GENETIC DIVERSITY OF THE MAIZE GERMPLASM CONSERVED AT THE NATIONAL GENE BANK OF KENYA

by

Isaac Nguthiru Kamweru

A thesis submitted to the faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Master of Science in Plant and Soil Sciences

Fall 2018

GENETIC DIVERSITY OF THE MAIZE GERMPLASM CONSERVED AT THE NATIONAL GENE BANK OF KENYA

by

Isaac Nguthiru Kamweru

Approved: _____

Randall J. Wisser, Ph.D.
Professor in charge of thesis on behalf of the Advisory Committee

Approved: _____

Erik H. Ervin, Ph.D.
Chair of the Department of Plant and Soil Sciences

Approved: _____

Mark Rieger, Ph.D.
Dean of the College of Agriculture and Natural Resources

Approved: _____

Douglas J. Doren, Ph.D.
Interim Vice Provost for Graduate and Professional Education

ACKNOWLEDGEMENTS

My deep appreciation goes to Jim and Marcia Borel for their generous gift to the University of Delaware, Borel Global Fellows Program. The funding received and the unwavering support is very much appreciated. I thank the management of the Alliance for a Green Revolution in Africa (AGRA) and the Kenya Agricultural and Livestock Research Organization (KALRO) for their administrative support.

I will forever remain indebted and thankful to Dr. Randall J. Wisser for his scholarly advice, research directions and insightful discussions. I gratefully acknowledge members of my thesis committee; Drs. Carl J. Schmidt, Nicole M. Donofrio and Murenga G. Mwimali for their intellectual contributions and overall guidance. I wish also to place on record my sincere gratitude to the University of Delaware fraternity, particularly faculty and staff members from the College of Agriculture and Natural Resources for sharing their expertise, knowledge and valuable academic support.

Special thanks go to my wonderful lab colleagues especially Michael Dumas and Scott Davis for their technical support and cooperation. Working with Teclemariam, Heather, Susan, Todd, Wax, Benta, Lydia, Kamau, Ndung'u, enhanced my research skills. I very much appreciate your time and explanations!

DEDICATION

To my dear wife Enid Kawira and son Theo Wagura with more gratitude and affection than I can well put down here.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AGRA: Alliance for a Green Revolution in Africa

CIMMYT: *Centro Internacional de Mejoramento de Maiz y Trigo*

DNA: Deoxyribonucleic Acid

FAO: Food and Agriculture Organization

GATK: Genome Analysis Tool Kit

GBS: Genotyping by Sequence

IFAD: International Fund for Agricultural Development

KALRO: Kenya Agricultural and Livestock Research Organization

MAF: Minor Allele Frequency

MLND: Maize Lethal Necrosis Disease

NGBK: National Gene Bank of Kenya

PCR: Polymerase Chain Reaction

PHATE: Potential of Heat-diffusion for Affinity based Transition Embedding

SNP: Single Nucleotide Polymorphism

ABSTRACT

Maize (*corn*) is the third most important crop after wheat and rice globally. In Kenya, its per capita consumption is estimated at 103 kilograms per year and its inadequacy due to biotic and abiotic stresses puts the nation's food security at risk. Maize breeders in effort to develop newer lines and hybrids combinations with better yield have been consistently inter-crossing a few elite lines. Farmers have increasingly replaced landraces with a few high yielding varieties. The conclusive overview is that the genetic base of the maize breeding pool is constricting, rendering the maize germplasm vulnerable to future threats such as extirpating crop pests and diseases as well as the adverse effects of a deteriorating global crop environment. The raw material for developing newer varieties that are better yielding and climate resilient includes gene bank accessions. Exploration of this collection for utilization in maize improvement however is impeded by failures in accession documentation, agronomic evaluation and characterization of genetic variability. The objectives of this study were to i) characterize the genetic diversity of the gene bank maize collection ii) investigate level of genetic relatedness among accessions. In the current study, 768 samples were genotyped via high-throughput sequencing of restriction fragments (GBS). Results show that for a period of 16 years, the proportion of polymorphic loci remained constant, alleles at minor frequency decreased significantly (p-value<.001) and heterozygosity increased slightly by 0.1%. Proportion of heterozygotes and alleles at minor frequency were found to be

highly correlated (r^2=-79, P=0.006), indicating that reduction of alleles at minor frequency caused an increase in the proportion of heterozygotes. Pairwise shared allele distance for most gene bank accessions (≥95%) fell between 0.20 and 0.30. Only 130 out of 147,696 comparisons (0.08%) had a shared allele distance <0.1 which indicated low levels of redundancy and high genetic distances among nearly all pairwise comparisons (99.9%). Potential of Heat-diffusion for Affinity-based Transition Embedding was used to capture local connections between SNP markers and reveal two clear genetic sub-divisions from admixed ancestry, which supported the hypothesis of a genetically narrow based maize germplasm due to the founder effect.

Keywords: Maize | Genetic diversity | Single Nucleotide Polymorphisms | Genotyping by Sequencing | Heterozygosity | Minor Allele Frequency | Shared allele distance | PHATE.

Chapter 1

INTRODUCTION

1.1 Maize Agriculture in Kenya

Sub-Saharan Africa (SSA) accounts for 21% of the global maize per capita consumption, but produces less than 7% of the total harvest (Ranum et al., 2014). Within the distribution of maize yields across the globe, averaged at the sub-continental scale, SSA is an extreme low-value outlier with yields of approximately 1.5 tons per hectare against the global mean of 5 tons per hectare (Shiferaw et al., 2011). A unique "Green Revolution" in SSA with maize as its most visible expression seem to have stalled as evidenced by a slight increase in yield of 1.4% over the last 50 years (FAO, IFAD, & WFP, 2015). On-farm production constraints, drought, insect pests, and diseases have been attributed to the low yield increases while climate change is projected to exacerbate the problem by causing frequent crop failure, increased rural poverty and food insecurity (Demombynes and Kiringai, 2011). The agricultural sector in Kenya constitutes about 27% of the gross domestic product, 60% of the export earnings and has been referred to as the backbone of the economy (Government of Kenya, 2010); previous studies estimated that a1% growth rate in the agricultural sector corresponds to 1.6% growth in the overall economy (Economic Review of Agriculture, 2015).

In terms of food security, maize is the most relevant staple for 96% of Kenya's 43 million population, contributing an estimated 68% of their daily per capita cereal consumption (Schroeder et al., 2013). Despite efforts to promote productivity enhancing technologies among smallholder farmers who dominate maize agriculture in Kenya, total maize output lags behind the ever-increasing consumer demand due to the rapid population growth and depleting resources required for increased production. The transformation of the country into a net maize grain importer and the rising import bill has caught the attention of government policy planners, agricultural economists, and plant breeders.

Although calls for a unique "Green Revolution" - with maize as its most visible expression - are gaining momentum in Kenya, use of fertilizer and use of improved seeds among smallholder farmers remains below the recommended threshold (De Groote et al., 2005). According to Hassan et al (1998), most maize varieties grown by farmers are on average 23 years old (year since release) partly due to the increased use of recycled farm seeds, low adoption of new varieties and the disjoint between local demand and supply of farmer preferred hybrids. To bring the country to food self-sufficiency, the government of Kenya initiated institutional and legislative reforms as well as strategies for revitalizing growth in the agricultural sector in a 10 year plan launched in 2004 and entrenched in the Kenya VISION 2030 blueprint (Government of Kenya, 2007). The critical need of harnessing and disseminating agricultural technological innovations to smallholder farmers with a view to increasing yield for key food staples in Kenya is currently a priority research agenda. The maize collection preserved at the national gene bank of Kenya is a critical natural resource at the center of yield improvement agenda.

Apparently, its vast potential has not been extensively utilized in maize breeding activities due to lack of reliable characterization information (Wambugu and Muthamia, 2009). Although the genetic contribution of traditional farmer's populations to the accumulation of favorable alleles in today's elite germplasm is evident, little is known about the underlying genomics of the maize germplasm that is now preserved ex situ.

The national gene bank of Kenya has made considerable efforts to characterize agro-morphological traits of about 37% of the total plant germplasm it holds, but no single accession has been characterized using molecular markers (Genetics Resources Research Institute, 2015). Agro-morphological trait characterization is known to be time and resource intense effort and yields low quality information because expression of some morphological markers are subject to environmental influence which leads to inconsistent descriptors (Govindaraj et al., 2015).

Gene banks play a fundamental role in diversifying maize cropping systems. Their importance in plant breeding can only be compared to that of a library or a collection of sources of valuable information. A library with unclassified resources that lack titles or keywords or any other relevant information will rarely be used and its value will be insignificant. Passport information, evaluation and characterization data is of great importance to plant breeders seeking to improve yield and other agronomic traits of importance.

This study therefore seeks to describe the genetic diversity of maize germplasm preserved at the National Gene Bank of Kenya using molecular markers and provide

baseline genetic information for utilization in maize breeding activities and germplasm curatorial management.

1.2 Problem Statement

Maize plays a critical role in Kenya's food equation and its inadequacy puts the nation's food supply at risk (Omoyo et al., 2015). With maize being the most relevant staple for millions of Kenyans, a major share of subsistence food requirement to meet the ever increasing consumer demand and surplus for the market has to come from the maize crop. In an effort to develop new lines and hybrids with better yield, maize breeders have been consistently inter-crossing a few elite lines (Prassana, 2012).

This system of mating however, constricts the genetic base of the maize breeding pool and slows down maize breeders' response to future threats such as extirpating crop pathogens and other environmental constraints associated with maize production (Reynolds et al., 2015). Farmers who traditionally maintained highly diverse landraces in situ have increasingly replaced primitive stocks and populations with highly bred maize cultivars (Ndiso et al., 2013). Maintaining and utilizing a genetically narrow based breeding pool increases the vulnerability of the maize germplasm to biological invasions and environmental stresses (Keneni et al., 2012).

Although gene bank collections provides a reliable source of unique germplasm, these collections largely remain underutilized even in other countries (Nass et al., 1993). Maize breeders for example cite the intensive breeding efforts required to bring unimproved germplasm of unknown pedigree and phenotypic adaptation to the desired levels of agronomic performance (Shimelis et al., 2012). Under-utilization of gene bank

accessions has also been attributed to lack of quality, reliable characterization information in addition to infrequent regeneration of accessions that leads to low seed viability levels and availability of limited sample sizes for distribution (Crossa et al., 1994). Assessment of the state of global plant genetic resources revealed the need to regenerate more than 50% of all gene bank accessions (FAO, 1996).

Although there are close to 50,000 accessions of plant germplasm are conserved at the National Gene Bank of Kenya, only about 8% of this extensive collection has been distributed to national research programs, universities, commercial companies and farmers in the last 15 years (Genetic Resources Research Institute, 2015).

The Kenya national strategy on genetic resources (2016-2020) proposed the application of molecular markers to characterize gene bank collection with a view to provide knowledge on the molecular basis of plant processes which can benefit conventional breeding (Genetic Resources Research Institute, 2015). Current advances in plant molecular biology are expanding the scope of maize genetic resources exploration and utilization. Molecular markers are increasingly being used to screen gene bank accessions in efforts to resolve taxonomic relationships (Hajibabaei et al., 2007), assess genetic diversity within and among accessions, eliminate redundant accessions(Treuren and Hintum, 2003) and identify genetically dissimilar accessions in pre-breeding programs (Shimelis et al., 2012).

1.3 Justification

Tropical maize germplasm is phenotypically and genetically diverse, and is endowed with genetic variability that can be utilized in the generation of private value to the farmer as well as public value to the society (Goodman, 2005). According to Romay et al (2013), unused natural variation in germplasm collections can be exploited to lift yield barriers and accelerate targeted plant breeding. Warburton et al (2002), cited unique germplasm as an important source of finding putatively useful variation for improving maize grain yield and other agronomic traits of importance. According to Munyiri et al (2010), most gene bank populations have distinct phenotypic and genetic identities and could be a critical source of useful genes 'lost' during the development of elite germplasm.

Although there are close to 50,000 maize accessions globally, 27,000 of which are conserved at CIMMYT headquarters in Mexico (Global Crop Diversity Trust, 2007), less than 1% of this global collection is used in maize breeding activities for example in the United States (Hoisington et al., 1999). Despite the fact that the genetic diversity found in United States maize germplasm is a small fraction of the total in global maize collection, yield gains has been attributed to maize genetic improvement across large interconnected breeding networks (Rubenstein et al., 2005).

Cultivated maize is therefore not endowed with all the desirable genes and gene complexes essential in developmental, reproductive and adaptation pathways and this raises concerns about its vulnerability to biotic and abiotic stresses especially in the face of a changing climate (Laborda et al., 2005). Gowda et al (2015), screened a diverse

global panel of close to 70,000 genotypes for resistance to Maize Lethal Necrotic Disease (MLND) in Kenya and reported that greater than 95% of this germplasm was susceptible to MLND.

The nested association mapping population created by crossing B73 reference line to a set of diverse founder lines has been successfully used for example to dissect the genetic architecture of complex traits in maize for example mapping genes associated to flowering time (Buckler et al., 2009), southern blight leaf disease resistance (Kump et al., 2011) and tolerance to insect herbivores (Meihls et al., 2012).

Predictive breeding holds potential value in benefiting future maize breeding activities and programs by downsizing time and other resources expended on maize genetic improvement and enhancement (Shimelis et al., 2012). Molecular characterization provides important decision making input information for use in germplasm base broadening and conservation programs. Breeders are continuously searching for novel sources of tolerance and resistance traits to mitigate future threats of biological invasions and environmental challenges and genetic information can be used to accelerate evaluation of accessions and introgression of traits into locally adapted germplasm.

1.4 Research Objectives

The research objectives of this study were to;

(i) To characterize the genetic diversity of the maize germplasm conserved at the national gene bank of Kenya based on a spatiotemporal scale.

(ii) To determine genetic distances among and between gene bank maize accessions

Redundant accessions increase the cost of ex situ management but their contribution in terms of useful variation is negligible. Reduction of gene bank redundancies will provide maize breeders with an objective overview of gene bank genetic diversity available for utilization. Elimination of duplicated accessions will help rationalize curatorial management.

(iii) To investigate the historical origin of gene bank accessions using SNP marker technology.

1.5 Research Background and Hypothesis

Prior to the advent of molecular characterization techniques, gene bank collections were mainly characterized based on their agro-morphological traits. Molecular characterization techniques are currently being used to provide a reliable and a more accurate analysis of large collections preserved in national seed repositories compared to agro-morphological based techniques. Genetic fingerprinting facilitates the exploration of maize genetic resources for use in maize improvement and optimizes chances of discovering genetically unique accessions for use as parents in maize hybrid development which is the key objective in maize breeding programs. While integration of both marker technology and detailed phenotypic characterization will stimulate an increase in the use of gene bank accessions to achieve its hoped-for impact, access to quality genomic information is expected to provide an invaluable resource for transforming and rationalizing curatorial management. This study posit that;

(i) Genetic diversity of the in situ maize populations, conserved at the National Gene Bank of Kenya is geographically and temporally structured.

(iii) In situ populations is genetically narrow based due to the founder effect. The basis of all varietal hybrids grown by farmers in Kenya traces back to two founder parents.

Chapter 2

LITERATURE REVIEW

2.1 The State of Plant Genetic Resources in Kenya

About 14 million species and 6.6 billion populations are estimated to compose the existing biodiversity of our planet, but close to 27,000 species are lost every year (Hughes et al., 1997). Kenya has an estimated number of 35,000 known species of flora and fauna held by a range of habitats (Government of Kenya, 2001). The continuous and significant loss of genetic diversity in most crops eliminates species, genes and biological traits which potentially might alter the functioning of local ecosystems (Cardinale et al., 2012). This has stimulated a growing interest in the conservation of the wealth in plant biodiversity. Genetic erosion, attributed to population pressure on agricultural land, wild habitats degradation and climate change poses a severe threat to plant genetic resources which forms the biological basis of food and nutritional security (Kiambi et al., 2010). Country report on the state of plant genetic resources for food and agriculture in Kenya for instance highlighted the disappearance of "*Githigu*" (Kikuyu) and "*Makondo*" (Luhya) traditional maize varieties (Wambugu and Muthamia, 2009).

