Molecular Breeding **10:** 181–191, 2002.

Variation of DNA fingerprints among accessions within maize inbred lines and implications for identification of essentially derived varieties. I. Genetic and technical sources of variation in SSR data

Martin Heckenberger[1], Martin Bohn[1,*], Janet S. Ziegle[2], Larry K. Joe[2], Joan D. Hauser[2], Michelle Hutton[2] and Albrecht E. Melchinger[1]

*[1]Institute of Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim, 70593 Stuttgart, Germany; [2]Celera AgGen, 1756 Picasso Avenue, Davis, CA 95616, USA; *Author for correspondence (e-mail: mbohn@uiuc.edu; fax: ++49 711 459 2343)*

Received 11 October 2001; accepted in revised form 19 March 2002

Key words: Essentially derived varieties, Genetic distances, Intra-varietal variation, Reproducibility, SSRs

Abstract

Genetic distances (GDs) based on molecular markers are important parameters for identifying essentially derived varieties (EDVs). In this context information about the variability of molecular markers within maize inbred lines is essential. Our objectives were to (1) determine the variation in the size of simple sequence repeat (SSR) fragments among different accessions of maize inbreds and doubled haploid (DH) lines, (2) attribute the observed variation to genetic and marker system-specific sources, and (3) investigate the effect of SSR fragment size differences within maize lines on the GD between maize lines and their consequences for the identification of essentially derived varieties. Two to five accessions from nine inbred lines and five DH lines were taken from different sources or drawn as independent samples from the same seed lot. Each accession was genotyped with 100 SSR markers that evenly covered the whole maize genome. In total, 437 SSR fragments were identified, with a mean of 4.4 alleles per locus. The average polymorphic information content (PIC) was 0.58. GD estimates between two accessions of the same genotype ranged from 0.00 to 0.12 with an average of 0.029 for inbred lines and 0.001 for DH lines. An average of 11.1 SSRs was polymorphic between accessions of the same inbred line due to non-amplification (8.1 SSRs), heterogeneity (4.0 SSRs) or unknown alleles (2.6 SSRs). In contrast to lab errors, heterogeneity contributed considerably to the observed variation for GD. In order to decrease the probability to be suited for infringing an EDV threshold by chance, we recommend to increase the level of homogeneity of inbred lines before applying for plant variety protection.

Introduction

According to the International Union for the Protection of new Varieties of Plants (UPOV) convention, a variety is deemed to be essentially derived from an initial variety if it is (i) predominantly derived and (ii) clearly distinguishable from the initial variety and (iii) genetically conform to the initial variety (UPOV 1991). The genetic conformity between initial and essentially derived varieties is considered to be the main important element within the concept of essentially derived varieties (EDV) (ASSINSEL 1999). Therefore, this concept should be based on genotypic information and the genetic distance (GD) between varieties is one of the key parameters to distinguish between essentially derived and independent varieties.

Genetic distances based on molecular marker data proved to be adequate estimates for the pedigree relationships in all major crops. Especially in maize, numerous studies yielded significant correlations between GDs obtained by molecular markers and the coefficient of coancestry (Lübberstedt et al. 2000; Smith et al. 1997). For this reason, molecular markers, particularly amplified fragment length polymorphisms (AFLPs) and simple sequence repeats (SSRs),

were proposed to be an appropriate tool to verify essential derivation in plant varieties (Smith et al. 1991; Knaak et al. 1996; ASSINSEL 2000).

Plant breeders did not yet implement the EDV concept into their breeding procedures due to the lack of suitable crop specific thresholds. EDV thresholds have to be crop specific to take into account the varying degree of polymorphism among the different crop species. For example, the proportion of alleles at marker loci alike in state between unrelated varieties is much smaller in maize than in barley or tomato (Qi and Lindhout 1997; Bernardo et al. 1997; Grandillo et al. 1999).

Because of the legal consequences scientifically reliable criteria have to be developed for the discrimination of essentially derived and independent varieties. Therefore, highest accuracy and reproducibility of GD estimates are mandatory. Potential reproducibility problems were investigated by Jones et al. (1997), who reported scoring differences of up to 2 bp among SSR fragments. In addition, information on the stability of molecular marker data over several generations of maintenance breeding is scanty.

The objectives of this study were to

- determine the variation in the size of SSR fragments among different accessions of maize inbreds and doubled haploid (DH) lines,
- attribute the observed variation to genetic and marker system-specific sources, and
- investigate the effect of SSR fragment size differences within maize lines on the GD between maize lines and their consequences for the identification of essentially derived varieties.

Material & methods

Plant materials

For nine maize inbred lines, six from the flint and three from the dent pool, and five DH lines from the dent pool, two to five accessions per line were fingerprinted (Table 1). Accessions were obtained from different generations of maintenance breeding conducted by the University of Hohenheim (UHOH) and three commercial breeding companies (B1–B3). Two accessions per DH line were obtained by drawing two independent samples of 20 kernels out of the same seed lot. The DH lines were derived from the cross of inbred lines RG2302 and 69117 by *in-vivo*-haploid induction (Deimling et al. 1997) with a subsequent colchicine treatment for chromosome doubling.

Maintenance breeding at the UHOH was performed by ear to row selection starting with the selfing of a single S_5 or S_6 ear per inbred line (Simmonds and Smartt 1999). In the further course of this procedure, one ear per row was selected per generation for maintaining the inbred line, whereas the other ears of the same homogeneous row were bulked for seed production. Thus, all individuals of one inbred line traced back to a single S_5 or S_6 ear.

