Theor Appl Genet (2005) 111: 906–913
DOI 10.1007/s00122-005-0016-1

ORIGINAL PAPER

Jochen C. Reif · Sonia Hamrit · Martin Heckenberger
Wolfgang Schipprack · Hans Peter Maurer
Martin Bohn · Albrecht E. Melchinger

Genetic structure and diversity of European flint maize populations determined with SSR analyses of individuals and bulks

Received: 24 November 2004 / Accepted: 31 May 2005 / Published online: 30 July 2005

Abstract Landraces of maize represent a valuable genetic resource for breeding and genetic studies. Using simple sequence repeat (SSR) markers, we analysed five flint maize populations from Central Europe that had played an important role in the pre-hybrid era in Germany. Our objectives were to (1) investigate the molecular genetic diversity within and among the populations based on the SSR analysis of individuals, (2) compare these results of the SSR analysis based on individuals with those based on bulks, (3) examine genotype frequencies for deviations from Hardy–Weinberg equilibrium (HWE) at individual loci, and (4) test for linkage disequilibrium (LD) between pairs of loci within populations. Thirty individuals and their bulked DNA per population were fingerprinted with 55 SSR markers. Across all populations, 46.7% of the SSR markers deviated significantly from HWE, with an excess of homozygosity in 97% of the cases. This excess of homozygosity can largely be explained by experimental errors during the amplification of SSRs apart from genuine genetic causes. Allele frequencies of the SSR analyses of individuals and bulks were significantly correlated ($r = 0.85$, $P < 0.01$), suggesting that SSR analysis of bulks is very cost-effective for large-scale molecular characterisation of germplasm collections. No evidence for genome-wide LD among pairs of loci was observed, indicating that the populations are well suited for high resolution association mapping studies.

Communicated by D. A. Hoisington

J. C. Reif · S. Hamrit · M. Heckenberger · W. Schipprack
H. P. Maurer · A. E. Melchinger (✉)
Institute of Plant Breeding, Seed Science,
and Population Genetics, University of Hohenheim,
70593 Stuttgart, Germany
E-mail: melchinger@uni-hohenheim.de
Fax: +49-711-4592343

M. Bohn
Crop Science Department, University of Illinois,
S-110 Turner Hall MC 046 1102 S Goodwin Ave,
Urbana Champaign, IL 61801, USA

Introduction

Maize was first introduced into Europe by Columbus from the West Indies to southern Spain in 1493. Later on, maize germplasm was also imported from various other regions of the New World. In particular, North American flint populations have played a key role in the adaptation of the crop to the cool climatic conditions of Central Europe (Rebourg et al. 2003). Since the introduction of maize to Europe five centuries ago, based on theoretical arguments one might expect that genetic diversity was on one hand enhanced by mutation, meiotic recombination, as well as new introgressions and on the other reduced by genetic drift as well as natural and artificial selection (Falconer and Mackay 1996). This resulted in numerous open-pollinated varieties (OPVs) adapted to specific regional growing conditions and needs of farmers.

After World War II, traditional OPVs of maize in Central Europe have largely been replaced by hybrids generated by crossing adapted European flint lines with high-yielding US dent lines (Schnell 1992). The flint lines were extracted from a few European OPVs. Therefore, it can be conjectured that those OPVs, which did not serve as germplasm source for the current elite flint breeding pool, may contain useful untapped allelic variation. Furthermore, in the first phase of hybrid breeding in Central Europe the main focus was on grain but later on fodder usage. Thus, OPVs may contain favourable alleles for fodder usage of maize absent in the elite germplasm.

Open-pollinated varieties conserved in seed banks must be well-characterised for efficient management and effective exploitation. The advent of PCR-based molecular markers such as SSRs has provided an opportunity for fine-scale genetic characterisation of germplasm collections. Since SSR markers are highly polymorphic (Smith et al. 1997), easy to generate, and highly repeatable (Warburton et al. 2002), they can be used for large-scale investigations as needed for the characterisation of genetic resources (Powell et al. 1996).

Most molecular marker studies investigated individual genotypes. In contrast, some authors (Dubreuil et al. 1999; Rebourg et al. 2001, 2003; Gauthier et al. 2002) analysed bulks of individuals with restriction fragment length polymorphism (RFLP) markers and estimated the marker allele frequencies from band intensities. This RFLP analysis of bulks was recommended as a cost-effective strategy for large-scale grouping of OPVs. To our knowledge, no study is available investigating the use of bulks in SSR analyses.

Association mapping studies have been proposed as a promising approach to detect genes and alleles of interest in germplasm collections (Lynch and Walsh 1997). The mapping resolution of association studies is determined by the extent of LD across the genome. LD depends on the genealogy of the germplasm. In addition, drift and selection within populations can cause LD. Therefore, the genomic structure of LD must empirically be determined before initiating association studies, because it can vary among different germplasm. While LD was investigated in tropical and subtropical maize populations (Reif et al. 2004), no information is available about LD in temperate OPVs.