"*Githigu*" was a high altitude landrace with strong purple kernel pigmentation and was believed to have been brought into the country by missionaries before World War I from Peru (Harrison, 1970).To fully integrate maize genetic resources in food and agricultural development, national and regional gene banks have been established *ex situ* where maize seeds are preserved outside their area of growth and *in situ*, where maize plants and their wild relatives are maintained in natural preserves to facilitate the natural processes of evolution and adaptation (Institute of Economic Affairs, 2011a).

2.2 The National Gene Bank of Kenya (NGBK)

Kenya has a rich plant diversity emanating from its geographical placement and varying topography. NGBK holds approximately 50,000 accessions of plant germplasm representing about 165 families, 893 genera and 2000 species (Genetics Resources Research Institute, 2015). Germplasm repository for the top 15 plant species indicates that there is a fair representation of all the major food and forage crop species (Table 1).

Table 1: Gene bank accessions for the top crop species

Crop name	Scientific name	Number of accessions	
		1996	2006
Sorghum	*Sorghum bicolor*	5,333	5,649
Common bean	*Phaseolus vulgaris*	3,305	3,428
Finger millet	*Eleusine coracana*	2,823	2,852
Sesame	*Sesamum indicum*	1,671	1,677
Maize	*Zea mays*	1,423	1,792
Rice	*Oryza sativa*	712	1,004

(Source: Country report on the state of plant genetic resources for food and agriculture, 2009).

About 60% of Kenya's plant heritage was assembled through in-country collection missions and the rest from at least 137 countries via donations, transfer, and exchange programs (Government of Kenya, 2013a). Only a fraction of the total collection (37%) has been characterized for agro-morphological traits due to financial constraints. No accession(s) has been characterized using molecular marker technology (Genetic Resources Research Institute, 2015).

The collection of maize at the national seed bank of Kenya comprises of traditional farmers' populations or landrace in addition inbred lines deposited by breeders. It is important to note that the National Gene Bank of Kenya also holds duplicated maize accessions from other countries such as Germany, France and Somalia for safety and as part of its global mandate. Subsets of this maize collection (from in-country collection missions) has been characterized for agro-morphological traits to facilitate processing of diverse seed requests by the maize breeding network.

Limited technological capacity and low quality of passport and evaluation data are some of the identified constraints hampering detailed characterization, increased use of gene bank maize accessions (Mutegi et al., 2005). Traditional farmer populations are generally perceived to be repositories of unique allelic variation because they are genotypes resulting from years of local selection by the maize farming communities. The need to conserve plant biodiversity for future utilization in crop improvement led to the establishment of the National Gene Bank of Kenya in 1988. To implement a systematic countrywide collection and conservation of maize genetic resources, the NGBK worked in collaboration with regional maize research programs and academic institutions (Nakhauka, 2009).

2.2.1 In-situ Management of Maize Accessions

In situ collections of the maize accessions was made directly from farmers' field during the harvest season, their "granaries" and local markets based on certain distinctive morphotypes. Aware that all maize accessions could not be collected for preservation due to space limitation of storage cells at the gene bank, priority was given to local materials/landraces and products of domestic breeding from their very beginnings. Efforts were also made to obtain samples that represented a good variation within and between accessions. For each accession, collection site passport data information was recorded and included the longitude, altitude, elevation, source/origin/district and any other important information.

Collected maize seeds were catalogued using unique gene bank accession numbers for long term conservation (base collection) or medium-term (active collection for distribution) conservation. From each accession, samples were drawn and their germination viability tested. Efficient exploration of the maize collection for genetically unique germplasm with putative utilization in maize breeding activities requires a clear understanding of its historical origin, genetic relationships among and within accessions and population subdivisions.

Kamau et al (2017), in "dealing with farmers' ethnolinguistic differences when collecting crop diversity on farm" imply that farmers' social-cultural diversity impact the morphological characteristics of sorghum landraces - the basis of ex situ collection mission. Some sorghum morphotypes were peculiar to some ethnolinguistic groups and

morphological differentiation explained by ethnolinguistic group was higher for landraces than improved varieties.

Even in a similar agro-climatic zones, ethnolinguistic diversity has been shown to influence the spatial genetic diversity of sorghum varieties (Labeyrie et al., 2014a). Collection of crop diversity therefore require an in-depth understanding of the dynamism of the social-organization of the maize farming communities and germplasm management in situ. Traditional farmers' seed selection criteria, their cultural seed exchange networks and their ethnolinguistic diversity has been shown to influence the ecological distribution of maize morphotypes in Mexico (Perales et al., 2005). The impact social-cultural dynamism on in-situ management of plant genetic resources however, remains understudied in Kenya.

2.3 Origin, Evolution and Description of Maize

Maize (*Zea mays* L.) is an annual, monoecious, single-stalk plant that belongs to the grass family Gramineae, subfamily Panicoideae and Maydae tribe (Edwards, 2011). It was domesticated 8,000 years ago in central Mexico from a short, bushy, wild grass species called teosinte according to Hake and Ross-Ibarra, (2015). Typical maize growth requirements includes; 500mm to 1200mm of rainfall, 12^0C soil temperature or greater during germination and 18 ^{0}C to 32 ^{0}C during growth, development to physiological maturity (Mandal, 2014). During life cycle of maize, a large number of diverse proteins function spatially and temporally to confer programmed plant development. Maize is predominantly a cross-pollinating species and according to Shiferaw et al. (2011), it's a

classical genetic model for plant research due to its moderate genome size, out-breeding reproductive system and broad morphological and geographical adaptability.

Natural cross-pollination played a major role in structuring global maize populations while continuous selective interventions by human beings and hybridization has led to isolated evolutionary pools that differentiate into maize races adapted to various agro-climatic regions of the world (Mandal, 2014).

Maize was first introduced at the coastal lowlands of Kenya by the Portuguese traders and Arab explorers during the World War I (Miracle, 1966). The Portuguese settlers grew maize along the Swahili Coast and with the development of the inland slaving routes, maize slowly became a staple in the local diet of the Swahili trade caravans.

Typical Kenyan diet during this period was dominated by millet, sorghum, tubers, and legumes commonly found in traditional food systems. The British conquest led to the establishment of large plantations of white maize because of the premium price the grain fetched in the British starch market (Smale and Jayne, 2003). Colonial maize farmers enjoyed good links with policy makers and successfully lobbied for maize legislation and fiscal policies which weakened African farmers' position in maize agriculture.

The domestic demand for maize grew because local farmers left their farms to work in crop plantations, mines and industrial plants. Colonial landowners used maize as the main form of food source for their African farm laborers and gave maize rations as "in-kind" payments. Grafting of large scale plantations of maize into the local food economy led to the diffusion and transition of maize into a major food crop in Kenya.

The crop became widely cultivated among the locals due to its ease of cultivation and adaptation to a wide range of agro-climatic conditions (Hassan et al., 1998).

According to Smale and Jayne, (2003), "people got used to what they ate" and most people preferred maize over millet and sorghum. Meals prepared my mashing potatoes, green vegetables and boiled maize and beans also became popular. Among the agro-pastoral communities, maize stalks were used to feed livestock while cobs were used as a source of domestic fuel. Although, maize selection for improved yield and disease resistance by individual farmers and isolated farms took place as far back as the 1920's (Harrison, 1970), the government of Kenya established the first maize breeding program at Kitale in 1955 to develop synthetic varieties for small-scale farmers and hybrid varieties for large scale farmers.

Using conventional maize breeding, inbred lines were developed from the well-adapted "Kenya Flat White complex" which was composed of varieties that traces back to from South Africa that originally came from the Southern U.S.A such as the "Hickory King", "White Horse tooth", "White Pearl" and others (Eberhart & Harrison, 1973). Inbred lines were then formed into synthetic type varieties. Kitale synthetic II (KSII) was released in 1961 and became the most popular synthetic variety grown on about half of the large scale farms (Harrison, 1970).

The "Kenya Flat White Complex" had a rather narrow genetic base due to the similarity of origin of its ancestral lines (Harrison, 1970). This limited further extraction of inbred lines for hybrid development. Rockefeller Foundation and USAID facilitated the exchange of germplasm between continents as well as research experience concerning

hybrid genetics, nothing significant had come out of earlier testing of temperate germplasm with Kitale Synthetic II (Gerhart, 1975).

Evaluations of maize germplasm from the Latin America was expected to give significant results due to similarity of Mexico and Colombia agro-climatic conditions to those found in Kenya and the fact that this germplasm came from the center of maize origin. In 1959, the chief maize breeder at Kitale program imported a large collection of maize genotypes from Mexico and Colombia gene banks to "widen the genetic base" of the maize breeding pool (Harrison, 1970).

After pre-selection for late maturing, high altitude and disease resistant types, an outstanding cross was made in 1961 between Kitale synthetic II (KSII) and unimproved Ecuadorian landrace (Ecu. 573; *Montana Ecuatoriano*). This landrace was late maturing, high altitude variety with the longest ears in Ecuador, but lacked the yellow endosperm (Ramirez et al., 1960).

The varietal cross between KSII and Ecu.573 known as H611 was released in 1964 and according to Gerhart (1975), became widely adopted by small and large scale farmers "at rates faster than among farmers in the U.S. Corn Belt during the 1930s-40s". This was due to the effective linkages of research extension and a strong commercial seed enterprise of the Kenya Seed Company (KSC) formed by large scale farmers in 1956.

In 1963, the Kenya Seed Company entered into an agreement with government of Kenya to produce and distribute maize seeds. The former Kenya Agricultural Research Institute (KARI) seed unit controlled the quality of maize seed production while the Ministry of Agriculture was supposed to manage agricultural extension services.

Through better seed marketing, KSC controls about 80% of the formal maize seed in Kenya. Old but popular maize varieties are still grown by farmers across agroecological zones although on smaller shares of maize area. Hassan et al (1998), reported that about 12 maize hybrids were grown by farmers in 1992, but this number has dramatically increased due to the liberalization of the seed sector in the mid 1980's.

According to the national crop variety list, 11 out of the 14 varietal hybrids released between 1964 and 1989 when the gene bank became operational were based on Hybrid 611 (Table 2). This suggest a genetically narrow based maize germplasm that is subject to the maize founder effect.

Table 2 A list of maize varieties released in Kenya between 1964 and 1989.

Variety name	Year Released	Owner/Licensee	Special attributes
H632	1964	KARI/KSC	Large kernels, Dent
H622	1965	KARI/KSC	Large kernels, Dent
H511	1967	KARI/KSC	Medium maturity
KAT CB	1967	KARI/KSC	Early maturing
H512	1970	KARI/KSC	Large kernels
H614D	1986	KARI/KSC	Semi-flint
H611D	1986	KARI/KSC	Frost tolerant
H612D	1986	KARI/KSC	Semi-flint
H613D	1986	KARI/KSC	Semi-flint
PH1 (Pwani hybrid)	1989	KARI/KSC	Drought tolerant

(Source; National Crop Variety List - Kenya)

Figure 1 below traces the history of the maize collection conserved at the gene bank to "Kenya Flat White Complex" made of varieties that traces back to South Africa but originally from Southern USA.

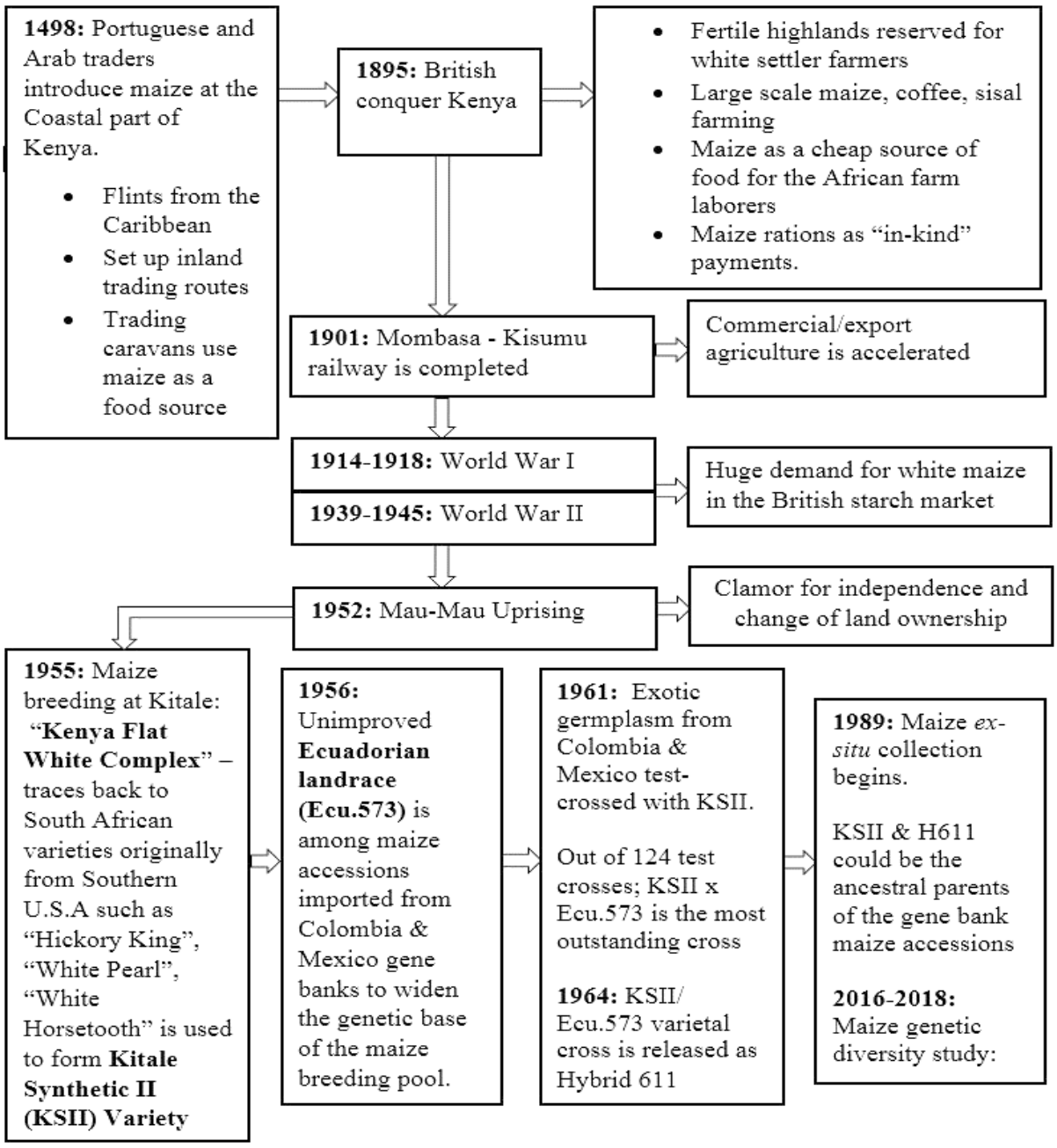


Figure 1: Historical origin of the gene bank maize collection.

2.4 Maize Breeding Programs and Sites

Food crop research institutes at the Kenya Agricultural and Livestock Research Organization are characterized by robust maize breeding programs. Regional maize breeding programs facilitated the collection of the maize germplasm preserved at the national seed bank. Embu maize breeding program was established in 1965 in the central part of Kenya to develop medium-maturing maize varieties. It is located on latitude 0^0 30' S and longitude 37^0 27'E, and 1510 meters above sea level. The program serves an area that receives an average annual rainfall of 1100 mm. EM11-133 and EM 12-210 are some of the important founder lines developed from Embu composites in 1960's.