All accessions of inbred lines were phenotypically homogeneous in field observation trials according to the regulations of the German Plant Variety Office and showed at least 95% of identical bands in an analysis of storage proteins conducted by isoelectric focusing following the rules of the International Seed Testing Association. Phenotypic evaluation was done at the UHOH experimental station at Eckartsweier, Germany, and the storage protein analyses were performed by the national agricultural research institute (LUFA) at Augustenberg, Germany, in 1999.

SSR analyses

DNA fingerprinting was performed with a standard marker set of 100 publicly available SSR markers that provide an even coverage of the maize genome (Figure 1). Fifteen seeds per accession were planted in a single pot for DNA extraction. Equal quantities of leaf material of 10 plants per accession were harvested at the 3 to 4 leaf stage. The leaf material was mixed and DNA was extracted using a modified Hexadecyltrimethyl-ammonium bromide (CTAB) procedure (Saghai Maroof et al. 1984). Electrophoresis was performed with an ABI Prism™ 377 DNA Sequencer using 5% polyacrylamid gels with 96 lanes. Internal fragment size standards were used in each lane to increase accuracy of DNA fragment size determination. The size of each DNA fragment was determined automatically by using the GeneScan® software and assigned to specific alleles by the Genotyper® software.

The 100 SSRs were selected based on robust single-locus amplification, absence of null alleles, high degree of polymorphism, and high reproducibility of the results. Seventy of the 100 SSRs contained dinucleotide repeat motifs, whereas the other 30 markers consisted of tri- to octa-nucleotide repeats. The development of the SSR set and the SSR analyses

Table 1. List of genotypes fingerprinted with 100 SSRs.

			Accessions		
Line	Type[†]	Pool	No.	Type[‡]	Source
D146	IL	Flint	3	1991, 1998, B3	bulk
D149	IL	Flint	4	1994, 1998, B1, B3	bulk
D171	IL	Flint	4	1994, 1998, B1, B3	bulk
D503	IL	Flint	2	1991, 1998	bulk
DK105	IL	Flint	4	1970 (1988)[§], 1980 (1988), 1991, 1996	bulk
UH002	IL	Flint	3	S_6, $S_{6:11}$, $S^{\#}_{6:11}$	ear
D06	IL	Dent	5	1988, 1994, 1998, B1, B2	bulk
UH200	IL	Dent	3	S_5, $S_{5:10}$, $S_{5:9}$	ear
UH300	IL	Dent	3	S_5, $S_{5:9}$, $S_{5:9}$	ear
RG2302[¶]	IL	Dent	1		
69117	IL	Dent	1		
941118[††]	HY	Dent	1		
ZS264	DH	Dent	2	2 repetitions[‡‡]	ear
ZS265	DH	Dent	2	2 repetitions	ear
ZS337	DH	Dent	2	2 repetitions	ear
ZS467	DH	Dent	2	2 repetitions	ear
ZS595	DH	Dent	2	2 repetitions	ear

[†] Line derivation: IL, inbred line; DH, doubled haploid line; HY, F_1 hybrid.
[‡] Accession types: year, year of seed maintenance performed by the UHOH maize program; B1–B3, line-maintenance performed by commercial breeders B1, B2, and B3.
[§] Maintenance breeding performed in 1970 and 1980 combined with one selfing generation in 1988 to maintain seed viability.
[#] Seeds of one S_5 or S_6 plant and two different S_9, S_{10}, or S_{11} plants derived from the particular S_5 or S_6 plant were used.
[¶] Parental lines of F_1 hybrid 941118.
[††] Parental source of all ZS lines.
[‡‡] Two independent samples from the second selfing generation of the same seed lot were fingerprinted.

were performed by Celera (1756 Picasso Avenue, Davis, CA 95616, USA).

Distinction of different cases of SSR results for various causes of variation

In order to distinguish different causes of variation for SSR results, the following distinction of disjunctive cases for each possible pairwise comparison of accessions of the same inbred or DH line was developed (Table 2).

Case 1 = The two accessions of a given line are homogeneous for the same allele A. This case is denoted as the normal case.

Case 2 = Both accessions are heterogeneous for the same alleles A and B.

Case 3 = Both accessions have the same allele A and one accession has an additional allele B.

Case 4 = One accession is homogeneous for allele A and one accession is heterogeneous carrying alleles B and C.

Case 5 = Both accessions are heterogeneous with only allele A in common.

Case 6 = Both accessions are heterogeneous with no allele in common.

Case 7 = One accession is heterogeneous and a missing value or null allele M occurs in the other accession.

Case 8 = One accession is homogeneous and a missing value or null allele M occurs in the other accession.

Case 9 = A missing value or null allele M occurs in both accessions.

Case 10 = The two accessions are homogeneous for different alleles.