The main goal of our study was to use SSRs for characterising the population genetic structure of five important flint maize populations from Central Europe, as a basis for an efficient use of this germplasm in breeding programs. Our objectives were to (1) investigate the molecular genetic diversity within and among populations based on SSR analysis of individuals, (2) compare these results of the SSR analysis based on the individuals with those based on bulks, (3) examine genotype frequencies for deviations from Hardy–Weinberg equilibrium (HWE) at individual loci, and (4) test for LD between pairs of loci within populations.

Materials and methods

Plant material

Five maize populations from Central Europe (Table 1) that had played an important role in the pre-hybrid era in Germany (Oettler et al. 1976) were investigated. The five OPVs are typical gene bank materials, where only limited information is available about the genealogy of the germplasm such as the number of plants originally collected per population. Each of the five populations was regenerated in the last generation by sowing a minimum of 20 rows with 24 seeds per row. All plants were shoot bagged and pollen from 10 rows was bulked to pollinate plants in the other 10 rows and vice versa. Thirty individuals were randomly chosen from seed of 400 ears harvested per population. The number of individuals used to represent an OPV is higher than commonly applied in diversity studies (e.g., Labate et al. 2003) because we also examined LD within each OPV. In addition, two US inbred lines (Mo17 and B73), as well as six commercial hybrids and their parents were used as controls.

SSR analyses

DNA was extracted by a modified CTAB procedure from each individual plant and from bulked leaf material from the 30 individuals per population (Saghai-Maroof et al. 1984). The 150 individuals and five bulks were fingerprinted with a set of 55 SSRs uniformly distributed across the maize genome (http://www.uni-hohenheim.de/jochreif/Marker/SSRs.html). PCR amplification was performed in 15 µl reaction containing the following reagents: 50 ηg of template DNA, 0.15 mM of each of the four dNTPs, 1×Taq polymerase buffer, 1 U Taq polymerase (Amersham Biosciences Europe GmbH, Freiburg, Germany), 2.5 mM $MgCl_2$, and 0.25 µM of each of the two primers. Amplification was performed in a Thermal Cycler (MWG BIOTECH AG, Primus HT, Ebersberg, Germany). Fragments were separated using polyacrylamide gels (ultra pure SequaGel-6, National Diagnostics, Atlanta, Georgia) run on an ALF express sequencer (Amersham Biosciences Europe GmbH, Freiburg, Germany). Fragment size data were recorded by software ALFwin version 2 (Amersham Biosciences Europe GmbH, Freiburg, Germany) and manually checked. The fragment with the highest fluorescent intensity was scored when SSR-primed products showed band stuttering.

Table 1 Description of the five Central European maize populations analysed in this study

Population	Abbreviation	Source and Germplasm description
Gelber Badischer Landmais	GB	University of Hohenheim. Collected in Rastatt, Germany, in 1949. Yellow flint type. Cylindrical ear shape
Maleksberger	MB	University of Hohenheim. Collected site and year are unknown. Yellow flint type
Mahndorfer	MD	University of Hohenheim. Collected in Hamersleben, Germany in 1959. Yellow flint type. Conical ear shape
Rheintaler	RT	Gene bank Switzerland. White flilnt type. Collection site and year are unknown. Long, slightly conical ears
Strenzfelder	SF	KWS SAAT AG. Collected in Bernburg, Germany, in 1959. Yellow flint type

Statistical analyses

The SSR allele frequencies of the bulks were estimated by dividing the peak area of each band by the cumulative peak area of all bands observed with the respective SSR marker for the particular population. We determined the number of alleles per locus for the entire set of 150 individuals as well as for each population separately, and examined the presence of population-specific alleles (further referred to as unique alleles). Additionally, the total gene diversity (H_T) across all populations and the gene diversity between individuals within each population (H_S) were calculated according to Nei (1987). Significant differences between H_S values were tested for the SSR analyses of the individuals by a Wilcoxon signed rank test (Hollander and Wolfe 1973). The coefficient of gene differentiation (G_{ST}) was calculated according to Nei (1987). G_{ST} is the relative differentiation of the populations. The fixation index F_{IS} for each marker was estimated separately for each population according to Nei (1987) as one minus the observed heterozygosity divided by the expected heterozygosity under HWE.

Deviations from HWE at individual loci were tested with an exact test described by Guo and Thompson (1992). LD between all pairs of loci was estimated for each population by calculating the composite disequilibria coefficient and by applying the chi-square test statistics (Weir 1996). The LD analyses were performed with the GDA 1.0 software (Lewis and Zaykin 1999). For testing the global null hypothesis that populations are in LD, the Bonferroni correction for multiple tests (Snedecor and Cochran 1980) was applied.

To examine technical problems associated with non-detection of heterozygous SSR loci, the genotypes of the fingerprinted hybrids were compared with the expected hybrid genotypes predicted from the genotypes of the parental inbreds (further referred to as predicted hybrids). Differences among genotypes and SSR markers for susceptibility of non-detection of heterozygous loci were tested with a contingency table chi-square test (Patefield 1981).

