GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Genetic diversity, population structure, and association mapping of agronomic traits in waxy and normal maize inbred lines

K.J. Sa[1], J.Y. Park[1,2], S.H. Choi[1], B.W. Kim[1], K.J. Park[2] and J.K. Lee[1]

[1]Department of Applied Plant Sciences,
College of Agriculture and Life Sciences,
Kangwon National University, Chuncheon, South Korea
[2]Maize Experiment Station,
Gangwon Agricultural Research and Extension Services,
Hongcheon, South Korea

Corresponding author: J.K. Lee
E-mail: jukyonglee@kangwon.ac.kr

Genet. Mol. Res. 14 (3): 7502-7518 (2015)
Received October 31, 2014
Accepted April 7, 2015
Published July 3, 2015
DOI http://dx.doi.org/10.4238/2015.July.3.26

ABSTRACT. Understanding genetic diversity, population structure, and linkage disequilibrium is a prerequisite for the association mapping of complex traits in a target population. In this study, the genetic diversity and population structure of 40 waxy and 40 normal inbred maize lines were investigated using 10 morphological traits and 200 simple sequence repeat (SSR) markers. Based on a population structure analysis, the 80 maize inbred lines were divided into three groups: I, II, and admixed. Significant marker-trait associations were identified between the markers and the 10 morphological traits, which were studied according to the model used to confirm the association. Using a general linear model, the lowest R^2 value (9.03) was detected in umc1139, which was associated with ear number, and the highest (43.97) was in umc1858, which was associated with plant height. Using a mixed linear model, the lowest R^2 value (18.74) was in umc1279, which was

associated with ear weight; the highest (27.66) was in umc1858, which was associated with 100-kernel weight. The SSR markers identified in the present study may serve as useful molecular markers for selecting important yield and agronomic traits. These results will be useful for marker-assisted selection in maize breeding programs, to help breeders choose parental lines and markers for crosses.

Key words: SSR marker; Genetic diversity; Population structure; Waxy and normal maize inbred line; Marker-trait association

INTRODUCTION

Maize is one of the most important agricultural crops in the world. Based on the starch composition of the seed's endosperm, maize can be divided into two types: normal (non-waxy) and waxy. The main difference between normal and waxy maize is the texture or starch content of the grain. The texture of the endosperm of waxy maize is one of its unique features. It contains only branched-chain starches, and consists of over 99% amylopectin. In contrast, the starch of normal maize is composed of about 75% amylopectin and 25% amylase (Nelson and Rines, 1962). Normal maize is widely cultivated for use in food and feed. Waxy maize is a special type of maize that is cultivated for food production in China and Korea.

The use of molecular marker-based techniques in genetic studies has advanced remarkably in recent years. Among the various types of molecular marker, simple sequence repeats (SSRs) or microsatellites, which are short regions containing tandemly repeated copies of 1-6 nucleotide fragments (Rafalski et al., 1996), are considered to be one of the most suitable for assessing genetic diversity because of their reliability, reproducibility, and discrimination (Akagi et al., 1997; Enoki et al., 2002). SSR markers work well in inbred maize lines, which contain a high level of allelic variation, in order to gain information about genetic diversity, genetic relationships, and population structure. Such data are of fundamental importance for the improvement and development of new cultivars, in planning crosses for hybrids or inbred-line development, assigning lines to heterotic groups, and protecting the plant germplasm (Pejic et al., 1998).

In plant breeding programs, determining the genetic basis of agronomic traits is a very important scientific problem for crop improvement (Pasam et al., 2012). There are two methods to identify genomic regions related to important traits: 1) quantitative trait loci (QTL) mapping based on linkage within segregating populations, and as a result of crosses between bi-parents with contrasting phenotypes and genotypes (Skot et al., 2005); and 2) association mapping using linkage disequilibrium (LD) between markers and agronomic traits of interest (Flint-Garcia et al., 2005; Yu and Buckler, 2006). Recently, association mapping based on LD has been used to identify genes that control important traits, and has been used in human genetics (Khoury et al., 2009). These methods have successfully been applied to the analysis of many crops (Zhu et al., 2008), e.g., rice (Borba et al., 2010), maize (Mezmouk et al., 2011), barley (Lorenz et al., 2010), and pea (Kwon et al., 2012).

The consumption of waxy maize is increasing in Korea, as the population transitions from a traditional rice-based diet to a Western, meat-based diet. A large collection of inbred

lines and maize of varying origins, both from local farmers and from other countries, has been assembled at the Maize Experiment Station, which is operated by Gangwon Agricultural Research and Extension Services. Because most lines in the collection have not been, or are, rarely utilized in breeding programs, genetic characterization is needed to ensure the long-term success of maize breeding programs that incorporate this material. The analysis of the collection also provides an opportunity for testing novel genetic methodologies.

In this study, we conducted the association mapping of 200 SSR markers and 10 agronomic traits in 40 waxy and 40 normal maize inbred lines. Our focus was to define the population structure, as well as the genetic diversity and relationships, of fundamental agronomic traits in a relatively large collection of defined plant material. These data will be of great use for future maize breeding programs.

MATERIAL AND METHODS

Plant materials and phenotypic analysis

The inbred lines used in these experiments, with their codes and entry numbers, pedigrees, and sources, are listed in Table 1. The inbred lines were obtained from the Maize Experiment Station, and had been originally collected from Korea, China, and other countries. The lines evaluated here were elite inbred lines, which had been cultivated for a number of years at the station (e.g., 97S6040 had been cultivated for 6 years, 96S7003 for 7 years, 98S8004 for 8 years, and 05YS9011 for 9 years). Ten agronomic traits were evaluated in 2010: the distance from the soil level to the base of the tassel (plant height, PH), the distance from the soil level to the base of the main ear (ear height, EH), leaf width (LW), ear length (EL), ear breadth (EB), the number of rows per ear (ER), the number of ears (EN), the yield of fresh ears without husks (ear weight, EW), the weight of 100 fresh kernels (100 KW), and the distance between the upper and lower pericarp surfaces (pericarp thickness, PT) (Table 2). Differences between the waxy and normal maize inbred lines were tested for significance by the Student *t*-test. Statistical analysis was performed using Microsoft Office Excel 2010.

DNA extraction and SSR amplification

Genomic DNA was extracted from young leaves using the method of Dellaporta et al. (1983), with minor modifications. A total of 200 microsatellite markers, distributed at about 20 loci per chromosome across the 10 maize chromosomes, were used for analyzing genetic diversity, population structure, and association mapping. All of the SSR markers were derived from MaizeGDB (http://www.maizegdb.org/).

SSR amplification was conducted in a total volume of 30 μL, which contained 20 ng genomic DNA, 1X PCR buffer, 0.3 μM forward and reverse primers, 0.2 mM dNTPs, and 1 U *Taq* Polymerase (BioTools). The PCR program consisted of a 5-min initial denaturation at 94°C, followed by two 1-min denaturation steps at 94°C, a 1-min annealing step at 65°C, and a 2 min extension at 72°C. After the second cycle, the annealing temperature was decreased by 1°C increments following every second cycle, until a final annealing temperature of 55°C was reached. The last cycle was repeated 20 times. Upon completion of all of the cycles, the final extension at 72°C was lengthened to 10 min.