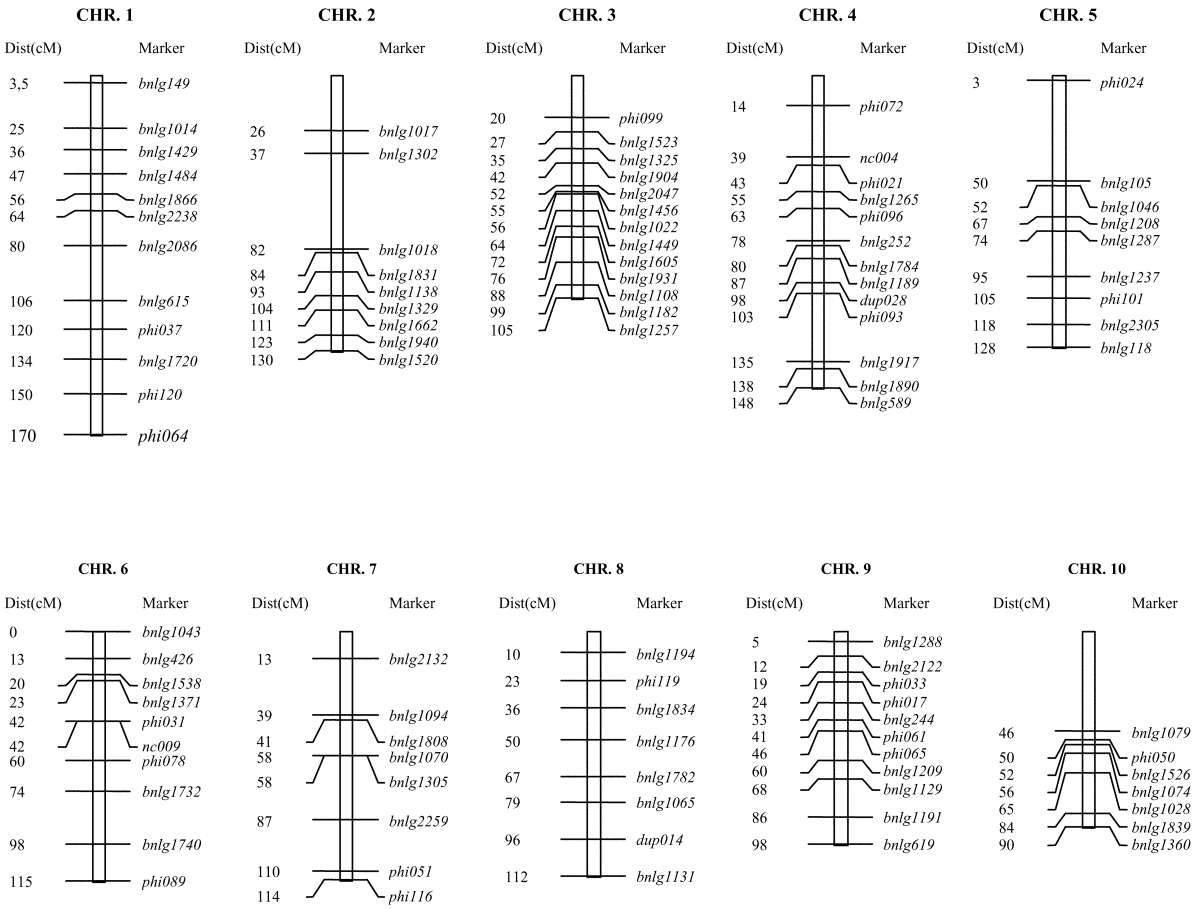

Figure 1. SSR markers used in the present study and their map positions.

Table 2. Distinction of different cases of SSR results for various causes of variation. Letters A, B, C, and D represent different alleles for a given marker locus, M represents a missing value. Accessions 1 and 2 denote two accessions of the same inbred or DH line.

	Case†									
Accession	1	2	3	4	5	6	7	8	9	10
1	A	A+B	A	A	A+B	A+B	A+B	A	M	A
2	A	A+B	A+B	B+C	A+C	C+D	M	M	M	B

† For a detailed description of the cases see Materials and Methods.

Statistical analyses

The polymorphic information content (PIC) was calculated for each SSR marker according to the formula of Botstein et al. (1980),

$$PIC = 1 - \left(\sum_{i=1}^{n} p_i^2 \right) - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2 p_i^2 p_j^2,$$

where p_i and p_j are the frequencies of the i^{th} and j^{th} allele of a given marker, respectively. Genetic distances were calculated using the Dice coefficient (Nei and Li 1979). In the case of missing values, *i.e.*, one or several primer pairs did not yield an amplification product in one accession, the corresponding alleles of the other accession were not used for GD calculation. The cluster analysis was performed with the distance matrix using the UPGMA method (Nei et al. 1983). The reliability of the cluster was assessed by applying a bootstrap procedure (Efron 1979).

The fit of observed heterogeneity with the values from expected heterozygosity in S_5 or S_6 and S_9, S_{10}, or S_{11} generations was evaluated with a χ^2 test. Heterogeneity was defined as the number of marker loci that were not homozygous. Differences between dinucleotide repeats and SSRs with larger repeat motifs with regard to PIC values, number of alleles per marker, level of heterogeneity, number of fragment

size differences, and number of null alleles were tested with a *t*-test.

PIC values were calculated using the Cervus computer program (Marshall et al. 1998), estimation of GDs and cluster analysis were performed with the NTSYS-PC software package (Rohlf 1989). The bootstrap procedure was carried out with the Winboot computer program (Yap and Nelson 1996). The SAS software package was used for all other statistical calculations (SAS Institute 1988).

Results

Characterization of markers

A total of 437 SSR alleles were identified. The number of alleles per marker varied from 1 to 9 with an average of 4.4 alleles per marker. Only marker *bnlg1605* on chromosome 3 was monomorphic across all accessions. PIC values of polymorphic markers varied from 0.25 to 0.82 with an average of 0.58.

