GENETIC ANALYSIS OF TRADITIONAL ETHIOPIAN HIGHLAND MAIZE (*Zea mays* L.) USING MOLECULAR MARKERS AND MORPHOLOGICAL TRAITS: IMPLICATION FOR BREEDING AND CONSERVATION

By

YOSEPH BEYENE AYDAGN

Submitted in partial fulfillment of the requirements for the degree of

Philosophiae Doctor:

PLANT BREEDING /GENETICS

In the Faculty of Natural and Agricultural Sciences,

Department of Genetics and Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria

Pretoria

July 2005

Supervisor: Dr. Alexander A. Myburg

Co-Supervisor: Prof. Anna-Maria Oberholster

DECLARATION

I declare that the dissertation, which I here by submit for the degree of Doctor of Philosophy in Plant Breeding/Genetics at the University of Pretoria, is my own work and has not been previously submitted by me for a degree at another University.

_____ _____

TABLE OF CONTENTS ……..……..…**Pages**

Preface…….……..….viii

Acknowledgments…….....……...ix

List of Tables…….…….…..…..xi

List of Figures …….…..….xiv

List of abbreviations…….xvii

Summary……...…..…..….xix

CHAPTER 1.....**1**

GENERAL INTRODUCTION.....**1**

1.1 Maize breeding in Ethiopia: Historical overview **2**

1.2 Importance of maize in the highlands of Ethiopia.....**3**

1.3 Objectives and outline of the study **4**

CHAPTER 2.....**7**

LITERATURE REVIEW **7**

2.1 Maize is an important crop for genetic analysis **8**

2.1.1 Diversity in maize 9

2.1.2 Maize genome evolution 10

2.1.3 Wild relatives of maize 11

2.2 Methods for assessing genetic variation 12

2.2.1 Morphology and pedigree data 14

2.2.2 Molecular markers 15

2.3 Use of pooled DNA samples in the study of genetic variation 20

2.4 Correlation between phenotypic and molecular markers distance 21

2.5 Statistical measures for assessing genetic diversity 23

2.5.1 Types of distance measures 23

2.5.2 Multivariate methods 24

2.5.3 Clustering methods 26

2.5.4 Partitioning of variation 27

2.6 Gene mapping/tagging 28

2.6.1 Conventional method of QTL detection 29

2.6.2 Bulk segregant analysis 30

2.6.3 Association mapping 31

2.6.4 Comparative genetic mapping 32

2.7 Marker-assisted selection and breeding 33

2.7.1 Introgression of desirable genes 35

2.8 Conclusions 36

CHAPTER 3 ..... **39**

PHENOTYPIC DIVERSITY FOR MORPHOLOGICAL AND AGRONOMIC TRAITS IN TRADITIONAL ETHIOPIAN HIGHLAND MAIZE ACCESSIONS **39**

3.1 ABSTRACT **40**

3.2 INTRODUCTION **41**

3.3 MATERIALS AND METHODS **42**

3.3.1 Plant materials and field evaluation 42

3.3.2 Statistical analysis 43

3.4 RESULTS **43**

3.4.1 Morphological and agronomic variability 43

3.4.2 Phenotypic and genotypic coefficients of variation, heritability and genetic advance as percent of the mean 44

3.4.3 Genotypic and phenotypic correlations 46

3.4.4 Cluster analysis 48

3.4.5 Principal component analysis 50

3.5 DISCUSSION **51**

CHAPTER 4 ..... **56**

BULKED-AFLP ANALYSIS OF GENETIC DIVERSITY AMONG

TRADITIONAL ETHIOPIAN HIGHLAND MAIZE ACCESSIONS.....56

4.1 ABSTRACT.....57

4.2 INTRODUCTION.....58

4.3 MATERIALS AND METHODS.....60

4.3.1 Plant materials and DNA extraction.....60

4.3.2 AFLP analysis.....62

4.3.3 Gel electrophoresis and scoring.....63

4.3.4 Data analysis.....64

4.4 RESULTS.....66

4.4.1 Detection limit of bulked AFLP analysis.....66

4.4.2 Marker polymorphism.....68

4.4.3 Distribution of bands across the three agroecologies.....71

4.4.4 Genetic dissimilarity of maize accessions.....72

4.4.5 Partitioning of genetic variation.....74

4.5 DISCUSSION.....74

CHAPTER 5.....80

ANALYSIS OF GENETIC DIVERSITY OF ETHIOPIAN HIGHLAND MAIZE ACCESSIONS USING SSR MARKERS.....80

5.1 ABSTRACT 81

5.2 INTRODUCTION 82

5.3 MATERIALS AND METHODS 84

5.3.1 Plant materials and DNA extraction 84

5.3.2 Simple sequence repeats primer selection 86

5.3.3 PCR amplification and gel electrophoresis 86

5.3.4 Data analysis 87

5.4 RESULTS 88

5.4.1 Detection limit of pooled DNA and marker polymorphism 88

5.4.2 Genetic dissimilarity and cluster analysis 91

5.4.3 Partitioning of genetic variation 92

5.5 DISCUSSION 93

CHAPTER 6 98

A COMPARATIVE STUDY OF MOLECULAR AND MORPHOLOGICAL METHODS OF DESCRIBING GENETIC RELATIONSHIPS IN MAIZE 98

6.1 ABSTRACT 99

6.2 INTRODUCTION 100

6.3 MATERIALS AND METHODS 102

6.3.1 Field evaluation and data recording 102

6.3.2 Plant materials DNA extraction 102

6.3.3 AFLP analysis 103

6.3.4 SSR analysis 103

6.3.5 Statistical analysis 104

6.4 RESULTS 105

6.4.1 Morphological variability 105

6.4.2 Variation in molecular markers 107

6.4.3 Distribution of dissimilarity coefficients 109

6.4.4 Correlations between dissimilarity matrices 110

6.5 DISCUSSION 114

CHAPTER 7 118

ASSOCIATION OF SIMPLE SEQUENCE REPEATS WITH QUANTITATIVE TRAITS IN ETHIOPIAN HIGHLAND MAIZE ACCESSIONS AND THE EFFECT OF ADMIXTURE 118

7.1 ABSTRACT 119

7.2 INTRODUCTION 120

7.3 MATERIALS AND METHODS 122

7.3.1 Selection of plant materials 122

7.3.2 SSR analysis 122

7.3.3 Statistical analysis 124

7.4 RESULTS 125

7.4.1 Phenotypic trait variance 125

7.4.2 Markers and phenotypic trait associations 126

7.4.3 Detection of chromosomal regions affecting the quantitative traits 129

7.4.4 Allelic frequency in contrasting agroecologies 131

7.4.5 Prediction of quantitative variation using significant markers 133

7.6 DISCUSSION 136

CHAPTER 8 141

CONCLUSIONS AND RECOMMENDATIONS 141

Use of bulked leaf samples for genetic diversity studies 143

REFRENCES 149

PREFACE

This work was conducted at the Department of Genetics and Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, South Africa. The project involved laboratory, glasshouse and field experiments. Various molecular marker techniques and quantitative genetics approaches were employed to accurately unravel the extent of genetic diversity and genetic relationships among traditional Ethiopian highland maize accessions with the view of establishing a strategic maize improvement program in the highlands of Ethiopia.

This dissertation is based on the following chapters, which were published, accepted or submitted for publication.

1. Phenotypic diversity for morphological and agronomic traits in traditional Ethiopian highland maize accessions (*South African Journal of Soil and Plant 2005, 22: 100-105)*
2. Genetic diversity in traditional Ethiopian highland maize accessions assessed by AFLP markers and morphological traits (*Journal of Biodiversity and Conservation, in press*).
3. Genetic diversity among traditional Ethiopian highland maize accessions assessed by simple sequence repeat (SSR) markers (*Journal of Genetic Resources and Crop Evolution, in press*).
4. A comparative study of molecular and morphological methods of describing genetic relationships in maize (*African Journal of Biotechnology 2005,4:586-595)*
5. Association of simple sequence repeats with quantitative traits in Ethiopian highland maize accessions and the effect of admixture.

ACKNOWLEDGMENTS

It is my pleasure to thank the many people who contributed to the success of my PhD studies.

I wish to express my sincere gratitude to my major advisor, Dr. Alexander A. Myburg for his guidance, encouragement, moral support and friendship throughout my study period. Moreover, he always made time to talk science and share his experience. I would also like to thank my co-advisor, Prof. Anna-Maria Oberholster for her helpful advice and guidance.

I am indebted to the Alemaya University Agricultural Research and Training Project (AU-ARTP) for providing me a PhD fellowship. I am also grateful to the International Foundation for Science (IFS) for funding for molecular analysis.

My gratitude goes to the University of Pretoria; Forestry and Agricultural Biotechnology Institute (FABI) especially people in the Forest Molecular Genetics Laboratory for their support and friendship. In particular, I would like to thank Dr. Solomon Kebede, a post doctorial fellow, for data analysis, reading the manuscripts and moral support throughout the study period.

I am also grateful to the Maize Improvement Program, Alemaya University, Ethiopia, for the technical assistance with field trial and to Dr. Habtamu Zelleke for skilful collaboration with field observations. I would like to thank Dr. T. Hole and Dr. M.

Warburton, CIMMYT, Mexico, for critical reading and helpful suggestions on Chapter 3 and 4, respectively.

Tekalign Tsegaw, Getu Beyene, Abubeker Hassen, Yibekal Alamayehu, Arega Demelashi, Tesfaye Lemma and many other people have wonderfully demonstrated how to enjoy living life. They have been a constant source of strong moral support as a colleague and as a friend.

Lastly but most importantly, I wish to express my deepest appreciation to my beautiful wife, Elsa and my parents. They have given me a tremendous amount of love, encouragement, unwavering support, and sacrifice that has allowed me to complete my study successfully. I will never be able to repay them, but my accomplishments will always be their accomplishments.

LIST OF TABLES

Table 2.1 Comparison of the most common used marker systems in plant breeding. 20

Table 3.1 Means, standard deviation of the means (St Dev), ranges and mean squares for 15 agro-morphological traits measured in 180 maize accessions 45

Table 3.2 Estimates of phenotypic and genotypic coefficients of variability, heritability and genetic advance as percent of mean 46

Table 3.3 Genotypic (above diagonal) and phenotypic (below diagonal) correlation coefficients between 15 agro-morphological traits measured in 180 maize accessions 47

Table 3.4 Eigenvector, eigenvalues, individual and cumulative percentage of variation explained by the first five principal components (PC) after assessing agro-morphological traits in 180 maize accessions 51

Table 4.1 Traditional Ethiopian highland maize accessions used in the study 61

Table 4.2 Relationships between AFLP band frequencies in individual plants and representation in bulks of 15 individuals based on the average of three primer pair combinations 67

Table 4.3 Degree of polymorphism and average polymorphism information content for the eight AFLP primer combinations used to analyze the 62 Ethiopian maize accessions 69

Table 4.4 Distribution of AFLP bands expressed as the percentage of accessions that carry a particular band in each agroecology 71

Table 4.5 Partitioning of the total genetic variation of traditional Ethiopian highland maize accessions into within and between agroecologies variation 74

Table 5.1 Summary of microsatellite, bin number, repeat unit, number of alleles per locus and the polymorphism information content (PIC) of the different agroecologies of traditional Ethiopian highland maize accessions..... 90

Table 5.2 Partitioning of the total genetic variation into within and between agroecologies of traditional Ethiopian highland maize accessions..... 93

Table 6.1 Summary statistics of the agro-morphological traits measured in 62 traditional Ethiopian highland maize accessions 106

Table 6.2 Eigenvectors, eigenvalues, individual and cumulative percentage of variation explained by the first four principal components (PC) after assessing morphological traits in 62 traditional Ethiopian highland maize accessions..... 107

Table 6.3 Level of polymorphisms and informativeness obtained with AFLP and SSR markers in 62 traditional Ethiopian highland maize accessions 108

Table 6.4 Mean, standard deviation and range of Nei and Li dissimilarity coefficients (calculated using AFLP and SSR markers) and Euclidean distance (calculated using morphological traits). The total sample of all accessions in this study is shown followed by accessions collected from the three agroecologies..... 110

Table 6.5 Correlation between dissimilarity matrices obtained with different marker types 111

Table 7.1 The genotypic correlation coefficients among the six traits of the 62

traditional Ethiopian highland maize accessions used in study 126

Table 7.2 The 20 SSR loci used and the designation of the polymorphic alleles scored at each locus 127

Table 7.3 Accessions name, their observed and predicted performance for ear traits in traditional Ethiopian highland maize accessions 133

Table 7.4 Accessions name, their observed and predicted performance for morphological traits in traditional Ethiopian highland maize accessions 135

LIST OF FIGURES

Figure 3.1 Mean values of morphological traits for traditional Ethiopian highland maize accessions for the four clusters generated through UPGMA as clustering techniques. Clusters having the same letter within each trait are not statistically significant at $p = 0.05$. Number in parenthesis (in the legend) indicates accessions grouped in each cluster. 48

Figure 3.2 Mean values of agronomic traits for traditional Ethiopian highland maize accessions for the four clusters generated through UPGMA as clustering techniques. Clusters having the same letter within each trait are not statistically significant at $p = 0.05$. Number in parenthesis (in the legend) indicates accessions grouped in each cluster. 49

Figure 4.1 An example of the LI-COR AFLP image generated using E-AAC/M-CGG primer combinations showing banding patterns of 15 individuals (1-15) and their bulked leaf samples (B). A molecular marker is indicated at the beginning and at the end of the gel. Arrows show examples of bands present in individuals and absent in bulked samples 66

Figure 4.2 Typical a LI-COR AFLP image produced by selective amplification using the E-ACG/M-CGG primer combination in 62 maize accessions. Lanes 1 and 64 are IRDye 700 molecular weight standards (LI-COR Biosciences).....70

Figure 4.3 Frequency distribution of the genetic dissimilarity of pairs of 62 Ethiopian highland maize accessions 72

Figure 4.4 Dendrogram of traditional Ethiopian highland maize accessions derived by Ward' minimum variance method from the dissimilarity matrix of AFLP data .73

Figure 5.1 Map showing the 62 maize accessions that were collected from different highlands of Ethiopia and used in the present study. The approximate location of each collection site is indicated in the map by square points 85

Figure 5.2 The SSR 3% gel image of individual plants (lane 1-15) and the bulk sample (lane B) of the accession 'Baw 10' amplified with locus *phi034*. M is molecular ladder (Gene Ruler 100-bp DNA ladder Plus, Frementas). Arrow shows the band present in individual plants and absent in the bulked sample 88

Figure 5.3 Frequency distribution of the genetic dissimilarity pairs of 62 Ethiopian maize accessions based on 20 SSR loci.....91

Figure 5.4 Ward minimum variance based dendrogram generated from the Nei and Li dissimilarity matrix showing relationships among 62 traditional Ethiopian highland maize accessions. 92

Figure 6.1 Frequency distribution of genetic dissimilarity among pair-wise combinations of 62 traditional Ethiopian highland maize accessions based on morphology, AFLP and SSR data.....109

Figure 6.2 Dendrogram of traditional Ethiopian highland maize accessions derived by UPGMA from the dissimilarity matrix of the morphological data.....112

Figure 6.3 Dendrogram of the 62 traditional Ethiopian highland maize accession based on the Ward minimum variance method applied to the dissimilarity matrix

generated by Nei and Li dissimilarity coefficients of the pooled AFLP and SSR data 113

Figure 7.1 Dendrogram resulting from cluster analysis of standardized morphological and agronomic data from 180 maize accessions using Euclidean distance and UPGMA clustering. Sixty-two accessions were selected for molecular analysis (indicated by horizontal bars as root of the dendrogram) 123

Figure 7.2 Normal probability plot for (A) days to silking; (B) plant height; (C) days to maturity; (D) kernels/row; (E) seed weight and (F) yield showing that all the residuals fall within the confidence limits of the normal probability plot 128

Figure 7.3 Mean allelic frequencies of SSR alleles selected for adaptation in contrasting agroecologies in Ethiopian highlands 132

LIST OF ABBREVIATIONS

°C	Degree celsius
µl	Microliter
µg	Microgram
µM	Micromolar
AEI	Assay efficiency index
AFLP	Amplified fragment length polymorphism
AMOVA	Analysis of molecular variance
AOD	Analysis of distance
bp	Base pair
CA	Cluster analysis
CIMMYT	International Maize and Wheat Improvement Center
cm	Centimeter
cM	Centimorgan
CSA	Central statistical authority
DNA	Deoxyribonucleic acid
dNTPs	The mixtures of four deoxynucleotides triphosphate
DSK	Days to silking
DTS	Days to tasseling
DYM	Days to maturity
e.g.	Example
EARO	Ethiopian Agricultural Research Organization
ERD	Ear diameter
ERH	Ear height
ERL	Ear length
EST	Expressed sequence tag
FAPRI	Food and Agricultural Policy Research Institute
g	Gram
GA	Genetic advance
GCV	Genotypic coefficient of variability
GS_{NL}	Genetic similarity base on Nei and Li
GS	Genetic similarity
GS_J	Genetic similarity based on Jaccard
GS_{MR}	Genetic similarity based on Modified Roger's distance
GS_{SM}	Genetic similarity Simple matching
h	Hour
h^2	Broad sense heritability
ha	Hectare
HCl	Hydrochloric acid
IRDye	Infrared day
Kg ha^{-1}	Kilogram per hectare
KCl	Potassium chloride
KLR	Kernels per row
LD	Linkage disequilibrium
LFL	Leaf length

LFW	Leaf width
MA	Millampere
MAS	Marker assisted selection
masl	Meters above sea level
Mb	Mega billion
$MgCl_2$	Magnesium chloride
min	Minute
mM	Millimolar
mm	Millimeter
mmt	Million metric tons
NIL	Near isogenic line
NMSA	National metrological service agency
no.	Number
^{ns}	Not significant
PC	Principal component
PCA	Principal component analysis
PCR	Polymerase chain reaction
PCV	Phenotypic coefficient of variation
PIC	Polymorphism information content
PLH	Plant height
QTL	Quantitative trait loci
R/L	Restriction ligation
R^2	Multiple regression coefficient
RAPD	Random amplified polymorphic DNA
RFLP	Restriction fragment length polymorphism
RWN	Rows per ear
s	Second
SDW	Seed weight
SNP	Single nucleotide polymorphism
Spp.	Species
SSR	Microsatellite or simple sequence repeat
St Dev	Standard deviation of the means
t	Ton
UPGMA	Unweighted paired group method using arithmetic averages
V	Volt
V_B	Between group variation
V_T	Total variation
V_W	Within group variation
W	Watt
w/v	Weight volume ratio
YD	Yield

GENETIC ANALYSIS OF TRADITIONAL ETHIOPIAN HIGHLAND MAIZE (*Zea mays* L.) USING MOLECULAR MARKERS AND MORPHOLOGICAL TRAITS: IMPLICATION FOR BREEDING AND CONSERVATION

By

Yoseph Beyene Aydagn

Supervisor: Dr. Alexander A. Myburg

Co-supervisor: Prof. Anna-Maria Oberholster

Department: Genetics

Degree: PhD

SUMMARY

Knowledge of the genetic variation of crop collections is essential for their efficient use in plant breeding programs. The Ethiopian Highland Maize Germplasm Collection Mission was launched throughout the highlands of Ethiopia in 1998 and 287 traditional maize accessions were collected from farmers' fields. To date, no information was available on the morphological and genetic diversity in this important collection. Various molecular marker techniques and quantitative genetics approaches were applied to accurately unravel the extent of phenotypic and genetic diversity, to study patterns of morphological and molecular variation and to determine association of molecular markers with quantitative trait variation, with the view of designing a sound breeding program and management strategy for maize in the highlands of Ethiopia.

The morphological study confirmed that traditional Ethiopian highland maize accessions contain large amounts of variation for agro-morphological traits. The broad

trait diversity observed among the accessions suggested ample opportunities for the genetic improvement of the crop through selection directly from the accessions and/ or the development of inbred lines for a future hybrid program. Selection practices followed by local farmers are mostly consistent within agroecology and gave rise to morphologically distinct maize accessions in different agroecologies. This underscores the importance of considering farmers' knowledge of diversity in the collection and evaluation of local accessions.

The results of amplified fragment length polymorphism (AFLP) and microsatellite or simple sequence repeat (SSR) marker analyses showed that bulking leaf samples from 15 individual plants per out-bred accession is an effective means of producing representative profiles of individual plants, thereby reducing the cost of DNA extraction and subsequent marker analysis of open-pollinated varieties. Cluster analyses based on AFLP and SSR data showed that most of the accessions collected from the Northern agroecology were genetically distinct from the Western and Southern accessions suggesting that differentiation for adaptive traits for drought conditions may have occurred in the Northern accessions. However, there was very little genetic differentiation between the Western and Southern accessions suggesting gene flow between the two agroecologies and recent introduction of similar improved varieties in these agroecoogies . In both marker systems, high mean genetic diversity was observed among the traditional Ethiopian highland maize accessions. This is possibly due to (i) the continuous introduction of maize from abroad by different organizations; (ii) genetic variation generated through farmers management practices; and (iii) the presence of different environmental conditions in the highlands of

Ethiopia to which local landraces may have been adapted.

The correlation between the morphological dissimilarity matrix and the matrices of genetic dissimilarity based on SSR and AFLP markers were 0.43 and 0.39, respectively (p = 0.001 in both cases). The correlation between SSR and AFLP dissimilarity matrices was 0.67 (p = 0.001). These significant correlations indicate that the three independent sets of data likely reflect the same pattern of genetic diversity, and validate the use of the data to calculate the different diversity statistics for Ethiopian highland maize accessions**.** From this study, three groups of maize accessions with distinctive genetic profiles and morphological traits were identified that will be useful for future collection, conservation and breeding programs of maize for the highlands of Ethiopia.

A pilot association study using SSR markers and quantitative trait variation indicated that molecular markers could be useful to identify genetic factors controlling earliness, tallness, grain yield and associated traits, which could be exploited by various breeding schemes. The analytical tools outlined in this dissertation can be a useful tool in managing genetic variation of open-pollinated crops and will aid in the conservation of unique genetic diversity. Production stability and global food security are linked to the conservation and exploitation of worldwide genetic resources and this research attempts to add to that body of knowledge.

Key words: AFLP markers, association mapping, bulked analysis, clustering, correlation, Ethiopia, genetic diversity, genetic resources, heritability, highland maize, quantitative traits, phenotypic diversity, regression analysis, SSR markers

CHAPTER 1

GENERAL INTRODUCTION

1.1 Maize breeding in Ethiopia: Historical overview

Maize was first introduced to Ethiopia by the Portuguese in the 16th or 17th century (Hafnagel, 1961). Since its introduction, it has gained importance as a food and feed crop in Ethiopia. Averages of the 2000/2001 national production estimates of the Central Statistical Authority (CSA, 2001) indicate that maize, with 1.40 million hectares and 2.52 million tons, accounts for about 20.9% of the total area and 32.6% of the gross annual grain production. Maize is one of the cereals that provide calorie requirements in the traditional Ethiopian diet. It is prepared and used as unleavened bread, roasted and boiled green ears, parched mature grain porridge and in local drinks like 'tella', 'borde' and 'areke' (Mulatu *et al*., 1992). Apart from these uses, maize leaves are fed to animals, while dry stalks are used as fuel and for the construction of fences and huts.

Maize growing areas in Ethiopia are broadly classified into four ecological zones based on altitude and annual rainfall (EARO, 2000). These are (1) the high altitude moist zone, which receive 1200 to 2000 mm rainfall and is at an altitude of 1700 to 2400 meters above sea level (masl), (2) the mid-altitude moist zone (1200-2000 mm rainfall and an altitude of 1000 to 1700 masl), (3) the low-altitude moist zone (less than 1000 masl and 1200-1500 mm rainfall), and (4) the moisture stress zone (from 500 to 1800 masl and receive less than 800 mm rainfall).

Maize breeding in Ethiopia has been ongoing since the 1950's and has passed through three distinctive stages of research and development (Mulatu *et al*., 1992). These are (1) from 1952 to 1980, the main activities were the introduction and evaluation of

maize materials from different part of the world for adaptation to local conditions, (2) from 1980 to 1990, the work was focused on evaluation of inbred lines and development of hybrid and open-pollinated varieties, and (3) from 1990 to present, the main activities were (a) extensive inbreeding and hybridization, (b) development of early maturing or drought tolerance cultivars, and (c) collection and improvement of maize with adaptation to highland agroecologies. As a result, various improved hybrids and open-pollinated varieties were released for large-scale production, especially for mid-altitude zones. The highland maize breeding program was started in 1998 in collaboration with the International Maize and Wheat Improvement Center (CIMMYT). The initial objectives of the program were facilitating collection, evaluation and documentation of locally important highland maize accessions in eastern African countries.

1.2 Importance of maize in the highlands of Ethiopia

The highland areas of Ethiopia constitute 36% of the total land area, over 90% of the crop land and support about 88 and 70% of the human and the livestock population, respectively (CSA, 1998). In these regions, 20% of the total land was devoted to maize cultivation and more than 30% of small-scale farmers depend on maize for their livelihood (CSA, 1998). In light of the rapidly increasing human population and expansion of agriculture into the highland areas of Ethiopia, maize has been selected as one of the national commodity crops (due to its high yield potential and wide adaptation) for the food self-sufficiency program of the country (Mulatu *et al.*, 1992).

Maize cultivars that are used in the highland regions of Ethiopia are well adapted, but low yielding, open-pollinated varieties developed by local farmers. Many of these varieties resulted from centuries of planting, harvesting and selection. The highland maize varieties may be grouped into a number of completely or partially isolated populations, which may each be adapted to different highland conditions. Unfortunately, formal breeding programs have had little success in developing improved varieties for these diverse agroecologies. In the past decade, only two improved open-pollinated maize varieties were developed for the highland zone (Twumasi-Afriyie *et al*., 2001).

In view of the above, the Highland Maize Germplasm Collection Mission was launched throughout the highlands of Ethiopia in collaboration with CIMMYT in 1998 (Twumasi-Afriyie *et al*., 2001). As part of this project, 287 maize accessions were collected from farmers' fields throughout the highland regions of Ethiopia. Currently, there is no information on the extent of morphological and genetic variation among these accessions.

1.3 Objectives and outline of the study

Molecular markers are rapidly being adopted by crop improvement researchers globally as effective and appropriate tools for basic and applied studies in crop plants. Use of molecular markers in maize breeding ranged from facilitating appropriate choice of parents for crosses, to mapping/tagging of gene blocks associated with economically important traits often termed as quantitative trait loci (QTLs).

In this PhD study, Amplified Fragment Length Polymorphism (AFLPs), microsatellites or Simple Sequence Repeats (SSRs) and agro-morphological traits were employed for genetic analysis of traditional Ethiopian highland maize accessions. The general objectives of this study were (1) to classify the highland maize accessions into distinct groups based on genetic profiles and morphological traits, (2) to test the utility of bulking DNA (by bulking leaf samples) for large-scale genetic characterization of open-pollinated varieties, (3) to determine the correlation between estimates of genetic diversity measured by AFLPs, SSRs and morphological traits, and (4) to study population level association of molecular markers with quantitative trait variation in diverse maize accessions. The specific objectives and research methodologies for each experiment is presented in detail under each chapter.

In chapter 2, the importance of maize for genetic studies, methods and measures for assessing genetic variation, and application of molecular markers in crop improvement are reviewed. In chapter 3, the phenotypic diversity of traditional Ethiopian highland maize accessions is presented. Multivariate statistics were employed to classify the accessions into similar phenotypic groups. In addition, different statistical measures such as phenotypic and genotypic variability, broad sense heritability and genetic advance were calculated. This chapter allowed the selection of representative maize accessions from all agroecologies and phenotypic classes that were used for subsequent molecular analyses.

In chapter 4, bulked-AFLPs were employed to study genetic diversity and relationships of selected maize accessions. The results of this chapter highlight the use

of bulked leaf samples for large-scale genetic characterization of open-pollinated varieties, which are commonly cultivated in developing countries. These investigations are continued in chapter 5. In this chapter, the utility of analyzing microsatellites on agarose gels and bulking leaf samples for large-scale diversity study is investigated. The results indicated that bulking leaves from 15 randomly selected individuals per accession proved to be useful for the creation of representative templates for DNA extraction, thereby reducing the cost of DNA extraction and subsequent genotyping.

In chapter 3, 4 and 5, detailed phenotypic and marker information were generated for a selected set of traditional Ethiopian highland maize accessions. The question is how can this information be used in an efficient way for phenotypic and genetic classification of the materials? The answer was highlighted in chapter 6, which deals with the agreement between distance estimates based on molecular markers and morphological traits. The results showed that significant correlations between different data set and validate the use of these data to calculate the different diversity statistics for Ethiopian highland maize accessions. In chapter 7, the association of SSRs with quantitative trait variation was investigated in the selected set of highland maize accessions. The information obtained from this chapter will be useful for future association genetic studies in similar populations. Finally, chapter 8 presents a general discussion of the implication of the results of this study for breeding, future collections and conservation of maize in the highlands of Ethiopia.

CHAPTER 2

LITERATURE REVIEW

2.1 Maize is an important crop for genetic analysis

Maize is a member of the grass family, Poaceae. This family represents a range of genome and structural complexity ranging from diploid species with a genome size of 415 Mb in rice to 16,000 Mb in hexaploid wheat. Maize is diploid (2n = 20). Its chromosomes contain 2.5 billion base pairs and it lies somewhere in the middle of grass genome size and complexity (Gaut *et al*., 2000). Maize has been a major focus of biotechnology research because of its economic importance and naturally occurring high polymorphism.

Maize is cultivated worldwide, at latitudes varying from the equator to slightly above 50 degrees north and south, from sea level to over 3000 meters elevation, in cool and hot climates, and with growing cycles ranging from 3 to 13 months (CIMMYT, 2000). Of the major grain crops, maize has the largest total annual grain production in the world (590.5 million metric tons, mmt) followed by wheat (567.7 mmt) and rice (380.3 mmt) and its average yield per hectare (4.3 t) is more than 60% higher than that of either wheat or rice (FAPRI, 2003). In addition to its direct use as food and feed, maize also has various industrial processing products, such as wet milling (e.g. starch, and oil), dry milling (e.g. meal and flour), and fermentation and distillation (e.g. alcohol and whisky).

Maize is one of the crop species with the highest level of molecular polymorphism. For certain loci, over 5% nucleotide diversity has been reported (Henry and Damerval, 1997). Nucleotide diversity is measured as the average sequence divergence between any two individuals for a given locus (Buckler and Thornsberry,

2002). Maize is also used as model to study genome size evolution because of its polyploidy origin and the abundance of transposons (Bennetzen and Devos, 2002). Transposons are segments of DNA that can move around to different positions in the genome. In the process, they may cause rearrangements or insertion and deletions and thereby change the amount of DNA in the genome. The use of transposons to isolate genes has made it possible to overcome the difficulty of working with a large and mostly repetitive genome. Because of this, maize is used as one of the model crops for studying the biology of cereals, which provide over 70% of the total human caloric intake worldwide (FAPRI, 2003).

2.1.1 Diversity in maize

Zea is the genus for maize (*Zea mays ssp. mays*) and its wild relatives, teosinte. There are four species (*Zea diploperennis*, *Zea perennis*, *Zea luxurians*, and *Zea mays*) in this genus and they are all native to Mexico and Central America (Doebley and Iltis, 1980). Four sub-specific taxa have been recognized in *Z. mays*, namely, *Z. mays ssp. huehuetenangensis*, *Z. mays ssp. mays*, *Z. mays ssp. mexicana*, and *Z. mays ssp. parviglumis*. Among the three wild subspecies of *Zea mays*, *Z mays ssp. parviglumis* is thought to be the progenitor of cultivated maize (Doebley *et al.,* 1984; Wang, *et al*., 1999). Similarly, using microsatellite-based phylogenetic analysis of 264 of maize and its progenitor, (Matsuoka *et al*., 2001) showed that maize was domesticated from *spp. parviglumis* in southern Mexico about 9,000 years ago.

Genetic diversity studies in maize have shown that maize is highly variable both

within and across populations. Sequencing of the *adh1* locus in maize, in *Z. mays* ssp. *parviglumis* (the maize progenitor), and in *Zea luxurians* (a distant maize relative), showed that maize retained 77% of the diversity of *parviglumis* and has more diversity than *Z. luxurians* (Eyre-Walker *et al.*, 1998). Maize molecular diversity is roughly 3 to 10 fold higher than that of other domesticated grass crops (Buckler *et al.*, 2001). This is probably the result of several factors: (1) the diversity of environments, culture, production system and the type of consumption of maize (Aguirre *et al*., 1998), which facilitated the development of different maize types throughout the world; (2) the high level of out-crossing. Maize is a monoecious crop with male and female reproductive parts that are physically separated, which facilitates out-crossing. This favors continuous gene exchange between neighboring plants and in some cases, with their wild relatives; (3) the existence of chromosomal duplications. In maize chromosomal duplications are extensive, which provides new mutational opportunities for creating greater phenotypic variability (Helentjaris *et al*., 1988); and (4) the presence of transposons and retrotransposon elements (Bennetzen and Devos, 2002). Breeders rely on the natural diversity found within crop species for selection and improvement of qualitative and quantitative traits. As a result, maize yields have increased to 55 fold higher than its progenitor (Buckler *et al*., 2001).

2.1.2 Maize genome evolution

Gene or whole genome duplications (polyplodization), insertions of viral DNA, microsatellite and heterochromatin expansions, and transposon insertions all add to nuclear genomes (Bennetzen and Devos, 2002). The maize genome contains extensive

chromosomal duplication and repetitive DNA. Most repetitive DNA in the maize genome comprises retrotransposon elements (a mobile segment of DNA, which uses RNA as a template for replication), and these repetitive DNA comprise 50% of the genome (Gaut *et al*., 2000). Repetitive DNA is defined as DNA with more than 100 copies per genome. The repetitive DNA of maize can further be categorized as 20% highly repetitive (over 800,000 copies per genome) and 40% middle repetitive (over 1000 copies per genome; Hake and Walbot, 1980).

Polyploidization is a major force in plant genome evolution. It has been estimated that 50-70% of flowering plants have experienced chromosome doubling at least once in their evolutionary history (Wendel, 2000). Restriction fragment length polymorphism mapping studies have shown that many markers map to two or more chromosomal locations (Helentjaris *et al*., 1988). Ahn and Tanksley (1993) reported that 72% of the single-copy rice genes are duplicated in the maize genome. Gaut (2001) found that 60-82% of the maize genome is statistically significant for colinearity (shared markers in shared order) and nearly a third of the genome may be even in multiple copies. These findings of extensive chromosomal duplication in maize have been interpreted as evidence for a polypolid origin of the genome (Helentjaris *et al*., 1988).