Table 1. Derivations of 80 waxy and normal maize inbred lines used in this study.

Entry No.	Pedigree	Source	Entry No.	Pedigree	Source
WMIL01	97S6040	Landrace, Pyeongchang-gun, Gangwon-do	NMIL01	00hf1	Eongdan14
WMIL02	98S7007	Chalok 1 / W7031	NMIL02	00hf11	P3525
WMIL03	98S8004	Landrace, Unknown	NMIL03	00hf17	N2BE / B73
WMIL04	98S8034	Landrace, Ulleung-gun, Gyongsangbuk-do	NMIL04	00hf22	Unknown
WMIL05	98S8044	W84-9067 / A632wx	NMIL05	00hf25	Hwaseong 1
WMIL06	98S8007	W9060 / A632wx	NMIL06	00hf28	Pioneer synthetic
WMIL07	99S8003	Chalok 1 / W7094	NMIL07	00hf29	P3352
WMIL08	99S8015	Cultivar, Kaset Khao	NMIL08	00hf33	P3790
WMIL09	00S8007	IT90185, RDA genebank	NMIL09	00hf36	Eongdan14
WMIL10	01S8025	Mo401wx / KW1	NMIL10	00hf41	P3352
WMIL11	01S8069	Daehakchal / KW14	NMIL11	hc1	Beijing Acod
WMIL12	02S8008	Landrace, Yeongju-si, Gyongsangbuk-do	NMIL12	hc5	Ho-5
WMIL13	02S8014	Landrace, Boeun-gun, Chungcheongbuk-do	NMIL13	hc2	NK487
WMIL14	02S8056	Chalok 1 / KW7	NMIL14	hc3	Nk692
WMIL15	02BS8005	Landrace, Namwon-si, Jeollabuk-do	NMIL15	hc4	NK698
WMIL16	03S8001	KW7 / KW8	NMIL16	hc6	8112
WMIL17	03S8013	KW2 / KW7	NMIL17	NC300	Unknown
WMIL18	03S8060	KW7 / Hoengseong landrace	NMIL18	HF1	Unknown
WMIL19	03S8064	KW7 / Inje landrace	NMIL19	HF2	Unknown
WMIL20	03BS8016	Unknown / Yungil landrace	NMIL20	CML177	Unknown
WMIL21	04S8008	KW7 / Hongcheon landrace	NMIL21	B84	Unknown
WMIL22	05S8004	Daehakchal / Chalok 2	NMIL22	KS85	Unknown
WMIL23	05S8019	Landrace, Unknown	NMIL23	KS118	Unknown
WMIL24	05BS8005	96A099 / 96A059	NMIL24	SIM6	Maysin collection
WMIL25	01S5071	Landrace, Jeollanam-do	NMIL25	EPM6	Unknown
WMIL26	01S6011	KW7 / KW8	NMIL26	Oh43	Unknown
WMIL27	01S6067	Daehakchal / KW7	NMIL27	Wf9	Unknown
WMIL28	01S8030	Landrace, Hoengseong-gun, Gangwon-do	NMIL28	00hf35	P3790
WMIL29	01S8034	Daehakchal / KW14	NMIL29	07S8004	IP144
WMIL30	01BS8045	Landrace, Hwaseong-si, Kyunggi-do	NMIL30	07S8009	IP152
WMIL31	01BS8068	Landrace, Tongyeong, Gyongsangnam-do	NMIL31	07S8011	IP161
WMIL32	HW3	W9060 / A632wx	NMIL32	07S8016	00Pop A (Early)
WMIL33	HW4	Landrace, Anseong-si, Kyunggi-do	NMIL33	06S8001	ISU pop T-C 8644-27 / ISU Pop 5
WMIL34	HW7	Landrace, Yanggu-gun, Gangwon-do	NMIL34	06S8008	9071 / 6B-6
WMIL35	HW8	Landrace, Hwacheon-gun, Gangwon-do	NMIL35	06S8013	ISU Inb.1368 / (B87/B73-12) B#
WMIL36	KW1	Landrace, Gosung-gun, Gangwon-do	NMIL36	06S8019	8321-18 / 12B-2
WMIL37	KW7	Landrace, Pyeongchang-gun, Gangwon-do	NMIL37	06S8030	EV43-SR / 9B-5
WMIL38	05YS9014	Landrace, Changchun	NMIL38	06S8042	IB89A-D14 1368 / ISUINB 7B-1
WMIL39	05YS9126	Landrace, Longjing	NMIL39	05S8011	96KPC midearly / early2
WMIL40	05YS9129	Landrace, Longjing	NMIL40	05S8027	S133

Electrophoresis and fragment detection

Five microliters of the final reaction product were mixed with 10 μL electrophoresis loading buffer (98% formamide, 0.02% BPH, 0.02% xylene C, and 5 mM NaOH). After denaturation and immediate cooling, 2 μL of the sample was loaded onto a 6% denaturing (7.5 M urea) acrylamide-bisacrylamide (19:1) gel in 1X TBE buffer, and electrophoresis was conducted at 1800 V and 60 W for 120 min. The separated fragments were then visualized by silver staining kit (Promega, USA).

Data analysis

The number of alleles, allele frequency, major allele frequency (MAF), gene diversity (GD), and polymorphic information content (PIC) were estimated using PowerMarker version 3.25 (Liu and Muse, 2005). Genetic similarities (GS) were calculated for each pair of lines us-

ing the Dice similarity index (Dice, 1945). The similarity matrix was then used to construct a dendrogram based on an unweighted-pair-group-method-using-arithmetic-averages algorithm (UPGMA), with the help of sequential agglomerative hierarchical non-overlapping clustering in NTSYSpc version 2.1 (Rohlf, 1998).

Population structure among the 80 lines was evaluated by the model-based program STRUCTURE 2.2 (Pritchard et al., 2000), in order to confirm the genetic structure. The STRUCTURE program was run five times for each *K* value, ranging from 1 to 10, using the admixture model with a burn-in of 100,000 and a run length of 100,000. An average likelihood value, LnP(D), was calculated across all runs for each *K*. The ad-hoc criterion (ΔK) of Evanno et al. (2005) was used to determine the most probable *K* value, in order to compensate for the overestimation of subgroup number by STRUCTURE. A run of estimated numbers of the subgroups showing maximum likelihood was used to assign inbred lines that had membership probabilities of ≥ 0.80 to subgroups. Inbred lines with membership probabilities of <0.80 were assigned to the admixed group (compare to Stich et al., 2005).

Association mapping was performed for the marker-trait association using TASSEL 3.0 (Bradbury et al., 2007). We used two models to confirm the marker-trait association: a general linear model (Q GLM) and a mixed linear model (Q+K MLM). The Q GLM method was performed using a Q-matrix derived from the STRUCTURE program. The number of permutations was set at 10,000, to obtain P values for marker significance of 0.05 and 0.01. The Q+K MLM method used a kinship K matrix, and a population structure Q matrix at $P < 0.05$ and $P < 0.01$. The K matrix was created in the SPAGeDi software (Hardy and Vekemans, 2002) by calculating kinship coefficients, using the method of Loiselle et al. (1995).

