RESEARCH ARTICLE

# Genetic diversity of Swiss maize (*Zea mays* L. ssp. *mays*) assessed with individuals and bulks on agarose gels

**Tobias W. Eschholz · Roland Peter · Peter Stamp · Andreas Hund**

**Abstract** About 65 years ago, more than 150 Swiss maize landraces (*Zea mays* L. ssp. *mays*) of the flint type were collected and conserved *ex situ*. Due to the climatically and culturally diverse environment of the Alps, a considerable genetic diversity of this material was assumed. To prove this, an efficient method was required to carry out genetic profiling of all the accessions in the Swiss Gene Bank. Simple sequence repeat marker (SSR) profiling in combination with the visualization of the polymerase chain reaction (PCR) products on agarose gels was chosen. Here a set of 19 different landrace accessions was analyzed to: (i) investigate their genetic diversity, (ii) investigate and display the population structure and (iii) determine whether DNA bulks rather than single plants can be used for such analyses. Four repeated samples of one accession were found to be much closer to one another than to the rest of accessions. Furthermore, specific alleles were identified for several accessions. The PCR products of the bulked DNA samples represented only a small part of the variation revealed by the analysis of individuals. Loci with four base repeat motifs performed better in the analysis of bulks than loci with other repeat motifs. The correlation between genetic distance matrices, based on the analysis of individuals and bulks, respectively, was significant. Thus, the single plant approach allowed for sufficient differentiation of accessions, and DNA bulks visualized on agarose gels led to correlated genetic distances although a limited number of alleles were detected. Although the limited resolution of agarose gels likely causes some bias, profiling of larger sets with the individual plant approach appears feasible and more informative compared to the bulk analysis we conducted.

**Keywords** Agarose gels · Bulk analysis · Genetic distance · Flint maize (*Zea mays* L. ssp. *mays*) · Landraces · SSR profiling

T. W. Eschholz
Pioneer Hibred Northern Europe, Muenstertaelerstr. 26, 79427 Eschbach, Germany

R. Peter · P. Stamp · A. Hund (✉)
Institute of Plant Sciences, ETH Zurich, Universitaetstr. 2, 8092 Zurich, Switzerland
e-mail: andreas.hund@ipw.agrl.ethz.ch

## Introduction

Maize of the flint type (*Zea mays* L. ssp. *mays*) has been grown in Switzerland for centuries. The geographic diversity of Switzerland and resulting potential isolation were the reasons to expect genetic diversity between accessions. In particular, early North American Flint populations have played a key role in the adaptation of the crop to the cool climatic conditions of central Europe and probably to the northern alpine valleys in Switzerland (Rebourg et al. 2003; Camus-Kulandaivelu et al. 2006; Dubreuil

et al. 2006). There is no genetic documentation of the origin of flint types grown in the southern Alpine valleys of Switzerland. However, a first investigation of genetic distances between Swiss accessions by use of DNA bulks revealed promising differences between accessions and even regions of origin (Eschholz et al. 2006). In addition to that, morphological data indicate that the southern Swiss accessions originated from southern European flint types described by Rebourg et al. (2003) and Brandolini (1970).

Landraces are open-pollinated varieties (OPVs), which have been multiplied and selected by farmers *in situ* (Zeven 1998); they are an important part of the genetic resources for maize (Carvalho et al. 2004). Since the number of OPVs in gene banks is large, they can only be used efficiently, if information about their characteristics is available. DNA profiling of maize OPVs with neutral markers may shed light on their genetic relationship and OPVs can be grouped for efficient use (Van Hintum and Van Treuren 2002).

The high level of genetic diversity in maize can be visualized with simple sequence repeat marker (SSR) polymorphisms displayed on agarose gels (Taramino and Tingey 1996; Senior et al. 1998; Sharopova et al. 2002; Taller and Bernardo 2004; Laborda et al. 2005) that are more cost-efficient but have a lower resolution compared to acrylamide gels or capillary sequencers. A number of agarose manufacturers offer high resolution agarose and recommend 3–4% gels for discriminating PCR fragments between 50 and 300 base pairs (bp), which is a resolution high enough to separate microsatellite products. The assessment of genetic diversity in maize is generally based on the analysis of individual plants. If high throughput is demanded, for example in studies, in which a huge number of accessions is analysed, then the investigation of bulked DNA is an attractive option. The feasibility of bulk analysis was demonstrated for the first time with restriction fragment length polymorphism (RFLP) markers (Dubreuil et al. 1999). Later bulk methods were established for AFLP (Beyene et al. 2006b), SSRs on acrylamide gels (Reif et al. 2005) as well as agarose gels (Beyene et al. 2006a). Generally bulks of DNA from 15 individuals were found to be informative. Drawbacks of the analysis with bulked DNA samples are the potential non-detection of alleles present at low

frequencies as well as the loss of information concerning the amount of heterozygosity within samples (Reif et al. 2005). In landrace accessions the variability between individuals within an accession is an important aspect as two accessions may contain the same alleles but differ by the frequency of alleles.

The single plant analysis should provide an insight into the distribution of genetic diversity within individuals, within accessions, and between accessions. To verify the possibility of further simplification of such a diversity study we challenged the analysis of single plants with an analysis of DNA bulks on agarose gels using the same ten SSR markers on DNA bulks of the very same accessions. The performance of our approach was evaluated by the possibility to identify samples of the same accession that were tested in the analysis of individual plants. The success in redetection of alleles in the analysis of bulks was reported. Based on preceding studies with maize (Reif et al. 2005; Beyene et al. 2006a) the redetection of most frequent alleles and a significant correlation between the genetic distance matrices from the analysis of individuals and the analysis of bulks were expected. It was our goal to test an assessment of the genetic diversity between accessions of a crop with a simple, efficient approach on agarose gels.

