphic level of tested microsatellites in sorghum accessions

All 39 selected SSR loci were polymorphic and revealed a total of 941 alleles in the 1108 sorghum genotypes. The loci, number of alleles, their major allele frequency, gene diversity and PIC values are given in Table 4.3. The number of alleles per locus ranged from eight (MSbCIR276, MSbCIR329, and Xcup61) to 56 (XGap206) with a mean value of 24.1. PIC varied from 0.27 (Xcup61) to 0.94 (XGap206) with an average of 0.74. XGap206 had the highest gene diversity (0.94) whereas Xcup61 had the lowest (0.31). The highest major allele frequency was obtained in Xcup61 (0.81) while the lowest value was found for MSbCIR238 (0.13).

4.5.2 Extent of genetic diversity in sorghum

Estimates of genetic diversity parameters for sorghum gene pools using various diversity parameters are shown in Table 4.4. Almost 59% of the detected alleles were determined to be rare (present in less than 5% of the genotypes). The number of total alleles per country ranged from 163 (Rwanda) to 448 (Sudan). In terms of rare alleles, Sudan had the highest percentage (63.6%) whereas Eritrea had the lowest (52.5%). Observed heterozygosity varied from 0.114 (Uganda) to 0.371 (Eritrea) with a mean value of 0.232. Sudan had the highest gene diversity (0.69) followed by Ethiopia (0.65). Rwanda had the lowest gene diversity (0.33). The same trend could be observed for allelic richness. Private allele richness ranged from 0.44 to 1.88 and was highest in Ethiopia. According to Wilcoxon's matched-pairs signed-rank test, private allele richness, gene diversity, and allelic richness were tested for significance between countries (Appendices 2, 3, and 4 respectively).

Table 4.3 Polymorphic parameters of microsatellites used in the study

Marker	Major allele frequency	No of alleles	Gene diversity	PIC
Gpsb067	0.4551	36	0.7535	0.7362
Gpsb123	0.4031	12	0.7154	0.6718
MSbCIR246	0.4121	11	0.6660	0.6077
MSbCIR262	0.5305	11	0.6578	0.6214
MSbCIR300	0.3805	13	0.7878	0.7648
MSbCIR329	0.5232	8	0.6584	0.6188
XGap206	0.1768	56	0.9423	0.9401
XGap84	0.2295	41	0.9094	0.9042
SbAGB02	0.2606	51	0.8889	0.8816
Xcup02	0.3360	17	0.7990	0.7744
Xcup14	0.3153	12	0.7754	0.7417
Xcup53	0.5363	13	0.6681	0.6413
Xcup61	0.8119	8	0.3084	0.2655
Xcup63	0.6450	9	0.5363	0.4956
XtXp010	0.2039	21	0.8878	0.8781
XtXp015	0.2808	19	0.8376	0.8197
XtXp040	0.4323	16	0.7390	0.7077
XtXp057	0.2034	36	0.9017	0.8945
XtXp145	0.5094	55	0.7303	0.7248
Xisep0310	0.6996	9	0.4790	0.4477
MSbCIR223	0.3212	16	0.8011	0.7754
MSbCIR238	0.1324	45	0.9353	0.9318
MSbCIR240	0.3433	28	0.7204	0.6688
MSbCIR248	0.5453	13	0.6664	0.6439
MSbCIR276	0.3526	8	0.7894	0.7624
MSbCIR283	0.2287	39	0.8632	0.8500
MSbCIR286	0.6075	11	0.5906	0.5603
MSbCIR306	0.3933	9	0.7687	0.7409
XGap72	0.2368	25	0.8592	0.8443
XtXp012	0.1435	45	0.9312	0.9272
XtXp021	0.3526	29	0.8310	0.8184
XtXp114	0.3300	18	0.7458	0.7007
XtXp136	0.5204	15	0.6387	0.5877
XtXp141	0.2293	36	0.9067	0.9010
XtXp265	0.2218	46	0.9145	0.9098
XtXp273	0.4601	23	0.7552	0.7403
XtXp278	0.6975	13	0.4940	0.4748
XtXp320	0.1628	31	0.9086	0.9018
XtXp321	0.1584	37	0.9299	0.9260
Mean	0.3790	24.1282	0.7613	0.7385

PIC = Polymorphic information content.

Table 4.4 Genetic diversity parameters

Country	No of individuals	Total no of alleles	Allelic richness	Rare alleles %	Private allele richness	Gene diversity	Observed heterozygosity
Sudan	208	448	8.04	63.6	1.80	0.685	0.142
Ethiopia	189	362	6.81	56.6	1.88	0.653	0.321
Kenya	189	359	6.57	62.1	0.95	0.569	0.239
Eritrea	140	308	6.06	52.5	1.32	0.561	0.371
Uganda	118	308	6.07	59.4	1.77	0.537	0.114
Burundi	165	291	5.42	60.5	0.44	0.466	0.178
Rwanda	99	163	3.35	56.6	0.53	0.330	0.260
Total	1108	941					
Mean	158	320	6.05	58.8	1.24	0.543	0.232

4.5.3 Genetic structure of sorghum accessions

4.5.3.1 Analysis of molecular variance

The outcome of the partitioning of genetic diversity within and among countries using AMOVA is presented in Table 4.5. The analysis indicated that all variance components were highly significant and that the bulk of variation was partitioned within countries (68.1%) compared to among countries (32%).

Table 4.5 Analysis of molecular variance among and within populations

Source of variation	Sum of squares	Variance components	Percentage variation
Among countries	6675.372	3.524	31.95***
Within countries	16577.241	7.504	68.05***
Total	23252.612	11.028	

All sources of variation were significant at $p < 0.001$.

4.5.3.2 F_{ST} based genetic variation

Genetic differentiation between countries based on F_{ST} was high and highly significant (F_{ST}=0.32; $p<0.001$) which supported the result obtained using AMOVA. Results from distance method pairwise differentiation between countries are shown in Table 4.6. The highest level of differentiation was between Rwanda and Eritrea (0.474) and the lowest between Sudan and Kenya (0.192).

Table 4.6 Estimates of pairwise genetic differentiation (F_{ST}) between countries

Country	Burundi	Eritrea	Ethiopia	Kenya	Rwanda	Sudan
Burundi						
Eritrea	0.35087					
Ethiopia	0.37483	0.32956				
Kenya	0.23981	0.34117	0.33435			
Rwanda	0.30309	0.47377	0.45944	0.39771		
Sudan	0.24307	0.24835	0.25050	0.19234	0.34887	
Uganda	0.31395	0.39579	0.33720	0.31389	0.40197	0.28785

4.5.4 Genetic variation within and between countries

To obtain a graphical demonstration of the relationships between individual sorghum accessions, a PCoA was performed based on the dissimilarity matrix. The two main eigen values explained 9.8% and 5.7% of the total variance, respectively. The PCoA presented in Figure 4.1 revealed two distinct clusters according to the centre of origin of accessions along axis 1, namely the central region (Kenya, Burundi, Uganda, and Rwanda) to the left of the axis, and the eastern region (Sudan, Ethiopia, and Eritrea) to the right. Accessions from most of the countries tended to cluster together and seven distinct clusters could be detected. All accessions from Rwanda clustered closely together with only a few clustering with accessions from Burundi. Accessions from Eritrea also formed a distinct group, although some accessions clustered somewhat away from the rest of the accessions. Accessions from Ethiopia also tended to cluster together and separately from accessions from the other six countries. Although accessions from Uganda, Kenya, and Sudan formed distinct clusters, some overlapping of accessions from these three countries was observed. Furthermore, although most accessions from Burundi clustered together, many accessions clustered within accessions from Uganda and to some extent, Kenya.

The genetic relationships within sorghum accessions as determined by NJ analysis are presented in Figure 4.2. Sorghum genotypes clustered into six main groups according to their geographical origin. Accessions from Kenya, Ethiopia, Eritrea, and Uganda each formed one main cluster, while two main groups were observed for Sudan and three for Burundi. Accessions from Rwanda grouped within accessions from Burundi. A few accessions from Burundi clustered with Ugandan accessions while another group clustered closely together with accessions from Kenya.

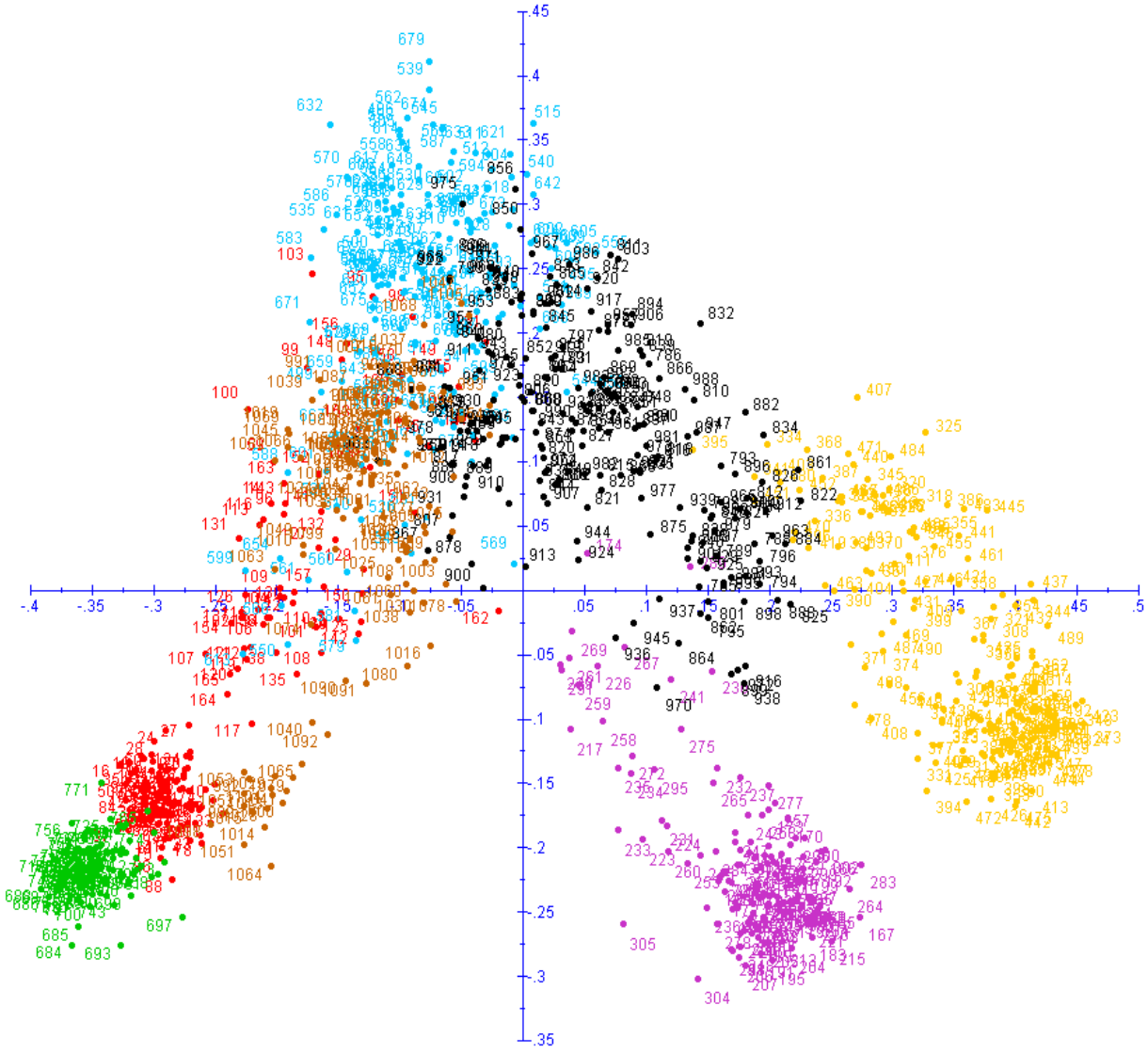

Figure 4.1 Biplot of the axis 1 and 2 of the principle coordinate analysis based on the dissimilarity of 39 SSR markers among 1108 sorghum accessions.

Eritrea, Ethiopia, Sudan, Kenya, Uganda, Burundi, and Rwanda.

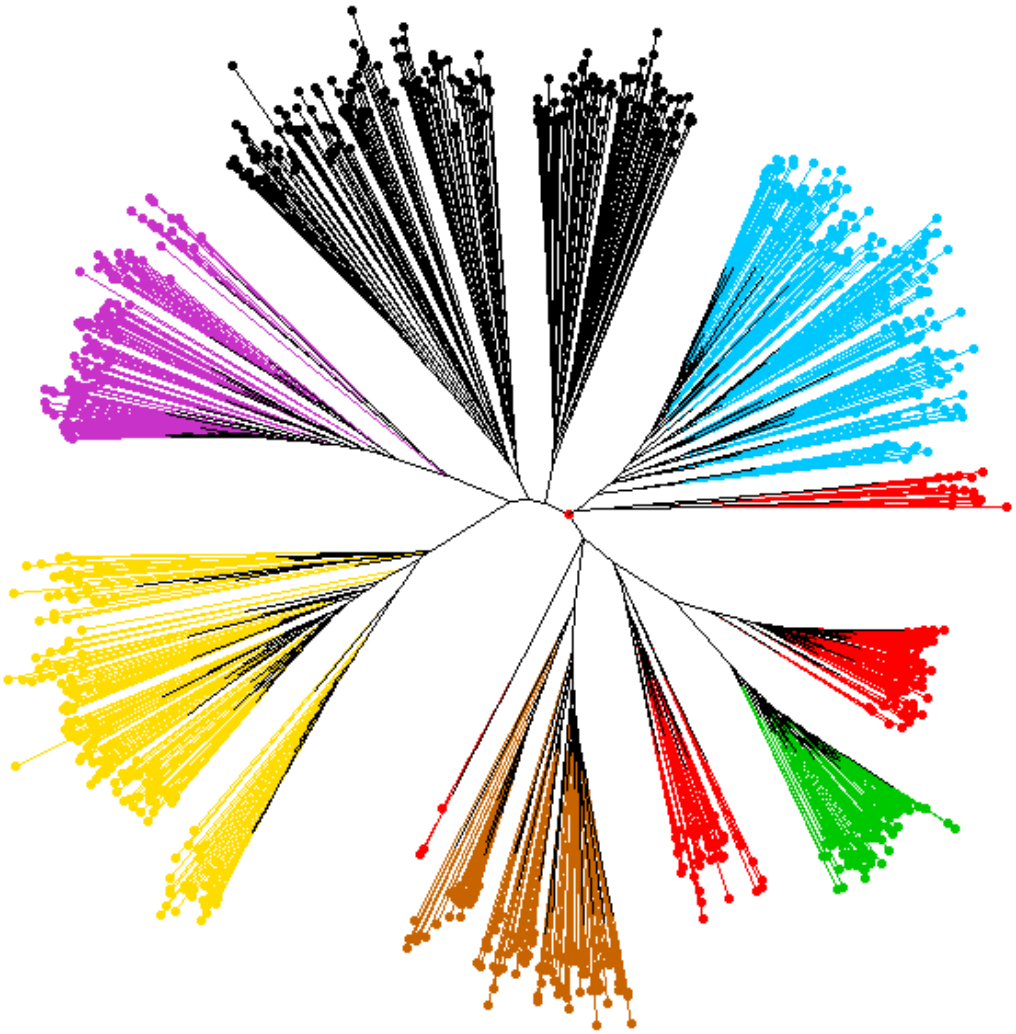

Figure 4.2 **Neighbour-joining cluster analysis dendrogram showing the genetic relationship among 1108 sorghum accessions using 39 SSR markers based on simple matching index.**

Eritrea, Ethiopia, Sudan, Kenya, Uganda, Burundi, and Rwanda.

4.5.5 Bayesian model-based cluster analysis

According to the method described by Evanno *et al.* (2005), the initial STRUCTURE analysis identified $K = 6$ to be the most appropriate number of populations (Figure 4.3).