2.1.3 Wild relatives of maize

The genus *Zea* consists of four species of which only *Zea mays* ssp. *mays* L. is economically important. The other *Zea* species, referred to as teosintes, are wild grasses (Doebley, 1990). These species, mostly perennials, contain a number of useful

genes. Attempts to transfer apomixis (asexual reproduction of a plant through seed) genes from *Tripsacum* to maize have been pursued for a number of years (Leblanc *et al*., 1995; Grossniklaus *et al*., 1998), and consequently patents on apomictic maize have been published (Savidan *et al*., 1998; Eubanks, 2000). Apomixis may be of great significance to the maize growing world. In developing countries, many farmers cannot take advantage of hybrid technology because hybrid seed is either unavailable or unaffordable. If apomixis could be applied in maize, farmers would have the opportunity to recycle hybrid seed from generation to generation, thereby avoiding the cost of buying new seed each season.

In Africa, the parasitic weed *Striga spp*. is a significant pest of maize and sorghum. Little resistance has been found within cultivated maize (Hoisington *et al*., 1999). A potential valuable source of resistance to *Striga hermonthica* may lie in the genetic potential of a wild relative of maize (Tanksley and McCouch, 1997). In conclusion, wild relatives of maize represent significant untapped genetic resources for the improvement of maize.

2.2 Methods for assessing genetic variation

Knowledge of the genetic variation in crop collections is essential for their efficient use in breeding programs, as well as to establish new collections and conservation strategies. Exploiting natural variation is very important for several reasons: (1) genetic variability in crops is advantageous, because it allows the crop to adapt to new biotic and abiotic stresses, (2) many landraces and wild relatives of crop plants contain desirable genes that

confer resistance to pests and diseases, and control quality traits. For example, approximately, 40-80% of the yield gain in maize, wheat and barely has been obtained from genetic improvements of these crops (Evans and Evans, 1993; Hallauer and Miranda, 1988). Consequently, large numbers of varieties are being collected around the world in an effort to conserve the genetic variation and provide access to valuable material for plant breeders. The international center for maize and wheat improvement maintains 17,000 maize accessions (CIMMYT, 2000). As the number of accessions increases, it becomes more difficult to avoid the inclusion of duplicate or at least very similar accessions. Evaluation of numerous, highly similar accessions not only wastes plant breeding resources but likely reduced the chance of identifying the truly unique and valuable accessions. In addition, field evaluation of the whole collection for a variety of traits is difficult because it is laborious and time consuming.

To evaluate and utilize these collections, it is necessary to identify a smaller subset or core collection that likely represents most of the genetic variation in the entire collection. Brown (1989a & b) suggested that at least 70% of the alleles present in the entire collection would be represented in a core collection comprised of at least 10% of the accessions, provided that the selection of the core collection is carried out systematically to capture most of the diversity. To assemble a core collection, numerical methods may be useful for directing the selection of accessions. The data could be agro-morphological performance, pedigree relationships or molecular marker information (genetic diversity).

2.2.1 Morphology and pedigree data

Morphological traits were among the earliest genetic markers used in germplasm management (Stanton *et al*., 1994) but they have a number of limitations, including low polymorphism, low heritability, late expression, and vulnerability to environmental influences (Smith and Smith, 1992), which in turn may affect the estimation of genetic relationships. Therefore, to be useful, morphological measurements should be accomplished in replicated trials. This may be expensive and time consuming. However, if the traits are highly heritable, morphological markers are one of the choices for diversity studies because the inheritance of the marker can be monitored visually without specialized biochemical or molecular techniques. In maize, qualitative and quantitative traits have been used to establish core collection (Taba *et al*., 1998) and to study phenotypic diversity (Alika *et al*., 1993; Lucchin *et al*. 2003).

To quantify the relationship based on pedigree information, Malecot (1948) presented the coefficient of co-ancestry (*f*), also known as the kinship coefficient or the coefficient of parentage. Pedigrees of varieties are defined as a complete record of relationships traced back to landraces and wild relatives. This measure estimates the probability that two randomly drawn, homologous genes (alleles) from each of two individuals are identical by descent. The measure based on Mendelian inheritance and probability is calculated under several assumptions: (1) the absence of selection, mutation, migration and drift, (2) regular diploid meiosis, and (3) no relationship for individuals without a verified common ancestor (Melchinger, 1993).

Several common features of plant breeding programs cause departures from these assumptions because of (1) intense selection, (2) drift due to small sample sizes, and (3) unknown or incorrect pedigree records (Messmer *et al*., 1993). Despite this, it has been widely used in self-pollinated crop species such as barley, wheat, soybean and peanut to examine the level of genetic diversity and identify major groupings of related cultivars (Martin *et al.,* 1991). Accurate estimation of genetic similarity by co-ancestry requires reliable and detailed pedigree records. However, for many maize inbreds and their progenitors, pedigree records tracing back more than two generations are rare or incomplete and calculation of co-ancestry for maize is not feasible (Messmer *et al*., 1993).

2.2.2 Molecular markers

Molecular markers are useful tools for assessing genetic diversity among germplasm compared with morphology and pedigree information because they are not affected by environmental factors. A molecular marker is a variant of DNA or a protein which can be detected and whose inheritance can be monitored reliably (Jones *et al*., 1997). Compared with morphological and pedigree information, molecular markers reveal differences among genotypes at the DNA level and thus provide a more direct, reliable and efficient tool for germplasm conservation and management. As a result, researchers are adopting molecular markers as valuable tools for genetic diversity studies in many crops.

In past decades, marker systems such as Restriction Fragment Length Polymorphisms (RFLPs), Randomly Amplified Polymorphic DNA (RAPDs, Welsh and McClell, 1990; Williams *et al*., 1990), Amplified Fragment Length Polymorphisms (AFLPs, Vos *et al., 1995*), microsatellites or Simple Sequence Repeats (SSRs, Tautz, 1989), single nucleotide polymorphisms (SNPs) and others have been developed and applied to a range of crop species. In general two basic types of marker systems are available (1) those that rely on hybridization between a probe and homologous DNA segment within the genome, and (2) those that use polymerase chain reaction (PCR) to exponentially amplify genome segments between arbitrary or specific oligonucleotide premier sites (reviewed by Karp *et al*., 1996; Jones *et al.,* 1997; Kumar, 1999).

RFLP analysis was one of the first techniques to be used widely to detect variation at sequence level. It examines the variation in size of specific DNA fragments following digestion with a restriction enzyme. RFLP is co-dominant and hence we can distinguish the heterozygote from homozygote individual (Helentjaris *et al*., 1985). RFLPs have been used in maize to investigate pedigree relationships among inbreds and to assign them to heterotic groups (Melchinger, 1993; Dubreuil *et al*., 1996), investigating genetic diversity and relationships (Pejic *et al.,* 1998; Rebourg *et al*., 2001; Gauthier *et al*., 2002) and for the development of genetic maps (Helentjaris *et al*., 1986; Gardiner *et al.,* 1993; Coe *et al*., 1995). However, a disadvantage of RFLPs is that large quantities of DNA are required, which limits the number of marker assays that can be performed on an individual plant and the technique is difficult to automate. As a result, it is increasingly substituted by other marker techniques based on the polymerase chain reaction (PCR) such as RAPDs, AFLPs and SSRs (Jones *et al*.,

1997), because these systems allow essentially unlimited marker assays per individual.

RAPD technology is another procedure used to detect nucleotide sequence variation This PCR-based technique requires neither cloning nor sequencing of DNA. It can detect several loci simultaneously. Short (8-12 bp) arbitrary primer sequences are used to amplify DNA, usually resulting in presence/absence of polymorphisms. Moeller and Schaal (1999) studied genetic variation among 15 Native American maize accessions and found an average polymorphism of 70.7% for the 11 primers analyzed. Although, RAPD analysis is easy, inexpensive and fast, its reproducibility is problematic due to the short primers being easily affected by low annealing temperatures (Demeke *et al.,* 1997; Karp *et al*., 1997).

The AFLP technique combines the restriction site recognition element of RFLP analysis with the exponential amplification aspects of PCR-based markers. It is similar to RAPD analysis, but the primer consists of a longer fixed portion (about 15 bp) and a short (2-4 bp) random portion. The fixed portion gives the primer stability (and hence enhances repeatability) and the random portion allows it to detect a specific subset of loci. Other advantages of the AFLP technique include: (i) no sequence information is required, (ii) the PCR technique is fast, and (iii) it has a very high multiplex ratio (up to 100 genetic loci may be simultaneously analyzed per experiment). This makes it suitable for large-scale genetic diversity studies in crop species. However, AFLP and RAPD are dominant markers, which prohibits the identification of heterozygote from homozygote. This makes AFLP and RAPD less

informative than other co-dominant markers (e.g. RFLP and SSR markers).

In maize, AFLP markers have been employed (i) to investigate the relationship between genetic distances and hybrid performance for yield (Ajmone-Marsan *et al.,* 1998; Melchinger *et al.,* 1998), (ii) to study the genetic similarity of inbreds (Pejic *et al*., 1998; Lubberstedt *et al.,* 2000; Vuylsteke *et al.,* 2000b), (iii) to identify chromosomal regions involved in hybrid performance and heterosis (Vuylsteke *et al*., 2000a) and for construction of genetic linkage map (Vuylsteke *et al.,* 1999).

SSR markers has been a marker system of choice for population genetic studies, because it combines many desirable properties including co-dominance, high variability, rapid and simple assays, and uniform genome coverage (Powell *et al*., 1996). In addition, automated PCR-based technique, which enables high-throughput data collection and good analytical resolution at a low cost, has been developed for microsatellites (Mitchell *et al*., 1997). Because of these qualities, it is frequently applied in genetic diversity studies in maize inbred lines and out-bred populations (Senior *et al*., 1998; Matsuoka *et al*., 2002; Warburton *et al*., 2002; Pinto *et al*., 2003) and to identify and map quantitative trait loci (QTLs) for grain yield and yield components in maize (Thornsberry *et al*., 2001; Mohammadi *et al*., 2002).

One advantage of microsatellite analysis is the large number of polymorphisms that the method reveals per locus, increasing the informativeness of SSR markers. A locus in maize can have up to 16 alleles (Warburton *et al*., 2002). The high allelic diversity is a product of their high rate of stepwise mutation due to replication slippage

(Levinson and Gutman, 1987). The stepwise mutation model assumes that alleles mutate back and forth by small number of repeats, and thus the same allelic state are created repeatedly over time. An alternative model is the infinite alleles model (Ohta and Kimura, 1973), which assumes that each mutation creates a new allele in the populations. Matsuoka *et al*. (2002) reported that out of 46 maize microsatellite loci analyzed on all the diploids of *Zea* and 101 maize inbreds, only two followed stepwise allelic distribution, while four were nearly stepwise, 13 mixed (stepwise and continuous), eight nearly continuous and 19 continuous.

In recent years, SNPs (single base pair positions at which different sequence alternatives exist between two individuals) have become an increasingly important class of molecular marker due to its abundance (present in all parts of the genome) and amenability to fully automated genotyping (micro-array procedures have been developed for automatically scoring hundreds of SNP loci simultaneously at a low cost per sample). A high throughput assay for the detection and validation of SNPs were developed in maize. These techniques allow the rapid production of valuable information on the genetic relationships among maize varieties.

The marker system of choice depends on the objective of the study, skills and facilities available in the laboratory. The relative advantages and disadvantages of these techniques are summarized in Table 2.1.

Table 2.1 Comparison of the most common used marker systems in plant breeding

Characteristics	RFLPs	RAPDs	AFLPs	SSRs	SNPs
DNA required (µg)	10	0.02	0.5-1.0	0.05	0.05
DNA quality	High	High	Moderate	Moderate	High
PCR-based	No	Yes	Yes	Yes	Yes
Level of polymorphism	High	Moderate	High	Very high	High
Ease of use	Not easy	Easy	Easy	Easy	Easy
Amenable to automation	Low	Moderate	High	High	High
Reproducibility	High	Unreliable	High	High	High
Development cost	Low	Low	Moderate	High	High
Cost per analysis	High	Low	Moderate	Low	Low

2.3 Use of pooled DNA samples in the study of genetic variation

Genetic variation is important in the process of crop improvement and is also the basis of genetic fingerprinting. Accordingly, there has been an interest in studying genetic variation through the introduction of different DNA-based marker techniques. Although most marker techniques are relatively simple and rapid, the large number of individual plants that need to be processed may limit the application of DNA-based marker analysis of entire germplasm collections. Because DNA-based marker analysis is quite expensive, the total cost of any such project usually limits the number of genotypes that can be analyzed. The problem is most acute for out-crossing species. Crossa *et al*. (1993) showed that for out-bred maize varieties, with 48 individuals per population, with 5 loci and 5 alleles per locus, there is a 95% probability of detecting all alleles with a frequency of 0.05 or greater. If only 24 individuals are analyzed, only alleles with frequency of 0.12 or greater can be detected at this level of

probability. Hence, genotyping of open-pollinated species using DNA-based markers is 24 to 48 times more expensive than that of self-pollinated species.

One approach to overcome this limitation is to analyze one, or several, bulked samples per population, rather than individual plants. Bulking of DNA samples not only drastically reduce the number of samples that need to be processed, but also results in dilution of rare alleles (Michelmore *et al*., 1991), and therefore simplifies the marker profile of an individual population. Bulking strategies provide a means for large-scale diversity analysis in out crossing plant species. Guthridge *et al*. (2001) in their genetic diversity analysis of perennial ryegrass using AFLP recommended pooling of DNA samples in order to ensure equivalent representation in the AFLP template. Furthermore, parallel studies in white clover (Kolliker *et al*., 2001) demonstrated that bulking at the leaf stage is effective in producing representative profiles of varieties. Both studies have found that bulks of 20 individuals for perennial grass were adequate to study within and between population variations. Similarly, using bulk RFLP methods (two 15-plant bulks per population), Rebourg *et al*. (2001) described the genetic relationships among the 131 European maize populations. All of these results indicate the feasibility of bulking DNA or leaf samples from 15-30 individuals per accession/population as a cost efficient and effective means of characterizing open-pollinated crops.

2.4 Correlation between phenotypic and molecular markers distance

The use of different molecular markers to evaluate genetic diversity may reveal

different patterns of variation due to inherent differences among marker systems. Differences detected by molecular markers are not necessarily correlated with phenotypic variation, because molecular markers can potentially cover the entire genome (coding as well as non-coding regions), and most of the genome is composed of non-coding DNA, it is reasonable that the majority of differences detected by molecular markers are from non-coding regions, while phenotypic differences are brought to specific genes or coding regions. Therefore, the combination of morphological and molecular information is required to describe correctly the relationships among genotypes.

Different researchers have studied the relationship between marker and morphological information. Theoretical results of Burstin and Charcosset (1997) suggested that the relationship between morphological and marker distance is most likely triangular. This means close genetic relationships correspond with close morphological relationships, whereas distant genetic relationships can correspond with both close and distant morphological relationships. In many cases, the correlation between distances based on morphology and molecular markers are not straightforward to interpret. Consequently, a combination of morphological and molecular analyses may be the most useful to understand all aspects of genetic variation within a species or populations. Based on this approach, Rebourg *et al*. (2001) classified European maize populations into six major groups that were consistent with the origin of the populations.

2.5 Statistical measures for assessing genetic diversity

Classifying genotypes into clusters based on molecular markers and agro-morphological traits for studying genetic and phenotype diversity is a common practice. Once the morphological traits or molecular profiles have been generated, various genetic distance measures have been proposed. Genetic distance is defined as any quantitative measure of genetic difference calculated between individuals, populations or species (Beaumont *et al*., 1998).

2.5.1 Types of distance measures

Genetic distance between individuals can be calculated by various statistical measures depending on the type of data. The Euclidean distance (straight-line) and squared Euclidean distance are two commonly used measures of dissimilarity between individuals based on morphological data. Dissimilarity coefficients estimate the distance or unlikeness of two individuals, the larger the values the more different the two individuals. While similarity indices measures the amount of closeness between two individuals, the larger the value the more similar the two individuals. For molecular data, the commonly used measures of genetic similarity (GS) are (i) Nei and Li's (1979) coefficient (GS_{NL}), (ii) Jaccard's (1908) coefficient (GS_J), (iii) Simple matching coefficient (GS_{SM}) (Sokal and Michener, 1958), and (iv) Modified Roger's distance (GS_{MR}). Genetic distances between two individuals *i* and *j* determined by these measures can be obtained as follows:

$$GD_{NL} = 1-[2N_{11}/(2N_{11}+N_{10}+N_{01})]$$

$GD_J = 1-[N_{11}/(N_{11}+N_{10}+N_{01})]$

$GD_{SM} = 1-[(N_{11}+N_{00})/(N_{11}+N_{10}+N_{01}+N_{00})$

$GD_{MR} = 1-[(N_{11}+N_{10})/2N]^{0.5}$

Where N_{11} is the number of bands/alleles present in both individuals; N_{00} is the number of bands absent in both individuals; N_{10} is the number of bands present only in individual *i*; N_{01} is the number of bands present only in individual *j*; and N represents the total number of bands. The GS_{NL} formula excludes bands absent in both individuals, which cannot be necessarily attributed to a common cause. In contrast, GD_{SM} gives equal weight to mismatches and matches of bands in both individuals (Link *et al*., 1995).

Most researchers use more than one measure of genetic distance to analyze a given data set. In such case, it is important to test the correlation between matrices derived from different distance measures. One such test is the Mantel test (Mantel, 1967). The Mantel test can be performed on dissimilarity or similarity matrixes and can be applied to different types of variables. This is especially important for the analysis of genetic diversity, where various types of data sets (e.g. morphological, biochemical or molecular markers) may be used to assess the relationships among individuals. The significance of correlation can be tested via permutation procedure (Manly, 1991).

2.5.2 Multivariate methods

Multivariate techniques, which simultaneously analyze multiple measurements on

each individual under study, are widely used in analysis of genetic diversity in morphological and molecular marker data. Among these methods, cluster and principal components analyses are most commonly used. Cluster analysis refers to a group of multivariate techniques, whose primary purpose is to group individuals based on the characteristics they possess so that individuals with similar descriptions are mathematically gathered into the same cluster (Hair *et al*., 1995). The resulting cluster of individuals should then exhibit high within cluster homogeneity and high between cluster heterogeneity. Principal component analysis (PCA) is defined as a method of data reduction to clarify the relationships between two or more characters and to divide the total variance of the original characters into a limited number of uncorrelated new variables (Wiley, 1981). PCA can be used to drive a two-dimensional scatter plot of individuals, such that the geometrical distance among individuals in the plot reflect the genetic distances among them with minimal distortion. Aggregations of individuals in such a plot will reveal sets of genetically similar individuals (Warburton and Crossa, 2000).

Principal Coordinate Analysis (PCO) is another data reduction method commonly used by breeders and geneticists. The goal of PCO is to permit the positioning of objects in a space of reduced dimensionality while preserving their distance relationships as much as possible. The value of PCO is that it permits the use of all types of variables, provided that a coefficient of appropriate type has been used to compute the resemblance half-matrix. PCO differs from PCA in the way in which the data swarm is constructed to begin with. In PCO, the points are not plotted in an *s*-dimensional coordinate frame. Instead, dissimilarities are calculated between every

possible pair of objects, and the points plotted in such a way as to make the distance between every pair of points as nearly as possible equal to their dissimilarity. One could argue that PCO is necessarily inferior to PCA because in PCA each point is placed exactly where it ought to be, whereas in PCO each point is only approximated based on a best-fit model of the dissimilarities.

2.5.3 Clustering methods

There are broadly two types of clustering methods: (1) distance-based methods, in which a pair-wise distance matrix is used as input for clustering analysis (Johnson and Wichern, 1992). The result can be visualized as a tree or dendrogram in which clusters may be identified, and (2) model-based methods in which observations from each cluster are assumed to be random draws from some parametric model, and inference about parameters corresponding to each cluster and cluster memberships of each individual are performed jointly using maximum-likelihood or Bayesian methods.

Distance-based methods can be further categorized into hierarchical and non-hierarchical. Hierarchical clustering is performed by a serious of successive mergers (agglomerative) of groups of individuals. The most similar individuals are first grouped and these initial groups are merged according to their similarities. The Unweighted Paired Group Method using Arithmetic averages (UPGMA, Sneath and Sokal, 1973) and Ward minimum variance methods (Ward, 1963) are the most commonly used agglomerative hierarchical clustering methods. The non-hierarchical

clustering procedures, also known as K-means clustering methods, are based on sequential threshold approaches for assigning individuals to specific clusters after the number of clusters to be formed is specified (Everitt, 1980). This method is rarely used for genetic diversity study because the lack of prior information about the optimal number of clusters that is required for accurate assignment of individuals.

2.5.4 Partitioning of variation

When a set of populations is investigated, the amount of genetic variability can be expressed at different hierarchical levels, e.g., between agroecologies, between populations within agroecologies and within populations. For molecular data, the analysis of molecular variance (AMOVA, Excoffier *et al*., 1992) has been widely used (Warburton *et al*., 2002; Reif *et al*., 2003) for estimation of the variance components among and within the group. AMOVA is based on squared Euclidean distances among individuals, and assumes that the studied populations are in Hardy-Weinberg equilibrium. A similar method is known as analysis of distance (AOD, Van-Eeuwijk and Baril, 2001), which can be applied for any distance matrix be it Euclidean or not, and for any type of marker (morphological, molecular or a mixture) using the following formula:

$$d^2_{gi;g'i'} = \sum_{m=1}^{M} \left(x_{mgi} - x_{mg'i'}\right)^2,$$

Where d^2 is the distance between an individual *i* in group *g* and an individual *i'* in group *g'* for marker m (from 1 to *M).* Accordingly, the squared distance is the sum of the squared differences between individual accessions over all variables. The total variation (V_T) can be shown to be equal to the sum of all squared pair-wise distances between individual accessions (over all groups), divided by the total number of accessions

(overall groups). The within group variation (V_W) is the sum over groups of the sum of squared pair-wise distances within a group divided by the group size. The between group variation (V_B) can be obtained by subtraction, $V_B = V_T - V_W$.

2.6 Gene mapping/tagging

The identification and use of major genes controlling quantitative traits (QTLs) have been a major focus in maize breeding. For simply inherited traits, a difference between parents in one or two genes may explain nearly 100% of the differences among the progeny. However, many agriculturally important traits such as yield, quality and plant height show continuous variation among individuals and such traits are termed quantitative traits. In maize, many quantitative traits have been extensively investigated using conventional biometric approaches (Hallauer and Miranda, 1988). The concept of detecting QTLs was developed more than 80 years ago (Sax, 1923). However, the development of genetic markers has had great impact in the field of quantitative genetics, mainly for identifying the chromosomal segments or individual genes underlying a quantitative trait (Kumar, 1999; Stuber *et al*., 1999, Bouchez *et al*., 2002).

Molecular markers also permit plant breeders to correctly map or place the various genes that condition complex agronomic traits. Mapping is putting markers in order, indicating the relative genetic distances between them, and assigning them to their linkage groups on the basis of the recombination values from all their pair-wise combinations (Jones *et al*., 1997). Genetic mapping is essential for effective

manipulation of important genes, QTL detection, comparative mapping, detection of chromosome duplications and marker-assisted selection. Genetic linkage maps have been constructed for maize using RFLPs (Helentjaris *et al*., 1986; Gardiner *et al*., 1993; Coe *et al*., 1995; Lee *et al*., 2002), AFLP (Vuylsteke *et al*., 1999), and SSRs (Sharopova *et al*., 2002). Also, sequenced cDNAs (also known as ESTs, expressed sequence tags) become a source of molecular markers and have now been integrated into maize genetic maps (Causse *et al*., 1996; Davis *et al*., 1999). These maps make it possible to locate genes and map QTLs in the maize genome.

2.6.1 Conventional method of QTL detection

QTL mapping is defined as association between observed trait values and the presence/absence of alleles of markers that have been mapped onto a linkage map. The established methods of detecting QTLs involves the selection of two parents that differ distinctly in a particular character and then determination of association between markers and that character in $F_{2,}$ and backcross progeny. If the correlation between the phenotype and alleles of the marker is significantly different from zero, then a QTL is detected.

In maize, a genetic map has been constructed from a large recombinant inbred line (RIL) population to increase the resolution of mapping (Lee *et al*., 2002). RILs are produced by inbreeding individual F_2 progeny up to six times to make them homozygous at any locus. Each RIL is thus fixed for short linkage blocks of parental alleles (Burr *et al*., 1988). RIL populations constitute permanent mapping populations

and can be used by different researchers in varying environments and the information can be added to a common database. As a result, the public maize-breeding sector has been able to develop detailed QTL and single gene maps for a number of traits (reviewed in Hoisington and Ribaut, 1998; Tuberosa *et al*., 2002). Despite these efforts, the resolution for many QTL maps is still several centimorgans (cM), corresponding to hundreds of genes.

Major shortcomings of QTL detection experiments include: (1) the limited number of recombination events per generation results in poor resolution, (2) only two alleles at any given locus can be studied simultaneously, (3) the number of location and effects of the identified QTLs vary according to the genetic background of the population, and (4) it is neither cheap nor fast (Tuberosa *et al*., 2002; Flint-Garcia *et al*., 2003). However, a number of alternative approaches of QTL detection are available whose application can contribute to partially circumvent some of the limitations discussed above.

2.6.2 Bulk segregant analysis

A cheaper and faster alternative to conventional QTL detection is bulk segregant analysis (BSA, Michelmore *et al*., 1991), which has been shown to work well with genes having major effects and that obviate the need for constructing detailed genetic map. For BSA of the trait of interest, parental lines are chosen that differ in their expression and crossed, and F_2 or RIL populations are generated which will segregate for the trait. The population is then phenotyped to identify individual plants or lines

having high or low expression of the trait. Two DNA bulks are prepared, one from the 'high' individuals and the other from 'low' individuals, and analyzed for allelic frequency with molecular markers. With dominant markers, only a few individuals are required in each bulk. The probability of an unlinked locus being polymorphic between two bulks of 10 individuals was calculated to be 2×10^{-6} (Michelmore *et al*., 1991). However, when using co-dominant markers (such as RFLPs and SSRs) with pools of genetically diverse individuals, where several marker alleles may be present, at least 50 individuals need to be combined to make each bulk (Quarrie *et al*., 1999)

The BSA can be used whether the individuals come from a single segregating population or from pools of genetically diverse individuals, such as variety mixtures or composite populations of out-breeding species such as maize. In maize, BSA has been efficiently used for the identification of QTLs for flowering time and yield (Tuberosa *et al*., 1998; Quarrie *et al*., 1999). However, a major shortcoming of BSA is that no information is provided on the distance of the QTL from the polymorphic marker, therefore, markers obtained in a BSA need to be mapped with standard approaches.

2.6.3 Association mapping

Another approach for QTL detection is association mapping. It is a population-based method used to identify marker-trait relationships based on linkage disequilibrium (LD, Remington *et al*., 2001). Linkage disequilibrium or allelic association is defined as the nonrandom association of alleles at different loci. Linkage refers to the

correlated inheritance of loci through the physical link on a chromosome, whereas LD refers to the correlation between alleles in a population.

Association and quantitative trait locus (QTL) studies suggested that the maize gene *Dwarf8* might affect the quantitative variation of maize flowering time (Thornsberry *et al*., 2001). In wheat this gene has contributed to yield increments seen in the 'Green Revolution' varieties and the *Arabidopsis* ortholog has been shown to play a role in regulating flowering time variation (Wilson *et al*., 1992). Similarly, association-mapping studies using RAPDs on genetically diverse rice germplasm (Virk *et al*., 1996) have identified markers associated with a number of characters, such as flowering time and panicle length.

The potential advantages of association mapping over conventional mapping are (i) only polymorphisms with extremely tight linkage to a locus with phenotypic effects are likely to be significantly associated with the trait in a randomly mating population, providing much finer resolution than genetic mapping (Remington *et al*., 2001), (ii) QTLs for any quantitative trait can be studied in the same investigation (Vuylsteke *et al*., 2000a), and (iii) detection of QTLs that vary across a wide spectrum of the gremplasm rather than just between two parental lines (Virk *et al*., 1996).

2.6.4 Comparative genetic mapping

One of the applications of genetic mapping is the comparison of genome colinearity and synteny within and between related crop species. Colinearity is defined as the conservation of gene content and order between two or more species, while synteny is

defined as the conservation of linkage on chromosomes, in the absence of a defined order (Bennetzen and Devos, 2002). Because of their conserved genetic nature, some DNA markers can be used in genetic mapping of the species of origin and closely related species. For example, species of the Poaceae (maize, sorghum, rice, oat and wheat) share conserved gene collections. Of 150 maize RFLP markers tested, only one failed to hybridize to sorghum DNA (Hulbert *et al*., 1990). About 85% of rice, oat and barely cDNA clones showed hybridization to maize DNA (Ahn and Tanksley, 1993). Hence, the same set of RFLP probes derived from a single species can be used for genetic mapping in related species. Thus, it is possible to compare linkage maps and determine whether the order of markers along the linkage groups is conserved across species.

However, detailed comparisons of genome colinearity and synteny can only be accomplished by comparative physical mapping and sequencing. This will provide new insights into gene and genome evolution, and are powerful tools for gene isolation and characterization. One approach of QTL cloning in maize is based on the identification and mapping of a large number of ESTs whose mapping will provide candidate genes for the QTLs (Davis *et al*., 1999). The maize genome will be sequenced providing the ultimate resources of candidate genes for QTL mapping and cloning.

2.7 Marker-assisted selection and breeding

Conventional plant breeding is time consuming and very dependent on environmental conditions. Breeding a new variety takes eight to twelve years and even then the

release of an improved variety may not be granted. Hence, breeders are interested in new technologies that could make this procedure more efficient. When selection is based on genetic information through the application of molecular markers it is called marker-assisted selection (MAS). Marker-assisted selection is based on the concept that it is possible to infer the presence of an allele of a gene from the presence of a marker allele that is tightly linked to the gene.

MAS improves selection for quantitative traits because (1) DNA markers can be assayed at the seedling stage, permitting one to make selections before many traits are expressed, thus reducing the number of individuals which must be grown to maturity, (2) many traits may be more accurately selected for by using genotypes at DNA markers than by relaying solely on phenotype which may be due to either genotype or environment, and (3) unlike phenotypic traits, genetic markers can be reliably assayed in non-target environments such as the growth chamber or greenhouse, permitting more rapid progress in breeding.

In principle, once QTLs have been identified, introgression of the favorable alleles and their pyramiding into elite germplasm (e.g. parental lines, populations, etc.) becomes possible through MAS (Ribaut and Hoisington, 1998, Stuber *et al*., 1999). However, only a few successful applications of MAS for improvement of quantitative traits have been described (Ragot *et al*., 2000: Ribaut *et al*., 2000; Bouchez *et al*., 2002) due to mainly to weak associations (in terms of genetic distance) between markers and target QTLs and high cost of MAS (Stuber *et al*., 1999; Moreau *et al*., 2000, Tuberosa *et al*., 2002).

2.7.1 Introgression of desirable genes

Another application of marker-assisted selection is the introgression of desirable genes from wild species into an elite variety. Tanksley and McCough (1997) proposed that wild or unimproved accessions may harbor important genes that can significantly improve yield and other important traits when introgressed into adapted cultivars with the use of DNA markers. In conventional plant breeding, backcross breeding is a well-known method for the introgression of desirable genes from a donor lines into recipient lines. Such components can be transferred to elite cultivated materials by repeated backcrossing. One of the disadvantages of this method is that other genes may also be transferred along with the genes that control the target trait, which may reduce yield or quality of the desired varieties. By the use of markers linked to specific QTLs it is possible to introgress specific regions of the genome that confer desirable quantitative characteristics to an elite variety (Tanksley *et al*., 1996; Harjes *et al*., 1999).

Given the results already produced in maize at the molecular level: linkage map (Davis *et al*., 1999; Lee *et al*., 2002), and QTL analysis (Veldboom *et al.,* 1994; Ribaut and Hoisington, 1998; Tuberosa *et al*., 1998), marker- assisted selection for maize improvement is becoming more and more efficient (Stuber *et al*., 1999; Bouchez *et al*., 2002; Tuberosa *et al.,* 2002).

2.8 Conclusions

Plant breeding relies on genetic variation and uses selection to improve plant productivity. Over 50% of agricultural productivity in the world has been achieved through traditional plant breeding (Kumar, 1999). However, as the human population increases and the expected reduction of available arable land due to climate and human intervention continues, it may be necessary to accelerate the rate at which genetic improvement is achieved.

Modern biotechnology provides new tools that can facilitate the development of improved plant breeding methods and expand our knowledge of plant genetics. The knowledge that is obtained with these new tools can be used to enhance food security throughout the world. Particularly, DNA markers have the potential to enhance the operation of plant breeding programs ranging from fingerprinting of genetic stocks, assessment of genetic diversity, increasing the efficiency of selection, to comparative mapping and manipulation of QTLs. Despite this potential, the current application of crop biotechnology is almost nil in most African countries, because of the lack of resources, trained personnel and infrastructure in this field.

The primary resource of plant breeding programs in Africa is the genetic variability available within landraces or primitive varieties. The success of crop improvement is highly dependent on the power and efficiency with which this genetic variability can be manipulated. However, in many African countries plant breeders still use morphological traits to study genetic diversity and genetic relationships among genotypes.

Morphological differences do not always reflect genetic differences, because of genotype x environment interaction. As a result, the potential of making genetic progress is slow. Therefore, in the future it is necessary to use DNA markers that will provide more rapid and precise information on the extent of genetic diversity and genetic relationships among genotypes than phenotypic selection.

Recent advances in automated marker technology have presented the possibility of efficiently applying marker-assisted selection at the scale of modern plant breeding. Conventional QTL studies are commonly based on the phenotypic and molecular analysis of single genotypes (individual plants or progenies) of a mapping population mostly derived from the cross of inbred lines. However, cheaper and faster alternatives to conventional QTL detection (e.g. BSA and association mapping) can be effectively used in African crops, which do not require the development of RILs and mapping populations.

Large-scale explorations of plant genomes will rapidly narrow the gap in knowledge between the model crops (rice and maize) and lesser-studied African crops (e.g. sorghum, millet and teff). Based on the intensive study of genes for important agronomic characters in rice or maize, it may be possible to make rapid developments of these traits in sorghum and millet breeding. Therefore, simple PCR-based markers (SSR, RAPD and AFLP) are an appropriate entry point to genomics for many African countries, including Ethiopia.