The modified Rogers' distance (MRD) between populations was calculated according to Wright (1978) and Goodman and Stuber (1983) using allele frequencies obtained from the SSR analyses of bulks and individuals. For the latter, standard errors of MRD estimates were calculated by using a bootstrap procedure with resampling over markers and individuals. Correlation between the distance matrices of the SSR analyses of individuals and bulks were tested with the Mantel test (Legendre and Legendre 1998). Associations among the five populations were revealed by principal coordinate analysis (PCoA) (Gower 1966) based on MRD values.

Genetic relationships among individual genotypes from the five populations were analysed with a model-based clustering approach using software package STRUCTURE (Pritchard et al. 2000). We followed the suggestions of Pritchard et al. (2000) and did not run STRUCTURE to determine the numbers of populations but assumed five populations. Six runs of STRUCTURE were performed, setting the burn-in time to 100,000 and replication number to 1,000,000. The run with the highest maximum likelihood was used to assign individual genotypes to the five populations.

A simulation study was conducted to examine the influence of non-detection of alleles in SSR analysis of bulks on the associations among the populations. Therefore, the individuals of each of the five populations were divided in several subsamples and it was assumed that alleles below frequencies of 0.05, 0.10, 0.15, 0.20, 0.25, and 0.30 required undetected in scoring. Based on this modified allele frequencies of the populations, MRDs were calculated and the correlation with the original MRD values were determined. The sampling was repeated 100 times and results were averaged.

The statistical analyses were carried out with the Plabsoft software (Maurer et al. 2004), which is implemented as an extension to the statistical software R (Ihaka and Gentleman 1996).

Results

Genetic diversity within populations

The 55 SSR markers were polymorphic across all 150 genotypes in the analysis of the individuals. For the SSR analysis of bulks, five loci were monomorphic across the five populations. We found a significantly higher average number of alleles per locus in the SSR analysis of individuals (5.9) than in the analysis of bulks (3.2) (Table 2). The majority of the alleles (83%), which remained undetected in the analysis of bulks, had an estimated allele frequency below 0.2 in the analysis of individuals (Fig. 1). We observed in the five OPVs seven alleles, which were present in the analysis of the bulks but absent in the analysis of the individuals. The number of unique alleles per population ranged for the SSR analysis of individuals from 9 (RT) to 28 (MD) and for the analysis of bulks from 9 (RT) to 17 (MB). The correlation of allele frequencies within populations between the SSR analysis of individuals and bulks was highly significant ($r = 0.85$, $P < 0.01$). In addition, strong associations were observed between the gene diversity values H_S based on the SSR analysis of individuals and bulks. The lowest gene diversity value within the five populations was observed for RT in both types of analyses. The coefficient of gene differentiation G_{ST} between the populations was lower for the SSR analyses of individuals (0.22) than for bulks (0.51).

Relationships between populations

MRD values between pairs of populations for the SSR analyses of individuals and bulks were significantly ($r = 0.90$, $P < 0.05$) correlated (Table 3) according to the Mantel test. MRD between populations based on the

Table 2 Genetic diversity within and among five Central European maize populations, Gelber Badischer Landmais (GB), Maleksberger (MB), Mahndorfer (MD), Rheintaler (RT), and Strenzfelder (SF) revealed by SSR analyses of individuals (I) and bulks (B)

Population	No. of individuals	Avg. no. of alleles per locus		Monomorphic loci		No. of unique alleles[a]		Gene diversity[b]		Fixation index (F_{IS})
		I	B	I	B	I	B	I	B	I
GB	30	3.5	1.7	3	26	18	10	0.41[a]	0.27	0.27
MB	30	3.3	1.7	2	21	22	17	0.43[a]	0.30	0.35
MD	30	3.8	1.6	3	26	28	12	0.45[a]	0.25	0.48
RT	30	2.7	1.4	6	36	9	9	0.31[b]	0.16	0.47
SF	30	3.4	1.5	2	27	20	10	0.45[a]	0.23	0.25
Total/Mean	150	5.9	3.2	0	5			0.53	0.50	0.34

[a]Number of unique alleles with respect to the total number of alleles found in all five populations
[b]Gene diversity values followed by the same letters are not different at the 0.01 probability level

SSR analysis of individuals averaged 0.39 and ranged from 0.32 (GB × SF) to 0.43 (RT × SF) with significant differences ($P < 0.01$) between MRD estimates. MRD values based on the SSR analysis of bulks averaged 0.56 and ranged from 0.49 (GB × SF) to 0.60 (RT × SF). The average correlation between MRD based on SSR analysis of individuals and between MRD values based on simulated SSR analysis of bulks decreased with increasing frequencies of alleles undetected in scoring (Table 4).

