Theor Appl Genet (2005) 111: 723–730
DOI 10.1007/s00122-005-2057-x

ORIGINAL PAPER

**Benjamin Stich · Albrecht E. Melchinger
Matthias Frisch · Hans P. Maurer
Martin Heckenberger · Jochen C. Reif**

Linkage disequilibrium in European elite maize germplasm investigated with SSRs

Received: 3 February 2005 / Accepted: 21 April 2005 / Published online: 5 July 2005

Abstract Information about the extent and genomic distribution of linkage disequilibrium (LD) is of fundamental importance for association mapping. The main objectives of this study were to (1) investigate genetic diversity within germplasm groups of elite European maize (*Zea mays* L.) inbred lines, (2) examine the population structure of elite European maize germplasm, and (3) determine the extent and genomic distribution of LD between pairs of simple sequence repeat (SSR) markers. We examined genetic diversity and LD in a cross section of European and US elite breeding material comprising 147 inbred lines genotyped with 100 SSR markers. For gene diversity within each group, significant ($P < 0.05$) differences existed among the groups. The LD was significant ($P < 0.05$) for 49% of the SSR marker pairs in the 80 flint lines and for 56% of the SSR marker pairs in the 57 dent lines. The ratio of linked to unlinked loci in LD was 1.1 for both germplasm groups. The high incidence of LD suggests that the extent of LD between SSR markers should allow the detection of marker-phenotype associations in a genome scan. However, our results also indicate that a high proportion of the observed LD is generated by forces, such as relatedness, population stratification, and genetic drift, which cause a high risk of detecting false positives in association mapping.

Introduction

Linkage mapping has become a routine tool for the identification of quantitative trait loci (QTL) in plants (for review, see Alpert and Tanksley 1996; Stuber et al. 1999). However, this procedure has major limitations, including high running costs (Parisseaux and Bernardo 2004) and a poor resolution in detecting QTL, and that with biparental crosses of inbred lines only two alleles at any given locus can be studied simultaneously (Flint-Garcia et al. 2003). An alternative, promising approach is association mapping, which has been successfully applied in human genetics to detect QTL coding for simple as well as complex diseases (Corder et al. 1994; Kerem et al. 1989). This method uses the linkage disequilibrium (LD) between DNA polymorphisms and genes underlying traits. The application of association mapping to plant breeding appears to be a promising approach to overcome the limitations of conventional linkage mapping (Kraakman et al. 2004).

The applicability and resolution of association mapping depends on the extent and structure of LD within the population under consideration. Major differences between human and plant breeding populations are expected, because in the latter (1) random mating is the exception rather than the rule and selfing is prevalent (Bernardo 2002), (2) effective population sizes are generally small (Ching et al. 2002), and (3) strong selection pressure is applied. In addition, association mapping requires detailed knowledge about genetic and phylogenetic relationships of the materials investigated because a disregarded population structure may cause spurious associations (Pritchard et al. 2000).

In contrast to animals and humans, little information is available on LD in plants, with most research being done in *Arabidopsis thaliana* and maize. Nordborg et al. (2002) examined patterns of LD of *A. thaliana* and reported that broad-based populations showed a lower level of LD than narrow-based populations. The same trend was observed in maize. In broad-based germplasm, intragenic LD decreased within 200 base pairs to r^2 values smaller than 0.25 (Tenaillon et al. 2001). Remington et al. (2001) obtained similar results with single nucleotide polymorphism (SNP) markers, and genome-wide LD for simple sequence repeat (SSR)

Communicated by R. Bernardo

B. Stich · A. E. Melchinger (✉) · M. Frisch · H. P. Maurer
M. Heckenberger · J. C. Reif
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

markers was significant ($P < 0.01$) among 10% of the marker pairs. In narrow-based elite germplasm, no decay of LD within a few hundred basepairs was observed (Ching et al. 2002). In a very diverse set of elite germplasm, Liu et al. (2003) detected a high level of LD between 94 SSR markers. To our knowledge, no information is available on LD in European maize germplasm, and no earlier study has examined the forces conserving and generating LD in elite maize inbreds.

The objectives of our research were to (1) investigate genetic diversity within germplasm groups of elite European maize inbred lines, (2) examine the population structure of elite European maize germplasm, and (3) determine the extent and genomic distribution of LD between pairs of SSR markers.

Materials and methods

Plant material and molecular markers

A cross-section of European and US elite breeding material comprising 147 inbred lines was used in this study. The European inbreds belong to the flint and dent germplasm groups. The US lines originated from the Iowa stiff stalk (SS) and non-stiff stalk (NSS) germplasm groups (Table 1) (pedigree see supplemental table at http://www.uni-hohenheim.de/~stich/pedigree/Tab_S1.html). For a direct comparison of the genetic diversity and LD estimates with those of Liu et al. (2003), the same set of 100 SSR markers was analyzed by the same company (Celera, 1756 Picasso Avenue, Davis, CA 95616, USA) with identical protocols for PCR and allele calling procedure. As in the study of Liu et al. (2003), we dropped six loci, each which had either a mean within-line heterozygosity of more than 0.10 or a proportion of missing data surpassing 0.20. A list of the SSR loci with their chromosomal locations has been deposited as a supplemental table at http://www.uni-hohenheim.de/~stich/markers/Tab_S2.html. Map positions of all SSRs are based on the Celera linkage map (unpublished data).