## Material and methods

### Genetic material

Among more than 150 Swiss flint maize (*Zea mays* L. ssp. *mays*) landrace accessions stored in the Swiss Gene Bank, a diverse set of 19 accessions was selected, including one that originated in Germany (127D = Gelber Badischer Landmais, Table 1). To ensure diversity of the 19 accessions, the selection was based on differences in root development, leaf greenness and shoot dry weight under cool conditions in growth chambers (Peter et al. 2006; Peter 2007) as well as on the region of origin (site of collection) of the landrace accessions (Fig. 1). Each accession was regarded as a subpopulation of the whole population under investigation. Six of the accessions were continuously conserved *in situ*, i.e., collected from the farmers' fields after 1988 and maintained *in situ*

**Table 1** List of the maize landrace accessions used for profiling

| ETH code[a] | Var. number[b] | Name[c] | Donor org.[d] | Collection date | Kernel col.[e] |
|---|---|---|---|---|---|
| 059LV | 13001059 | Eschenbach | ACW | 1949 | y |
| 055LV | 13001055 | Tuggen | ACW | 1942 | w/y |
| 089LV | 13001089 | Wangen | ACW | 1943 | y/sr/w |
| 007PR | 13001007 | Tartar | ACW | 1942 | y/b/w |
| 134PR | 13001134 | Almens | ACW | n.a. | y |
| 125PV | 13001125 | Brusio | ACW | 1941 | y/r |
| 018RV | 13001018 | Pfaeffers | ACW | 1941 | y |
| 103RV | 13001103 | Au | ACW | 1944 | w |
| 179RV | 13001179 | Vetsch, Buchs | VRRM | n.a. | w |
| 174RV | 13001174 | Balgach | VRRM | n.a. | w/y/b |
| 151RV | 13001151 | Mels | VRRM | n.a. | w |
| 176RV | 13001176 | Oberriet | VRRM | n.a. | w/y/b |
| 073TM | 13001073 | Gerra | ACW | n.a. | y (d) |
| 075TM | 13001075 | Vogorno | ACW | 1944 | y |
| 198TM | n.a. | Bleniotaler | SGE | 1990 | y/r |
| 425TM | n.a. | Verzasca-Mais | SGE | 1988 | y |
| 023VS | 13001023 | Vouvry | ACW | n.a. | y(d) |
| 025VS | 13001025 | Drone | ACW | n.a. | y/r |
| 127D | 13001127 | Gelber Badischer Landmais | ACW | n.a. | y |

Accessions are ordered by region of collection

[a] Codes used to name the accessions based on the last three digits of the var. number and their regions of origin: PR = Posterior Rhine, LV = Linth valley, PV = Poschiavo valley, RV = Rhine valley, TM = Tessin and Mesolcino, VS = Valais, D = Germany

[b] Entry number in the public Swiss national database (http://www.bdn.ch)

[c] Name commonly used for the accession

[d] Swiss seed donor organisations: ACW (Agroscope Changins Wädenswil, CH-1260 Nyon 1), VRRM (Verein Rheintaler Ribelmais, CH-9465 Salez), SGE (Sortengarten Erschmatt, CH-3597 Erschmatt)

[e] Kernel colour: y = yellow, b = blue/grey, r = red, w = white, v = violet/black red, (d) = dark

by private organizations (VRRM and SGE; Table 1). The remainder were conserved *ex situ* for more than 50 years. One accession was sampled four times for methodological reasons (025VS a-d). A total of 50 seeds of each accession were grown to multiply the seed in the field. From the harvest of each accession, a random sample was planted for DNA extraction to gain 10 samples. The controls were a typical flint × dent hybrid ('Magister', Syngenta, Basel, Switzerland) and a contrasting tropical highland hybrid ('CMS939083' [CMS], CIMMYT, El Batán, Mexico).

DNA extraction

Leaf material from 10 seedlings per accession was harvested individually when two to three leaves were fully developed. The leaves were cut, immediately put into liquid nitrogen, lyophilized for 48 h and stored at room temperature until further processing. DNA extraction from the dried material of 10 plants per accession was done with an extraction kit (Nucleo Spin Plant, Macherey Nagel GmbH & Co. KG, Düren, Germany) according to a modified standard protocol: The incubation time was extended to 80 min and the final elution was performed in two steps in 75 μl elution buffer. The DNA concentration of the resulting extracts was measured with a photometer. DNA quality was checked with gel electrophoresis on 0.7% agarose gels. Stock solutions were diluted to 10 ng/μl before PCR was carried out. The diluted fraction was transferred to 1.2-ml megatiter plates (Treff AG, Degersheim, Switzerland) with a sealing mat (Treff AG) and stored at 8°C until further processing. For the bulk test, equal
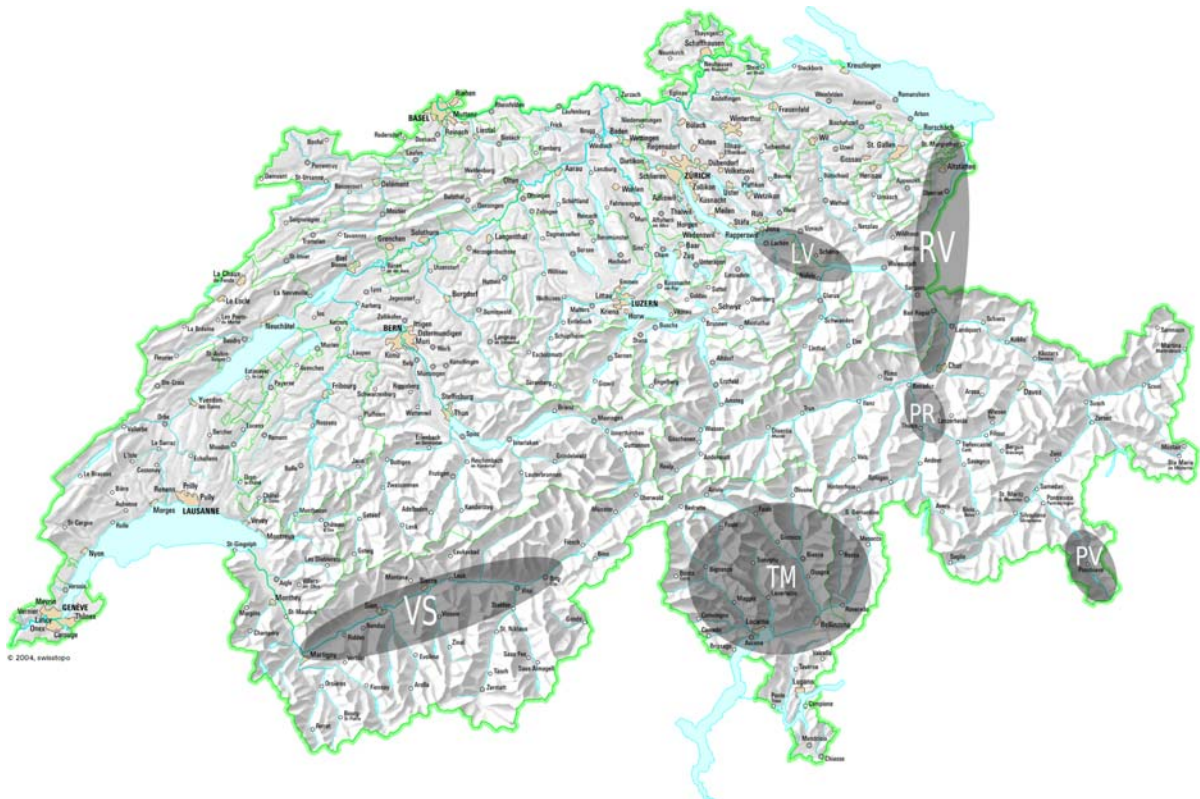
**Fig. 1** Regions of origin are shown by grey transparent ovals. White letters refer to the code in Table 1 (LV = valley, PR = Posterior Rhine, PV = Poschiavo valley, TM = Tessin and Mesolcino, VS = Valais). K606-01 © 2004 swisstopo

amounts of 10 ng/µl diluted DNA from the individuals representing one accession were mixed, and aliquots of the mixture were used for PCR.