RESULTS

Phenotypic analysis and correlation analysis

The phenotypic characteristics of the inbred lines are summarized in Table 2. We found that most of the agronomic traits exhibited differences between the two types of maize, and the average values were greater in the normal than in the waxy inbred lines. A correlation analysis was performed to confirm the genetic relationships between the agronomic traits and the inbred lines. Most of the traits were positively correlated with each other, except for EN and LW, which were negatively correlated. Among all of the possible trait combinations, those between PH and EH (0.853), EB and EL (0.708), and EB and EN (0.745) had higher correlation coefficients than did the others. EW, in particular, was highly correlated with the other nine traits, with P values ranging from 0.01 to 0.05.

Genetic variation and diversity among inbred waxy and normal maize lines

A total of 200 SSR loci were used to evaluate GD and the genetic relationships among the 80 maize inbred lines (Table 3). All of the SSR loci were confirmed in 1610 alleles. The number of alleles per locus ranged from 2 to 31, and the average number of alleles per locus was 8.05 (Figure 1A). The average GD was 0.72, and ranged between 0.16 and 0.93 (Figure 1B). The average PIC value was 0.68, and ranged between 0.15 and 0.92. The average MAF was 0.40, and ranged between 0.13 and 0.91 (Figure 1C, Table 3). Of the 1610 alleles, 324

Table 2. Correlation coefficients, means, and standard deviations for 10 agronomic traits in waxy and normal maize inbred lines.

	Plant height (PH)	Ear height (EH)	Leaf width (LW)	Ear length (EL)	Ear breadth (EB)	Ear row (ER)	Ear number (EN)	Ear weight (EW)	100 Kernel weight (100 KW)	Pericarp thickness (PT)
PH	-	0.853**	0.444**	0.370**	0.244*	0.311**	0.119	0.463**	0.440**	0.300**
EH		-	0.284*	0.274*	0.2	0.346**	0.116	0.399**	0.274*	0.144
LW			-	0.143	0.097	0.118	-0.092	0.419**	0.547**	0.165
EL				-	0.708**	0.154	0.662**	0.577**	0.321**	0.17
EB					-	0.473**	0.745**	0.620**	0.265*	0.109
ER						-	0.235*	0.365**	0.025	0.077
EN							-	0.402**	0.079	0.041
EW								-	0.466**	0.228*
100 KW									-	0.334**
Mean ± SD	166.7 ± 32.1	81.9 ± 24.7	7.5 ± 1.5	11.9 ± 2.9	3.3 ± 0.7	12.2 ± 1.8	9.1 ± 1.7	475.1 ± 227.0	22.5 ± 5.2	45.6 ± 20.1
Min.	95.3	38.7	4.8	2.1	0.8	7.6	2	19	8.3	15.6
Max.	235.3	149.0	11.7	18.6	4.5	16.4	11	1404	38.3	102.2
Mean ± SD (W40)	146.8 ± 24.6	68.7 ± 19.9	6.7 ± 0.9	11.1 ± 3.0	3.2 ± 0.8	11.7 ± 2.1	9.0 ± 1.9	361.3 ± 159.7	19.6 ± 3.8	36.3 ± 13.2
Mean ± SD (F40)	186.7 ± 25.8	95.0 ± 22.1	8.3 ± 1.6	12.6 ± 2.6	3.4 ± 0.6	12.7 ± 1.4	9.2 ± 1.5	589.0 ± 228.6	25.3 ± 4.9	55.0 ± 21.5

**Significant at $P < 0.01$. *Significant at $P < 0.05$.

private alleles (20.1%) were detected in only one of the 80 inbred lines. The frequency of rare alleles (frequency of less than 0.05) was 44.3% (714 alleles), whereas intermediate (frequency of between 0.05 and 0.5) and abundant alleles (frequency greater than 0.5) accounted for 53.0% (854 alleles) and 2.60% (42 alleles), respectively, of the total (Figure 2). To clearly understand the genetic variation in the waxy and normal lines, we also analyzed allele number, GD, and PIC. Table 3 summarizes these values for the 200 SSR loci in the two types of maize. The average numbers of alleles were 6.34 and 6.54 in waxy and normal lines, respectively. The average GD, PIC, and MAF values were 0.66, 0.62, and 0.46, respectively, for the waxy lines; for the normal lines these values were 0.69, 0.65, and 0.43, respectively (Table 3). We also estimated the number of specific alleles. Most of the 1610 alleles were distributed evenly between the waxy and normal lines, but there were 303 alleles that were only in the waxy lines and 342 alleles that were only in the normal lines.

Table 3. Total number of alleles and genetic diversity index for 200 simple sequence repeat loci in two types of maize.

Parameter	Total inbred lines (N = 80)	Waxy inbred lines (N = 40)	Normal inbred lines (N = 40)
No. of alleles	1610	1268	1307
Mean	8.05	6.34	6.54
Range	2-31	2-14	2-24
Gene diversity	0.72	0.66	0.69
Min.	0.16	0.1	0.05
Max.	0.93	0.89	0.93
PIC	0.68	0.62	0.65
Min.	0.15	0.09	0.05
Max.	0.92	0.88	0.93
Major allele frequency	0.4	0.46	0.43
Min.	0.13	0.2	0.13
Max.	0.91	0.95	0.98

PIC = polymorphic information content.

Population structure and cluster analysis

The LnP(D), calculated using the STRUCTURE program, was not clear for K values ranging from 1 to 10, which were calculated from five replicate sets. Therefore, to estimate the number of subgroups we applied the ad-hoc measure ΔK, as suggested by Evanno et al. (2005). For all of the lines, the highest ΔK value was at $K = 2$ (Figure 3). Based on a membership probability threshold of 0.8 (Wang et al., 2008), the lines were divided into three groups: I, II, and admixed. Fourteen waxy lines were assigned to group I, and group II contained 5 waxy and 40 normal lines. The admixed group had 21 waxy lines, with a membership threshold of <0.8 (Figure 4). A dendrogram constructed from the UPGMA analysis is presented in Figure 4. The 80 lines were clearly classified into two groups, based on their grain texture, and they had a GS of 0.25. Group I included 40 waxy maize inbred lines, and group II included 40 normal maize inbred lines (Figure 4).