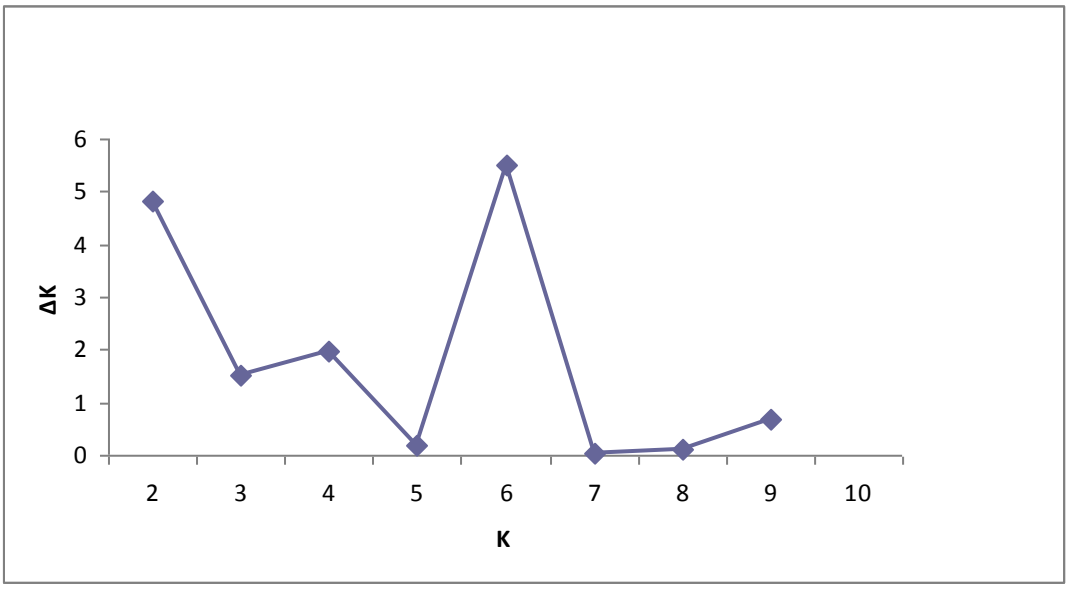

Figure 4.3 Evanno's Δ*K* statistic for *K*=1 to *K*=10. The modal value is at *K*=6.

The Bayesian model-based cluster analysis at $K=6$ was successful to identify distinct differentiation among sorghum accessions based on their country of origin, which confirmed results obtained from the PCoA and NJ analyses. Sorghum genotypes were grouped into six populations mainly according to their geographical origin (Figure 4.4). Accessions from Eritrea, Ethiopia, Kenya, Sudan, and Uganda formed part of one specific population each (green, blue, pink, yellow, and navy respectively). Accessions from Burundi and Rwanda belonged to the same population (red). Accessions from Burundi were assigned to two different populations (red and pink) that corresponded to individuals of Rwanda (red) and Kenya (pink).

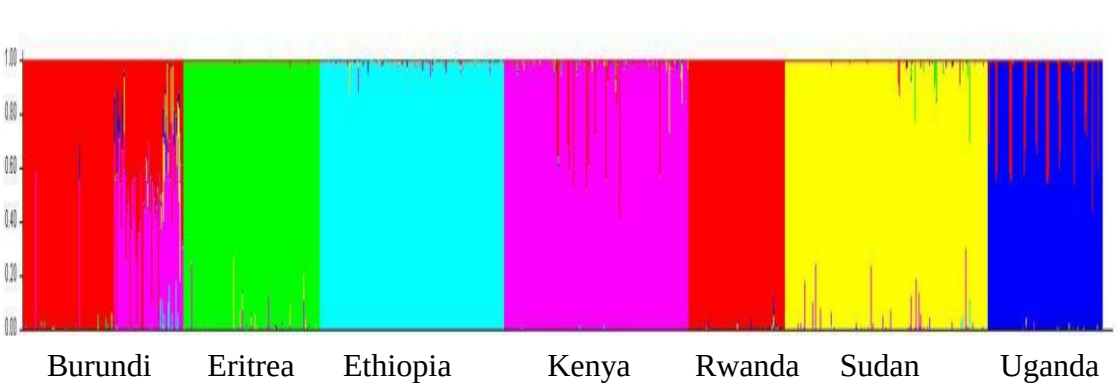

Figure 4.4 **Bar plot of the estimated genetic structure at *K*=6 using the default STRUCTURE parameters with the individuals ordered by country of origin. Each individual is represented by a vertical line which is partitioned into coloured segments that represent its proportion of genome in *K* (coloured) clusters.**

4.6 Discussion

The highest observed major allele frequency of an individual allele per locus was 0.81 that was higher than the value of 0.52 from a study done by Ghebru *et al*. (2002). With regard to the number of alleles (8-56) observed for most of the loci in this study, it was in agreement with the range reported for sorghum by Dje *et al*. (2000) of 14-24, and 2-23 for maize (Senior *et al*., 1998). However, it was higher than reported in other sorghum studies of 3-9 by Ghebru *et al*. (2002), 4-10 by Anas and Yoshida (2004) and 2-10 by Ali *et al*. (2008) and 2-8 for wheat (Ahmad, 2002). The higher values reported in the current study are due to the bigger sample size and wider geographic origin of accessions studied.

Smith *et al*. (2000) demonstrated that the PIC of an SSR marker provides an estimate of the discriminatory power of that SSR marker by taking into account not only the number of alleles that are detected but also the relative frequencies of those alleles. In the present study the mean PIC value for SSR markers was 0.74 while Smith *et al*. (2000) and Ali *et al*. (2008) reported PICs of 0.645 and 0.400 respectively in sorghum using SSR markers. The high PIC value detected in this study indicates that the SSR marker system has sufficient resolution to be used for genetic studies.

Most of the SSR markers used in this study revealed a high discriminatory power. The allelic diversity was likely due to the high levels of polymorphism of the markers. The

high levels of allelic variability but low levels of heterozygosity observed in this study corresponded with a previous SSR marker study of five Guinea-race accessions by Dje *et al*. (2000) and 100 Guinea-race accessions by Folkertsma *et al*. (2005) and fits with the predominantly inbreeding nature of sorghum. The high level of polymorphism of a number of markers of the SSR loci permitted the selection of a group of six markers, whose alleles in combinations offered exclusive genotyping for all 1108 accessions. These loci were XGap206, MSbCIR238, Xtxp012, Xtxp141, XGap84, and Xtxp265. If possible, combinations of two or more of these primer pairs in a single PCR reaction mix might facilitate fast fingerprinting of sorghum accessions.

This study furthermore revealed a high level of rare alleles (59%), which is in line with Folkertsma *et al*. (2005) who detected 50% rare alleles when they studied the pattern of genetic diversity in Guinea-race sorghum and Casa *et al*. (2005) who detected 64% rare alleles. The high value of rare alleles could be an indication of the relatively high mutation rate of SSR loci. It might also be due to the big sample size used as well as seven countries being sampled. Both the PCoA and NJ cluster analysis indicated that accessions from the different countries clustered separately, indicating the uniqueness of genotypes from each country, which can probably be explained by the high level of private alleles.

This study, using sorghum accessions from seven different countries, detected relatively low levels of genetic diversity [average gene diversity (GD) per country=0.543; average GD per locus=0.761] compared to values observed in other studies at regional scale. For example, Dje *et al*. (1999) reported a GD=0.83 for a sample from five regions in Morocco, and Uptmoor *et al*. (2003) reported a GD=0.59 for 23 landraces from southern Africa. Multiple origins for domesticated sorghum, cross-pollination between selected races, and outcrossing between domestic cultivars and highly variable wild species all are considered to be factors contributing to the extensive genetic diversity observed in sorghum (Doggett, 1988).

The current study detected high levels of genetic variability within countries (68.1%) using AMOVA analysis, suggesting that an increased sample sizes per country might even detect further variation. This might suggest that results from this study underestimated the genetic diversity present within the region. The F_{ST} value observed in

this study was relatively lower (0.32) compared to those observed in previous studies. Dje *et al*. (2000) reported a F_{ST} of 0.68 in a world collection (25 accessions) on the basis of only three different SSR loci, and Ghebru *et al.* (*2002)* reported a F_{ST} of 0.50 among 28 Eritrean sorghum. Dje *et al*. (1999) reported that low values of F_{ST} might be caused either by the occurrence of frequent gene flow among countries, for instance as a consequence of seed exchanges among farmers, or by a restriction of the intensity of genetic drift due to a high effective population size. The lower F_{ST} value of the current study could be explained by the fact that the study took place *in situ* at a large scale (1108 accessions), with the potential for high gene flow among accessions that are planted in close proximity in mixed fields. However, the significant differentiation detected in this study among countries suggested the existence of biological barriers to gene flow, including environmental pressures and biological traits of the plant, such as its mating system. Human factors furthermore affect the dynamics of diversity by influencing gene flow, drift and selection.

Genetic relationships observed among sorghum accessions using distance methods, pairwise differentiation, PCoA, NJ analysis, and Bayesian model-based cluster analysis confirmed the differentiation of accessions according to their country of origin. Even though there was a low value of differentiation (F_{ST}=0.32) among country groups, there was high enough levels of genetic variation among accessions to separate them (as was observed for F_{ST} pairwise distances). Clustering of accessions from Rwanda between two groups of accessions from Burundi supports the idea that the germplasm represents an independent event of domestication or that introgression with wild genotypes has occurred (Ellstrand, 2003). PCoA results also revealed the genetic similarity between pairs of accessions from Sudan with some of their cultivated relatives from Kenya. This could be caused either by the occurrence of frequent gene flow among these two countries, for instance as a consequence of seed exchanges among farmers, or by a restriction of the intensity of genetic drift due to a high effective population size.

Additionally, the PCoA analysis of the 1108 accessions in the present study produced significant groupings that support the earlier examination of east African (Sudan, Ethiopia, and Eritrea) and central African (Kenya, Uganda, Burundi, and Rwanda) collections, from earlier studies done by Harlan and De Wet (1972), Harlan *et al.* (1976), Stemler *et al*. (1977), Doggett (1988), and Doggett and Parasada Roa (1995). Moreover,

Ejeta *et al.* (1999) reported that inspection of the degree of relationship of accessions with their geographic areas of origin indicated that the total genetic variation was attributable to discrepancy among regions. In spite of the limited differentiation among regions in their study, the extent of genetic diversity within and among regions showed some trends. They found that the fact that southern African germplasm was represented by a large number of accessions, it exhibited the lowest level of genetic diversity, suggesting a narrow genetic base for accessions from this region. In contrast, West Africa exhibited a high level of genetic diversity, with the smallest number of accessions. Genetic diversity in central and eastern Africa, as well as for accessions from the Middle East was as high as that observed for accessions from West Africa. This supported the findings from this study that variation among countries was highly significant (32%).

NJ clustering and Bayesian model-based cluster analysis grouped accessions according to their country of origin. Several factors could have contributed to this detected pattern. The predominantly autogamous breeding system of sorghum can contribute towards explaining patterns of genetic diversity and structure observed. Secondly, environmental, biological, cultural, and socio-economic factors all play a role in farmer's decisions to choose or keep a particular sorghum cultivar at any given time. Farmers make decisions on how much of each accession to plant each year, the percentage of seed or germplasm to save from their own stock and the percentage to buy or exchange from other sources. Each of these decisions affects the genetic diversity of crop cultivars and is linked to a complex set of environmental and socio-economic influences (Hobbs and Huenneke, 1992).

4.7 Conclusions

In conclusion, this study addressed the pattern of genetic diversity of sorghum landraces at a large scale. High levels of diversity were detected in central and eastern Africa. It was suggested that despite sorghum's predominantly autogamous mating system, countries clustered totally separately with no integration, thus underlining the role of farmers' practices in the maintenance of landrace identity and genetic diversity. This data suggests that molecular markers are suitable to assess genetic diversity and to identify diverse sources in crop germplasm collections.

4.8 References

Agrama HA, Tuinstra MR. 2003. Phylogenetic diversity and relationships among sorghum lines using SSRs and RAPDs. African Journal of Biotechnology 2:334-340.

Ahmad M. 2002. Assessment of genomic diversity among wheat genotypes as determined by simple sequence repeats. Genome 45:646-651.

Akkaya MS, Bhagwat AA, Cregan PB. 1992. Length polymorphisms of simple sequence repeat DNA in soybean. Genetics 132:1131-1139.

Aldrich PR, Doebley J. 1992. Restriction fragment variation in the nuclear and chloroplast genomes of cultivated and wild *Sorghum bicolor*. Theoretical and Applied Genetics 85:293-302.

Aldrich PR, Doebley J, Schertz KF, Stec A. 1992. Patterns of allozyme variation in cultivated and wild *Sorghum bicolor*. Theoretical and Applied Genetics 85:451-460.

Ali ML, Rajewski JF, Baenziger PS, Gill KS, Eskridge KM, Dweikat I. 2008. Assessment of genetic diversity and relationship among a collection of US sweet sorghum germplasm by SSR markers. Molecular Breeding 21:497-509.

Anas YT and Yoshida T. 2004. Genetic diversity among Japanese cultivated sorghum assessed with simple sequence repeat markers. Plant Production Science 7:217-223.

Appa-Rao S, Prasada Rao KE, Mengesha MH, Gopal Reddy V. 1996. Morphological diversity in sorghum germplasm from India. Genetic Resources and Crop Evolution 43:559-567.

Ayana A, Bryngelsson T, Bekele E. 2000. Genetic variation of Ethiopian and Eritrean sorghum [*Sorghum bicolor* (L.) Moench] germplasm assessed by random amplified polymorphic DNA (RAPD). Genetic Resources and Crop Evolution 47:471-482.

Barnaud A, Deu M, Garine E, McKey D, Joly H. 2007. Local genetic diversity of sorghum in a village in northern Cameroon: Structure and dynamics of landraces. Theoretical and Applied Genetics 114:237-248.

Beta T, Corke H. 2001. Genetic and environmental variation in sorghum starch properties. Journal of Cereal Science 34:261-268.

Briggs KG. 1991. Spatial variation in seed size and seed set on spikes of some Canadian spring wheat cultivars. Canadian Journal of Plant Science 71:95-103.

Brown SM, Hopkins MS, Mitchell SE, Senior ML, Wang TY, Duncan RR, Gonzalez-Candelas F, Kresovich S. 1996. Multiple methods for the identification of polymorphic simple sequence repeats (SSRs) in sorghum [*Sorghum bicolor* (L.) Moench]. Theoretical and Applied Genetics 93:190-198.

Brown SM, Hopkins MS, Mitchell SE, Senior ML, Wang TY, Duncan RR, Gonzalez-Candelas F, Kresovich S. 1996. Multiple methods for the identification of polymorphic simple sequence repeats (SSRs) in sorghum [*Sorghum bicolor* (L.) Moench]. Theoretical and Applied Genetics 93:190-198.

Cadee N. 2000. Genetic and environmental effects on morphology and fluctuating asymmetry in nestling barn swallows. Journal of Evolutionary Biology 13:359-370.

Casa AM, Mitchell SE, Hamblin MT, Sun H, Bowers JE, Paterson AH, Aquadro CF, Kresovich S. 2005. Diversity and selection in sorghum: simultaneous analyses using simple sequence repeats. Theoretical and Applied Genetics 111:23-30.

Cui YX, Xu GW, Magill CW, Schertz KF. 1994. A low-copy number sorghum DNA sequence that detects hypervariable *Eco*RV fragments. Theoretical and Applied Genetics 89:64-69.

Dahlberg JA. 2000. Classification and characterization of Sorghum. In: Sorghum origin, history, technology and production. Edited by Smith CW, Frederiksen RA. Wiley and Sons, New York, pp. 99-130.

Deu M, Rattunde F, Chantereau J. 2006. A global view of genetic diversity in cultivated sorghum using a core collection. Genome 49:168-180.

Dje Y, Ater M, Lefebvre C, Vekemans X. 1998. Patterns of morphological and allozyme variation in sorghum landraces of northwestern Morocco. Genetic Resources and Crop Evolution 45:541-548.

Dje Y, Forcioli D, Ater M, Lefebvre C, Vekemans X. 1999. Assessing population genetic structure of sorghum landraces from north-western Morocco using allozyme and microsatellite markers. Theoretical and Applied Genetics 99:57-163.