Molecular markers

The SSRs were chosen from the Maize Genetics and Genomics Database (MGDB, Lawrence et al. 2004). To simplify genotyping, those markers (one per chromosome) were selected that showed large differences in allele size based on MGDB information. A total of 23 markers were first tested for their ability to reveal easily scorable polymorphism in individuals of a randomly selected landrace accession. When applied to all accessions markers were replaced if they did not show amplification in more than one accession or in more than 20% of all plants. A total of 10 markers were used in this study (Table 2). PCR conditions were adjusted for each individual marker to ensure reliable amplification. Annealing

temperatures ranged from 57 to 65°C (Table 2). PCR reactions were done with a thermal cycler (Mastercycler, Eppendorf, Hamburg, Germany) using TAQ polymerase and dNTP sets, all from Eppendorf. The thermal cycler program consisted of an initial denaturation step at 95°C for 2 min followed by 35 cycles of 15–60 s at 94°C, 15–60 s at annealing temperature and 30–60 s elongation at 72°C. The final step was an additional elongation period of 5 min at 72°C.

Genotyping

The PCR products were visualized on 4% agarose gels (Agarose 50–1000 bp, Roth, Karlsruhe, Germany) as described by Senior and Heun (1993). The gels were run in 1x TBE buffer (Sambrook et al. 1989) at 150 V for 105 min and stained in an ethidium bromide (EtBr) solution (1 mg/l). PCR products of 10 individual plants were followed by

**Table 2** Primers used for profiling are shown

| Description | | | | | Per locus statistics of analysis of individuals | | | | Performance of the analysis of bulks | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Name | Repeat motif | Annealing temp.[a] | Bin | Size[b] | Major allele frequency | Genotype no. | PIC | Av. allele no. | Av. allele no. bulks | Nulls in bulks[e] | % Artefacts in bulks |
| dupssr12 | (AC)15 | 58 | 1.08 | 67–154 | 0.37 | 18 | 0.73 | 10 | 1.38 | 1 | 30.77 |
| phi109642 | ACGG | 60 | 2.04 | 119–166 | 0.44 | 12 | 0.67 | 7 | 1.38 | 3 | 6.90 |
| umc1970 | TATG[c] | 62 | 3.01 | 130–184 | 0.29 | 16 | 0.74 | 10 | 1.33 | 2 | 28.57 |
| bmc1337 | (AG)21 | 60 | 4.1 | 127–228 | 0.28 | 30 | 0.84 | 15 | 1.57 | 2 | 50.00 |
| phi087 | ACC | 62 | 5.06 | 135–187 | 0.25 | 28 | 0.87 | 17 | 1.24 | 5 | 61.54 |
| ZAG249 | 2 Bases[d] | 57 | 6.01 | 107–165 | 0.28 | 31 | 0.84 | 11 | 1.43 | 0 | 43.33 |
| umc1125 | (CTCG)5 | 65 | 7.04 | 78–108 | 0.57 | 8 | 0.46 | 5 | 1.67 | 2 | 9.09 |
| dupssr14 | (CT)3 T (CT)6 (CA)16 | 59 | 8.09 | 67–116 | 0.26 | 22 | 0.85 | 12 | 0.71 | 7 | 46.67 |
| umc1033 | (GA)25 | 59.5 | 9.02 | 95–235 | 0.40 | 18 | 0.74 | 11 | 0.91 | 8 | 42.11 |
| umc1930 | (GT)8 | 62 | 10.04 | 90–143 | 0.29 | 28 | 0.77 | 10 | 1.71 | 5 | 25.71 |
| Mean | | | | | 0.34 | 21.1 | 0.75 | 10.8 | 1.3 | 3.5 | 34.5 |
| SD | | | | | 0.03 | 2.52 | 0.04 | 1.01 | | | |
| 2.5% 1.b.[bb] | | | | | 0.29 | 16.00 | 0.67 | 8.80 | | | |
| 97.5% u.b. | | | | | 0.41 | 26.00 | 0.81 | 12.80 | | | |

Primer name, sequence, bin and repeat motif are given according to MGDB (URL: http://www.maizegdb.org). Summary data were used to compare the performances of the analysis of individual plants and the analysis of bulks. Possible null alleles were treated as missing data

[a] Annealing temperatures adjusted for every primer pair

[b] 95% confidence intervals based on 1,000 bootstraps

[c] according to Vigouroux et al. (2005)

[d] according to Smith et al. (1997)

[e] number of accessions where alleles could not be detected by the analysis of bulks

the products of bulked DNA of the same individuals and a negative control (purified water instead of DNA). A 20 Bp ladder (Bio-Rad Laboratories, Hercules CA, USA) was used as size standard. Bulked DNA of each accession was included in this study to test its reliability to reveal allelic variability among samples of accessions when agarose gel is used. For accession 025VS, two bulks were created each consisting of 20 plants.

Each digital photograph (Doc Print, Vilber Lourmat, Cedex, France) consisted of half a gel row and had a resolution of 72 dots per inch. Digital images were analyzed using the software Gel-Pro Analyzer 4.5 (Media Cybernetics, Silver Spring, MD, USA). Allele size was determined with Gel-Pro Analyzer by comparison with the 20 Bp ladder with a minimal difference of 1 Bp between alleles. In Gel-Pro Analyzer the minimum band separation for differentiation of alleles was set to 1.5% of the ratio of distance migrated to total length of gel lanes (Rf).