Association mapping using the Q GLM and Q+K MLM models

At a significance level of 0.05, we found that 126 SSR markers were associated with the phenotypic traits using the Q GLM model, and 46 SSR markers were associated with them

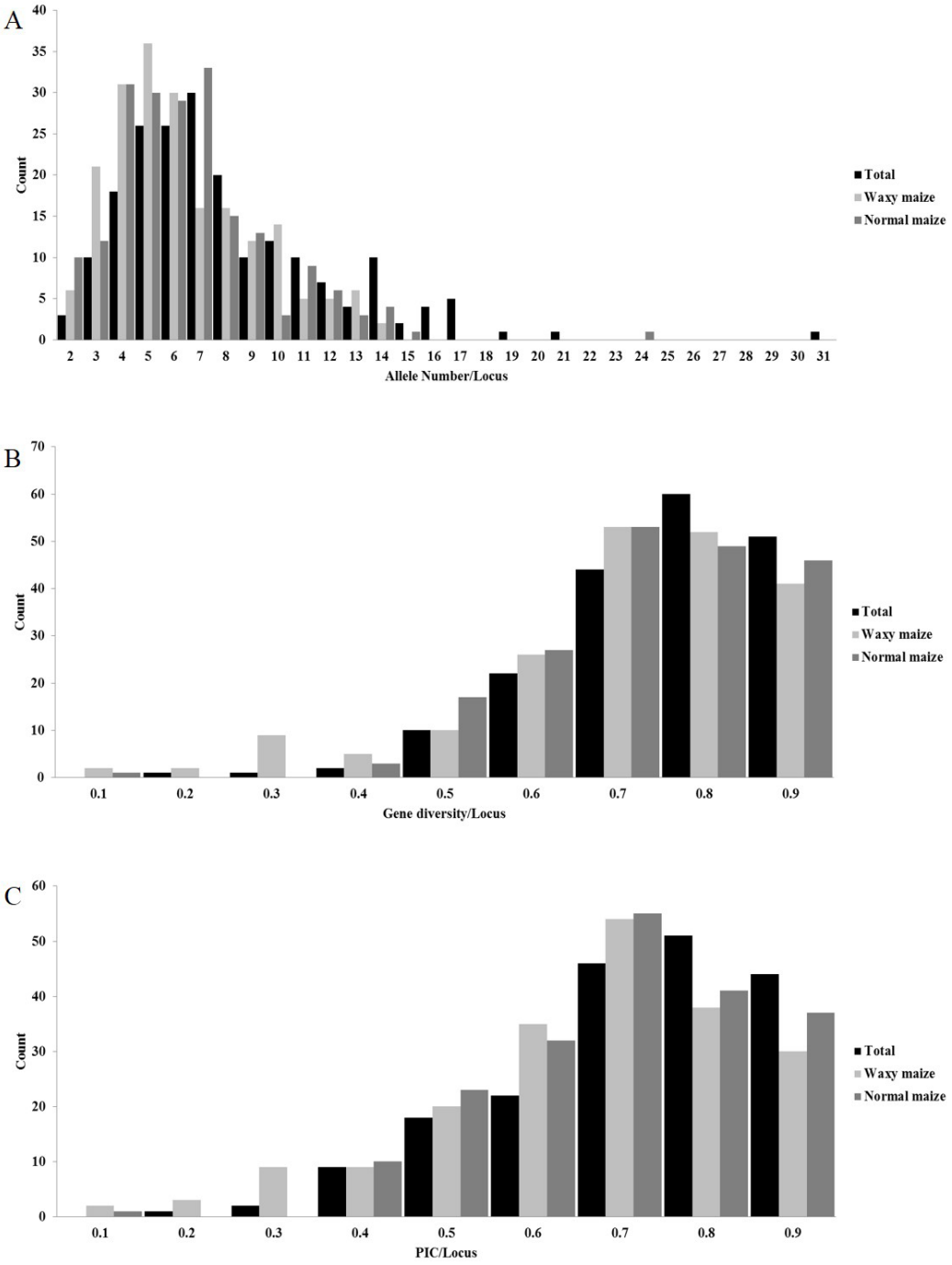

Figure 1. Frequency of allele number **(A)** gene diversity **(B)** and polymorphic information content (PIC) per locus **C.** in waxy and normal maize inbred lines.

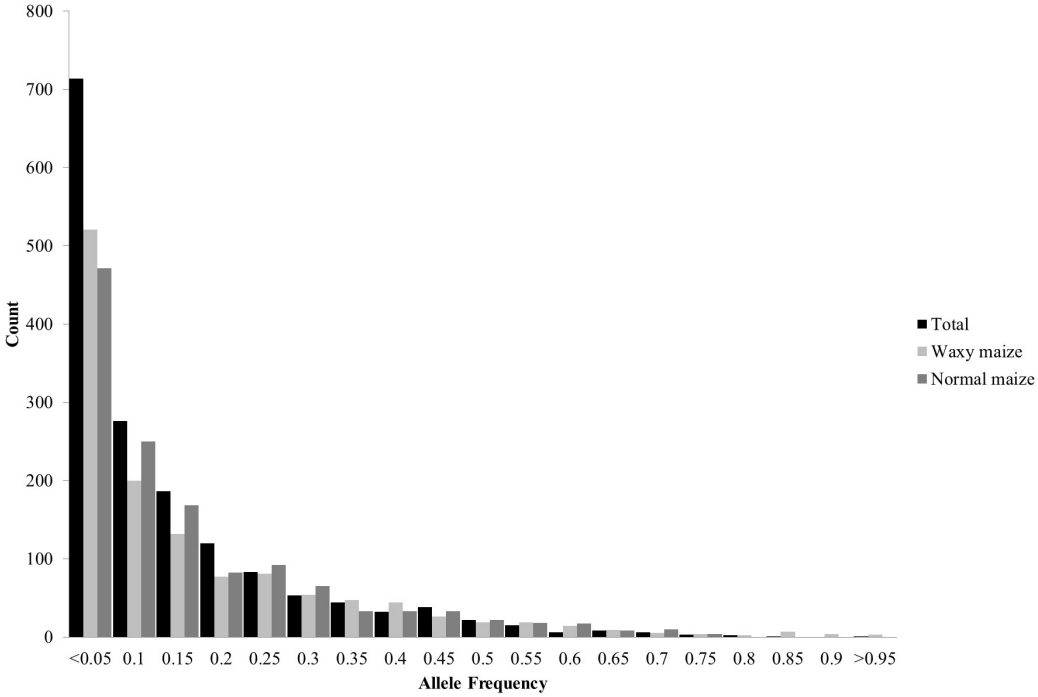

Figure 2. Histogram of allele frequencies in waxy and normal maize inbred lines.

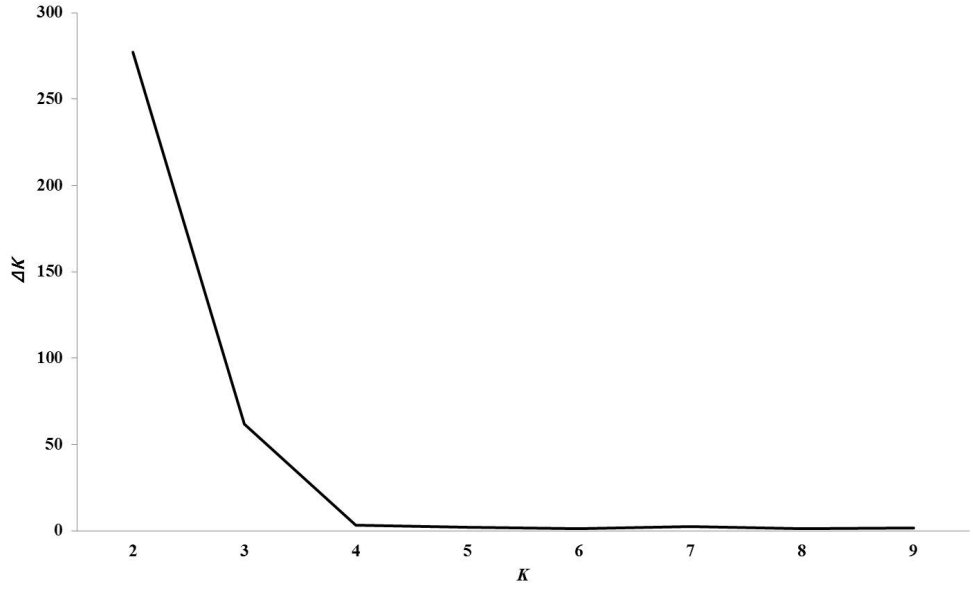

Figure 3. Rate of change in the log probability of data between true *K* values (*ΔK*).