Genetic relationships of accessions within maize lines

The GD between two accessions of the same line ranged between 0.00 and 0.12 for inbred lines and between 0.00 and 0.01 for DH lines. The mean GD between accessions of the same line varied from 0.01 to 0.08 for inbred lines (Table 3). The average of the GD means was 0.03 for inbred lines and 0.00 for DH lines. The dendrograms obtained from UPGMA cluster analyses on the basis of GD estimates resulted in a clear separation of flint and dent inbred lines (Figure 2).

Sources of variation

Case 1: Normal case

For different accessions of a given inbred line, the average number of markers displaying the normal case varied from 78.7 for DK105 with 6 pairwise comparisons to 95.0 for D503 with 1 pairwise comparison (Table 4). Across the sets of DH lines, an average number of 92.0 loci showed the normal case, ranging from 87 for ZS265 to 98 for ZS337. Thirty-five SSR markers showed the normal case across all pairwise comparisons of accessions within all inbred or DH lines.

Table 3. Means, minima, and maxima of genetic distances (GD) between accessions of the same inbred line.

Line	No. of accessions	GD between accessions		
		Mean	Min	Max
D146	3	0.09	0.07	0.12
D149	4	0.03	0.04	0.02
D171	4	0.02	0.01	0.03
D503	2	0.01	0.01	0.01
DK105	4	0.04	0.02	0.06
UH002	3	0.03	0.03	0.04
D06	5	0.01	0.00	0.03
UH200	3	0.03	0.03	0.05
UH300	3	0.03	0.02	0.05
Total/Mean	31	0.029		

Cases 2–7: Heterogeneity

The number of marker loci heterogeneous for at least one accession per inbred or DH line varied from 1 to 13 for inbred lines and from 2 to 4 for DH lines. The observed level of heterogeneity for S_5 and S_6 lines was not significantly different from the expected heterozygosity levels. For S_9, S_{10}, and S_{11} lines, the observed level of heterogeneity was significantly ($P < 0.05$) higher than expected and was not significantly different from the observed heterogeneity for S_5 and S_6 lines. An average of 3.1 and 0.2 heterogeneous marker loci (Cases 3–7) was found for inbred and DH lines, respectively (Table 4). An average of 1.0 marker loci for inbred lines and 2.2 for DH lines showed a Case 2 type of heterogeneity with no effect on the variation of GD estimates. Four loci showed three DNA fragments for one accession at a particular locus (included in Cases 3–5).

Accessions of DH lines showed a GD of 0.0 with one exception. One accession of line ZS 264 carried allele 236 at marker locus *bnlg2122* whereas the second accession showed alleles 236 and 254.

Cases 7–9: Occurrence of null alleles or missing values

Forty-one markers showed no amplification product for at least one of all 44 fingerprinted accessions. Across all accessions of one specific inbred line, a mean of 10.3 SSRs yielded no amplification product. Across all DH lines, the mean number of markers without amplification product was 5.6 (data not shown). In addition, several cases of non-amplification in only one of the two repetitions of a DH line were detected.