For individual plants, the results from Gel-Pro Analyzer were visualized as count histograms of the reported allele sizes in base pairs, allowing for identification of "natural" allele classes at the peak positions in Bp (bins). Alleles were assigned to these "bins" before further processing of the data. Alleles were assigned to different bins when the bins showed clearly separate peaks in the histogram. In doubtful cases when alleles fell in regions with low frequency alleles and thus poorly defined bins, arbitrarily alleles were combined when their size differences were <3% of the allele sizes. To investigate the abundance of possible null alleles in our material, plants were considered to carry null alleles if more than 50% of the individuals of one accession (>5 individuals) showed non-amplification for one marker but their DNA showed amplification for the majority of the other markers. Therefore, non-amplification in these cases could not be due to poor DNA quality. For the bulks, the allele frequencies were determined by dividing the intensity of a specific band, measured with Gel-Pro Analyzer 4.5, by the sum of the intensity of all the bands present in the lane of this accession's bulk. For comparability of the analysis, alleles from bulks were assigned to the bins defined in the analysis of individuals. In general electrophoresis was repeated when alleles were not clearly visible and PCR was repeated if no alleles were present in more than one complete accession on a gel.

Statistics

$F$-statistics (Wright 1965; Weir and Cockerham 1984) were calculated and the test for Hardy–Weinberg Equilibrium (HWE) was computed according to Guo and Thompson (1992) to determine significant differences between the observed and expected levels of heterozygosity. The summary statistics, namely the average number of alleles, average gene diversity (or expected heterozygosity), average observed heterozygosity, polymorphic information content (Botstein et al. 1980) and inbreeding coefficient, were calculated. The average gene diversity was calculated as

$$\widehat{D}_l = \left(1 - \sum_{u=1}^{k} \widetilde{p}_{lu}^2\right) \qquad (1)$$

where $\widehat{D}_l$ is the common biased estimator of gene diversity, $\widetilde{p}_{lu}$ the population frequency of the $u$th allele at the $l$th locus and $k$ the total number of alleles in the population (Weir 1996). The inbreeding coefficient ($f$) was computed according to the method of moments (Weir and Cockerham 1984; Weir and Hill 2002), as implemented in Powermarker (Liu and Muse 2005). To obtain summary statistics, 40 individuals of 025VS were combined to one sample to investigate the effects of differences in the sample size. For the analysis of bulks, the 40 individuals of 025VS were represented with two bulks of 20 individuals for comparison. The average of the two bulks of 025VS was used to compare the summary statistics of individuals and bulks.

To calculate the genetic distances, potential null alleles were treated as missing data to avoid clustering caused by shared null alleles. The Modified Rogers' distance ($d_W$) was calculated based on allele frequencies (Wright 1978). Hierarchical cluster analysis (HCA) was performed using the $d_W$ matrix applying Wards hierarchical grouping method (Ward 1963) as agglomeration method. A Mantel test was computed with 10,000 permutations to compare the genetic distance matrices of the analysis of individuals with the analysis of the bulks (Mantel 1967) both based on the binned alleles. All calculations were performed with Powermarker 3.25 (Liu and Muse 2005) with the exception of the Ward dendrogram, which was constructed using R 2.0.1 software (R development core team 2004).

## Results

### Molecular markers

In the analysis of individual plants with the 10 SSR markers tested on 220 samples, 108 alleles were detected within the entire population, and the number of alleles per locus ranged from 5 to 17 with an average of 10.8 (Table 2). With an average PIC of 0.75 the selected markers appeared suitable for the differentiation of genotypes (Table 2). The average amount of possible null alleles per locus was 4.9% with a maximum of 18.4% for locus dupssr14 and the average amount of missing data was 9.8% with a maximum of 17% at locus bmc1337 (data not shown).

### Profiling of single plants

On average, 2.9 alleles per landrace accession and locus were found, ranging from 1.9 for 425TM to 3.6 for 007PR and 134PR (Table 3). Accession 025VS, with 40 instead of 10 individuals, had an average number of 4.1 alleles. About 22 alleles (20%) were specific for one accession (specific alleles) 13 accessions contained specific alleles and at the maximum, three specific alleles were detected in 176RV from the Rhine valley (Table 3). The average $F_{ST}$ among the accessions was 0.36 (0.32–0.41 95% CI), indicating a low correlation of allele frequencies among different accessions. $F_{IS}$ was 0.60 (0.44–0.74 95% CI) on average signalizing moderately high correlation of the alleles within accessions.

The gene diversity among the landraces, based on 10 sampled plants, ranged from 0.25 (425TM) to 0.57 (007PR), with an average of 0.44. 025VS, with four times more sampled individuals, showed the greatest gene diversity (0.60, Table 3). The average observed heterozygosity was 0.20, with values ranging from 0.14 (179RV, 125PV and 023VS) to 0.37 (134PR). The inbreeding coefficient (f), ranged from 0.41 (425TM) to 0.70 (059LV and 179RV) with an average of 0.58 (Table 3). If the whole population was tested, then all loci showed significant deviation from HWE. For the accessions, an average 45.7% (range: 33–67%) of the polymorphic loci showed deviation ($P = 0.01$) from HWE, always with excess homozygosity.

The genetic distance $d_W$ between the different accessions ranged from 0.113 (between 151RV and 174RV) to 0.231 (between 059LV and 425TM) (data not shown). As expected, $d_W$ values of the four repeated samples (025VSa to d) were lower, ranging from 0.065 to 0.089. Thus, in this study, a $d_W$ value of 0.09 or less described identical accessions. The replicated samples of 025VS produced enough statistical weight to form their own cluster in the Ward dendrogram (Fig. 2). 'Magister' together with 425TM were closely related to the German accession (127D), which indicated representation of northern European flint in this cluster. By contrast, most material from the southern valleys of Tessin (TM) and from the Poschiavo valley (PV) was found in the cluster harbouring the Mexican hybrid, with only two exceptions (425TM and 198TM). The remaining material was situated in two different clusters, both harbouring more accessions from the Rhine valley than from anywhere else. One subcluster therein contained maize from the Linth valley and 198TM.