In the highland areas of Ethiopia, maize is the most important crop grown by

subsistence farmers and it is an important "hunger breaking crop" due to the fact that it is often consumed green. It is hypothesized that many of the highland maize varieties have been geographically isolated for long periods of time and may have accumulated specific genetic adaptations for highland conditions. Therefore, it is necessary to study the genetic diversity and genetic relationships among these accessions using morphological and molecular markers in order to (a) understand the distribution of genetic variation in different highland regions, (b) better conserve the genetic variation contained in them and (c) facilitate their use in new, dedicated breeding programs for highland maize.

CHAPTER 3

PHENOTYPIC DIVERSITY FOR MORPHOLOGICAL AND AGRONOMIC TRAITS IN TRADITIONAL ETHIOPIAN HIGHLAND MAIZE ACCESSIONS

South African Journal of Plant and Soil 2005, 22: 100-105

3.1 ABSTRACT

Farmers in the highlands of Ethiopia have developed locally adapted maize varieties for more than 300 years. In order to assess the phenotypic diversity among traditional Ethiopian highland maize accessions, a total of 180 accessions were evaluated for agro-morphological traits in a replicated randomized complete block design. The accessions varied significantly for all of the measured traits. Cluster analysis revealed the presence of four major clusters. Accessions collected from the different regions were distributed over all the phenotypic clusters, reflecting wide variation within a particular region, but low differentiation among regions. The first principal component, which explained 40.4% of the total variation, was due to days to tasseling and silking, plant and ear height, leaf length and days to maturity. Traits directly selected by farmers (yield, kernels per row, rows per ear, and ear height) had the highest phenotypic coefficients of variation (PCV), whereas indirectly selected traits (ear diameter, days to tasseling and silking) showed lower PCV values. Number of kernels per row had high heritability and genetic advance as percent of the mean and could be used as selection criterion to increase grain yield. Overall, the study indicated the existence of ample trait diversity in highland maize accessions, which can be exploited by hybridization and selection.

Key words: Ethiopia, correlation, heritability, highland maize, phenotypic diversity

3.2 INTRODUCTION

Since its early domestication in Mexico, maize was introduced to many regions of the world where it has become adapted to a wide range of climates and agronomic conditions. It is believed that maize was first introduced to Ethiopia in the 16th or 17th century (Hafnagel, 1961). Since its introduction, it has gained importance as a food and feed crop in Ethiopia. Averages of the 2000/2001 national production estimates of the Central Statistical Authority (CSA, 2001) indicate that maize, with 1.4 million ha and 2.52 million t, accounts for about 20.9% of the total area and 32.6% of the gross annual grain production.

The highland zone of Ethiopia covers 20% of the total land devoted to maize cultivation and more than 30% of small-scale farmers in this region depend on maize for their livelihood (CSA, 1998). Despite its importance in the highlands of Ethiopia, only two improved open-pollinated maize varieties were developed in the past 10 years (Twumasi-Afriyie *et al*., 2001). The cultivars grown by highland farmers may not be as productive as commercial hybrids but they possess many useful adaptive traits, which have helped them to thrive in the difficult highland environments. They could, therefore, be useful material in the development of superior hybrids or open-pollinated varieties suited to this region.

In view of this fact, the Highland Maize Germplasm Collection Mission was launched in 1998 throughout the highlands of Ethiopia in collaboration with CIMMYT (Twumasi-Afriyie *et al*., 2001). As part of this project, 287 maize accessions were

collected from farmers' fields throughout the highland regions of Ethiopia. Currently, there is no information on the extent of morphological and genetic variation among these accessions.

The objectives of this study, therefore, were (i) to evaluate and characterize these accessions for agro-morphological traits, (ii) to assess the extent of phenotypic and genotypic variability, heritability (broad sense) and expected genetic advance, and (iii) to classify and identify groups of similar accessions by means of cluster and principal component analysis. This information will be useful to identify genotypic groups for breeding purposes and to select a representative sample for molecular marker analysis.

3.3 MATERIALS AND METHODS

3.3.1 Plant materials and field evaluation

A total of 180 maize accessions collected from the Northern, Southern and Western highlands of Ethiopia were used in this study. The accessions were grown at Alemaya University in Ethiopia during the 2002 main cropping season in a randomized complete block design with two replications. Each accession was grown in two row plots. Each row had 25 plants, which constituted 44444 plants ha $^{-1}$, which is recommended for the testing site. From each accession, 20 competitive plants were selected at random to record 15 agro-morphological traits (Table 3.1).

3.3.2 Statistical analysis

The mean values of sampled observations for 15 agro-morphological data were analyzed using SAS (SAS, 1993). Statistical measures of variability such as genotypic coefficient of variability (GCV), phenotypic coefficient of variability (PCV), broad sense heritability (h^2), genetic advance as percent of the mean (GA), and genotypic and phenotypic correlations were calculated according to Singh and Chaudhary (1977). The standardized mean values (mean of each trait was subtracted from the data values and divided by the standard deviation) were used to perform cluster analysis (CA) and principal component analyses (PCA) using NCSS 2000 (Jerry, 2000) statistical software. The unweighted pair group method with arithmetic average (UPGMA) was used as clustering technique.

3.4 RESULTS

3.4.1 Morphological and agronomic variability

Analysis of variance revealed highly significant differences among the accessions for all the traits studied (Table 3.1). There was a wide range of expression across the accessions for all of the traits, including a 79 day range in maturity, 28 day range in male flowering, 155 cm range in plant height and 251 g range in 1000 seed weight. Grain yield exhibited one of the widest ranges (424.2 to 7313.2 kg ha^{-1}), possibly due to the specific adaptation of these accessions to various highland environments.

3.4.2 Phenotypic and genotypic coefficients of variation, heritability and genetic advance as percent of the mean

The PCV, GCV, h^2 and GA are presented in Table 3.2. Ear height showed the highest genotypic variability followed by number of kernels per row and yield ha^{-1}. The lowest GCV was recorded for ear diameter. The differences between GCV and PCV for all traits, except yield ha^{-1}, leaf width, 1000 seed weight and ear length were small indicating that these traits were less influenced by environment. High h^2 estimates were noted for the morphological traits (number of leaves, days to maturity, tasseling and silking and plant height). The lowest h^2 estimate of 17% was recorded for yield ha^{-1}. The GA that could be expected from selecting the top 5% of the accessions, varied from 10.1% for leaf width to 38.7% for ear height.

Table 3.1 Means, standard deviation of the means (St Dev), ranges and mean squares for 15 agro-morphological traits measured in 180 maize accessions

Traits	Mean	St Dev	Minimum	Maximum	Variety mean square
Days to tasseling	65.8	5.8	48.5	76.0	67.2**
Days to silking	71.1	5.3	56.5	80.5	55.8**
Plant height (cm)	218.3	28.5	155.0	310.0	1621.2**
Ear height (cm)	124.8	28.2	71.5	274.5	1584.0**
Leaf length (cm)	71.6	8.3	49.5	100.8	136.7**
Leaf width (cm)	9.00	0.8	6.4	12.8	1.3**
Number of leaves	6.1	0.3	5.2	6.8	0.1*
Foliage rating	6.1	1.0	3.0	7.0	2.0**
Days to maturity	144.4	15.0	108.0	186.5	452.2**
Ear diameter (cm)	3.9	0.3	3.3	4.9	0.5*
Ear length (cm)	18.1	1.7	11.2	22.0	11.4**
Rows per ear (no)	10.9	1.8	6.5	14.0	2.5**
Kernels per row (no)	28.6	5.8	18.0	41.0	30.1**
1000 seed weight (g)	298.1	36.0	159.0	410.0	2804.0*
Yield (kg ha $^{-1}$)	2841.0	17.1	424.2	7313.2	1.9**

** & * Significant at $p = 0.01$, and $p = 0.05$, respectively.

Table 3.2 Estimates of phenotypic and genotypic coefficients of variability, heritability and genetic advance as percent of mean

Traits	PCV (%)	GCV (%)	h^2 (%)	GA (%)
Days to tasseling	9.3	8.3	78.5	17.0
Days to silking	7.9	7.0	77.8	14.3
Plant height (cm)	14.2	11.8	70.1	24.4
Ear height (cm)	25.8	18.8	53.0	38.7
Leaf length (cm)	13.5	9.2	45.8	18.9
Leaf width (cm)	11.6	4.9	17.7	10.0
Number of leaves (no)	13.9	12.9	86.9	26.6
Foliage rating	19.3	12.4	40.9	25.4
Days to maturity	10.8	9.9	84.1	20.4
Ear diameter (cm)	7.3	4.9	44.7	10.1
Ear length (cm)	12.2	5.7	21.6	11.7
Rows per ear (no)	19.4	13.2	46.4	27.2
Kernels per row (no)	22.0	18.3	69.5	37.8
1000 seed weight (g)	15.7	6.7	18.1	13.8
Yield (kg ha^{-1})	40.6	16.7	17.0	13.5

3.4.3 Genotypic and phenotypic correlations

There were significant genotypic and phenotypic correlations among the various traits (Table 3.3). At genotypic level yield was negatively and significantly correlated with all of the morphological traits, but positively and significantly correlated with all of the agronomic traits. When correlations between morphological traits were taken into account, all of the values between various trait pairs were significant and positive at both phenotypic and genotypic levels. Seed weight appears to contribute substantially to yield at the genotypic level, as the strongest positive and significant correlation of yield was recorded for this trait (r= 0.67).

Table 3.3 Genotypic (above diagonal) and phenotypic (below diagonal) correlation coefficients between 15 agro-morphological traits measured in 180 maize accessions

	DTS [a]	DSK	PLH	ERH	LFL	LFW	NRL	ERD	ERL	RWE	KLR	DYM	SDW	FGR	YDH
DTS		0.99	0.74**	0.74**	0.81**	0.95**	0.46**	0.19**	0.62**	0.22**	0.19**	0.74**	0.19**	0.83**	-0.48**
DSK	0.96**		0.76**	0.76**	0.83**	1.01**[s]	0.02 [ns]	0.18**	0.63**	0.24**	0.22**	0.76**	0.20**	0.81**	-0.46**
PLH	0.53**	0.51**		0.99**	0.97**	0.93**	0.49**	0.33**	0.70**	0.52**	0.36**	0.73**	0.41**	0.88**	-0.44**
ERH	0.45**	0.44**	0.77**		0.99**	0.93**	0.47**	0.42**	0.60**	0.56**	0.40**	0.76**	0.65**	0.86**	-0.61**
LFL	0.49**	0.47**	0.63**	0.53**		1.34**	0.52**	0.34**	0.68**	0.52**	0.36**	0.85**	0.51**	0.95**	-0.59**
LFW	0.40**	0.41**	0.40**	0.33**	0.46**		0.09 [ns]	0.67**	0.94**	0.56**	0.49**	0.84**	0.44**	1.02**	-0.64**
NRL	0.14*	0.02 [ns]	0.21**	0.10 [ns]	0.20**	0.08 [ns]		0.26**	0.41**	0.43**	0.24**	0.41**	0.30**	0.83**	0.29**
ERD	0.10[ns]	0.12 [ns]	0.22**	0.24**	0.14*	0.12 [ns]	0.03 [ns]		0.22**	0.40**	0.51**	0.18**	0.77**	0.19**	0.23**
ERL	0.19 [**]	0.21**	0.25**	0.26**	0.16*	0.05 [ns]	0.04 [ns]	0.23**		0.53**	0.39**	0.53**	0.68**	0.50**	0.17*
RWE	0.15*	0.17*	0.29**	0.27**	0.21**	0.22**	0.06 [ns]	0.32**	0.04 [ns]		0.95**	0.29**	0.64**	0.24**	0.23**
KLR	0.17*	0.19**	0.27**	0.27**	0.21**	0.17*	0.07 [ns]	0.34**	0.24**	0.65**		0.18**	0.60**	0.06 [ns]	0.41**
DYM	0.58**	0.60**	0.57**	0.52**	0.53**	0.32**	0.21**	0.09 [ns]	0.26**	0.15*	0.14*		0.39**	0.78**	-0.35**
SDW	0.05 [ns]	0.05 [ns]	0.12 [*]	0.13*	0.10 [ns]	0.16*	0.05 [ns]	0.31**	0.17*	0.19**	0.19**	0.14*		0.20**	0.67**
FGR	0.47**	0.50**	0.52**	0.45**	0.48**	0.33**	0.23**	0.12 [ns]	0.25**	0.08 [ns]	0.06 [ns]	0.49**	0.19**		-0.36**
YLD	-0.19 [**]	-0.19 [**]	-0.08 [ns]	-0.05 [ns]	-0.15*	-0.29**	0.04 [ns]	0.25**	0.08 [ns]	0.14*	0.22**	-0.17*	0.04 [ns]	-0.13*	

[a] DTS - days to tasseling, DSK - Days to silking, PLH - Plant height (cm), ERH - ear height, LFL - Leaf length (cm), LFW - Leaf width (cm), ERD - Ear diameter (cm), ERL - Ear length (cm), RWN - rows per ear (no), KLR - Kernels per row (no.), DYM - Days to maturity, SDW - 1000 seed weight (g), YDH - Yield per hectare (kg).

* Significant at $p = 0.05$, ** significant at $p = 0.01$, [ns] not significant

3.4.4 Cluster analysis

The trait means for the four clusters generated by UPGMA as clustering technique are given in Figure 3.1 and Figure 3.2. Cluster I was the biggest with 73 accessions, 90% of them were collected from the Southern and Western regions of Ethiopia. Accessions in this cluster expressed high values for morphological traits (days to tasseling, silking and maturity) and for all agronomic traits (ear diameter, ear length, rows per ear, kernels per row and grain yield). Cluster II contained 68 accessions (42.6, 27.9 and 29.5%, collected from Northern, Western and Southern Ethiopia, respectively). Accessions in this cluster expressed high values in days to tasseling, silking and maturity (Figure 3.1) but low mean values for all of the agronomic traits (Figure 3.2).

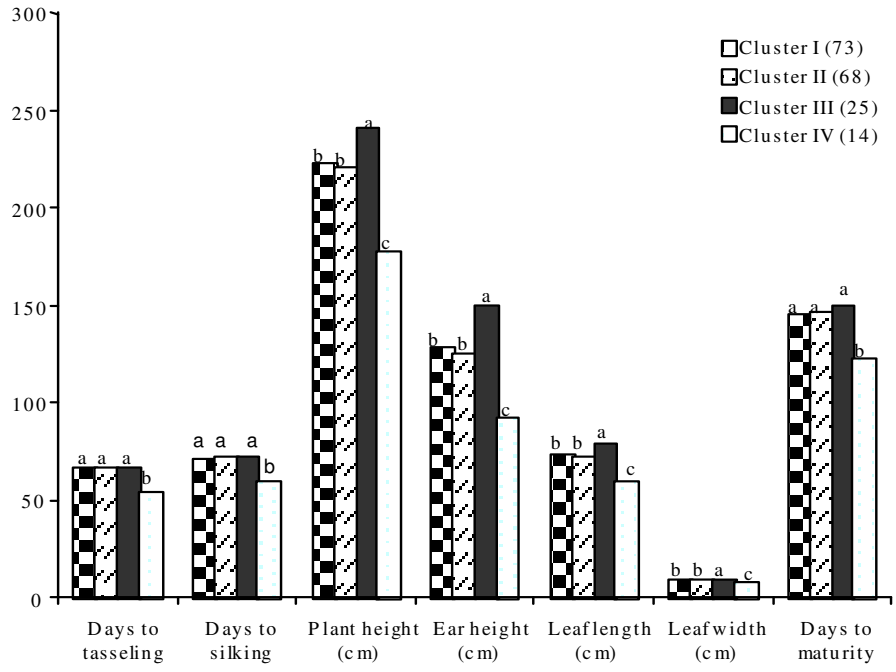

Figure 3.1 Mean values of morphological traits for traditional Ethiopian highland maize accessions for the four clusters generated through UPGMA as clustering techniques. Clusters having the same letter within each trait are not statistically significant at $p = 0.05$. Number in parenthesis (in the legend) indicates accessions

grouped in each cluster.

Cluster III contained 25 accessions (collected equally from Northern, Western and Southern Ethiopia) were tall and late maturing plants that had broad and long leaves. This cluster also gave the lowest mean values for all of the agronomical traits. In contrast, Cluster IV, comprising of 14 accessions, had low mean values for all of the morphological traits. High values were observed for all of the agronomic traits. Seventy percent of the accessions in cluster IV were collected from the Northern part of the country.

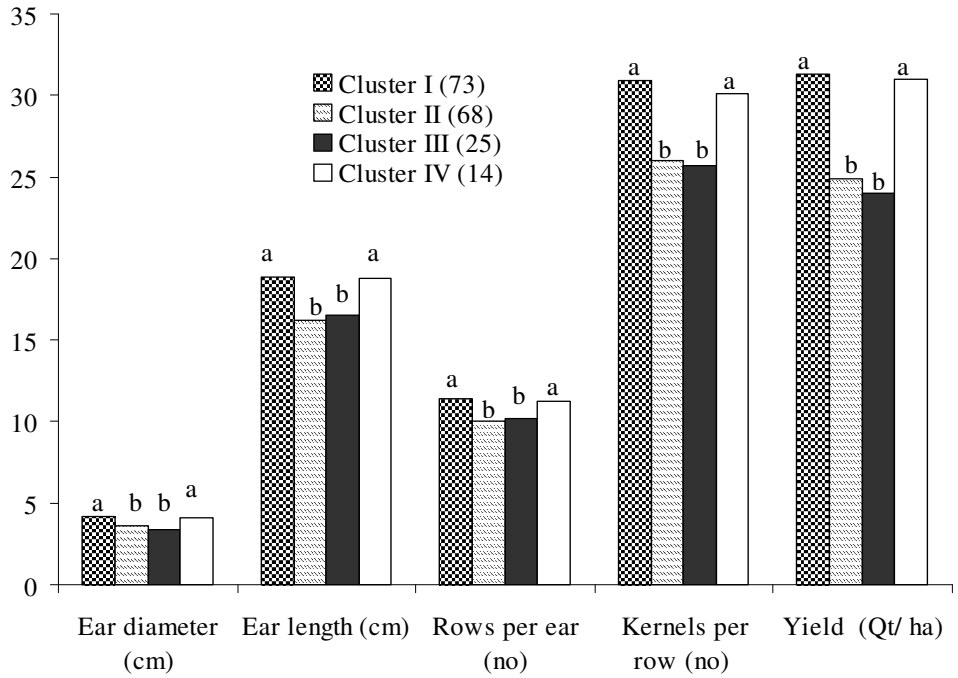

Figure 3.2 Mean values of agronomic traits for traditional Ethiopian highland maize accessions for the four clusters generated through UPGMA as clustering techniques. Clusters having the same letter within each trait are not statistically significant at p = 0.05. Number in parenthesis (in the legend) indicates accessions grouped in each cluster.

3.4.5 Principal component analysis

The first five principal components explained 75.1% of the total variation, with the first three components, with eigenvalues higher than 1.0, accounting for 62.8% of total variation (Table 3.4). Morphological traits such as days to tasseling and silking, plant and ear height, leaf length and days to maturity, were the major discriminatory traits associated with the first principal components axis, which accounted for 40.4% of the total variation, while agronomic traits (number of kernels per row, number of rows per ear, 1000 seed weight, ear diameter and yield) were important traits associated with the second principal component, which accounted for 15% of the total variation. The third principal component, which explained 7.4% of the total variation, was dominated by number of leaves, leaf width and grain yield.

Table 3.4 Eigenvector, eigenvalues, individual and cumulative percentage of variation explained by the first five principal components (PC) after assessing agro-morphological traits in 180 maize accessions

Traits	PC1	PC2	PC 3	PC 4	PC 5
Days to tasseling	-0.33	0.19	0.02	-0.04	-0.29
Days to silking	-0.33	0.18	0.02	-0.04	-0.30
Plant height (cm)	-0.35	0.01	0.01	-0.06	0.03
Ear height (cm)	-0.34	-0.02	-0.05	0.03	0.04
Leaf length (cm)	-0.34	0.07	-0.07	-0.06	0.17
Leaf width (cm)	-0.28	0.06	-0.37	0.08	0.23
Number of leaves	-0.13	-0.02	0.61	-0.49	0.38
Foliage rating	-0.30	0.15	0.26	0.04	0.16
Days to maturity	-0.32	0.13	0.11	0.04	-0.06
Ear diameter (cm)	-0.14	-0.40	0.00	0.29	0.19
Ear length (cm)	-0.19	-0.15	0.28	0.42	-0.49
Rows per ear	-0.18	-0.43	-0.29	-0.39	-0.02
Kernels per row	-0.17	-0.48	-0.23	-0.29	-0.21
1000 seed weight (g)	-0.13	-0.33	0.09	0.50	0.42
Yield (kg ha $^{-1}$)	0.09	-0.42	0.43	-0.06	-0.26
Eigenvalue	6.1	2.3	1.1	1.0	0.9
Individual variation in %	40.4	15.0	7.4	6.5	5.8
Accumulated variation in %	40.4	55.4	62.8	69.3	75.1

3.5 DISCUSSION

Knowledge of the existing genetic variation and association between various agro-morphological traits and their heritability is vital for any breeding program. The accessions collected from different highlands of Ethiopia showed considerable variability for all examined morphological and agronomic traits (Table 3.1). Similarly,

Lucchin *et al*. (2003) found significant differences within and between populations for all the traits measured in their study aimed to characterize 20 Italian maize populations for 34 morphological and agronomic traits. The broad range in the means of accessions for the various traits implies great possibility for the development of inbred lines, hybrid and/or open-pollinated varieties. The wide range in days to maturity (108 to 186.5) for example, suggest flexibility for the development of cultivars for the various highlands of Ethiopia with differing rainfall and length of growing season.

Genetic traits such as the genotypic coefficient of variability, heritability and genetic advance provide estimates of genetic variation of quantitative traits. Of all the traits evaluated in this study, grain yield appears to combine high values of PCV, intermediate GCV and low h^2 (Table 3.2). This is in agreement with the report of (Rebourg *et al*., 2001) in maize. Hallauer and Miranda (1988) summarized numerous estimates of heritability in maize. These range from less than 0.3 for grain weight and kernel depth to between 0.5 and 0.7 for plant height, ear height, kernel row number and days to flower. The relatively low estimates for yield indicate that selection for this trait in maize would be more difficult. In contrast, high h^2 values with increments in the range of about 69.5 to 86.9% were noted for most morphological traits. Greater than 60% for heritability for plant and ear height has been previously reported in maize by Rebourg *et al*. (2001). On the other hand, number of kernels per row exhibited moderate PCV, GCV and high h^2 and GA as percentage of the mean (Table 3.2), which indicated that it is under additive genetic control. Simple selection of the plants bearing higher number of kernels per row may lead to success in improving the

trait up to 37.8%. This result was in agreement with previous studies of Arias *et al*. (1999) and Kumar and Kumar (2000). Therefore, highland maize breeders in Ethiopia should give more importance to kernels per row as selection criteria to increase grain yield.

The majority of the genotypic correlation coefficients were positive and highly significant (Table 3.3). However, only correlation coefficients greater than 0.71 or smaller than – 0.71 have been suggested to be biologically important (Skinner *et al*., 1999), as more than 50% of the variation in one trait is predicted by the other (Snedecor and Cochran, 1980). In this study, such important correlations (at genotypic level) were found between days to 50% tasseling and days to 50% silking (0.99), plant height and ear height (0.99), plant height and leaf length (0.97), ear diameter and 1000 seed weight (0.77), and number of rows per ear and number of kernels per row (0.95). The genotypic correlations between morphological traits (plant height, ear height and days to maturity) with yield were negative and significant (Table 3.3), indicating that the possibility of developing high yielding varieties with short plant height, medium ear height and early maturing.

Unlike the present study where the accessions were grouped into a few clusters (Figure 3.1 and 3.2), Taba *et al*. (1998) showed the formation of 12 non-overlapping clusters by evaluating 249 Caribbean maize accessions. The limited clustering observed in this study may be attributed to several factors: (i) maize being open-pollinated, there is continuous gene exchange between adjacent fields, (ii) local farmers acquire new seeds from distant sources to meet their requirements, and (iii)

there is a continuous seed supply by research organizations, non-governmental organizations and agricultural offices in these regions. Accessions collected from the Western and Southern regions (receiving high annual rainfall and long growing period) had the highest mean values for morphological traits. On the other hand, accessions collected from the Northern region (characterized by low annual rainfall and a short growing period) expressed the lowest mean values for all of the morphological traits. This result suggests that rainfall and growing season are the most important environmental factors differentiating traditional Ethiopian highland maize accessions. A similar result was reported by Ayana and Bekele (2000) in their study of morphological variation in sorghum collected from Ethiopia and Eritrea and noted that regional mean for plant height and days to maturity increases from North to South and from East to West, which followed the rainfall, temperature and seasonal patterns in Ethiopia (Tato, 1964).

The existence of broad morphological and agronomic diversity among the Ethiopian highland maize accessions is further substantiated by principal component analysis (Table 3.4), which indicated that the major contributing traits to the total variation are fairly distributed across morphological and agronomical traits. The major role of morphological traits in phenotypic variation is consistent with the work of Alika *et al*. (1993).

This study confirmed that traditional Ethiopian highland maize accessions display large amounts of variation for studied agro-morphological traits. The broad trait diversity evident among the maize accessions suggests ample opportunities for the

genetic improvement of the crop through selection directly from the accessions and/ or the development of inbred lines for future hybrid programs. Grouping accessions into morphologically similar, and most likely genetically similar groups (Souza and Sorrells, 1991) is helpful for selecting parents for crossing. In addition, the study allowed the selection of representative accessions from different areas of Ethiopia, which will be studied using molecular markers. Therefore, the grouping of accessions by phenotypic diversity in the present study and the data from AFLP and SSR markers analysis will be used to classify the highland maize accessions into genetically related groups, which could be used for various breeding, collection and conservation programs in the highlands of Ethiopia.

CHAPTER 4

BULKED-AFLP ANALYSIS OF GENETIC DIVERSITY AMONG TRADITIONAL ETHIOPIAN HIGHLAND MAIZE ACCESSIONS

Journal of Biodiversity and Conservation 2005 (in press)

4.1 ABSTRACT

In the highland regions of Ethiopia the heterogeneity of the land, climate, and soil favors the presence of a large number of landraces. A representative sample of 62 traditional Ethiopian highland maize accessions was analyzed using amplified fragment length polymorphism (AFLP) markers, to determine the degree of genetic diversity and relationships among these accessions and to study agroecological variation. Eight *Eco*RI/*Mse*I primer combinations were used. Of a total of 650 AFLP markers that were scored, 89.5% were polymorphic. Pair-wise genetic dissimilarity estimates revealed dissimilarity coefficients ranging from 0.32 to 0.69, with a mean of 0.57. Cluster analysis grouped most accessions collected from the Northern highlands into one major cluster while the Western and Southern accessions clustered together. In addition, variation partitioning revealed that only 9% of the total genetic variation was found between agroecologies, whereas 91% was found within agroecologies in Ethiopia. This finding may be explained by long distance seed exchange, continuous seed introduction and gene flow between the two agroecologies. The consistency of the results between clustering method and variation partitioning showed that AFLP markers accurately revealed genetic structure among maize accessions. Overall, the AFLP marker analysis indicated the existence of ample genetic diversity in highland maize accessions, which can be exploited by hybridization and selection.

Keywords: AFLP markers, bulked analysis, genetic resources, highland maize

4.2 INTRODUCTION

Maize was first introduced in Ethiopia in the 16th or 17th century (Hafnagel, 1961). Since its introduction, it has gained importance as a food and feed crop. Currently, it is the second most important crop, exceeded only by teff [*Eragrostis tef* (Zucc) Trotter] in terms of production area. However, it exceeds all other cereals in terms of annual production and yield ha $^{-1}$ (EARO, 2000). Maize is one of the cereals that provide most of the calorie requirements in the traditional Ethiopian diet. It is prepared and used as unleavened bread, roasted and boiled green ears, parched mature grain porridge and in local drinks like 'tella', 'borde' and 'areke' (Mulatu *et al*., 1992). Apart from these uses, maize leaves are fed to animals, while dry stalks are used as fuel and for the construction of fences and huts.

Ethiopia is a diverse country in terms of altitude, temperature, rainfall and soil types. Such diversity is apparent within even a short distance in a given locality. Diverse environments result in the presence of diverse vegetation, crop species and varieties in farmers' fields in most parts of the country (Vavilov, 1951). Maize varieties that are used in the highland regions of Ethiopia are well adapted, but are generally low-yielding open-pollinated varieties developed by local farmers. Many of these varieties resulted from centuries of planting, harvesting and selection. The highland maize varieties may be grouped into a number of completely or partially isolated populations, which may each, be adapted to different highland conditions (Chapter 3). Unfortunately, formal breeding programs have had little success in developing improved varieties for the diverse agroecological conditions in the highlands of

Ethiopia. In the past decade, only two improved open-pollinated maize varieties were developed for the highland zone (Twumasi-Afriyie *et al*., 2001).

Effective plant breeding and crop improvement programs depend on the availability of genetic diversity. Landraces are the original source of variation for breeding programs and are still the major source for new breeding programs in many developing countries. To assess the genetic diversity present in Ethiopian highland maize accessions, seed samples were collected from 287 highland locations of Ethiopia (Twumasi-Afriyie *et al*., 2001). A recent field study revealed that these accessions are highly variable for morphological and agronomic characteristics (Chapter 3). However, morphological variation does not always reflect real genetic variation because of genotype x environment interaction and the largely unknown genetic control of polygenic morphological and agronomic traits (Smith and Smith, 1992).

Molecular markers can be used to study the genetic diversity and genetic relationships among maize accessions directly at the DNA level. Amplified Fragment Length Polymorphism (AFLP) markers (Vos *et al*., 1995) have gained importance in crop genetic analyses, mainly due to the high multiplex ratio of this marker system. AFLP markers have been extensively used to study genetic diversity in maize inbred lines (Lubberstedt *et al*., 2000; Vuylsteke *et al*., 2000b). Although the technique is relatively simple and rapid, the large number of individual plants that need to be processed may limit AFLP analysis of cross-pollinated species like maize. One approach to overcome this limitation is to analyze one, or several, bulked samples per

accession, rather than individual plants. Kolliker *et al*. (2001) demonstrated that bulking equal amounts of leaf material before DNA extraction is an effective approach to produce representative AFLP marker profiles in white clover.

In this study the results of a genetic diversity analysis of Ethiopian highland maize accessions using bulked AFLP markers analysis will be reported. The objectives of this study were: (i) to assess the amount of genetic diversity and relationships among the highland maize accessions and (ii) to understand the distribution of genetic variation within and among agroecologies. The suitability of bulking leaf samples for genetic diversity of maize accessions has been investigated using both individual plants and their bulked samples.

4.3 MATERIALS AND METHODS

4.3.1 Plant materials and DNA extraction

A total of 62 traditional Ethiopian highland maize accessions were used for this study (Table 4.1). Previously, a representative subset of 180 of the 287 maize accessions collected from different highland regions in Ethiopia was analyzed for 15 morphological and agronomic traits (Chapter 3). Principal component and cluster analyses grouped these 180 accessions into four main clusters. The 62 accessions were chosen from the four clusters to represent the different agroecologies of Ethiopia and the range of morphological and agronomic variation observed in the field. For each of the 62 accessions, genomic DNA was extracted from leaf discs, harvested from 15 three-week

old plants (one 10-mm leaf disc per plant). For two accessions, individual DNA samples were also isolated from the 15 plants used for bulked sampling. DNA was extracted using the QIAGEN DNeasy plant Mini Kit, (QIAGEN, GmbH, Hilden) and homogenization was performed using the FP-120 FastPrep instrument (QBiogene, Carlsbad, CA, USA; Myburg *et al*., 2001). DNA quantity and quality was determined on 0.8% (w/v) agarose gel electrophoresis using known quantities of lambda DNA as a concentration standard.

Table 4.1 Traditional Ethiopian highland maize accessions used in the study

No	Accession	Collection site	Major agroecology	Altitude [a]
1	Ad-1-01	Gonder	North	2360
2	Ad-1-03	Armachew	North	2771
3	Ad-1-9-6	Adi Arkay	North	1837
4	Ad-1-9-8	Adi Arkay	North	1741
5	Ad-1-1-16	Armachew	North	2527
6	Ad-1-1-17	Armachew	North	1850
7	Ad-1-2-20	Armachew	North	1765
8	Ad-1-3-21	Armachew	North	2354
9	Ad-1-4-26	Dembia	North	2133
10	Ad-1-3-32	Dembia	North	2100
11	Ad-1-3-35	Chilga	North	1900
12	Ad-3-6-40	Gondar	North	2105
13	Ad-3-6-42	Fogera	North	1930
14	Ad-3-7-45	Farta	North	2400
15	Ad-3-7-46	Farta	North	2674
16	Ad-3-7-50	Este	North	2728
17	Ad-4-11-55	Sera	North	2544
18	Ad-5-13-59	Yilmana	North	2266
19	Ad-5-13-60	Yilmana	North	2300
20	Ad-5-13-61	Yilmana	North	2432
21	Ad-5-14-64	HuletEynes	North	1980
22	Ad-5-16-67	HuletEynes	North	2512
23	Ad-5-17-69	GoneraSiso	North	2654
24	Ad-5-17-68	GoneraSiso	North	2651
25	Ad-5-17-70	GoneraSiso	North	2668
26	Ad-5-18-71	Debrework	North	2598
27	Ad-5-18-72	Enemay	North	2474

28	Ad-5-19-76	Awabel	North	2554
29	Ad-5-21-79	Gozamin	North	2529
30	Ad-4-24-81	Gozamin	North	2383
31	Ad-6-28-89	Quarit	North	2000
32	Ad-6-28-92	Sekela	North	2500
33	Ad-6-28-94	Awi	North	1580
34	Ad-6-26-96	Awi	North	1714
35	Ad-1-31-101	Banja Awi	North	2200
36	Aw-03	Merka	South	1950
37	Aw-10	Agere Mariam	South	2180
38	Aw-13	Kofele	South	2500
39	Aw-17	Hitosa	South	2230
40	Aw-18	Boloso	South	1950
41	Aw-21	Arero	South	2160
42	Aw-25	Agere Mariam	South	2290
43	Aw-29	Agere Mariam	South	2200
44	Aw-33	Tiyo	South	2300
45	Aw-35	Tiyo	South	2515
46	Aw-41	Agere Mariam	South	2200
47	Aw-44	Ejere	South	2300
48	Aw-54	Merka	South	2145
49	Baw-01	Wolemra	West	2260
50	Baw-10	Wolemra	West	1800
51	Baw-11	Dendi	West	2290
52	Baw-12	Ambo	West	2270
53	Baw-13	Weliso	West	2300
54	Baw-14	Jeldu	West	2010
55	Baw-15	Becho	West	2225
56	Baw-17	Sululta	West	2350
57	Baw-18	Sululta	West	2350
58	Baw-20	Ambo	West	2280
59	Baw-22	Dega	West	2250
60	Baw-28	Bedele	West	1880
61	Baw-30	Ambo	West	2305
62	Baw-33	Limu	West	2080

[a] Meter above sea levels

4.3.2 AFLP analysis

AFLP template preparation was performed using AFLP template preparation kits from LI-COR Biosciences (LI-COR, Lincoln, NE, USA) according to the manufacturers' instructions, except that 10 µl diluted R/L mix, 2.0 mM $MgCl_2$ and 1.0

U Taq polymerase were used in the preamplification step. Polymerase chain reactions (PCRs) were performed using a BIO-RAD iCycler (Version 3.021, BIO-RAD Laboratories, Inc.). The preselective amplification cycle profile was as follows: incubation for 10 s at 72°C, followed by 30 cycles of denaturation for 10 s at 94°C, annealing for 30 s at 56°C, and extension for 1 min at 72°C with a 1 s per cycle increasing extension time. Selective amplification was performed on 1:20 diluted (in SABAX water) preselective amplification products with the following cycling profile: 13 cycles of 2 min at 94°C, 30 s at 65°C (reduced by 0.7°C per cycle), and 1 min at 72°C; followed by 20 cycles of 10 s at 94°C, 30 s at 56°C, and 1 min (extended 1 s per cycle) at 72°C. The preselective and selective amplification primer pairs all had two and three-nucleotide extensions at the 3' end, respectively. In all reactions only the *Eco*RI primers were 5'labelled with infrared dyes (IRDye 700 or IRDye 800, LI-COR). Initially, eight accessions were chosen to test the amplification successes of different primer combinations. The polymorphism rates and the total number of scorable fragments were evaluated in these eight accessions with 32 primer combinations. Eight primer combinations (Table 4.3) with the highest polymorphism rates and large numbers of clearly scorable fragments were selected to analyze the full set of 62 accessions.

