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

121

/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.
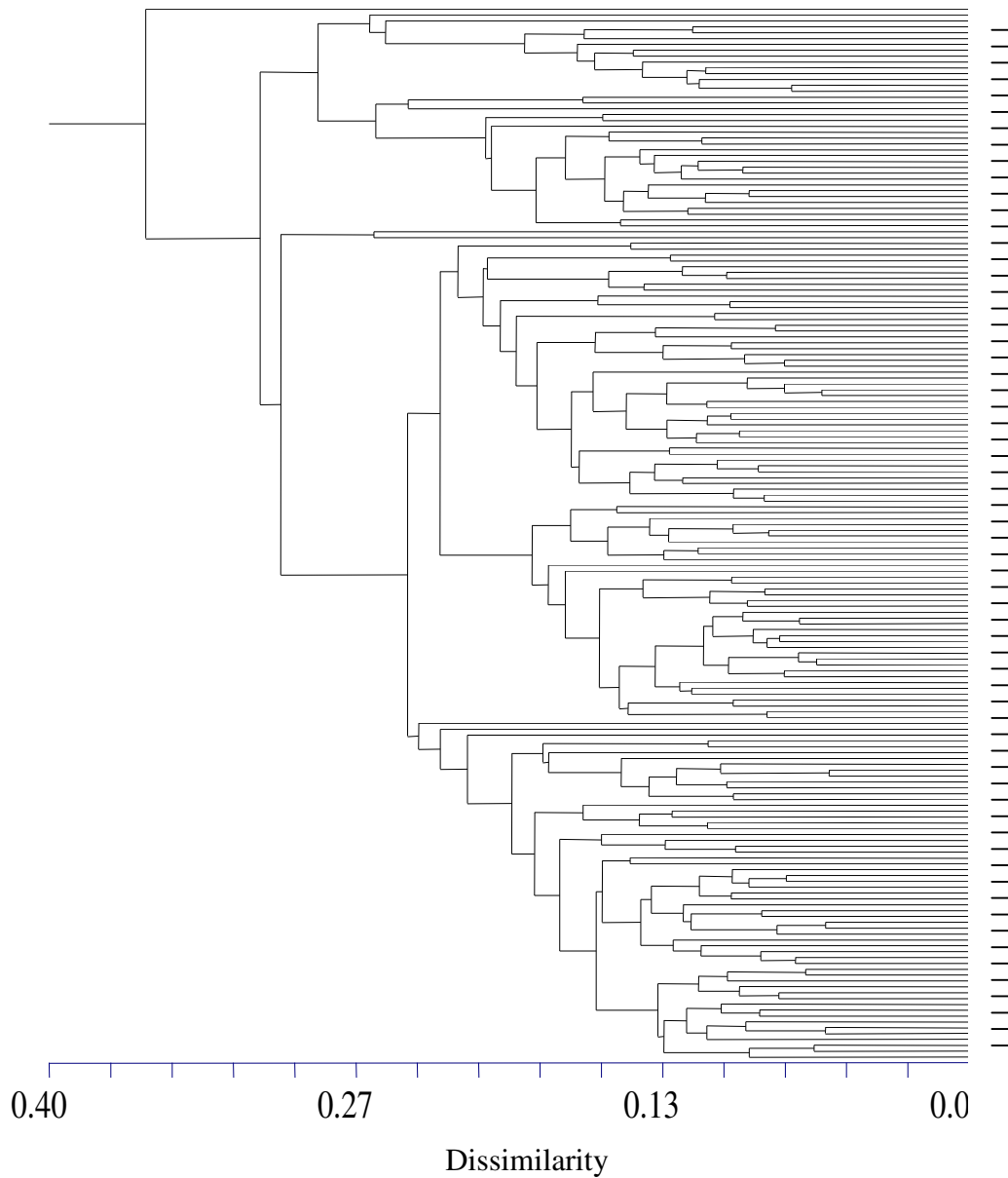

Figure 7.1 Dendrogram resulting from cluster analysis of standardized morphological and agronomic data from 180 maize accessions using Euclidean distance and UPGMA clustering. Sixty-two accessions were selected for molecular analysis (indicated by horizontal bars as root of the dendrogram)

123

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).