### Profiling of bulks

In the analysis of bulks, 79 different alleles were detected. Comparing the presence of these alleles per accession with the 108 alleles detected in the analysis of the individuals led to 598 comparisons; in 184 (30.8%) of these comparisons alleles were confirmed in the bulks (Fig. 3). These alleles had an average frequency of 0.46 in the analysis of the individuals. Of the alleles with a frequency below 0.2 in the analysis of individuals only a fifth were detected in the bulks, while a third of the alleles with a frequency of 0.2–0.5 and roughly half of the alleles with a frequency of greater than 0.5 were detected in the bulks (Fig. 3). In 025VS treated as two samples of 20 plants the rarest allele had a frequency of 0.02 in the analysis of individuals but the allele with the lowest frequency that was detected in the analysis of bulks had a frequency of 0.24 in the analysis of individuals. About 15% of the bands detected in the analysis of bulks were artefacts, i.e., the bulks showed bands that were not detected in any of the individuals of an accession in 89 of the 598 comparisons. Phi109642 and umc1125 both with four base repeat motifs produced the lowest amount of artefacts, whereas the highest number of artefacts was recorded for phi087

**Table 3** Summary of genetic diversity within 19 Swiss landraces and two hybrids (MAG, CMS) including a comparison of the analysis of individuals (I) and the analysis of bulks (B)

| Analysis of individuals | | | | | | Analysis of individuals vs. analysis of bulks | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| OTU[a] | Major allele frequency | No. of genotypes | Gene diversity | Heterozygosity | $f$[b] | Av. no. of alleles per locus | | No. of specific alleles | |
| | | | | | | Individuals | Bulks | I | B |
| CMS | 0.74 | 2.60 | 0.33 | 0.13 | 0.65 | 2.3 | 1 | 1 | 1 |
| Magister | 0.82 | 1.70 | 0.23 | 0.26 | −0.04 | 1.6 | 1.3 | 0 | 1 |
| 055LV | 0.61 | 3.20 | 0.49 | 0.19 | 0.66 | 3.0 | 1.5 | 1 | 3 |
| 059LV | 0.64 | 3.10 | 0.46 | 0.16 | 0.70 | 3.0 | 1.4 | 2 | 1 |
| 089LV | 0.61 | 3.30 | 0.49 | 0.18 | 0.67 | 2.8 | 1 | 0 | 0 |
| 007PR | 0.54 | 3.70 | 0.57 | 0.33 | 0.46 | 3.6 | 1.4 | 1 | 1 |
| 134PR | 0.54 | 4.30 | 0.55 | 0.37 | 0.39 | 3.6 | 2 | 0 | 2 |
| 125PV | 0.71 | 2.80 | 0.39 | 0.14 | 0.68 | 2.6 | 0.9 | 2 | 0 |
| 018RV | 0.74 | 2.70 | 0.34 | 0.17 | 0.55 | 2.3 | 0.8 | 1 | 0 |
| 103RV | 0.65 | 3.40 | 0.45 | 0.17 | 0.66 | 3.0 | 0.8 | 1 | 0 |
| 151RV | 0.51 | 3.70 | 0.54[c] | 0.25[c] | 0.59 | 3.2 | 1.3 | 2 | 2 |
| 174RV | 0.54 | 3.30 | 0.53 | 0.31 | 0.46 | 3.4 | 1.4 | 0 | 0 |
| 176RV | 0.62 | 3.40 | 0.46 | 0.18 | 0.65 | 3.2 | 1.4 | 3 | 0 |
| 179RV | 0.68 | 3.20 | 0.42 | 0.14 | 0.70 | 2.7 | 1.3 | 1 | 0 |
| 023VS | 0.70 | 3.10 | 0.39 | 0.14 | 0.67 | 2.5 | 1.3 | 2 | 0 |
| 073TM | 0.61 | 3.20 | 0.48 | 0.19 | 0.64 | 2.9 | 1.4 | 1 | 1 |
| 075TM | 0.74 | 3.00 | 0.36 | 0.19 | 0.50 | 2.7 | 1.4 | 0 | 3 |
| 198TM | 0.69 | 2.70 | 0.42 | 0.20 | 0.57 | 2.7 | 1.7 | 2 | 0 |
| 425TM | 0.72 | 2.10 | 0.25[c] | 0.16[c] | 0.41 | 1.9 | 1.3 | 0 | 1 |
| 127D | 0.67 | 3.30 | 0.42 | 0.23 | 0.51 | 2.7 | 1.6 | 2 | 1 |
| Mean[d] | 0.64 | 3.19 | 0.44 | 0.20 | 0.58 | 2.9 | 1.3 | 1.2 | 0.8 |
| 025VS | 0.52 | 5.00 | 0.60 | 0.20 | 0.67 | 4.10 | 1.5 | 5 | 2 |

Availability of data for individuals per accession ranged from 0.79 (151RV) to 0.98 (023VS) with a mean of 0.89. Ten individuals per accession were sampled; only 025VS was represented by 40 individuals

[a] See Table 1 for a description of the operational taxonomic unit (OTU)

[b] Inbreeding coefficient

[c] Based on nine loci due to missing data

[d] Mean of Swiss accessions only

(three bases) and bmc1337 (two bases, Table 2). On average 77% of these artefacts were caused by wrongly placing an allele in the adjacent allele class.

For three accessions both sampling methods showed non-amplification. In 28 cases (5%) in the analysis of bulks the amplification failed for an accession that showed amplification products in the analysis of individuals. The Mantel test, calculated on the basis of the genetic distance matrices of the analysis of the individuals and the analysis of the bulks, resulted in a positive correlation of 0.25 ($P = 0.007$).