Katumani agricultural experiment stations was established in 1975 to develop soil and water management technologies that will help improve crop varieties tolerate water stress in dry and semi-arid lands. The center is located on 1^0 35' S and 37^0 32'E and 1600 meters above sea level in an eco-geography with ferrosols soils which are highly weathered and leached. Katumani composite A and B were the first improved maize varieties released in 1966 and 1968 respectively.

Mtwapa maize breeding program is located on the coastal lowland part of Kenya on latitude 3^0 56' S, longitude 39^0 44' E at an elevation of 15 meters above sea level. The program was established in 1955 but released Pwani Hybrid I in 1989, replacing an open pollinated variety called the coast composite. The area has mainly sandy soils and receives an average annual rainfall of approximately 900 mm to 1200 mm.

Kakamega food crop research institute focuses on cereals especially maize, wheat sorghum, and millet. It is located on 0^0 16' N and 37^0 14' E, at an elevation of

1585meters above sea level. The area receives on average an annual rainfall of 1916 mm and has soils that are deep, friable, basaltic loam, fertile and well drained. Inbred lines and hybrids are screened in this station for gray leaf spot, striga- parasitic weed, common rust and northern leaf blight disease. Kakamega maize breeding program has released hybrids such as KH633A, KH634A, KM20077, KM20084 and KM20090 characterized as high yielding and resistant to gray leaf spot and northern leaf blight diseases.

The National Crop Variety List of the Ministry of Agriculture indicates that 338 maize varieties were registered between 1964 and 2017. Multisite collaborations enhances project planning, execution and rapid dissemination of research findings to the smallholder farmer level across the country where maize is grown.

Well designed and properly implemented maize breeding programs across multi-locations with a diversity of agroecological zones ensures the availability of large, diverse samples with sufficient statistical power to detect significant associations and outcomes. A wide genetic base of the breeding pool across maize breeding programs and sites is important as it promotes chances of successful breeding of solutions to new pests and emerging diseases.

Chapter 3

MATERIALS AND METHODS

3.1 Study Area

3.2 Classification of Agroecological Zones in Kenya

Kenya is an equatorial country in East Africa and lies between latitudes 4^0N and 4^0S and longitudes 34^0E and 42^0E. The country's total land mass of approximately 587,000 km^2 is divided into seven agro-ecological zones using a moisture index that is based on the annual rainfall expressed as a percentage of potential of evaporation according to Sombroek et al (1982). An agro-ecological zone refers to a land resource mapping unit, defined in terms of climate, temperature regimes, topography and soil type. Agroecological zones (AEZ) 1, II, and III represents about 12% of the total land area has a moisture index of more than 50% whereas AEZ IV, V, VI, and VII account for an estimated 88% of the total land area and has a moisture index of below 50% (Figure 2). Regions classified as agro-ecological zone I has no direct importance in agricultural production as they are mainly confined to mountains and immediate surroundings (Obiero and Onyando, 2013).

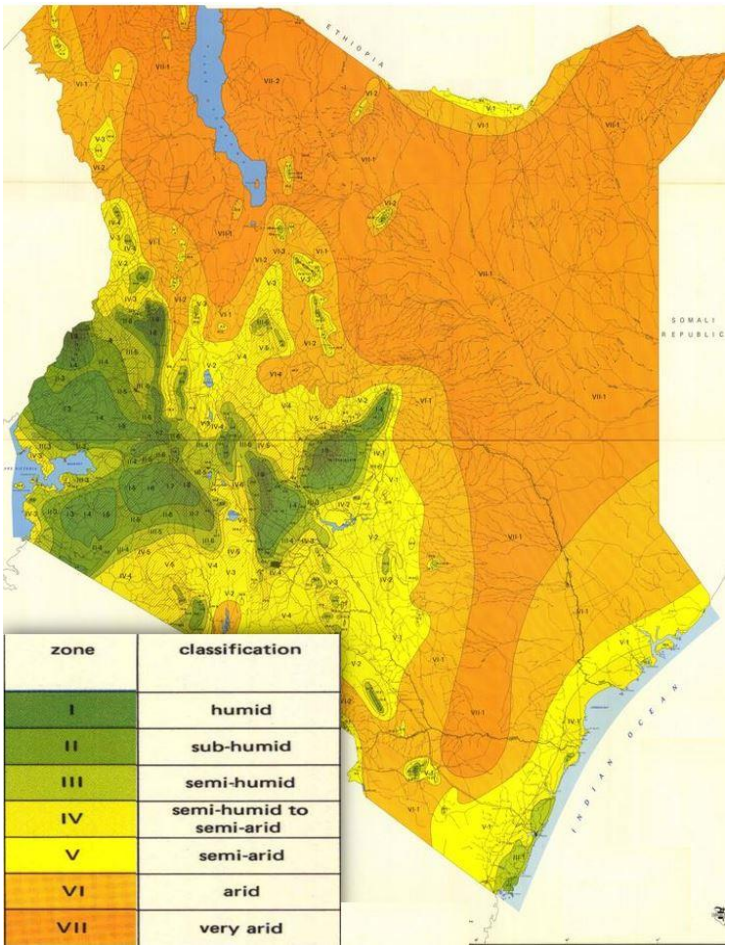
Table 3: Agroecological zones in Kenya

Zone	Moisture Index *R/Eo* (%)	Annual rainfall (mm)	Climatic designation	Risk of maize crop failure (%)	Area (Km^2)
I	>80	1100 - 2700	Humid	0-1	800
II	65 - 80	1000 - 1600	Sub-humid	1-5	53,000
III	50 - 65	800 -1400	Semi-humid	5-10	53,000
IV	40 - 50	600 -1100	Transitional	10-25	48,200
V	25 - 40	450 - 900	Semi-Arid	25-75	300,000
VI	15-25	150 - 350	Arid	75-95	112,000
VII	<15	<150	Very Arid	95-100	

(Source: Sombroek et al., 1982, *R* = Average annual rainfall; *Eo* = Average annual potential evaporation; Transitional = Semi-humid to Semi-arid)

Sub humid agroecological zones (Table 3) are restricted to Kenyan highlands with forests and grasslands and occur between 1,980 and 2,700 meters above sea level. Semi-humid agro-ecological zones occurs at elevations of between 900 and 1,800 meters above sea level and are the most significant in terms of crop-livestock agricultural systems. Transitional agro-ecological zones (semi-humid to semi-arid) receives an annual rainfall of about 500-1000mm. Laikipia, Naivasha and Machakos districts in Kenya as well as Central and Southern Coast Province are some of the regions designated as agroecological zone IV. Agroecological zone V, is semi-arid and occurs at lower elevations and is usually inhabited by low trees and shrubs. Northern Baringo, Turkana, lower Makueni and vast parts of North Eastern province are some of the regions classified as agro-ecological zone. Regions designated as agroecological zone VI are arid and are generally considered as semi-desert. Rainfall in these regions is below 250mm and is quite unreliable. Marsabit, Turkana, Mandera and Wajir districts are some of the

regions that characterize this agro-ecological zone. Regions classified as agroecological zone VII are the driest part of the country and can be illustrated by the Chalbi desert in Marsabit district. Pastoralists who inhabit the salt desert use it as a source of mineral lick for livestock during the rainy season. Figure 2 below shows the agroecological zone map of Kenya



(Source; Sombroek et al., 1982).

Figure 2: Agroecological zone map of Kenya

Arid and semi-arid zones lack fully developed soils and receive an annual average rainfall of less than 250mm. These regions are predominantly inhabited by pastoralist and agro pastoralists and support more than 50% of the country's livestock population. With more than 80% of the country classified as arid to semi-arid, Kenya is one of the water deficient countries in the world with a water area that is 2.3% of the total area. Whereas agricultural productivity can be improved through intensified irrigation, groundwater resources for exploitation are unevenly distributed in time and space and have not been assessed and quantified (Government of Kenya, 2010).

Maize is adapted to a whole range of climatic and ecological extremes and is therefore the most extensively cultivated crop in Kenya grown alongside other subsistence crops like beans, potatoes and bananas. Approximately more than 80% of the country's annual maize production is obtained from semi-humid and transitional agroecological zones. Different varieties maize are grown as determined by the prevailing rainfall and temperature conditions of these ecological zones. However, the risk of crop failure due to increased frequency of dry spells and uneven rainfall distribution is high even in agroecological zones classified as sub-humid, semi-humid and transitional.

3.3 Sampling the Maize Collection

To establish protocols for sampling, tissue processing and exchange from Kenya to the University of Delaware, maize seeds samples were randomly sampled from a short-term maize seed bank located at Katumani experimental station in Machakos district. This collection was mainly composed of inbred lines from previous maize projects coordinated by the KALRO maize program across multiple locations (agroecological

maize breeding sites). Ten maize seeds were randomly sampled from among 107 different accessions. Different traits of interest associated with these accessions included; early maturing, drought tolerance, stem borer resistance and low nitrogen tolerance. The seeds were grown at KALRO biotechnology center and tissues samples were obtained from an emerging or fully extended leaf at V3 stage according to the collar method described by Nielsen (2014). Leaf tissues were collected from eight random plants per accession.

Tissues were frozen and freeze dried before shipping to the University of Delaware via the United States Department of Agriculture, Animal and Plant Health Inspection unit. After a stringent quality control criteria, 191 DNA samples were genotyped by GBS (genotyping by sequencing) at Delaware Biotechnology Institute (DBI) Sequencing and Genotyping Center.

In order to sample the total maize collection (1,782 accessions) conserved at the National Gene Bank of Kenya, passport data information was collated and cross referenced as far as its practical to help implement a stratified sampling strategy. Accessions with missing passport information (collection site longitude, latitude, elevation, district of origin, gene bank accession number, date of collection, amount of seeds available for distribution) were excluded from the study. Passport data was also used to ensure a fair representation of sample from different geographical sources and collection dates.

During 2007 and 2013 collection period, gene bank curatorial and conservation focus was on other food crops in line with a prevailing theme and resources allocation. As

a result, limited or no maize accessions were collected for in situ conservation. Figure 3 shows the total number of maize accessions deposited at the national gene bank between the year 1989 and 2014.

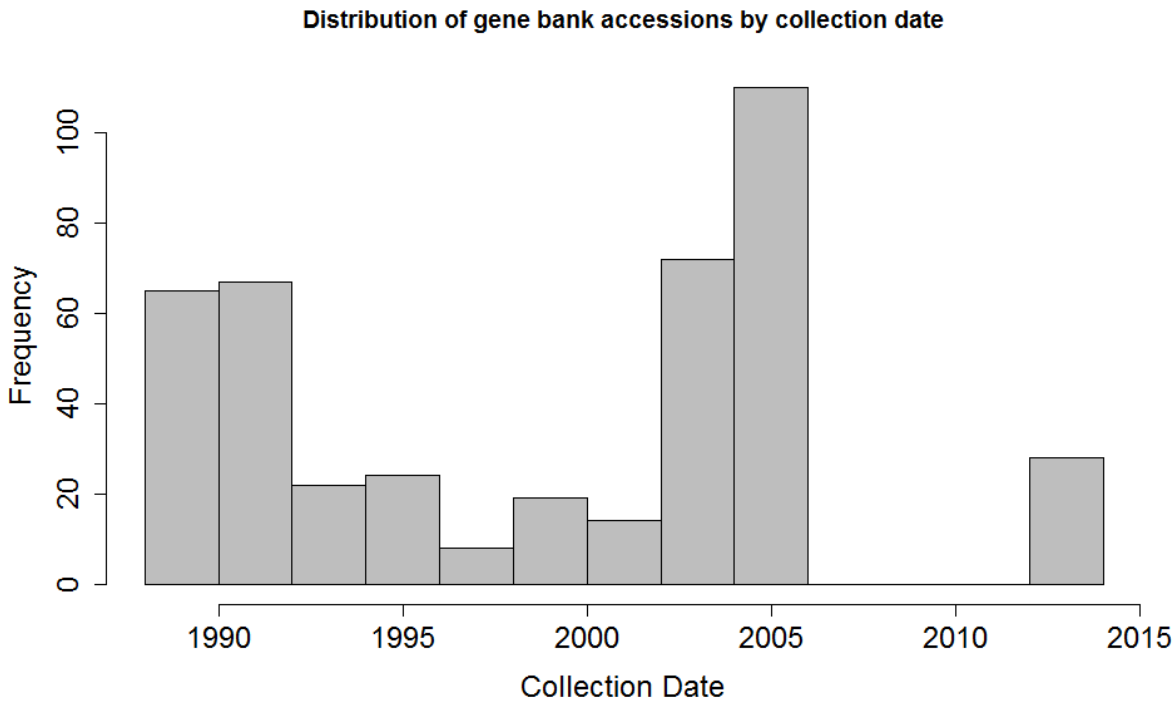


Figure 3: In situ maize collection (1989 - 2014).

The gene bank collection was also stratified based on their geographical sources of origin into six non-overlapping subgroups or strata (Figure 4). This is because regional maize breeding programs served a number of administrative districts with a diversity of agro-climatic conditions. In Central, Coast, Eastern and Western/Rift-Valley/Nyanza former provincial administrations, maize research and breeding activities are undertaken at Embu, Mtwapa, Katumani and Kakamega agricultural experiment stations respectively (Figure 4). Embu station in Central province for example serves eight administrative

districts (Mbeere, Embu, Kirinyaga, Nyeri, Meru North, Meru South, Meru Central and Tharaka) with a total land area of 18,000km^2, a population of over 3.6 million people and approximately 730,000 households. Maize is grown in all districts of former provinces but districts in the former Rift-Valley province are the largest producers of maize followed by districts in the former Nyanza province and Western.

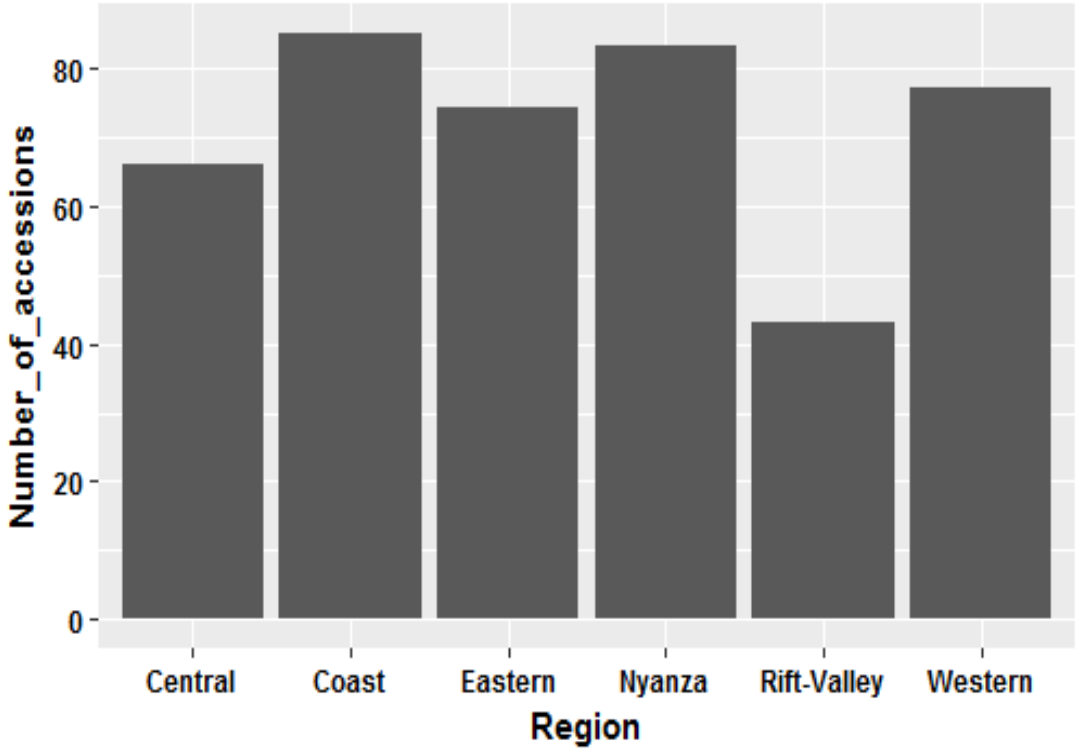


Figure 4: Geographical sources of origin of the gene bank accessions.