Table 3 Modified Rogers' distances (MRD) between five Central European flint populations based on SSR analyses of individuals (above diagonal) (average standard error 0.03) and bulks (below diagonal)

	GB	MB	MD	RT	SF
GB		0.41	0.35	0.43	0.32
MB	0.57		0.38	0.42	0.40
MD	0.53	0.56		0.38	0.35
RT	0.59	0.57	0.57		0.43
SF	0.49	0.61	0.53	0.60	

GB Gelber Badischer Landmais; *MB* Maleksberger; *MD* Mahndorfer; *RT* Rheintaler; *SF* Strenzfelder

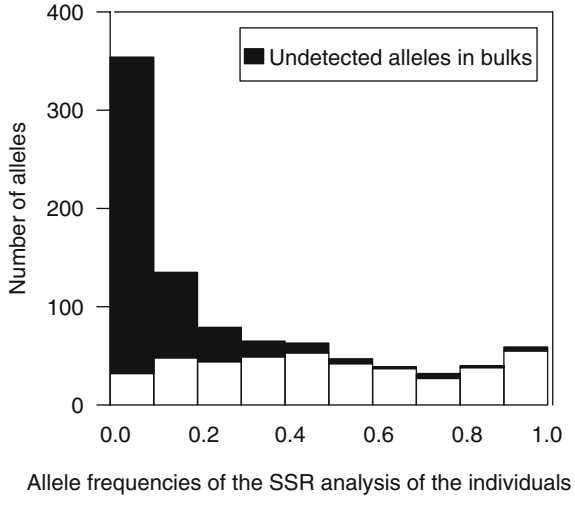


Fig. 1 Number of alleles observed in the SSR analysis of individuals and number of alleles that remained undetected in the analysis of bulks of five Central European maize populations

The first three principal coordinates (PC) explained a total of 83.5% and 81.6% of the molecular variation among the populations based on the SSR analyses of individuals and bulks, respectively (Fig. 2). In both PCoAs, the three populations RT, MB and MD formed distinct clusters, separate from GB and SF. In the model-based clustering, the individual genotypes were grouped into their respective populations except for one genotype of SF and RT, which clustered with 55% and 85% probability, respectively, to MD (Fig. 3).

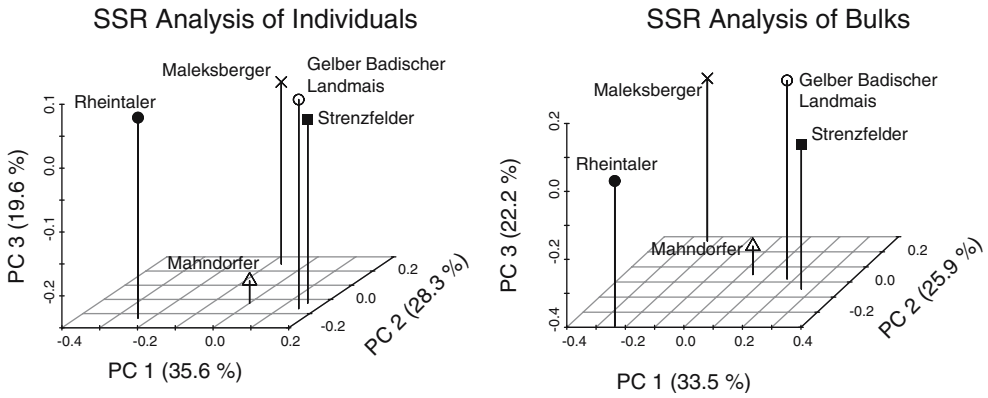
Population genetic analysis

The percentage of loci with significant ($P < 0.01$) deviations from HWE varied from 34.6% (GB) to 59.6% (MD), with an average of 46.7% (Fig. 4). In 97% of the cases, deviations from HWE were attributable to an excess of homozygosity. F_{IS} values ranged from 0.25 (SF) to 0.48 (MD) (Table 2). The percentage of SSR loci pairs in linkage disequilibrium ($P < 0.01$) averaged 1%

Table 4 Average correlations between modified Rogers' distances (MRD) based on SSR analysis of individuals and MRD values based on simulated SSR analysis of bulks

Threshold for allele frequencies	Individuals per subsample (N)			
	30	15	10	5
< 0.05	0.999	0.999	1.000	1.000
< 0.10	0.966	0.992	0.998	1.000
< 0.15	0.968	0.983	0.989	0.994
< 0.20	0.968	0.971	0.979	0.994
< 0.25	0.959	0.967	0.976	0.982
< 0.30	0.811	0.926	0.956	0.979

For the simulated SSR analysis individuals of each of the five populations were divided into subsamples of size (N) and it was assumed that alleles below a certain threshold remained undetected in the scoring of the bulks



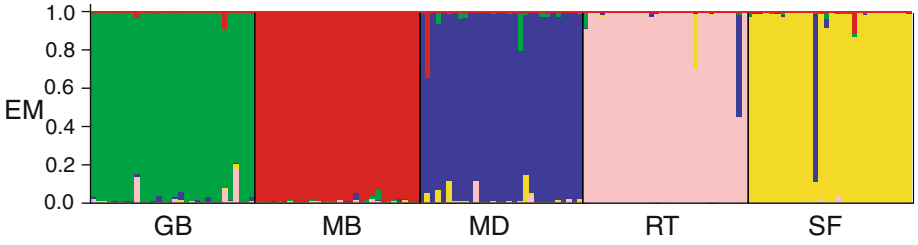


Fig. 2 Principal coordinate analysis of the five Central European maize populations based on modified Rogers' distances calculated from the SSR allele frequencies within populations for the SSR analyses of individuals and bulks. PC1, PC2, and PC3 denote the first, second, and third principal coordinate, respectively