Dje Y, Heuertz M, Lefébvre C, Vekemans X. 2000. Assessment of genetic diversity within and among germplasm accessions in cultivated sorghum using microsatellite markers. Theoretical and Applied Genetics 100:918-925.

Doggett H. 1988. Sorghum. Longman Scientific and Technical Essex, England, pp. 512.

Doggett H, Prasada Rao KE. 1995. Sorghum. In: Evolution of Crop Plants (2^{nd} edtion). Edited by Smartt J, Simmonds NW. Longman Group, Burnt Mill., pp. 173-180.

Doldi ML, Vollmann J, Lelly T. 1997. Genetic diversity in soybean as determined by RAPD and microsatellite analysis. Plant Breeding 116:331-335.

Ejeta G, Goldsbrough PB, Tuinstra MR, Grote EM, Menkir A, Ibrahim Y, Cisse N, Weerasuriya Y, Melakeberhan A, Shaner CA. 1999. Molecular marker applications in sorghum: Paper presented at Workshop on Application of Molecular Markers, Ibadan, Nigeria, 16-20 August.

Ellstrand NC. 2003. Dangerous liaisons - when cultivated plants mate with their wild relatives. Johns Hopkins University Press Baltimore MD, pp. 244.

Evanno S, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14:2611-2620.

Excoffier L, Laval LG, Schneider S. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1:47-50.

Folkertsma RT, Fredrick H, Rattunde W, Chandra S, Raju SG, Hash TC. 2005. The pattern of genetic diversity of *Guinea*-race *Sorghum bicolor* (L.) Moench landraces as revealed with SSR markers. Theoretical and Applied Genetics 111:399-409.

Geleta N. 1997. Variability and association of morpho-agronomic characters with reference to highland sorghum [*Sorghum bicolor* (L.) Moench] landraces of Hararghie, Eastern Ethiopia. MSc. Dissertation. Alemaya University of Agriculture, Alemaya, Ethiopia.

Geleta N, Labuschagne MT. 2005. Qualitative traits variation in sorghum (*Sorghum bicolor* (L.) Moench) germplasm from eastern highlands of Ethiopia. Biodiversity and Conservation 14:3055-3064.

Gepts P. 1993. The use of molecular and biochemical markers in crop-evaluation studies. In: Evolutionary biology. Edited by Hecht MK. vol. 27. Plenum press, New York, pp. 51-94.

Ghebru B, Schmidt RJ, Bennetzen JL. 2002. Genetic diversity of Eritrean sorghum landraces assessed with simple sequence repeat (SSR) markers. Theoretical and Applied Genetics 105:229-236.

Harlan JR, De Wet JMJ. 1972. A simplified classification of cultivated sorghum. Crop Science 12:172-177.

Harlan JR, De Wet JMJ, Stemler ABL. 1976. The origins of African plant domestication. Mouton Press, The Hague, Netherlands, pp. 498.

Hartl DL, Clark G. 1997. Principles of population genetics. Sinauer Associates, Inc. Sunderland, pp. 452.

Hobbs RJ, Huenneke LF. 1992. Disturbance, diversity, and invasion: implications for conservation. Conservation Biology 6:324-337.

Kalinowski S. 2005. HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. Molecular Ecology Notes 5:187-189.

Kresovich S, Szewe-Mcfadden AK, Bliek SM, McFerson JR. 1995. Abundance and characterization of simple-sequence repeats (SSRs) isolated from size-fractioned genomic library of *Brassica napus* L. (rapseed). Theoretical and Applied Genetics 19:206-211.

Kubik C, Sawkins M, Meyer WA, Gaut BS. 2001. Genetic diversity in seven perennial ryegrass (*Lolium perenne* L.) cultivars based on SSR markers. Crop Science 41: 1565-1572.

Lagercrantz U, Ellegran H, Anderson L, 1993. The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Research 21:1111-1115.

Liu K, Muse S. 2004. POWERMARKER: New genetic data analysis software Version 3.0. Free programme distributed by the author over the internet from http://www. Powermarker.net.

Mace ES, Hutokshi K, Buhariwalla HK, Crouch JH. 2004. A high-throughput DNA extraction protocol for tropical molecular breeding programs. Plant Molecular Biology Reporter 21:459a-459f.

Menz MA, Klein RR, Unhruh NC, Rooney Wl, Klein PE, Mullet JE. 2004. Genetic diversity of public inbreds of sorghum determined by mapped AFLP and SSR markers. Crop Science 44:1236-1244.

Mitchell SE, Kresovich S, Jester CA, Hernández J, Szewc-McFadden AK. 1997. Application of multiplex and fluorescence-based, semi-automated allele sizing technology for genotyping plant genetic resources. Crop Science 37:617-624.

Morell MK, Peakall R, Appels R, Preston LR, Lloyd HL. 1995. DNA profiling techniques for plant variety identification. Australian Journal of Experimental Agriculture 35:807-819.

Nkongolo KK, Nsapato L. 2003. Genetic diversity in *Sorghum bicolor* (L.) Moench accessions from different ecogeographical regions in Malawi assessed with RAPDs. Genetic Resources and Crop Evolution 50:149-156.

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

Petit RJ, Mousadik AE, Pons O. 1998. Identifying populations for conservation on the basis of genetic markers. Conservation Biology 12:844-855.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945-959.

Redfearn DD, Buxton DR, Devine TE. 1999. Sorghum intercropping effects on yield, morphology, and quality of forage soybean. Crop Science 39:1380-1384.

Saghai-Maroof MA, Biyashev RM, Yang GP, Zhang Q, Allard RW. 1994. Extraordinarily polymorphic microsatellites DNA in barley: Species diversity, chromosomal locations and population dynamics. Proceedings of the National Academy of Science USA 91:5466-5470.

Schug MD, Hutter CM, Wetterstrand KA, Gaudette MS, Mackay TFC, Aquadro CF. 1998. The mutation rates of di-, tri-, and tetra-nucleotide repeats in *Drosophila melanogaster*. Molecular Biology and Evolution 15:1751-1760.

Senior ML, Murphy JP, Goodman MM, Stuber CW. 1998. Utility of SSRs for determining genetic similarities and relationships in maize using an agarose gel system. Crop Science 38:1088-1098.

Smith JSC, Kresovich S, Hopkins MS, Mitchell SE, Dean RE, Woodman WL, Lee M, Porter K. 2000. Genetic diversity among elite sorghum inbred lines assessed with simple sequence repeats. Crop Science 40:226-232.

Smith JSC, Smith OS. 1992. Fingerprinting crop varieties. Advances in Agronomy 47: 85-140.

Stemler ABL, Harlan JR, de Wet JMJ. 1977. The sorghums of Ethiopia. Economic Botany 31:446-460.

Tao Y, Manners JM, Ludlow MM, Henzell RG. 1993. DNA polymorphisms in grain-sorghum (*Sorghum bicolor* (L.) Moench). Theoretical and Applied Genetics 86:679-688.

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

Uptmoor R, Wenzel W, Friedt W, Donaldson G, Ayisi K. 2003. Comparative analysis on the genetic relatedness of *Sorghum bicolor* accessions from Southern Africa by RAPDs, AFLPs and SSRs. Theoretical and Applied Genetics 106:1316-1325.

Vierling RA, Xiang Z, Joschi CP, Gilbert ML, Nguyen HT. 1994. Genetic diversity among elite sorghum lines revealed by restriction fragment length polymorphisms and random amplified polymorphic DNAs. Theoretical and Applied Genetics 87:816-820.

VSN International Ltd. 2007. Genstat Discovery Edition 3. VSN International Ltd. Hernel Hempstead, UK.

Weber JL. 1990. Informativeness of human $(dC\text{-}dA)_n$ (dG-dT)n polymorphisms. Genomics 7:524-530.

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358-1370.

Chapter 5

Comparison of morpho-agronomical and SSR markers for estimating genetic diversity in sorghum

5.1 Abstract

Comparison of results of different methods for estimating genetic diversity is important to evaluate their usefulness in plant breeding and germplasm conservation. In this study, morpho-agronomical and microsatellite or SSR markers were used to evaluate 659 sorghum accessions from Burundi, Ethiopia, Rwanda, Sudan, and Uganda for genetic diversity and also discrimination power of the techniques. The mean morpho-agronomical dissimilarity (0.43 with a range of 0.002-0.81) was high in comparison to dissimilarities calculated using SSR markers (0.21 with a range of 0.15-0.39). The correlation between the morpho-agronomical and the genetic matrix dissimilarity based on SSR data was 0.47 (p=0.001). Results from this study confirmed a clear relationship between molecular and morphological estimates. Although relationships determined using molecular data were different from those determined using morpho-agronomical traits, it remains a useful way to assess diversity for breeding purposes even though the more detailed genetic relationships may be unclear. Morphological traits are largely influenced by the environment thus there is more chance of fluctuation with change in environmental factors while genetic structure is more consistent and not influenced by the environment. Therefore the strategy of combining molecular and morpho-agronomical traits would be best to study genetic diversity of sorghum accessions.

5.2 Introduction

Knowledge of genetic diversity is an important factor in the improvement of crop productivity as well as for conservation of genetic resources (Dean *et al*., 1999; Simioniuc *et al*., 2002). Saffdar *et al*. (2009) reported that existence of adequate genetic diversity in the germplam is important for improvement and efficient use of available material. Conventionally, data on agronomical, morphological and physiological plant traits are used to estimate genetic diversity. However, molecular markers are also now available for authentication and reliable studies of genetic diversity.

Assessment of genetic diversity based on morpho-agronomical characteristics has limitations, since most morphological characters are greatly influenced by environmental factors and the developmental stage of the plant (Morell *et al*., 1995). In contrast, molecular markers based on DNA sequence polymorphism are independent on environmental conditions and show a higher level of polymorphism. In addition, an unambiguous, reliable, fast and cost-effective assessment of genetic diversity is important for determining the uniqueness and distinctiveness of the phenotypic and genetic constitution of genotypes to protect breeder's intellectual property rights (Franco *et al*., 2001).

Molecular markers showing polymorphism among closely related genotypes include RAPDs (Williams *et al*., 1990; Zannou *et al*., 2008), SSRs (Sonnante *et al*., 1994; Gupta and Varshney, 2000; Diouf and Hilu, 2005), and AFLPs (Vos *et al*., 1995; Gillaspie *et al*., 2005). SSR markers are preferred due to their co-dominant inheritance, locus specificity, and multi-allelic character. Hence, they have been recognized as useful genetic markers in many plant species (Cregan *et al*., 1999; Goulão *et al*., 2001). It has been used to differentiate genetic diversity present in elite inbred genotypes and cultivated races of sorghum (Brown *et al*., 1996; Dean *et al*., 1999; Dje *et al*., 2000; Smith *et al*., 2000).

No particular technique is best for the study of genetic diversity in germplasm collections (Singh *et al.,* 1991; Beer *et al*., 1993; Liu and Furnier, 1993; Zhang *et al*., 1993; de Oliveira *et al*., 1996). Different methods test genetic variation at different levels leading to variation in the genetic data generated as well as the superiority of the information content.

In sorghum, information on the estimates of genetic variation has been significantly improved with the implementation of molecular techniques. Seed storage proteins (Shechter and DeWet, 1975), isozymes (Aldrich *et al.,* 1992; Morden *et al*., 1989; 1990), RFLP (Aldrich and Doebley, 1992), AFLP (Uptmoor *et al*., 2003; Perumal *et al*., 2007), SSR (Taramino *et al*., 1997; Kong *et al.,* 2000), and RAPD (Menkir *et al.,* 1997; de Oliveira *et al*., 1996) have been used to estimate genetic variation in sorghum. These markers vary in the level at which they identify genetic variation (and hence extent of genome coverage), extent of polymorphism, number of loci, molecular basis of the

polymorphism and amenability to statistical estimation of population genetics parameters (Gepts, 1995; Hamrick and Godt, 1997).

In terms of combining morpho-agronomical and SSR data, Geleta and Labuschagne (2005) used AFLP, SSR, and morpho-agronomical markers to evaluate 45 sorghum accessions in terms of genetic diversity assessment and discrimination power. Their results demonstrated that both AFLP and SSR-based data matrices differentiated more distinctly between the 45 accessions than morpho-agronomical trait data and genetic diversity estimates from morpho-agronomic traits were not well suited for clarifying more complex relationships but was adequate for estimating the overall pattern of genetic variation among accessions.

SSR analysis was used in this study to complement the previous studies based on morpho-agronomical characters. The objective of this study was to compare the use of morpho-agronomical and SSR markers to assess genetic diversity in sorghum accessions from eastern Africa.

5.3 Materials and methods

5.3.1 Plant material

A total of 659 sorghum accessions from five countries namely Burundi, Ethiopia, Rwanda, Sudan and Uganda were used. Accessions were selected based on the availability of both morpho-agronmical and molecular data for each of the selected accessions.

5.3.2 Methods

5.3.2.1 Morpho-agronomic traits

Data collection methods described for 13 qualitative and five quantitative traits in Chapter 3, sections 3.3.3.2 and 3.3.4.2 were used.

5.3.2.2 SSR analysis

A total of 39 SSR loci (Chapter 4, Table 4.2) were used for the molecular diversity study. Primer sequence information, SSR repeat motifs, PCR amplification conditions and visualization of amplified fragments are discussed in Chapter 4.

5.3.3 Statistical analysis

Data generated in Chapters 3 and 4 were imported into R software version 2.8.1 (R Development Core Team, 2008) to calculate dissimilarity matrices. In order to group accessions based on both morpho-agronomical (qualitative and quantitative data) and molecular data, cluster analysis was conducted using the Euclidean distance matrix using UPGMA. From these dissimilarity distance matrices, the mean genetic distances, and distribution of dissimilarity values, were calculated.

The relationships between the Euclidean distance matrix based on Euclidean distance square obtained with morpho-agronomical data and the simple matching coefficient matrix obtained with SSR markers were analyzed using the approach developed by Mantel (1967) using Pearson's correlation coefficient. The principle of this approach is to calculate the sum of the cross product of the distance matrices and to compare this sum with the value expected according to the null hypothesis (no difference between the distance matrices).

5.4 Results

5.4.1 Distribution of dissimilarity coefficients

Comparison of dissimilarity coefficients (minimum, maximum and mean) for 659 accessions of sorghum generated from morpho-agronomical data, molecular markers, and combined data (clustering using the Euclidean dissimilarity coefficient) is presented in Table 5.1.

Table 5.1 Minimum, maximum and mean Euclidean dissimilarity coefficients for morpho-agronomical, SSR, and combined data

Marker type	Minimum	Maximum	Mean
Morpho-agronomical	0.002	0.81	0.43
SSR	0.150	0.39	0.21
Combined data	0.324	0.95	0.68

The dissimilarity coefficients based on morpho-agronomical data ranged from 0.002 to 0.81 with an overall mean of 0.43. Based on SSR data, coefficients ranged from 0.15 to 0.39 with an average of 0.21. For combined data, coefficients ranged from 0.32 to 0.95 with a mean value of 0.68.

5.4.2 Correlations between dissimilarity matrices

In order to compare the extent of agreement between dendrograms derived based on Euclidean dissimilarity coefficient from morpho-agronomical and SSR data, a distance matrix was constructed for each assay and compared using the Mantel matrix correspondence test. A highly significant positive correlation was found between morpho-agronomical and SSR data (r=0.47, p=0.001). This indicated that the two independent data sets likely reflected the same pattern of genetic diversity and validated the use of these data to calculate the different diversity statistics for eastern African sorghum accessions.

