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.

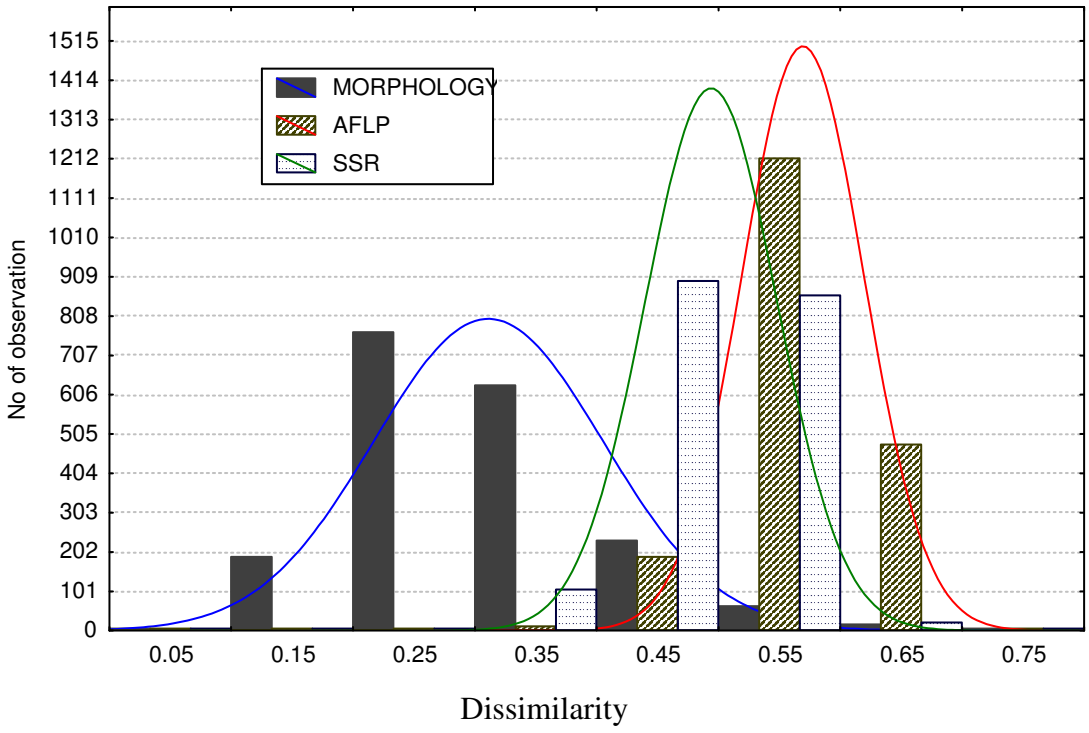

Figure 6.1 Frequency distribution of genetic dissimilarity among pair-wise combinations of 62 traditional Ethiopian highland maize accessions based on morphology, AFLP and SSR data

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).