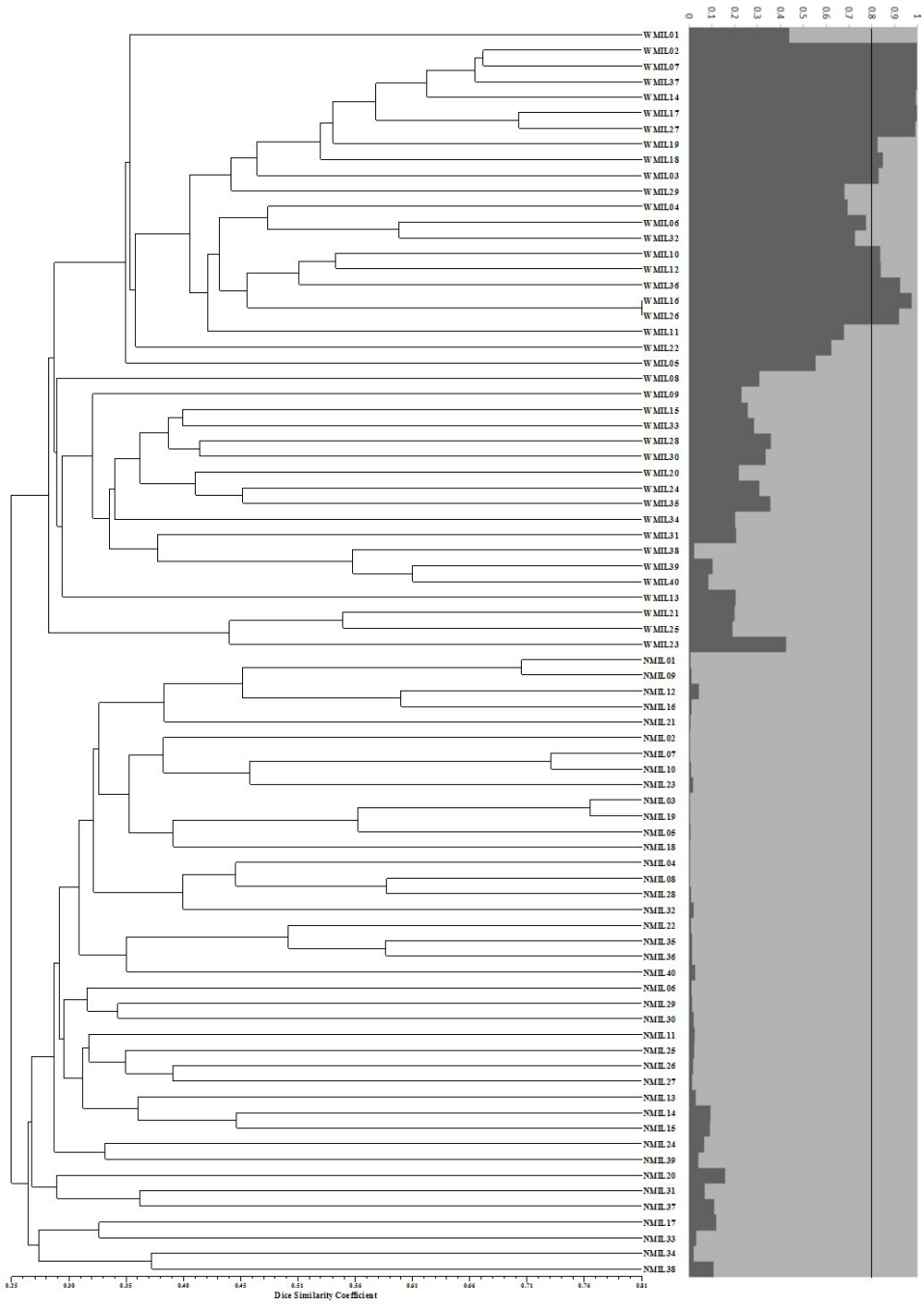

Figure 4. Unweighted pair group methods using arithmetic average algorithm dendrogram and population structure based on 200 simple sequence repeat (SSR) markers. WMIL = waxy maize inbred lines; NMIL = normal maize inbred lines.

using the Q+K MLM model (data not shown). However, when the significance level was increased to 0.01, 72 markers were associated with the traits using the Q GLM model. Two SSR markers, phi099 and umc1858, were associated with six traits, while 33 markers (umc1514, bnlg1036, bnlg125, dupssr30b, umc1079, umc1542, bnlg1605, umc1030, umc1949, umc2376, umc1031, umc1294, umc1667, umc1682, umc2365, umc1192, umc1447, umc2022, umc2161, umc1014, umc1250, umc1883, umc1134, umc1154, umc1159, bnlg1812, umc1139, umc1974, umc1040, umc1657, bnlg1028, bnlg1079, and umc2632) were only associated with one trait (PH, EH, LW, ER, EN, EW, 100 KW, or PT) (Table 4). Among the 141 marker-trait associations in the Q GLM model, the lowest R^2 value (9.03) was detected in umc1139, which was associated with EN, while the highest (43.97) was found in umc1858, which was associated with PH (Table 4). Four SSR markers were associated with four phenotypic traits in the Q+K MLM model, and each marker was related to only one trait. In the Q+K MLM model, the lowest R^2 value (18.74) was in umc1279, which was associated with EW, and the highest (27.66) was in umc1858, which was associated with 100 KW (Table 5). We also compared the associated traits that overlapped the Q GLM and Q+K MLM models. There were four marker-trait associations detected in both models at a level of significance of $P < 0.01$: umc1858, umc1645, umc2215, and umc1279, which were associated with 100 KW, EH, EN, and EW, respectively (Table 5).

DISCUSSION

In order to develop a new cultivar or elite inbred line, knowledge of the genetic diversity, genetic relationships, and population structure of the breeding materials is important for breeding programs. In addition, understanding the structure of the analyzed population is critical for association mapping (Flint-Garcia et al., 2005). In this study, 200 SSR loci (about 20 loci per chromosome), covering the whole maize genome, were used to detect genetic diversity in 40 waxy and 40 normal inbred lines. A total of 1610 alleles were detected with an average number of 8.05 alleles per locus, and the average GD and PIC values were 0.72 and 0.68, respectively (Table 3). We compared phenotypic variation and diversity between waxy and normal maize.