## Discussion

### Methodological considerations

To assess the suitability of our approach for profiling large numbers of accessions, we investigated a set of 19 accessions with 10 plants per accession. The PCR products of the SSR loci selected for this analysis did not show signs of stuttering in any genotype on the described agarose gels probably due to the resolution achieved. 025VS, sampled with 40 individuals, revealed an increase in the number of genotypes,
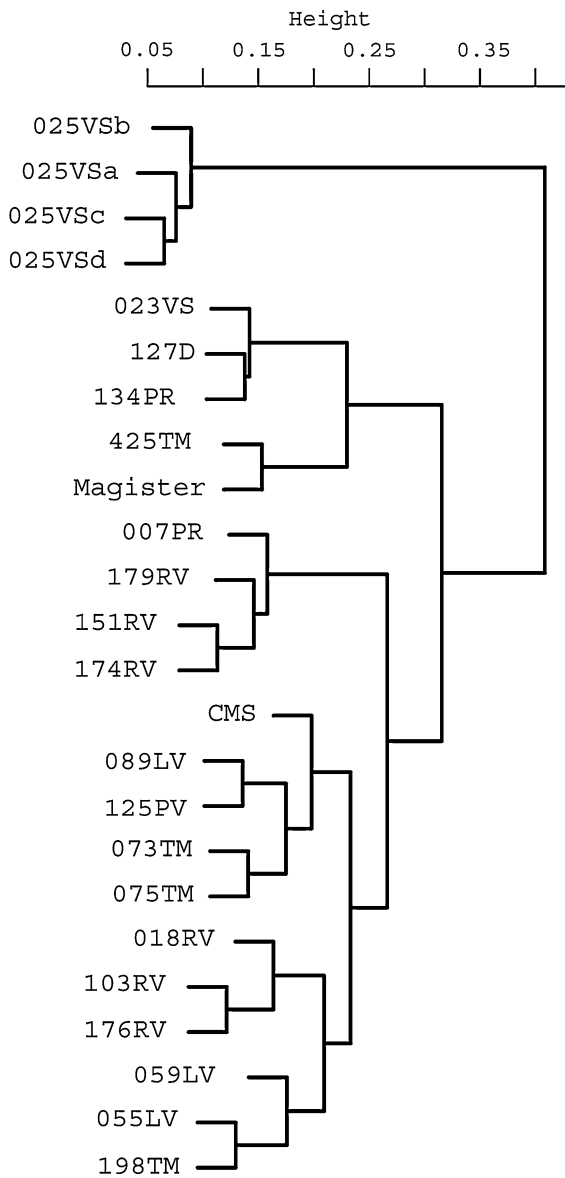
Fig. 2 Relationship among the evaluated accessions as established by hierarchical cluster analysis (Ward) based on the pairwise modified Rogers' distances from the analysis of individuals. 025VSa-d are four repeated samples of 10 plants from the same accession
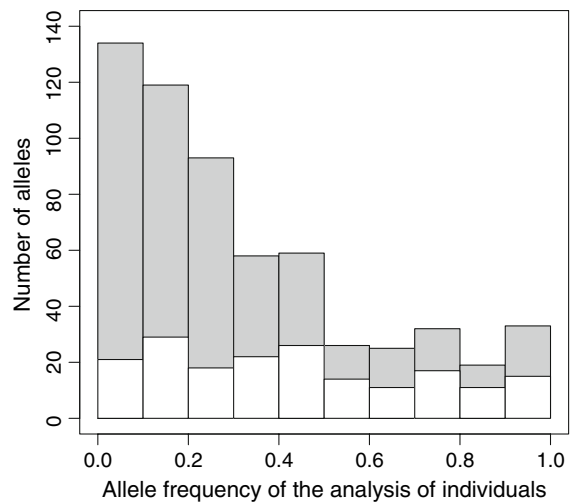


Fig. 3 Amount of detected alleles per allele frequency class for the single-plant analysis and the portion of alleles not detected in the bulks (grey bars)

gene diversity, average number of alleles per locus, as well as in the number of specific alleles above the range of the accessions sampled with 10 individuals as could be expected. The frequency of the major alleles and the inbreeding coefficient of 025VS indeed were within the range of the other accessions. The appropriate number of plants selected for fingerprinting is critical to the detection of rare

alleles. However, the detection of rare alleles is not necessarily the most important aim of a genetic characterization. Based on the literature, 10 plants per accession are sufficient for fingerprinting a large and reasonably diverse set of landraces (Matsuoka et al. 2002; Ho et al. 2005). In maize populations, a variation in the number of individuals per sample (5–23) did not result in significantly different data for the genetic distance or for the average number of alleles per locus and accession (Labate et al. 2003). Using eight plants per accession, Ho et al. (2005) achieved the same result as Reif et al. (2005) i.e., the effect of the most common alleles was sufficient to differentiate between maize populations. Our results seemed to support this finding. One reason for this might be that larger samples do not contribute additional information since additional alleles are more likely to represent deleterious genes with a low frequency and which contribute little to the genetic variation (Nei 1975).

Apart from the number of individuals the number of probed loci is crucial for a reliable estimation of the genetic distances among accessions. Based on publications in the field of molecular ecology it was concluded that there is large benefit when 10 SSRs instead of five are used to calculate genetic distance. A further increase in the number of loci, however, did not have strong effects on the standard deviation of pair wise genetic distances (Koskinen et al. 2004). A simulation study of the number of alleles required to

estimate genetic distances led to a similar conclusion, that a sufficient number of alleles can be generated by using few loci with many alleles (Kalinowski 2002). Because the results of population differentiation tend to depend more on the set of microsatellites than on decreases in the number of loci (Koskinen et al. 2004), we did not select microsatellite loci with only one class of repeat elements. Instead we included dinucleotide repeat loci was well as higher and complex repeats (Table 2). Our selection of markers was based on a screening process similar to what was described by Lu and Bernardo (2001).

To account for the comparably low precision of agarose gels for band separation a binning procedure was applied as described in an NRC report on forensic genetics (1992). This strategy appeared still useful as it was discussed recently for automated genotyping (Jayashree et al. 2006; Amos et al. 2007). By inclusion of four independent samples of the same accession we had control over the calculation of genetic distance based on the binned alleles. The genetic distance between the four repetitions was much smaller than any other pair wise genetic distance in this study. Therefore, it appeared that no pair of accessions could be considered to be identical or to be derived from the same ancestral population a short time ago. Consequently binning and sampling had a weak effect on the genetic distance values and the accessions were all considered as different.

To control the process of determining allele classes, six individuals of eight flint by flint single cross hybrids of a Swiss maize breeding program (Delley Semences et Plantes SA, Delley, Switzerland) were genotyped in the same way as the landrace accessions (data not shown). The average number of genotypes detected per hybrid was 1.18, i.e., close to the expected hybrid uniformity. As the artefacts present in the bulks were often caused by misplacing alleles, the binning process probably should have been even more conservative. Also errors likely could have been prevented by using SSRs with four base repeat motifs only.

Analysis of genetic diversity

*Profiling of single plants*

The specific alleles (20%) we found are comparable to the amount of specific alleles (7%) found in four maize landrace accessions from Mexico and Venezuela as well as those (20%) found in wild subspecies of maize (Matsuoka et al. 2002). Our value was high because a diverse set of Swiss maize was chosen, which also explains the detected fixation effects. An overall $F_{ST}(\theta)$ of 0.36 indicates a high level of partitioning of allelic variation due to the restricted gene flow between the different accessions (or subpopulations); since a $\theta$ value of 0 indicates panmixis and a value of 1 indicates complete isolation of the different accessions. Labate et al. (2003) reported an $F_{ST}$ value of 0.15 for Corn Belt Dent in contrast to 0.35 found by González Ugalde (1997) for American Northern Flint. The Swiss accessions also fall into the flint category, and their $F_{ST}$ is comparable to that of American Northern Flint. The average gene diversity of 0.44 (Table 3) for the Swiss landrace accessions is in accordance with studies of accessions of Central and North American flint and dent accessions and is close to typical values for American Northern Flint (Matsuoka et al. 2002; Ho et al. 2005; Reif et al. 2006).