5.4.3 Clustering based on morpho-agronomical and SSR markers

The dendrogram generated based on morpho-agronomical data (Euclidean dissimilarity coefficient) showed four major clusters at the mean genetic dissimilarity of 0.08 (Figure 5.1). Cluster group 1 consisted of 15% of all accessions and contained accessions from Burundi, Ethiopia, and Sudan, but mainly accessions from Ethiopia (89%). Forty six percent of all accessions from Ethiopia clustered in group 2. Cluster group 2 contained 39% of all accessions and contained accessions from all countries, with 72%, 62% and 46% of accessions from Sudan, Uganda, and Ethiopia, respectively. Cluster group 3 contained 18% of accessions from all counties expect Burundi, and mainly contained accessions from Rwanda, with 94% of accessions from Rwanda clustering in this group. Cluster group 4 contained 28% of the total accessions, containing accessions from Burundi, Sudan, and Uganda but consisted mainly of accessions from Burundi (74%). Eighty three percent of Burundi's accessions clustered within group 4. Group 2 was the most diverse, containing 46% accessions from Ethiopia, 72% from Sudan, and 62% from Uganda while the other three groups mainly consisted of accessions from a single country. The exact distribution of accessions within each cluster is given in Table 5.2.

A dendrogram generated from the molecular data (based on the Euclidean dissimilarity coefficient) is presented in Figure 5.2. The UPGMA cluster analysis revealed four

clusters at the mean genetic dissimilarity of 0.025. The first cluster contained 381 accessions (58%), including all accessions from Burundi, Rwanda and Uganda while 4% of the accessions in this group were from Ethiopia. The second cluster contained 173 accessions (26% of all accessions), all from Ethiopia, with dissimilarity values of 0.025. The third and fourth cluster contained accessions from Sudan (3% and 13% of all accessions, respectively). The exact distribution of accessions within each cluster is given in Table 5.3.

The dendrogram derived from the combined data using the Euclidean dissimilarity coefficient showed four major clusters at the mean genetic dissimilarity of 0.04 (Figure 5.3). All accessions from Sudan were found in group 1 and 2 (1% and 15% of all accessions, respectively). The third cluster contained 177 accessions (27%), all from Ethiopia. A total of 377 accessions (57%) from Burundi, Rwanda, Uganda, and Ethiopia were accumulated in group 4. All accessions from Burundi and Uganda clustered in this group. The exact distribution of accessions within each cluster is given in Table 5.4.

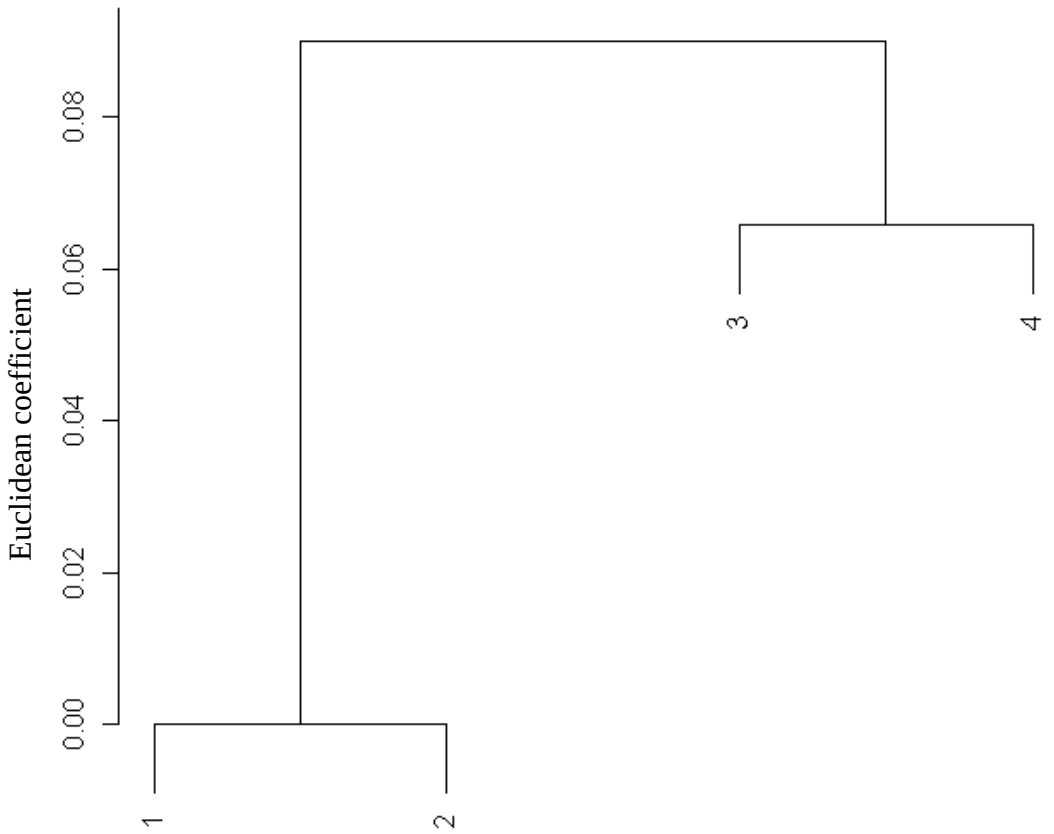

Figure 5.1 Dendrogram showing four cluster groups amongst the 659 sorghum accessions based on morpho-agronomical data.

Table 5.2 Distribution of the 659 sorghum accessions into four clusters based on morpho-agronomical data and clustering per country

Country	Cluster				Total per country
	Group 1	Group 2	Group 3	Group 4	
Burundi	10	17	0	**136**	163
Ethiopia	**89**	**87**	12	0	188
Rwanda	0	5	**85**	0	90
Sudan	1	76	17	11	105
Uganda	0	70	7	36	113
Total	100	255	121	183	659

Table 5.3 Distribution of the 659 sorghum accessions into four clusters based on SSR markers data and clustering per country

Country	Cluster				Total per country
	Group 1	Group 2	Group 3	Group 4	
Burundi	**163**	0	0	0	163
Ethiopia	15	**173**	0	0	188
Rwanda	90	0	0	0	90
Sudan	0	0	**20**	**85**	105
Uganda	113	0	0	0	113
Total	381	173	20	85	659

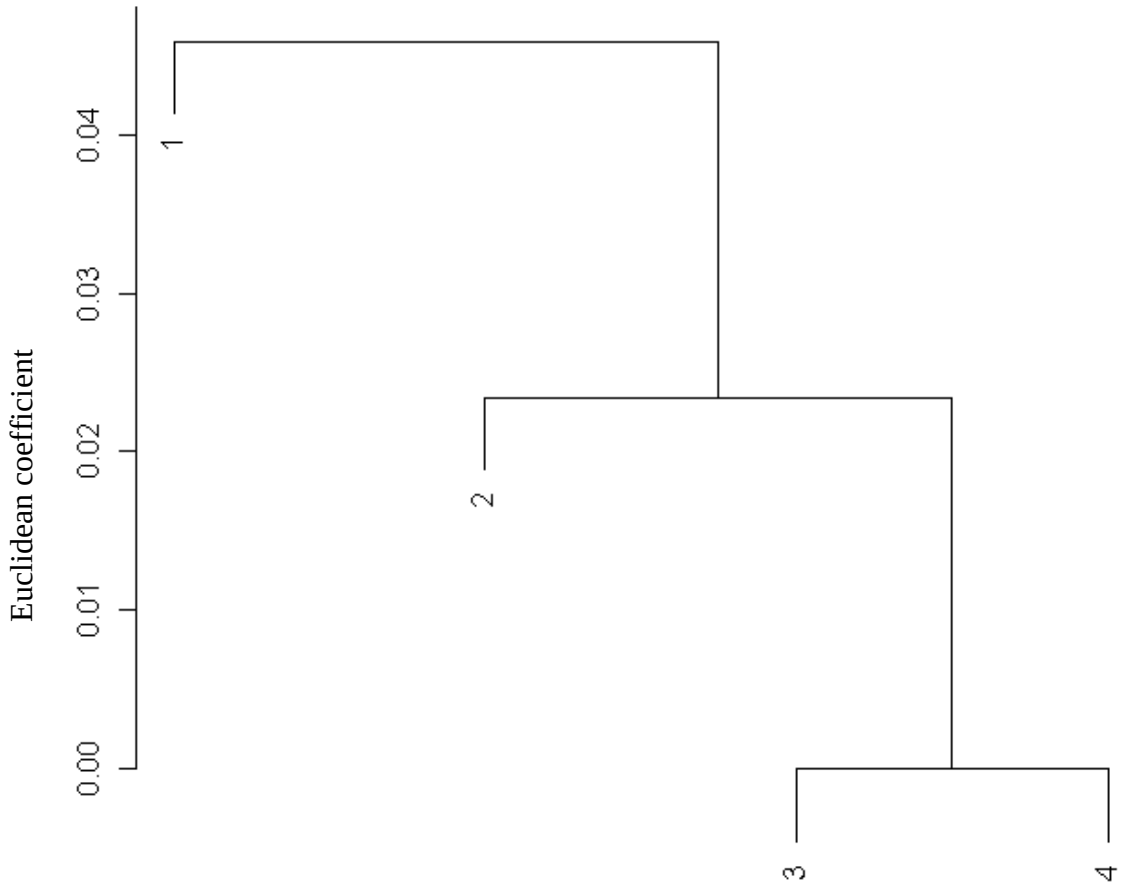

Figure 5.2 Dendrogram showing four cluster groups among the 659 sorghum accessions based on SSR data.

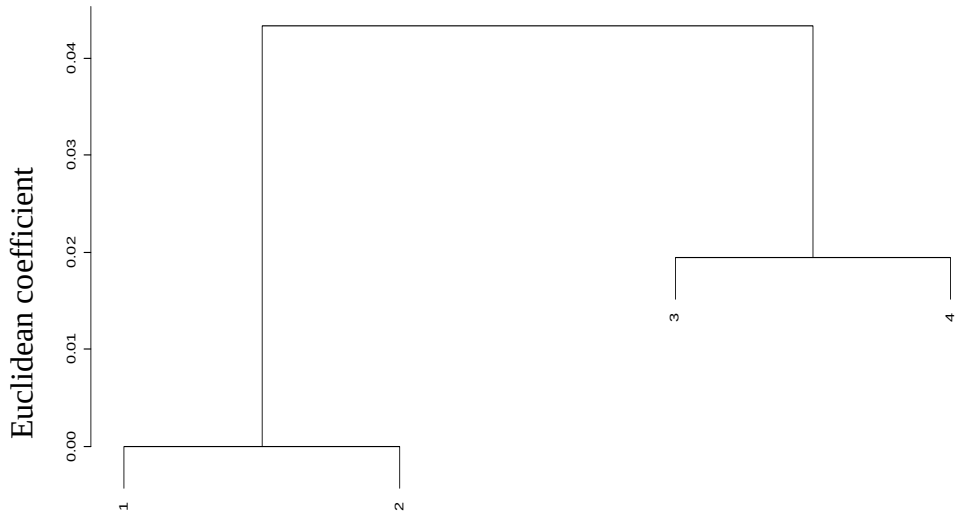

Figure 5.3 Dendrogram showing four cluster groups among the 659 sorghum accessions based on combined data.

Table 5.4 Distribution of the 659 sorghum accessions into four clusters based on combined data and clustering per country

Country	Cluster				Total per country
	Group 1	Group 2	Group 3	Group 4	
Burundi	0	0	0	**163**	163
Ethiopia	0	0	**177**	11	188
Rwanda	0	0	0	90	90
Sudan	9	**96**	0	0	105
Uganda	0	0	0	113	113
Total	9	96	177	377	659

Based on morpho-agronomical data, accessions from Burundi were found in all clusters except in group 3, but 83% of the accessions clustered within one group (group 4) with a small percentage of accessions clustering with accessions from Sudan and Uganda. However, accessions from Burundi accumulated mainly in one group based on SSR marker and combined data. Accessions from Ethiopia clustered in three of the four groups based on morpho-agronomical data but mainly clustered into two closely linked groups based on SSR marker data and mainly into a single group based on combined data. Accessions from Rwanda clustered into two groups based on morpho-agronomical data with 94% in a single group. In the SSR and combined dendrograms all accessions from Rwanda grouped into a single group. Sudanese accessions were spread across all clusters based on morpho-agronomical data with 72% in a single group together with accessions from all other countries. Based on SSR marker data and combined data, they grouped into two closely linked groups. Accessions from Uganda were present in three of the four groups based on morpho-agronomical data with 62% in a single group. However, all accessions from Uganda grouped into a single group based on SSR as well as combined data and mainly clustered with accessions from Burundi and Rwanda.

The dendrograms based on SSR data alone was very similar to the dendrogram based on combined data. This might be attributed to the higher number of data points generated by SSR analysis compared to morpho-agronomical data. However, morpho-agronomical analysis detected higher levels of diversity between accessions in each country as well as between countries. More accessions from different countries clustered together based on morpho-agronomical data compared to using SSR or combined data. Furthermore, accessions from Burundi, Rwanda, and Uganda were more similar to each other based on SSR data compared to morpho-agronomical data.

5.5 Discussion

The distribution of dissimilarity values for morpho-agronomical and genetic dissimilarity differed significantly. Morpho-agronomical dissimilarity covered a greater range (0.002-0.81), while SSR data revealed a lower dissimilarity range (0.15-0.39). In other words, SSR data detected 61-85% similarity while morpho-agronomical data detected a much bigger similarity of 19.0-99.8%. Although morpho-agronomical markers detected a high level of similarity between certain accessions (99.8%), they detected much lower levels of similarity in other accessions (19%) compared to SSR data. Regardless of this, morpho-

agronomical traits are helpful for preliminary assessment since they can be rapidly assessed, are uncomplicated, and can be used as a universal approach for assessing genetic diversity among morphologically distinguishable accessions.

The significance of correlation of the matrices (r=0.47) based on morpho-agronomical and SSR data using the Mantel test showed that a good association existed between phenotypic and genetic distance. In a similar study, Tatineni *et al*. (1996) reported a high correlation between RAPD and morphological characters. Geleta and Labuschagne (2005) also found that morphological and SSR markers were significantly related. However, morphological and AFLP data showed non-significant correlation. Geleta *et al*. (2006) furthermore reported a significant correlation between SSR and morphological traits among 45 accessions using ten SSR primers. This is in contrast to Ayana (2001) who detected no genetic relationship (allozyme plus RAPD data) with quantitative agro-morphological traits. Dhalberg *et al*. (2002) reported an insufficient relationship between RAPD markers and agronomic descriptors. Morphological traits are more influenced by the environment and accordingly there is a higher possibility of fluctuation with change in environmental factors while genetic structure is steadier with less influence of the environment (Romero *et al*., 2009).

The two dendrograms based on either morpho-agronomical and SSR data both separated the 659 sorghum accessions into four cluster groups. The outputs of each of these clusters showed some similarities. For instance, in the SSR dendrogram accessions from Burundi, Rwanda, and Uganda clustered together while accessions from Rwanda and Burundi also clustered together in the morpho-agronomical dendrogram. Accessions from Ethiopia and Sudan also tended to cluster together in both dendrograms. One of the main differences between the two dendrograms is the clustering of accessions from Uganda. In the morpho-agronomical dendrogram they mainly clustered with accessions from Ethiopia and Sudan while they clustered with accessions from Burundi and Rwanda based on SSR data. Another difference between the morpho-agronomical and SSR dendrograms is that accessions from each country mainly clustered into one main group in the SSR based dendrogram while most accessions from all countries were spread over two to four groups in the morpho-agronomical dendrogram. Outputs of the dendrogram based on combined data were almost identical to the dendrogram based on SSR data alone. It furthermore revealed some similarities to morpho-agronomical and SSR dendrograms since

accessions from Burundi, Rwanda, and Uganda clustered together. The main differences between the three dendrograms are the clustering of accessions from Sudan. They mainly clustered with Burundi, Rwanda, Ethiopia and Uganda in the morpho-agronomical dendrogram while they were absent in the same clustered group in both SSR and combined dendrograms. However, in the morpho-agronomical dendrogam, 73% of the Sudanese accessions clustered with 94% of the Ethiopian accessions in groups 1 and 2 and this relationship was also seen in the SSR and combined dendrograms where the closest group to the Sudanese accessions where the group containing accessions from Ethiopia.