Fig. 3 Estimated population structure of 150 individuals from five Central European maize populations. Each individual is represented by a *thin vertical line*, which is partitioned into five coloured segments that represent the individual estimated membership (*EM*) to the five clusters. *GB* Gelber Badischer Landmais, *MB* Maleksberger, *MD* Mahndorfer, *RT* Rheintaler, and *SF* Strenzfelder

(Table 5). We observed almost as much inter-chromosomal ($P < 0.01$) as intra-chromosomal LD. Application of the Bonferroni correction for multiple tests revealed that LD within the populations was not significant.

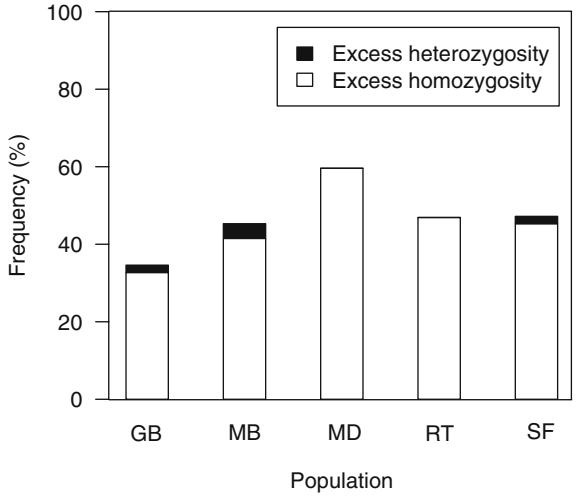


Fig. 4 Frequency of SSR loci with significant ($P \leq 0.01$) Hardy–Weinberg equilibrium tests in five Central European maize populations: Gelber Badischer Landmais (*GB*), Maleksberger (*MB*), Mahndorfer (*MD*), Rheintaler (*RT*), and Strenzfelder (*SF*)

Discussion

Molecular diversity within and among the five maize populations based on SSR analysis of individuals

The average number of 5.9 alleles per SSR in our study is in close agreement with the findings of Senior et al. (1998) (5.2), but lower than reported by Matsuoka et al. (2002) (6.9), Labate et al. (2003) (6.5), and Liu et al. (2003) (21.7). The lower number of alleles per locus is most likely attributable to the reduced genetic diversity in the plant material investigated (only Central European flint populations). Additionally, exclusion of di-nucleotide repeat SSRs in this survey, which tend to be more polymorphic than SSRs with longer-repeat motifs (Vigouroux et al. 2002), could also have caused the differences in the average number of alleles.

Table 5 Percentage of SSR loci pairs in linkage disequilibrium (% LD) at the 0.01 probability level for inter- (Inter) and intra-chromosomal (Intra) comparisons within five Central European maize populations (*GB* Gelber Badischer Landmais; *MB* Maleksberger; *MD* Mahndorfer; *RT* Rheintaler; *SF* Strenzfelder)

	Inter		Intra		Total	
	NT	%LD	NT	%LD	NT	%LD
GB	1211	2	109	0	1320	2
MB	1262	1	115	1	1377	1
MD	1083	1	109	0	1192	1
RT	978	2	90	0	1068	2
SF	1260	1	116	3	1376	1

NT refers to the number of tests

The total gene diversity of the five flint populations in our survey (0.53) was similar to the gene diversity of the Stiff Stalk pool (0.51) but lower than the gene diversity of the tropical (0.68) and the non-Stiff Stalk pool (0.68) when di-nucleotide SSRs were excluded (Liu et al. 2003). Thus, our results indicated that the original Central European flint germplasm had possessed a limited genetic variability. This can be due to (1) a bottleneck during the introduction of maize into Europe and/or (2) a loss of genetic diversity during the selective adaptation to European conditions (Rebourg et al. 2003). Among the five populations investigated, RT showed by far the lowest gene diversity, number of alleles per locus, number of unique alleles per locus, and the highest number of monomorphic loci (Table 2). This may reflect the intrinsic narrow genetic base of RT. Moreover, the low genetic diversity could have been caused by a bottleneck during the maintenance of this population in the seed bank.

PCoA based on MRD showed that two populations, GB and SF, clustered closely together (Fig. 2). Although GB was collected in South Germany and SF in East Germany (Table 1), a closer examination of possible links between the two populations suggests that SF is most likely a derivative of GB because the breeder of SF in East Germany worked before in the Rastatt region (South Germany), where GB was traditionally cultivated (W. Schmidt, personal communication). Nevertheless, the model-based cluster analysis revealed that drift, migration, and selection have led to allele frequency changes resulting in a clear separation of individuals from GB and SF (Fig. 3).