4.3.3 Gel electrophoresis and scoring

An equal volume of loading solution (LI-COR) was added to each selective amplification reaction. Samples were denatured at 95°C for 3 min and placed on ice for 10 min before loading. A volume of 0.8 µl was loaded with an 8-channel syringe

(Hamilton, Reno, Nevada) onto 25-cm 8% Long Ranger gels (BMA, Rockland, ME, USA). Electrophoresis and detection of AFLP fragments were performed on LI-COR IR^2 (model 4200S) automated DNA analyzers. The electrophoresis parameters were set to 1500V, 40 mA, 40 W, 50°C, and a scan speed of 3. The run-time was set to 4 h and gel images were saved as TIF files for further analysis. The gel images were scored using a binary scoring system that recorded the presence and absence of bands as 1 and 0, respectively. Semi-automated scoring was performed with $SAGA^{MX}$ (Version 3.2, LI-COR) and followed by manual editing to make adjustments to the automated score where necessary. A locus was scored as polymorphic when the frequency of the most common allele (band present or absent) was less than 0.97 (absent or present in at least two individuals). Bands with the same mobility were considered as identical products (Waugh *et al*., 1997), receiving equal values regardless of their fluorescence intensity.

4.3.4 Data analysis

The binary data were exported into Microsoft Excel and formatted for use in the NCSS statistical software package. The average Polymorphic Information Content (PIC) for dominant markers was calculated according to Riek *et al*. (2001) by the following formula:

$PIC = 1 - \left[f^2 + (1 - f^2)^2 \right]$; where f is the frequency of the marker in the data set. Agroecology-specific alleles (private alleles) were recorded, if present for all primer pair combinations. Chi-squared test was used to test the significance difference in band frequency among agroecologies. Genetic similarity between accessions was

calculated according to Nei and Li (1979) using the formula:

$s_{ij} = \frac{2a}{(2a+b+c)}$, where s_{ij} is the similarity between two accessions i and j, a is the number of bands present in both i and j, b is the number of bands present in i and absent in j and c is the number of bands present in j and absent i. This formula excludes bands absent in both individuals, which cannot be necessarily being attributed to a common cause (Kolliker *et al*., 2001). Genetic dissimilarity was calculated as $1 - s_{ij}$. Cluster analysis was performed on the genetic dissimilarity matrix using Ward's minimum variance method (Ward, 1963). The Ward method was found to be a more suitable clustering technique than UPGMA as it avoided chaining effects that are often observed with UPGMA (Dubreuil *et al*., 1996). Genetic dissimilarity between accessions was also calculated based on Euclidean distances (Sneath and Sokal, 1973), which allowed us to estimate genetic variance within and among agroecologies according to Van Eeuwijk and Baril (2001) with the following formula:

$$d^2_{gi;g'i'} = \sum_{m=1}^{M} \left(x_{mgi} - x_{mg'i'}\right)^2,$$

Let d^2 is the distance between an individual *i* in group *g* and an individual *i'* in group *g'* for marker (from 1 to *M)*. Accordingly, the squared distance is the sum of the squared differences between individual accessions over all variables. The total variation (V_T) can be shown to be equal to the sum of all squared pair-wise distances between individual accessions (over all groups), divided by the total number of accessions (overall groups). The within group variation (V_W) is the sum over groups of the sum of squared pair-wise distances within a group divided by the group size. The between group variation (V_B) can be obtained by subtraction, $V_B = V_T - V_W$.

4.4 RESULTS

4.4.1 Detection limit of bulked AFLP analysis

The detection threshold of AFLP bands was tested in two 15-plant bulks by analyzing individuals and their bulked leaf samples separately. A total of 120 polymorphic markers were scored in the two bulked samples (15 individual/accession) and 30 individual samples using three selective primer combinations. The analysis of the AFLP profiles generated from bulked leaf samples demonstrated that most of the bands present in individual plants were present in AFLP profiles from bulked leaf samples (Figure 4.1).

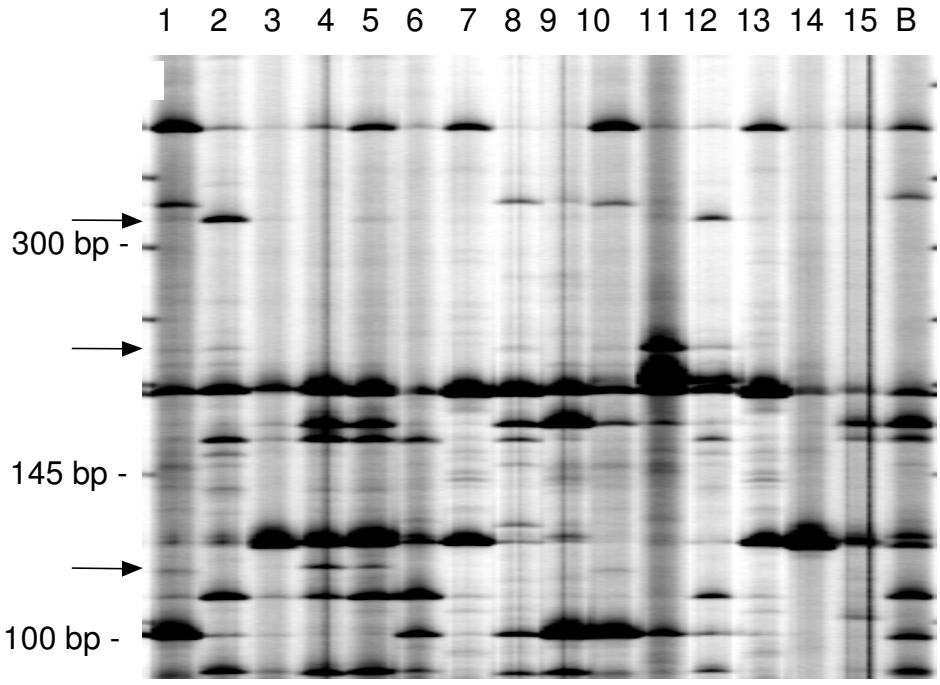

Figure 4.1 An example of the LI-COR AFLP image generated using E-AAC/M-CGG primer combinations showing banding patterns of 15 individuals (1-15) and their

bulked leaf samples (B). A molecular marker is indicated at the beginning and at the end of the gel. Arrows show examples of bands present in individuals and absent in bulked samples

The relationships between band frequency and presence and absence in the bulks of 15 plants of the two accessions (average of three primer pairs) is presented in Table 4.2 Bands which are only present in single plant (7%), are present in bulks of accession 'Baw-01' (58%), and 'Ad-1-03' (46%). High proportion (83%) of bands which are only present in less than 20% (3 out of 15) of the individual plants were present in bulks across the two accessions. However, all bands that are shared by more than 50% of the individual plants were represented in bulked samples in both accessions across the three primer pair combinations (Table 4.2).

Table 4.2 Relationships between AFLP band frequencies in individual plants and representation in bulks of 15 individuals based on the average of three primer pair combinations

		Accession		
	Band frequency in individual plants	'Baw-01'	'Ad-1-03'	Average of the two accessions
% Bands represented in bulk	Upto7%	0.58	0.46	0.52
	7-20%	0.82	0.83	0.83
	21-50%	0.85	0.92	0.89
	50-75%	100	100	100
	75-100%	100	100	100
% Total bands in individuals present in bulks		90	89	89.5

4.4.2 Marker polymorphism

To assess the genetic diversity of the 62 highland maize accessions, a total of 650 AFLP bands with fragment sizes ranging from 52 to 720 bp were generated using eight selective AFLP primer pair combinations. Of these, 89.5% were polymorphic among the 62 accessions (Table 4.3). The number of polymorphic bands per primer combination ranged from 56 to 98 with an average number of 72.8 (Table 4.3). PIC values for primer enzyme combinations ranged from 0.279 to 0.370, with an overall mean of 0.325. A typical LI-COR AFLP image generated using E-ACG/M-CGG primer combination across 62 maize accessions is presented in Figure 4.2.

Table 4.3 Degree of polymorphism and average polymorphism information content for the eight AFLP primer combinations used to analyze the 62 Ethiopian maize accessions

No	Primer combination [a]	Total number of bands	Number of polymorphic bands	% of polymorphic bands	PIC
1	E-AGG/ M-CAG	73	64	87.7	0.370
2	E-ACG/ M-CCG	72	66	91.6	0.320
3	E-ACA/ M-CGA	109	98	89.9	0.279
4	E-ACA/ M-CCC	86	76	88.0	0.321
5	E-AAC/ M-CAC	72	68	94.5	0.359
6	E-ACG/ M-CGG	74	68	91.8	0.321
7	E-AAC/ M-CCG	69	56	81	0.327
8	E-AAC/ M-CGG	95	86	90.5	0.300
	Total	650	582	Na.	Na.
	Mean	81.3	72.8	89.5	0.325

[a] E, *Eco*RI & M, *Mse*I

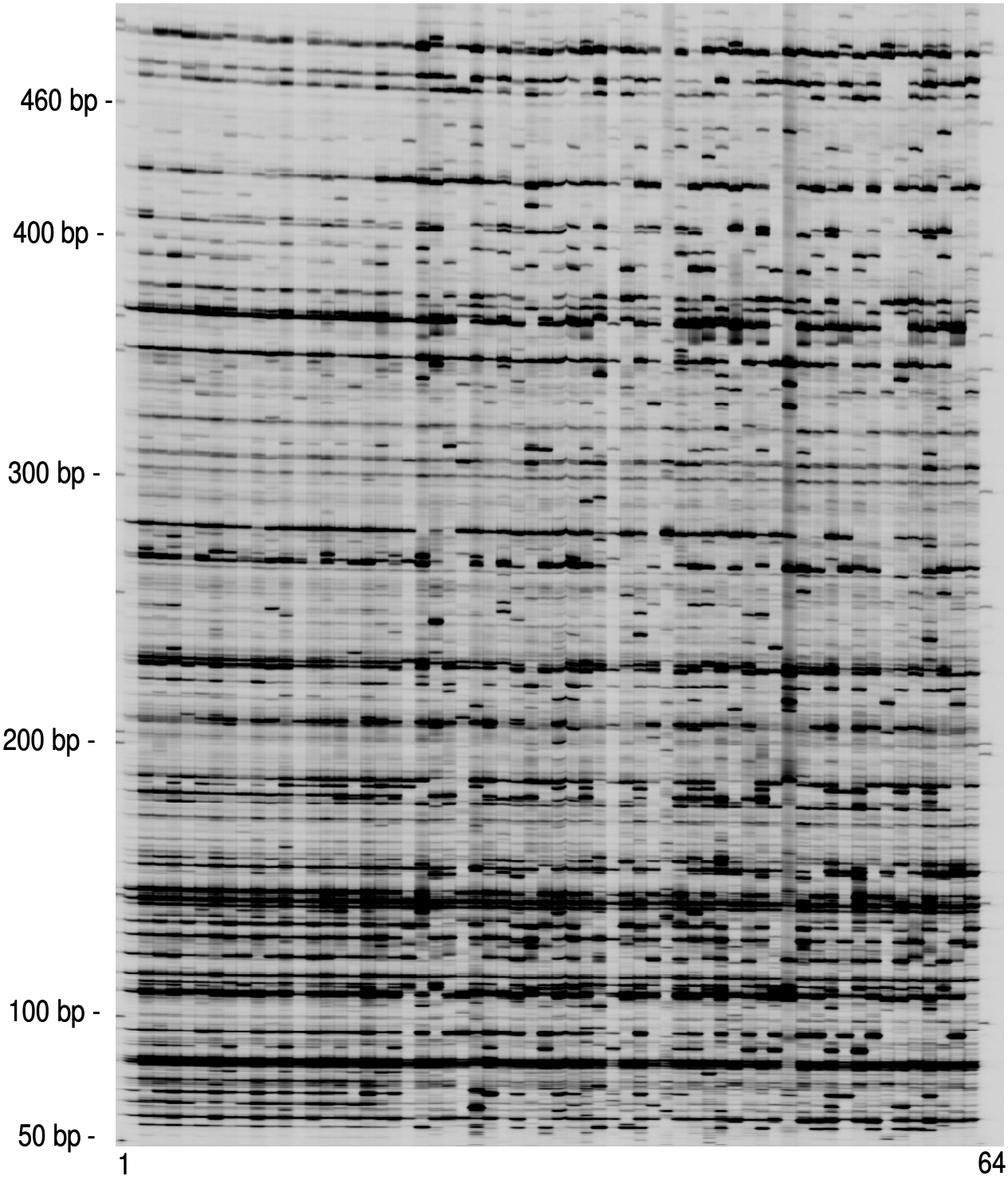

Figure 4.2 Typical a LI-COR AFLP image produced by selective amplification using the E-ACG/M-CGG primer combination in 62 maize accessions. Lanes 1 and 64 are IRDye 700 molecular weight standards (LI-COR Biosciences)

4.4.3 Distribution of bands across the three agroecologies

Table 4.4 shows the total number of bands and their frequency for each agroecology. According to Chi-square tests, accessions collected from the Northern agroecology were significantly (p = 0.01) different from the Southern and Western agroecologies with respect ot rare bands, which present up to 25% of the accessions. There were significant difference (p = 0.01) between the Northern and Western agroecologies with respect to bands that present up to 50% of the accessions. However, there was no significant difference between the Western and Southern agroecologies in any of the bands frequencies. Comparing the three agroecologies simultaneously, they differ significantly in all band frequencies except bands that occurred at 50-70% frequencies. Among 650 markers 73 bands were unique to the Northern agroecology. However, the Western and Southern agroecologies had 6, and 5 unique bands, respectively.

Table 4.4 Distribution of AFLP bands expressed as the percentage of accessions that carry a particular band in each agroecology

	Northern agroecology		Southern agroecology		Western agroecology	
Frequency of bands	No. of bands	% of bands from the total	No. of bands	% of bands from the total	No. of bands	% of bands from the total
< 5	73	0.11	6	0.01	5	0.01
5-15	152	0.24	84	0.15	73	0.13
15-25	59	0.09	122	0.21	135	0.24
25-50	114	0.18	131	0.23	163	0.29
50-75	121	0.19	110	0.19	106	0.19
75-100	120	0.19	119	0.21	89	0.16
Total	639		572		571	

4.4.4 Genetic dissimilarity of maize accessions

In order to evaluate the molecular diversity of the 62 accessions, the pair-wise genetic dissimilarity was calculated for the 1891 pairs of Ethiopian highland maize accessions (Figure 4.3). Pair-wise dissimilarity ranged from 0.32 (accession no. 1 and accession no. 13 from the Northern agroecology) to 0.69 (accession no. 1 and accession no. 40 from the Northern and Southern agroecologies, respectively) with an overall mean of 0.57. More than 71% of the pair-wise comparisons exhibited genetic dissimilarity higher than 0.5.

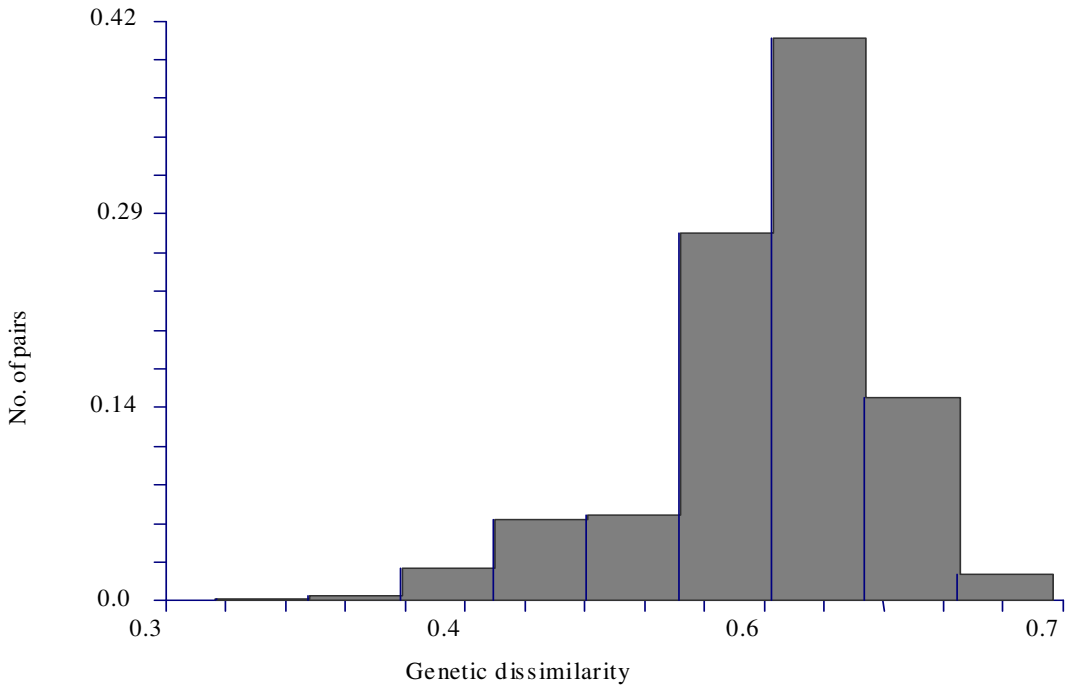

Figure 4.3 Frequency distribution of the genetic dissimilarity of pairs of 62 Ethiopian highland maize accessions

To visualize the relationships among the accessions, a dendrogram was generated

from the dissimilarity matrix using Ward minimum variance as clustering method (Figure 4.4). The Ward method combines the two clusters in each step whose fusion leads to the smallest increase in the Euclidean sum of squares within groups, thus leading to a maximized variance within groups and a minimized variance within groups. This was particularly important for the accessions, since the variance within accessions was already very high. The dendrogram showed three major clusters. Cluster I consisted of 20 accessions, all collected from the Northern agroecology. Cluster II consisted of 21 accessions and 85.7% were collected from Southern and Western agroecologies. Cluster III contained 21 accessions, collected from the three agroecologies.

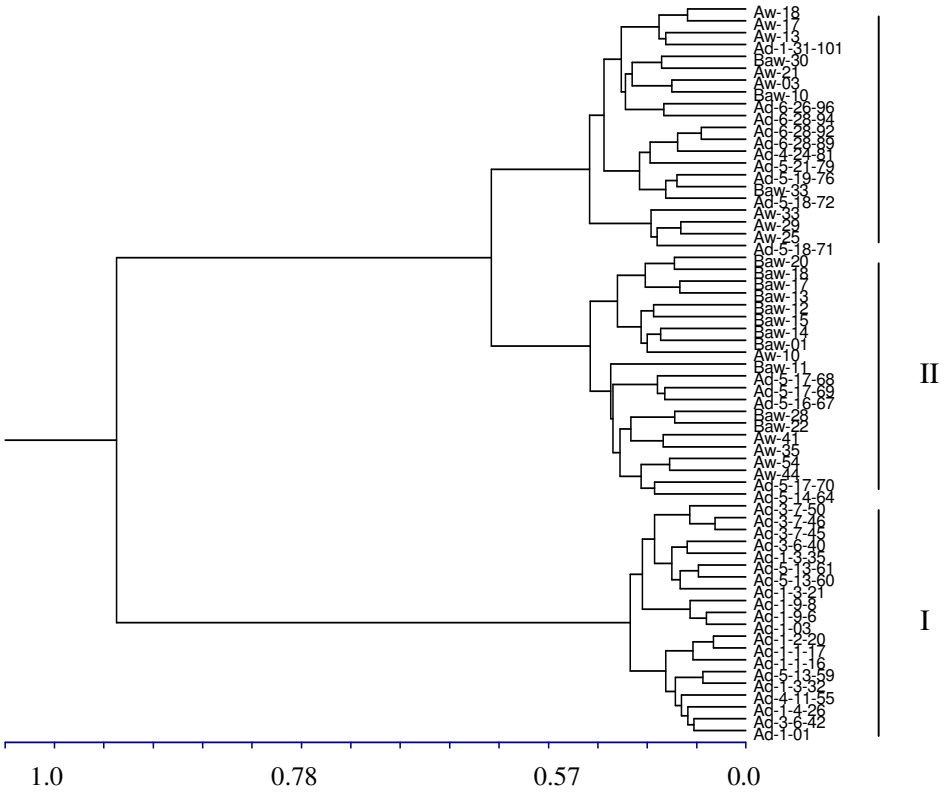

Figure 4.4 Dendrogram of traditional Ethiopian highland maize accessions derived by

Ward' minimum variance method from the dissimilarity matrix of AFLP data

4.4.5 Partitioning of genetic variation

Partitioning the total genetic variability using the analysis of distance method (van-Eeuwijk and Baril, 2001) revealed that 91% of the total variation was found within and the remaining 9% among agroecologies (Table 4.5) suggesting limited genetic differentiation among the agroecologies. Furthermore, partitioning of the within agroecology variation into different agroecologies, indicated that accessions collected from the Northern agroecology contributed 54.3% of the total variation. The Southern and the Western agroecologies contributed only 21.7 and 24% to the total variability, respectively (Table 4.5).

Table 4.5 Partitioning of the total genetic variation of traditional Ethiopian highland maize accessions into within and between agroecologies variation

Sources of variation	Number of accession	Variance component	Percentage of variation	Mean genetic distance
Total	62	10		0.57
Within	-	9.1	91.0	
Northern	35	4.9	54.3	0.57
Southern	13	1.0	21.7	0.43
Western	14	2.2	24.0	0.45
Among	-	0.9	9.0	

4.5 DISCUSSION

Knowledge about genetic diversity and relationships among diverse germplasm is

useful for plant breeders. It supports their decisions on the selection of parents for crossing and is helpful to widen the genetic basis of breeding programs. A genetic improvement program has been initiated at Ambo, Ethiopia in collaboration with CIMMYT with the goal of producing improved maize cultivars for the highlands of Eastern African countries. To initiate the improvement effort, a basic understanding of the genetic diversity and relationships among the highland maize accessions was considered essential. In this chapter, the results of a genetic diversity analysis of 62 traditional Ethiopian highland maize accessions collected from different highlands of Ethiopia using bulk-AFLP markers analysis will be reported.

Molecular markers are a very efficient approach to rapidly attain genetic diversity estimates to be used in various breeding programs and policies. However, the cost of DNA extraction and subsequent analysis is a major consideration that limits the use of molecular markers for large-scale genetic diversity study. This is particularity true for comparisons of open-pollinated crops where many individuals per population should be sampled (to capture the genetic variability within a single population) and cannot be used for the routine characterization of large germplasm collections. One approach to overcome this problem is the use of bulking leaf/DNA samples per population rather than several individual per population. Bulking strategies provide a means of large-scale diversity analysis in cross-pollinated crop species (Kolliker *et al*., 2001; Rebourg *et al*., 2001).

To determine the detection limits and suitability of bulked AFLP analysis for the measurement of genetic diversity in maize accessions, the results of AFLP gel profiles

from 15 individual plants and their bulked leaf samples was compared. The AFLP patterns obtained from bulked leaf samples were highly representative of the AFLP patterns obtained from individual plants of the same accession (Figure 4.2). The detection limit found in this study varied somewhat among primer combinations. Some bands that appeared to be of low frequency (less than 3 out of 15) were detectable in bulked samples, while other present in higher frequency (up to 6 out of 15) were not detected in bulked samples (Table 4.2). This observation was in agreement with the work of Kolliker *et al.* (2001), who reported that some bands that were present in individual white clover plants at a higher frequency (20-50%) were absent in bulked samples, whereas some bands present in individual plants at low frequency (less than 5%) were represented in bulked samples. The authors speculated that complex competition processes during PCR amplification in AFLP technique could be the reason for this. In a previous study two bulk of 15 individuals/population has been used in study of European maize populations using RFLP markers (Rebourg *et al*., 2001). In this study, on average 89.5% of the bands, which were scored in individual plants, were present in bulked samples suggesting that pooling leaf samples before DNA extraction is an effective means of producing representative profiles of individual plants. This allowed higher sample throughput during DNA extraction and minimized reagent cost for genetic diversity estimation among Ethiopian highland maize accessions. Similarly, Kolliker *et al.* (2001) have demonstrated that bulking at leaf stage is effective in producing representative profiles in white clovers using AFLP analysis.

The proportion of polymorphic bands (89.5%, Table 4.3) obtained in the present study is high compared to the work of Lubberstedt *et al.* (2000) who reported an average

polymorphism rate of 84% in early European maize inbred lines selected from different heterotic groups. As in other studies, AFLP analysis in Ethiopian traditional maize accessions detected many polymorphic bands and is an efficient method for diversity study. With a single combination of selective primers, the average number of bands detected was 81.3. Considering the technical simplicity and sensitivity to DNA polymorphism of this technique, it is advantageous for studies of open-pollinated species than other similar techniques. The average PIC value (0.325, Table 4.3) in this study was also close to the high end of the range (0.29 to 0.33) previously reported for maize (Lubberstedt *et al*., 2000; Vuylsteke *et al*., 2000b). The high PIC values may be due to the fact that we prescreened AFLP primer combinations and selected the eight primer combinations (Table 4.3) with the highest polymorphism rates and largest numbers of clearly scorable fragments.

The range and average dissimilarity based on AFLP data (range: 0.32 to 0.69, mean: 0.57) observed in the present study were similar to that reported by Rebourg *et al*. (2001; range: 0.106 to 0.793, average of 0.55) for European maize populations. The average genetic diversity among Ethiopian highland maize accessions is therefore as diverse as observed among maize populations collected from the whole of Europe. The high genetic diversity observed among the traditional Ethiopian highland maize accessions suggests ample opportunity for the development of improved varieties for different highland parts of Ethiopia. This might be due to the nature of the materials used in the study. The highland accessions are open-pollinated varieties developed by local farmers over centuries and there has been a continuous introduction of seed into these regions. Another factor might be the sampling strategy employed in this study,

which maximized the geographical and morphological range among the 62 selected highland maize accessions.

The dendrogram revealed that the Northern accessions are more differentiated compared to the Western and Southern accessions (Figure 4.3). The separation of the Northern accessions from the rest of the accessions might be (a) due to strong selection by local farmers for adaptation to the drier growing conditions and as result they had large number of unique bands (73, Table 4.4), (b) due to the little introduction of high yielding and uniform varieties into this agroecology, and (c) due to the general restriction of seed movement into the Northern region from other agroecologies due to geographical isolation. There was, however, no differentiation between accessions collected from the Western and Southern agroecologies. This was also supported by Chi-square tests where there was no significant difference in band distribution between the Western and Southern accessions (Table 4.4). The reasons might be the result of several factors: Firstly, in these agroecologies, maize is the staple food (mainly as porridge) and hence local farmers have selected similar accessions suitable for food properties. Secondly, these two agroecologies are physically in close proximity and there might be gene flow between farmers' varieties. Finally, there have been continuous introductions of high yielding and uniform varieties released in the surrounding intermediate regions by government and non-government extension programs (Sasakawa-Global2000, 2002). The average yield of introduced varieties (5.6 t/ha) was three times higher than that obtained from traditional maize varieties (1.2 t/ha). All these activities together with the tradition of local framers to acquire seeds from distant places might be the reasons for the low

genetic differentiation between the Southern and the Western agroecologies.

To our knowledge no published data so far available on the use of bulked AFLP for genetic diversity study of maize accessions. These results indicate that bulked AFLP analysis can be successfully applied to study the genetic diversity and relationships among maize accessions/landraces.

CHAPTER 5

ANALYSIS OF GENETIC DIVERSITY OF ETHIOPIAN HIGHLAND MAIZE ACCESSIONS USING SSR MARKERS

- ***Journal of Genetic Resources and Crop Evolution 2005 (in press)***

5.1 ABSTRACT

Over the past three centuries, maize has become adapted to complex environmental conditions in the highlands of Ethiopia. We analyzed 62 traditional Ethiopian highland maize accessions, using 20 simple sequence repeat (SSR) markers, to assess genetic diversity among these accessions and to understand the within and among agroecologies genetic variation. The average number of alleles was 4.9 per locus and the average polymorphism information content was 0.61. Pair-wise genetic dissimilarity coefficients ranged from 0.27 to 0.63 with a mean of 0.49. Of the total 98 alleles detected in the traditional Ethiopian highland maize accessions, 26 alleles were found to be specific to Northern agroecology, while accessions from the Western and Southern agroecologies each had only two region specific alleles. Eight individual alleles were almost fixed (>90%) in the Northern accessions while 11 and 12 alleles were fixed in Southern and Western accessions, respectively. Therefore, genes at or linked to these alleles might be contributed towards one or more traits determining adaptation to specific environmental conditions. Ward minimum variance cluster analysis grouped most of the accessions from the Northern agroecology into three major clusters and all of the Southern and Western accessions into another clusters suggest that genetic variability was low amongst the Western and Southern accessions because they show the least variation and clustered together. Maize accessions grown in the drier regions of the Northern agroecology might have accumulated favorable genes to drought tolerance which could be exploited for the development of varieties specifically adapted to the region or regions with similar agroecology.

Key words: Clustering, Ethiopia, genetic diversity, highland maize, SSR markers

5.2 INTRODUCTION

Maize (*Zea mays* L.) was introduced in Ethiopia more than three centuries ago (Hafnagel, 1961) and is grown mainly for human consumption. Since then, it has been grown in the lowland, mid-altitude and highland parts of the country. According to the Central Statistical Authority (CSA, 2001), maize is grown on 1.4 million hectares of land, which is about 21% of the cultivated area in Ethiopia. The national average yield of maize (1.9 t ha^{-1}) is well below the world average (EARO, 2000). Maize cultivars that are used in the highland regions of Ethiopia are well adapted, but low yielding open-pollinated varieties developed by local farmers. Many of these varieties resulted from centuries of planting, harvesting and selection. The highland maize varieties may be grouped into a number of completely or partially isolated populations, which may each be adapted to different highland conditions.

To assess the diversity present in these materials, the Highland Maize Germplasm Collection Mission was launched throughout the different highlands of Ethiopia in 1998 in collaboration with CIMMYT (Twumasi-Afriyie *et al*., 2001). As part of this project, 287 maize accessions were collected from farmers' fields. Recent field study revealed that these accessions are highly variable for morphological and agronomic characteristics (Chapter 3). However, morphological variation does not always accurately reflect the real genetic variation because of genotype x environment interaction (Smith and Smith, 1992).

Molecular markers can reveal differences among accessions at the DNA level and thus provide a more direct, reliable and efficient tool for germplasm conservation and management. Microsatellite, or simple sequence repeat (SSRs) markers, have been frequently applied in genetic diversity studies in maize inbred lines and populations (Matsuoka *et al*., 2002; Warburton *et al*., 2002; Pinto *et al*., 2003). A method for detecting SSR marker polymorphism using the less costly and more widely available agarose gel system was suggested (Senior and Heun, 1993). Using the agarose gel system, Senior *et al*. (1998) used 70 SSR primers on 94 U.S. maize inbred lines and was able to group these lines into nine clusters that corresponded to major maize heterotic groups or endosperm types. Similarly, Pinto *et al*. (2003) using 30 SSR loci on agarose gel were able to measure and compare the genetic diversity in tropical maize populations and synthetics and concluded that mean number of alleles per locus, proportion of polymorphic loci and gene diversity were greater in the synthetic 'IG-3' than 'IG-4'.

In this study, SSR polymorphism among traditional Ethiopian highland maize accessions was analyzed using agarose gel electrophoresis according to Senior *et al*. (1998). The objectives of this study were to estimate the level of genetic diversity and relationships among the accessions and to understand the within and among agroecologies genetic variation. The information will be useful to identify genetically related genotypes for future maize improvement and to design conservation strategies in the highlands of Ethiopia.

5.3 MATERIALS AND METHODS

5.3.1 Plant materials and DNA extraction

The total of 62 traditional Ethiopian highland maize accessions were used for this study (Table 4.1). Previously, a representative subset of 180 of the 287 maize accessions collected from different highland regions in Ethiopia were analyzed for 15 morphological and agronomic traits. Principal component and cluster analyses grouped these 180 accessions into four main clusters (Chapter 3). The 62 accessions were chosen from four phenotypic clusters to represent the different agroecologies and the range of agro-morphological variation observed in the field. The regions in which the maize accessions were collected represent almost all of the highlands of Ethiopia (Figure 5.1). The agroecologies vary in altitude (meter above sea level: Northern, 1600-2400; Western, 1800-2500; Southern 1750-2650), annual rainfall (average in mm: Northern, 600-1200; Western, 1000-2000; Southern, 1350-1600), average temperature (minimum and maximum mean air temperature in °c: North, 15-30; Western, 12-27; Southern, 6-27) and growing periods (in months: Northern, 4-5, Western, 5-8, Southern, 4-8; NMSA, 2000).