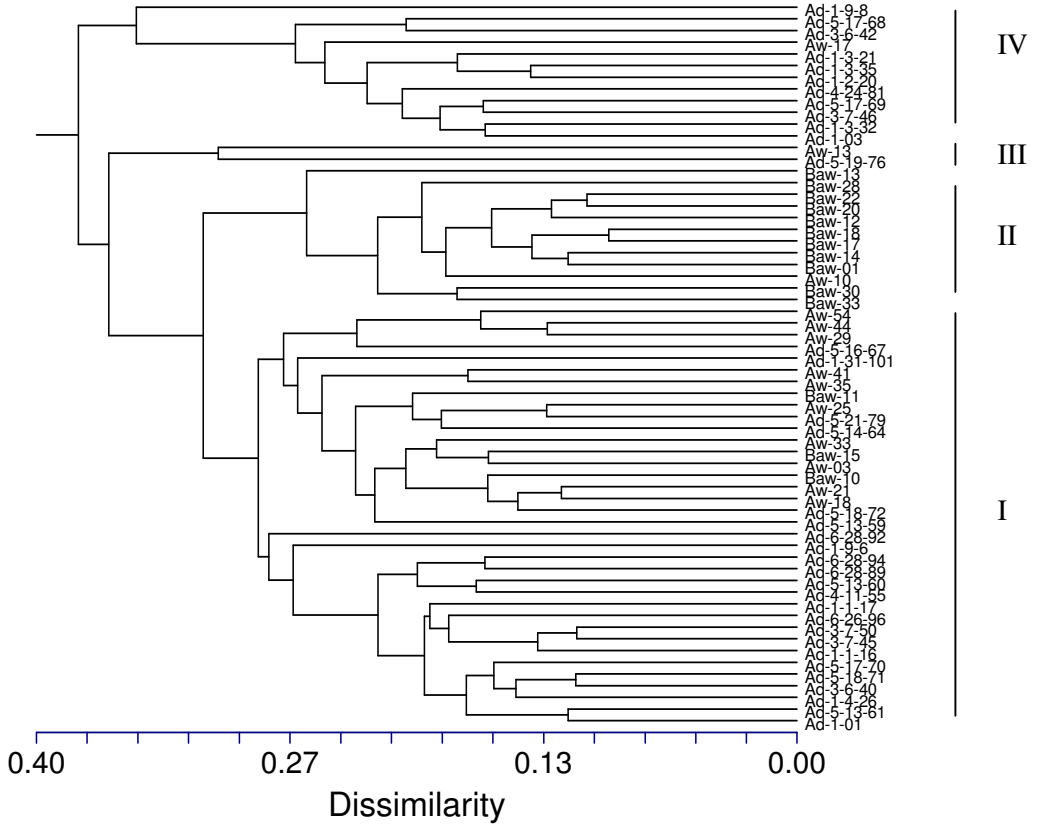

Figure 6.2 Dendrogram of traditional Ethiopian highland maize accessions derived by UPGMA from the dissimilarity matrix of the morphological data

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.

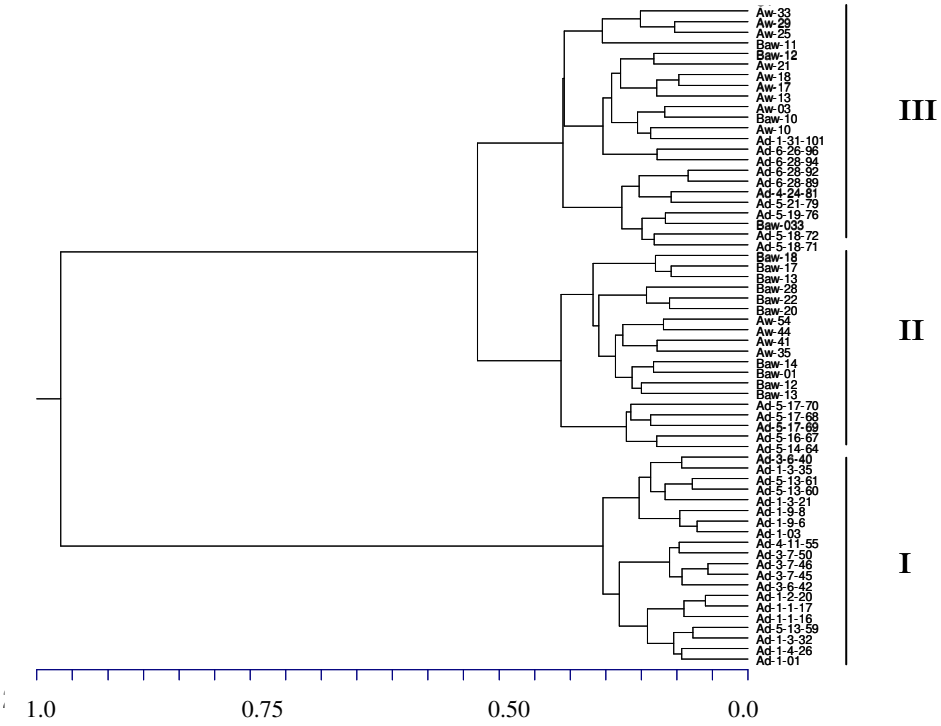

Figure 6.3 Dendrogram of the 62 traditional Ethiopian highland maize accession based on the Ward minimum variance method applied to the dissimilarity matrix generated by Nei and Li dissimilarity coefficients of the pooled AFLP and SSR data

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.