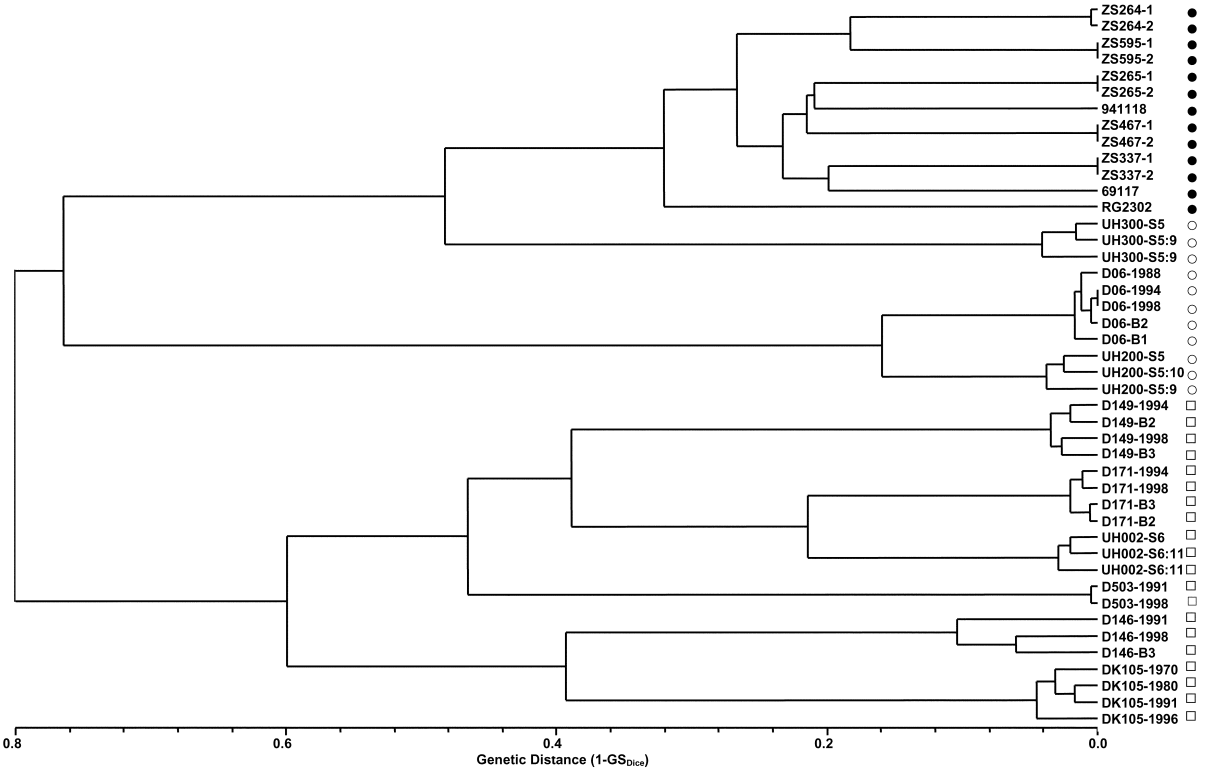

Figure 2. Associations among accessions of maize inbred lines revealed by UPGMA cluster analysis based on genetic distances calculated from SSR data. Asterisks (*) at the forks indicate that the group right of the fork was found for at least 95% of 1000 bootstrap runs. DH lines are marked by filled circles (●). Flint and dent lines are marked with squares (□) and circles (● ○), respectively.

Case 10: New alleles

For inbred lines the number of loci with homozygous unknown alleles averaged 3.9. Homozygous unknown alleles between accessions of the same line were not detected for DH lines (Table 5). The differences between fragment sizes of accessions of the same inbred lines was 1 bp in 14 cases and larger than 1 bp in 25 cases. For six of the 14 differences scored to 1 bp, the exact software-detected difference was 0.4 bp and smaller. The exact difference was 0.9–1.2 bp for the other eight 1 bp differences. The difference between exact and rounded values for allele calling was at 840 out of the 4561 data points 0.6 bp and higher. At 14% of the marker data points the difference was 0.6–0.8 bp, and between 0.8 and 1.2 bp at 4.3% of the data points. Unknown alleles increased the variation of GD within accessions of the same line. The fragment sizes of all accessions of a given line were rounded only in a few cases in the same wrong direction with no effect on the variation of GD estimates within the line. In addition, non-parental bands were detected at six marker loci for DH lines.

Genetic distance between lines

Due to the fact that 2 to 5 accessions per line were fingerprinted, GD values of up to 20 pairwise comparisons of accessions between the same two lines were calculated. The range for GD values calculated for each pairwise combination of accessions between the same two lines varied from 0.00 to 0.07 with an average range of 0.02. The mean range of GD between two inbred lines was significantly higher *($P < 0.01$)* than the mean range of GD between two DH lines.

Comparison of di-nucleotide and longer repeats

Di-repeat SSRs yielded a significantly greater number of alleles per marker ($P < 0.01$) and number of unknown alleles ($P < 0.05$) than SSRs with longer repeat motifs. The observed differences between these two SSRs groups were not significant for PIC values, the number of missing values, and the amount of heterogeneity.

Table 4. Occurrence of different cases for each pairwise comparison of accessions of the same line analyzed in this study.

Line	No. of pairwise comparisons	Case									
		1	2	3	4	5	6	7	8	9	10
		Average no. of loci per pairwise comparison									
Inbred lines											
D146	3	78.7	1.3	5.0	0.7	1.3	0.0	1.0	3.7	4.7	3.7
D149	6	84.7	1.8	4.0	0.0	0.0	0.0	0.3	5.2	2.0	2.0
D171	6	87.3	1.0	0.7	0.0	0.0	0.0	1.3	5.3	3.2	1.2
D503	1	95.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	4.0	0.0
DK105	6	78.7	0.7	1.7	0.0	1.0	0.0	0.5	9.0	6.0	2.5
UH002	3	88.0	1.0	3.7	0.0	0.0	0.0	0.3	2.3	4.0	0.7
D06	10	91.5	1.1	1.0	0.0	0.0	0.0	0.0	1.6	4.0	0.8
UH200	3	90.3	1.0	2.7	0.0	0.0	0.0	0.0	2.0	2.0	2.0
UH300	3	85.3	0.7	1.3	0.0	0.0	0.0	0.0	6.0	4.0	2.7
Mean		**86.6**	**1.0**	**2.3**	**0.1**	**0.3**	**0.0**	**0.4**	**3.9**	**3.8**	**1.7**
DH-lines											
ZS264	1	91.0	3.0	1.0	0.0	0.0	0.0	0.0	0.0	5.0	0.0
ZS265	1	87.0	2.0	0.0	0.0	0.0	0.0	0.0	9.0	2.0	0.0
ZS337	1	98.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ZS467	1	92.0	2.0	0.0	0.0	0.0	0.0	0.0	3.0	3.0	0.0
ZS595	1	92.0	2.0	0.0	0.0	0.0	0.0	0.0	2.0	4.0	0.0
Mean		**92.0**	**2.2**	**0.2**	**0.0**	**0.0**	**0.0**	**0.0**	**2.8**	**2.8**	**0.0**
Total Mean		**88.5**	**1.4**	**1.6**	**0.1**	**0.2**	**0.0**	**0.3**	**3.5**	**3.4**	**1.1**

Table 5. Number and category of differences in fragment size at the same locus within the same line group for Case 10.