For each of the 62 accessions, genomic DNA was extracted from young leaves, harvested in bulk (one 10 mm leaf disc per plant) from 15 three-week old plants. For two accessions, individual DNA samples were also isolated from the 15 plants used for bulked samples. DNA was extracted using the QIAGEN DNeasy plant Mini Kit, (QIAGEN, GmbH, Hilden) and homogenization was performed using the FP-120

FastPrep instrument (QBiogene, Carlsbad, CA, USA, Myburg *et al*., 2001). DNA quantity and quality was determined on 0.8% (w/v) agarose gel electrophoresis using known quantities of lambda DNA as concentration standard.

Figure 5.1 Map showing the 62 maize accessions that were collected from different highlands of Ethiopia and used in the present study. The approximate location of each collection site is indicated in the map by square points

Original map source: (http:/www.1uptravel.com/worldmaps/Ethiopia.html)

5.3.2 Simple sequence repeats primer selection

A total of 105 SSR primers were selected from previous studies (Senior *et al*., 1998; Matsuoka *et al*., 2002: Warburton *et al*., 2002) and from the public Maize GDB (http://www.agron.missouri.edu/ssr_probes/ssr.htm) based on their high polymorphism information content and chromosome locations (at least 10 SSRs per chromosome, data not shown). The 105 SSRs were assayed in eight diverse highland maize accessions, which were expected to represent a high level of genetic diversity due to difference in collection sites and morphological traits. A final set of 20 SSR primers (Table 5.1), which gave consistent and easily scorable bands across the eight accessions were chosen for further analyses.

5.3.3 PCR amplification and gel electrophoresis

Polymerase chain reactions (PCRs) were performed in 15 µl reaction mixes consisting of 50 ng template DNA, 0.4 mM dNTPs, 0.4 µM SSR primers (forward and reverse), 0.1 mM $MgCl_2$, 0.5 U Taq polymerase (Roche) and 1X reaction buffer (10 mM Tris-HCl, 50 mM KCl, 1.5 mM $MgCl_2$). PCR reactions were performed in a BIO-RAD iCycler (Version 3.021, BIO-RAD Laboratories, Inc.) with the following touch-down PCR program: an initial denaturation at 94°C for 2 min, followed by 10 cycles of 30 s at 94°C, 45 s at 65°C (reduced by 1°C per cycle), and 1:30 min at 72°C; followed by 28 cycles of 30 s at 94°C, 30 s at 55°C, and 1:30 min at 72°C. A final extension step of 72°C for 15 min was performed. The SSR amplification products were resolved on 3% agarose gels (1:1 mixes of Molecular Screening, MS-8, agarose, LSS-Gibco and

molecular grade agarose, Gibco-BRL, a cheaper alternative with similar resolution, Pinto *et al*., 2003) in 0.5X TBE buffer. Gels were run in a large format (23 x 40 cm) horizontal gel system (Model A3-1, Owl Separation Systems, Portsmouth, NH, USA) at 150V for 3.5 h and were photographed under UV light (Geldoc. BIO-RAD Laboratories, Inc) after ethidium bromide staining.

5.3.4 Data analysis

The alleles of each marker were binary coded using 1 for presence or 0 for absence within each marker class. Data were recorded as a binary matrix by assigning a molecular weight to each allele in comparison to 50 and/or 100-bp molecular weight ladder. The different polymorphic alleles scored from each locus were designated by the name that consisted of the locus name followed by a number starting from one for the heaviest locus (Table 5.1). The exact molecular size of each allele was not determined. The discriminatory potential of each locus considering all accessions, was determined by the polymorphism information content (PIC) according to Powell *et al*. (1996). Agroecology-specific alleles (private alleles) and common alleles were recorded, if present for each locus and for all SSR loci.

Genetic distances between accessions were calculated based on the formula of Nei and Li (1979). Cluster analysis was performed on the genetic dissimilarity matrix using Ward's minimum variance method (Ward, 1963). The Ward method was found to be a more suitable clustering technique than UPGMA as it avoided chaining effects that are often observed with UPGMA (Dubreuil *et al*., 1996). The total genotypic

variation of the accessions within and among agroecologies was calculated based on the squared Euclidean distance according to the formula given by Van Eeuwijk and Baril (2001). All of the data analyses were performed using the software package PHYLIP version 3.5c (Felsenstein, 1993) and NCSS-2000 (Jerry, 2000).

5.4 RESULTS

5.4.1 Detection limit of pooled DNA and marker polymorphism

The detection threshold of SSR alleles was tested in 15-plant bulks by analyzing separate individuals and their bulked DNA samples (Figure 5.2). The majority of alleles present in at least two or more individuals were generally present in the bulks. This seemed to be the detection limit for alleles across all of the SSR loci.

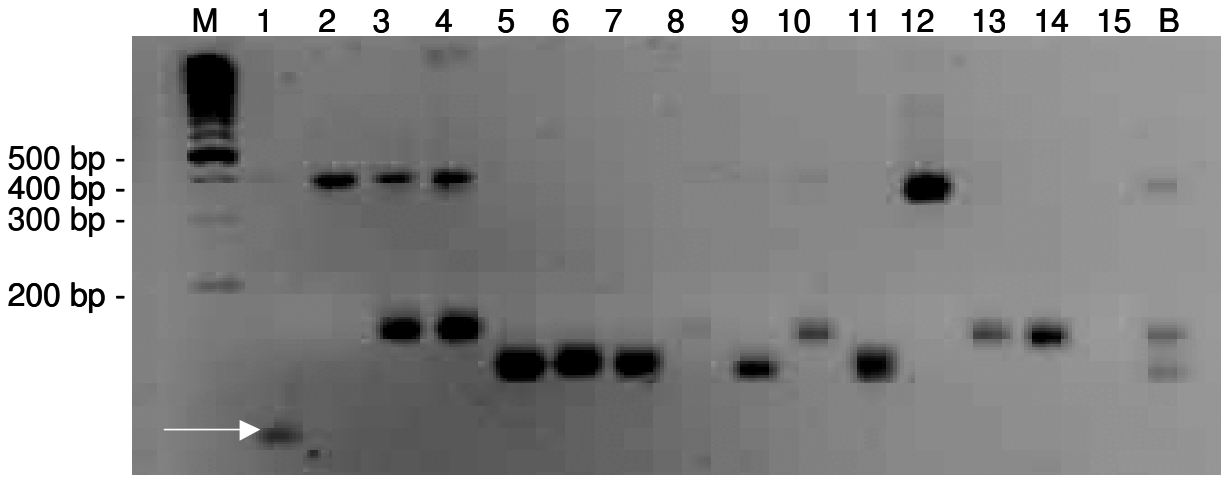

Figure 5.2 The SSR 3% gel image of individual plants (lane 1-15) and the bulk sample (lane B) of the accession 'Baw 10' amplified with locus *phi034*. M is molecular ladder (Gene Ruler 100-bp DNA ladder Plus, Frementas). Arrow shows the band present in individual plants and absent in the bulked sample

Indices of genetic variability assessment among the 62 traditional Ethiopian highland maize accessions are given in Table 5.1. Except for chromosomes 4 and 7, which had three SSR markers and chromosomes 6 and 8, which had one SSR marker, the rest of the chromosomes were represented by two SSR markers, which provided a good coverage of genome wide variation. A total of 98 alleles were detected, an average of 4.9 alleles per locus. The majority (75%) of the SSR loci had 3-5 alleles. However, a few loci, namely *phi042, umc2040, phi026, phi45312, bnlg182, nc003* and *bnlg2190* had 6-10 alleles per locus (Table 5.1). The PIC ranged from 0.06 (*umc1357*) to 0.76 (*nc003*) with a mean of 0.61 for the entire collection. The average PIC values were 0.61 for the Northern, 0.51 for the Southern and 0.57 for the Western agroecologies.

Approximately 26.5% of the SSR alleles were found to be unique (exclusive alleles that are found only in a single agroecology) to the Northern region. The Western and Southern collections had only two specific alleles each (Table 5.1). In the Northern agroecology, eight individual alleles (*umc1632-1, phi042-6, phi021-5, phi054-4, phi037-3, umc1357-3, phi015-2 and umc1537-3*) were almost fixed in all accessions. In Southern accessions, 12 alleles namely: *nc003-3, umc2190-1, umc1632-1, phi042-6, bnlg182-6, phi034-2, bnlg2190-3, phi021-5, phi054-4, umc1357-3, phi015-2 and umc2129-2*) were fixed while in the Western accessions 11 alleles scored from 10 SSR markers (*nc003-7, umc21901*, *umc1632-1, phi042-6, bnlg2190-3, phi021-5, phi054-4, umc1357-3, phi015-2 and umc1537-3)* were fixed. The percentage of shared alleles among the Northern and the Southern, Northern and Western, and Southern and Western agroecologies were 63.5, 72.9 and 82.5%, respectively.

Table 5.1 Summary of microsatellite, bin number, repeat unit, number of alleles per locus and the polymorphism information content (PIC) of the different agroecologies of traditional Ethiopian highland maize accessions

			Entire collection		Northern agroecology		Southern agroecology		Western agroecology	
SSR locus	Bin[a] No.	Repeat unit	No. of alleles	PIC	No.of alleles	PIC	No. of alleles	PIC	No.of alleles	PIC
bnlg182	1.03	Unknown	7	0.72	7(3)[b]	0.72	4	0.70	4	0.71
phi037	1.08	AG	3	0.66	3	0.66	3	0.42	3	0.60
nc003	2.06	AG	8	0.76	8(3)	0.81	3	0.70	5	0.71
umc2129	2.07	CGC	5	0.59	4	0.55	3	0.58	4(1)	0.65
phi453121	3.0	ACC	6	0.71	6(2)	0.75	3	0.53	4	0.65
umc2152	3.09	TG	3	0.66	3	0.64	3	0.66	3	0.59
phi021	4.03	AG	5	0.66	5(1)	0.67	4	0.66	4	0.61
phi026	4.05	CT	6	0.75	6(2)	0.76	2	0.35	4	0.74
phi079	4.05	AGATG	3	0.59	3	0.58	3	0.57	3	0.61
umc1537	5.07	TCG	3	0.60	3	0.62	2	0.50	3	0.59
umc1153	5.09	TCA	5	0.60	4	0.50	5(1)	0.66	4	0.57
umc2040	6.05	CGC	6	0.74	6(2)	0.74	3	0.57	4	0.64
umc1632	7.01	AGC	3	0.56	3	0.55	3	0.58	3	0.57
phi034	7.02	CCT	4	0.75	4(1)	0.58	3	0.56	3	0.54
umc2190	7.06	CCT	3	0.33	3	0.44	1	0	3	0.24
phi015	8.08	AAAC	5	0.68	5(2)	0.67	3	0.65	3	0.66
phi042	9.04	CATA	6	0.59	4	0.57	5	0.60	6(1)	0.62
umc1357	9.05	CTG	3	0.06	3(2)	0.10	1	0	1	0
phi054	10.0	AG	4	0.66	4(1)	0.66	4(1)	0.65	3	0.64
bnlg2190	10.1	AG	10	0.59	10(7)	0.70	3	0.32	3	0.39
Total			98		94(26)		61(2)		70(2)	
Mean			4.9	0.61	4.7	0.61	3.1	0.51	3.5	0.57

[a], the bin no contains linkage group and genetic interval information. Each of the 10 maize linkage groups is divided into approximately 10 bins (Maize GDB **http://www.agron.missouri.edu/ssr_probes/ssr.htm**.). At least one SSR loci was sampled in each maize chromosome.

[b], Numbers in the bracket is alleles unique to specific agroecology.

5.4.2 Genetic dissimilarity and cluster analysis

The genetic dissimilarity coefficients for the 1891 pairs of accessions showed a normal distribution (Figure 5.3). It ranged from 0.27 (accessions no. 1 and 2) both collected from Northern agroecology to 0.63 for (accessions no. 1 and 50) collected from the Northern and Western agroecologies, respectively, with an overall mean of 0.49 (Figure 5.3).

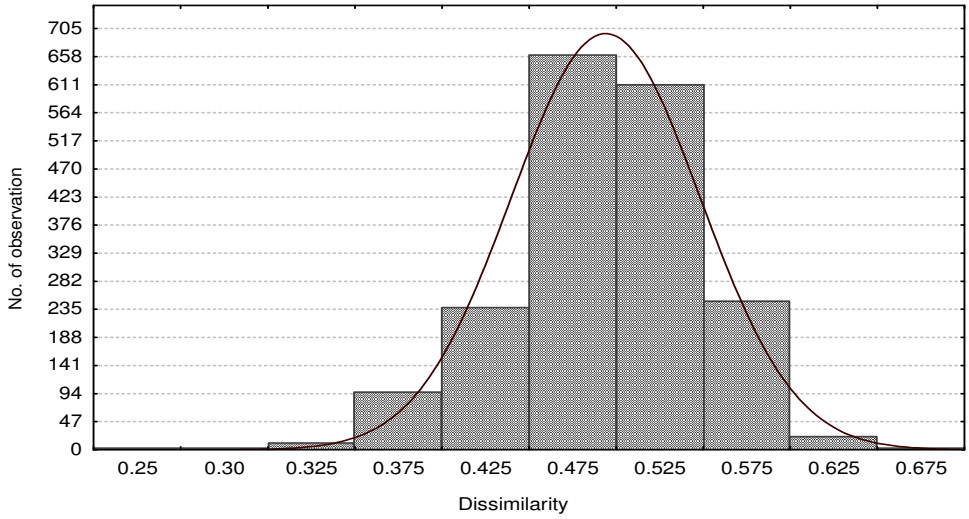

Figure 5.3 Frequency distribution of the genetic dissimilarity pairs of 62 Ethiopian maize accessions based on 20 SSR loci

The Ward cluster analysis revealed that the 62 accessions were grouped into five clearly defined clusters at a mean genetic dissimilarity of 0.49 (Figure 5.4). Cluster I and II contained 11 and 9 accessions, respectively and all of them were collected from the Northern agroecology. Cluster III consisted of 11 accessions, of which ten were collected from the Northern, and one from the Western agroecology. Almost all

(84%) of the accessions collected from the Western and Southern agroecologies were grouped in cluster IV. Cluster V consisted of 7 accessions, collected from the Northern, Southern and Western agroecologies.

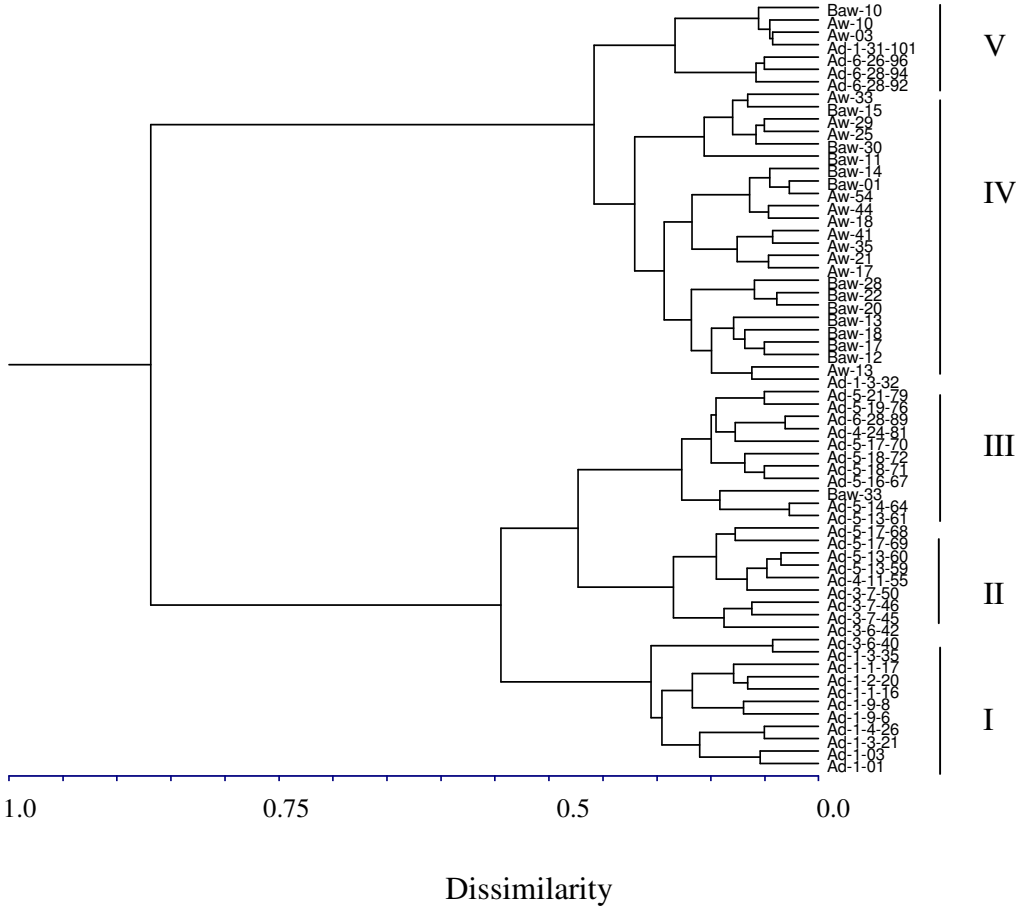

Figure 5.4 Ward minimum variance based dendrogram generated from the Nei and Li dissimilarity matrix showing relationships among 62 traditional Ethiopian highland maize accessions.

5.4.3 Partitioning of genetic variation

The analysis of distance method revealed that 89.6% of the total variation was found within and the remaining 10.4% among agroecologies (Table 5.2) further supporting the finding of very little differentiation among the agroecologies. Furthermore, partitioning

of the within agroecologies variation into different agroecologies, indicated accessions collected from the Northern agroecology, contributed 56.5% of the total variation. The Southern and the Western agroecologies contributed only 18.5 and 14.6% to the total variability, respectively.

Table 5.2 Partitioning of the total genetic variation into within and between agroecologies of traditional Ethiopian highland maize accessions

Sources of variation	Number of accession	Variance component	Percentage of variation	Mean genetic distance
Total	62	7.5		0.49
Within	-	6.8	89.6	
Northern	35	4.3	56.5	0.50
Southern	13	1.1	14.6	0.42
Western	14	1.4	18.5	0.46
Among	-	0.9	10.4	

5.5 DISCUSSION

Knowledge about diversity and genetic relationships among landraces is important in crop improvement strategies. Molecular markers reveal differences at the DNA level and thus provide direct, reliable and efficient tools for germplasm conservation and management. In this study, 20 polymorphic SSR loci were used to estimate the genetic relationships among representatives of 62 traditional Ethiopian highland maize accessions. Cluster analysis showed that accessions collected from the Northern agroecology were distinct from the Western and Southern accessions. However, there was no distinction between Western and Southern accessions.

The analysis of individuals included in the bulk showed that an allele is easily detectable in the bulked sample when present in a proportion greater than 1 out of 15 (7%) of the individuals within DNA pooled-samples. This result is consistent with the sensitivity of detection (0.05) reported by Michelmore *et al*. (1991) in bulked segregant analysis. Using the pooled DNA strategy (two 15-plant bulks per population), Rebourg *et al*. (2001) genotyped 23 RFLP loci in 131 European maize populations and found an average number of 9.1 alleles per locus. They concluded that there was high genetic diversity and strong differentiation between populations. The efficient screening of SSR polymorphism on an agarose gel system, coupled with the high detection limit of the bulked DNA sample strategy appeared to be promising as a method for characterizing open-pollinated maize varieties, which are mainly grown in developing countries, at relatively low cost.

The 20 SSR loci displayed high levels of polymorphism among the traditional Ethiopian highland maize accessions. Barbosa-Neto *et al*. (1997) have reported that marker loci should be chosen uniformly over the genome in genetic diversity studies, avoiding biases due to sampling and increasing precision of genetic similarity. Therefore, the analysis of 20 SSR loci, which were sampled from all chromosomes, appeared to be effective for assessment of genetic diversity among traditional Ethiopian highland maize accessions. According to Senior *et al*. (1998), five SSR loci were adequate to give distinctive fingerprints for 94 U.S. maize inbred lines.

The average number of alleles and PIC detected in the 62 accessions (Table 5.1) was

higher than those reported by Pinto *et al*. (2003) for tropical maize populations and synthetics. The high average number of alleles and PIC observed in this study might be due to the fact that we prescreened 105 SSR primers and selected the 20 SSR primers (Table 5.1) with the highest polymorphism information content. This is in agreement with the work of Enoki *et al*. (2002) who screened 100 SSR primers and found the mean PIC value of 0.69 with 60 selected SSR primers.

The mean and range of genetic dissimilarity observed in this study were comparable to those reported by Pinto *et al*. (2003) for tropical populations and synthetics. This might be due to the sampling techniques, which maximized geographical and morphological range among the highland maize accessions. Bogyo *et al*. (1990) stated that sampling germplasm collections across diverse environments based on morphological variation is considered to be the most effective way for capturing genetic diversity. The genetic distance values among the accessions confirmed that accessions from different agroecologies are genetically more dissimilar than those originating from the same agroecology. This observation is in agreement with Henandez (1985) who found that local farmers in different regions independently developed maize accessions for desired traits such as yield, kernel color and food properties.

Most of the maize accessions collected from the Northern agroecology (cluster I, II and III, Figure 5.4) were separated from the rest of agroecologies indicating a distinctly different genetic background from the other agroecologies. This may be partly due to adaptation to different climatic conditions and restriction of seed

movement from other agroecologies due to geographical isolation. Within the Northern agroecology where the accessions were collected, there is a shortage of rainfall and a short growing period, and therefore farmers have selected the accessions for these climatic conditions over centuries. This is further illustrated by the occurrence of a high level of unique alleles in the Northern agroecology (Table 5.1). Unique alleles are valuable because they indicate the presence of novel genetic variation. Moreover, it seems that higher percentages of specific alleles is a characteristics of tropical germplasm. A higher percentage of private alleles were observed in tropical inbreds when compared to the US, Canadian and European ones (Matsuoka *et al*., 2002). The SSR markers have been shown to be under the influence of natural selection (Saghai-Maroof *et al*., 1994) and unique SSR alleles within the Northern agroecology could have been selected in drought-stressed environments. This study also clearly indicated that there are close genetic relationships between accessions collected from the Western and Southern agroecologies. The differences between these agroecologies as indicated by 20 SSR loci were only 17.6%.

Partitioning of total genetic variability into within agroecologies and among agroecologies using analysis of distance (AOD) further demonstrated high (89.4%) variability found within agroecologies and only 10.6% accounted for among agroecologies differentiation (Table 5.2). The low differentiation might be the result of several factors: Firstly, in these two regions maize is the stable food (mainly as porridge) and hence local farmers select similar accessions suitable for food and testing properties. Secondly, these two agroecologies were identified as the high potential maize growing areas because of similar environmental conditions (high

rainfall, fertile soils and the long growing period) by the National Maize Improvement Center of Ethiopia (EARO, 2000). Finally, the two regions are physically in close proximity, and there might be gene flow among farmer's varieties.

In conclusion, accessions from the Northern agroecology may be used as base materials for the development of improved varieties for the drier parts in the highlands of Ethiopia. From a conservation perspective, sampling many accessions from all agroecologies would be an effective way of capturing genetic variation for future collections. Moreover, seeds should be collected from the Western and Southern agroecologies before the existing diversity is lost as result of the introduction of high yielding and uniform varieties in the neighboring areas.

CHAPTER 6

A COMPARATIVE STUDY OF MOLECULAR AND MORPHOLOGICAL METHODS OF DESCRIBING GENETIC RELATIONSHIPS IN MAIZE

- *African Journal of Biotechnology 2005, 4:586-597.*

6.1 ABSTRACT

The comparison of different methods of estimating the genetic diversity could define their usefulness in plant breeding and conservation programs. In this study, a total of 15 morphological traits, eight AFLP-primer combinations and 20 simple sequence repeat (SSR) loci were used (i) to study the morphological and genetic diversity among 62 selected highland maize accessions, and (ii) to assess the level of correlation between phenotypic and genetic distances. The analysis of variance of the morphological data revealed significant differences among accessions for all measured traits. The mean morphological dissimilarity (0.3 with a range of 0.1-0.68) was low in comparison to dissimilarity calculated using SSR markers (0.49 with a range 0.27-0.63) and AFLP markers (0.57 with a range 0.32-0.69). The correlation between the morphological dissimilarity matrix and the matrices of genetic dissimilarity based on SSR and AFLP markers was 0.43 and 0.39, respectively (p = 0.001 in both cases). The correlation between SSRs and AFLPs dissimilarity matrices was 0.67 (p = 0.001). This congruence indicates that both marker systems are equally suited for genetic diversity study of maize accessions. Cluster analysis of morphological and marker distances indicate that Northern accessions are genetically differentiated from the Western and Southern accessions in the highlands of Ethiopia. From this study three groups of maize accessions with distinct genetic profiles and morphological traits were identified, which will be useful for a future collection and breeding of maize for the highlands of Ethiopia.

Key words: AFLP, correlation, phenotypic diversity, SSR

6.2 INTRODUCTION

The development of improved varieties of crop plants in breeding and selection programs depends on the existence of genetic diversity on which selection can act. Knowledge of genetic variation and relationships between accessions or genotypes is important: (i) to understand the genetic variability available and its potential use in breeding programs, (ii) to estimate any possible loss of genetic diversity, (iii) to offer evidence of the evolutionary forces shaping the genotypic diversities, and (iv) to choose genotypes to be given priority for conservation (Thormann *et al*., 1994). Characterization of genetic resource collections has been greatly facilitated by the availability of a number of molecular marker systems. Morphological traits were among the earliest markers used in germplasm management, but they have a number of limitations, including low polymorphism, low heritability, late expression, and vulnerability to environmental influences (Smith and Smith, 1992).

On the other hand, DNA markers do not have such limitations. They can be used to detect variation at the DNA level and have proven to be effective tools for distinguishing between closely related genotypes. Different types of molecular markers have been used to assess the genetic diversity in crop species, but no single technique is universally ideal. Therefore, the choice of the technique depends on the objective of the study, financial constraints, skills and facilities available. Amplified fragment length polymorphisms (AFLPs) and microsatellites, or simple sequence repeats (SSRs) are the most frequently used molecular markers in the analysis of genetic resources, because they can be automated and so have great potential in large-

scale genetic diversity studies. The chromosomal locations of SSR markers are frequently known, thus providing additional information in genetic diversity studies and on the other hand, AFLPs have a high multiplex ratio, offering a distinctive advantage when genome coverage is a major issue (Pejic *et al*., 1998).

Powell *et al*. (1996) examined the utility of RFLP, RAPD, AFLP and SSR markers for soybean germplasm analysis by evaluating information content (expected heterozygosity), number of loci simultaneously analyzed per experiment (multiplex ratio) and effectiveness in assessing relationships between accessions. In this study SSR markers had the highest expected heterozygosity, while AFLP markers had the highest effective multiplex ratio.

The Highland Maize Germplasm Collection Mission was launched throughout the highlands of Ethiopia in 1998 in collaboration with CIMMYT (Twumasi-Afriyie *et al*., 2001). As part of this project, 287 maize accessions were collected from farmers' fields throughout the highland regions of Ethiopia. One hundred and eighty of these accessions were studied for morphological diversity (Chapter 3) and thereby generated baseline data for future breeding and molecular studies. In Chapter 4 the authors performed various statistical analyses on AFLP data to assess the genetic diversity and differentiation among 62 selected traditional Ethiopian highland maize accessions. This was also confirmed by SSR data generated on the same 62 maize accessions (Chapter 5). Nevertheless, the relationships among the morphological diversity, AFLP diversity and SSR diversity have not been investigated.

The objectives of this study were thus (i) to investigate genetic diversity and relationships among 62 selected highland maize accessions using morphological, AFLP and SSR markers, (ii) to assess the correlation between distance estimates based on morphological traits and molecular markers, and (iii) to classify the accessions into groups based on a combination of molecular profiles and morphological traits.

6.3 MATERIALS AND METHODS

6.3.1 Field evaluation and data recording

A total of 62 maize accessions collected from the Northern, Southern and Western highlands of Ethiopia were used in this study (Table 4.1 and Figure 5.1). The accessions were grown at Alemaya University in Ethiopia during the 2002 main cropping season in a randomized complete block design with two replications. Each accession was grown in two row plots. Each row had 25 plants, which constitute 44444 plants per hectare recommended for the testing site. From each accession, 20 plants were selected at random to record 15 morphological traits.

6.3.2 Plant materials DNA extraction

All 62 accessions were fingerprinted with AFLP and SSR markers. All plants used in this molecular analysis were generated from seed and grown in the greenhouse. As this study did not aim to estimate the degree of heterozygosity and heterogeneity

within the accessions, 15-plant bulks were analyzed in order to represent the genotypic variability present within each maize accession. DNA was extracted using the QIAGEN DNeasy plant Mini Kit, (QIAGEN, GmbH, Hilden).

6.3.3 AFLP analysis

Eight AFLP primer combinations were used in this study (E-AGG/M-CAG, E-ACG/M-CCG, E-ACA/M-CGA, E-ACA/M-CCC, E-AAC/M-CAC, E-ACG/M-CGG, E-AAC/M-CCG and E-AAC/M-CGG). These were chosen from 32 primer combinations tested in a previous study (Chapter 4) with eight Ethiopian highland maize accessions (which were expected to represent a high level of genetic diversity due to difference in collection sites and morphological traits). The selection was based on amplification success, high polymorphism and the total number of scorable fragments. AFLP analysis was performed according to Vos *et al*. (1995) using AFLP template preparation kits from LI-COR Biosciences (LI-COR, Lincoln, NE, USA) with little modification as described in detail in Chapter 4. Fragments were separated by polyacrylamide gel electrophoresis using LI-COR 420S DNA analyzers (for details see Chapter 4).

6.3.4 SSR analysis

Twenty SSR loci (Table 5.1), which had previously been shown to display easy to read banding patterns on agarose gels (Chapter 5), which mapped to different linkage groups and which displayed a high degree of polymorphism in maize (Senior *et al*.,

1998; Warburton *et al*., 2002) were used for the molecular diversity study. Information about primer sequences, SSR repeat motifs, chromosomal location, PCR amplification conditions, gel electrophoreses and data scoring are discussed in Chapter 5.

6.3.5 Statistical analysis

Analysis of variance was performed for all measured traits in order to test the significance of variation among accessions. The standardized traits mean values (mean of each trait was subtracted from the data values and the result divided by the standard deviation) were used to perform principal component and cluster analyses using NCSS 2000 software (Jerry, 2000). To group the accessions based on morphological dissimilarity, cluster analysis was conducted on the Euclidean distance matrix with the unweighted pair group method based on arithmetic averages (UPGMA).

For molecular diversity analysis, matrices of binary data were constructed with rows equal to accessions, and columns equal to distinct molecular marker fragments (bands in the case of AFLP and alleles for SSR). For the 62 maize accessions, the body matrix contained zeros and ones, corresponding to the absence or presence of marker band/alleles, respectively. Dissimilarity matrices were constructed from the binary data with Nei and Li (1979) similarity coefficients. From these matrices of dissimilarity coefficients, the mean genetic distances, standard deviations and distribution of dissimilarity values were calculated. Finally, to determine the efficiency of each marker type in detecting polymorphisms, the assay efficiency

index, AEI (Pejic *et al*., 1998; AEI = BP/T, where BP is the total number of polymorphic fragments detected and T is the total number of marker assays performed), and the proportion of polymorphic fragments (total number of polymorphic fragments detected/total number of fragments detected) were calculated. The average Polymorphism Information Content (PIC) for AFLP markers was calculated according to Riek *et al*. (2001), while for SSR markers it was calculated according to Powell *et al*. (1996).

The relationships between the Euclidean distance matrix based on morphology and the Nei and Li distance matrices obtained with AFLP and SSR markers were analyzed using the approach developed by Mantel (1967). The principle of this approach is to calculate the sum of the cross product of the distance matrices and to compare this sum with the value expected according to a null hypothesis (no difference between the distance matrices). All the data analyses were performed using the software package NCSS-2000 (Jerry, 2000) and Statistical for Windows (1995).

6.4 RESULTS

6.4.1 Morphological variability

The analysis of variance revealed highly significant differences among accessions for all of the traits suggesting that there was a high degree of phenotypic diversity among the accessions (Table 6.1). Grain yield, plant and ear height and days to maturity showed wide variation, while number of leaves, leaf width and ear diameter showed a narrower

range of phenotypic variation (Table 6.1).

Table 6.1 Summary statistics of the agro-morphological traits measured in 62 traditional Ethiopian highland maize accessions

Traits	Mean	St Dev	Minimum	Maximum
Days to tasseling	65.1	3.2	51.5	76.0
Days to silking	71.5	3.0	58.0	80.5
Plant height (cm)	217.8	14.4	161.0	288.0
Ear height (cm)	125.9	26.3	74.0	227.5
Leaf length (cm)	71.3	9.1	51.8	100.8
Leaf width (cm)	9.1	1.0	6.4	12.7
Numbers of leaf	6.1	0.3	5.2	6.6
Foliage rating	6.2	0.9	4	7.0
Days to maturity	143.8	7.8	108	167.5
Ear diameter (cm)	3.9	0.2	3.3	4.6
Ear length (cm)	18.1	2.2	14.5	22.7
Rows per ear	10.7	1.5	7	13.9
Kernels per row	27.42	3.6	18	36.9
1000 seed weight (g)	295.8	41.3	229	410.0
Yield (kg ha $^{-1}$)	2645.4	195.4	1305.2	42.82.3

The first four principal components (PCs), which had eigenvalue higher than one, explained a total of 71.8% of the phenotypic variation (Table 6.2). In the first PC, which explained 42.1% of the total variation, the most important traits were plant and ear height, leaf length and days to tasseling and silking. Number of rows per ear also appeared to be important in the first PC. In the second PC, which explained 12.6% of the total variation, predominant traits were ear traits (yield, ear length, ear diameter and kernels/row) and foliage rating. The third principal component, which accounted for 10.5% of the total variation, was dominated by traits such as number of leaves, ear

diameter, yield and ear length, while days to maturity, leaf width and number of leaves were important delineating traits associated with the fourth principal component, which accounted for 6.7% of the total variation.