125

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

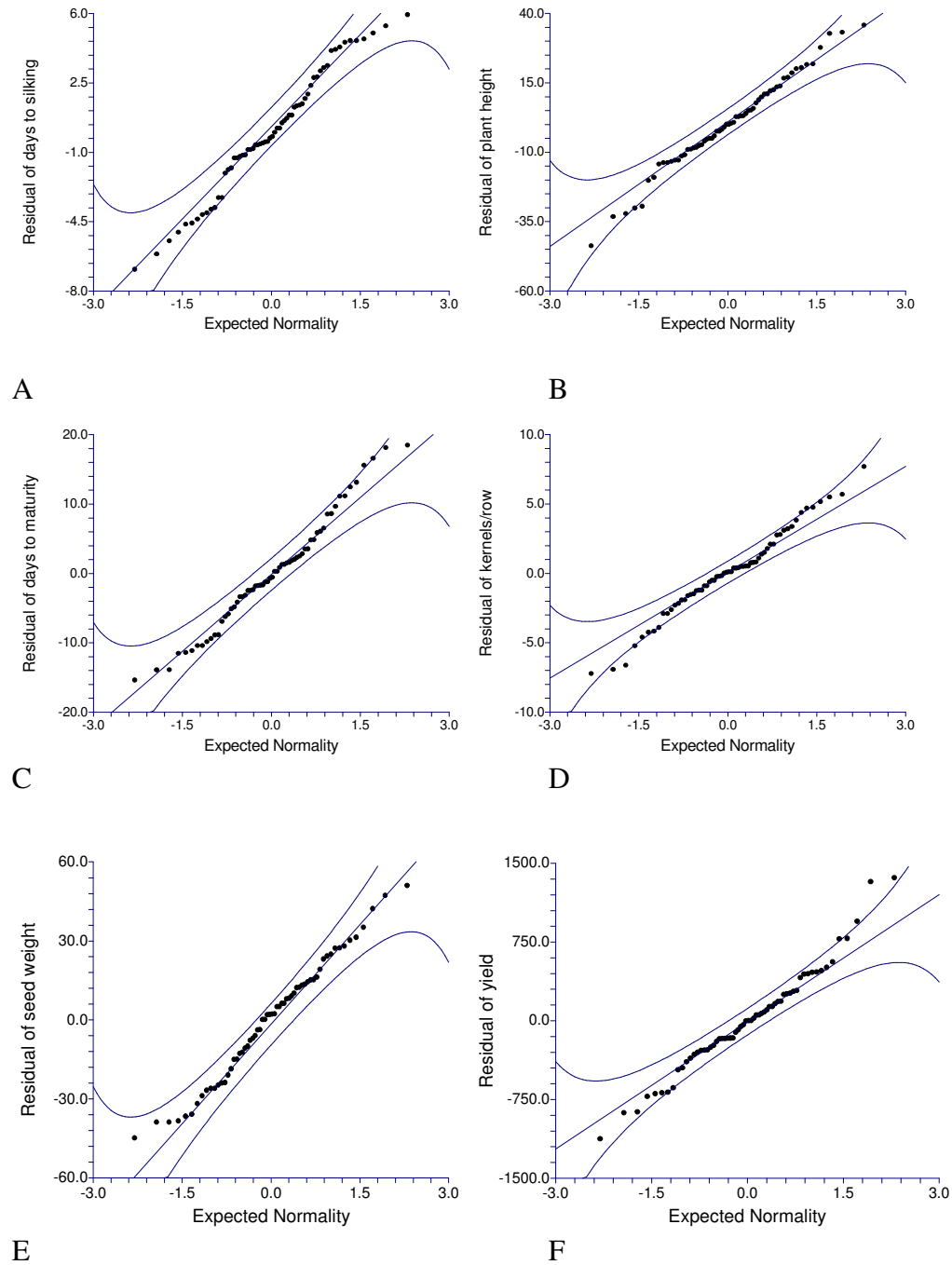| SSR	Allele	SSR	Allele
nc003	1 to 8	phi054	1 to 5
umc2190	1 to 3	phi453121	1 to 5
umc1632	1 to 3	umc2152	1 to 3
phi042	1 to 6	umc2040	1 to 6
bnlg182	1 to 7	phi037	1 to 3
phi079	1 to 3	umc1357	1 to 3
phi034	1 to 4	phi015	1 to 5
phi026	1 to 6	umc2129	1 to 5
bnlg2190	1 to 9	umc1537	1 to 3
phi021	1 to 5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

129

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

130

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).
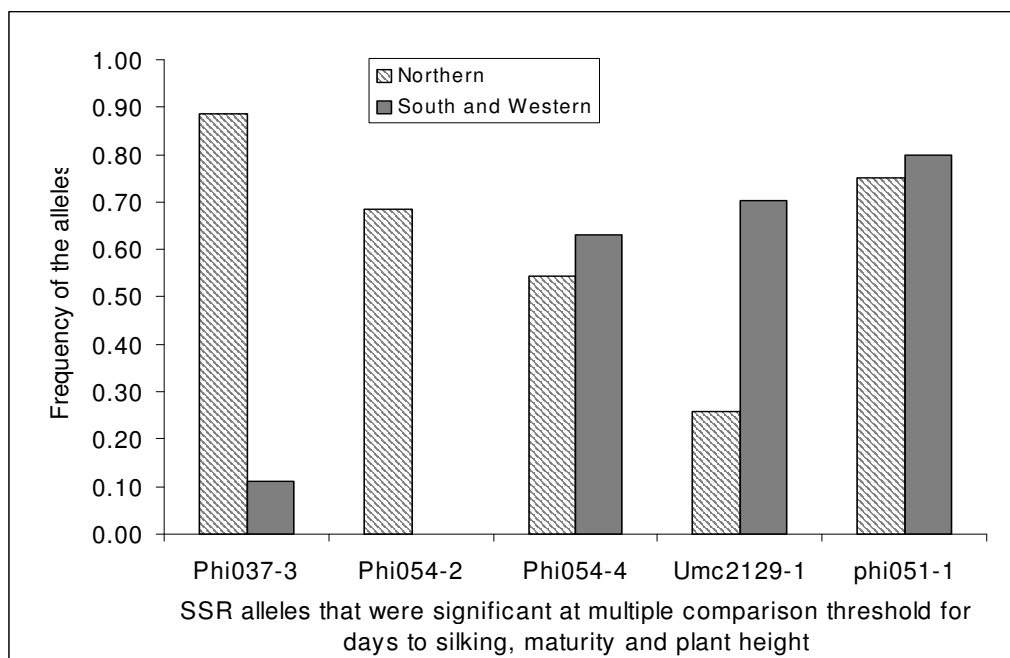

Figure 7.3 Mean allelic frequencies of SSR alleles selected for adaptation in contrasting agroecologies in Ethiopian highlands

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.