According to the phenotypic variation analysis, 10 agronomic traits exhibited differences between the two types of line. Normal maize lines had much higher biomass values for most of the morphological traits than the waxy lines (Table 2). The average number of alleles and GD were 6.34 and 0.66, respectively, in the waxy lines, and were 6.54 and 0.69, respectively, in the normal lines (Table 3). These results indicate that genetic variation in the normal lines was much higher than in the waxy lines. Choice of germplasm is critical to the success of association mapping. As previously suggested by Malosetti et al. (2007), good examples of plant populations for use in genetic association mapping are breeding and gene bank collections of cultivars, breeding lines, germplasm, etc. The materials used in our study were breeding lines, and the variation and GD of these maize materials indicated that they were well suited for this type of genetic analysis.

Inbred lines of waxy and normal maize often have a complex genetic background. Therefore, understanding the population structure of inbred lines is very important for maize improvement by genetic crosses (Wang et al., 2008). In our study, the initial population structure pattern was not clearly distinguishable between the waxy and normal maize groups. In the population structure analysis, the maximum ΔK value was at $K = 2$ and the 80 inbred

Table 4. List of significant markers detected using a general linear model.

Locus	Chr	Trait	P value*	R^2	Locus	Chr	Trait	P value	R^2
umc1514	1	100 KW	0.007	21.1	umc2022	5	PT	3.65E-04	24.7
umc1553	1	LW	2.74E-04	34.3	umc2161	5	PH	0.0072	29.9
		100 KW	0.004	27.6	bnlg1617	6	LW	0.0053	38.7
umc1991	1	LW	0.009	14.2			100KW	0.003	40.3
		EW	0.0043	16.0	phi364545	6	LW	0.0073	17.0
		100 KW	0.0057	15.3			PT	0.0072	17.1
		PT	0.0057	15.3	umc1014	6	100KW	0.0051	21.9
umc2215	1	EL	0.0076	14.6	umc1250	6	PT	0.009	18.4
		EB	0.0024	17.4	umc1595	6	LW	0.0061	9.4
		EN	5.25E-05	25.7			EW	0.0062	9.3
		100 KW	0.0013	18.8			100KW	0.0014	12.5
bnlg1036	2	EN	0.0029	35.7	umc1883	6	EH	0.0074	20.9
bnlg125	2	100 KW	0.0013	40.9	umc2059	6	PH	0.0019	22.5
dupssr30b	2	PH	0.0096	36.7			EH	5.49E-04	25.5
umc1079	2	100 KW	0.0038	34.5			LW	0.0067	19.4
umc1185	2	PH	3.32E-04	28.4	bnlg1808	7	PH	1.85E-04	33.7
		ER	0.0082	20.6			EH	0.0033	26.8
umc1542	2	ER	0.0014	19.1			LW	6.02E-04	31.0
umc1823	2	PH	0.0054	32.8			EW	0.0062	25.1
		LW	5.90E-04	38.5			100KW	0.0021	27.9
umc1934	2	EH	0.0098	20.2	umc1134	7	PT	0.009	16.3
		LW	0.0083	20.6	umc1154	7	PH	0.006	25.8
umc2402	2	PH	7.19E-04	21.1	umc1159	7	EW	0.0065	24.5
		EH	4.27E-04	22.3	umc1295	7	PH	3.22E-04	32.8
bnlg1019a	3	PH	1.47E-04	39.1			LW	0.0044	26.4
		100 KW	0.0055	30.0	umc1944	7	PH	0.0094	21.3
bnlg1601	3	PH	0.0018	31.2			PT	4.19E-04	29.9
		EH	0.0085	27.0	umc2328	7	LW	0.0026	36.8
bnlg1605	3	PT	0.007	21.6			PT	0.0066	34.2
phi099	3	PH	3.09E-05	25.7	umc2332	7	LW	9.35E-04	24.0
		EH	0.0013	17.4			ER	0.0045	20.2
		LW	0.0077	13.0	bnlg1812	8	PH	0.0089	27.2
		EW	0.0072	13.2	umc1139	8	EN	0.0079	9
		100 KW	0.0032	15.1	umc1858	8	PH	2.17E-07	44.0
		PT	1.63E-04	22.1			EH	9.31E-07	41.4
phi374118	3	PH	0.0041	26.9			LW	7.63E-04	27.5
		LW	0.0097	24.5			ER	0.0021	24.9
		100 KW	0.003	27.7			EW	2.67E-04	29.9
		PT	0.0048	26.5			100KW	5.12E-06	38.2
umc1030	3	LW	0.0028	25.7	umc1974	8	EW	0.0018	27.4
umc1136	3	PH	0.0073	22.0	bnlg279	9	PH	0.0032	35.0
		ER	0.0063	22.4			EL	0.0027	35.5
umc1949	3	EH	3.54E-04	28.6			EN	0.0053	33.6
umc1973	3	LW	0.0013	21.8			100KW	0.0043	34.2
		100 KW	0.0089	16.9	phi065	9	100KW	2.65E-04	22.1
umc2000	3	PH	0.0031	25.2			PT	1.15E-05	28.5
		LW	0.0027	25.6	umc1040	9	PH	0.0046	32.8
umc2101	3	EB	0.0033	23.2	umc1279	9	LW	3.28E-05	29.9
		EN	0.001	26.1			EW	9.50E-05	27.7
umc2376	3	EW	0.0063	29.3	umc1634	9	PH	0.0031	19.4
umc1031	4	100 KW	0.0017	38.1			EH	0.0027	19.7
umc1058	4	PH	2.80E-05	29.6	umc1657	9	LW	0.0059	32.5
		EH	5.04E-04	23.4	umc2213	9	PH	0.0069	12.1
		LW	0.0045	18.2			100KW	1.81E-04	20.1

Continued on next page

Table 4. Continued.

Locus	Chr	Trait	P value*	R^2	Locus	Chr	Trait	P value	R^2
		EW	0.0033	19	bnlg1028	10	LW	0.0052	20.8
		100 KW	0.0067	17.2	bnlg1079	10	PT	0.0031	29.8
umc1101	4	PH	9.58E-04	25.4	bnlg2190	10	PH	5.61E-04	36.9
		EH	0.0048	21.3			EH	0.0087	29.7
		PT	1.36E-04	29.9	umc1176	10	PH	4.91E-04	23.5
umc1294	4	100 KW	0.0073	20.2			EH	2.04E-04	25.4
umc1667	4	PH	0.0096	26			100KW	0.0097	16.3
umc1682	4	PH	0.0042	30			PT	0.002	20.2
umc1871	4	LW	0.0082	18.6	umc1556	10	EW	0.0034	22.9
		100 KW	0.0097	18.2			100KW	0.0014	25.0
umc2365	4	PT	0.0055	24.3	umc1645	10	PH	1.29E-06	35.9
bnlg1237	5	PH	0.0046	24.4			EH	1.68E-06	35.4
		LW	0.001	28.1			100KW	0.0078	17.1
		EW	2.87E-05	36.1	umc1785	10	LW	0.0096	11.5
		PT	8.17E-04	28.7			PT	0.0083	11.8
umc1192	5	EN	0.009	25.9	umc2632	10	EW	0.0053	21.0
umc1447	5	PH	0.0031	22.1					

*Significant at $P < 0.01$.

Table 5. List of common significant markers detected using Q GLM and Q+K MLM models.