Among accessions from each stratum (province), a fixed proportion of entries (~38 accessions) were randomly sampled to avoid giving undue weight to large collections from similar geographical sources. Within each accession, ten seeds were randomly sampled to help assay intra-accession variability. The resulting materials sampled from the gene bank germplasm was mainly composed of populations. These

populations were of historical origin and were associated with traditional farming systems. Although locally adapted, most populations lacked formal improvement across breeding networks as they were products of continuous selection by farmers based on their adaptability to agro-climatic conditions, prolificacy, yield and resistance to biotic and abiotic stresses (Prassana, 2012).

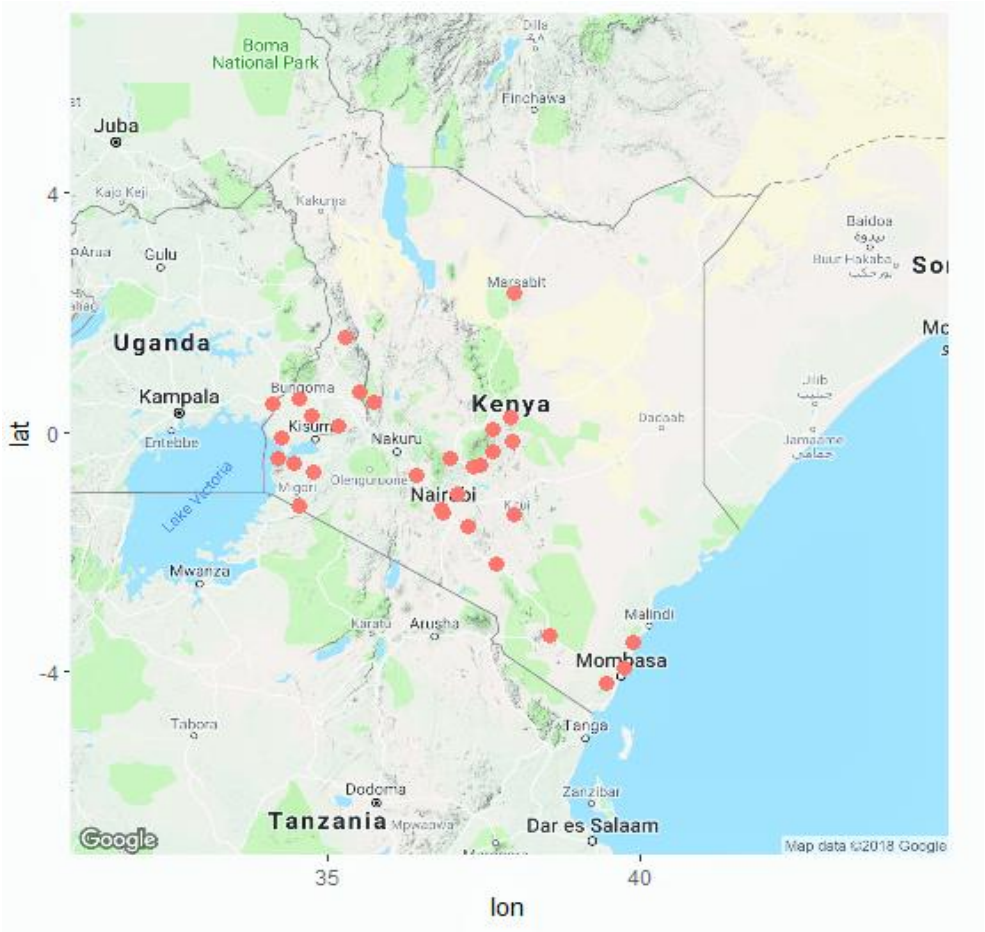


Figure 5: Geographical sources of origin of the study samples.

The collection site information (name of the district, latitude, longitude, elevation) from where each of these samples were collected was used to determine the agroecological zone associated with each accession. It was important to associate the

sampled accessions with their respective agroecological zone because genotypes respond to environmental variations in different ways and analysis of genetic diversity based on the accession agroecological zone of origin was expected to be very informative.

3.4 Plant Material

Wet paper towel germination test was used to determine the viability of the sampled seeds. Maize seedlings were grown under 12 hours of light at 30 ^{0}C and 12 hours of darkness at 24 ^{0}C with a relative humidity of approximately 65%. For every accession, twelve seedlings were then maintained for 21 days and tissues were obtained from eight random plants per accession. Envelopes labelled with a unique study code that was linked to the actual gene bank accession number were used to collect leaf tissues from an emerging or fully extended leaf at V3 stage and following the collar method described by Nielsen (2014).

Sample codes for the study samples were different from the accession code that exists for the particular stock in the gene bank and was developed using the following syntax;

I. BGFXXXXXX-YY; (Where BGF is the abbreviation of Borel Fellows Program and XXXXXX was a five to six digit number representing the source and accession number)
II. YY; represented independent samples from the same accession.

Leaf tissue samples were obtained from eight random seedlings within each accession to help assay intra-accession variability. Leaf tissues collected in Ziploc bags were double bagged before being frozen at - 80 ^{0}C for 36 hours. To freeze dry the leaf

tissues, the lyophilizer was set down to a temperature of -60^{0}C and pulled to a good vacuum (10 microns Hg) before loading the samples. The vacuum was then set at 100 micron Hg and a condenser temperature of -60^{0}C. All samples were freeze dried for 72 hours and thereafter stored in a sealed Ziploc bag at room temperature. Lyophilized samples were taken to the Kenya Plant Health Inspectorate department for phytosanitary inspection before shipping to the University of Delaware via United States Department of Agriculture, Animal and Plant Health Inspection unit.

3.5 Leaf Tissue Processing

Genomic DNA was extracted using an in-house Wisser Laboratory protocol. A standard hole-punch were used to obtain an average of ten leaf punches from each sample. Tissues from each sample were placed in individual wells for DNA extraction using Qiagen DNeasy 96 plant tissue kit and following Biomek 2000 protocol. A negative control (water sample), a blank and one repeatability control (B73 maize reference line sample) were included in each extraction plate. Maize leaf tissues obtained from CML277, Hp301, Mo17, P39, Tx303, and B73 maize founder lines for North America corn breeding programs were also incorporated as study samples for experimental validation. A total number of 1,312 leaf tissue samples were processed.

3.6 DNA Extraction

DNeasy Kit provided a fast and easy way of extracting and purifying DNA from lyophilized maize tissues. Leaf discs of approximately 20mg were obtained from lyophilized tissue and was put into a 2mL centrifuge tube containing a single grinding bead and stored at -80^{0}C for 3-4 hours. Each plate was dipped in liquid nitrogen for 5

minutes. Geno Grinder 2000 machine was set to a 1X rate at 700 strokes per minute for 30 seconds while grinding the samples. The samples were cooled in liquid nitrogen for 5 minutes, and were ground for the second time while switching their positions on the Geno Grinder. Lysis mixture was poured into the full reservoir at the A5 position and Plate 1 was placed at B5 location. The robot transferred 420 µL of lysis buffer to each well of plate 1, the plate was re-capped, shaken for 15 seconds, centrifuged at 1000 rotations per minute (rpm) and incubated at 65^0C for 20 minutes. Plate 1 was then positioned at location A5 and the robot transferred 143 µL of buffer P3 to each well. The plate was re-capped, shaken for 15 seconds, spun for up to 3000 rpm and frozen at -20^0c for 10 minutes. The samples were centrifuged at 4700 rpm for 6 minutes.

A newly labelled Greiner 1mL master block was placed at the B5 position and the robot transferred 400 µL of supernatant to a new labelled plate. 600 µL of AW1 buffer was added, shaken for 15 seconds and spun at 3000 rpm. Plate 1 was placed at the B4 position, the robot transferred 1 mL to filter plate 1 at A4. ~80mL of wash buffer AW2 was then added to the full reservoir at A6. The vacuum pump was turned on and the robot transferred 800 µL of buffer AW2 to filter plate 1. To remove contaminants such as proteins and polysaccharides two wash steps were carried out. Filter plate 1 was withdrawn from vacuum collar, sandwiched with empty elution micro tube plate and spun at 4700 rpm for 17 minutes. Elution buffer was placed in left-hand reservoir at the A5 position, the eluate was transferred after about 12 minutes. Upon incubation for 20 seconds to optimize elution, the robot transferred 35 µL buffer AE to filter plate 1 and another 50 µL. Air pore tape was then placed on filter plate 1 before incubation at room temperature for 4 minutes. Filter plate 1 was sandwiched with PCR plate and centrifuged

at 4700 rpm for 3 minutes. The collection plate was sealed with aluminum tape and stored at -20^{o}C. Biomek platform has functions for serial dilution, shaking, temperature control and DNA extraction procedure was repeated in identical configurations but using plates 2 – 14.

Quant-iT Pico green dsDNA assay kit (Thermo Fischer Scientific, Inc., MA) and Spectra-Max plate reader were used for fluorometric quantification of multiplex libraries. To achieve precise quantifications at low concentrations, the average between two replicates per sample were used to help mitigate pipette inaccuracies and slight variations in concentrations. DNA was normalized to achieve equal concentrations of 200ng Gel electrophoresis was used to assess genomic DNA quality.

3.7 Genotyping by Sequencing (GBS)

In this study, 200ng of purified and normalized DNA was digested using two types of restriction endonucleases, *Ngo*MIV and C*sp*6I simultaneously. These enzymes are produced by certain bacteria and are known to provide defense mechanism against invading viruses by cutting foreign DNA. Restriction enzymes are cutsmart and cleave DNA molecules at near or specific sequence of base pairs. A restriction enzyme master mix was prepared and aliquoted into 1mL well plates, per reaction. *Ngo*MIV and C*sp*6I are type II restriction endonucleases that recognize specific short DNA sequences and carry out endonucleolytic cleavage of DNA to give specific double-stranded fragments with terminal 5'- phosphates. The recognition site is usually a palindromic sequence 4-8 base pairs in length. C*sp*6I restriction enzyme recognizes G^ TAC site while *Ngo*MIV recognizes (G^CCGGC) site and both work best at 37 ^{0}C (Poland et al., 2012).

Restriction associated sequence polymorphism (RASP) barcoded adapters design were developed for ligation Temperature cycle ligation included 300 cycles of 30s at 10 ^{0}C, 30s at 30 ^{0}C and heat inactivation of the ligase at 65^{0}C for 30 minutes (Lund et al., 1996).

Adapters with sample-specific molecular barcodes were ligated with T4 DNA ligase to the ends of the digested DNA to differentiate the samples computationally. This allowed for multiplexing of 192 samples per library. Equal volumes of each ligate was pooled and purified using AMPure beads and following manufacturer's recommended protocol for standard clean-up. BluePippin (Sage Science) was used for size selecting the pooled ligate while target enrichment of the size restricted fragments (350 – 425 base pairs) was carried out through amplification by polymerase chain reaction (PCR) using Phusion High Fidelity Master Mix. Quant-iT PicoGreen® dsDNA assay kit (Thermo Scientific) was used to quantify multiplex libraries before sequencing on Illumina Hi-sequation 2500 platform at the Delaware Biotechnology Institute (DBI) Sequencing and Genotyping Center.

Illumina platform has unlimited sequencing capability as well as optimal configurations and the available optics, fluidics and enzymatics can operate at multiples of a million nucleotides per second. A total of 761 samples including 26 experimental controls from North America maize germplasm were genotyped by GBS. Genotyping by Sequencing is a highly multiplexed system for constructing reduced representation libraries for Illumina next-generation sequencing (Elshire et al., 2011). Multiplexing and use of inexpensive barcoding reduce sample handling as well as polymerase chain reaction and purification steps.

With the current advances in next generation sequencing (NGS) and high throughput (HTP) platforms, a large, diverse population of maize can therefore be genotyped without experimental bottlenecks by manipulating multiplexing and restriction-mediated genomic reduction levels.

3.7.1 GBS Data Pre-processing

Leveraging GBS power for analyzing fundamental genetic variation in maize, bioinformatics tools and computational pipelines were used to process the multi-locus genotype data. Reduced representation "RedRep" computational pipeline (https://github.com/UD-CBCB/RedRep) designed for the analysis of double-digest reduced representation genomic libraries was used to deconvolute barcoded sequences stored in a FASTQ metafile. This metafile contained sample identifiers, barcode sequences and restriction site information.

RedRep pipeline utilized a collection of open source tools, utilities and custom scripted workflow to transform raw data into a format that could easily be interrogated further using genetic variation analysis softwares as well as R program algorithms. Genome Analysis Toolkit (GATK) Haplotype Caller was used to call sequence variants using the hidden markov model likelihood function to estimate posterior probability of allele frequency at each locus (McKenna et al., 2010).

In the current study, all SNP data (761 genotypes by 11,139 variants) were scored using Illumina Bead studio genotyping software. SNPs call rate per sample averaged 91% and a consistent rate was observed for technically duplicated samples. The highest call rate (93%) was obtained for B73 samples. Sequence variation in form of Single

Nucleotide Polymorphisms (SNPs) were stored in a variant call format (VCF) file. A suite of VCF tools were implemented to process the resulting VCF file to provide a high-density SNP map for downstream genetic diversity analysis.

A stringent VCF quality filtering criteria was used to remove a) variants that were absent at in silico digest loci b) variants called at repeat loci (greater than 96%) c) water samples used as controls d) variants with less than 85% call rate e) variants that were not bi-allelic and f) insertions and deletions. All calls that did not meet specific threshold for SNP and genotype properties were removed using the VCF tools. The nucleotides present at a certain marker and individual were then encoded to comply with the formatting required for genetic diversity algorithms and programs.

Chapter 4

RESULTS AND DISCUSSION

4.1 Genetic Diversity Indices

Traditional farmer practices such as cultural seed exchange, agro-climatic variables and geographical factors such as farmland fragmentations are some of the important mechanism that impact the genetic structure of in situ maize populations. According to Rabbi et al (2010), there are higher chances of gene flow among geographically proximate populations. In the current study, the proportion of polymorphic sites, heterozygosity, allele frequencies, shared allele distances, and population sub-divisions indicators were used to characterize the genetic diversity of gene bank accessions on a spatiotemporal scale.

To investigate the extent of genetic relatedness among accessions and across agroecological zones, pairwise comparisons of the shared allele distances among accessions were evaluated. Shift in allelic proportions was also investigated. This is because alleles conferring selective advantages may become fixed while deleterious alleles may become extinct over time due to genetic drift, natural selection or artificial selection. Potential of Heat-diffusion for Affinity based Transition Embedding (PHATE) was used to capture local connections between SNP markers to delineate the population structure (Moon et al., 2017).

4.1.1 Single Nucleotide Polymorphisms

Single Nucleotide Polymorphisms (SNPs) are defined as base substitutions at a single nucleotide position. According to Mammadov et al. (2012), SNPs are the most abundant form of genetic variation in maize and therefore provides a highly informative genotyping assay with multiple sites for investigating genetic variation. SNPs markers have gained tremendous applications and prospects in crop genetic diversity studies due to their pliability for high throughput automation approaches and cost effectiveness.

According to Turakulov & Easteal (2003), the power to delineate population structure even by conventional clustering methods is highly dependent on the density of markers utilized. It is the ability of SNPs to detect population differentiation that Semagn et al (2012), detected three genetic groups that were consistent with pedigree information when they genotyped 450 maize inbred lines developed and widely used in CIMMYT breeding programs in Kenya and Zimbabwe using 1,065 SNP markers.