The high $f$ values are surprising since effects of inbreeding should be excluded because the material was randomly mated (Hardy 1908; Weinberg 1908; Stern 1943). Significant influences of genetic origin, such as homogamy, artificial subgrouping of individuals and selection favouring homozygotes during directed pollination, were not expected. However, studies by Reif et al. (2005) and Labate et al. (2003) as well as studies not based on SSRs also reported excessive homozygosity. In our study, the following possible errors made in the laboratory may have led to the overestimation of homozygotes: (1) the threshold for the band intensity may have been too strict to enable detection of weak alleles, (2) heterozygous loci carrying a null allele were scored as homozygous and (3) insufficient resolution prevented small differences in allele size from detection. When using agarose gel electrophoresis, the signal caused by a PCR product may be hidden by weak fluorescence or contamination of the gel. This may make it difficult to detect bands of low intensity (1). The presence of an average of 4.9% null alleles per locus may have led to a tendency towards homozygosity excess, because hybrids of amplifying alleles and null alleles would be scored as homozygotes (Reif et al. 2005). However, in the presence of fewer than 5% null alleles per locus this effect was not found when

maize landraces, teosinte and 101 inbred lines were compared (Matsuoka et al. 2002), this source of error should therefore be negligible (2). Two alleles with size differences of less than three base pairs within one plant may have led to the detection of only one bright band (3) causing artificially decreased heterozygosity and a reduced number of alleles per accession. Additionally preferential amplification during PCR may have prevented alleles from detection, although we tried to avoid this by adjusting PCR conditions to each marker individually. In comparable studies the average number of alleles per accession of flint landraces ranged from 2.60 (Labate et al. 2003) to 5.9 (Reif et al. 2005).

The low genetic distances between the 025VS subsamples indicated that a threshold of $d_W$ defined on the basis of repeated samples may be suitable for identifying doubles. The average genetic distance for Swiss material (mean: 0.17) is small compared to material from Mexico (mean: 0.35) and Germany (mean: 0.39) (Reif et al. 2005, 2006). Similar to our findings, genetic differences between open-pollinated populations of maize tend to be rather small, while genetic diversity within populations was found to be high (Warburton et al. 2002; Reif et al. 2004).

The fact that 425TM was placed next to 'Magister' indicated the possibility of contamination of material preserved in situ with alleles of modern hybrids; this was supported by early flowering of 425TM (data not shown). However, 425TM may also have been derived from maize material from Italy, for example, which shares alleles with the parents of 'Magister'. It is assumed that 'Magister' consists in part of breeding material based on Northern European flint because of its excellent early vigour; therefore its proximity to the German accession is not surprising. Southern accessions conserved in the gene bank usually clustered together with CMS from the Mexican highlands, which supports the hypothesis of a relationship between Southern Swiss material and maize from Italy (probably Italian Orange Flint) that was very similar to maize from Central and South America (Rebourg et al. 2003; Dubreuil et al. 2006). One cluster consisted of Swiss accessions only, most of the material coming from the cooler northern Alps. The presence of material from the Valais in the same cluster as the northern European accessions made sense because the Valais is considered as part of the central Alps and geographically clearly distinct from the southern part of the country. The Rhine valley accessions present in different clusters indicated the possible existence of different types of Rhine valley maize. It remains to be determined why accessions from the Linth valley clustered together with accessions from south of the Alps, but generally, there is a clear structure within the Swiss landrace population that is probably partly due to the geography of the region from which the accessions were collected.

### Profiling of bulks

The bulk DNA approach was reported to be a relatively inexpensive strategy for large-scale grouping (Dubreuil et al. 1999), and its application to SSRs was, therefore, investigated (Reif et al. 2005; Beyene et al. 2006a). The applicability of the bulk approach was tested further and optimized when applied to a set of European and American accessions (Dubreuil et al. 2006). It was concluded that the genotyping of a single bulk of 15 plants per accession on polyacrylamide gels was sufficiently accurate. Instead of using the established polyacrylamide gel electrophoresis (PAGE) system, we applied an established, visualization method (Senior et al. 1998) with lower technical demands. Compared to Beyene et al. (2006a) who had a similar approach, our use of a 20 Bp ladder as size standard was meant to increase precision of band size estimation. The decision in favour of this visualization method was based on the finding that most maize SSR loci are highly polymorphic (Smith et al. 1997) and often do not fit the stepwise mutation model (Matsuoka et al. 2002), which led to the assumption that differences in allele size often are sufficiently large to be displayed on agarose. This assumption was tested by first screening loci for detectable polymorphism in our system of choice using a randomly selected accession. In a pilot study (Eschholz et al. 2006) the bulk approach was tested on a large set of accessions from Switzerland. There, valuable hints were gained on the repeatability of a bulk approach that were followed up in detail here.

The number of artefacts caused by bulking of DNA may have influenced the results of the bulk analysis. It seems to be more important, however, that the bulk analysis often did not detect alleles. A limited number of alleles were detected with a frequency below 0.2.

This indicated that the alleles present with high frequency in the analysis of individuals, were more likely to be detected by bulk analysis. Furthermore, the detection of frequent alleles in bulks on agarose gels is lower than was reported for bulk analysis with polyacrylamide gel electrophoresis (Reif et al. 2005). However, the Mantel test calculated on the genetic distance matrices based on individuals and bulks resulted in a significant positive correlation as reported before (Reif et al. 2005). Thus it can be concluded that the two distance matrices were not independent of one another. Based on the amount of non-detection of alleles per accession the detection of alleles in the analysis of individuals was more reliable than the analysis of the bulks on agarose gels. But bulks detected a considerable amount of the more frequent alleles.

## Conclusions

Based on the analysis of single plants with a set of 10 selected SSR markers visualized on agarose gel, the diverse Swiss accessions were grouped according to origin; at the same time subsamples of a randomly selected accession were identified. Thus, with this approach, differentiation was sufficient for the profiling of a larger set of accessions using single-plant analyses. The amount of information on the frequency and distribution of alleles was higher in the analysis of individual plants. The use of DNA bulks with SSR products visualized on agarose gels generally appeared feasible. We recommend using SSRs with repeat motifs of four bases or more to increase precision of the analysis of bulks.