Table 6.2 Eigenvectors, eigenvalues, individual and cumulative percentage of variation explained by the first four principal components (PC) after assessing morphological traits in 62 traditional Ethiopian highland maize accessions

Traits	PC1	PC2	PC 3	PC 4
Days to tasseling	-0.32	-0.20	0.03	-0.23
Days to silking	-0.32	-0.19	0.00	-0.21
Plant height (cm)	-0.37	-0.09	-0.06	0.08
Ear height (cm)	-0.36	-0.09	-0.08	0.12
Leaf length (cm)	-0.34	-0.06	-0.01	-0.13
Leaf width (cm)	-0.24	0.10	-0.13	-0.40
Number of leaves	-0.06	0.03	0.59	-0.31
Foliage rating	-0.15	0.49	0.17	-0.22
Days to maturity	-0.15	0.02	0.29	0.67
Ear diameter (cm)	-0.21	0.39	-0.37	0.11
Ear length (cm)	-0.22	0.39	-0.34	0.07
Rows per ear	-0.33	-0.10	0.14	0.14
Kernels per row	-0.14	0.35	0.21	0.23
1000 seed weight (g)	-0.28	-0.14	0.27	0.13
Yield (kg ha -1)	0.09	0.44	0.36	-0.12
Eigenvalue	6.31	1.88	1.55	1.00
Individual variation %	42.05	12.57	10.51	6.67
Accumulated variation %	42.05	54.61	65.12	71.79

6.4.2 Variation in molecular markers

The 62 traditional Ethiopian highland maize accessions were fingerprinted with AFLP

and SSR markers. The levels of polymorphism detected with both marker systems and polymorphism information content are reported in Table 6.3. Both molecular marker systems were able to uniquely discriminate each accession. The total number of bands was 650 based on eight AFLP primer combinations, and 98 alleles were detected for 20 SSR loci. All 20 SSR loci and 89.5% of AFLP bands were polymorphic (Table 6.3). The average number of scored bands was 81.3 for AFLP primer combinations and ranged from 69 (E-AAC/M-CCG) to 109 (E-ACA/M-CGA). The mean number of alleles per SSR locus was 4.9, ranging from 3-10 (Table 6.3). The PIC values for primer enzyme combinations of AFLP ranged from 0.279 to 0.370, with an overall mean of 0.325. For SSR analysis this value ranged from 0.06 (*umc1357*) to 0.76 (*nc003*) with a mean of 0.61. The assay efficiency index of AFLPs was far superior to that of SSRs (AEI = 72.6 *vs*. 4.9), but the proportion of polymorphic fragments was higher for SSRs (Table 6.3).

Table 6.3 Level of polymorphisms and informativeness obtained with AFLP and SSR markers in 62 traditional Ethiopian highland maize accessions

Parameters/marker	AFLPs	SSRs
Number of primer pairs	8	20
Number of fragment detected	650	98
Number of polymorphic fragments	582	98
Polymorphic fragments per marker	56-98	3-10
Assay efficiency index (AEI)	72.6	4.9
Proportion of polymorphic fragments	89.5	1
Polymorphism information content per marker	0.279 - 0.37	0.06 - 0.76

6.4.3 Distribution of dissimilarity coefficients

A histogram of pair-wise dissimilarity for 62 traditional Ethiopian highland maize accessions generated from molecular markers and morphological data is presented in Figure 6.1 and a comparison of dissimilarity coefficients (range, mean and standard deviation) is presented in Table 6.4. The dissimilarity coefficients based on morphology ranged from 0.1 to 0.68 with an average of 0.3. Based on SSR, these values ranged from 0.27 to 0.63 with an overall mean of 0.49. For AFLP, it ranged from 0.32 to 0.69 with an overall mean of 0.57. More than 71% of AFLP- based pair-wise comparisons exhibited genetic dissimilarity higher than 0.5.

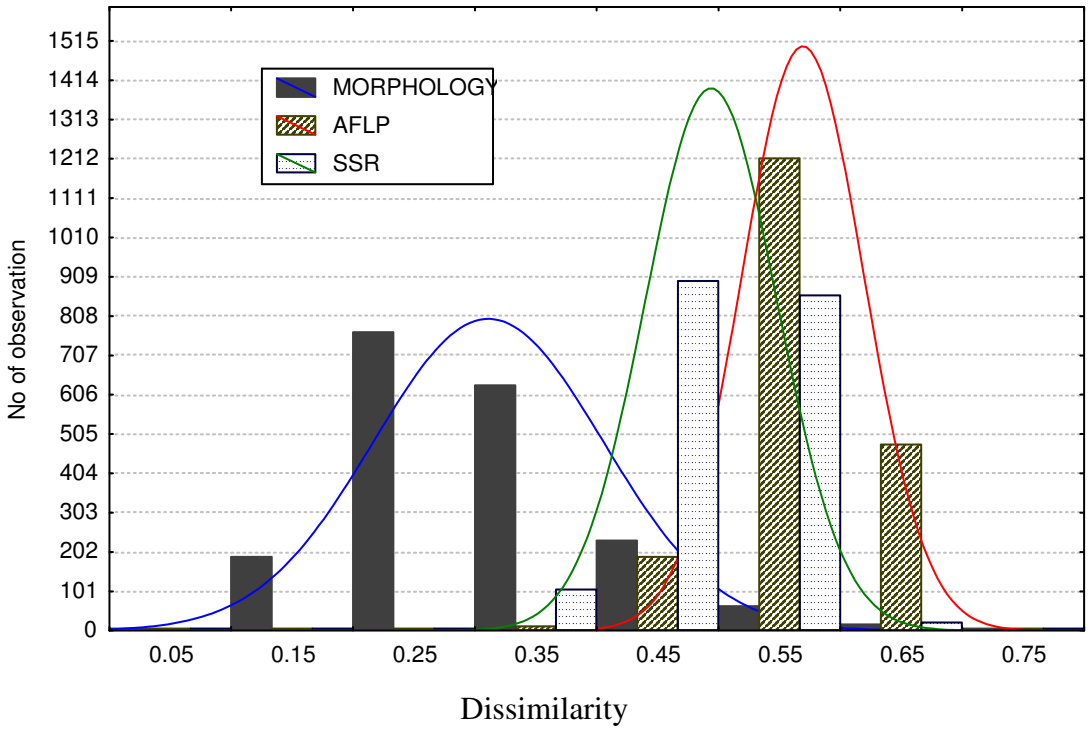

Figure 6.1 Frequency distribution of genetic dissimilarity among pair-wise combinations of 62 traditional Ethiopian highland maize accessions based on morphology, AFLP and SSR data

Table 6.4 Mean, standard deviation and range of Nei and Li dissimilarity coefficients (calculated using AFLP and SSR markers) and Euclidean distance (calculated using morphological traits). The total sample of all accessions in this study is shown followed by accessions collected from the three agroecologies

Parameters	Accessions	Morphological	AFLPs	SSRs
Mean	Entire collection	0.30	0.57	0.49
	Northern	0.28	0.58	0.51
	Southern	0.23	0.54	0.43
	Western	0.29	0.57	0.46
Standard deviation	Entire collection	0.09	0.05	0.05
	Northern	0.08	0.06	0.06
	Southern	0.09	0.04	0.05
	Western	0.10	0.04	0.05
Range	Entire collection	0.1-0.68	0.32-0.69	0.27-0.63
	Northern	0.12-0.57	0.32-0.66	0.27-0.65
	Southern	0.12-0.54	0.44-0.64	0.34-0.51
	Western	0.10-0.37	0.47-0.65	0.32-0.59

6.4.4 Correlations between dissimilarity matrices

In order to compare the extent of agreement between dendrograms derived from morphology, SSRs and AFLPs, a distance matrix was constructed for each assay and compared using the Mantel matrix correspondence test (Table 6.5). A highly significant positive correlation was found between the two molecular data sets (r = 0.67;

$p = 0.001$). The AFLP data was significantly correlated with the morphological data ($r = 0.39$, $p = 0.001$), and the SSR data was also correlated with morphological data ($r = 0.43$; $p = 0.001$). The correlation between morphological data with that of molecular data (AFLP +SSR, joint analysis) was high and significant ($r = 0.54$; $p = 0.001$). The significant correlations indicate that these three independent sets of data likely reflect the same pattern of genetic diversity and validate the use of these data to calculate the different diversity statistics for Ethiopian highland maize accessions**.**

Table 6.5 Correlation between dissimilarity matrices obtained with different marker types

	Morphology	SSR	AFLP	AFLP +SSR (joint analysis)
Morphology		0.43	0.39	0.54
SSR			0.67	

** Mantel test, $p = 0.001$

A dendrogram generated from the standardized morphological data is presented in Figure 6.2. The UPGMA cluster analysis revealed four clusters at the mean genetic dissimilarity of 0.3. The first cluster contained 36 accessions, most collected from the Northern agroecology. Short plants and early maturity characterized accessions in this group. The second cluster contained 12 accessions, of which 11 were collected from the Western and one from the Southern agroecology. Accessions in this group had tall plants and ear heights. This group also had the maximum yield ha $^{-1}$. The third cluster contained only two accessions with dissimilarity values of 0.4. The fourth cluster

contained 11 accessions collected from all three agroecologies, and there was one outlier (AD-1-9-8) that did not fall into any cluster (Figure 6.2).

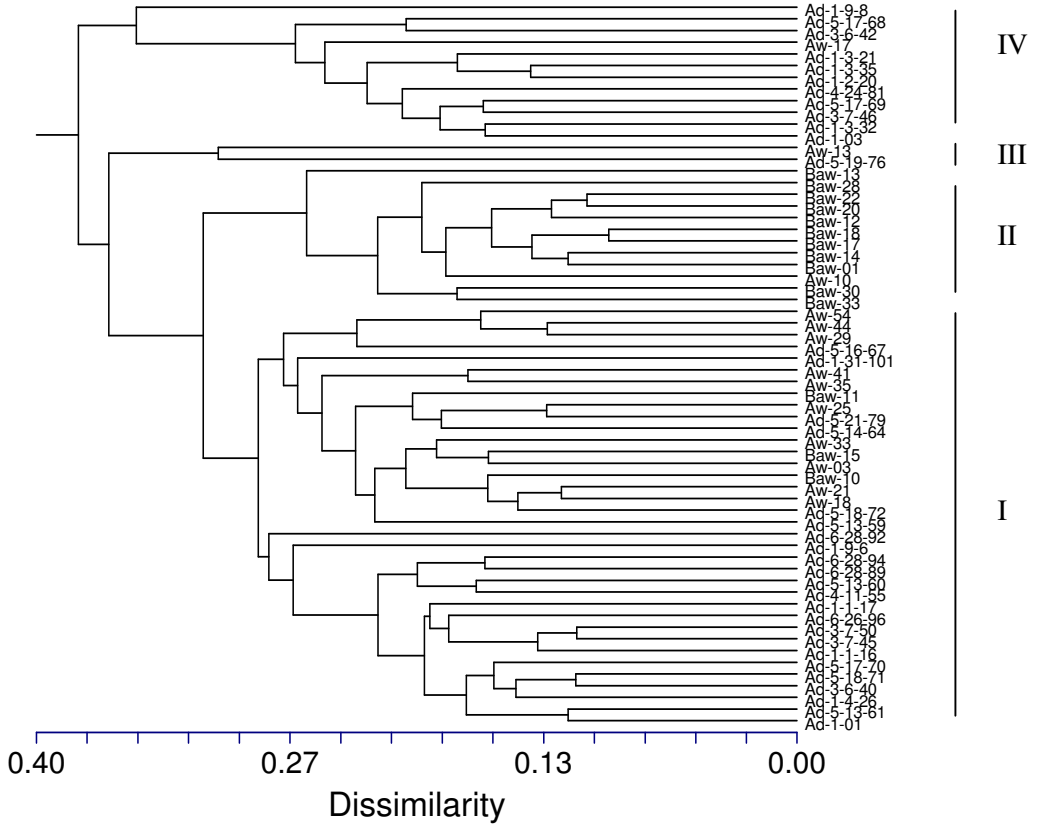

Figure 6.2 Dendrogram of traditional Ethiopian highland maize accessions derived by UPGMA from the dissimilarity matrix of the morphological data

The dendrogram generated based on a combined SSR and AFLP data set showed three major clusters (Figure 6.3). Cluster I consisted of 20 accessions, all collected from the Northern agroecology. Accessions in this cluster had short plant height (average 178.5 cm) and matured earlier (average 123 days) than any of the other clusters. Cluster II consisted of 19 accessions collected from three agroecologies. Accessions in this cluster

were tall plants (on average 220 cm) and they needed more than 150 days to reach maturity. The group also had the highest mean values for ear traits (18.2 cm in ear length, 30 kernels per row, 11 rows per ear and 3884 kg ha $^{-1}$ in grain yield). Cluster III contained 23 accessions, characterized by tall and late maturing plants that had broad and long leaves. This cluster also had the lowest mean values for all of the ear traits.

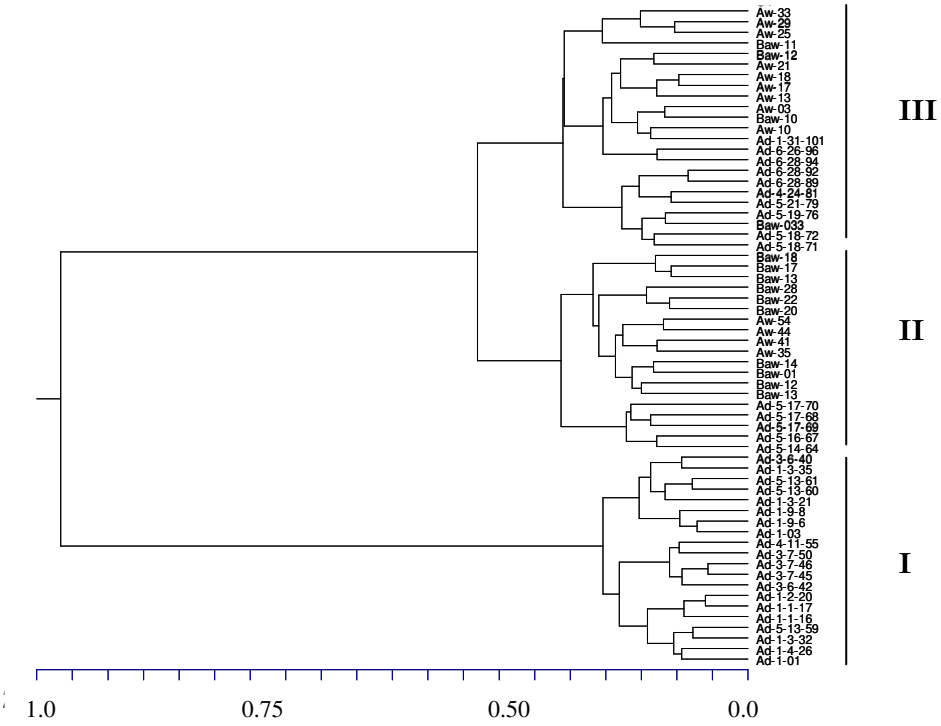

Figure 6.3 Dendrogram of the 62 traditional Ethiopian highland maize accession based on the Ward minimum variance method applied to the dissimilarity matrix generated by Nei and Li dissimilarity coefficients of the pooled AFLP and SSR data

6.5 DISCUSSION

In this study, AFLP and SSR markers and morphological traits were used to characterize a set of 62 traditional Ethiopian highland maize accessions collected throughout the highlands of Ethiopia. There was high and significant correlation between the SSR and AFLP data. This congruence indicates that the two techniques are equally suited for the analysis of genetic diversity in maize. This study allowed us to distinguish three groups of maize accessions with distinctive genetic profiles and morphological traits, which will be useful for breeding, collection and conservation strategies in the highlands of Ethiopia.

The 62 accessions represent genetic diversity in a much larger set of 287 accessions collected from different highland regions of Ethiopia. The broad range in the means of accessions for the various traits implies great potential for the development of improved open-pollinated varieties, inbred lines and hybrids for these regions. The existence of broad morphological and agronomical diversity among the highland maize accessions is further substantiated by principal component analysis (Table 6.2), which indicated that the total variation was fairly distributed across all of the morphological and agronomical traits.

In this study, SSR marker polymorphism was screened on agarose gel system, which is less costly and more widely available (Senior and Heun 1993; Senior *et al*., 1998). In this study, higher genetic diversity values were obtained for AFLP than for SSR (Table 6.3). The reason might be the difference in marker screening systems (agarose gels for

SSR and polyacrylamide gels for AFLP) and data collection procedures (automated for AFLP, manual scoring of alleles for SSR). Acrylamide gels have greater resolving power than agarose gels. This is especially true for dinucleotide repeats whose amplification products are difficult to resolve on agarose gels and inaccuracies in scoring due to the production of stutter bands. Unlike the present study, lower mean value of genetic dissimilarity for AFLP than SSR were reported by Pejic *et al*. (1998) for maize and by Uptmoor *et al*. (2003) for sorghum genotyped by automatic DNA sequencers in polyacrylamide gels.

As in other studies, AFLP analysis in Ethiopian traditional maize accessions detected many polymorphic bands and is an efficient method for diversity study. With a single combination of selective primers, the average number of bands detected was 81.3 per accession, of which 89.5% were polymorphic. As expected, the assay efficiency index of AFLPs was far superior to that of SSRs (72.6 *vs*. 4.9, Table 6.3). However, the proportion of polymorphic fragments was 10% higher for SSRs (Table 6.3). The high AEI of the AFLP markers is due to the large number of loci detected per AFLP primer combination. The low AEI of SSR can, however, be increased by using multiplexing techniques, whether in PCR reaction or at the time of loading (Mitchell *et al*. 1997). In addition, more than 2000 SSRs have already been mapped onto maize chromosomes so that the genome can be uniformly sampled, which increases the precision of genetic diversity estimates and is useful for locating quantitative trait loci (QTLs).

The distribution of values for morphological dissimilarity and genetic dissimilarity (calculated with SSRs and AFLPs) differed substantially (Table 6.4). The

morphological dissimilarity covered a greater range, but was significantly skewed towards small values (Figure 6.1). Comparing the two marker types, although there was little difference in the range, SSRs had the lowest dissimilarity values while the AFLPs data were skewed towards higher values (71% of the values were higher than 0.5). These data suggest that SSR marker can better differentiate pairs of accessions than AFLPs that show a low level of genetic variation between them. The subsets of the sample show that accessions collected from the Northern agroecology were on average more dissimilar than accessions collected from the Western and Southern agroecologies as measured by SSRs and AFLPs (Table 6.4). This partly reflects the frequent introduction of high yielding and uniform varieties in the surrounding intermediate altitudes of the Western and Southern regions, which might have replaced some traditional varieties.

To provide an objective comparison, correlations between distance matrices calculated on the basis of AFLP, SSR and morphological data were examined using a Mantel test (Table 6.5). As shown in Table 6.5, the correlation between SSR and morphological data was higher than between AFLP and morphological data. This might be because the frequency of SSR was significantly higher in ESTs (transcribed regions) than in genomic DNA across all species (Morgante *et al*., 2002). The results suggest that SSR markers may be a better choice for marker-traits association genetic studies of open-pollinated maize accessions than AFLP. Working with 16 ryegrass varieties Roldan-Ruiz *et al*. (2001) reported a correlation value of $r = -0.06$ between AFLP and 15 morphological characters. In comparison with ryegrass, traditional Ethiopian highland maize accessions appear to be environmentally more stable, as suggested by the higher agreement between

phenotypic and molecular distances and indicates that the observed phenotypic variation was at least partly caused by genetic factors. The correlation between the two molecular markers was higher than the morphology (Table 6.5). Therefore, when compared with DNA fingerprinting techniques, morphological traits are relatively less reliable and efficient for precise discrimination of closely related accessions and analysis of their genetic relationships. Despite this limitation, morphological traits are useful for preliminary evaluation because it is fast, simple, and can be used as a general approach for assessing genetic diversity among morphologically distinguishable accessions.

This study allowed us to distinguish three groups of maize accessions, with distinctive genetic profiles and morphological traits. The first group constitutes the early maturing, short-statured accessions (Figure 6.3, cluster I), which were collected from the Northern agroecology from which they probably acquired earliness. The second group includes the tall, high yielding varieties (Figure 6.3, cluster II), which are currently the most important landraces grown in the Southern and Western parts of Ethiopia. The third group includes tall, late maturing and low yielding accessions (Figure 6.3, cluster III), which are being cultivated in some parts of the Northern, Western and Southern highlands of Ethiopia. Therefore, accessions from the Northern agroecology may be used as base materials for the development of improved varieties for the drier parts in the highlands of Ethiopia, because these accessions are able to grow and produce under very harsh environmental conditions (drought, poor soils, excessive radiation, etc), and have adaptation traits (e.g. short flowering, short ear and plant height and narrow leaf), while accessions from the Western and Southern agroecologies can be used for the development of high yielding varieties suitable for high potential maize growing regions of Ethiopia.

CHAPTER 7

ASSOCIATION OF SIMPLE SEQUENCE REPEATS WITH QUANTITATIVE TRAITS IN ETHIOPIAN HIGHLAND MAIZE ACCESSIONS AND THE EFFECT OF ADMIXTURE

7.1 ABSTRACT

The use of molecular markers to identify quantitative trait loci has become a key approach in modern plant breeding programs. In this study, 62 maize accessions collected from different highlands of Ethiopia were analyzed using 20 microsatellite markers and 98 marker alleles were detected. Multiple regression analysis was carried out to detect association between SSR alleles and grain yield, 1000-seed weight, number of kernels per row, days to silking, plant height and days to maturity. Seven SSR alleles were found to be significantly associated with grain yield. For each of the six traits, marker alleles explained between 53 and 71% of the phenotypic variation. Marker loci were identified that were associated with important traits for adaptation to specific environments. Genomic regions represented by SSR alleles *phi054-2 and phi037-3* were strongly selected in water stress areas (on average 94.3% of accessions in the drier Northern region had these alleles), whereas SSR alleles *umc2129-1*and *phi054-4* were selected in high rainfall areas (81.5% of accessions in the Western and Southern agroecologies had these alleles). Genes linked to these alleles could contribute towards one or more traits determining adaptation to specific agroecologies. These results can be used as a basis for future population based association studies in African maize populations. In addition, such studies will also be useful for the development of improved varieties using conventional and/ or marker assisted breeding suitable for specific target environments.

Key words: Association mapping, maize, quantitative traits, regression analysis, SSRs

7.2 INTRODUCTION

Maize (*Zea mays* L.) is an excellent crop to investigate the location and behavior of genetic factors that underlie quantitative trait variation. Many quantitative traits in maize have already been extensively investigated using conventional biometric approaches (Hallauer and Miranda, 1988). Genetic markers have received extensive attention as a tool to improve knowledge about the genetics of various traits, and to enhance breeding efficiency (Stuber, 1989). Earlier studies by Stuber *et al*. (1982) and Edwards *et al*. (1987) have shown the utility of co-dominant molecular markers (isozymes) in identifying quantitative trait loci (QTLs) and following their transmission in sexual crosses. These early studies were limited by the relatively small number of markers available at that time.

With the advent of polymerase chain reaction (PCR) technology, there has been an explosion of DNA markers that made it possible to solve various problems in plant genetics (reviewed in Kumar, 1999). Microsatellites, or simple sequence repeats (SSRs) are one of the PCR-based molecular marker systems that have been extensively used in plant genetic studies because it combines many desirable properties including co-dominance, high variability, rapid and simple assays, and uniform genome coverage (Powell *et al*., 1996). As a result, it is frequently applied in genetic diversity studies in maize inbred lines and populations (Senior *et al*., 1998; Matsuoka *et al*., 2002; Warburton *et al*., 2002), to identify and map QTLs for grain yield and yield components in maize (Thornsberry *et al*., 2001; Mohammadi *et al*., 2002). In maize, thousands of SSRs have been developed and mapped (Senior and

Heun, 1993; Senior *et al.*, 1996; Sharopova *et al*., 2002). These properties of SSRs make them a good marker system for association mapping studies.

In crop genetic studies, recombinant inbred lines have been used very successfully for mapping QTLs to 10–30 cM regions (Stuber *et al*., 1999). Association studies based on linkage disequilibrium (LD) have also been found to be useful for the identification of the actual genes represented by QTLs (Remington *et al*., 2001). These authors suggested that only polymorphisms with extremely tight linkage to a locus with phenotypic effects were likely to be significantly associated with the trait in large randomly mating populations, providing much finer resolution than genetic mapping. Essentially this means that single nucleotide polymorphisms (SNPs) or SSRs inside genes need to be used for association genetic studies. However, association mapping using RAPDs and genetically diverse rice germplasm (Virk *et al*., 1996) allowed the prediction of performance in other rice germplasm for six quantitative traits with 97.6% accuracy. These results suggest that genome-wide markers such as SSRs may indeed allow accurate ranking of germplasm for associated traits. It was also suggested that potential advantages of population level association mapping over conventional mapping include (i) the detection of QTLs that vary across a wide spectrum of the germplasm rather than just between two parental lines, and (ii) the fact that QTLs for many quantitative traits can be studied in the same experimental populations.

In the present study, associations between discrete alleles of SSR loci and phenotypic variation of six quantitative traits were examined. The results from these genotype

/phenotype associations were compared with published information from conventional mapping studies in maize. In addition, the possible effect of admixture on association mapping was evaluated in this study.

7.3 MATERIALS AND METHODS

7.3.1 Selection of plant materials

Previously, we analyzed 180 maize accessions collected from different highland regions in Ethiopia for 15 morphological and agronomic traits (Chapter 3). Principal component and cluster analyses grouped these 180 accessions into four main clusters. For the present study 62 accessions were chosen from the dendrogram (Figure 7.1) by stratification from all clusters to represent the different agroecologies of Ethiopia and the range of morphological and agronomic variation observed in the field. From each accession, 20 plants were selected at random to record plant height (cm), number of kernels per row and 1000 seed weight (g), while days to 50% silking (days from sowing to the stage when 50% plants have emerged silk) days to 50% maturity (days from sowing to the stage when 50% plants form black layer at the tip of the kernels) and yield (kg ha^{-1}) were recorded on plot bases.

7.3.2 SSR analysis

The 62 accessions were genotyped with 20 polymorphic SSR markers. The procedures for DNA extraction, primer selection, PCR reaction, gel electrophoreses,

data scoring and the SSR primers used (including repeat motifs, chromosomal location) are described in detail in Chapter 5.

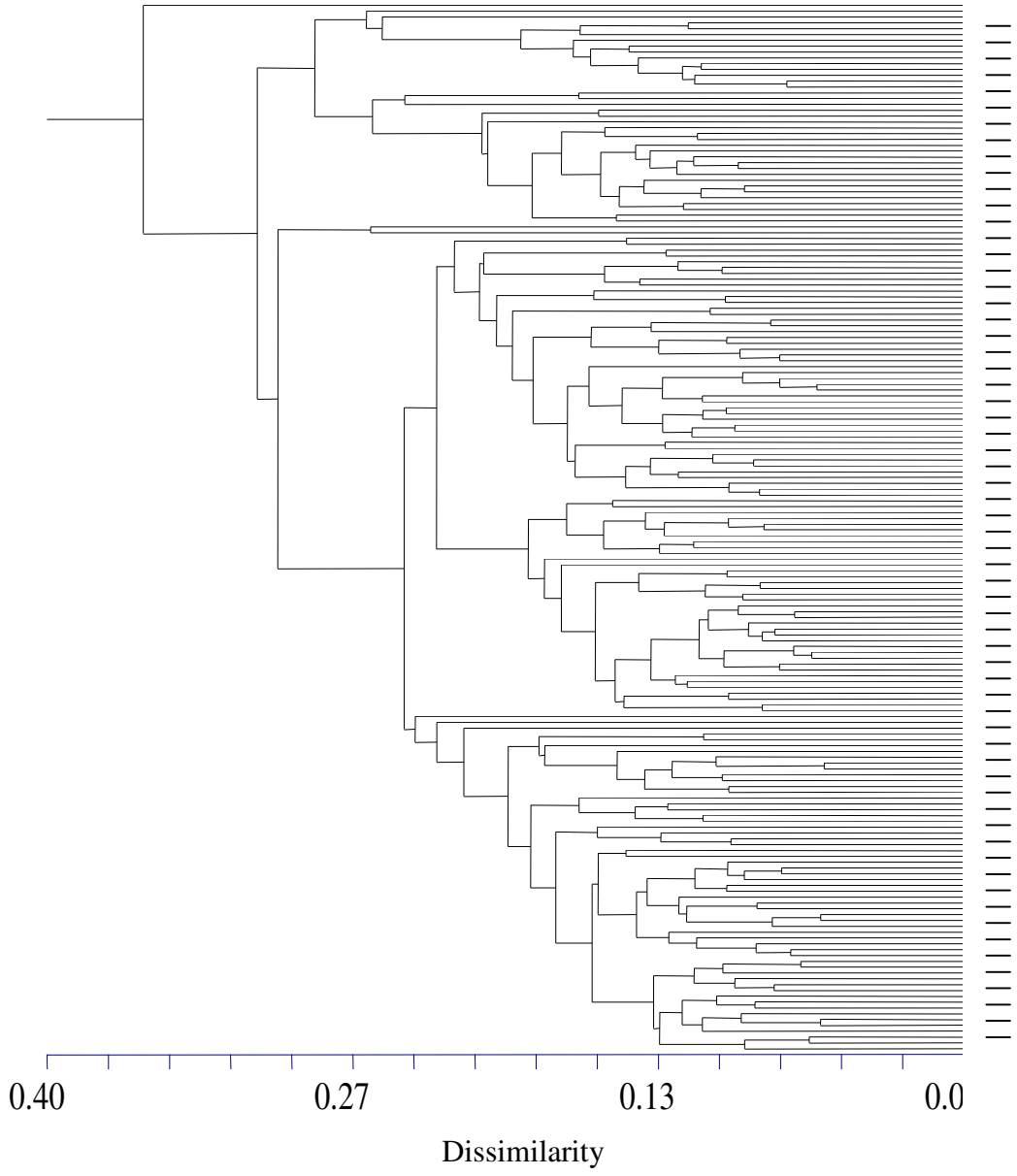

Figure 7.1 Dendrogram resulting from cluster analysis of standardized morphological and agronomic data from 180 maize accessions using Euclidean distance and UPGMA clustering. Sixty-two accessions were selected for molecular analysis (indicated by horizontal bars as root of the dendrogram)

7.3.3 Statistical analysis

Analysis of variance was performed for all measured traits in order to test the significance of variation among accessions. Genotypic correlation coefficients were calculated for all pair-wise combination of traits. Ninety- eight polymorphic SSR alleles were recorded as a binary matrix of 0 for absence and 1 for presence. To identify possible significant alleles with phenotypic variance, a two-step regression approach was used. In the first step, a multivariate variable selection procedure was employed to find those alleles showing significant effect on the phenotypic traits and fulfilling the assumptions of multiple linear regression models (normality, linearity, constant variance, negligible multi-colinearity and increased predictability). In the second step, full regression model was established with selected markers as follows: $Y = b_0 + b_1X_1 + b_2X_2 + \ldots + b_jX_j + e_i$.

This model relates the variation in the dependent variable (Y = accession means of the phenotypic trait) to a linear function of the set of independent variables X_j, representing the SSR alleles. The b_j terms are the partial regression coefficients that specify the empirical relationships between Y and X_j, and e is the residual unexplained variation in Y that includes environmental variation.

To correct experimental-wise Type I error rate of 0.05 at Bonferroni multiple comparison threshold (Neter *et al*., 1990), the appropriate significance level for a single test should be adjusted to 0.05/98 = 0.0005. However, the more stringent the significance level, the greater the bias in detecting QTLs having larger effects (Georges *et al*., 1995;

Stuber *et al*., 1999). Such QTLs have high heritabilities and can easily be manipulated by traditional breeding practices. Therefore, a significance level of alpha at 0.01 was used as a general indicator of the associations between SSR markers and quantitative traits variation to give more emphasis on those chromosomal regions that show a relatively minor effect. The multiple regression coefficients (R^2) indicate the proportion of the total phenotypic variation explained jointly by the SSR alleles. To predict the phenotypic variation, we used the 'leave out one at a time' method where multiple regression analysis was applied to 61 of the 62 accessions at a time and the trait value of the 62 th accession predicted. All statistical analyses were performed using NCSS (Jerry, 2000) statistical software.

7.4 RESULTS

7.4.1 Phenotypic trait variance

Analysis of variance revealed that there were highly significant differences among the mean trait values of the accessions for all the traits studied, suggesting that a high degree of genetic diversity for the traits (Table 6.1). Six traits, which represented different developmental stages in maize, were analyzed for marker traits association (Table 7.1). Days to silking was positively and significantly correlated with plant height, days to maturity, number of kernels per row and 1000 seed weight. Grain yield was positively correlated with 1000 seed weight and number of kernels per row but negatively with plant height (Table 7.1).

Table 7.1 The genotypic correlation coefficients among the six traits of the 62 traditional Ethiopian highland maize accessions used in study

Traits	Days to silking	Plant height	Days to maturity	Kernels per row	Seed weight	Yield
Days to silking	1	0.76**	0.75**	0.29*	0.19*	-0.19ns
Plant height		1	0.83**	0.47**	0.36**	-0.47**
Days to maturity			1	0.37**	0.41**	-0.15ns
Kernels per row				1	0.35**	0.70**
1000 seed weight					1	0.87**
Yield ha $^{-1}$						1

* Significant at $p = 0.05$, ** significant at $p = 0.01$, ^{ns} = not significant

7.4.2 Markers and phenotypic trait associations

A total of 98 polymorphic SSR alleles were scored for the purpose of the present analysis (Table 7.2). To eliminate possible correlations among SSRs alleles, which can cause false positives in the detection of marker-trait association, and that affect the estimation of multiple gene effects; three SSRs alleles (*bnlg182-1*, *phi079-3* and *umc2129-5*) were removed during the first step of the variable selection procedure because they showed multicollinearity (as determined by their eigenvalues, which were near to zero, data not shown). The normal probability plots (one of multiple regression assumptions used to test normality) for the six traits showed that all of the residuals fall within the 95% confidence limits (Figure 7.2).

