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.

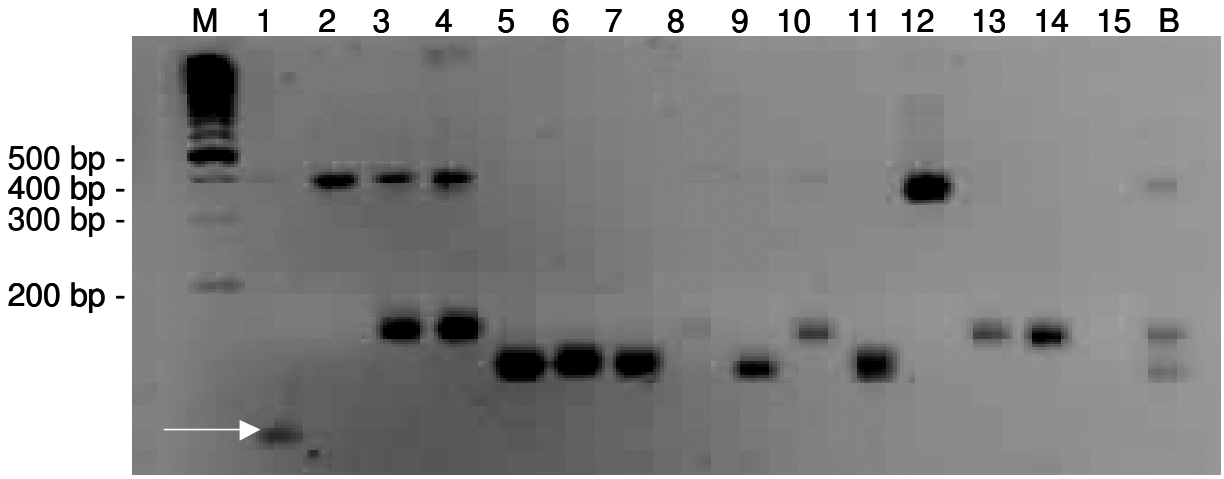

Figure 5.2 The SSR 3% gel image of individual plants (lane 1-15) and the bulk sample (lane B) of the accession 'Baw 10' amplified with locus *phi034*. M is molecular ladder (Gene Ruler 100-bp DNA ladder Plus, Frementas). Arrow shows the band present in individual plants and absent in the bulked sample

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).

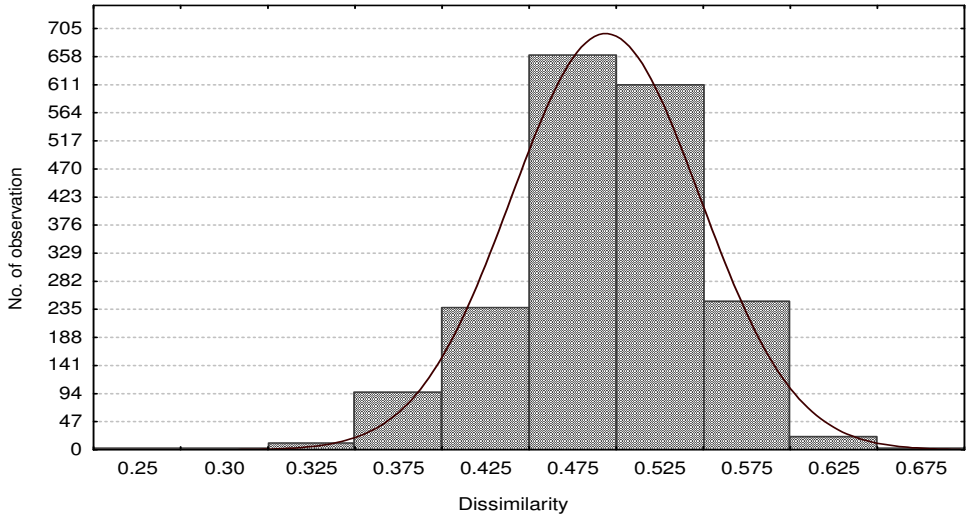

Figure 5.3 Frequency distribution of the genetic dissimilarity pairs of 62 Ethiopian maize accessions based on 20 SSR loci

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.

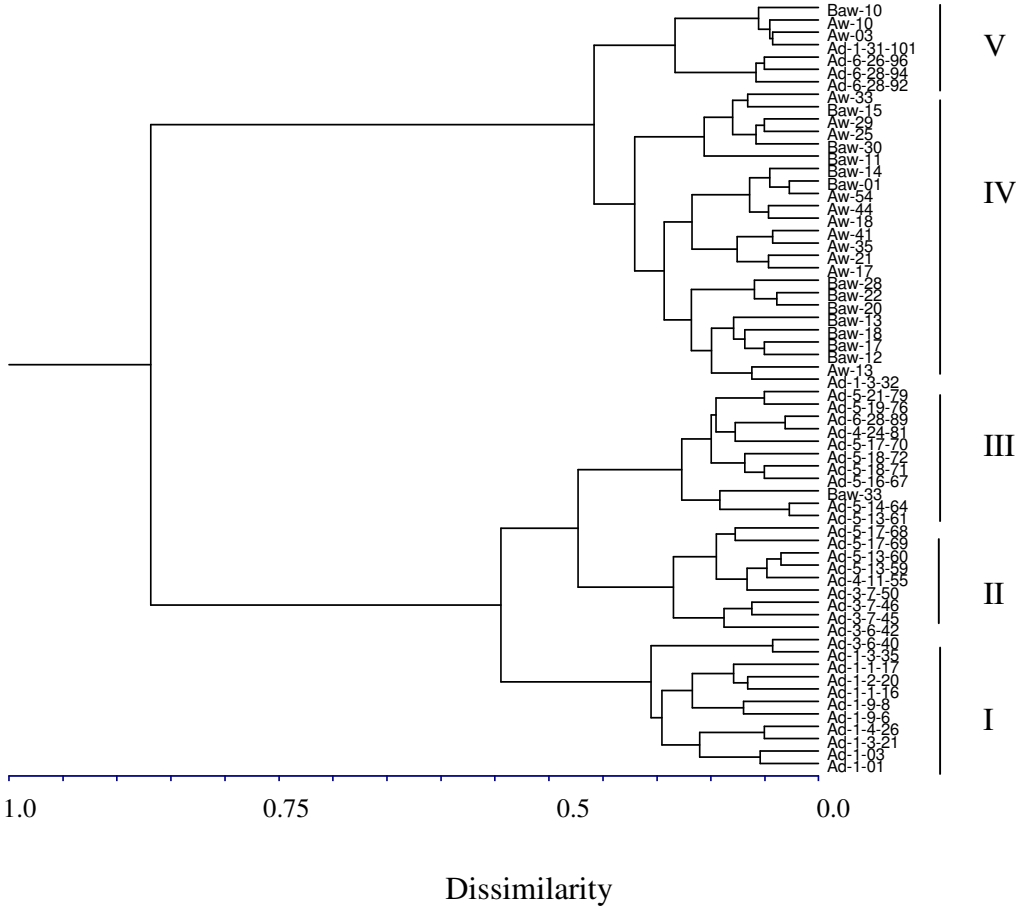

Figure 5.4 Ward minimum variance based dendrogram generated from the Nei and Li dissimilarity matrix showing relationships among 62 traditional Ethiopian highland maize accessions.

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.