Masuka et al., (2017) reported separation of lines by pedigree and origin after genotyping 67 selected CIMMYT maize hybrids in East and Southern Africa using 258,038 SNP markers. SNP markers detection and validation has been implemented in the dissection of complex traits such as flowering time in maize (Buckler et al., 2009) as well as genomic prediction of resistance to maize lethal necrotic disease (Gowda et al., 2015). In current study, 11,065 SNP markers were used to genotype 712 samples and were expected to be informative besides providing a good level of genetic resolution to elucidate the structure of the population under study.

Results presented and inferences made in this study are therefore based on the analysis of all SNP data (11,065 SNP markers scored over 712 individuals). Out of 11,139 SNP markers, 74 were identified from scaffolds and could not be anchored to any of the 10 chromosomes. These markers were not informative and were excluded from downstream analysis. SNP markers per chromosome ranged from 1,631 in chromosome 1 to 687 in chromosome 10 and averaged 1,106 SNPs per chromosome (Table 4).

Table 4: SNP marker distribution along the 10 maize chromosomes

Chromosome	SNP Markers	Chromosome	SNP Markers
1	1,631	6	802
2	1,361	7	984
3	1,332	8	936
4	1,083	9	877
5	1,372	10	687
Total	11,139		
Average	1,107		
Unmapped	74		

4.1.2 Polymorphic Loci

A loci was define as polymorphic if the frequency of one of its alleles was less than or equal to 0.95. The proportion of polymorphic loci per individual was quantified across time and space. To determine if this proportion shifted over years of in-situ management by the maize farming communities who lived in different agroecological

zones and ex-situ management by the National Gene bank of Kenya since collection, linear regression and ANOVA tests were implemented. Between 1990 and 2006 the proportion of polymorphic loci ranged from 96% to 98% and averaged 97% over the collection period (Figure 6).

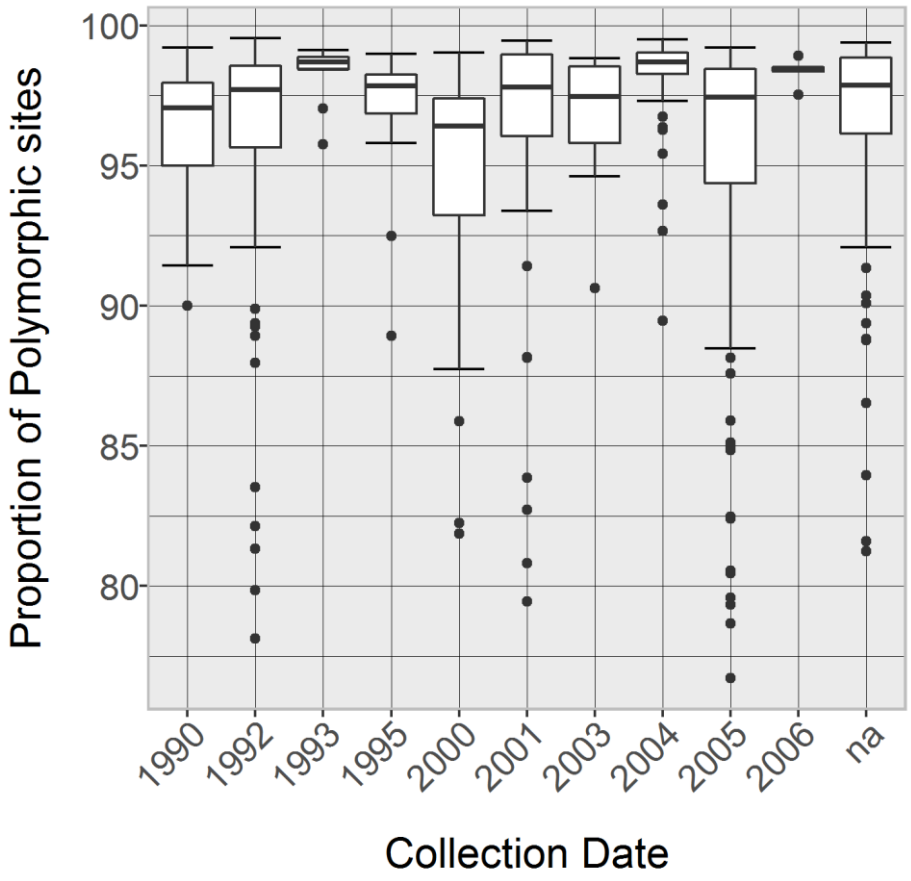


Figure 6. Proportion of the polymorphic loci across time in years

Linear regression analysis was implemented to model the relationship between the proportions of polymorphic loci and collection dates in years and to determine if the two variables were associated.

Results showed that the slope of the linear regression was not significant different from zero (r^2=0.002), P=0.7). Collection dates could only explain 0.002% of the variability of polymorphic loci.

Analysis of variance (ANOVA) is the most commonly used technique for comparing the means of groups ANOVA was implemented to test the null hypothesis that the mean number of polymorphic loci were the same across agroecological groups. It was important to check if the assumptions of normal distribution were met before implementing the test.

A non-significant p-value from the Levene's test indicated that the assumptions for normality and homogeneity of variances were not violated (p-value=0.6). ANOVA results showed that the mean differences were not statistically significant (p-value=0.6). From the results, it can be concluded that there was no gain or reduction on the proportion of polymorphic loci over years of in-situ across agroecological zones or ex-situ management at the National Gene Bank of Kenya. The proportion of polymorphic loci captured at different time points therefore remained the same.

4.1.3 Allele Frequencies

Mechanisms that cause changes in allele frequencies and subsequent departure of allelic proportions from what is expected under Hardy-Weinberg theorem in a population includes natural selection, genetic drift and gene flow (Luikart et al., 1998). Due to these forces, certain genotypes survive, reproduce and pass their alleles to the next generation. The frequency of the favored allele's increases over time and generations and advantageous alleles may become fixed in the population.

Rare alleles are most found in heterozygotes and their frequency may increase over generations to reach high enough frequencies to become common or even fixed. Fixation of one particular allele in a population causes loss of the other and over time heterozygosity decays towards zero. According to Romero et al (2017), allele frequencies can also fluctuate to reach a state of equilibrium – where sub-populations within a meta-population attain allele frequency values that are equal to average frequencies across the meta-population. As populations diverge genetically, allele frequencies change and populations accumulate differences. Over generations, a state of genetic homogeneity may be attained which basically reduces genetic diversity. Of the 11,065 markers detected, 33% had a MAF of less than 0.05 and were considered as rare, while the rest had a MAF of more than 0.2 and were considered as having normal allele frequency (Figure 7).

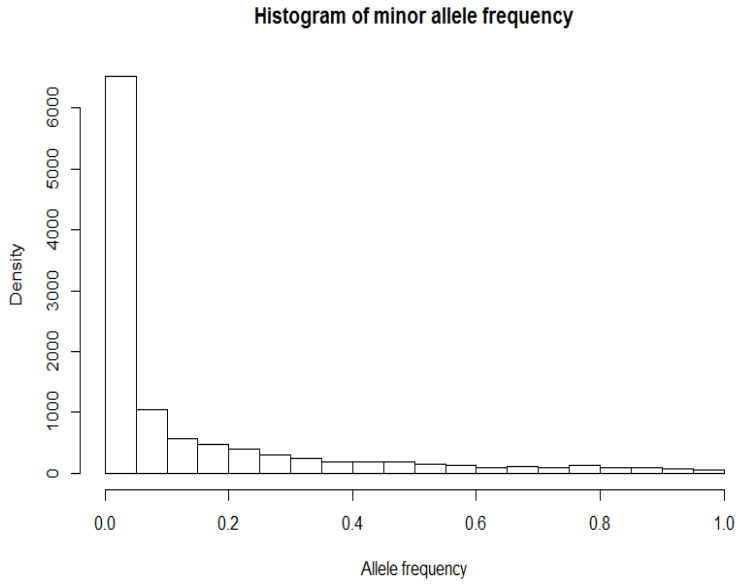


Figure 7: Minor allele frequencies across 11,065 SNP markers.

Figure 8 shows minor allele frequencies distributions among gene bank accessions which is not consistent with the expectations of a random mating population. The density of the distribution was found at lower values than higher values. Minor allele frequencies ranged from 13.2% in BGF179_5 to 33% in BGF141_4 and averaged 16.5% among accessions.

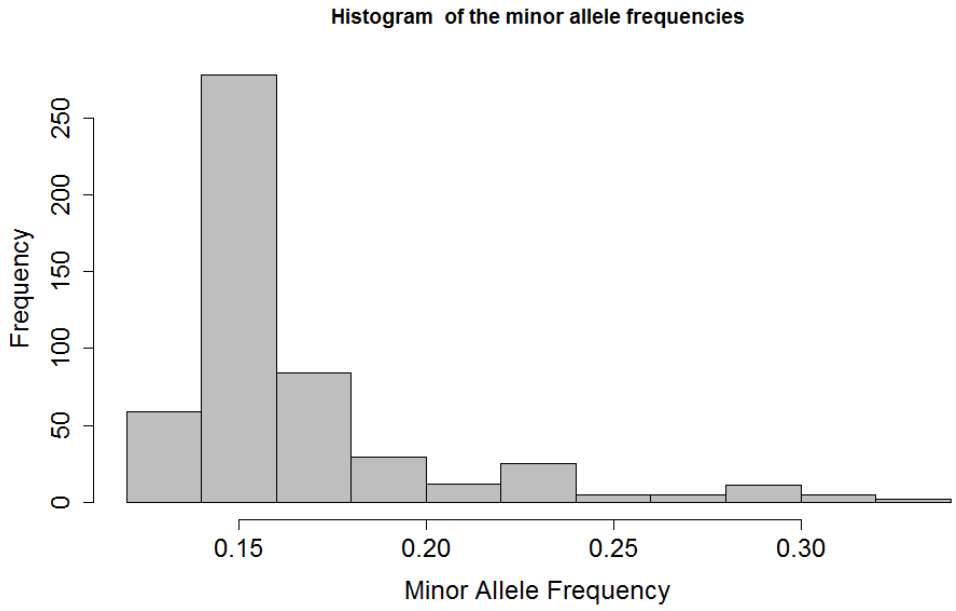


Figure 8. Minor allele frequencies distributions among gene bank accessions.

Minor allele frequencies ranged from 13.6% in 2004 to 14.2% in 1990 and averaged 13.7% over a period 16 years (Figure 9). Results from the linear regression analysis showed that the slope (606) was significantly different from zero (r^2= 0.17, p-value=<.001).

The direction of the slope (-) indicated that the frequency of the minor alleles decreased per unit time in years (Figure 9). R-Squared value indicated that the collection dates explained about 17 % of the variability in the frequency of the minor alleles.

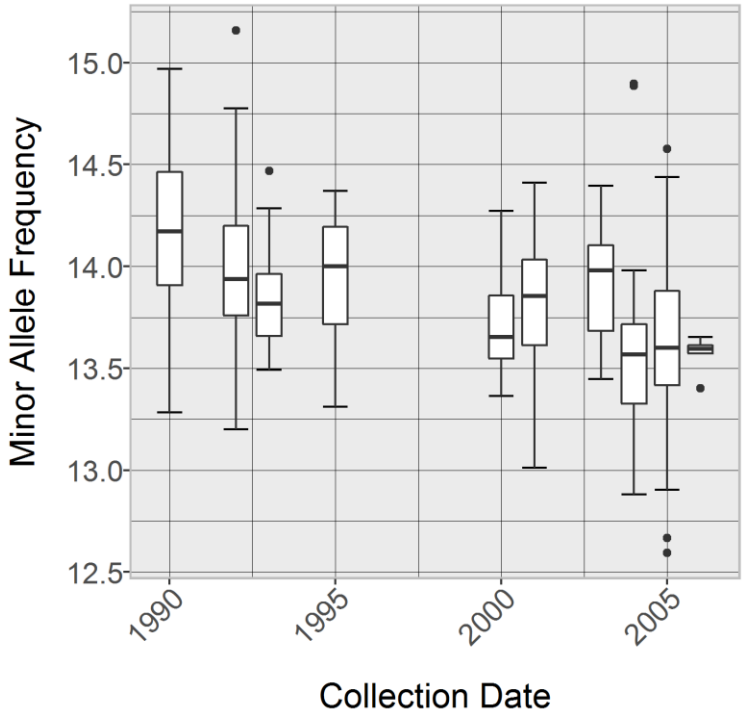


Figure 9. Distribution of the minor allele frequencies across time in years.

According to Luikart et al. (1998), alleles conferring selective advantages may become fixed in a population over time while deleterious alleles may become extinct due to forces such as genetic drift or artificial selection regimes in breeding programs or even unconscious selection by traditional farmers. Genetic drift tend to push smaller demes towards different allele frequencies and allele fixation more quickly than would take place in a larger undifferentiated population. This is particularly a concern in small

threatened populations in which fixation of deleterious alleles can reduce population viability and raise the risk of extinction.

Analysis of variance (ANOVA) was used to evaluate whether the minor allele frequency means were significantly different across agroecological groups (Figure 10). Overall significance test showed there were significant mean differences (p-value<.001).

Tukey Kramer test for multiple comparison was implemented to identify the pair groups whose mean differences were significantly different. Mean differences were statistically significant when AEZ4 group was compared to AEZ2 (p=<.001) and AEZ3 (p=<.001).

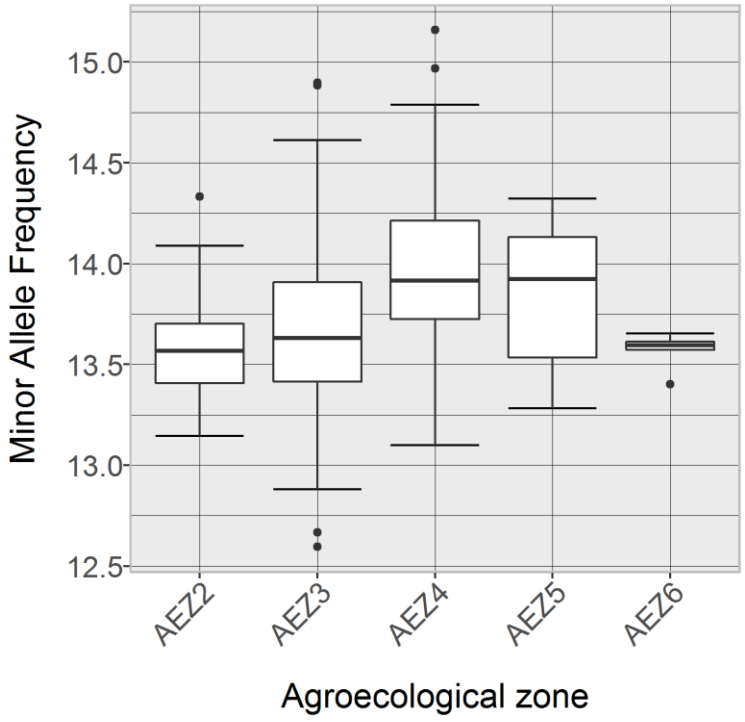


Figure 10. Distribution of the minor allele frequencies across agroecological zones.

The relationship between the proportion of heterozygotes and alleles at minor frequency was found to be highly correlated and significant (correlation coefficient =-0.79, p-value=0.006) (Figure 11). This indicate that an increase in the frequency of alleles at minor frequency caused a decrease in the proportion of heterozygotes and vice versa. (Figure 11). It can be concluded that the frequency of minor alleles decreased over time especially in accessions originating from agroecological zones II, III, and IV.