No.	Locus	Chr	Trait	Q GLM		Q+K MLM	
				P value*	R^2	P value*	R^2
1	umc2215	1	EN	5.25E-05	25.7	0.0049	20.8
2	umc1858	8	100 KW	5.12E-06	38.2	0.0096	27.7
3	umc1279	9	EW	9.50E-05	27.7	0.0019	18.7
4	umc1645	10	EH	1.68E-06	35.4	0.0060	20.3

*Significant at $P < 0.01$.

lines could be divided into two distinct groups (I and II), and a third admixed group (Figure 4). However, the maize lines were clearly divided into two major groups by the UPGMA clustering analysis (Figure 4). In general, the population structure and clustering pattern of any domesticated crop species is influenced by the natural history of pre-domesticated ancestral populations, as well as by the agricultural breeding system and complexity of the breeding practices conducted during the plant's history (Xie et al., 2008). However, in the case of maize, which is subject to out-crossing, the population structure, clustering pattern, and genetic background may be more complex than in a self-crossing species such as rice, wheat, or soybean.

Furthermore, in the two types of maize studied here, waxy maize arose from normal maize by a single mutation at some point during its history. Therefore, it is important to examine several inbred lines in a study of the population structure of waxy and normal genotypes in inbred maize, as we have conducted in the present study.

The identification of genes that control important agronomic traits is essential for maize breeding programs. During the past few decades, QTL mapping has been used in many studies to detect genes that control phenotypic traits (Yan et al., 2006; Ma et al., 2007; Li et al., 2009). Recently, association mapping has been used as an alternative to QTL mapping, because it is effective in detecting marker-trait associations in LD data (Zhu et al., 2008; Borba et al., 2010; Lorenz et al., 2010; Mezmouk et al., 2011; Kwon et al., 2012).

When designing genetic experiments, the first consideration should be the probability

of detecting a given genetic variation. Therefore, population size and composition are two of the most important factors for detecting genetic associations (Spencer et al., 2009). In our study, we used 80 inbred lines, with equal numbers of waxy and normal types. According to the results of the phenotypic analysis, there were significant differences in 10 agronomic traits between the two types (Table 2). Therefore, our association panel should be suitable for association mapping.

A second design consideration should be the marker density used in the experiment (Mackay et al., 2009), which is of even greater importance in whole-genome association studies. Linkage is the main factor that influences LD based on SSR loci (Zhang et al., 2012). In our study, 200 SSR loci (about 20 SSR per chromosome), which were distributed across the 10 maize chromosomes, were used. Although whole-genome association studies have the potential to identify genetic polymorphisms that underlie important agronomic traits, false positives (Type-I error) are a major problem, and can lead to spurious associations in population structure and unequal relatedness (K) measures (Zhang et al., 2010). To avoid false positives, we used a general linear model based on a Q-matrix (Q GLM), and a mixed linear model based on a Q and K matrix (Q+K MLM) (Tables 4 and 5). Population structure analysis using the Q GLM model identified 141 marker-trait associations; in contrast, only four associations were found using the Q+K MLM model, based on population structure and kinship. In general, the Q+K MLM method identifies relatively fewer significant associations (Yu et al., 2005; Kwon et al., 2012).

Some of the SSR markers used in our study have been used previously in other QTL or association mapping studies (Table 6). For example, the seven SSR markers (umc2215, umc1823, umc1294, umc1058, bnlg1812, umc1657, and bnlg2190) that were used in our study have also been used in association analysis studies (Zhang et al., 2012). Zhang et al. (2012) found that umc2215 (located on chromosome 1) is linked to four traits (number of days of tasseling/anthesis, EW, and grain weight per ear); we found that the SSR was linked to EL, EB, EN, and 100 KW. In addition, Zhang et al. (2012) found that on chromosome 2 umc1823 (related to PH and LW in our study) is linked to kernel ratio, and umc1294 (related to 100 KW) on chromosome 4 is linked to the number of days of silking. In our study, umc1058 on chromosome 4 was related to PH, EH, LW, EW, and 100 KW, which was linked to the number of kernels per row by Zhang et al. (2012). On chromosome 8, bnlg1812 was related to PH in our study, and was linked to grain embryo length ratio by Zhang et al. (2012). In our study, umc1657 on chromosome 9 was related to LW, and was linked to the number of days of tasseling/anthesis by Zhang et al. (2012). We found bnlg2190 on chromosome 10 to be related to PH and EH; by Zhang et al. (2012) it was found to be linked to the number of days of silking/tasseling/anthesis and cob diameter. We found umc2215 on chromosome 1 to be related to EL, EB, and EN, and umc1645 on chromosome 10 to be related to PH and EH; Li et al. (2010) found that they are associated with six traits (grain yield per plant, 100 KW, EL, ER, ear diameter, and number of kernels per row). In our study, umc1136 on chromosome 3 was related to PH and ER, but was associated with 100 KW in the QTL analysis conducted by Liu et al. (2010). We found that umc1883 (chromosome 6) and bnlg1812 (chromosome 8) were related to EH and PH, respectively, whereas Liu et al. (2011) reported that they are associated with grain filling rate, using QTL mapping. In addition, we have identified two SSR markers, umc2215 (related to EL, EB, and EN) and bnlg1812 (related to PH), which are also considered to be important markers for QTL and association mapping (Liu et al., 2010; Zhang et al., 2012). Several markers were detected in our study that have not yet been reported, and may be useful in future studies.

Table 6. Comparisons between our results and those of other studies.

Marker	Chr	Traits detected in this study	Marker-trait association or QTL mapping in other studies	Reference
umc2215	1	EL, EB, EN, 100 KW	The number of days of tasseling/anthesis ear weight, grain weight per ear grain yield, 100 kernel weight, ear length, row number per ear, ear diameter	Zhang et al. (2012) Li et al. (2010)
umc1823	2	PH, LW	Kernel ratio	Zhang et al. (2012)
umc1136	3	PH, ER	100KW	Liu et al. (2010)
umc1294	4	100KW	Number of days of silking	Zhang et al. (2012)
umc1058	4	PH, EH, LW, EW, 100 KW	The number of kernels per row	Zhang et al. (2012)
umc1883	6	EH	Grain filling rate	Liu et al. (2011)
bnlg1812	8	PH	Grain embryo length ratio Grain filling rate	Zhang et al. (2012) Liu et al. (2011)
umc1657	9	LW	The number of days of tasseling/anthesis	Zhang et al. (2012)
bnlg2190	10	PH, EH	The number of days of silking/tasseling/anthesis. cob diameter	Zhang et al. (2012)
umc1645	10	PH, EH	Row number per ear, kernel number per row, ear diameter	Li et al. (2010)

Marker-trait studies in maize inbred lines could provide a useful alternative to association mapping for marker-assisted selection (MAS) in breeding programs. Therefore, the detection and confirmation of loci associated with yield and agronomic traits may provide greater opportunities for maize breeders to control quality by MAS. The present study has demonstrated the utility of SSR analysis for the study of GD and population structure in waxy and normal maize inbred lines, and these data should help in optimizing the choice of parents for crossing combinations, as well as in selecting markers for MAS for the improvement of maize.