Comparison of SSR analysis of individuals versus bulks in genetic diversity studies

Although a large number of alleles with low frequency remained undetected in the SSR analysis of bulks (Fig. 1), the average number of alleles per locus within the populations was significantly correlated between the SSR analyses of individuals and bulks ($r = 0.92$, $P < 0.05$). This strong association can be explained by a homogeneous non-detection of alleles across the five populations. As the non-detection of alleles mainly depends on the allele frequency in the populations (Fig. 1), the application of average number of alleles as diversity measure is highly questionable for SSR analysis of bulks.

The low number of alleles detected in the SSR analysis of bulks but absent in the SSR analysis of the individuals suggests that only a small fraction of the amplified bands are an artifact. While most alleles with a frequency below 0.2 remained undetected in the SSR analysis of bulks in our survey (Fig. 1), Dubreuil et al. (1999) reported in a RFLP study of bulks that alleles could be detected as long as their frequency was greater than 0.05. The higher sensitivity in the latter survey compared to our study may be due to the lacking PCR step in the RFLP analysis: during the PCR reaction, alleles with low frequency may not be amplified due to competition for polymerase and other reagents. Another reason could be the different number of individuals used per bulk. While we used one bulk of 30 individuals per population, Dubreuil et al. (1999) took three bulks of ten individuals per population. The results of the simulated SSR analysis of bulks strongly suggests that even if alleles with frequencies below 0.25 were not scored, the correlation between MRD values of SSR analysis of individuals and bulks is high (Table 4). Consequently, associations among such diverse populations can be accurately displayed. Under the assumption of a constant number of individuals analysed, the precision in estimating allele frequencies increases with decreasing number of individuals per bulk. Thus, for closely related germplasm a higher number of individuals per bulks is advantageous to correctly display the associations among the populations. In addition, it is of interest to examine whether bulking should be done before or after DNA extraction. Bulking before DNA extraction, as applied in this study and by Dubreuil et al. (1999), offers big advantages with regard to time and cost effectiveness, but could also cause biases due to an unequal DNA contribution of individuals.

Our results clearly showed that the allele frequencies of a population could be estimated with high accuracy from the band intensities of SSR profiles of bulks. This suggests that the SSR analysis of bulks allows to (1) measure the gene diversity within populations and (2) examine the genetic relatedness among populations (Table 2, Fig. 2). The higher G_{ST} value in the SSR analysis of bulks compared to individuals indicates that differences between populations are emphasised due to a concentration on alleles with high frequencies. This hypothesis has to be verified with a larger sample of populations. Summarising, SSR analyses of bulks promise to be a cost-effective method to characterise a large number of populations in seed banks and, consequently, facilitate the conservation and use of genetic resources.

Population genetic analysis

In a large population with no selection, mutation, or migration, HW proportions are attained after one generation of random mating (Hardy 1908; Weinberg 1909). Thus, considering the procedure to maintain the five OPVs, it was expected that the populations were in HWE. Nevertheless, all five maize populations deviated significantly from HWE at a large number of SSRs (Fig. 4) and showed an excess of homozygous loci (Table 2). The deviations from HWE are in agreement with previous studies with random mating populations using (1) PCR-based markers (Reif et al. 2004), (2) RFLP markers (Dubreuil and Charcosset 1998; Labate et al. 2000), or (3) protein markers (Kahler et al. 1986; Revilla et al. 1997). However, F_{IS} values were higher

when PCR-based marker systems rather than protein or RFLP markers were used.

The excess of homozygosity of the five maize populations used in this study can be related to the following genetic causes: (1) positive assortative mating between individuals (homogamy), (2) artificial subgrouping of individuals within populations, and (3) selection favouring homozygotes. Positive assortative mating can be caused by genetic variation in flowering. In this case, only SSRs closely linked to QTL for flowering time should display a higher degree of homozygosity than expected under HWE. However, LD between the 55 SSRs and QTL for flowering time is unlikely due to the uniform distribution of the markers across the entire maize genome. Artificial subgrouping of genotypes within populations was not observed in the PCoA of individuals (data not shown) and, thus, cannot explain the excess of homozygosity. Selection favouring homozygotes is unlikely in maize, where fitness increases with heterozygosity due to positive heterosis for plant vigour. Summarising, genetic causes provide at best a partial explanation for the observed deviation from HWE.

To examine whether technical problems may have caused the excess of homozygosity, hybrids were compared with predicted hybrids. The high percentage of heterozygous loci in the predicted hybrids, which were homozygous in the hybrids (28.6%), strongly suggests that the excess of homozygosity is related to the experimental errors during the SSR assays. Assuming that the error rate of the hybrids also applies to the populations, this results in an average F_{IS} value of 0.29, which is similar to the observed total F_{IS} value (0.34). Thus, the excess of homozygosity in our study can largely be explained by experimental errors apart from genuine genetic causes.