Line	No. of accessions	Fragment size differences				Total
		1 bp	2 bp	3 bp	⩾4 bp	
		No. of differences				
D146	3	6	1	1	4	12
D149	4	1	2	0	2	5
D171	4	0	0	1	1	2
D503	2	0	0	0	0	0
DK105	4	5	1	0	2	8
UH002	3	0	1	0	1	2
D06	5	0	1	0	1	2
UH200	3	1	1	0	2	4
UH300	3	1	0	0	3	4
DH-lines	10	0	0	0	0	0
Total	41	14	7	2	16	39

Discussion

The use of flint and dent lines as well as the commonly employed methods of maintenance breeding make our study representative for maize breeding in Europe. In addition, SSR analyses were performed using publicly available SSR primers with a semi-automatic high-throughput system, which reduces human errors and subjectivity to a minimum. This should result in a higher accuracy of GD estimates, a prerequisite for EDV identification.

The degree of polymorphism in our line set was in close agreement with results reported by Smith et al. (1997). PIC values and the average number of alleles per marker were of similar size in both studies. However, in an analysis of genetic diversity among 33 inbred lines from the US corn belt, Pejic et al. (1998) found a substantially higher number of alleles than in the previous studies. The observed discrepancy can be explained by sampling effects caused by different inbred line sets that were fingerprinted with different sets of SSR markers.

Causes of variation

We observed a considerable variation for GD among different accessions of the same inbred line. This variation can be explained by genetic and technical reasons.

Genetic reasons

Mutations within the SSR primer region may yield null alleles, whereas a mutation between the primer regions may result in new alleles. The natural mutation rate for genomic non-repetitive DNA is estimated to range from 10^{-8} to 10^{-6} per locus and generation (Drake et al. 1998; Allard 1999). However, SSRs showed higher mutation rates than non-SSR regions ranging from approximately 10^{-6} per locus and generation for *Saccharomyces cerevisiae* (Sia et al. 2000) up to 10^{-3} in the pipefish *Syngnathus typhle* (Jones et al. 1999). The mutation rate of SSRs was found to be dependent on the repeat type, the repeat number, and the sequence of the repeat motif or the flanking sequence (Schloetterer 2000). Mutations within SSR markers were mostly insertions and deletions of mainly complete repeats (Twerdi et al. 1999). For maize, no information about the mutation rate of SSR loci is yet available. However, if their mutation rate is also higher than for non-SSR regions, mutations cannot be neglected as a cause of genetic variation between accessions of the same line. Unequal crossover in SSR regions is another genetic reason for the unexpected variation in GD as reported in wheat (Plaschke et al. 1995).

Segregation from S_5 and S_6 to the particular S_9, S_{10}, or S_{11} generations was the cause of genetic variation between accessions of the same line at four loci. At these loci, the fingerprinted S_5 or S_6 accession was heterogeneous and the corresponding S_9, S_{10}, or S_{11} accessions were homogeneous, each with one allele of the particular S_5 or S_6 accession. However, for S_9, S_{10}, and S_{11} lines the observed level of heterogeneity was significantly ($P < 0.05$) higher than expected for these selfing generations.

Bulking during maintenance breeding can be one reason for this unexpectedly high level of heterogeneity in highly inbred lines. Ears of each row not used for generating the next generation were bulked for seed production of the particular line. Because of the segregation due to residual heterozygosity, this procedure may have resulted in a mixture of genotypes that were homogeneous for different parental alleles. For samples drawn out of the bulk (Table 1), these effects are not negligible. In addition, genotyping a bulk of ten individuals can lead to variation, when certain regions of the genome still segregate. To avoid this, the individuals should be genotyped independently.

Another possible cause for the variation of GD values within the same line is contamination by foreign pollen during maintenance breeding (Smith et al. 1997). However, all accessions were homogeneous in field observation trials and isoelectric focusing. In addition, deviations from the normal case at numerous loci would be expected in case of a contamination. Therefore, contamination with foreign pollen can be excluded as a cause for the observed high level of heterogeneity in advanced generations.

Technical reasons

In this study, DH lines were used to distinguish between genetic and technical reasons for variation of GD within accessions of the same line. If the fingerprints of two samples of the same DH line are not identical, only technical reasons might explain this finding. The observation of heterozygous DH lines could be explained by heteroduplex bands, especially in those four cases with three alleles at a single locus. It is known that heteroduplex bands emerge by annealing of two DNA fragments of unequal sequence or length (Hatcher et al. 1993). This mismatch usually tends to retard the migration of DNA of the heteroduplex band during electrophoresis (Nataraj et al. 1999). In addition, the intensity of heteroduplex bands is supposed to fall between the corresponding homoduplex bands. However, adopting this definition, only two of the four above mentioned bands, would be heteroduplex.

Artificial stutter bands could also have led to variation in GD estimates within the inbred lines. Especially SSRs with a di-repeat motif are known to show stutter bands caused by a ‘loop’ of 2 bp in the strand of the template (Smith et al. 1997). Therefore, stutter bands appear 2 bp shorter than the main band. The software program “Genotyper” automatically identifies stutter bands based on their migration distance and the intensity of the particular bands. However, intense stutter bands could have been wrongly scored as non-stutter bands and independent bands with low intensity could have been incorrectly identified as stutter bands.

Compared to a study of Murigneux et al. (1993) performed with RFLPs, our results showed an increased level of heterogeneity between DH accessions based on SSRs. This finding can be explained by artificial SSR bands caused by heteroduplex or stutter bands. In addition, mutations due to colchicine treatment used for chromosome doubling and tissue culture steps may also have caused the observed heterogeneity (Marhic et al. 1998).