## References

Amos W, Hoffman JI, Frodsham A, Zhang L, Best S, Hill AVS (2007) Automated binning of microsatellite alleles: problems and solutions. Mol Ecol Notes 7:10–14

Beyene Y, Botha A-M, Myburg AA (2006a) Genetic diversity among traditional Ethiopian highland maize accessions assessed by simple sequence repeat (SSR) markers. Genet Resour Crop Evol 53:1579–1588

Beyene Y, Botha A-M, Myburg AA (2006b) Genetic diversity in traditional Ethiopian highland maize accessions assessed by AFLP markers and morphological traits. Biodivers Conserv 15:2655–2671

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic-linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

Brandolini A (1970) Razze Europee di Mais. Maydica 15:5–27

Camus-Kulandaivelu L, Veyrieras J-B, Madur D, Combes V, Fourmann M, Barraud S, Dubreuil P, Gouesnard B, Manicacci D, Charcosset A (2006) Maize adaptation to temperate climate: relationship between population structure and polymorphism in the Dwarf8 gene. Genetics 172:2449–2463

Carvalho VP, Ruas CF, Ferreira JM, Moreira RMP, Ruas PM (2004) Genetic diversity among maize (Zea mays L.) landraces assessed by RAPD markers. Genet Mol Biol 27:228–236

Council NR (1992) DNA typing: statistical bases for interpretation. DNA technology in forensic science. National Academy Press, Washington, pp 74–96

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

Dubreuil P, Warburton M, Chastanet M, Hoisington D, Charcosset A (2006) More on the introduction of temperate maize into Europe: large-scale bulk SSR genotyping and new historical elements. Maydica 51:281–291

Eschholz TW, Peter R, Stamp P, Hund A (2006) Swiss maize landraces – their diversity and genetic relationships. Acta Agronomica Hungarica 54:321–328

González Ugalde WG (1997) Genetic characterisation of Northern Flints and Flours maize (Zea mays L. spp. mays) with isozyme, SSR and morphological markers. Diss Abstr Int 58(11B):5706, Iowa State University, Ames

Guo SW, Thompson EA (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

Ho JC, Kresovich S, Lamkey KR (2005) Extent and distribution of genetic variation in US maize: historically important lines and their open-pollinated Dent and Flint progenitors. Crop Sci 45:1891–1900

Jayashree B, Reddy P, Leeladevi Y, Crouch J, Mahalakshmi V, Buhariwalla H, Eshwar KE, Mace E, Folksterma R, Senthilvel S, Varshney R, Seetha K, Rajalakshmi R, Prasanth VP, Chandra S, Swarupa L, SriKalyani P, Hoisington D (2006) Laboratory information management software for genotyping workflows: applications in high throughput crop genotyping. BMC Bioinformatics 7:383

Kalinowski ST (2002) How many alleles per locus should be used to estimate genetic distances? Heredity 88:62–65

Koskinen MT, Hirvonen H, Landry P-A, Primmer CR (2004) The benefits of increasing the number of microsatellites utilized in genetic population studies: an empirical perspective. Hereditas 141:61–67

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

Laborda P, Oliveira K, Garcia A, Paterniani M, de Souza A (2005) Tropical maize germplasm: what can we say about its genetic diversity in the light of molecular markers? Theor Appl Genet 111:1288–1299

Lawrence CJ, Dong Q, Polacco ML, Seigfried TE, Brendel V (2004) MaizeGDB, the community database for maize genetics and genomics. Nucleic Acids Res 32:393–397

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

Lu H, Bernardo R (2001) Molecular marker diversity among current and historical maize inbreds. TAG Theor Appl Genet 103:613–617

Mantel N (1967) Detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220

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

Nei M (1975) Molecular population genetics and evolution. North-Holland Publishing Company, Amsterdam

Peter R (2007) Early vigour of Swiss maize landraces (*Zea mays* L.) in cool environments – agronomic performance and root characteristics. ETH, Zurich

Peter R, Eschholz TW, Stamp P, Liedgens M (2006) Swiss maize landraces – early vigour adaptation to cool conditions. Acta Agronomica Hungarica 54:329–336

R development core team (2004) R: a language and environment for statistical computing. R foundation for statistical computing, Vienna

Rebourg C, Chastanet M, Gouesnard B, Welcker C, Dubreuil P, 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, Peter Maurer H, Bohn M, Melchinger AE (2005) Genetic structure and diversity of European flint maize populations determined with SSR analyses of individuals and bulks. Theor Appl Genet 111:906–913

Reif J, Warburton ML, Xia XC, Hoisington DA, Crossa J, Taba S, Muminovic J, Bohn M, Frisch M, Melchinger AE (2006) Grouping of accessions of Mexican races of maize revisited with SSR markers. Theor Appl Genet 113:177–185

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, New York

Senior ML, Heun M (1993) Mapping maize microsatellites and polymerase chain reaction confirmation of the targeted repeats using a CT primer. Genome 36:884–889

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

Sharopova N, McMullen MD, Schultz L, Schroeder S, Sanchez-Villeda H, Gardiner J, Bergstrom D, Houchins K, Melia-Hancock S, Musket T, Duru N, Polacco M, Edwards K, Ruff T, Register JC, Brouwer C, Thompson R, Velasco R, Chin E, Lee M, Woodman-Clikeman W, Long MJ, Liscum E, Cone K, Davis G, Coe EH (2002) Development and mapping of SSR markers for maize. Plant Mol Biol 48:463–481

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

Stern C (1943) The Hardy-Weinberg law. Science 97:137–138

Taller JM, Bernardo R (2004) Diverse adapted populations for improving northern maize inbreds. Crop Sci 44:1444–1449

Taramino G, Tingey S (1996) Simple sequence repeats for germplasm analysis and mapping in maize. Genome 39:277–287

Van Hintum TJ, Van Treuren R (2002) Molecular markers: tools to improve genebank efficiency. Cell Mol Biol Lett 7:737–744

Vigouroux Y, Mitchell S, Matsuoka Y, Hamblin M, Kresovich S, Smith JSC, Jaqueth J, Smith OS, Doebley J (2005) An analysis of genetic diversity across the maize genome using microsatellites. Genetics 169:1617–1630

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

Ward JH Jr (1963) Hierarchical grouping to optimize an objective function. J Am Stat Assoc 58:236–244

Weinberg W (1908) Über den Nachweis der Vererbung beim Menschen. Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg 64:368–382

Weir BS (1996) Genetic data analysis II. Sinauer, Sunderland

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370

Weir BS, Hill WG (2002) Estimating F-statistics. Annu Rev Genet 36:721–750

Wright S (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution 19:395–420

Wright S (1978) Evolution and the genetics of populations. Vol 4. Variability within and among natural populations. University of Chicago, Chicago, 580 pp

Zeven AC (1998) Landraces: a review of definitions and classifications. Euphytica 104:127–139