The reason behind the differences observed between dendrograms based on morpho-agronomical data and SSR data is that the morphological traits are controlled by a subset of the genomic regions, while most molecular markers sample random genomic regions (Williams *et al*., 1990; Joyee *et al*., 1999; Dahlberg, 2000) most of which are likely to be related to the morphological traits. As a result, markers like SSRs may accurately assay the degree of genetic change distinguishing two genomes, but they may not necessarily reflect the divergence in terms of changes in traits of agronomic importance, which are subjected to selective modifications.

In general, the outcome of this study showed that by means of the SSR molecular technique, a big set of informative data could be generated in less time than with morpho-agronomical analysis. In addition, when molecular markers are used in combination with morpho-agronomic traits to evaluate genotypes, it should be possible to get a significantly smaller subset of marker-fragments that can be used in combination with accessible morpho-agronomic data to improve classification of genotypes compared to using only quantitative or only qualitative traits. The present results imply that although morpho-agronomical characterization is influenced by the environment and is time consuming in general, among other disadvantages, in relation to SSRs it can still be an important and practical means of making progress in germplasm evaluation by conservationists and breeders. Clustering of accessions from Rwanda together with accessions from Burundi confirmed results obtained using NJ and Bayesian based clustering in the previous chapter and could be caused either by the occurrence of frequent gene flow among these two countries, for instance as a consequence of seed exchanges among farmers, or by a restriction of the intensity of genetic drift due to a highly effective population size. This

could be a potentially important source of germplasm for further crop improvement in these countries.

5.6 Conclusions

Characterization of sorghum accessions at DNA level can help identify genetically representative, non-redundant sets of germplasm for sorghum breeding and conservation purposes. As observed from the significant correlation coefficient obtained between genetic distance values from the two marker techniques (morpho-agronomical and SSR), all have shown a comparable genetic diversity level. It further indicates that diversity assessed using molecular markers may efficiently represent the genetic diversity in morpho-agronomical traits. Although the relationships determined by molecular data were different to those identified using morpho-agronomical traits, the latter is still useful in assessing genetic diversity for the purpose of breeding selection on condition that the genotypes under investigation are not too closely related. Molecular techniques have a clear advantage over morpho-agronomical traits in elucidating complex relationships, especially of genotypes sharing morpho-agronomical traits or coming from the same geographic location.

This data will help to establish a sorghum core collection in east Africa. The use of population diversity based approaches and SSR assessments can be used to identify genomic regions of interest.

5.7 References

Aldrich PR, Doebley J. 1992. Restriction fragment variation in the nuclear and chloroplast genomes of cultivated and wild *Sorghum bicolor*. Theoretical and Applied Genetics 85:293-302.

Aldrich PR, Doebley J, Schertz KF, Stec A. 1992. Patterns of allozyme variation in cultivated and wild *Sorghum bicolor*. Theoretical and Applied Genetics 85:451-460.

Ayana A. 2001. Genetic diversity in sorghum (*Sorghum bicolor* (L.) Moench) germplasm from Ethiopia and Eritrea. Ph.D. thesis. Department of Biology, Science Faculty, Addis Ababa University, Addis Ababa, Ethiopia.

Beer SR, Goffreda J, Phillips TD, Murphy JP, Sorrells ME, 1993. Assessment of genetic variability in *Avena sterilis* using morphological traits, isozymes, and RFLPs. Crop Science 33:1386-1393.

Brown SM, Hopkins MS, Mitchell SE, Senior ML, Wang TY, Duncan RR, Gonzalez-Candelas F, Kresovich S. 1996. Multiple methods for the identification of polymorphic simple sequence repeats (SSRs) in sorghum [*Sorghum bicolor* (L.) Moench]. Theoretical and Applied Genetics 93:190-198.

Cregan PB, Jarvick T, Bush AL, Shoemaker RC, Lark KG, Kahler AL, Kaya N, Van Toai TT, Lohnes DG, Chung J, Specht JE. 1999. An integrated genetic linkage map of the soybean genome. Crop Science 39:1464-1490.

Dahlberg JA. 2000. Classification and characterization of Sorghum. In: Sorghum origin, history, technology and production. Edited by Smith CW, Frederiksen RA. Wiley and Sons, New York, pp. 99-130.

Dahlberg JA, Zhang X, Hart GE, Mullet JE. 2002. Comparative assessment of variation among sorghum germplasm accessions using seed morphology and RAPD measurements. Crop Science 42:291-296.

de Oliveira AC, Richter T, Bennetzen JL. 1996. Regional and racial specificities in sorghum germplasm assessed with DNA markers. Genome 39:579-587.

Dean RE, Dahlberg JA, Hopkins MS, Mitchell SE, Kresovich S. 1999. Genetic redundancy and diversity among 'Orange' accessions in the U.S. national sorghum collection as assessed with simple sequence repeat (SSR) markers. Crop Science 39:1215-1221.

Diouf D, Hilu KF. 2005. Microsatellites and RAPD markers to study genetic relationships among cowpea breeding lines and local varieties in Senegal. Genetic Resources and Crop Evolution 52:1067-1067.

Dje Y, Heuertz M, Lefébvre C, Vekemans X. 2000. Assessment of genetic diversity within and among germplasm accessions in cultivated sorghum using microsatellite markers. Theoretical and Applied Genetics 100:918-925.

Franco J, Crossa J, Ribaut JM, Betran J, Warburton ML, Khairallah M. 2001. A method for combining molecular markers and phenotypic attributes for classifying plant genotypes. Theoretical and Applied Genetics 103:944-952.

Geleta N, Labuschagne MT. 2005. Qualitative traits variation in sorghum [*Sorghum. bicolour* (L.) Moench] germplasm from eastern highland of Ethiopia. Biodiversity and Conservation 14:3055-3064.

Geleta N, Labuschagne MT, Viljoen CD. 2006. Genetic diversity analysis in sorghum germplasm as estimated by AFLP, SSR and morpho-agronomical markers. Biodiversity and Conservation 15:3251-3265.

Gepts P. 1995. Genetic markers and core collections. In: Core collections of plant genetic resources. Edited by Hodgkin T, Brown AHD, van Hintum TJL, Morales EAV. International Board of Plant Genetic Resources Institute (IBPGRI), A Wiley-Sayce Publication, pp. 127-146.

Gillaspie Jr AG, Hopkins MS, Dean RE. 2005. Determining genetic diversity between lines of *Vigna ungiculata* subspecies by AFLP and SSR markers. Genetic Resources and Crop Evolution 52:245-247.

Goulão L, Monte-Corvo L, Oliveira CM. 2001. Phenetic characterization of plum cultivars by high multiplex ratio markers: Amplified fragment length polymorphisms and inter-simple sequence repeats. Journal of American Society for Horticultural Science 126:72-77.

Gupta PK, Varshney RK. 2000. The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113:163-185.

Hamrick JL, Godt JW. 1997. Allozyme diversity in cultivated crops. Crop Science 37:26-30.

Joyee TA, Abberton MT, Michaelson-Yeates TPT, Forster JW. 1999. Relationships between genetic distance measured by RAPD-PCR and heterosis in inbred lines of white clover (*Trifolium repens* L.). Euphytica 107:159-165.

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

Liu Z, Furnier GR. 1993. Comparison of allozyme, RFLP, and RAPD markers for revealing genetic variation within and between trembling aspen and big tooth aspen. Theoretical and Applied Genetics 87:97-105.

Mantel N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Research 27:209-220.

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

Morden CW, Doebley JF, Schertz KF. 1989. Allozyme variation in Old World races of *Sorghum bicolor* (Poaceae). American Journal of Botany 76:247-255.

Morden CW, Doebley JF, Schertz KF. 1990. Allozyme variation among the spontaneous species of *Sorghum* section Sorghum (Poaceae). Theoretical and Applied Genetics 80:296-304.

Morell MK, Peakall R, Appels R, Preston LR, Lloyd HL. 1995. DNA profiling techniques for plant variety identification. Australian Journal of Experimental Agriculture 35:807-819.

Perumal R, Renganayaki K, Menz MA, Katilé S, Dahlberg J, Magill CW, Rooney WL. 2007. Genetic diversity among sorghum races and working groups based on AFLP and SSR. Crop Science 47:1375-1383.

R Development Core Team. 2008. R: A language and environment for statistical computing. R. Foundation for Statistical Computing, Vienna, Austria (http://www.R-project. org).

Romero G, Adeva C, Battad II Z. 2009. Genetic fingerprinting: advancing the frontiers of crop biology research. Philosophy of Science Letter 2:8-13.

Saffdar H, Ashfaq M, Hameed S, Ul Haque I, Mujeeb-Kazi A. 2009. Molecular analysis of genetic diversity in elite II synthetic hexaploid wheat screened against Barley yellow dwarf virus. African Journal of Biotechnology 8:3244-3250.

Shechter Y, de Wet JMJ. 1975. Comparative electrophoresis and isozyme analysis of seed proteins from cultivated races of sorghum. American Journal of Botany 62:254-261.

Simioniuc D, Uptmoor R, Friedt W, Ordon W. 2002. Genetic diversity and relationships among pea cultivars (*Pisum sativum* L.) revealed by RAPDs and AFLPs. Plant Breeding 121:429-435.

Singh SP, Gutierrez JA, Molina A, Urrea C, Gepts P. 1991. Genetic diversity in cultivated common bean: II. Marker-based analysis of morphological and agronomic traits. Crop Science 31:23-29.

Smith JSC, Kresovich S, Hopkins MS, Mitchell SE, Dean RE, Woodman WL, Lee M, Porter K. 2000. Genetic diversity among elite sorghum inbred lines assessed with simple sequence repeats. Crop Science 40:226-232.

Sonnante G, Stockton T, Nodari RO, Velasquez VLB, Gepts P. 1994. Evolution of genetic diversity during domestication of common bean (*Phaseolus vulgaris* L.). Theoretical and Applied Genetics 89:629-635.

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

Tatineni V, Cantrell RG, Davis DD. 1996. Genetic diversity in elite cotton germplasm determined by morphological characters and RAPDs. Crop Science 36:186-192.

Uptmoor R, Wenzel W, Friedt W, Donaldson G, Ayisi K, Ordon F. 2003. Comparative analysis on the genetic relatedness of *Sorghum bicolor* accessions from Sourthern Africa by RAPDs, AFLPs and SSRs. Theoretical and Applied Genetics 106: 1316-1325.

Vos P, Hogers R, Bleeker M, Reijans M, Van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research 23:4407-4414.

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey S. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research 18:6531-6535.

Zannou A, Kossou DK, Ahanchede A, Zoundjihekpon J, Agbicodo Struik PC, Sanni A. 2008. Genetic variability of cultivated cowpea in Benin assessed by random amplified polymorphic DNA. African Journal of Biotechnology 7:4407-4414.

Zhang Q, Saghai-Maroof MA, Kleinhofs A. 1993. Comparative diversity of analysis of RFLPs and isozymes within and among populations of *Hordeum vulgare* ssp. *spontaneum*. Genetics 134:909-916.

Chapter 6

General conclusions and recommendations

Sorghum is essential to diets of poor people in the semi-arid tropics where droughts cause frequent failures of other crops. In Africa, sorghum is still mainly a subsistence food crop. There is no doubt that sorghum is essential for the achievement of food security and for the establishment of sustainable production systems in the semi-arid regions of eastern Africa. It grows in areas where the annual rainfall is in the range of 500-700 mm per year. Therefore, most countries in east Africa where sorghum is a major arable crop is arid and areas are at risk of desertification.

High levels of genetic diversity in plants are necessary to improve utilization of a broader range of varieties to meet the increased food demands. It provides farmers and plant breeders with options to develop, through selection and breeding, new and more productive crops that are adapted to changing environments.

Germplasm collections have become an important part of several breeding programmes at both national and international levels. Advances in the development of new and acceptable crop varieties and hybrids are highly dependent on the diversity of resource materials available to breeders. One of the main objectives of germplasm collections is to gather and preserve the genetic diversity in order to ensure its sustained feasibility to meet the needs of different users.

Morpho-agronomical and molecular marker techniques have different advantages for assessing genetic relations. Studies that combine the two methods can thus use the advantages of both techniques. The current study was mostly motivated by the current need to quantify and understand the partitioning of genetic diversity of sorghum germplasm collections from east Africa, to study the genetic structure and distribution of diversity within and between countries, and to study the combination of morpho-agronomical and molecular markers in order to provide a complete picture of the genetic variation among sorghum accessions from eastern Africa.

A good understanding of the morpho-agronomical diversity related to accessions' distribution in east Africa is an important tool for efficient management of crop genetic resources. This study, using 18 morpho-agronomical characters, indicated a wide range of variation recorded in both qualitative and quantitative characters and for some accessions variation was also detected within accessions. The 1013 studied accessions revealed that an impressive range in morpho-agronomical diversity existed among east African sorghum accessions. This study has recognized and identified both desirable and undesirable traits which can be used to improve sorghum production for many purposes. Results observed may help sorghum researchers in the choice of sorghum accessions to plant in each country during crop selection experiments and trials in farmer's fields. The information obtained by this study will facilitate the conservation and utilization of the materials studied.

For all SSR loci studied, a high level of genetic polymorphism across all country groups was observed, which confirmed that the sorghum microsatellite kit is an excellent tool to access diversity of this crop. The diversity assessment of the east African collection with 39 SSR markers revealed that there is a high level of genetic diversity in the collection and also showed a low genetic differentiation between Sudanese accessions and those from Kenya, suggesting exchange of material between them. The high amount of diversity found in the east African collection could be explained by the large amount of diversity present in the original populations, as well as heterozygosity. The differentiation patterns in sorghum are important for the development of sorghum breeding programmes. The evaluation of the genetic variation of accessions from different countries will help confirm or find new differentiation patterns. Sorghum diversity studies have been focused on country level, and little is known about sorghum genetic diversity present at a regional level. Thus it is suggested that future work include the systematic collection of new accessions on a large scale as part of a conservation strategy for sorghum genetic resources for the east African region. The high level of genetic diversity found in the east African collection have important applications in the *ex situ* and *in vitro* sorghum conservation programmes.

Challenges for the future

There continues to be a considerable need for research on many aspects of the extent and distribution of genetic diversity in east Africa:

i) Investigate the ways in which farmer management practices and ecological or geographic factors interact to determine sorghum population structure.

ii) Study the impact of ecogeographic factors and domestication events, and socio-economic, cultural, and political factors on the diversity found in sorghum germplasm.

iii) Investigate such factors as the distribution of allelic variation within and between populations, predominantly with respect to multi-allelic associations and the significance of linkage disequilibrium in determining the importance of linkage and allelic associations in sorghum.

iv) Study the extent and effect of introgression between sorghum and their wild relatives.

v) Study multi-location trials for evaluation of germplasm for important agronomic characters like photoperiod sensitivity, forage yield and its related traits, responses to biotic stresses, etc. to identify locally adapted material for use in breeding programmes.

vi) There is a need for future germplasm characterization that combines both the classical phenotypical characterization and biotechnological tools, to ensure a more complete and informative characterization that reveals the true genetic diversity of accessions in terms of nutritional quality.

Summary

Eastern Africa, where sorghum is a significant arable crop, is arid and areas are at risk of desertification. Consequently, many valuable landraces of sorghum are at risk for genetic erosion. Genetic diversity plays a vital role in the success of any breeding programme. This study aimed to investigate the genetic diversity and genetic relationships in germplasm accessions among east African countries using morpho-agronomical and simple sequence repeat (SSR) markers in order to (i) understand the extent of genetic variation in different countries, and (ii) quantify the genetic structure and how the diversity is distributed among and within countries. The extent of regional patterns of phenotypic diversity was assessed in 1013 accessions using 13 qualitative and five quantitative traits. A total of 1108 sorghum accessions from Sudan, Kenya, Uganda, Ethiopia, Eritrea, Rwanda, and Burundi were assessed using 39 SSR markers. Results on qualitative and quantitative traits data showed that there was a high level of morpho-agronomical diversity among accessions studied. Analysis of variance revealed highly significant differences between accessions pooled over countries and between countries. Based on SSR data, Sudan had the highest genetic diversity and Rwanda the lowest. Levels of genetic diversity differed significantly, with most of the diversity being partitioned more within than between countries. Results suggested that despite sorghum's predominantly autogamous mating system, countries clustered totally separately with almost no integration and therefore emphasized the role of farmers' practices in the preservation of landrace identity and the favouring of genetic diversity. The morpho-agronomical and molecular marker data showed a high level of variation among accessions, and indicated that sorghum populations studied were a mixture of a large number of different genotypes. Therefore, future germplasm collection should take all levels of variation into consideration.