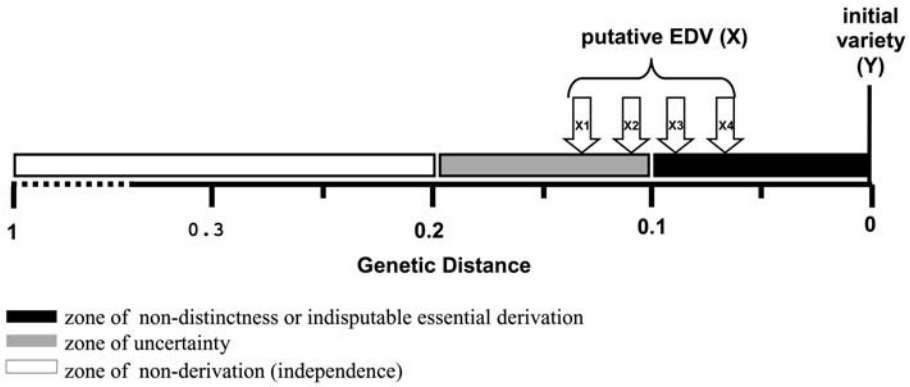

Figure 3. The concept of different thresholds and the consequences of variation between accessions of the same line for the identification of EDVs.

An additional source of variation within an inbred line are unexpected effects due to slippage of the DNA polymerase during the PCR. Therefore, whole or partial repeats could be added or removed from the template and thus yielding genotyping errors (Palsboll et al. 1999; Schloetterer and Tautz 1992).

Software imperfections resulted in a further increase of variation of GD estimates. Genotyping errors of the software Genotyper® caused about half of the 1 bp differences. The software Gene Scan® determined the DNA fragment size using a sizing curve with an accuracy of 0.1 bp. These fragment size measures were employed by the software Genotyper® to assign the fragments to specific alleles. However, reasons for the genotyping errors remain unknown as the algorithm of the software is not publicly available. For those 1 bp differences that could not be assigned to genotyping errors, slippage effects are a possible reason. In addition, small variations in the concentration of the gel, buffer or the voltage of the run could lead to 1 bp differences. To check the repeatability of these 1 bp differences, further studies are required, using several repetitions of single accessions, by sequencing the DNA of the particular fragments, or by genotyping individuals instead of bulks.

Implications for the identification of EDVs

The results of this study demonstrated that lab error and heterogeneity caused variation of GD estimates between different accessions of the same inbred line. In order to assess the implications of the lab error on the identification of EDVs, we first assume the following hypotheses:

H_0: Lines X and Y carry for a specific marker locus the same marker band.

H_1: Lines X and Y carry for a specific marker locus different marker bands.

Based on these hypotheses a Type I error, *i.e.*, two bands were scored as different although they were identical, and a Type II error, *i.e.*, two bands were scored as identical although they were different, can be distinguished. In the case of highly related lines, lab errors result in an overestimation of GD(X, Y) and the hypothesis that X and Y are highly related will be rejected too frequently. However, this study showed that for SSRs employing a semi-automated gel and scoring system, lab errors accounted only for a minor proportion of the detected variation of GD among accessions of the same inbred line and are, therefore, negligible.

The impact of heterogeneity on EDV identification can be exemplified using the following scenario. An individual of line X is used for the development of line Y. A genotypic fingerprint of line Y will be compared with fingerprints of different accessions of line X. Due to the possible varying GD values among accessions of X, GD(X,Y) will also vary conditional on the used accession of X. If the range of GD values between accessions of the same line is large, it might be possible that a true EDV could be judged as independently derived or a truly independent variety as essentially derived just by genotyping different accessions of each variety (Figure 3). In contrast to lab errors that were of negligible importance, heterogeneity reached considerable levels in some cases, *e.g.*, inbred line D146 (Table 3).

Our results illustrate the crucial importance of increased levels of homogeneity within new lines be-

fore applying for plant breeder's rights. If lines are heterogeneous, further selfing generations should be performed. A more rapid procedure would be the prescreening of newly developed lines with molecular markers. Our study showed that the variation of GD estimates within maize lines was lowest among DH lines. If DH techniques can efficiently be applied in a breeding program, this technique might be less laborious and costly than a routine fingerprinting with a marker set covering the entire genome.

Acknowledgements

The authors wish to thank Semundo (Hadmersleben, Germany) for supplying the DH lines.

References

Allard R.W. 1999. Principles of Plant Breeding. John Wiley & Sons, New York.

ASSINSEL 1999. Consolidation of ASSINSEL position papers on protection of biotechnological inventions and plant varieties.

ASSINSEL 2000. DUS testing: Phenotype vs. Genotype. ASSINSEL, Position Paper adopted at the Rome Congress in May 2000.

Bernardo R., Murigneux A., Maisonneuve J.P., Johnsson C. and Karaman Z. 1997. RFLP-based estimates of parental contribution to F_2 and BC_1 derived maize inbreds. Theor. Appl. Genet. 94: 652–656.

Botstein D., White R.L., Skolnick M. and Davis R.W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32: 314–331.

Deimling S., Röber F. and Geiger H.H. 1997. Methodology and genetics of *in vivo* haploid induction in Maize. Vortr. Pflanzenzüchtg. 38: 203–224.