Table 7.2 The 20 SSR loci used and the designation of the polymorphic alleles scored at each locus

SSR	Allele	SSR	Allele
nc003	1 to 8	phi054	1 to 5
umc2190	1 to 3	phi453121	1 to 5
umc1632	1 to 3	umc2152	1 to 3
phi042	1 to 6	umc2040	1 to 6
bnlg182	1 to 7	phi037	1 to 3
phi079	1 to 3	umc1357	1 to 3
phi034	1 to 4	phi015	1 to 5
phi026	1 to 6	umc2129	1 to 5
bnlg2190	1 to 9	umc1537	1 to 3
phi021	1 to 5	umc1153	1 to 8

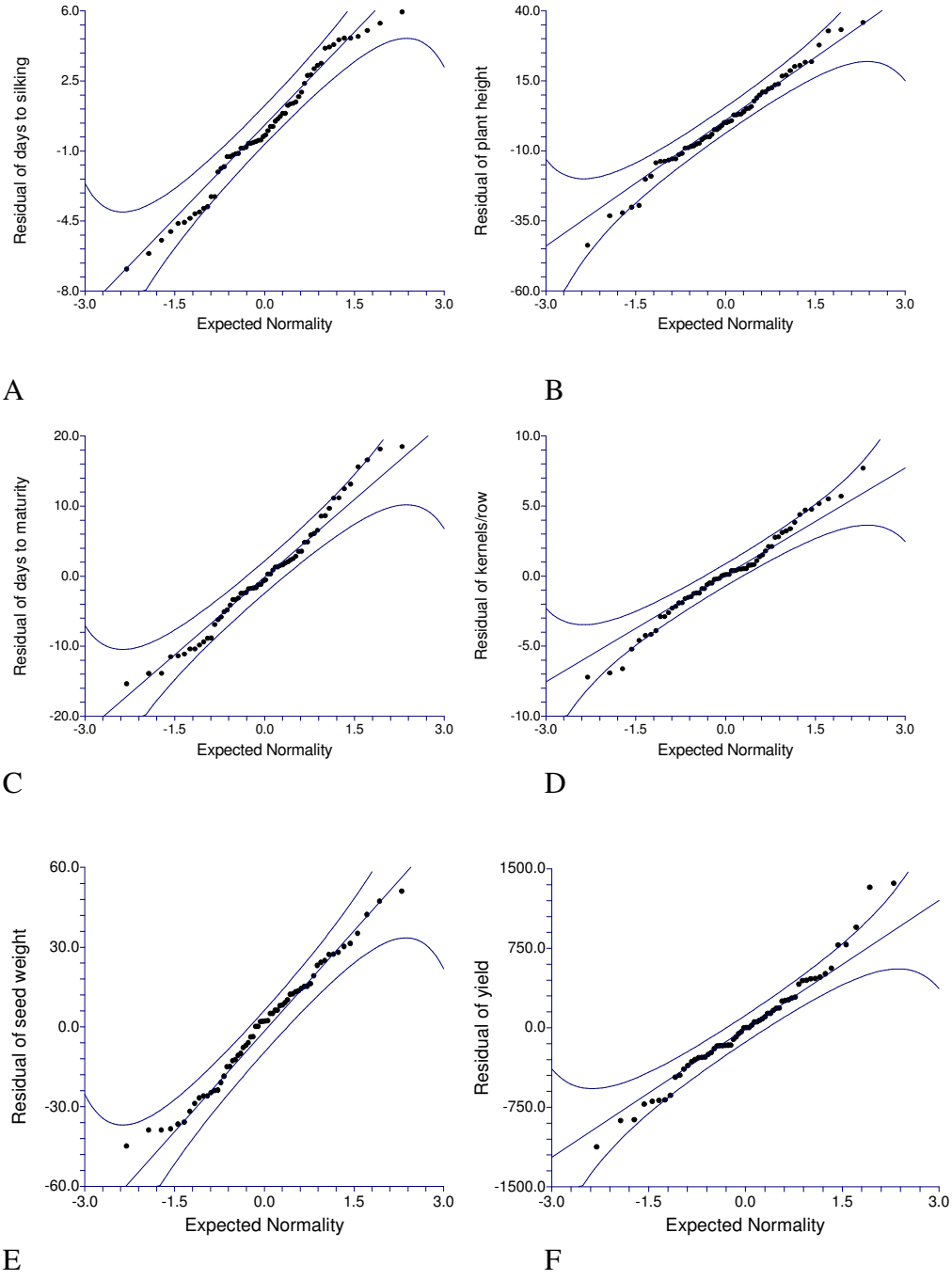

Figure 7.2 Normal probability plot for (A) days to silking; (B) plant height; (C) days to maturity; (D) kernels/row; (E) seed weight and (F) yield showing that all the residuals fall within the confidence limits of the normal probability plot

7.4.3 Detection of chromosomal regions affecting the quantitative traits

Seven SSR alleles, located in six chromosomes (1,4,7,8,9 and 10) were significantly associated with grain yield and explained 53% of the phenotypic variation (Table 5.1 and Table 7.3). Among the chromosomal regions associated with grain yield, the SSR marker on chromosome 1 (*phi037-3*) was also associated with days to maturity and plant height, but in different directions (Table 7.3 & 7.4). The correlations between yield and plant height was negative and significant, while yield and maturity was negative but not significant (Table 7.1). Six SSR alleles were associated with kernels per row (Table 7.3). These alleles explained 65% of total phenotypic variability of the trait (Table 7.3). Five SSR alleles distributed on four chromosomes (1,2, 3 and 6) were associated with 1000 seed weight (Table 5.1 & 7.3). Most of the SSRs alleles associated with seed weight had positive effect and explained jointly 57% of the total phenotypic variation. The correlation between seed weight and yield was significant and high (r = 0.87, Table 7.1).

Nine SSR alleles were significantly associated with days to silking and explained 57% of the total phenotypic variation (Table 7.4). Nine SSR alleles, located in six chromosomes (1,2,3,4,7 and 10) were significantly associated with plant height and explained 60% of the total phenotypic variation (Tables 5.1 and 7.4). Locus *phi054-2* (located on chromosome 1) explained 50.4% of the total variation. Interestingly, this marker also explained 11.5% of days to silking (Table 7.4). The correlation between plant height and days to silking was significant and high (r = 0.76). Three SSR alleles (*umc2129-1, umc2190-1* and *phi037-3*) associated with plant height were also

associated with days to maturity. The correlation between plant height and maturity was 0.83. Ten SSR alleles were significantly associated with days to maturity (Table 7.4). These limited set of alleles explained the majority (71%) of the total phenotypic variation.

Table 7.3 Regression coefficients, Fisher's test of the goodness of fit of the model and R^2 value of for ear traits in traditional Ethiopian maize accessions

Kernels/row		Seed weight		Yield	
Marker[1]	Regression coefficient[2]	Marker	Regression coefficient	Marker	Regression coefficient
phi021-2	-2.7	nc003-4	17.9	phi034-3	4.9*
umc2040-3	2.4	bnlg2129-7	65.8*	phi026-2	-8.8
phi037-3	5.6*	phi453121-3	74.0*	bnlg2190-4	-8.7*
phi015-2	-3.5	umc2040-3	-37.0*	bnlg2190-9	-15.7*
phi015-4	-11.9*	phi037-3	37.7*	phi037-3	5.2
umc2190-4	-2.7			umc1357-1	-7.0*
				phi015-1	12.3
Intercept	33.9		290.9		26.2
F (model)	0.001		0.000000		0.000000
R^2 (model)	0.65		0.57		0.53

[1] See table 7.1 for key to markers

[2] Each regression coefficient is significant at p= 0.01 levels.

*, Significant at multiple comparison thresholds

Table 7.4 Regression coefficients, Fisher's test of the goodness of fit of the model and R^2 value of for morphological traits in traditional Ethiopian maize accessions

Days to silking		Plant height		Days to maturity	
Marker[1]	Regression coefficient[2]	Marker	Regression coefficient	Marker	Regression coefficient
phi054-2	-11.4	nc003-5	60.0	umc2190-1	-10.9
phi054-3	-2.9	umc2190-1	-19.0	umc1632-3	17.7*
phi054-4	11.0*	umc2190-2	-22.8	phi026-1	24.7*
phi037-3	-6.8	bnlg182-4	23.9	phi054-4	18.8
phi015-5	2.7	phi021-2	-17.3	umc2040-2	-13.8*
umc2129-1	4.3*	phi054-2	-50.4	umc2040-5	8.9
umc1153-1	6.0	phi453121-1	13.3	phi037-3	-13.6*
umc1153-3	4.2	phi037-3	-32.8*	umc1357-1	16.7*
umc1153-4	-3.4*	umc2129-1	22.6*	umc2129-1	10.6*
				umc1153-4	-5.1
Intercept	49.9		240.0		120.8
F (model)	0.00001		0.00001		0.0001
R^2 (model)	0.57		0.60		0.71

[1] See table 7.1 for key to markers

[2] each regression coefficient is significant at the $p = 0.01$ level

*, Significant at multiple comparison thresholds

7.4.4 Allelic frequency in contrasting agroecologies

The frequency of some of the significant alleles varied according to selection environments (Figure 7.3). In the Northern agoecology, there were higher frequencies

of *phi037-3* and *phi054-2* alleles. Conversely the frequency of *umc2129-1* was higher in the Western and Southern agroecologies. No significant difference between agroecologies was detected for allele *phi015-1* and *phi054-4*, which had high frequencies in both agroecologies (Figure 7.3). The flowering traits (number of days to silking, plant height and days to maturity) were different between the Northern accessions and the Western and Southern accessions. Generally, plants from the Northern accessions had lower trait values for all these traits. The Northern accessions also differed very much in allele frequencies at *phi037-3* and *phi054-2*, *umc2129-1* and *phi054-4,* suggesting that these accessions were differentiated from the other agroecologies. The results indicates that the presences of admixture in experimental populations (i.e. population substructure).

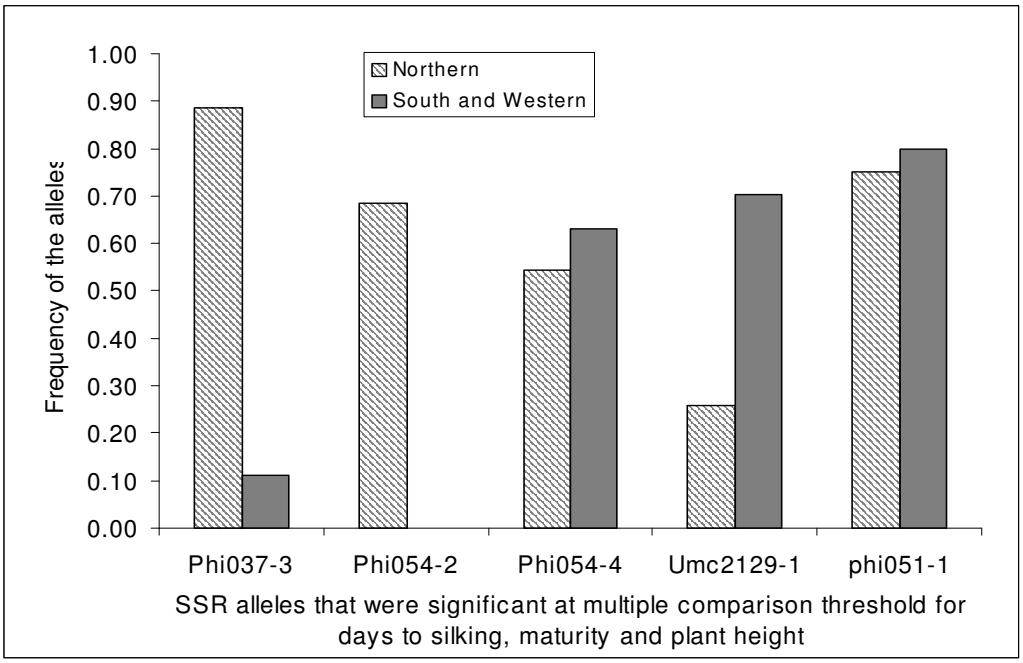

Figure 7.3 Mean allelic frequencies of SSR alleles selected for adaptation in contrasting agroecologies in Ethiopian highlands

7.4.5 Prediction of quantitative variation using significant markers

Predictions of the performance of accessions for each of the six traits were made based upon their marker profiles (Table 7.5 and 7.6). The maximum number of accessions with the observed values fall out side the 95% confidence interval of the predicted values for any trait was 3 out of 62 accessions (Table 7.5 and 7.6). Across the six traits, in nine cases the observed values fall out side of the prediction. These differences accounted for only 2.4% of the total 372 accessions/traits combinations to which the models were fitted. The correlation between the observed and predicted values of each trait was highly significant and ranged from 0.75 (seed weight) to 0.85 (for yield) with an overall mean of 0.80.

Table 7.3 Accessions name, their observed and predicted performance for ear traits in traditional Ethiopian highland maize accessions

	Yield		Kernels/row		Seed weight	
Accessions	Observed	Predicted	Observed	Predicted	Observed	Predicted
Ad-1-01	2926.1	2744.3	26.0	25.9	323	310.6
Ad-1-03	2387.0	2744.3	26.0	23.3	289	292.8
Ad-1-9-6	1582.8	1861.3	23.0	25.6	229	255.0
Ad-1-9-8	1961.6	1861.3	21.0	22.9	317	320.9
Ad-1-1-16	2796.0	2744.3	26.0	25.6	245	255.0
Ad-1-1-17	2490.2	2744.3	24.0	25.6	249	255.0
Ad-1-2-20	2443.6	2744.3	21.0	25.6	260	255.0
Ad-1-3-21	2632.1	2744.3	19.0	25.6	240	255.0
Ad-1-4-26	3153.2	2944.3	29.0	25.6	229	255.0
Ad-1-3-32	3274.7	3441.4	18.0	22.2	**306**	**255.0**
Ad-1-3-35	2714.5	2553.8	26.0	25.6	257	255.0
Ad-3-6-40	2082.5	2253.8	22.0	22.2	263	255.0
Ad-3-6-42	3570.5	3127.3	24.0	25.9	266	292.8
Ad-3-7-45	2334.4	2253.8	23.0	22.5	268	292.8
Ad-3-7-46	2250.1	2253.8	21.0	22.5	282	292.8
Ad-3-7-50	2463.4	2744.3	24.0	23.3	299	292.8
Ad-4-11-55	3904.9	3617.9	23.0	23.3	320	292.8

Ad-5-13-59	4079.9	3617.9	28.0	23.3	308	292.8
Ad-5-13-60	3866.1	3617.9	19.0	23.3	280	292.8
Ad-5-13-61	3333.4	3617.9	26.0	28.3	283	255.0
Ad-5-14-64	2412.1	2734.9	30.0	25.6	248	255.0
Ad-5-16-67	4064.1	3617.9	23.0	19.8	295	292.8
Ad-5-17-69	2051.0	2051.0	**19.0**	**25.9**	274	310.6
Ad-5-17-68	2498.8	2734.9	23.0	22.5	325	310.6
Ad-5-17-70	3219.5	2944.3	22.0	23.3	292	310.6
Ad-5-18-71	2928.0	2744.3	26.0	22.2	283	272.9
Ad-5-18-72	2276.2	2744.3	27.0	25.6	343	329.5
Ad-5-19-76	1995.6	1861.3	19.0	20.5	410	385.1
Ad-5-21-79	2812.6	2744.3	26.0	25.6	309	347.4
Ad-4-24-81	2353.0	2744.3	25.0	25.6	281	272.9
Ad-6-28-89	2576.4	2744.3	20.0	22.2	273	272.9
Ad-6-28-92	**1347.5**	**2225.0**	**24.0**	**31.2**	270	291.0
Ad-6-28-94	2924.4	3098.6	22.0	22.0	293	291.0
Ad-6-26-96	2358.5	2225.0	28.0	28.5	267	291.0
Ad-1-31-101	2234.7	2422.1	28.0	28.5	285	308.8
Aw-03	**3703.5**	**2922.1**	29.0	28.5	328	308.8
Aw-10	1708.4	1708.4	31.0	33.9	311	308.8
Aw-13	1316.3	1839.1	26.0	31.2	339	308.8
Aw-17	2299.8	2039.1	29.0	28.5	264	308.8
Aw-18	2655.6	2744.3	34.0	28.3	308	272.9
Aw-21	2943.1	2922.1	31.0	33.9	351	308.8
Aw-25	2281.4	2922.1	30.0	33.9	277	308.8
Aw-29	4243.6	2922.1	36.0	33.9	333	308.8
Aw-33	3199.3	2922.1	32.0	31.2	373	356.8
Aw-35	2053.4	2225.0	37.0	33.9	321	308.8
Aw-41	2480.9	2225.0	33.0	31.2	280	308.8
Aw-44	4282.3	3922.1	31.0	31.2	322	308.8
Aw-54	3699.8	3322.1	34.0	28.5	270	308.8
Baw-01	2225.7	2422.1	30.0	28.5	314	291.0
Baw-10	2882.5	2922.1	34.0	33.9	315	308.8
Baw-11	1749.2	2631.6	33.0	33.9	300	291.0
Baw-12	2680.4	2334.5	34.0	31.2	270	308.8
Baw-13	2472.1	2922.1	36.0	33.9	296	291.0
Baw-14	3004.2	2950.9	**36.0**	**28.3**	308.5	320.9
Baw-15	3241.6	3441.4	33.0	28.3	288	272.9
Baw-17	2386.2	2225.0	32.0	31.2	336	308.8
Baw-18	1847.9	1642.0	30.0	31.2	276	291.0
Baw-20	1305.2	1531.6	30.0	31.2	301	308.8
Baw-22	2168.0	2225.0	33.0	33.9	273	308.8
Baw-28	2598.4	2339.1	34.0	33.9	309	308.8
Baw-30	2058.8	2225.0	35.0	33.9	356	308.8
Baw-33	**2229.5**	**3098.6**	34.0	28.8	360	328.7

*Values in bold indicate cases in which the observed value falls out side of the 95% confidence interval of the value predicted by the multiple locus model.

Table 7.4 Accessions name, their observed and predicted performance for morphological traits in traditional Ethiopian highland maize accessions

Accession	Silking		Plant height		Days to maturity	
	Observed	Predicted	Observed	Predicted	Observed	Predicted
Ad-1-01	64.5	64.0	200	203.7	143.0	142.7
Ad-1-03	66.5	67.0	182.5	184.1	126.0	131.9
Ad-1-9-6	67.5	64.7	188.5	203.7	137.0	123.9
Ad-1-9-8	58.0	63.5	162.5	186.4	115.0	123.9
Ad-1-1-16	75.0	73.7	195	186.4	135.0	128.9
Ad-1-1-17	72.0	67.4	231.5	227.7	135.0	123.9
Ad-1-2-20	63.5	64.8	161	186.4	122.0	128.9
Ad-1-3-21	66.5	67.4	186.5	184.1	108.0	121.9
Ad-1-4-26	69.5	70.8	201.5	203.7	136.0	137.8
Ad-1-3-32	64.0	64.0	176	176.0	130.0	131.8
Ad-1-3-35	62.0	66.1	185	190.4	110.0	123.9
Ad-3-6-40	67.5	63.2	211.5	214.3	142.0	123.9
Ad-3-6-42	63.5	62.9	207.5	216.2	125.0	123.7
Ad-3-7-45	76.5	75.1	218.5	235.8	143.0	153.4
Ad-3-7-46	70.5	70.8	191	213.2	141.0	142.7
Ad-3-7-50	75.5	70.8	220	193.6	143.0	131.9
Ad-4-11-55	72.5	70.8	217	206.9	168.0	149.5
Ad-5-13-59	67.5	70.8	237.5	226.6	142.0	142.7
Ad-5-13-60	71.0	70.8	234.5	226.6	145.0	142.7
Ad-5-13-61	67.0	70.8	214	203.7	143.0	142.7
Ad-5-14-64	74.0	75.1	219	226.4	144.0	153.4
Ad-5-16-67	66.5	67.4	177	190.4	136.0	137.7
Ad-5-17-69	67.5	69.6	174	184.1	122.0	131.9
Ad-5-17-68	67.5	70.8	180.5	206.9	123.0	131.9
Ad-5-17-70	70.5	75.1	207	193.4	145.0	142.5
Ad-5-18-71	71.5	71.7	209.5	213.0	145.0	148.4
Ad-5-18-72	72.0	67.4	231	231.0	142.0	144.5
Ad-5-19-76	73.5	74.7	231.5	195.7	146.0	148.4
Ad-5-21-79	70.5	71.7	192	193.4	144.0	137.5
Ad-4-24-81	71.5	70.8	197.5	170.8	127.0	131.9
Ad-6-28-89	76.0	75.1	230	213.0	162.0	153.4
Ad-6-28-92	67.0	71.1	243.5	247.1	166.0	169.2
Ad-6-28-94	72.5	71.1	237.5	223.2	165.0	176.2
Ad-6-26-96	69.5	70.3	225	205.9	162.0	146.4
Ad-1-31-101	69.0	70.8	198	205.9	145.0	143.5
Aw-03	74.0	72.1	230.5	228.5	164.0	161.2
Aw-10	68.5	69.1	275.5	245.8	159.0	154.2
Aw-13	74.5	70.3	256	223.2	164.0	154.4
Aw-17	63.0	67.4	192	205.9	132.0	147.4

Aw-18	68.5	69.0	205	213.0	141.0	143.5
Aw-21	71.5	67.4	209.5	223.2	147.0	143.5
Aw-25	70.0	70.5	209	226.2	133.0	129.5
Aw-29	67.0	73.9	**200.5**	**245.8**	134.0	145.4
Aw-33	75.5	70.5	227.5	226.2	163.0	161.0
Aw-35	63.5	67.4	236.5	236.6	148.0	146.4
Aw-41	65.5	67.4	208	223.2	136.0	146.4
Aw-44	64.0	69.0	**199**	**245.8**	137.0	140.4
Aw-54	71.0	70.8	191.5	205.9	136.0	134.7
Baw-01	77.5	75.1	250	228.5	158.0	163.1
Baw-10	66.5	67.1	207.5	226.2	144.0	145.3
Baw-11	69.0	75.1	209.5	245.8	125.0	136.5
Baw-12	76.0	74.7	242.5	245.8	139.0	145.3
Baw-13	77.0	72.5	245.5	245.8	158.0	162.2
Baw-14	74.5	75.1	236.5	213.0	158.0	145.6
Baw-15	75.5	69.6	206	203.7	152.0	152.6
Baw-17	76.0	75.1	264.5	265.8	163.0	162.2
Baw-18	80.5	75.1	278.5	245.8	168.0	162.2
Baw-20	78.5	75.4	288	268.7	162.0	145.4
Baw-22	76.5	73.7	282.5	236.6	160.0	151.5
Baw-28	78.5	75.1	274.5	259.2	158.0	159.2
Baw-30	74.5	71.2	246.5	245.8	167.0	162.2
Baw-33	73.5	78.1	254	250.1	158.0	156.2

* Values in bold indicate cases in which the observed value falls out side of the 95% confidence interval of the value predicted by the multiple locus model.

7.6 DISCUSSION

The association between molecular markers and quantitative traits is vital in choosing parents for crossing, mapping studies and for marker-assisted selection. The application of this approach in maize is feasible due to the large number of molecular markers linked to specific traits (Hoisington and Ribaut, 1998). In this study, multiple regressions analyses were carried out to test the association between SSR markers and phenotypic performance in traditional Ethiopian highland maize accessions. This results suggest that SSRs can be used to study population level association, provided

that population structure is accounted for. In addition, agroecology-specific markers were identified and genes linked to these alleles could be involved in adaptation to agroecology.

Falconer and Mackay (1996, p. 357) designated QTL explaining 10% of the phenotypic variance or their standardized effects exceeding 0.5, respectively, as "large." In this study, the cumulative effects of the significant markers explained 53 to 71% of the phenotypic variation. Some markers explained substantial proportions of variation in different traits (two to five, Table 7.3 and 7.4). One possible reason for the associations between SSR alleles and quantitative traits could be linkage disequilibria involving chance association due to population admixture. Cluster analysis of SSR data (Chapter 5) suggests that there were differences in allelic frequencies between the agroecologies (20 out of 98 SSR alleles with extreme differentiation in allelic frequencies among agroecologies). However, genetic linkage between the SSR alleles and quantitative traits is mostly the explanation of the results found in this study. Marker *phi037-3* located on chromosome 1(bin no.1.08) explained significant variation in grain yield, 1000 seed weight and number of kernels /row. Similarly, using conventional QTLs analysis (Mohammadi *et al*., 2002) found that four markers in chromosome 1 had the highest genetic effect on yield and yield components in maize. Another SSR allele, *phi054-2*, which was mapped on chromosome 10 explained 50.4% and 11.4% of the total phenotypic variation for plant height and days to silking, respectively. This might be due to the high heritability of these traits (greater than 70%, Table 3.2) and the presence of major genes. However, for yield the maximum value explained by a single marker was only

15.7% of the total phenotypic variation. The possible explanation could be low heritability (17%, Table 3.2) of this trait.

In this study the detection of nine SSR markers distributed in six chromosomes affecting plant height (Table 5.1 and 7.4) was consistent with previous studies in maize. In 22 maize populations studied by 10 groups of researchers, 105 QTLs or mutations affecting plant height have been reported (Lin *et al*., 1995). These results indicate the complexity of this trait. In this study, the SSR allele *phi037-3* which is located on chromosome 1 (bin no. 1.08) was significantly associated with plant height and this allele was fixed in the Northern accessions (all Northern accessions were on average less that 200 cm tall), but almost null (10%) in the Western and Southern agroecologies (all accessions on average were more than 230 cm in height). This may be a result of close linkage of this allele to maize height mutants, *br*2, *an*1, and *br*1 (Coe and Neuffer, 1993). These mutants associated with interval 1.08-1.10 of the maize RFLP map. The same locus (*phi037-3*) was also significantly associated with days to silking. Using conventional QTL analysis, this chromosomal region (1.08-1.10) had significant effects on flowering time in maize (Veldboom *et al*., 1994). Another SSR allele *phi054-3* on chromosome 10 (bin no. 10.0) was significantly associated with days to silking. Koester *et al*. (1993) reported significant effects on flowering time associated with interval 10.08 of the maize RFLP map.

Earliness is an important adaptive trait for the Northern agroecology because the region is characterized by low moisture stress, poor soil and excessive radiation. In contrast long duration is important in the Western and Southern agroecologies because these regions have high rainfall, fertile soils and long growing period. This

was reflected, for example, in allelic frequencies differences at *phi037-3* and *umc2129-1.* Farmers in the Northern agroecology select for early maturity to synchronize growing cycles with the available moisture. As a result, accessions in this agroecology had higher frequency of allele *phi037-3,* which explained large proportion of phenotypic variation in days to silking and maturity in the negative direction (Table 7.4). In contrast, in the Western and Southern agroecologies farmers selected for long maturity and accessions in these agroecologies had fixed allele *umc2129-1*, which explained 4.3, 10.7 and 22.6 % of the variation in days to silking, days to maturity and plant height, respectively in the positive direction (Table 7.4). Therefore, the genomic regions defined by the *phi037-3* and *umc2190-1* may be potential targets for manipulation by marker-assisted selection for the development of varieties for moisture stress and high rainfall areas, respectively. However, not all alleles were selected for adaptation. For example, irrespective of environmental differences in agroecologies of Ethiopia, allele *phi015-1* (which explained 12.3% of the variation in grain yield) was the most common allele in all agroecologies and this allele is likely to confer significant improvements in grain yield.

The accessions involved in this study represent genetic diversity in a much larger set of 287 accessions collected from the different highlands of Ethiopia. Therefore, it could be possible to conclude that Ethiopian traditional maize accessions contain considerable useful genes both controlling earliness, tallness and grain yield and associated traits, which could be exploited by various breeding schemes. The strong selection pressures for adapted accessions in contrasting environments indicate that specific breeding programs are required for each region. The significant association

between molecular markers and quantitative traits is useful for selection of parents for crossing. In general, accessions from the Northern agroecology could be used for the development of drought tolerant varieties. On the other hand, accessions from the Southern and Western agroecologies could be useful for the development of long maturing high yielding varieties. In addition, accessions from the Northern agroecology in one hand and the Western and Southern accessions on the other hand represent phenotypic extremes for important traits and are also polymorphic for SSR loci that are linked to these traits**.** These accessions could be useful as parents for future crossing programs.

If genetic linkage is the main cause of the associations, then the use of molecular markers will improve selection of desirable genotypes. Furthermore, if diverse germplasm is characterized for important traits requiring specialized assessment conditions such as stress tolerance, then marker data can be efficient means of predicting suitable genotypes before field evaluation, which is costly and time consuming. Up to 5% of the predictions can in theory be expected to differ from the observed values (Virk *et al*., 1996). Therefore our predictions were highly reliable (97.6% accuracy). Using a linear model approach (Nuel *et al*., 2000) showed that 29% of field-trial savings could be obtained for less than 5% of errors in the pre-screening of maize varieties by utilizing marker information. The results of this study could be useful in planning of breeding programs for the improvement of various traits and provide preliminary information for marker-assisted selection.

CHAPTER 8

CONCLUSIONS AND RECOMMENDATIONS

Scope and objectives of the study

Genetic variation is of fundamental importance in many areas of basic and applied biology. It is a prerequisite for breeding, collection and conservation strategies, as well as being the basis for genetic fingerprinting. The development of DNA-based marker systems such as RFLP, RAPD, AFLP and SSR, in particular, has made it possible to access the genetic variation in virtually any organism. Most DNA–based genetic diversity studies in maize to date, have focused on inbred lines rather than on out-bred populations. This is because the cost and complexity of the analysis process is lower in inbred lines than in populations. In order to analyze large numbers of out-bred varieties or populations, bulking or pooling strategies have to be employed (Kolliker *et al*., 2001, Rebourg *et al*., 2001). However, very limited information is available on the use of bulked DNA samples to estimate genetic variability and genetic relationships for out-bred maize accessions using molecular markers.

In the highland areas of Ethiopia, maize contributes greatly to household food security. Maize cultivars that are used in the highland regions are well adapted, but low yielding open-pollinated varieties resulted from centuries of selection. These accessions may be grouped into a smaller number of reproductively isolated populations, which may each have accumulated specific genetic adaptations for different highland conditions. Effective plant breeding and crop improvement programs depend on the availability of crop genetic diversity. In the search for diverse breeding materials, landraces are usually the major source of variation. This study represents a detailed investigation of maize accessions collected from the highlands of Ethiopia using morphological traits and

molecular markers.

Morphological variability

The morphological study revealed a wide range of variation among the highland accessions, which could be grouped into distinct phenotypic clusters. This allowed the selection of representative accessions from different highland areas of Ethiopia and from different morphological classes for molecular marker analysis. Specific recommendations from this phase of the study include that highland maize breeders in Ethiopia should give importance to kernels per row as a selection criterion to increase grain yield, because simple phenotypic selection of plants bearing higher numbers of kernels per row may lead to trait improvement of up to 37.8%.

Use of bulked leaf samples for genetic diversity studies

Individual and bulked - AFLP analyses suggested that bulking equal amounts of leaf samples from 15 individual plants per accession was an effective means of producing representative profiles of the out-bred varieties, thereby reducing the cost of DNA extraction and subsequent marker analysis. Cluster analysis of AFLP data grouped most of the Northern accessions into one cluster, while the Western and Southern accessions were grouped together. Partitioning of the total genetic variability into within and between agroecologies revealed substantially higher variability within than between agroecologies. The consistency of the results between Wards clustering method and partitioning of variation showed that bulked-AFLP marker analysis is

useful for studying the genetic diversity and relationships of out-bred crop varieties or populations.

In another experiment, we validated the genetic relationships detected by AFLP analysis with bulked SSR analysis. Analysis of individuals included in the bulk showed that an allele is generally detectable in the bulked sample when present in a proportion greater than 1 out of 15 (7%) of the individuals within DNA pooled-samples. Similar to the result of the AFLP analysis, cluster analysis performed on SSR data showed that accessions collected from the Northern agroecology were clearly distinct from the Western and Southern accessions, but there was very little differentiation between the Western and Southern accessions. These results indicated that the two techniques revealed similar genetic patterns in traditional Ethiopian highland maize accessions and are useful to study genetic diversity in out-bred varieties or populations. However, bulked SSR analysis is more useful than bulked – AFLP because SSR map positions are frequently known, which is useful to associate traits of interest with molecular markers and to compare association study result across different studies.

Genetic diversity in traditional Ethiopian highland maize accessions

High mean genetic diversity values were obtained in the traditional Ethiopian maize accessions with both marker systems. The values of overall genetic diversity obtained in this study agreed with diversity values reported in other studies of out-bred maize populations collected from Europe using RFLP (Rebourg *et al.*, 2001 mean, 0.55) and

from tropical and sub tropical maize populations and synthetics using SSR data (Pinto *et al*., 2003 mean, 0.48). Therefore, it is possible to conclude that Ethiopian traditional maize accessions may be as diverse as European and tropical maize populations and synthetics. The presence of such variation suggests (1) adaptation of these accessions to many niches of Ethiopian highlands over a long period of time and (2) that highland farmers have maintained a wide diversity of maize accessions to meet their social, economic, cultural and ecological needs. In fact, farmers' selection for desirable agronomic traits may be the main force in shaping the genetic dynamics of the highland maize accessions. Landraces and farmers are interdependent and both are in need of each other for their survival (Appa Rao *et al*., 1998).

In all studies (morphological, AFLP and SSR), differences in genetic diversity measures were detected among agroecologies (Table 6.4). These results suggest that farmers' selection criteria are different for the different agroecologies in accordance with the specific environmental conditions and/or local environmental condition may have influenced the genotypic constitution of these accessions. Farmers in the Northern agroecology have selected plants with short stature, short flowering time and that mature earlier to synchronize with moisture stress and short growing periods which are prevalent in the Northern agroecology. However, farmers in the Western and Southern agroecologies have selected tall, broad leaf and late maturing maize accessions, because these regions have higher rainfall, more fertile soils and longer growing periods (EARO, 2000). This underscores the importance of considering indigenous knowledge of genetic diversity in attempts to collect and evaluate local accessions.

Implication for breeding and conservation

A combination of morphology and molecular analyses revealed three groups of maize accessions, which could be useful to design different breeding programs for the highlands of Ethiopia. The first group (Figure 6.3, I) constitutes the early maturing, short statured accessions which were collected from the Northern agroecology from which they probably acquired earliness. This group is of special interest to maize breeding in Ethiopia where drought is one of the key production constraints. Therefore, these materials could be used as basis for the development of drought tolerant cultivars and need to be given a special emphasis in further crossing programs. The second group (Figure 6.3, II) includes the tall, high yielding varieties which are currently the most important landraces grown in the Southern and Western parts of Ethiopia. Therefore, accessions in this group can be used for the development of high yielding varieties suitable for high potential maize growing regions of Ethiopia. The third group (Figure 6.3, III) includes tall, late maturing and low yielding accessions, which are being cultivated in some parts of the Northern, Western and Southern highlands of Ethiopia and might be used as source of fuel (dry stalk).

The highland collection of maize accessions is maintained at the Institute of Biodiversity Conservation (IBC), Ethiopia. These 287 landraces were collected from all over the country. These accessions contain high genetic diversity as can be seen from morphological and molecular diversity measured in this study. For example, the Northern accessions are able to grow and produce under very harsh environmental conditions (drought, poor soils, excessive radiation, etc) and have relatively good

yield. On the other hand, the Western and Southern maize accessions have evolved with the parasitic weed *Striga spp* for several hundred years and are able to produce higher yield. In Africa, this weed is a significant pest of maize and sorghum, and to date, little resistance has been reported in maize (Hoisington *et al*., 1999). Thus, Ethiopian highland maize accessions could be a source of variability for weed tolerance (S*triga spp*.), drought and cold tolerance. More accessions should be collected from all agroecologies to capture these valuable and unique germplasm.