Key words: diversity, eastern Africa, genetic structure, germplasm, morpho-agronomical, SSR, *Sorghum bicolor*.

Opsomming

Oostelike Afrika, waar sorghum ‘n belangrike verboude gewas is, is waterarm en gebiede is in gevaar om woestynagtig te word. Die gevolg is dat waardevolle sorghum landrasse deur genetiese erosie bedreig word. Genetiese diversiteit speel ‘n belangrike rol in die sukses van enige teelprogram. Die doel van hierdie studie was om die genetiese diversiteit en genetiese verwantskappe in kiemplasma genotipes uit oos Afrika lande te ondersoek deur van morfo-agronomiese en eenvoudig herhalende volgorde (SSR) merkers gebruik te maak ten einde (i) die omvang van genetiese variasie in verskillende lande te ondersoek en (ii) die genetiese struktuur te kwantifiseer en te bepaal hoe die diversiteit tussen en binne lande versprei is. Die fenotipiese diversiteitspatrone van verskillende streke is ondersoek deur van 13 kwalitatiewe en vyf kwantitatiewe eienskappe gebruik te maak. ‘n Totaal van 1108 sorghum genotipes van Sudan, Kenia, Uganda, Etiopië, Eritrea, Rwanda and Burundi is ondersoek deur van 39 SSR merkers gebruik te maak. Resultate gebaseer op kwalitatiewe and kwantitatiewe eienskapdata het aangetoon dat daar ‘n hoë vlak van morfo-agronomiese diversiteit tussen die bestudeerde genotipes was. Analise van variasie het hoogs betekenisvolle verskille tussen genotipes, gekombineer oor lande asook tussen lande, aangetoon. Gebaseer op SSR data het Sudan die hoogste en Rwanda die laagste genetiese diversiteit getoon. Vlakke van genetiese diversiteit het betekenisvol verskil en meeste van die variasie was meer binne as tussen lande verdeel. Resultate het getoon dat ten spyte van sorghum se hoofsaaklik outogame voortplantingsisteem, lande totaal apart gegroepeer het met amper geen integrasie nie. Dit het die rol van boere se praktyke in die bewaring van die identiteit van landrasse en die bevordering van genetiese diversiteit uitgewys. Die morfo-agronomiese en molekulêre merker data het ‘n hoë vlak van variasie tussen genotipes aangedui en aangetoon dat die bestudeerde sorghum populasies ‘n mengsel van ‘n groot aantal verskillende genotipes was. Toekomstige kiemplasma kolleksies behoort dus alle vlakke van variasie in ag te neem.

Sleutelwoorde: diversiteit, eenvoudig herhalende volgorde (SSR), genetiese struktuur, kiemplasma, morfo-agronomies, oostelike Afrika, *Sorghum bicolor*.

Appendix 1. Genotype names and countries used in this study

Sample no	Genotype	Country	Sample no	Genotype	Country
1	BRD1	Burundi	47	BRD54	Burundi
2	BRD2	Burundi	48	BRD55	Burundi
3	BRD3	Burundi	49	BRD56	Burundi
4	BRD4	Burundi	50	BRD57	Burundi
5	BRD5	Burundi	51	BRD58	Burundi
6	BRD8	Burundi	52	BRD59	Burundi
7	BRD9	Burundi	53	BRD61	Burundi
8	BRD10	Burundi	54	BRD62	Burundi
9	BRD11	Burundi	55	BRD63	Burundi
10	BRD12	Burundi	56	BRD64	Burundi
11	BRD13	Burundi	57	BRD65	Burundi
12	BRD14	Burundi	58	BRDMCB4	Burundi
13	BRD15	Burundi	59	BRD67	Burundi
14	BRDMCB5	Burundi	60	BRD68	Burundi
15	BRD18	Burundi	61	BRD69	Burundi
16	BRD20	Burundi	62	BRD70	Burundi
17	BRD22	Burundi	63	BRD71	Burundi
18	BRD23	Burundi	64	BRD72	Burundi
19	BRD24	Burundi	65	BRD74	Burundi
20	BRD25	Burundi	66	BRD75	Burundi
21	BRD26	Burundi	67	BRD76	Burundi
22	BRD27	Burundi	68	BRD77	Burundi
23	BRD28	Burundi	69	BRD78	Burundi
24	BRD29	Burundi	70	BRD79	Burundi
25	BRD30	Burundi	71	BRD80	Burundi
26	BRD31	Burundi	72	BRD81	Burundi
27	BRD33	Burundi	73	BRD82	Burundi
28	BRD34	Burundi	74	BRD83	Burundi
29	BRD35	Burundi	75	BRD86	Burundi
30	BRD36	Burundi	76	BRD87	Burundi
31	BRD37	Burundi	77	BRD88	Burundi
32	BRD38	Burundi	78	BRD89	Burundi
33	BRD39	Burundi	79	BRD90	Burundi
34	BRD40	Burundi	80	BRD91	Burundi
35	BRD41	Burundi	81	BRD92	Burundi
36	BRD42	Burundi	82	BRD93	Burundi
37	BRD43	Burundi	83	BRD94	Burundi
38	BRD44	Burundi	84	BRD95	Burundi
39	BRD45	Burundi	85	BRD96	Burundi
40	BRD46	Burundi	86	BRD97	Burundi
41	BRD47	Burundi	87	BRD98	Burundi
42	BRD48	Burundi	88	BRD99	Burundi
43	BRD49	Burundi	89	BRD100	Burundi
44	BRD51	Burundi	90	BRD101	Burundi
45	BRD52	Burundi	91	BRD102	Burundi

46	BRD53	Burundi	92	BRD103I	Burundi
93	BRD103II	Burundi	139	BRDGBRE50	Burundi
94	BRDGCBE01	Burundi	140	BRDGBRE51	Burundi
95	BRDGCBE02	Burundi	141	BRDGBRE52	Burundi
96	BRDGCBE03	Burundi	142	BRDGBRE53	Burundi
97	BRDGCBE04	Burundi	143	BRDSSB	Burundi
98	BRDGCBE05	Burundi	144	BRDSSR	Burundi
99	BRDGCBE06	Burundi	145	BRDSARIASO 14	Burundi
100	BRDGCBE07	Burundi	146	BRDGambela	Burundi
101	BRDGCBE08	Burundi	147	BRD5DX160	Burundi
102	BRDGCBE09	Burundi	148	BRDSVR8	Burundi
103	BRDGKDE10	Burundi	149	BRDSVR157	Burundi
104	BRDGKDE11	Burundi	150	BRDSRI	Burundi
105	BRDGKDE12	Burundi	151	BRD BLANC I	Burundi
106	BRDGKDE13	Burundi	152	BRD	Burundi
107	BRDGKDE15	Burundi	153	BRDW	Burundi
108	BRDGKDE16	Burundi	154	BRDD	Burundi
109	BRDGKDE17	Burundi	155	BRDGOALA	Burundi
110	BRDGMGE18	Burundi	156	BRDAgashari	Burundi
111	BRDGMGE19	Burundi	157	BRD76T#123	Burundi
112	BRDGMGE20	Burundi	158	BRDMEKO	Burundi
113	BRDGMGE21	Burundi	159	BRDTeshale	Burundi
114	BRDGMGE22	Burundi	160	BRDS35	Burundi
115	BRDGMGE23	Burundi	161	BRDCHIRO	Burundi
116	BRDGBRE24	Burundi	162	BRDETS2752	Burundi
117	BRDGBRE25	Burundi	163	BRDETS1176	Burundi
118	BRDGRTE26	Burundi	164	BRDMCB1	Burundi
119	BRDGRTE28	Burundi	165	BRDMCB2	Burundi
120	BRDGRTE29	Burundi	166	ERT4	Eritrea
121	BRDGRTE30	Burundi	167	ERT5	Eritrea
122	BRDGRTE31	Burundi	168	ERT7	Eritrea
123	BRDGRTE32	Burundi	169	ERT9	Eritrea
124	BRDGRTE33	Burundi	170	ERT10	Eritrea
125	BRDGRTE34	Burundi	171	ERT13	Eritrea
126	BRDGRYE35	Burundi	172	ERT14	Eritrea
127	BRDGRYE36	Burundi	173	ERT16	Eritrea
128	BRDGRYE37	Burundi	174	ERT17	Eritrea
129	BRDGCZE38	Burundi	175	ERT20	Eritrea
130	BRDGCZE39	Burundi	176	ERT21	Eritrea
131	BRDGCZE41	Burundi	177	ERT22	Eritrea
132	BRDGCZE42	Burundi	178	ERT23	Eritrea
133	BRDGCZE43	Burundi	179	ERT24	Eritrea
134	BRDGCZE44	Burundi	180	ERT25	Eritrea
135	BRDGCZE45	Burundi	181	ERT26	Eritrea
136	BRDGCZE46	Burundi	182	ERT28	Eritrea
137	BRDGCZE47	Burundi	183	ERT29	Eritrea
138	BRDGCZE49	Burundi	184	ERT30	Eritrea
185	ERT31	Eritrea	230	ERT82	Eritrea

186	ERT33	Eritrea	231	ERT83	Eritrea
187	ERT35	Eritrea	232	ERT84	Eritrea
188	ERT36	Eritrea	233	ERT85	Eritrea
189	ERT37	Eritrea	234	ERT86	Eritrea
190	ERT39	Eritrea	235	ERT87	Eritrea
191	ERT40	Eritrea	236	ERT88	Eritrea
192	ERT42	Eritrea	237	ERT89	Eritrea
193	ERT43	Eritrea	238	ERT91	Eritrea
194	ERT44	Eritrea	239	ERT93	Eritrea
195	ERT45	Eritrea	240	ERT94	Eritrea
196	ERT46	Eritrea	241	ERT95	Eritrea
197	ERT47	Eritrea	242	ERT96	Eritrea
198	ERT49	Eritrea	243	ERT97	Eritrea
199	ERT50	Eritrea	244	ERT98	Eritrea
200	ERT51	Eritrea	245	ERT99	Eritrea
201	ERT52	Eritrea	246	ERT100	Eritrea
202	ERT53	Eritrea	247	ERT101	Eritrea
203	ERT54	Eritrea	248	ERT102	Eritrea
204	ERT55	Eritrea	249	ERT103	Eritrea
205	ERT56	Eritrea	250	ERT104	Eritrea
206	ERT57	Eritrea	251	ERT105	Eritrea
207	ERT58	Eritrea	252	ERT106	Eritrea
208	ERT59	Eritrea	253	ERT107	Eritrea
209	ERT60	Eritrea	254	ERT108	Eritrea
210	ERT61	Eritrea	255	ERT109	Eritrea
211	ERT62	Eritrea	256	ERT110	Eritrea
212	ERT63	Eritrea	257	ERT111	Eritrea
213	ERT64	Eritrea	258	ERT112	Eritrea
214	ERT65	Eritrea	259	ERT113	Eritrea
215	ERT66	Eritrea	260	ERT114	Eritrea
216	ERT67	Eritrea	261	ERT118	Eritrea
217	ERT69	Eritrea	262	ERT120	Eritrea
218	ERT70	Eritrea	263	ERT121	Eritrea
219	ERT71	Eritrea	264	ERT122	Eritrea
220	ERT72	Eritrea	265	ERT124	Eritrea
221	ERT73	Eritrea	266	ERT125	Eritrea
222	ERT74	Eritrea	267	ERT126	Eritrea
223	ERT75	Eritrea	268	ERT127	Eritrea
224	ERT76	Eritrea	269	ERT128	Eritrea
225	ERT77	Eritrea	270	ERT129	Eritrea
226	ERT78	Eritrea	271	ERT131	Eritrea
227	ERT79	Eritrea	272	ERT132	Eritrea
228	ERT80	Eritrea	273	ERT133	Eritrea
229	ERT81	Eritrea	274	ERT134	Eritrea
275	ERT135	Eritrea	321	69232	Ethiopia
276	ERT136	Eritrea	322	69233	Ethiopia
277	ERT137	Eritrea	323	69234	Ethiopia
278	ERT138	Eritrea	324	69235	Ethiopia

279	ERT139	Eritrea	325	69236	Ethiopia
280	ERT140	Eritrea	326	239178	Ethiopia
281	ERT141	Eritrea	327	69249	Ethiopia
282	ERT142	Eritrea	328	69252	Ethiopia
283	ERT143	Eritrea	329	69262	Ethiopia
284	ERT144	Eritrea	330	69525	Ethiopia
285	ERT145	Eritrea	331	69527	Ethiopia
286	ERT146	Eritrea	332	69528	Ethiopia
287	ERT147	Eritrea	333	210974	Ethiopia
288	ERT148	Eritrea	334	210908	Ethiopia
289	ERT149	Eritrea	335	210949	Ethiopia
290	ERT151	Eritrea	336	210950	Ethiopia
291	ERT152	Eritrea	337	210953	Ethiopia
292	ERT153	Eritrea	338	210973	Ethiopia
293	ERT154	Eritrea	339	212636	Ethiopia
294	ERT155	Eritrea	340	212637	Ethiopia
295	ERT156	Eritrea	341	212639	Ethiopia
296	ERT160	Eritrea	342	212640	Ethiopia
297	ERT161	Eritrea	343	212644	Ethiopia
298	ERT163	Eritrea	344	212646	Ethiopia
299	ERT165	Eritrea	345	213353	Ethiopia
300	ERT168	Eritrea	346	215525	Ethiopia
301	ERT169	Eritrea	347	215526	Ethiopia
302	ERT170	Eritrea	348	215727	Ethiopia
303	ERT172	Eritrea	349	216734	Ethiopia
304	ERT174	Eritrea	350	216735	Ethiopia
305	ERT177	Eritrea	351	216736	Ethiopia
306	69111	Ethiopia	352	216738	Ethiopia
307	69114	Ethiopia	353	216741	Ethiopia
308	69115	Ethiopia	354	217702	Ethiopia
309	69121	Ethiopia	355	219983	Ethiopia
310	69208	Ethiopia	356	219999	Ethiopia
311	69209	Ethiopia	357	220004	Ethiopia
312	69211	Ethiopia	358	220009	Ethiopia
313	69212	Ethiopia	359	220012	Ethiopia
314	69213	Ethiopia	360	220015	Ethiopia
315	69226	Ethiopia	361	220018	Ethiopia
316	69227	Ethiopia	362	221730	Ethiopia
317	69228	Ethiopia	363	222879	Ethiopia
318	69229	Ethiopia	364	222880	Ethiopia
319	69230	Ethiopia	365	222885	Ethiopia
320	69231	Ethiopia	366	223247	Ethiopia
367	223487	Ethiopia	413	234120	Ethiopia
368	223489	Ethiopia	414	235447	Ethiopia
369	223495	Ethiopia	415	235448	Ethiopia
370	223503	Ethiopia	416	235449	Ethiopia
371	223505	Ethiopia	417	235453	Ethiopia
372	223519	Ethiopia	418	235454	Ethiopia