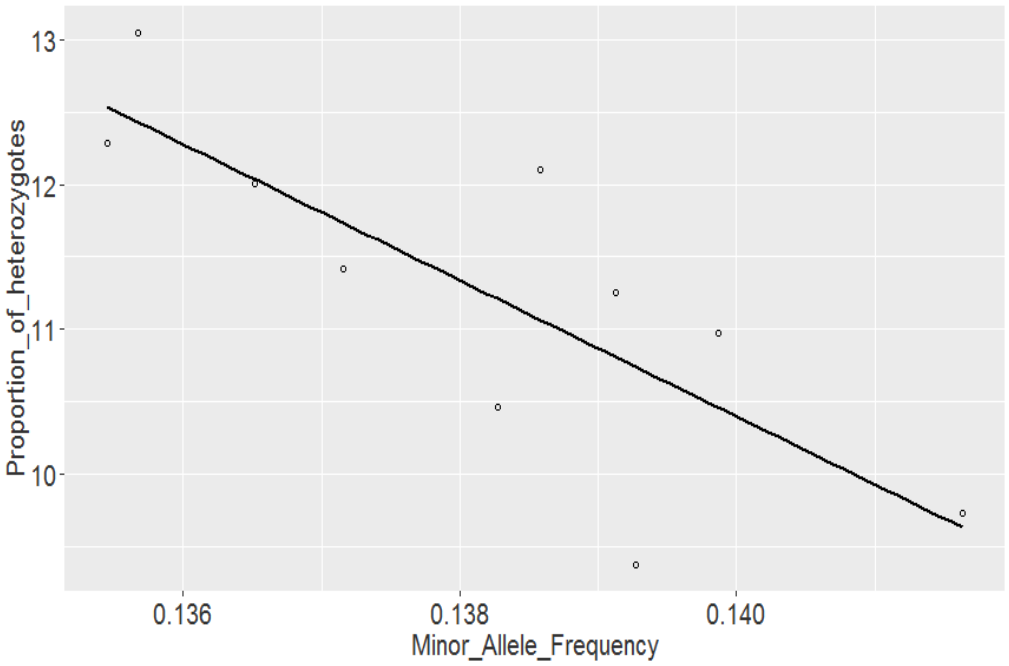


Figure 11. MAF and heterozygosity correlation.

While rare allele can serve as novel sources of unique alleles for use in maize improvement, their extinction may lead to loss of valuable genetic variation for use in future maize breeding. Gilding et al (2013), reported a low frequency allele in drought-adapted sorghum grain associated with increased digestibility of the cereal. Bio-fortification of maize with pro-vitamin A also involved the introgression of rare favorable

alleles into elite maize germplasm from natural maize populations (Harjes et al., 2008). Allelic diversity is therefore important for the long-term response to selection and survival of genotypes.

4.1.4 Heterozygosity

In traditional farming systems, it was common practice to cultivate different crop populations within the same field - which unconsciously favored gene flow (Barnaud et al., 2009). Recurrent gene flow may lead cultivation and maintenance of new genetic combinations, loss of genetic integrity, and extinction of smaller populations (Papa and Gepts, 2003). Whereas inter-mating of genetically differentiated populations increased heterozygosity, selection exerted by farmers contributes to inbreeding and consequently loss or reduction of heterozygosity.

Heterozygosity is proportional to the amount of genetic variance at a locus and is closely related to the polymorphic nature of each locus. The proportion of heterozygotes ranged from 9.7% in 1990 to 13.1% in 2006 and averaged 10.3% over the collection period.

Results from the linear regression analysis revealed a small magnitude of the slope (0.005) that was significantly different from zero (p-value=<.001). The direction of the slope positive and indicated that the proportion of heterozygotes increased per unit time. R-Squared value of 0.05 showed that collection dates (time in years) explained 0.05 % of the variability observed in the proportion of heterozygotes.

From these results, it can be concluded that the proportion of heterozygotes increased by a very small proportion (0.1%) over 16 years of in-situ management of maize accessions by traditional farmers (Figure 12). This contributes to an overall gain in genetic diversity.

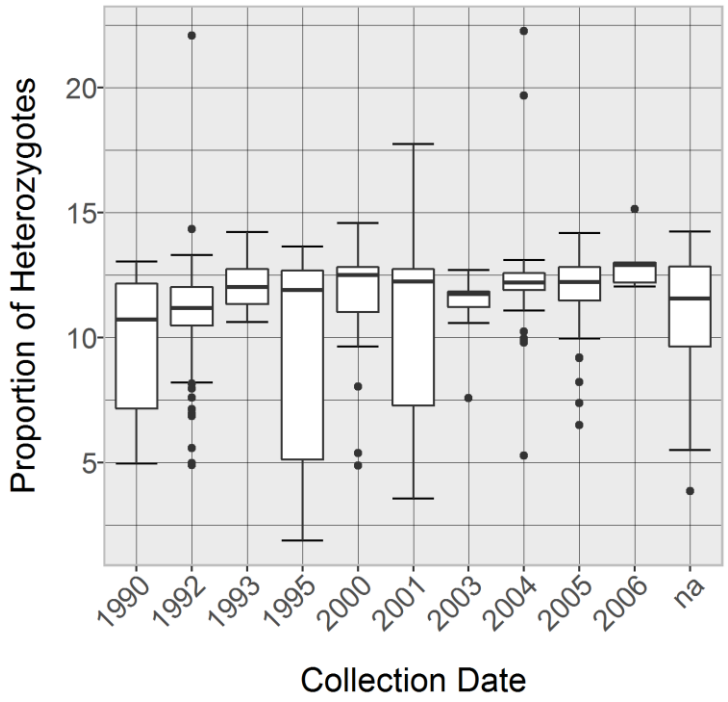


Figure 12. Distribution of the proportions of heterozygotes across time (years).

Test for normality and homogeneity of variances showed that the assumptions of normal distribution were violated (Levene's test, p-value=0.01). Kruskal-Wallis test is recommended when the assumptions of ANOVA are not met. Results from Kruskal-wallis test showed statistically significant mean differences of AEZ2 group was compared to that of AEZ6 (p =0.004).

Accessions associated with arid agroecological zone (AEZ6) within the gene bank sub-population had the highest proportion of heterozygotes (Figure 13).

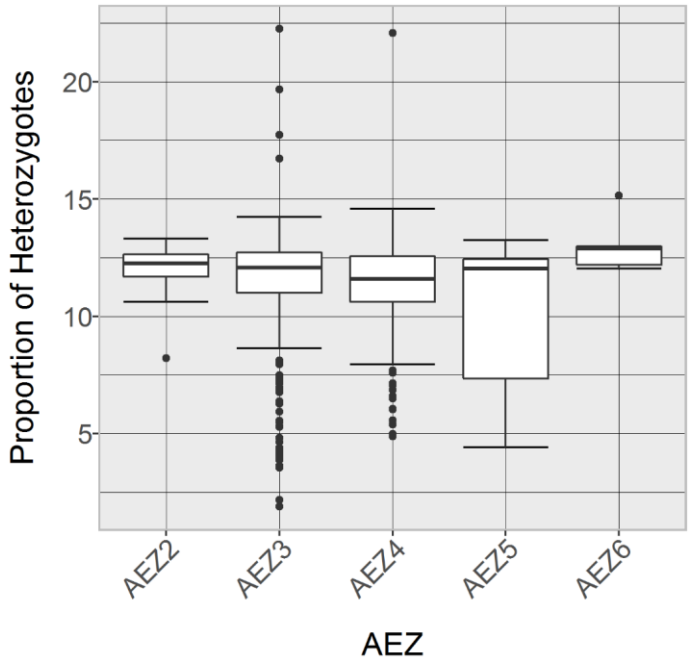


Figure 13: Proportion of heterozygotes across agroecological zones.

According to Westengen et al (2012), this could suggest presence of loci of critical importance for adaptive response to environmental stresses that characterize these agroecological zones. Due to heat and moisture stress, farmers in these zones are shifting to cultivating of drought tolerant crops such as sorghum at the expense of maize.

Among all gene bank accessions, BGF200 and BGF30 had the highest proportion of heterozygotes (0.22) and are seen as possible outliers in AEZ3 and AEZ4 (Figure 13). These accessions could also be earmarked for target selection as experimental entries during strategic research and genetic diversity studies. According to Kotze and Muller (1994), loci with critical importance for adaptive response to environmental changes

could be expected to correlate with high levels of average heterozygosity. This has also been attributed to the long term selection for adaptation or historic inter-mating of different populations.

4.1.5 Genetic Distances

To quantify genetic distances between and within accessions, pairwise allele sharing distance was computed using statistical algorithms of the R program (Bowcock et al., 1994). Pairwise genetic distance ranged from 0.004 to 0.35 and averaged 0.24. Only 130 out of 147,696 comparisons (0.08%) had a shared allele distance <0.1 which indicated low levels of redundancy among gene bank accessions and high genetic distance among nearly all pairwise comparisons (99.9%). Figure 14 below shows that the shared allele distance of most samples (>98%) fell between 0.2 and 0.3. Previous studies have reported variable genetic distances estimates ranging from 0.3 to 0.5 among inbred lines (Semagn et al., 2012) and 0.055 to 0.457 among double haploid maize lines (Ogugo et al., 2014)

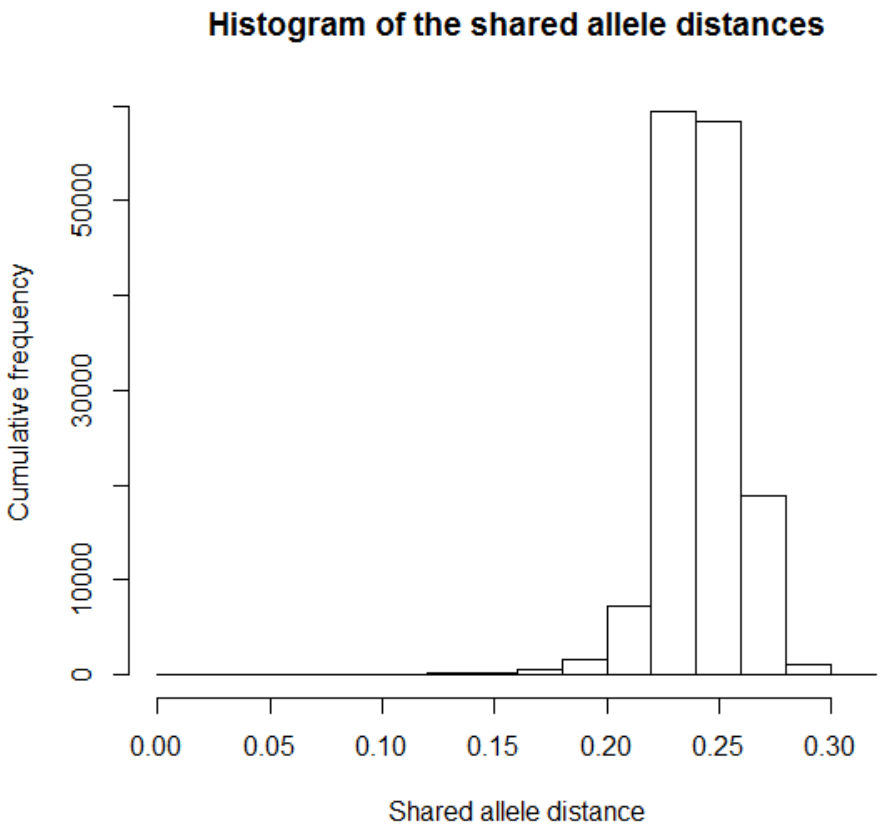


Figure 14: Pairwise comparisons of the shared allele distances among accessions.

Smaller values of shared allele distance (~0.01) between pairs indicated a high degree of relatedness or co-ancestry among accessions. This renders the maize germplasm susceptible to disease and pest outbreak or other adverse effects of climate change. While there are numerous approaches for maize inbred development, the most common method across breeding program in Kenya involves the inter-mating of existing elite materials or introgression of a desirable trait into an elite inbred line through back cross breeding. Accessions originating from similar agro-ecological zones showed a higher proportion of allele sharing than those from different agro-ecological zones

(Figure 15). This suggests the possibility of farmers saving farm seeds to plant in the next season or exchange of seeds between farmers within an agroecological zone.

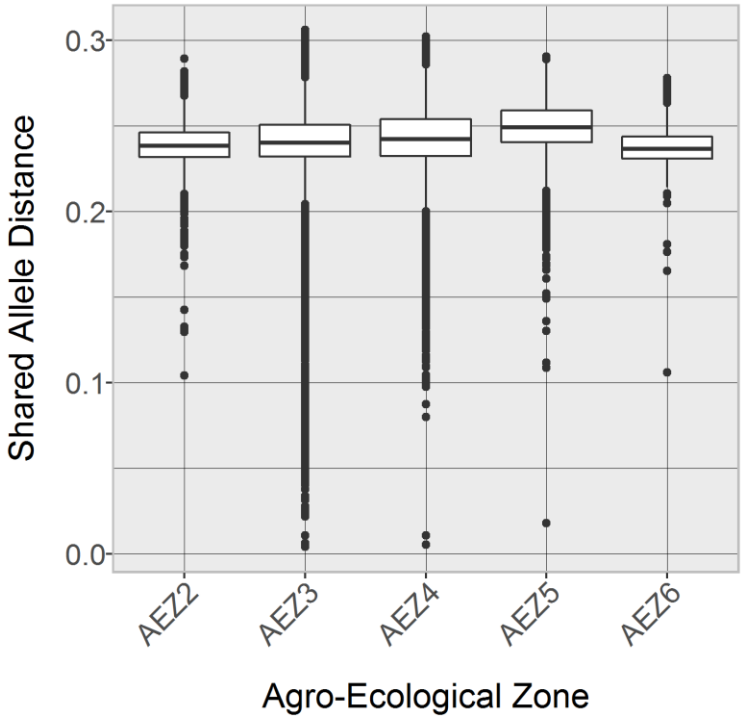


Figure 15: Pairwise shared allele distances across agro-ecological zones.

The maximum shared allele distance (~0.01) obtained between duplicated Mo17 reference lines was used to set the threshold distance for identification of putative duplicates or redundant accessions.

The basis of hybrid maize breeding is the identification of heterotic groups where breeders assign different lines to specific heterotic groups using testcrosses, pedigrees and morphological traits. Accessions originating from a similar genetic background are however, difficult to classify into these groups based on pedigree, agro-morphological or test cross information. To fully exploit heterosis and hybrid vigor, maize breeders ought

to inter-cross accessions that are genetically distant. Table 5 shows the shared allele distance of the top 5 most genetically dissimilar accessions across agro-ecological zones. Accessions separated by large genetic distance should be selected as parents during hybrid development. According to Tracy and Chandler (2006), crosses between genetically dissimilar lines produce better performing hybrids compared to those between genetically similar parents.

Table 5: Maximum pairwise shared allele distance across agroecological zones.

Agro-ecological Zone (AEZ)	Sample name	Dissimilar Pair	Shared allele distance
AEZ2	BGF45_7	BGF92_7	0.289124
AEZ3	BGF168_4	BGF200_3	0.305921
AEZ4	BGF152_8a	BGF183_4	0.302091
AEZ5	BGF167_1	BGF69_7	0.290226
AEZ6	BGF162_2	BGF168_4	0.2777

4.1.6 Population Structure

The embedding provided by PHATE (Potential of Heat-diffusion for Affinity based Transition Embedding) captured local connections between markers and aggregated them to reveal a global connectivity (Moon et al., 2017). Compared to other clustering methods such as Principal Coordinate Analysis (PCA), PHATE is able to reduce non-linear noise the data to reveal local progression paths, branches or splits in progressions paths, and their end states which may correspond to real decision points. Revelations of major or minor “evolutionary” trajectories, branches or clusters among accessions facilitates correct annotation and interpretation of PHATE visual. PHATE was

implemented using the R package (*phater*). *Phater* algorithm provided a dimensionality reducing method for visualizing trajectory structures in high-dimensional biological data. Visual inspection of the plot showed the emergence of two distinct progression paths of SNP markers from a central intersection. The intersection suggested the beginning or the ending of a genetic differentiation process. This is because the transition of data points could be either in forward or backward direction (Figure 16).