ACKNOWLEDGMENTS

Research supported by a Korea Research Foundation Grant funded by the Next-Generation BioGreen ‘‘21’’ Program (Plant Molecular Breeding Center, #PJ0080182014) of the Rural Development Administration, Republic of Korea, and Golden Seed Project (#213001-04-1-SBA10), Ministry of Agriculture, Food and Rural Affairs (MAFRA), Ministry of Oceans and Fisheries (MOF), Rural Development of Korea (RDA), and Korea Forest Service (KFS).

REFERENCES

Akagi H, Yokozaki Y, Inagaki A and Fujimura T (1997). Highly polymorphic microsatellites of rice consist of AT repeats, and a classification of closely related cultivars with these microsatellite loci. *Theor. Appl. Genet.* 94: 61-67.

Borba TCO, Brondani RPV, Breseghello F, Coelho ASG, et al. (2010). Association mapping for yield and grain quality traits in rice (*Oryza sativa* L.). *Genet. Mol. Biol.* 33: 515-524.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, et al. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23: 2633-2635.

Dellaporta SL, Wood J and Hicks JB (1983). A simple and rapid method for plant DNA preparation, version II. *Plant Mol. Biol. Rep.* 1: 19-21.

Dice LR (1945). Measures of the amount of ecologic association between species. *Ecology* 26: 297-302.

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

Evanno G, Regnaut S and Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14: 2611-2620.

Flint-Garcia SA, Thuillet AC, Yu JM, Pressoir G, et al. (2005). Maize association population: a high resolution platform for quantitative trait locus dissection. *Plant J.* 44: 1054-1064.

Hardy OJ and Vekemans X (2002). SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol. Ecol. Notes* 2: 618-620.

Khoury MJ, Bertram L, Boffetta P, Butterworth AS, et al. (2009). Genome-wide association studies, field synopses, and the development of the knowledge base on genetic variation and human diseases. *Am. J. Epidemiol.* 170: 269-279.

Kwon SJ, Brown AF, Hu J, McGee R, et al. (2012). Genetic diversity, population structure and genome-wide marker-trait association analysis emphasizing seed nutrients of the USDA pea (*Pisum sativum* L.) core collection. *Genes Genomics* 34: 305-320.

Li M, Guo XH, Zhang M, Wang XP, et al. (2010). Mapping QTLs for grain yield and yield components under high and low phosphorus treatments in maize (*Zea mays* L.) *Plant Sci.* 178: 454-462.

Li YL, Li XH, Li JZ, Fu JF, et al. (2009). Dent corn genetic background influences QTL detection for grain yield and yield components in high-oil maize. *Euphytica* 169: 273-284.

Liu K and Muse SV (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21: 2128-2129.

Liu XH, Zheng ZP, Tan ZB, Li Z, et al. (2010). Quantitative trait locus (QTL) mapping for 100-kernel weight of maize (*Zea mays* L.) under different nitrogen regimes. *Afr. J. Biotechnol.* 49: 8283-8289.

Liu ZH, Ji HQ, Cui ZT, Wu X, et al. (2011). QTL detected for grain-filling rate in maize using a RIL population. *Mol. Breeding* 27: 25-36.

Loiselle BA, Sork VL, Nason J and Graham C (1995). Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *Am. J. Bot.* 82: 1420-1425.

Lorenz AJ, Coors JG, Hansey CN, Kaeppler SM, et al. (2010). Genetic analysis of cell wall traits relevant to cellulosic ethanol production in Maize (*Zea mays* L.). *Crop Sci.* 50: 842-852.

Ma XQ, Tang JH, Teng WT, Yan JB, et al. (2007). Epistatic interaction is an important genetic basis of grain yield and its components in maize. *Mol. Breed.* 20: 41-51.

Mackay TFC, Stone EA and Ayroles JF (2009). The genetics of quantitative traits: challenges and prospects. *Nat. Rev. Genet.* 10: 565-577.

Malosetti M, van der Linden CG, Vosman B and van Eeuwijk FA (2007). A mixed-model approach to association mapping using pedigree information with an illustration of resistance to *Phytophthora infestans* in potato. *Genetics* 175: 879-889.

Mezmouk S, Dubreuil P, Bosio M, De'cousset L, et al. (2011). Effect of population structure corrections on the results of association mapping tests in complex maize diversity panels. *Theor. Appl. Genet.* 122: 1149-1160.

Nelson OE and Rines HW (1962). The enzymatic deficiency in the waxy mutant of maize. *Biochem. Biophys. Res. Commun.* 9: 297-300.

Pasam RK, Sharma R, Malosetti M, van Eeuwijk FA, et al. (2012). Genome-wide association studies for agronomical traits in a worldwide spring barley collection. *BMC Plant Biol.* 27: 12-16.

Pejic I, Ajmone-Marsan P, Morgante M, Kozumplick V, et al. (1998). Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSR, and AFLPs. *Theor. Appl. Genet.* 97: 1248-1255.

Pritchard JK, Stephens M and Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.

Rafalski JA, Vogel JM, Morgante M, Powell W, et al. (1996). Generating and using DNA markers in plants. In: Non-mammalian genomic analysis, a practical guide (Birren B and Lai E, eds.). Academic Press, San Diego, 75-134.

Rohlf FJ (1998). NTSYS- pc: Numerical taxonomy and multivariate analysis system, version 2.02. Exeter Software, Setauket, New York.

Skot L, Humphreys MO and Armstead I (2005). An association mapping approach to identify flowering time genes in natural populations of *Lolium perenne* (L.). *Mol. Breed.* 15: 233-245.

Spencer CC, Su Z, Donnelly P and Marchini J (2009). Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. *PLoS Genet.* 5: e1000477.

Stich B, Melchinger AE, Frisch M, Maurer HP, et al. (2005). Linkage disequilibrium in European elite maize germplasm investigated with SSRs. *Theor. Appl. Genet.* 111: 723-730.

Wang R, Yu Y, Zhao J, Shi Y, et al. (2008). Population structure and linkage disequilibrium of a mini core set of maize inbred lines in China. *Theor. Appl. Genet.* 117: 1141-1153.

Xie CX, Warburton M, Li MS, Li XH, et al. (2008). An analysis of population structure and linkage disequilibrium using multilocus data in 187 maize inbred lines. *Mol. Breed.* 21: 407-418.

Yan JB, Tang H, Huang YQ, Zheng YL, et al. (2006). Quantitative trait loci mapping and epistatic analysis for grain yield and yield components using molecular markers with an elite maize hybrid. *Euphytica* 149: 121-131.

Yu J and Buckler ES (2006). Genetic association mapping and genome organization of maize. *Curr. Opin. Biotechnol.* 17: 155-160.

Yu J, Pressoir G, Briggs WH, Bi IV, et al. (2005). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38: 203-208.

Zhang Q, Wu C, Ren F, Li Y, et al. (2012). Association analysis of important agronomical traits of maize inbred lines with SSRs. *AJCS* 6: 1131-1138.

Zhang Z, Ersoz E, Lai CQ, Todhunter RJ, et al. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* 42: 355-360.

Zhu C, Gore M, Buckler ES and Yu J (2008). Status and prospects of association mapping in plants. *Plant Genome* 1: 5-20.