Drake J.W., Charlesworth B., Charlesworth D. and Crow J.F. 1998. Rates of spontaneous mutation. Genetics 148: 1667–1686.

Efron B. 1979. Bootstrap methods: another look at the jackknife. Ann. Stat. 7: 1–26.

Grandillo S., Ku H.M. and Tanksley S.D. 1999. Identifying the loci responsible for natural variation in fruit size and shape in tomato. Theor. Appl. Genet. 99: 978–987.

Hatcher S.L., Lambert Q.T., Raymond L.T. and Carlson J.R. 1993. Heteroduplex formation: a potential source of errors from PCR products. Prenatal Diagnosis 13: 171–177.

Jones A.G., Rosenqvist G., Berglund A. and Avise J.C. 1999. Clustered microsatellite mutations in the pipefish *Syngnathus typhle*. Genetics 152: 1057–1063.

Jones C.J., Edwards K.J., Castaglione S., Winfield M.O., Sala F., Vandewiel C. et al. 1997. Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. Mol. Breed. 3: 381–390.

Knaak C., Förster J. and Jäger-Gussen M. 1996. "Abgeleitete" Sorten aus praktischer Sicht (in German). Bericht über die 47. Arbeitstagung 1996 der Saatzuchtleiter im Rahmen der "Vereinigung österreichischer Pflanzenzüchter" BAL Gumpenstein: 167–172.

Lübberstedt T., Melchinger A.E., Dussle C., Vuylsteke M. and Kuiper M. 2000. Relationships among early European maize inbreds: iv. Genetic diversity revealed with AFLP markers and comparison with RFLP, RAPD, and pedigree data. Crop Sci. 40: 783–791.

Marhic A., Anatoine-Michard S., Bordes J., Pollacsek M., Murigneux A. and Beckert M. 1998. Genetic improvement of anther culture response in maize: relationships with molecular, mendelian and agronomic traits. Theor. Appl. Genet. 97: 520–525.

Marshall T.C., Slate J., Kruuk L.E.B. and Pemberton J.M. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Mol. Ecol. 7: 639–655.

Murigneux A., Barloy D., Leroy P. and Beckert M. 1993. Molecular and morphological evaluation of doubled haploid lines in maize. 1. Homogeneity within DH lines. Theor. Appl. Genet. 86: 837–842.

Nataraj A.J., Olivos-Glander I., Kusukawa N. and Highsmith W.E. 1999. Single-strand conformation polymorphism and heteroduplex analysis for gel-based mutation detection. Electrophoresis 20: 1177–1185.

Nei M. and Li W.H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Soc. USA 76: 5269–5273.

Nei M., Tajima F. and Tateno Y. 1983. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. J. Mol. Evol. 19: 153–170.

Palsboll P.J., Berube M. and Jorgensen H. 1999. Multiple Levels of Single-Strand Slippage at Cetacean Tri- and Tetranucleotide Repeat Microsatellite Loci. Genetics 151: 285–296.

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

Plaschke J., Ganal M.W. and Roder M.S. 1995. Detection of genetic diversity in closely related bread wheat using microsatellite markers. Theor. Appl. Genet. 91: 1001–1007.

Qi X. and Lindhout P. 1997. Development of AFLP markers in barley. Mol. Gen. Genet. 254: 330–336.

Rohlf F.J. 1989. NTSYS-pc Numerical Taxonomy and Multivariate Analysis System. Exeter Publishing Co, Ltd, Setauket, NY.

Saghai Maroof M.A., Soliman K.M., Jorgensen R.A. and Allard R.W. 1984. Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. Proc. Natl. Acad. Soc. USA 81: 8014–8.

SAS Institute 1988. SAS/STAT User's Guide, Release 6.03 edn. SAS, Cary.

Schloetterer C. and Tautz D. 1992. Slippage synthesis of simple sequence DNA. Nucl. Acids Res. 20: 211–215.

Schloetterer C. 2000. Evolutionary dynamics of microsatellite DNA. Chromosoma 109: 365–371.

Sia E.A., Butler C.A., Dominska M., Greenwell P., Fox T.D. and Petes T.D. 2000. Analysis of microsatellite mutations in the mitochondrial DNA of *Saccharomyces cerevisiae*. Proc. Natl. Acad. Soc. USA 97: 250–255.

Simmonds N.W. and Smartt J. 1999. Principles of Crop Improvement. Blackwell Science, Oxford.

Smith J.S.C., Chin E.C.L., Shu H., Smith O.S., Wall S.J., Senior M.L. et al. 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.

Smith J.S.C., Smith O.S., Bowen S.L., Tenborg R.A. and Wall S.J. 1991. The description and assessment of distances between inbred lines of maize. III. A revised scheme for the testing of distinctiveness between inbred lines utilizing DNA RFLPs. Maydica 36: 213–226.

Twerdi C.D., Boyer J.C. and Farber R.A. 1999. Relative rates of insertion and deletion mutations in a microsatellite sequence in cultured cells. Proc. Natl. Acad. Soc. USA 96: 2875–2879.

UPOV 1991. International Convention for the Protection of New Varieties of Plants.

Yap I.V. and Nelson R.J. 1996. WinBoot: A Program for Performing Bootstrap Analysis of Binary Data to Determine Confidence Limits of UPGMA-Based Dendrograms. International Rice Research Institute, Manila, Philippines.