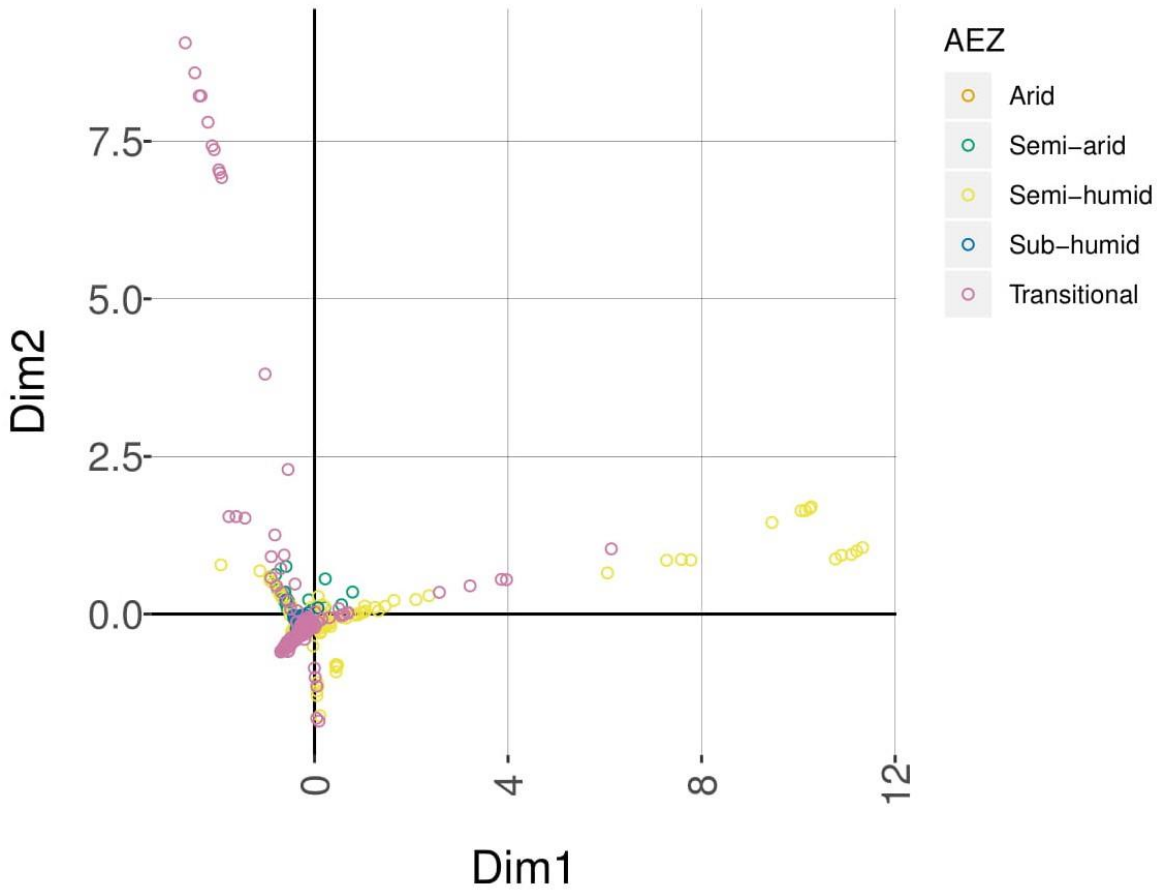


Figure 16: PHATE plot based on SNP marker data

Accessions from similar agroecological zones were fitted closer together as small clusters at various levels of the ordination plot while dissimilar ones were ordinated farther away from each other (Figure 16). Accessions associated with semi-humid and

transitional agroecological zones exhibited two divergent progressions. Accessions originated from arid, sub-humid and semi-arid agroecological zones were clustered at the origin of the two branches, suggesting which suggested an admixed ancestry and low population differentiation.

History of maize breeding in Kenya indicate that inbred lines were first derived from the well-adapted "Kenya Flat White complex" which traces back to varieties from South Africa that originally came from the Southern U.S.A such as the "Hickory King", "White Horse tooth", "White Pearl" and others. Synthetic varieties were then derived from crosses of artificially selected inbred lines. Due to the few number of valuable loci that are targeted in this breeding process the amount and distribution of genetic diversity is reduced.

To "widen the genetic base" of the maize breeding pool a large collection of maize genotypes were imported from Mexico and Colombia gene banks (Harrison, 1970). From 124 crosses, most outstanding cross was between Kitale synthetic II (KSII) variety and unimproved Ecuadorian landrace (Ecu. 573; *Montana Ecuatoriano*). The genetic dissimilarity of these two parents can be considered as a major factor in heterosis of the Hybrid 611 varietal cross which formed the basis of hybrid maize development in Kenya. With the inevitable modernization of agriculture, farmers who traditionally maintained highly diverse populations in situ increasingly replaced these local populations with highly bred maize cultivars. As a result, there is little or no genetic divergence from the two major branches which supposedly reflect the genetic polymorphism of KSII and Ecuador 573. Another interpretation could be that in situ

maize populations were initially composed of two genetically distinct populations that were continuously inbred by farmers to attain a state of homogeneity (Figure 16).

Maize is widely grown in semi-humid and transitional agroecological zones and its economic importance in these regions basically attracts a lot of research work and breeding activities which involves the screening of mainly inbred lines from different heterotic groups for tolerance or resistance traits. It is important to note that a large number of homozygous inbred lines have been developed through selfing and high levels of heterosis has been achieved by combining different inbred lines from various heterotic groups. Figure 17 shows that higher levels of genetic diversity exist among inbred lines compared to gene bank populations.

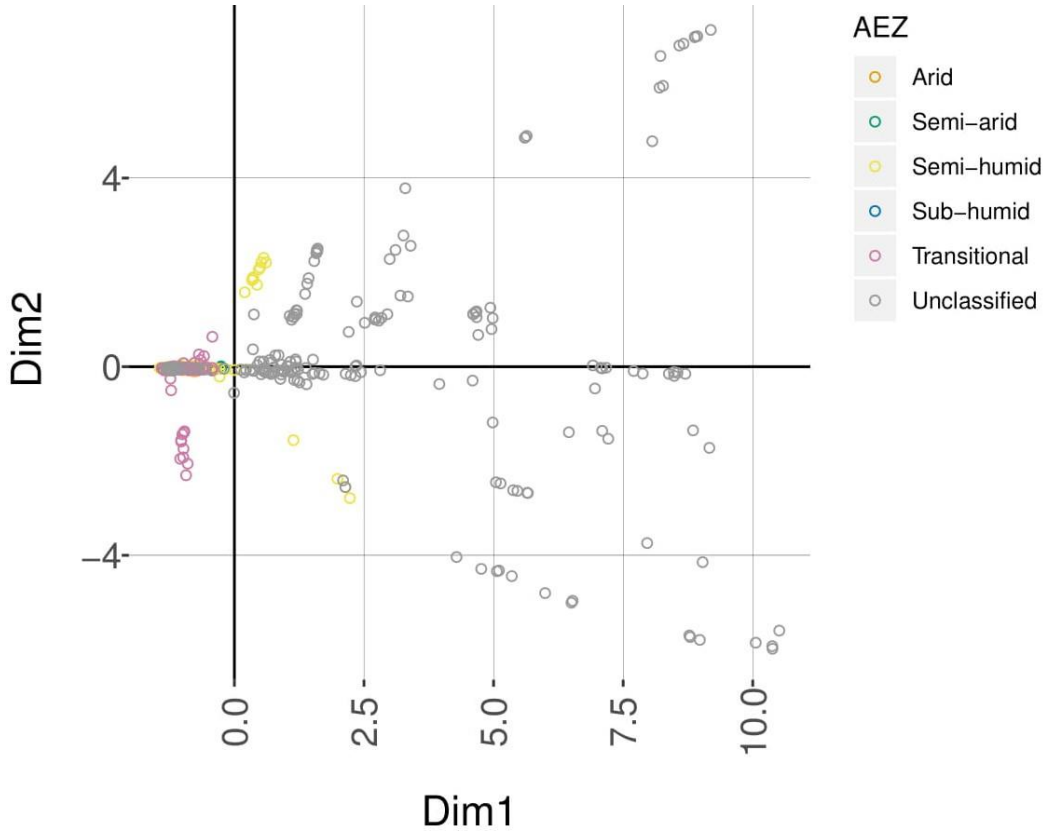


Figure 17: PHATE plot based on gene bank populations and inbred lines.

Results from PHATE suggest that the main source of genetic variability across regional maize breeding programs could be breeder to breeder seed exchange. This study suggests the need to reconsider the current maize breeding strategies. Some levels of genetic diversity should be maintained deliberately through a variety of technical and policy support programs such as giving incentives to farmers to maintain locally adapted landraces. This will promote targeted conservation of the most divergent genotypes and invite new thinking on the importance of utilizing this genetic diversity to mitigate adverse effects of biological and environmental stresses. In regions where maize

agricultural mechanization is feasible, this study suggests the systematic exploitation of advantages inbreeding and genetic uniformity.

4.2 Putative Duplicates

Global Plan Action for Conservation and Sustainable Utilization of Plant Genetic Resources for Food and Agriculture (FAO, 1996) suggested the development of core collections and reduction in duplicate accessions within and between gene bank collections. Redundancy in the gene bank collection may have different meanings depending on whether it is approached from a user (breeder) or conservationist perspective. Maize breeders for example will tend to focus more on genetic variation for particular trait(s) that are presently relevant to their current breeding objective. Gene bank curators, on the other hand will consider any amount of genetic variation even if it is of limited current use but has potential for relevance in the future.

While the contribution of genetically similar accessions to maize breeding is negligible, storage, regeneration, documentation and distribution of redundant accessions has financial implications on gene bank operations. According Treuren and van Hintum, (2003) marker-assisted reduction of gene bank redundancy can help rationalize curatorial management and enhance utilization of unique germplasm in maize breeding. Gene bank curators can utilize germplasm genetic information to assemble a limited set of accessions that reflect the genetic spectrum of the total collection. As a result, breeders can get a more objective overview of the genetic diversity conserved at the national repository and leverage genetic information to redefine their heterotic groups. In the current study, the maximum shared allele distance of (~0.01) between Mo17 duplicates

was used to set the threshold distance for identifyig other probable duplicates (Table 6). Genetic duplicates among and within gene bank accessions have been rarely found to be completely identical (SAD= 0) due to the high resolution power of SNP markers in detecting variation even in self-fertilizing crop species (Treuren and van Hintum, 2003). A more extensive and comprehensive gene bank study investigating the extent of genetic duplication and redundancy among all accessions is recommended to help earmark all duplicates for removal and re-classification of misclassified accessions.

Table 6: A list of the putative duplicates

Sample	Putative duplicate	SAD
BGF116_1	BGF116_5	0.005
BGF127_4	BGF127_5	0.006
BGFMUG57_7	BGFMUG57_8	0.003
BGF31_4	BGF31_8	0.01
BGF152_3	BGF152_4	0.01
BGF169_1	BGF169_4	0.01
Mo17_1a	Mo17_1b	0.009

Determination of probable duplicates should also be subjected to further verification using additional passport or evaluation data. This is because markers assayed at a particular time could be neutral in terms of their phenotypic effects in landrace populations but breeders might be interested in polymorphism within genes of interest.

4.3 Conclusions

Kenya has a rich reservoir of plant genetic diversity within and between plant species and across ecosystems which forms the biological basis for food and nutritional

security. With maize being the most relevant staple for millions of Kenyans, its inadequacy puts the nation's food supply at risk. Agricultural economists, policy makers and relevant government stakeholders agree that a major proportion of subsistence food requirement to meet the ever-increasing consumer demand and surplus for the market, has to come from the maize crop. The need to develop better yielding and climate resilient maize varieties has therefore been a priority research agenda. The raw material for developing newer varieties includes the shelf-ready maize genetic resources conserved at the National Gene bank of Kenya. Exploration of this collection for utilization in maize improvement however is impeded by lack of reliable molecular characterization information. As a result, accelerated breeding that could provide urgent solutions to challenges facing maize production in Kenya is impeded. In-depth knowledge on the levels and distribution of genetic diversity in the seed bank collection can optimize target selection of genetically dissimilar accessions for use in hybrid development or genetic diversity studies.

Results from the analysis of various genetic diversity indices across time and eco-geographies showed how in-situ and ex-situ management of the maize germplasm shaped the current patterns of maize genetic diversity. Although the proportion of polymorphic loci remained constant over years of in-situ management, the proportion of heterozygotes increased by 0.1% as the frequency of alleles at minor frequency decreased. The relationship between the proportion of heterozygotes and alleles at minor frequency was found to be negative and highly correlated (correlation coefficient=-79). This is particularly a concern in small threatened populations in which fixation of deleterious alleles can reduce population viability and raise the risk of extinction. Pairwise shared

allele distance for most gene bank accessions (>95%) fell between 0.20 and 0.30. This study recommends the incorporation of highly dissimilar accessions such as BGF168_4 and BGF200_3 as experimental entries during strategic research or genetic diversity studies. Accessions with maximum shared allele distance can be target selected to optimize choice of genetically diverse parents during hybrid maize development. Accessions originating from similar agro-ecological zones showed higher proportions of allele sharing than those from different agro-ecological zones and this suggested relatedness among gene bank accessions. Potential of Heat-diffusion for Affinity based Transition Embedding (PHATE) was used to capture local connections between SNP markers to delineate the population structure and revealed two genetic sub-divisions, which supported the hypothesis that the historical genomics of gene bank collection traces back to the genetic polymorphism of the two parents of hybrid 611 which formed the basis of all varietal hybrids in Kenya. Although farmers are increasingly replacing traditional populations with highly bred varieties, informal seed systems, neighboring practices and cultural seed exchanges has ensured the continued recycling and use of landraces and open pollinated varieties.

Integration of marker technology with agro-morphological trait characterization can help accelerate the development of maize hybrids with better yield and help reduce duplication of accessions in gene banks. This study suggest the expansion of molecular characterization and further evaluation gene populations to facilitate use across maize breeding networks. While modern biotechnological tools and approaches should be integrated with conventional maize breeding, the National Gene Bank of Kenya ought to developed policies and strategies that promote the evaluation, characterization and

improvement of traditional farmers' varieties in maize breeding activities. This will play a significant role in fast tracking the attainment of national food self-sufficiency envisaged in the Kenya VISION 2030.

4.4 Future Directions

Aware that all maize accessions could not be characterized due to resource and time limitations, a more direct continuation of this work could include a detailed phenotypic characterization of a core subset of the genotyped maize accessions. Historically important founder lines such as Kitale Synthetic II (KSII), Ecuadorian 573 landrace as well as the varietal hybrids derived from H611 could also be incorporated. This study also suggest the development and implementation of sampling strategies that could optimize collection of genetically diverse accession in future in situ collection missions.

A subset of the genotyped accessions could be grown in multi-locations across agroecological zones to investigate phenotypes that define ecological adaptation of maize accessions in these regions and determine whether selection environment enhance specific phenotypes such as flowering time, yield, plant height and resistance to agroecological prevalent pest and diseases. The scope of curatorial management could also be extended to include that of archiving genetic information. Phenotype data could be integrated with genotype data and accession as well as detailed passport information to build a training set from the gene bank maize accessions. This training set can be used as a model to predict the phenotypes of all gene bank accessions that will be genotyped in the future. This could also facilitate the assignment of phenotype breeding values to all

accessions in the gene bank. Future maize breeding activities and research could target select accessions for use as experimental entries using the predicted breeding values and save time and resources. This has the potential to accelerate maize improvement and improve maize breeders' response to threats such as disease pathogens and extirpating crop pests. Integration of both phenotypic data and genetic information could enhance use of maize genetic resources and promote curatorial management of the maize germplasm.

REFERENCES

Africa Agriculture Status Report. (2013). Focus on Staple Crops. Nairobi. Kenya: Alliance for a Green Revolution in Africa (AGRA).

Bailey-Serres, J., Fukao, T., Ronald, P., Ismail, A., Heuer, S., & Mackill, D. (2010). Submergence tolerant rice: SUB1's journey from landrace to modern cultivar. Rice, 3(2–3), 138–147.