Association mapping and the effect of admixture

The results of a preliminary association study using SSR marker alleles revealed putative associations of SSR markers with quantitative variation among traditional Ethiopian highland maize accessions. These results will assist the selection of divergent parental accessions for the development of inbred lines, thereby maximizing heterosis in future hybrid-breeding programs. Significant differences in alleles frequency were observed between the agroecologies (except between the Western and Southern) and most likely resulted in some cases from the selection for traits that contribute to overall performance and adaptation. The study also gives a preliminary method for selection of useful accessions prior to field evaluation, which is costly and time consuming. For example, the SSR allele *phi037-3* was associated negatively with flowering time, tallness and maturity. This allele also had positive and significant effects for grain yield and 1000 seed weight and number of kernels/row. Therefore, this allele can be used to develop short maturing plants without significantly reducing grain yield.

However, one of the problems of association genetic studies is the presence of population substructure (admixtures) as observed in this study, which causes the detection of false positives. In this study, two groups of population structure were observed. The first group was accessions collected from the Northern agroecology and the second group was accessions collected from the Southern and Western agroecologies. Out of 98 SSR alleles scored in this study, eight individual alleles were fixed in the Northern accessions and 12 alleles were fixed in the Western and Southern accessions (Chapter 5). Hence, these 20 alleles could be significantly associated (by chance) with any traits that are different between the two groups. Therefore, future population level association studies should consider the population structure in order to increase the power of detecting true markers associated with traits of economic importance.

In conclusion, this research is the first attempt to characterize the genetic diversity found within traditional Ethiopian highland maize accessions. Maize breeders in Ethiopia will benefit from knowledge of the genetic relationships of the highland maize accessions so that they can improve adaptive and agronomic traits through traditional hybridization and/or marker-assisted selection techniques. The analytical tools outlined in this dissertation can be useful for detecting genetic variation among open pollinated crop varieties and will aid in the conservation and preservation of unique genetic diversity present in germplasm collections. Production stability and global food security are linked to the conservation and exploitation of worldwide genetic resources and this research attempts to add to that body of knowledge.

REFERENCES

Aguirre, G., Bellon, M.R. and Smale, M. 1998. A regional analysis of maize biological diversity in Southeastern Guanajuato, Mexico. Pp. 22. CIMMYT Economics Working Paper 98-06. Mexico.

Ahn, S. and Tanksley, S.D. 1993. Comparative linkage maps of the rice and maize genomes. *Proc. Natl. Acad. Sci. USA* **90**: 7980-7984.

Ajmone-Marsan, P., Castiglioni, P., Fusari, F., Kuiper, M. and Motto, M. 1998. Genetic diversity and its relationship to hybrid performance in maize as revealed by RFLP and AFLP markers. *Theor. Appl. Genet*. **96**:219-227.

Alika, J.E., Aken'ova, M.E. and Fatokun, C.A. 1993. Variation among maize (*Zea mays* L.) accessions of Bendel State, Nigeria. Multivariate analysis of agronomic data. *Euphytica* **66**: 65-71.

Appa Rao, S., Mengesha, M.H., Gopal Reddy, V. and Prasada Rao, K.E. 1998. Collecting and evaluation of sorghum germplasm from Rwanda. *Plant Genet. Res. Newslett.* **114**: 26–28.

Arias, C.A.A., deSouza, C.L. and Takeda, C. 1999. Path coefficient analysis of ear weight in different types of progeny in maize. *Maydica* **44**: 251-262.

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

Barbosa-Neto, J.F., Hernandez, C.M., O'Donoughue, L.S. and Sorrells, M.E. 1997. Precision of genetic relationship estimates based on molecular markers. *Euphytica* **98**: 59-67.

Beaumont, M.A., Ibrahim, K.M., Boursot, P. and Bruford, M.W. 1998. Measuring genetic distance. Pp.315-325. In A. Karp *et al*. (ed.) Molecular tools for screening biodiversity. Chapman and Hall, London.

Bennetzen, J.L. and Devos, K. 2002. Plant synteny, colinearity and genome evolution. Encyclopedia of Life Sciences. Pp.1-5. Macmillan Publishers Ltd.

Bogyo, T.P., Porceddu, E. and Perrino, P. 1990. Analysis of sampling strategies for collecting genetic material. *Econ. Bot*. **34**: 11-86.

Bouchez, A., Hospital, F., Causse, M., Gallais, A. and Charcosset, A. 2002. Marker-assisted introgression of favorable alleles at quantitative trait loci between maize elite lines. *Genetics* **162**: 1945-1959.

Brown, A.H.D. 1989a. The case for core collections. Pp. 136-156. In A.H.D. Brown *et al*. (ed.), The use of plant genetic resources. Cambridge University press, Cambridge.

Brown, A.H.D. 1989b. Core collections: A practical approach to genetic resources management. *Genome* **31**: 818-824.

Buckler, E.S. and Thornsberry, J.M. 2002. Plant molecular diversity and application to genomics. *Curr. Opin .Plant Biol.* **5**: 107-111.

Buckler, E.S., Thornsberry, J.M. and Kresovich, S. 2001. Molecular diversity, structure and domestication of grasses. *Genet. Res.* **77**: 213-218.

Burr, B., Burr, F.A., Thompson, K.H., Albertsen, M.C. and Stuber, C.W. 1988. Gene mapping and recombinant inbreds in maize. *Genetics* **118**: 519- 526.

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

Causse, M., Santoni, S., Damerval, C., Maurice, A., Charcosset, A., Deatrick, J.

and de Vienne, D. 1996. A composite map of expressed sequences in maize. *Genome* **39**: 418-432.

CIMMYT, 2000. International Maize and Wheat Improvement Center (CIMMYT). CGIAR Research areas of research: Maize (*Zea mays* L.) http://www.cgiar.org/areas/maize.htm.

Coe, E. and Neuffer, M.G. 1993. Genetic loci and linkage map of corn (*Zea mays* L.) (2n = 20), Pp. 6157-6189. In S.J. O'Brien (ed.), Genetic Maps: Locus maps of complex genomes, 6th edition, Cold Spring Harbor Press, Cold Spring Harbor, NY.

Coe, E., Hancock, D., Kowaleski, S. and Polacco, M. 1995. Gene list and working maps. *Maize Genet. Coop. Newslett*. **69**: 191-267.

Crossa, J., Hernandez, C.M., Bretting, P., Eberhart, S.A. and Taba, S. 1993. Statistical genetic considerations for maintaining germplasm collections. *Theor. Appl. Genet.* **86**: 673-678.

CSA. 1998. Central Statistical Authority (CSA): Agricultural sample survey 1997/98. Report on farm management practices of 'belg' season . Vol. **III**. *Statistical bulletins 171*. Addis Ababa, Ethiopia.

CSA. 2001. Central Statistical Authority (CSA): Agricultural sample survey 2000/2001 on area production of major crops, private peasant holdings 'meher' season. Vol. **III**. *Statistical bulletins 190*. Addis Ababa. Ethiopia.

Davis, G.L., McMullen, M.D., Baysdorfer, C., Musket, T., Grant, D., Staebell, M., Xu, G., Polacco, M., Koster, L., Melia-Hancock, S., Houchines K., Chao, S. and Coe, E. 1999. A maize map standard with sequenced core markers, grass genome reference points and 932 expressed sequenced tagged

sits (ESTs) in a 1736-locus map. *Genetics* **152**: 1137-1172.

Demeke, T., Sasikumar, B., Hucl, P. and Chibbar, R.N. 1997. Random amplified polymorphic DNA (RAPD) in cereal improvement. *Maydica* **42**: 133-142.

Doebley, J.F. 1990. Molecular evidence for gene flow among *Zea* species. *Bio Science* **40**: 443- 448.

Doebley, J.F. and Iltis, H.H. 1980. Taxonomy of *Zea* (Gramineae): I. A subgeneric classification with a key to taxa. *Am. J. Bot.* **67**: 982-993.

Doebley, J.F., Goodman, M.M. and Stuber C.W. 1984. Isoenzymatic variation in *Zea* (Gramineae). *Systematic Bot.* **9**: 203-218.

Dubreuil, P., Dufour, P., Krejet, E., Causse, M., de Vienne, D., Gallais. A. and Charcosset, A. 1996. Organization of RFLP diversity among inbred lines of maize representing the most significant heterotic groups. *Crop Sci.* **36**: 790-799.

EARO. 2000. Ethiopian Agricultural Research Organization (EARO), Research strategy plan for maize. Addis Ababa, Ethiopia.

Edwards, M.D., Stuber, C.W. and Wendel, J.F. 1987. Molecular-marker facilitated investigations of quantitative traits in maize. I. Numbers, genomic distribution and types of gene action. *Genetics* **116**: 113-125.

Enoki, A., Sato, H. and Koinuma, K. 2002. SSR analysis of genetic diversity among maize inbred lines adapted to cold regions of Japan. *Theor. Appl. Gene.* **104**: 1270-1277.

Eubanks, M. 2000. International application published under the patent cooperation treaty (PCT)WO00/07434. Novel genetic materials for transmission into maize.

Evans, L.T. and Evans, T.L.T. 1993. Crop evolution, adaptation, and yield. Cambridge University Press; New York.

Everitt, B. 1980. Cluster analysis. 2nd edition, Halstead Press, New York.

Excoffier, L., Smouse, P.E. and Quattro, J.M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* **131**: 479-491.

Eyre-Walker, A., Gaut, R.L., Hilton, H. and Gaut, B.S. 1998. Investigation of the bottleneck leading to the domestication of maize. *Proc. Natl. Acad. Sci. USA* **95**: 4441-4446.

Falconer, D.S. and Mackay, T.F.C. 1996. Introduction to quantitative genetics. 4th edition, Longman Ltd. Group, London.

FAPRI. 2003. U.S. and World Agriculture Outlook. Food and Agricultural Policy Research Institute. Iowa State University. Ames, Ia. U.S. (http://www.fapri.iastate.edu/outlook2003/).

Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author, department of genetics, University of Washington, Seattle, Washington.

Flint-Garcia, S.A., Thornsberry, J.M. and Buckler, E.S. 2003. Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Biol.* **54**: 357-374.

Gardiner, J.M., Coe, E.H., Melia-Hancock, S., Hoisington, D.A. and Chao, S. 1993. Development of a core RFLP map in maize using an immortalized F_2 population. *Genetics* **134**: 917-930.

Gaut, B.S. 2001. Patterns of chromosomal duplication in maize and their implications for comparative maps of the grasses. *Gen. Research* **11**: 55-56.

Gaut, B.S., d'Ennequin, M.T., Peek, A.S. and Sawkins, M.C. 2000. Maize as a model for the evolution of plant nuclear genomes. *Proc. Natl. Acad. Sci. USA* **97**: 7008-7015.

Gauthier, P., Gouesnard, B., Dallard, J., Redaelli, R., Rebourg, C., Charcosset, A. and Boyat, A. 2002. RFLP diversity and relationships among traditional European maize populations. *Theor. Appl. Genet.* **105**: 91-99.

Georges, M., Nielsen, D., Mackinnon, M., Mishra, A., Okimoto, R., Pasquino, A.T., Sargeant, L.S., Sorensen, A., Steele, M.R., Zao, X., Womack, J.E. and Hoeschle, A. 1995. Mapping quantitative traits loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics* **139**: 907-920.

Grossniklaus, U., Koltunow, A. and van-Lookeren-Campagne, M. 1998. A bright future for apomixis. *Trends Plant Sci.* **3**: 415-416.

Guthridge, K.M., Dupal, M.P., Kolliker, R., Jones, E.S., Smith, K.F. and Forster, J.W. 2001. AFLP analysis of genetic diversity within and between populations of perennial ryegrass (*Lolium perenne* L.). *Euphytica* **122**: 191-201.

Hafnagel, H.P. 1961. Agriculture in Ethiopia. Food and Agricultural Organization of the United Nations, Rome, Italy.

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

Hake, S. and Walbot, V. 1980. The genome of *Zea mays*, its organization and homology to related grasses. *Chromosoma* **79**: 251-270.

Hallauer, A.R. and Miranda, J.B. 1988. Quantitative genetics in maize breeding.

Iowa State Univ. Press, Ames, Iowa.

Harjes, C.E., Smith, M.E., McCouch, S.R. and Tanksley, S.D. 1999. Advanced backcross QTL analysis and introgression of perennial teosinte (*Zea diploperennis*) alleles to maize. *Plant and Animal Genome* **VII** Conference, January 17-21, San Diego, California.

Helentjaris, T., Weber, D. and Wright, S. 1988. Identification of genomic locations of duplicate nucleotide sequences in maize by analysis of restriction fragment length polymorphisms. *Genetics* **118**: 353-363.

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

Helentjaris, T., Slocum, M., Wright, S., Schaefer, A. and Nienhuis, J. 1986. Construction of genetic linkage maps in maize and tomato using restriction fragment length polymorphisms. *Theor. Appl. Genet.* **72**: 761-769.

Henandez, Z.E. 1985. Maize and man in the greater Southwest. *Econ. Bot.* **39**: 416-430.

Henry, A.M. and Damerval, C. 1997. High rates of polymorphism and recombination at the opaque-2 locus in cultivated maize. *Mol. Gen. Genet.* **256**: 147-157.

Hoisington, D.A. and Ribaut, J. M. 1998. What can biotechnology offer Africa? - Concrete examples. Six Eastern and Southern Africa Regional Maize Conference. Pp. 12-17, Addis Ababa, Ethiopia.

Hoisington, D.A., Khairallah, M., Reeves, T., Ribaut, J.M., Skovmand, B., Taba, S. and Warburton M. 1999. Plant genetic resources: What can they

contribute toward increased crop productivity? *Proc. Natl. Acad. Sci. USA* **96**: 5937-5943.

Hulbert, S.H., Richter, T.E., Axtell, J.D. and Bennetzen, J.L. 1990. Genetic mapping and characterization of sorghum and related crops by means of maize DNA probes. *Proc. Natl. Acad. Sci. USA* **87**: 4251-4255.

Jaccard, P. 1908. Nouvelles researches sur la distribution florale. *Bull. Soc. Vaudoise Sci. Natl.* **44**: 223-270.

Jerry, H. 2000. Number Cruncher Statistical Systems (NCSS). Statistical software, Utah.

Johnson, A.R. and Wichern, D.W. 1992. Applied multivariate statistical analysis. 3rd edition, Prentice-Hall, Englewood Cliffs, NJ.

Jones, C.J., Edwards, K.J., Castiglion, S., Winfield, M.O., Sala, F., Van de Wiel, C., Bredemeijer, G., Vosman, B., Matths, M., Daly, A., Brettschneider, R., Bettini, P., Buitti, M., Maestri, E., Malcevschi, A., Marmirdi, N., Aert, R., Volkaert, G., Rueda, J., Linacero, R., Vazquez, A. and Karp, A. 1997. Reproducibility testing of RAPD, AFLP, and SSR markers in plants by a network of European laboratories. *Mol. Breed.* **3**: 381-389.

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

Karp, A., Seberg, O. and Buiatti, M. 1996. Molecular techniques in the assessment of botanical diversity. *Ann. Bot.* **78**: 143-149.

Karp, A., Edwards, K., Bruford, M., Vosman, B., Morgante, M., Seberg, O., Kremer, A., Boursot, P., Arctander, P., Tautz, D. and Hewitt, G. 1997. Newer molecular technologies for biodiversity evaluation: Opportunities and

challenges. *Nature Biotechnol.* **15**: 625-628.

Koester, R., Sisco, P. and Stuber, C. 1993. Identification of quantitative trait loci controlling days to flowering and plant height in two near-isogenic lines of maize. *Crop Sci***. 33**: 1209–1216.

Kolliker, R., Jones, E.S., Jahufer, M.Z.Z. and Forster, J.W. 2001. Bulked AFLP analysis for the assessment of genetic diversity in white clover (*Trifolium repens* L.). *Euphyitca* **121**: 305-315.

Kumar, L.S. 1999. DNA markers in plant improvement: An overview. *Biotechnology advances* **17**: 143-182.

Kumar, M.V.N. and Kumar, S.S. 2000. Studies on character association and path coefficient for grain yield and oil content in maize (*Zea mays* L.). *Ann. Agric. Res.* **21**: 73-78.

Leblanc, O., Grimanelli, D., González-de-León, D. and Savidan, Y. 1995. Detection of the apomictic mode of reproduction in maize-Tripsacum hybrids using maize RFLP markers. *Theor. Appl. Genet.* **90**: 1198-1203.

Lee, M., Sharopova, N., Beavis, W.D., Grant, D., Katt, M., Blair, D.L. and Hallauer, A.R. 2002. Expanding the genetic map of maize with intermated B73 x MO17 (IBM) population. *Plant Mol. Bio.* **48**: 453-461.

Levinson, G. and Gutman, G.A. 1987. Slipped-strand mispairing: A major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **4**: 203-221.

Lin, Y.R., Schertz, K.F. and Paterson, A.H. 1995. Comparative analysis of QTLs affecting plant height and maturity across the Poaceae, in refernce to interspecific sorghum population. *Genetics* **141**: 391- 411.

Link, W.C., Dickens, C., Singh, M., Schwall, M. and Melchinger, A.E. 1995.

Genetic diversity in European and Mediterranean fababean germplasm revealed by RAPD markers. *Theor. Appl. Genet.* **90**: 27-32.

Lubberstedt, T., Melchinger, A.E., Dußle, C., Vuylsteke, M. and Kuiper, M. 2000. Relationships among early European maize inbreds: IV. Genetic diversity revealed with AFLP markers and comparison with RFLP, RAPD, and pedigree data. *Crop Sci.* **40**: 783-791.

Lucchin, M., Baraccia, G. and Parrini, P. 2003. Characterization of flint maize (*Zea mays* L. convar. Mays) Italian landraces: Morpho-phenological and agronomic traits. *Genet. Res. Crop Evol.* **50**: 315-327.

Malecot, G. 1948. Les mathematiques de l'heredite. Masson and Cie Paris.

Manly, B.F.J. 1991. Randomization and Monte Carlo methods in biology. Chapman and Hall, New York.

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

Martin, J.M., Blake, T.K. and Hockett, E.A. 1991. Diversity among North American spring barley cultivars based on coefficients of parentage. *Crop Sci*. **31**: 1131-1137.

Matsuoka, Y., Mitchell, S.E., Kresovich, S., Goodman, M. and Dobeley J. 2002. Microsattelite in *Zea* –variability, patterns of mutations and use for evolutionary studies. *Theor. Appl. Genet.***104**: 436-450.

Matsuoka, Y., Vigouroux, Y., Goodman, M.M., Sanchez, J.G., Buckler, E. and Doebley, J. 2001. A single domestication for maize shown by multilocus microsatellite genotyping. *Proc. Natl. Acad. Sci. USA* **99**: 6080-6084.

Melchinger, A.E. 1993. Use of RFLP markers for analysis of genetic relationships

among breeding materials and prediction of hybrid performance. Pp. 621-627. In D.R. Buxton *et al*. (ed.) First international crop science congress. CSSA, Madison, WI.

Melchinger, A.E., Gumber, R.K., Leipert, R.B. and Vuysteke, M. 1998. Predication of testcross means and variance among F_3 progenies of F_1 crosses from testcross means and genetic distances of their parents in maize. *Theor. Appl. Genet.* **96**: 503-512.

Messmer, M.M., Melchinger, A.E., Herrmann, R.G. and Boppenmaier, J. 1993. Relationship among early European maize inbreeds: Comparison of pedigree and RFLP data. *Crop Sci.* **33**: 944-950.

Michelmore, R.W., Paran, I. and Kessli, R.V. 1991. Identification of markers linked to disease resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions using segregenat populations. *Proc. Natl. Acad. Sci. USA* **88**: 9828-9832.

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

Moeller, D.A. and Schaal, B.A. 1999. Genetic relationships among Native American maize accessions of the Great Plains assessed by RAPDS. *Theor. Appl. Genet.* **99**:1061-1067.

Mohammadi, S.A., Prasanna, B.M., Sudan, C. and Singh, N.N. 2002. A microsatellite marker based study of chromosomal regions and gene effects on yield and yield components in maize. *Cellu. Mol. Biol. Lett.***7**: 599-606.

Moreau, L., Lemarié, S., Charcosset, A. and Gallais, A. 2000. Economic efficiency of one cycle of marker-assisted selection. *Crop Sci.* **40**: 329-337.

Morgante, M., Hanafey, M. and Powell, W. 2002. Microsatellites are preferentially associated with non-repetitive DNA in plant genomics. *Nat. Genet.* **30**:194-200.

Mulatu, K., Bogale, G., Tolessa, B., Desalegn, Y. and Afeta A. 1992. Maize production trends and research in Ethiopia. In: B. Tolessa and J.R. Ransom (eds.). Proceeding of the first national maize workshop in Ethiopia. Addis Ababa, Ethiopia.

Myburg, A.A., Remington, D.L., O'Malley, D.M., Sederoff, R.R. and Whetten, R.W. 2001. High-throughput AFLP analysis using infrared dye-labeled primers and an automated DNA sequencer. *BioTechniques* **30**: 348-357.

Nei, M. and Li, W.H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA* **76**: 5269-5273.

Neter, J., Wasserman, W., and Kutner, M.H. 1990. Applied linear statistical models. 3rd edition, Irwin, Inc., Homewood, Illinois.

NMSA. 2000. National Meteorology Service Agency (NMSA), monthly weather report, Feb. 1980 to Dec. 2000. Addis Ababa, Ethiopia.

Nuel, G., Baril C.P. and Robin, S. 2000. Varietal distinctness assisted by molecular markers: A methodological approach. *Acta. Hort.* **546**: 65-71.

Ohta, T. and Kimura, M. 1973. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genet. Res.* 22:201-204.

Pejic, I., Ajmone-Marson, P., Morgante, M., Kozumplick, V., Castiglioni, P., Tarmino, G. and Motto, M. 1998. Comparative analysis of genetic similarity among maize inbred line detected by RFLPs, RAPDs, and AFLPs. *Theor. Appl. Genet.* **97**: 1248-1255.

Pinto, L.R., Vieira, M.L.C., de Souza, C.L.P. and de Souza, A.P. 2003. Genetic diversity assessed by microastellites in tropical maize populations submitted to a high-intensity reciprocal recurrent selection. *Euphytica* **134**: 277-286.

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

Quarrie, S.A., Lazic-Jancic, V., Kovacevic, D., Steed, A. and Pekic, S. 1999. Bulk segregant analysis with molecular markers and its use for improving drought resistance in maize. *J. Exp. Botany* **50**: 1299-1306.

Ragot, M., Gay, G., Muller, J.P. and Durovray, J. 2000. Efficient selection for adaptation to the environment through QTL mapping and manipulation in maize. Pp. 128-130. In: Ribaut, J.M. and Poland, D. (eds.) Molecular approaches for the genetic improvement of cereals for stable production in water-limited environments. A strategic planning workshop held at CIMMYT, Et Batan, Mexico, June 21-25, 1999.

Rebourg, C., Gouesnard, B. and Charcosset, A. 2001. Large-scale molecular analysis of traditional European maize populations: Relationships with morphological variation. *Heredity* **86**: 574-587.

Reif, J.C., Melchinger, A.E., Xia, X.C., Warburton, M.L., Hoisington, D.A., Vasal, S.K., Beck, D., Bohan, M. and Frisch. 2003. Use of SSRs for

establishing heterotic groups in subtropical maize. *Theor. Appl. Genet.* **107**: 947-957.

Remington, D. L., Thornsberry, J. M., Matsuoka, Y., Wilson, L. M., Whitt, S. R., Doebley, J. F., Kresovitch, S., Goodman, M. M. and Buckler, E. S. 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc. Natl. Acad. Sci. USA* **98**: 11479-11484.

Ribaut, J.M. and Hoisington, D. 1998. Marker-assisted selection: new tools and strategies. *Trends Plant Sci.* **31**: 236-239.

Ribaut, J.M ., Edmeades, G., Perotti, E. and Hoisington, D. 2002. QTL analyses, MAS results, and perspectives for drought-tolerance improvement in tropical maize. Pp. 131-136. In: Ribaut, J.M. and Poland, D. (eds.) Molecular approaches for the genetic improvement of cereals for stable production in water-limited environments. A strategic planning workshop held at CIMMYT, Et Batan, Mexico, June 21-25, 1999.

Riek, J.De., Calsyn, E., Everaet, I., van Bockstaele, E. and De Loose, M. 2001. AFLP based alternative for the assessment of distinctness, uniformity and stability in sugar beet varieties. *Theor. Appl. Genet.* **103**: 1254-1265.

Roldan-Ruiz, L., Van Eeuwijk, F.A., Gilliland, T.J., Dubreuil, P., Dillmann, C., Lallemand, J., De Loose, M. and Baril, C.P. 2001. A comparative study of molecular and morphological methods of describing relationships between perennial ryegrass (*Lolium perenne* L.) varieties. *Theor. Appl. Genet.* **103**: 1138-1150.

Saghai-Maroof, M.A., Biyashev, R.M., Yang, G.P., Zhang, Q. and Allard, R.W. 1994. Extraordinarily polymorphic microsatellite DNA in barely: Species

diversity, chromosomal locations and population dynamics. *Proc. Natl. Acad. Sci. USA* **91**: 5466-5470.

SAS. 1993. SAS/STAT users'. Version 6.09. SAS Institute Inc carry, NC.

Sasakawa-Global2000. 2002. Agricultural project in Ethiopia: Activities and output. An assessment report, Sasakawa-Global 2000, Addis Ababa, Ethiopia.

Savidan, Y., Grimanelli, D., Perotti, E. and Leblanc, O. 1998. International application published under the patent cooperation treaty (PCT) WO98/36090. Means for identifying nucleotide sequences involved in apomixis.

Sax, K. 1923. The association of size difference with seed-coat pattern and pigmentation in *Phaseolus Volgaris*. *Genetics* **8**: 552-560.

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

Senior, M.L., Murphy, J.P., Goodman, M.M. and Stuber, C.W. 1998. Utility of SSRs for determining genetic similarities and relationships in maize using an agarose gel system. *Crop Sci*. **38**: 1088-1098.

Senior, M.L., Chin, E.C.L., Lee, M., Smith, J.S.C. and Stuber, C.W. 1996. Simple sequence repeat markers developed from maize sequences found in the GENE BANK database: Map constriction. *Crop Sci.* **36**: 1676-1683.

Sharopova, N., McMullen, M.D., Schulz, L., Schroeder, S., Sancock-Villeda, H., Gardiner, J., Bergstrom, D., Houchins, K., Melia-Hancock, S., Musket, T., Duru, N., Polacco, M., Edwards, K., Ruff, T., Register, J.C., Brouwer, C., Thompson, R., Velasco, R., Chin, E., Lee, M., Woodman-

Clikeman, W., Long, M. J., Liscum, E., Cone, K., Davis, G. and JrCoe, E.H. 2002. Development and mapping of SSR markers for maize. *Plant Mol. Biol.* **48**: 463-481.

Singh, R.K. and Chaudhary, B.D. 1977. Biometrical methods in quantitative genetics analysis. Kalyani Publishers, New Delhi.

Skinner, D.Z., Bauchan, G.R., Auricht, G. and Huges S. 1999. A method for the efficient management and utilization of large germplasm collections. *Crop Sci.* **39**: 1237-1242.

Smith, J.S.C. and Smith, O.S. 1992. Fingerprinting crop varieties. *Adv. Agron.* **47**: 85-140.

Sneath, P.H.A. and Sokal, R.R. 1973. Numerical taxonomy. WH Freeman, San Francisco.

Snedecor, G.W. and Cochran, W.G. 1980. Statistical Methods. 7th edition, Iowa State Univ. Press, Ames.

Sokal, R.R. and Michener, C.D. 1958. A statistical method for evaluating systematic relationships. *Univ. Kansas Sci. Bull*. **38**: 1409-1438.

Souza, E. and Sorrells, M.E. 1991. Relationships among 70 American oat germplasm. I. Cluster analysis using quantitative characters. *Crop Sci.* **31**: 599-605.

Stanton, M.A., Stewart, J.M., Percival, A.E. and Wendel, J.F. 1994. Morphological diversity and relationships in A-genome cottons, *Gossypium arboreum* and *G. herbaceum*. *Crop Sci.* **34**: 519-527.

Statistical for Windows. 1995. Stat Soft. Inc., Computer program manual. Tulsa. OK: Stat Soft. Inc., 2325 East 13th Street Tulsa.

Stuber C.W. 1989. Marker-based selection for quantitative traits. Pp.31-49:In: Proc. XII Congr. EUCARPIA. Parey, Berlin, Hamburg.

Stuber, C.W., Goodman, M.M. and Moll, R.H. 1982. Improvement of yield and ear number resulting from selection at allozyme loci in a maize population. *Crop Sci.* 22:737-740.

Stuber, C.W., Polacco, M. and Senior, M.L. 1999. Synergy of empirical breeding, marker-assisted selection, and genomics to increase crop yield potential. *Crop Sci.* **39**: 1571-1583.

Taba, S., Diaz, J., Franco, J. and Crossa, J. 1998. Evaluation of Caribbean maize accessions to develop a core subset. *Crop Sci.* **38**: 1378-1386.

Tanksley, S.D. and McCouch, S.R. 1997. Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* **277**: 1063–1066.

Tanksley, S.D., Grandillo, S., Fulton, T.M., Zamir, D., Eshed, T., Petiard, V., Lopez, J. and Beck-Bunn, T. 1996. Advance backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium*. *Theor. Appl. Genet.* **92**: 213-224.

Tato, K. 1964. Rainfall in Ethiopia. *Eth. Geog. J.* **2**: 28-36.

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

Thormann, C.E., Ferreira, M.E., Camargo, L.E.A., Tivang, J.G. and Osborn, T.C. 1994. Comparison of RFLP and RAPD markers to estimating genetic relationships within and among cruciferous species. *Theor. Appl. Genet.* **88:** 973-980.

Thornsberry, J.M., Goodman, M.M., Doebley, J., Kresovich S., Nielsen, D. and

Edward, S.E. 2001. *Dwarf*8 polymorphisms associate with variation in flowering time. *Nat. Genet.* **28**: 286-289.

Tuberosa, R., Salvi, S. and Phillips, S. 1998. Bulked segregant analysis confirms the importance of the region near umc89a for days to pollen shed in maize. *Maize Gene. Newsl.* **72**: 71-72.

Tuberosa, R., Salvi, S., Sanguineti, M.C., Landi, P., Maccaferri, M. and Conti, S. 2002. Mapping QTLs regulating morph-physiological traits and yield: Case studies, shortcomings and perspectives in drought-stressed maize. *Annals Botany* **89**: 941-963.

Twumasi-Afriyie, S., Habatmu, Z., Kass, Z., Bayisa, Y. and Sewagegne, T. 2001. Development and improvement of highland maize in Ethiopia. pp.1-12. 2nd national maize workshop of Ethiopia, Addis Ababa, Ethiopia.

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

Van Eeuwijk, F.A. and Baril, C.P. 2001. Conceptual and statistical issues related to the use of molecular markers for distinctness and essential derivation. *Acta. Hort.* **546**: 35-53.

Vavilov, N.I. 1951. Phytogeographic basis of plant breeding: The origin, variation, immunity and breeding of cultivated plants. *Chron. Bot.* **13**: 13-54.

Veldboom, L.R., Lee, M. and Woodman, W.L. 1994. Molecular marker-facilitated studies in an elite maize populations: Linkage analysis and determination of QTL for morphological traits. *Theor. Appl. Genet.* **88**: 7-16.

Virk, P.S., Ford-Lioyed, B., Jackson, M.T., Pooni, H.S., Clemeno, T.P. and Newbury, H.J. 1996. Predicting quantitative variation within rice germplasm using molecular markers. *Heredity* **76**: 296-304.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., VanDe Lee, T., Hornes, M., Frijters, A., Pot, J., Poleman, J., Kuiper, M. and Zabeau, M. 1995. AFLP®: a new technique for DNA fingerprinting. *Nucleic Acids Res.* **23**: 4407-4414.

Vuylsteke, M., Kuiper, M. and Stam, P. 2000a. Chromosomal regions involved in hybrid performance and heterosis: Their AFLP-based identification and practical use in prediction models. *Heredity* **85**: 208-218.

Vuylsteke, M., Mank, R., Brugmans, B., Stam, P. and Kuiper, M. 2000b. Further characterization of AFLP data as a tool in genetic diversity assessments among maize (*Zea mays* L.) inbred lines. *Mol. Breed.* **6**: 265-276.

Vuylsteke, M., Mank, R., Antonise, R., Bastiaans, E., Senior, M.L., Stuber, C.W., Melchinger, A.E., Lubberstedt, T., Xia, X.C., Stam, P., Zabeau, M. and Kuiper M. 1999. Two high-density AFLP linkage maps of *Zea mays* L.: Analysis of distribution of AFLP markers. *Theor. Appl. Genet.* **99**: 921-935.

Wang, R.L., Stec, A., Hey, J., Lukens, L. and Doebley, J. 1999. The limits of selection during maize domestication. *Nature* **398**: 236-239.

Warburton, M.L. and Crossa, J. 2000. Data analysis in CIMMYT applied biotechnology center for fingerprinting and genetic diversity studies. CIMMYT, Mexico.

Warburton, M.L., Zianchun, X., Crossa, J., Franco, J., Melchinger, A.E., Frisch, M., Bohn, M. and Hoisington, D. 2002. Genetic characterization of

CIMMYT inbred maize lines and open pollinated populations using large scale fingerprinting methods. *Crop Sci.* **42**: 1832-1840.

Ward, J.H. 1963. Hierarchical grouping to optimize an objective function. *Am. Stat. Assoc. J.* **56**: 236-244.

Waugh, R., Bonar, N., Baird, E., Thomas, B., Graner, A. and Hayes, P. 1997. Homology of AFLP products in three mapping populations of barely. *Mol. Gen. Genet.* **255**: 311-321.

Welsh, J. and McClell, M. 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res.* **18**:7213-7218.

Wendel, J.F. 2000. Genome evolution in polyploids. *Plant Mol. Biol.* **42**: 225-249.

Wiley, E.O. 1981. Phylogenetics: The theory and practice of phylogenetics and systematics. John Wiley, New York.

Williams, J.G.K., Kubelick, A.R., Livak, K.J., Rafalski J.A. and Tingey, S.V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* **18**: 6531-6535.

Wilson, R.N., Heckman, J.W. and Somerville, C.R. 1992. Gebberellin is required for flowering in *Arabidopsis-thaliana* under short days. *Plant Physiol.* **100**: 403-408.