373	223520	Ethiopia	419	235456	Ethiopia
374	223525	Ethiopia	420	235457	Ethiopia
375	223533	Ethiopia	421	235459	Ethiopia
376	223534	Ethiopia	422	235461	Ethiopia
377	223579	Ethiopia	423	235463	Ethiopia
378	223589	Ethiopia	424	235464	Ethiopia
379	225836	Ethiopia	425	235468	Ethiopia
380	225837	Ethiopia	426	235469	Ethiopia
381	226046	Ethiopia	427	235482	Ethiopia
382	226049	Ethiopia	428	235617	Ethiopia
383	226050	Ethiopia	429	235618	Ethiopia
384	226052	Ethiopia	430	235619	Ethiopia
385	226051	Ethiopia	431	235789	Ethiopia
386	226056	Ethiopia	432	235790	Ethiopia
387	226058	Ethiopia	433	235791	Ethiopia
388	226063	Ethiopia	434	235792	Ethiopia
389	226064	Ethiopia	435	235793	Ethiopia
390	227083	Ethiopia	436	237260	Ethiopia
391	227085	Ethiopia	437	237265	Ethiopia
392	227086	Ethiopia	438	237268	Ethiopia
393	227090	Ethiopia	439	237269	Ethiopia
394	229894	Ethiopia	440	237272	Ethiopia
395	229895	Ethiopia	441	237275	Ethiopia
396	229899	Ethiopia	442	237276	Ethiopia
397	234064	Ethiopia	443	237279	Ethiopia
398	234069	Ethiopia	444	237289	Ethiopia
399	234070	Ethiopia	445	237291	Ethiopia
400	234071	Ethiopia	446	237294	Ethiopia
401	234077	Ethiopia	447	237301	Ethiopia
402	234084	Ethiopia	448	237304	Ethiopia
403	234090	Ethiopia	449	237306	Ethiopia
404	234101	Ethiopia	450	237311	Ethiopia
405	234102	Ethiopia	451	238379	Ethiopia
406	234107	Ethiopia	452	238380	Ethiopia
407	234108	Ethiopia	453	238387	Ethiopia
408	234111	Ethiopia	454	238394	Ethiopia
409	234113	Ethiopia	455	238395	Ethiopia
410	234114	Ethiopia	456	238397	Ethiopia
411	234118	Ethiopia	457	238399	Ethiopia
412	234119	Ethiopia	458	238402	Ethiopia
459	238404	Ethiopia	505	Akualem	Kenya
460	238415	Ethiopia	506	Lokilioko	Kenya
461	238420	Ethiopia	507	Edoidoi	Kenya
462	238423	Ethiopia	508	Naliba	Kenya
463	238424	Ethiopia	509	Ekiriente	Kenya
464	238429	Ethiopia	510	Loponoikal	Kenya
465	238432	Ethiopia	511	Nakualem	Kenya
466	238442	Ethiopia	512	Lokilioko	Kenya

467	238445	Ethiopia	513	Ikariboyer	Kenya
468	239126	Ethiopia	514	Longiro Akwaun	Kenya
469	239128	Ethiopia	515	Yolyolten	Kenya
470	239129	Ethiopia	516	Ngakot Sikiria	Kenya
471	239133	Ethiopia	517	Looyakes	Kenya
472	239147	Ethiopia	518	Namesek	Kenya
473	239119	Ethiopia	519	Atotoemug	Kenya
474	239165	Ethiopia	520	Lopook	Kenya
475	239167	Ethiopia	521	Nakosim Ekori	Kenya
476	239168	Ethiopia	522	Lokilioko	Kenya
477	239170	Ethiopia	523	Akuaaite	Kenya
478	239172	Ethiopia	524	Ngorotom	Kenya
479	239174	Ethiopia	525	Bulukunyang	Kenya
480	239176	Ethiopia	526	Emaritoit	Kenya
481	239186	Ethiopia	527	Longimakuan	Kenya
482	239193	Ethiopia	528	Alekuekaal	Kenya
483	239194	Ethiopia	529	Naliba	Kenya
484	239195	Ethiopia	530	Naseger Nyang	Kenya
485	239197	Ethiopia	531	AurienKori	Kenya
486	239202	Ethiopia	532	Lorengen	Kenya
487	239219	Ethiopia	533	Akitir Nyang	Kenya
488	239230	Ethiopia	534	Namesek	Kenya
489	239231	Ethiopia	535	Nakualem	Kenya
490	239232	Ethiopia	536	Ex-nyasiongo	Kenya
491	239236	Ethiopia	537	Ex-kuria	Kenya
492	239243	Ethiopia	538	Gakuru	Kenya
493	239244	Ethiopia	539	Ondigo	Kenya
494	239246	Ethiopia	540	Jowi Jawomo	Kenya
495	Napese	Kenya	541	Nyaimbo	Kenya
496	Napete Nyang	Kenya	542	Migogo	Kenya
497	Nangori Nyang	Kenya	543	Andiwo	Kenya
498	Lokabalabalati	Kenya	544	Gopari	Kenya
499	Nalireng	Kenya	545	Obamo	Kenya
500	Loponoikal	Kenya	546	Jowi Jawomo	Kenya
501	Ikorinaite	Kenya	547	Obamo	Kenya
502	Naliba	Kenya	548	Gopari	Kenya
503	Torukaee	Kenya	549	Nyaimbo	Kenya
504	Ekosim Ekori	Kenya	550	Migogo	Kenya
551	Gopari	Kenya	597	Ex-Kidera-2	Kenya
552	Kadero	Kenya	598	Ex-Kidera-1	Kenya
553	Jowi Jawomo	Kenya	599	Khumba	Kenya
554	Ex-Tabora	Kenya	600	Kizungu	Kenya
555	Jowi Jawomo	Kenya	601	Kitaita	Kenya
556	Gopari	Kenya	602	Ex-kuria/kisii?	Kenya
557	Migogo	Kenya	603	Muhawi	Kenya
558	Boke	Kenya	604	Kibiriti-B	Kenya
559	Andiwo	Kenya	605	Bishe	Kenya
560	Ex-kuria	Kenya	606	Kibiriti-Ngoma	Kenya

561	Ex-kuria	Kenya	607	Kibiriti-A	Kenya
562	Andiwo	Kenya	608	Kijuju-B	Kenya
563	Boke	Kenya	609	Kijuju-A	Kenya
564	Jowi Jawomo	Kenya	610	Ex-Busia	Kenya
565	Obamo	Kenya	611	Kambio(White)	Kenya
566	Nyaguage	Kenya	612	Kajerere	Kenya
567	Jowi Jawomo	Kenya	613	Megewa	Kenya
568	Andiwo	Kenya	614	Kambio(pigmented)	Kenya
569	Nyakosore	Kenya	615	Ex-malindi-A	Kenya
570	Andiwo	Kenya	616	Ex malindi-B	Kenya
571	Apundo	Kenya	617	Kimiiru(brown)	Kenya
572	Ochuti Rachar	Kenya	618	Kimiiru(brown)	Kenya
573	Nyachong Rawo	Kenya	619	Kimiiru(brown)	Kenya
574	Ochuti Makawar	Kenya	620	Kimiiru(light brown)	Kenya
575	Nolusu	Kenya	621	Kimiiru(light brown)	Kenya
576	Ex-Maiyakalo	Kenya	622	Kimiiru(brown)	Kenya
577	Namadete	Kenya	623	Kimiiru(light brown)	Kenya
578	Unyimbo	Kenya	624	Kimiiru(White)	Kenya
579	Nakhabado	Kenya	625	Kimiiru(light brown)	Kenya
580	Maebeye	Kenya	626	Munthano	Kenya
581	Nangalama	Kenya	627	Mubeeta Mutune	Kenya
582	Sabina-C	Kenya	628	Munthano	Kenya
583	Opunde	Kenya	629	Kaguru?	Kenya
584	Namayeye	Kenya	630	Munthano	Kenya
585	Namonibili	Kenya	631	Munthano	Kenya
586	Olusu	Kenya	632	Kaguru?	Kenya
587	Sabina-L	Kenya	633	Muthigwa	Kenya
588	Nyabuluri	Kenya	634	Mnka	Kenya
589	Nakhalori	Kenya	635	Mnka	Kenya
590	Nyayo	Kenya	636	Ex-iriga	Kenya
591	Nagugu	Kenya	637	Mnka	Kenya
592	Adala	Kenya	638	Mnka	Kenya
593	Oleuro	Kenya	639	Kimiiru(light brown)	Kenya
594	Ushalak	Kenya	640	Wagana(white)	Kenya
595	Ex-Nambale	Kenya	641	Wagana	Kenya
596	Karingan	Kenya	642	NO_ID	Kenya
643	Wagana	Kenya	689	N6	Rwanda
644	Muchuuri	Kenya	690	BM2	Rwanda
645	S 47	Kenya	691	Muhimpundu	Rwanda
646	N-8	Kenya	692	BM3	Rwanda
647	Susa	Kenya	693	N7	Rwanda
648	Damoga	Kenya	694	N8	Rwanda
649	IS25567	Kenya	695	BM4	Rwanda
650	L-5	Kenya	696	Muhimpun	Rwanda
651	URUKARAZA	Kenya	697	N9	Rwanda
652	LIVOYWA	Kenya	698	BM5	Rwanda
653	GAHUNDA	Kenya	699	N10	Rwanda
654	N-12	Kenya	700	N11	Rwanda

655	IS9119	Kenya	701	BM6	Rwanda
656	GATARANGA	Kenya	702	BM7	Rwanda
657	N-15	Kenya	703	N12	Rwanda
658	IS25542	Kenya	704	N13	Rwanda
659	ICSV1156BF	Kenya	705	N14	Rwanda
660	MUL	Kenya	706	Urukaraza	Rwanda
661	F6YQ212X	Kenya	707	N15	Rwanda
662	SDSH9003	Kenya	708	Ikinyaruka	Rwanda
663	KAT369XF6YQ212	Kenya	709	Gahunda	Rwanda
664	GADAM	Kenya	710	N18	Rwanda
665	MARIMANTI-CO1110	Kenya	711	BM29	Rwanda
666	F6YQ212XICSUIII(IN)	Kenya	712	N20	Rwanda
667	IS3697XICSVIII(IN)	Kenya	713	Amatega	Rwanda
668	SDS1848-3(ICSEL)6	Kenya	714	SVR96	Rwanda
669	SEREDOXESSUTI	Kenya	715	BM10	Rwanda
670	SERENAXESSUTI	Kenya	716	Cyatanombe/Nyakinama	Rwanda
671	KAK8474	Kenya	717	N21	Rwanda
672	TESO 6	Kenya	718	BM11	Rwanda
673	E525HRXESSUTI	Kenya	719	Abaresha	Rwanda
674	E525HRXESSUTI	Kenya	720	N22	Rwanda
675	KAK13	Kenya	721	N23	Rwanda
676	E525HRXESSUTI	Kenya	722	Gikoma	Rwanda
677	SERENAXESSUTI	Kenya	723	BM12	Rwanda
678	E525HRXESSUTI	Kenya	724	BR20	Rwanda
679	KAK8540	Kenya	725	Amabyiga	Rwanda
680	E525HRXESSUTI	Kenya	726	BM14	Rwanda
681	SERENAXIKHUMBA	Kenya	727	BM15	Rwanda
682	KAK2809	Kenya	728	Susa	Rwanda
683	PLOT 71	Kenya	729	BM17	Rwanda
684	N2	Rwanda	730	Gataraga	Rwanda
685	N3	Rwanda	731	BM20	Rwanda
686	BM1	Rwanda	732	BM21	Rwanda
687	Imbundi	Rwanda	733	BM22	Rwanda
688	N4	Rwanda	734	Mangoli	Rwanda
735	BM33	Rwanda	781	BM34	Rwanda
736	Sokanya	Rwanda	782	BM18	Rwanda
737	BM26	Rwanda	783	PI 569020	Sudan
738	Cyatanombe/Cyeru	Rwanda	784	HSD 2813	Sudan
739	BM29	Rwanda	785	HSD 3462	Sudan
740	BM30	Rwanda	786	PI 569909	Sudan
741	BM31	Rwanda	787	HSD 3471	Sudan
742	BM32	Rwanda	788	HSD 3449	Sudan
743	BM33	Rwanda	789	HSD 5315	Sudan
744	Ndamaga	Rwanda	790	HSD 2814	Sudan
745	N19	Rwanda	791	PI 569859	Sudan
746	MB35	Rwanda	792	HSD 3506	Sudan

747	MB36	Rwanda	793	HSD 5120	Sudan
748	MB39	Rwanda	794	HSD 5311	Sudan
749	Nyundo original	Rwanda	795	HSD 5211	Sudan
750	Muhimpundu Kinigi	Rwanda	796	PI 569866	Sudan
751	Gicamunkoni Nkuri	Rwanda	797	HSD 3458	Sudan
752	Gahunda	Rwanda	798	HSD 5411	Sudan
753	Muhimpundu Mukingo	Rwanda	799	HSD 4122	Sudan
754	Urukaraza	Rwanda	800	HSD 5389	Sudan
755	MB27	Rwanda	801	HSD 5349	Sudan
756	IS9200	Rwanda	802	PI 217837	Sudan
757	Igitukura	Rwanda	803	PI 569424	Sudan
758	Mbo S48	Rwanda	804	HSD 5371	Sudan
759	Amamikazi	Rwanda	805	PI 569862	Sudan
760	Bayashinyike	Rwanda	806	HSD 4735	Sudan
761	IS9203	Rwanda	807	PI 569910	Sudan
762	Buhuli	Rwanda	808	HSD 4132	Sudan
763	S80Nyirabuhuli	Rwanda	809	HSD 4958	Sudan
764	Ruheng S82	Rwanda	810	PI 569890	Sudan
765	S84	Rwanda	811	HSD 4950	Sudan
766	S85Nyirabuhuri	Rwanda	812	HSD 5204	Sudan
767	Byumba S86Nyirarunyogwe	Rwanda	813	PI 569380	Sudan
768	S87	Rwanda	814	PI 569452	Sudan
769	S88Nyirabuhuri	Rwanda	815	HSD 2795	Sudan
770	Rouge S89	Rwanda	816	HSD 2785	Sudan
771	S90	Rwanda	817	PI 569290	Sudan
772	S91Mbagara	Rwanda	818	HSD 2795	Sudan
773	S92Rugogoma	Rwanda	819	PI 569414	Sudan
774	S94Mbagara rouge	Rwanda	820	PI 270504	Sudan
775	S95	Rwanda	821	HSD 5112	Sudan
776	S96	Rwanda	822	HSD 5257	Sudan
777	IS9119	Rwanda	823	PI 569004	Sudan
778	IS2201 (R)	Rwanda	824	HSD 5302	Sudan
779	IS9202	Rwanda	825	HSD 3465	Sudan
780	MB4	Rwanda	826	HSD 5365	Sudan
827	HSD 5376	Sudan	873	HSD 3480	Sudan
828	HSD 2780	Sudan	874	HSD 3479	Sudan
829	HSD 5410	Sudan	875	HSD 5266	Sudan
830	HSD 3461	Sudan	876	HSD 4245	Sudan
831	HSD 2808	Sudan	877	HSD 3491	Sudan
832	PI 569172	Sudan	878	HSD 2818	Sudan
833	HSD 4101	Sudan	879	PI 569910	Sudan
834	HSD 5119	Sudan	880	HSD 2809	Sudan
835	PI 569062	Sudan	881	HSD 5114	Sudan
836	HSD 4722	Sudan	882	HSD 2800	Sudan
837	HSD 2790	Sudan	883	PI 569438	Sudan
838	HSD 5353	Sudan	884	HSD 5300	Sudan