Barnaud, A., Deu, M., Garine, E., Chantereau, J., Ouin, E., & Mckey, D. (2009). A weed – crop complex in sorghum: the dynamics of genetic diversity in a traditional farming system. Ameican Journal of Botany 96(10): 1869–1879.

Barnaud, A., Trigueros, G., Mckey, D., Joly, H. I., & Ii, M. (2008). High outcrossing rates in fields with mixed sorghum landraces: how are landraces maintained? Heredity. 445–452.

Buckler, E. S., J. B. Holland, P. J. Bradbury, C. B. Acharya, P. J. Brown et al., (2009). The genetic architecture of maize flowering time. Science 325: 714–718.

Cardinale, B. J., Duffy, J. E., Gonzalez, A., Hooper, D. U., Perrings, C., Venail, P., … Grace, J. B. (2012). Biodiversity loss and its impacton humanity, 0–9.

Crossa, J., Taba, S., Eberhart, S. A., Bretting, P., & Vencovsky, R. (1994). Practical considerations for maintaining germplasm in maize. Theoretical and Applied Genetics, 89(1), 89–95.

De Groote H, Owuor G, Doss C.R., Ouma J., Muhammad L., Danda K. (2005). The maize green revolution in Kenya revisited. Journal of Agricultural development Economics 2:32–49.

Demombynes G. & Kiringai Jane. (2011). The drought and food crisis in the horn of Africa Impacts and proposed responses for Kenya. Economic Premise; No. 71. World Bank. Washington, DC.

Eberhart S. A. and Harrison M. N. (1973). Progress from Half-Sib Selection in Kitale Station Maize East African Agricultural and Forestry Journal, Vol. 39. No.1, Pg. 12-16

Edwards, E. J. (2011). Rapid report: New grass phylogeny resolves deep evolutionary relationships and discovers C 4 origins. New Phytologist. 304–312.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE, 6(5), 1–10.

FAO, IFAD, & WFP. (2015). The State of Food Insecurity in the World: Meeting the 2015 international hunger targets: Taking stock of uneven progress. FAO, IFAD and WFP 2015.

FAO. (1996). Global Plan of Action for the Conservation and Sustainable Utilization of Plant Genetic Resources for Food and Agriculture and the Leipzig Declaration. International Technical Conference on Plant Genetic Resources, 63p.

Genetic Resources Research Institute. (2015). Kenya National Strategy on Genetic Resources within the context of climate change. KALRO. Kenya.

Gilding, E. K., Frère, C. H., Cruickshank, A., Rada, A. K., Prentis, P. J., Mudge, A. M., … Godwin, I. D. (2013). Allelic variation at a single gene increases food value in a drought-tolerant staple cereal. Nature Communications

Global Crop Diversity Trust. (2007). Global Strategy for the *Ex-situ* Conservation and Utilization of Maize Germplasm. Rome, Italy.

Goodman, M.M. (2005) Broadening the U.S germplasm base. Maydica 50:203-214.

Government of Kenya. (2001). Agriculture. Natural resource aspects of sustainable development in Kenya. Government Printers, Nairobi.

Government of Kenya. (2007a). Kenya Vision 2030: A Globally Competitive and Prosperous Kenya. Ministry of Planning and National Development and National Economic and Social Council. Government Printer.

Government of Kenya. (2010). National Climate Change Response Strategy. Ministry of Environment Water and Natural Resources. Nairobi, Kenya.

Govindaraj, M., Vetriventhan, M., & Srinivasan, M. (2015). Importance of Genetic Diversity Assessment in Crop Plants and Its Recent Advances: An Overview of Its Analytical Perspectives. Genetics Research International, 2015, 1-14.

Gowda, M., Das, B., Makumbi, D., Babu, R., Semagn, K., Mahuku, G., Prasanna, B. M. (2015). Genome-wide association and genomic prediction of resistance to maize lethal necrosis disease in tropical maize germplasm. Theoretical and Applied Genetics, 128(10), 1957-1968.

Hajibabaei, M., Singer, G. A. C., Hebert, P. D. N., & Hickey, D. A. (2007). DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. Trends in Genetics, 23(4), 167–172.

Harjes C.E., Rocheford T.R., Bai .L., Brutnell T.P., Kandianis C.B., Sowinski S.G., Stapleton A.E., Vallabhaneni R., Williams M., Wurtzel E.T. (2008). Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. Science 319: 330–333

Poland, J. A., Brown, P. J., Sorrells, M. E., & Jannink, J. (2012). Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. PLoS ONE, 7(2).

Hake, S., & Ross-Ibarra, J. (2015). Genetic, evolutionary and plant breeding insights from the domestication of maize. ELife. Vol 4. 1–8.

Harrison M.N. (1970). Maize Improvement in East Africa. In: Leakey C.L.A. (Ed.), Crop improvement in East Africa. Commonwealth agricultural Bull., Farnham Royal, Bucks, U.K.

Hassan, R. M., K. Njoroge, M. Njore, R. Otsyula, and A. Laboso. (1998). Adoption patterns and performance of improved maize in Kenya. In Maize technology development and transfer: A GIS application for research planning in Kenya, Hassan, ed. London: CAB International for the Overseas Development Institute.

Hoisington D, Khairallah M, Reeves T, Ribaut JM, Skovmand B, et al. (1999). Plant genetic resources: what can they contribute toward increased crop productivity? Proc Natl Acad. Sci. USA 96: 5937–5943.

Hughes J., Daily G. C., Ehrlich P. R. Population diversity: its extent and extinction. *Science* 1997, *278*, 689–692.

Institute of Economic Affairs. (2011a). Biodiversity conservation in Kenya. Trade notes. Nairobi, Kenya.

Kamau, J. I., Labeyrie, V., Njoroge, G. N., Wanjoya, A. K., Wambugu, P. W., Muthamia, Z. K., & Leclerc, C. (2017). Dealing with farmers' Ethnolinguistic differences when collecting crop diversity on-farm. *Plant Genetic Resources: Characterisation and Utilisation*, *15*(5), 400–408.

Keneni, G., Bekele, E., Imtiaz, M., & Dagne, K. (2012). Genetic Vulnerability of Modern Crop Cultivars: Causes, Mechanism and Remedies. *International Journal of Plant Research*, *2*(3), 69–79.

Kenya Agricultural Research Institute. (2000). Review of the national maize research programme. KARI Workshop proceedings. Kakamega, Kenya. pg. 40.

Kiambi D.K., Attere A.F., and Mgonja M. (2010). Plant Genetic Resources in Africa: Opportunities, Challenges and Priorities. Paper presentation for the 5th Agricultural Science Week and FARA General Assembly, held on 19-24th July 2010, Ouagadougou, Burkina Faso.

Kotzé A. and Muller, G.H. (1994). Genetic relationship in South African cattle breeds. In: Proceedings of the 5th world congress on genetics applied to livestock production, Guelph, Canada. University of Guelph, Guelph, Ontario, Canada. Volume 21: 413–416

Laborda, P. R., Oliveira, K. M., Garcia, A. A., M. E. A. G. Z. Paterniani, & Souza, A. P. (2005). Tropical maize germplasm: What can we say about its genetic diversity in the light of molecular markers? Theoretical and Applied Genetics, 111(7), 1288-1299.

Luikart, G., Allendorf, F. W., Cornuet, J. M., & Sherwin, W. B. (1998). Distortion of allele frequency distributions provides a test for recent population bottlenecks. Journal of Heredity, 89(3), 238–247.

Lund A.H., Mogens D., and Finn S.P. (1996). Increased cloning efficiency by temperature-cycle ligation. Nucleic acids research 24. 4: 800-801.

Mammadov, J., Aggarwal, R., Buyyarapu, R., & Kumpatla, S. (2012). SNP Markers and Their Impact on Plant Breeding. *International Journal of Plant Genomics*, *2012*, 728398.

Mandal, B. C. (2014). *Maize Breeding and Production Manual*: Food and Agriculture Organization of the United Nations.

Masuka, B. P., Van Biljon, A., Cairns, J. E., Das, B., Labuschagne, M., MacRobert, J., Semagn, K. (2017). Genetic diversity among selected elite CIMMYT maize hybrids in East and Southern Africa. Crop Science, 57(5), 2395–2404.

Mckenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Depristo, M. A. (2010). The Genome Analysis Toolkit : A Map Reduce framework for analyzing next-generation DNA sequencing data. 1297–1303.

Meihls, L. N., Kaur, H., & Jander, G. (2012). Natural Variation in Maize Defense against Insect Herbivores Natural Variation in Maize Defense against Insect Herbivores, *LXXVII*.

Ministry of Agriculture Livestock and Fisheries. (2015). Central Planning and Project Monitoring Unit Economic Review of Agriculture. ERA, 2015.

Miracle M. P., (1966). Maize in Tropical Africa. University of Wisconsin Press. Madison, Wis., U.S.A.: Pp. xvii, 327

Moon, K. R., Dijk, D. van, Wang, Z., Chen, W., Hirn, M. J., Coifman, R. R., Krishnaswamy, S. (2017). PHATE: A Dimensionality Reduction Method for Visualizing Trajectory Structures in High-Dimensional Biological Data. *BioRxiv*. 120378.

Munyiri S.W., Okari P., Mugo S.N., Otim M., Gibson P. and Mwololo J.K. (2010). Genetic diversity in maize landraces for resistance to *Chilo partellus* in Kenya. Second RUFORUM Biennial Meeting 20-24 September 2010, Entebbe, Uganda.

Mutegi E., Muthamia Z.K., Mutisya J and Muoki S. (2005). Study on the extent of utilization of plant genetic resources in Kenya. Agricultural Research Centre, Muguga South. Annual Report. 2005. Kenya Agricultural Research Institute.

Nakhauka, E. B. (2009). Agricultural biodiversity for food and nutrient security: the Kenyan perspective. International Journal of Biodiversity and Conservation, 1(7), 208–214.

Nass, L. L., Pellicano, I. J., & Candeira Valois, A. C. (1993). Utilization of Genetic Resources for Maize and Soybean Breeding in Brazil. Brasil. Genetics, 16(4), 983–988.

Ndiso J.B., Mugo S., Kibe A.M., Pathaka R.S. (2013) Phenotypic diversity in local coastal maize landraces in Kenya. International journal of Agricultural Sciences ISSN: 2167-0447 Vol 3 (10)

Nielsen, R. L. (2014). Determining Corn Leaf Stages. Retrieved March 3, 2016, from https://www.agry.purdue.edu/ext/corn/news/timeless/vstagemethods.html

Obiero, J. P. O., & Onyando, J. O. (2013). Climate. Developments in Earth Surface Processes, 16, 39–50.

Ogugo, V., Semagn, K., Beyene, Y., Runo, S., Olsen, M., & Warburton, M. L. (2014). Parental genome contribution in maize DH lines derived from six backcross populations using genotyping by sequencing. Euphytica, 202(1), 129–139.

Omoyo, N., Wakhungu, J., and Oteng‟i, S. (2015). Effects of climate variability on maize yield in the arid and semi-arid lands of lower Eastern Kenya. *Journal of Agriculture and Food Security, 4*(8), 1-13.

Papa, R., & Gepts, P. (2003). Asymmetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (Phaseolus vulgaris L.) from Mesoamerica, 239–250.

Perales, H. R., Benz, B. F., & Brush, S. B. (2005). Maize diversity and ethnolinguistic diversity in Chiapas, Mexico. *Proceedings of the National Academy of Sciences, 102*(3), 949–954.

Poulton, C., & Kanyinga, K. (2014). The Politics of Revitalizing Agriculture in Kenya. Development Policy Review, 32, s151-s172.

Prassana BM. (2012). Diversity in Global Maize Germplasm: Characterization and Utilization. Journal of Bioscience 37(5): 843-855.

Rabbi, I. Y., Geiger, H. H., Haussmann, B. I. G., Kiambi, D., Folkertsma, R., & Parzies, H. K. (2018). Impact of farmers' practices and seed systems on the genetic structure of common sorghum varieties in Kenya and Sudan, Plant Genetic Resources: Characterization and Utilization 8(2); 116–126.

Ramirez, R. E., Timothy, D. H., Diaz, E. B., & Grant, U. J. (1960). Races of maize in Bolivia, 167. Retrieved from Wellhausen. Razas de maiz en mexico.

Ranum, P., Peña-Rosas, J. P., & Garcia-Casal, M. N. (2014). Global maize production, utilization, and consumption. Annals of the New York Academy of Sciences, 1312(1).

Reynolds, T. W., Waddington, S. R., Anderson, C. L., Chew, A., True, Z., Cullen, A True, Z. (2015). Environmental impacts and constraints associated with the production of major food crops in Sub-Saharan Africa and South Asia, 795–822.

Romay, M. C., Millard, M. J., Glaubitz, J. C., Peiffer, J. A., Swarts, K. L., Casstevens, T. M., … Gardner, C. A. (2013). Comprehensive genotyping of the USA national maize inbred seed bank.

Romero Navarro, J. A., Willcox, M., Burgueño, J., Romay, C., Swarts, K., Trachsel, S... Buckler, E. S. (2017). A study of allelic diversity underlying flowering-time adaptation in maize landraces. Nature Genetics, 49(3), 476–480.

Rubenstein, D.K., Heisey, P., Shoemaker R., Sullivan, J. And Frisvold, G. (2005). Crop genetic resources: An economic appraisal. (USDA). Economic Information Bulletin No.2.s

Schroeder, C., Onyango, K. O., Nar Bahadur, R., Jick, N., Parzies, H., and Gemenet, D. (2013). Potentials of hybrid maize varieties for small-holder farmers in Kenya: A Review based on SWOT analysis. *African Journal of Food, Agriculture, Nutrition and Development, 3*(2), 7562-7568.

Semagn, K., Magorokosho, C., Vivek, B. S., Makumbi, D., Beyene, Y., Mugo S., Warburton, M. L. (2012). Molecular characterization of diverse CIMMYT maize inbred lines from eastern and southern Africa using single nucleotide polymorphic markers. BMC Genomics, 13(1), 113.

Shiferaw, B., Prasanna, B.M., Hellin, J. and Bänziger, M. (2011). Crops that feed the world. Past successes and future challenges to the role played by maize in global food security. Food Security 3(3):307-327.

Shimelis, H., & Laing, M. (2012). Timelines in conventional crop improvement: Pre-breeding and breeding procedures. Australian Journal of Crop Science, 6(11), 1542–1549.

Smale, M., & Jayne, T. (2003). Maize in Eastern and Southern Africa: "Seeds" of Success in Retrospect. Food Policy, (97), 1–79.

Smale, M., Byerlee D., and Jayne T., (2011). Maize revolutions in Sub-Saharan Africa: Forthcoming, World Bank, Washington, DC and Tegemeo Inst., Nairobi, Kenya.

Sombroek, W.G., Braun H.M., and van der Pour, B.J.A. (1982). Exploratory Soil Map and Agro-climatic Zone Map of Kenya. Report E1. National Agricultural Laboratories, Soil Survey Unit, Nairobi, Kenya.

Turakulov, R., & Easteal, S. (2003). Number of SNP loci needed to detect population structure. Human Heredity, 55(1), 37–45.

Van Treuren, R., & Van Hintum, T. J. L. (2003). Marker-assisted reduction of redundancy in germplasm collections: Genetic and economic aspects. Acta Horticulturae, 623, 139–149.

Wallace, J. G., Larsson, S. J., & Buckler, E. S. (2014). Entering the second century of maize quantitative genetics. Heredity, 112(1), 30–38.

Wambugu, P. W. and Muthamia Z.K. (2009). The State of Plant Genetic Resources for Food and Submitted To FAO Commission on Plant Genetic Resource for Food and Agriculture, (July 2009).

Westengen, O. T., Berg, P. R., Kent, M. P., & Brysting, A. K. (2012). Spatial Structure and Climatic Adaptation in African Maize Revealed by Surveying SNP Diversity in Relation to Global Breeding and Landrace Panels, *7*(10).