The following laboratory errors would lead to an overestimation of homozygosity: (1) alleles remained undetected due to competition for polymerase and other reagents during amplification in the PCR reaction, (2) the setting of the threshold of band intensity to detect alleles was too strict, and (3) a heterozygous locus carrying a null allele was scored as homozygous. The rate of null alleles can be estimated in an analysis of inbred lines and was reported to be generally low in elite Central European flint germplasm for the 55 SSRs used in this study (Reif et al. 2005). Additionally, a low threshold of band intensity to detect alleles was applied and, consequently, lab errors caused by the detection systems were minimised. Therefore, non-amplification of a second allele during the PCR reaction is presumably the major reason for the excess of homozygosity.

The significant differences among SSRs in the hybrids for non-detection of heterozygous loci in the contingency table chi-square test and the significant correlations between F_{IS} values of all pairwise population comparisons strongly suggest differences among SSRs in their susceptibility towards lab errors. We tested the possibility to pre-select SSRs less affected by lab errors, based on the comparison of hybrids with predicted hybrids. Therefore, SSRs heterozygous in the predicted hybrids have been considered and those markers have been selected that were also heterozygous in the hybrids. The selected SSRs had on average lower F_{IS} values in the five populations than all SSRs, and 60% of these markers showed no excess of homozygosity. Consequently, it would be prudent in future studies investigating HWE within populations to pre-select SSRs robust against lab errors, based on a test set of hybrids and their parental inbred lines. The high level of lab errors for detecting heterozygous loci have strong implications, for instance in mapping studies, and warrants further detailed investigations by fingerprinting a larger number of hybrids and their parental inbred lines with different PCR-based marker systems in different laboratories with different detection systems.

Technical problems during the SSR assays causing a systematic non-amplification of specific alleles could also influence the LD among marker pairs. To investigate whether specific alleles remained undetected during the SSR assays, it would be prudent in future studies to repeat the SSR analysis of the predicted hybrids. We assumed that the non-detection of alleles occurred randomly in heterozygous individuals, which would not influence LD detection. LD was tested by applying the composite disequilibrium coefficient (Weir 1996), because this method is free of the HWE assumption. We found that on average 1% of the two-locus disequilibrium tests were significant, which can be explained by Type I error alone. This is in agreement with a previous study with OPVs (Reif et al. 2004) but in apparent disagreement with surveys within elite breeding material (Liu et al. 2003; Remington et al. 2001). Thus, the absence of genome-wide LD in our study and the presence of LD in the surveys of Liu et al. (2003) and Remington et al. (2001) can be explained by the different type of plant materials used. The LD in OPVs is expected to be lower than in elite breeding germplasm due to (1) a lower selection intensity and (2) a large number of inter-mating generations since the establishment of the populations.

The lack of LD in the five flint populations indicates that the populations are well suited for high resolution association mapping studies. As a first step towards association mapping we suggest to decompose the OPVs into representative samples of homozygous lines by using, for example, doubled haploid lines. The decomposition has two advantages: (1) the resulting genotypes are immortal and can be subject to replicated tests for phenotyping and (2) the genetic load is reduced by eliminating deleterious alleles during the inbreeding process. The lines can then serve as a base population to identify favourable alleles with high resolution association mapping. This strategy would facilitate a systematic exploitation of the available genetic diversity in maize populations.

Acknowledgments This research was part of project 'Genetic Diversity in Agriculture: Temporal Flux, Sustainable Productivity and Food Security', which was supported by a research grant from the European Union in Framework 5 Quality of life and management of living resources program (QLRT-2000-00934). We

gratefully acknowledge the technical assistance of Ms. E. Kokai-Kota in generating part of the SSR data. The authors thank two anonymous reviewers for their valuable suggestions.

References

Dubreuil P, Charcosset A (1998) Genetic diversity within and among maize populations: a comparison between isozyme and RFLP loci. Theor Appl Genet 96:577–587

Dubreuil P, Rebourg C, Merlino M, Charcosset A (1999) Evaluation of a DNA pooled-sampling strategy for estimating the RFLP diversity of maize populations. Plant Mol Biol Rep 17:123–138

Falconer DS, TFC Mackay (1996) Introduction to quantitative genetics, 4th edn. Longman Group Ltd, London

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

Goodman MM, Stuber CW (1983) Races of maize: VI. Isozyme variation among races of maize in Bolivia. Maydica 28:169–187

Gower JC (1966) Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53:325–338

Guo S, Thompson E (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48:361–372

Hardy GH (1908) Mendelian proportions in a mixed population. Science 28:49–50

Hollander M, Wolfe DA (1973) Nonparametric statistical inference. Wiley, New York, pp 139–146

Ihaka R, Gentleman R (1996) A language for data analysis and graphics. J Comput Graph Stat 53:299–314

Kahler AL, Hallauer AR, Gardner CO (1986) Allozyme polymorphisms within and among open-pollinated and adapted exotic populations of maize. Theor Appl Genet 72:592–601

Kimura M, Ohta T (1978) Stepwise mutation model and distribution of allelic frequencies in a finite population. Proc Natl Acad Sci USA 75:2868–2872