839	PI 569890	Sudan	885	PI 569240	Sudan
840	HSD 5124	Sudan	886	HSD 3446	Sudan
841	HSD 4115	Sudan	887	PI 569857	Sudan
842	HSD 3469	Sudan	888	HSD 5271	Sudan
843	HSD 3505	Sudan	889	PI 569079	Sudan
844	PI 569905	Sudan	890	HSD 4201	Sudan
845	HSD 2799	Sudan	891	PI 217833	Sudan
846	HSD 5399	Sudan	892	HSD 2778	Sudan
847	HSD 3541	Sudan	893	HSD 5325	Sudan
848	HSD 5204	Sudan	894	HSD 4244	Sudan
849	PI 217727	Sudan	895	HSD 3449	Sudan
850	HSD 4120	Sudan	896	HSD 5216	Sudan
851	HSD 3473	Sudan	897	HSD 4209	Sudan
852	HSD 3499	Sudan	898	HSD 5189	Sudan
853	PI 570498	Sudan	899	HSD 4040	Sudan
854	HSD 2805	Sudan	900	PI 569403	Sudan
855	HSD 4111	Sudan	901	HSD 5121	Sudan
856	PI 217881	Sudan	902	PI 569855	Sudan
857	PI 569100	Sudan	903	PI 569139	Sudan
858	HSD 5377	Sudan	904	PI	Sudan
859	HSD 5194	Sudan	905	PI569028	Sudan
860	PI 569121	Sudan	906	PI569300	Sudan
861	HSD 5184	Sudan	907	MiloYellow	Sudan
862	HSD 5320	Sudan	908	PI563327	Sudan
863	PI 217799	Sudan	909	PI563298	Sudan
864	PI 569860	Sudan	910	PI568996	Sudan
865	PI 569199	Sudan	911	PI569299	Sudan
866	HSD 5218	Sudan	912	Gadamblia	Sudan
867	PI 569864	Sudan	913	HSD2779	Sudan
868	HSD 4179	Sudan	914	Ajeb Seido	Sudan
869	PI 569222	Sudan	915	PI570947	Sudan
870	HSD 2822	Sudan	916	PI569911	Sudan
871	HSD 4151	Sudan	917	PI570505	Sudan
872	PI 569230	Sudan	918	PI568990	Sudan
919	Eriana	Sudan	965	PI217841	Sudan
920	PI569164	Sudan	966	PI569808	Sudan
921	PI563304	Sudan	967	PI563321	Sudan
922	PI563312	Sudan	968	HSD5120	Sudan
923	PI569394	Sudan	969	PI569408	Sudan
924	HSD5275	Sudan	970	SRN	Sudan
925	Red Mugod	Sudan	971	PI217760	Sudan
926	PI569339	Sudan	972	PI569854	Sudan
927	PI217798	Sudan	973	PI570518	Sudan
928	PI569891	Sudan	974	PI569811	Sudan
929	GEW-37-13	Sudan	975	HSD5276	Sudan
930	PI569907	Sudan	976	PI217797	Sudan
931	Tall Milo	Sudan	977	PI217674	Sudan
932	PI569280	Sudan	978	Dabar Habashi	Sudan

933	TSS1	Sudan	979	PI563297	Sudan
934	Abu Shai	Sudan	980	PI569805	Sudan
935	PI563324	Sudan	981	PI569879	Sudan
936	PI569861	Sudan	982	SD1	Sudan
937	Arfa Gadamak	Sudan	983	PI563309	Sudan
938	PI570948	Sudan	984	PI569201	Sudan
939	PI570506	Sudan	985	PI563322	Sudan
940	Wad Akar	Sudan	986	PI563302	Sudan
941	PI568990	Sudan	987	PI569867	Sudan
942	PI569878	Sudan	988	PI569451	Sudan
943	Serina	Sudan	989	PI568998	Sudan
944	PI563323	Sudan	990	Feterita Arafa	Sudan
945	HSD3456	Sudan	991	GA 06/01	Uganda
946	PI569390	Sudan	992	GA 06/15	Uganda
947	PI569344	Sudan	993	GA 06/29	Uganda
948	HSD1529	Sudan	994	GA 06/42	Uganda
949	PI217784	Sudan	995	GA 06/54	Uganda
950	PI563314	Sudan	996	GA 06/72	Uganda
951	PI568994	Sudan	997	GA 06/92	Uganda
952	Regan	Sudan	998	MR-19	Uganda
953	PI569071	Sudan	999	HLS 03/17	Uganda
954	PI570453	Sudan	1000	HLS 03/053	Uganda
955	Dabar Zirazira	Sudan	1001	GA 06/97	Uganda
956	PI569925	Sudan	1002	HLS 03/059	Uganda
957	PI568989	Sudan	1003	GA 06/07	Uganda
958	PI568276	Sudan	1004	GA 06/21	Uganda
959	PI563311	Sudan	1005	GA 06/36	Uganda
960	PI563308	Sudan	1006	GA 06/48	Uganda
961	PI563313	Sudan	1007	GA 06/61	Uganda
962	PI569858	Sudan	1008	GA 06/62	Uganda
963	PI563306	Sudan	1009	GA 06/80	Uganda
964	PI569898	Sudan	1010	GA 06/106	Uganda
1011	LULU D	Uganda	1057	HLS 03/021	Uganda
1012	HLS 03/24	Uganda	1058	IS 25434	Uganda
1013	HLS 03/050	Uganda	1059	GA 06/04	Uganda
1014	HLS 03/047	Uganda	1060	GA 06/18	Uganda
1015	GA 06/02	Uganda	1061	GA 06/33	Uganda
1016	GA 06/16	Uganda	1062	GA 06/45	Uganda
1017	GA 06/43	Uganda	1063	GA 06/58	Uganda
1018	GA 06/55	Uganda	1064	GA 06/53	Uganda
1019	GA 06/73	Uganda	1065	GA 06/70	Uganda
1020	GA 06/93	Uganda	1066	GA 06/84	Uganda
1021	MR-20	Uganda	1067	MR-14	Uganda
1022	HLS 03/018	Uganda	1068	HLS 03/011	Uganda
1023	SERENA	Uganda	1069	HLS 03/044	Uganda
1024	HLS 03/025	Uganda	1070	GA 06/56	Uganda
1025	HLS 03/056	Uganda	1071	GA 06/12	Uganda
1026	GA 06/09	Uganda	1072	GA 06/26	Uganda

1027	GA 06/22	Uganda	1073	GA 06/27	Uganda
1028	GA 06/37	Uganda	1074	GA 06/40	Uganda
1029	GA 06/49	Uganda	1075	GA 06/46	Uganda
1030	GA 06/50	Uganda	1076	GA 06/59	Uganda
1031	GA 06/63	Uganda	1077	GA 06/77	Uganda
1032	GA 06/82	Uganda	1078	GA 06/96	Uganda
1033	GA 05/001	Uganda	1079	GA 06/103	Uganda
1034	AF 28	Uganda	1080	SEREDO	Uganda
1035	HLS 03/019	Uganda	1081	HLS 03/022	Uganda
1036	HLS 03/020	Uganda	1082	GA 06/68	Uganda
1037	HLS 03/012	Uganda	1083	GA 06/05	Uganda
1038	GA 06/03	Uganda	1084	GA 06/19	Uganda
1039	GA 06/17	Uganda	1085	GA 06/28	Uganda
1040	GA 06/31	Uganda	1086	GA 06/41	Uganda
1041	GA 06/57	Uganda	1087	GA 06/53	Uganda
1042	GA 06/94	Uganda	1088	GA 06/71	Uganda
1043	GA 06/83	Uganda	1089	GA 06/85	Uganda
1044	MR-11	Uganda	1090	MR-17	Uganda
1045	HLS 03/058	Uganda	1091	HLS 03/016	Uganda
1046	GA 06/25	Uganda	1092	HLS 03/045	Uganda
1047	GA 06/10	Uganda	1093	GA 06/64	Uganda
1048	GA 06/23	Uganda	1094	GA 06/13	Uganda
1049	GA 06/38	Uganda	1095	GA 06/14	Uganda
1050	GA 06/39	Uganda	1096	GA 06/20	Uganda
1051	GA 06/51	Uganda	1097	GA 06/35	Uganda
1052	GA 06/66	Uganda	1098	GA 06/47	Uganda
1053	GA 06/75	Uganda	1099	GA 06/60	Uganda
1054	GA 06/95	Uganda	1100	GA 06/78	Uganda
1055	SEKEDO	Uganda	1101	GA 06/79	Uganda
1056	EPURIPUR	Uganda	1102	GA 06/104	Uganda
1103	LULU T	Uganda	1149	TZA46	Tanzania
1104	HLS 03/023	Uganda	1150	TZA47	Tanzania
1105	HLS 03/ 040	Uganda	1151	TZA48	Tanzania
1106	GA 06/06	Uganda	1152	TZA49	Tanzania
1107	E36-1	Uganda	1153	TZA50	Tanzania
1108	B35	Uganda	1154	TZA51	Tanzania
1109	TZA1	Tanzania	1155	TZA52	Tanzania
1110	TZA2	Tanzania	1156	TZA53	Tanzania
1111	TZA3	Tanzania	1157	TZA54	Tanzania
1112	TZA4	Tanzania	1158	TZA55	Tanzania
1113	TZA8	Tanzania	1159	TZA56	Tanzania
1114	TZA9	Tanzania	1160	TZA57	Tanzania
1115	TZA10	Tanzania	1161	TZA58	Tanzania
1116	TZA11	Tanzania	1162	TZA59	Tanzania
1117	TZA12	Tanzania	1163	TZA60	Tanzania
1118	TZA13	Tanzania	1164	TZA61	Tanzania
1119	TZA14	Tanzania	1165	TZA62	Tanzania
1120	TZA15	Tanzania	1166	TZA63	Tanzania

1121	TZA16	Tanzania	1167	TZA64	Tanzania
1122	TZA17	Tanzania	1168	TZA65	Tanzania
1123	TZA18	Tanzania	1169	TZA66	Tanzania
1124	TZA19	Tanzania	1170	TZA67	Tanzania
1125	TZA20	Tanzania	1171	TZA68	Tanzania
1126	TZA21	Tanzania	1172	TZA69	Tanzania
1127	TZA22	Tanzania	1173	TZA70	Tanzania
1128	TZA23	Tanzania	1174	TZA71	Tanzania
1129	TZA24	Tanzania	1175	TZA72	Tanzania
1130	TZA27	Tanzania	1176	TZA73	Tanzania
1131	TZA28	Tanzania	1177	TZA74	Tanzania
1132	TZA29	Tanzania	1178	TZA75	Tanzania
1133	TZA30	Tanzania	1179	TZA76	Tanzania
1134	TZA31	Tanzania	1180	TZA77	Tanzania
1135	TZA32	Tanzania	1181	TZA78	Tanzania
1136	TZA33	Tanzania	1182	TZA79	Tanzania
1137	TZA34	Tanzania	1183	TZA80	Tanzania
1138	TZA35	Tanzania	1184	TZA81	Tanzania
1139	TZA36	Tanzania	1185	TZA82	Tanzania
1140	TZA37	Tanzania	1186	TZA83	Tanzania
1141	TZA38	Tanzania	1187	TZA84	Tanzania
1142	TZA39	Tanzania	1188	TZA85	Tanzania
1143	TZA40	Tanzania	1189	TZA86	Tanzania
1144	TZA41	Tanzania	1190	TZA87	Tanzania
1145	TZA42	Tanzania	1191	TZA88	Tanzania
1146	TZA43	Tanzania	1192	TZA89	Tanzania
1147	TZA44	Tanzania	1193	TZA90	Tanzania
1148	TZA45	Tanzania	1194	TZA91	Tanzania
1195	TZA92	Tanzania	1241	TZA140	Tanzania
1196	TZA93	Tanzania	1242	TZA141	Tanzania
1197	TZA94	Tanzania	1243	TZA142	Tanzania
1198	TZA95	Tanzania	1244	TZA143	Tanzania
1199	TZA96	Tanzania	1245	TZA144	Tanzania
1200	TZA97	Tanzania	1246	TZA145	Tanzania
1201	TZA98	Tanzania	1247	TZA146	Tanzania
1202	TZA99	Tanzania	1248	TZA147	Tanzania
1203	TZA100	Tanzania	1249	TZA148	Tanzania
1204	TZA101	Tanzania	1250	TZA149	Tanzania
1205	TZA102	Tanzania	1251	TZA150	Tanzania
1206	TZA104	Tanzania	1252	TZA151	Tanzania
1207	TZA105	Tanzania	1253	TZA152	Tanzania
1208	TZA107	Tanzania	1254	TZA153	Tanzania
1209	TZA108	Tanzania	1255	TZA154	Tanzania
1210	TZA109	Tanzania	1256	TZA155	Tanzania
1211	TZA110	Tanzania	1257	TZA156	Tanzania
1212	TZA111	Tanzania	1258	TZA157	Tanzania
1213	TZA112	Tanzania	1259	TZA158	Tanzania
1214	TZA113	Tanzania	1260	TZA159	Tanzania

1215	TZA114	Tanzania	1261	TZA160	Tanzania
1216	TZA115	Tanzania	1262	TZA161	Tanzania
1217	TZA116	Tanzania	1263	TZA162	Tanzania
1218	TZA117	Tanzania	1264	TZA163	Tanzania
1219	TZA118	Tanzania	1265	TZA164	Tanzania
1220	TZA119	Tanzania	1266	TZA165	Tanzania
1221	TZA120	Tanzania	1267	TZA166	Tanzania
1222	TZA121	Tanzania	1268	TZA167	Tanzania
1223	TZA122	Tanzania	1269	TZA168	Tanzania
1224	TZA123	Tanzania	1270	TZA169	Tanzania
1225	TZA124	Tanzania	1271	TZA170	Tanzania
1226	TZA125	Tanzania	1272	TZA171	Tanzania
1227	TZA126	Tanzania	1273	TZA172	Tanzania
1228	TZA127	Tanzania	1274	TZA173	Tanzania
1229	TZA128	Tanzania	1275	TZA174	Tanzania
1230	TZA129	Tanzania	1276	TZA175	Tanzania
1231	TZA130	Tanzania	1277	TZA176	Tanzania
1232	TZA131	Tanzania	1278	TZA177	Tanzania
1233	TZA132	Tanzania	1279	TZA179	Tanzania
1234	TZA133	Tanzania	1280	TZA180	Tanzania
1235	TZA134	Tanzania	1281	TZA181	Tanzania
1236	TZA135	Tanzania	1282	TZA182	Tanzania
1237	TZA136	Tanzania	1283	TZA183	Tanzania
1238	TZA137	Tanzania	1284	TZA184	Tanzania
1239	TZA138	Tanzania	1285	TZA185	Tanzania
1240	TZA139	Tanzania	1286	TZA186	Tanzania
1287	TZA187	Tanzania			
1288	TZA189	Tanzania			
1289	TZA190	Tanzania			
1290	TZA191	Tanzania			
1291	TZA192	Tanzania			
1292	TZA193	Tanzania			
1293	TZA194	Tanzania			
1294	TZA195	Tanzania			
1295	TZA198	Tanzania			
1296	TZA199	Tanzania			
1297	TZA200	Tanzania			

Appendix 2 Private allele richness probability differences

Country	Burundi	Eritrea	Ethiopia	Kenya	Rwanda	Sudan	Uganda
Burundi		NS	* * *	*	NS	* * *	*
Eritrea	0.087		NS	NS	*	NS	NS
Ethiopia	<0.01	0.293		NS	*	NS	NS
Kenya	0.018	0.411	0.110		NS	*	NS
Rwanda	0.648	0.048	0.004	0.104		* * *	*
Sudan	<0.01	0.108	0.983	0.040	<0.001		NS
Uganda	0.005	0.556	0.499	0.774	0.012	0.405	

NS=Not significant, *=Significant, * * *=Highly significant.

Appendix 3 Gene diversity probability differences

Country	Burundi	Eritrea	Ethiopia	Kenya	Rwanda	Sudan	Uganda
Burundi		*	* * *	* * *	* * *	* * *	* *
Eritrea	0.033		* *	* * *	* * *	* * *	NS
Ethiopia	<0.001	0.009		NS	* * *	NS	*
Kenya	<0.001	< 0.001	0.124		* * *	* * *	NS
Rwanda	<0.001	< 0.001	< 0.001	< 0.001		* * *	* * *
Sudan	<0.001	< 0.001	0.111	< 0.001	<0.001		* * *
Uganda	0.003	0.907	0.080	0.395	<0.001	< 0.001	

NS=Not significant, *=Significant, * * *=Highly significant.

Appendix 4 Allelic richness probability differences

Country	Burundi	Eritrea	Ethiopia	Kenya	Rwanda	Sudan	Uganda
Burundi		NS	* * *	*	* * *	* * *	NS
Eritrea	0.302		*	NS	* * *	* * *	NS
Ethiopia	<0.001	0.027		NS	* * *	*	*
Kenya	0.003	0.171	0.634		* * *	* * *	NS
Rwanda	<0.001	<0.001	<0.001	<0.001		* * *	* * *
Sudan	<0.001	<0.001	0.004	<0.001	<0.001		* * *
Uganda	0.427	0.897	0.012	0.16	<0.001	<0.001	

NS=Not significant, *=Significant, * * *=Highly significant.