Labate JA, Lamkey KR, Lee M, Woodman W (2000) Hardy-Weinberg and linkage equilibrium estimates in the BSSS and BSCB1 random mated populations. Maydica 45:243–255

Labate JA, Lamkey KR, Mitchell SH, Kresovich S, Sullivan H, Smith JSC (2003) Molecular and historical aspects of Corn Belt dent diversity. Crop Sci 43:80–91

Legendre P, Legendre L (1998) Numerical ecology, 2nd edn. Elsevier, Amsterdam

Lewis PO, Zaykin D (1999) Genetic data analysis. Computer program for the analysis of allelic data. Version 1.0. Distributed by the authors at (http://lewis.eeb.uconn.edu/lewishome/software.html). (verified 1 August 2004). Lewis Labs, Univ. of Connecticut, Storrs, CT

Liu K, Goodman M, Muse S, Smith JS, Buckler E, Doebley J (2003) Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. Genetics 165:2117–2128

Lynch M, Walsh B (1997) Genetics and analysis of quantitative traits. Sinauer Assoc, Sunderland, p 413

Matsuoka Y, Mitchell SE, Kresovich S, Goodman M, Doebley J (2002) Microsatellites in *Zea*—variability, patterns of mutations, and use for evolutionary studies. Theor Appl Genet 104:436–450

Maurer HP, Melchinger AE, Frisch M (2004) Plabsoft: software for simulation and data analysis in plant breeding. XVIIth EUCARPIA General Congress 2004, Tulln, Austria

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Oettler G, Schnell FW, Utz HF (1976) Die westdeutschen Getreide- und Kartoffelsortimente im Spiegel ihrer Vermehrungsflächen. In: Alleweldt G (ed) Hohenheimer Arbeiten Schriftenreihe der Universität Hohenheim Reihe:Pflanzliche Produktion. Verlag Eugen Ulmer, Stuttgart, Germany

Patefield WM (1981) Algorithm AS159. An efficient method of generating r×c tables with given row and column totals. Appl Stat 30:91–97

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

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

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

Rebourg C, Gouesnard B, Welcker C, Dubreuil P, Chastanet M, Charcosset A (2003) Maize introduction into Europe: the history reviewed in the light of molecular data. Theor Appl Genet 106:895–903

Reif JC, Xia XC, Melchinger AE, Warburton ML, Hoisington DA, Beck D, Bohn M, Frisch M (2004) Genetic diversity determined within and among CIMMYT maize populations of tropical, subtropical, and temperate germplasm by SSR markers. Crop Sci 44:326–334

Reif JC, Hamrit S, Heckenberger M, Schipprack W, Bohn M, Melchinger AE (2005) Trends in genetic diversity among European maize cultivars and their parental components during the past 50 years. Theor Appl Genet (in press)

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler E (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci USA 98:11479–11484

Revilla P, Vales MI, Malvar RA, Ordas A (1997) Allozyme frequencies, heterozygosity and genetic distances following S_1 recurrent selection in two synthetic maize populations. Theor Appl Genet 95:1057–1061

Saghai-Maroof MA, Soliman KM, Jorgenson R, Allward RW (1984) Ribosomal DNA spacer length polymorphisms in barley: Mendelian inheritance, chromosomal location and population dynamics. Proc Natl Acad Sci USA 81:8014–8018

Schnell FW (1992) Maiszüchtung und die Züchtungsforschung in der Bundesrepublik Deutschland. Vorträge Pflanzenzüchtung 22:27–44

Senior ML, Murphy JP, Goodman MM, Stuber CW (1998) Utility of SSRs for determining genetic similarities and relationships in maize using an agarose gel system. Crop Sci 38:1088–1098

Smith JSC, Chin ECL, Shu H, Smith OS, Wall SJ, Senior ML, Mitchell SE, Kresovich S, Ziegle J (1997) An evaluation of utility of SSR loci as molecular markers in maize (*Zea mays L.*): comparisons with data from RFLPs and pedigree. Theor Appl Gen 95:163–173

Snedecor GW, Cochran WG (1980) Statistical methods. Iowa State University Press, Ames

Vigouroux Y, Jaqueth JS, Matsuoka Y, Smith OS, Beavis WD (2002) Rate and pattern of mutation at microsatellite loci in maize. Mol Biol Evol 19:1251–1260

Warburton ML, Xianchun X, Crossa J, Franco J, Melchinger AE, Frisch M, Bohn M, Hoisington D (2002) Genetic characterization of CIMMYT inbred maize lines and open pollinated populations using large scale fingerprinting methods. Crop Sci 42:1832–1840

Weinberg W (1909) Über Vererbungsgesetze beim Menschen. Zeitschrift für induktive Abstammungs- und Vererbungslehre 1:377–393

Weir BS (1996) Genetic data analysis II, 2nd edn. Sinauer Associates Inc, Sunderland

Wright S (1978) Evolution and genetics of populations, Vol. IV. The Univ of Chicago Press, Chicago, IL, p 91