Statistical analyses

The modified Rogers distance (MRD) was calculated according to Wright (1978). Associations among the groups were revealed with principal coordinate analysis (Gower 1966) based on MRD estimates between pairs of inbred lines. The average number of alleles per locus and the number of group-specific alleles were determined for the germplasm groups and for various subsets within this collection. The total gene diversity was decomposed into gene diversity estimates between individual lines within each germplasm group according to Nei (1987). Confidence intervals for gene diversity estimates were obtained by a bootstrap procedure with resampling across markers. Nei's minimum distance was calculated according to Nei (1987).

A model-based approach implemented in software package STRUCTURE (Pritchard et al. 2000) was used to subdivide the group of flint and dent inbreds. Because STRUCTURE overestimates the number of subgroups when examining inbred individuals (Pritchard and Wen 2004), it is more reliable to choose the number of subgroups based on prior information. We set the number of subgroups to two based on prior knowledge of the pedigree structure. The established subgroups were further subdivided using separate runs of STRUCTURE for each subgroup as proposed by Pritchard and Wen (2004). The number of clusters was set for each subgroup to two. For each run, the burn-in time was 50,000 and the number of replications was 100,000, following the suggestion of Pritchard and Wen (2004). The run with the maximum likelihood was used to assign lines with membership probabilities of 0.80 or more to subgroups. Lines with membership probabilities less than 0.80 for both subgroups were assigned to a mixed group.

A permutation test using a Monte-Carlo procedure was applied to test for LD between pairs of SSR loci on a genome-wide scale (Lewis and Zaykin 2002; Weir 1996). An adaptive permutation procedure was used to reduce the computational effort: if the P value was smaller than 0.3 after 2,500 permutations, another 14,500 permutations were performed. If the P value was

Table 1 Grouping of the 147 maize inbred lines

Group	Subgroup	Inbreds
Flint	Subgroup 1	L1, L2, L3, L4, L5, L6, L7, L8, L9, L10, L13, L55, L56, L57, L58, L59, L62, L83, L172, L173, L174, L175, L176, L177, L178, L179, L180, L181, L182, L183, L184, L185, L186, L187, L188, L189, L190, L191, L191, L192, L193, L194, L195, L196, L197, L198, L199, L200, L201, L205, L206, L207, L208, L209, L210, L211, L212, L213, L229
	Subgroup 2	L66, L95, L96, L97, L98, L99, L100, L101, L102, L103, L104, L105, L106, L107
	Mixed	L11, L12, L15, L54, L61, L74, L89, L269, L272
Dent	Subgroup 1	L28, L29, L31, L33, L34, L39, L41, L65, L108, L244, L271, L289, L291, L293, L295, L297, L299, L301
	Subgroup 2	L24, L46, L47, L49, L50, L51, L52, L53, L64, L67, L68, L70, L72, L92, L110
	Mixed	L21, L23, L25, L27, L32, L35, L36, L37, L38, L42, L43, L44, L85, L88, L93, L109, L111, L112, L245, L246, L247, L248, L249, L275
SS		B101, B105, B108, B109
NSS		B98, B100, B102, B106, B107, LO977

Table 2 Average number of alleles per locus, number of group-specific alleles, and gene diversity for the four maize germplasm groups

Statistics	Overall	Flint	Dent	SS	NSS
Sample size	147	80	57	4	6
Alleles per locus	9.8	5.1	6.0	2.9	3.5
Group-specific alleles		186	183	52	73
Gene diversity[a]	0.68	0.50b	0.56a	0.43c	0.50b

[a] Gene diversity values followed by the same letters are not different at the 0.05 significance level according to a bootstrap procedure

then smaller than 0.075, another 933,000 permutations were performed. Test statistic D' for LD was calculated according to Tenesa et al. (2003). Since the power of the test for LD depends on the number of inbred lines analyzed per group, we used a resampling strategy to obtain comparable estimates: random samples were drawn from the group with the higher number of lines using a sample size equal to the number of lines of the smaller group. This procedure was repeated 50 times and the results were averaged. The sequential Bonferoni procedure was used for testing the genome-wide null hypothesis "all marker loci pairs are in linkage equilibrium" taking the multiple test problem for independent tests into account (Stahel 1995).

The ratio of the percentage of linked to unlinked loci pairs in LD was calculated, whereas linked loci were defined to be located on the same chromosome and unlinked loci were defined to be located on different chromosomes. Spearman rank correlation was calculated between D' estimates and the genetic map distance between markers. If within a chromosome region all pairs of adjacent loci were in LD, this region was referred to as a LD block. Gene diversity and Nei's minimum distance were calculated with POWERMARKER (Liu 2002). All other analyses were performed with software PLABSOFT (Maurer et al. 2004), which is implemented as an extension of the statistical software R (Ihaka and Gentleman 1996).

Results

The total number of alleles detected for the 94 SSR loci was 923, with the number of alleles per locus ranging from 3 to 22. The mean number of alleles per locus ranged from 2.9 for the four SS lines to 6.0 for the 57 dent lines (Table 2). The number of group-specific alleles was lowest (52) for the SS lines and highest (186) for the 80 flint lines. We observed a total gene diversity of 0.68. For gene diversity within each group, significant ($P < 0.05$) differences existed among groups, with values ranging from 0.43 (SS) to 0.56 (dent).

In principal coordinate analysis based on MRD estimates of all inbred lines, the first two principal coordinates explained 20% and 6% of the molecular variance (Fig. 1). The flint lines were clearly separated from the other lines with respect to the first principal

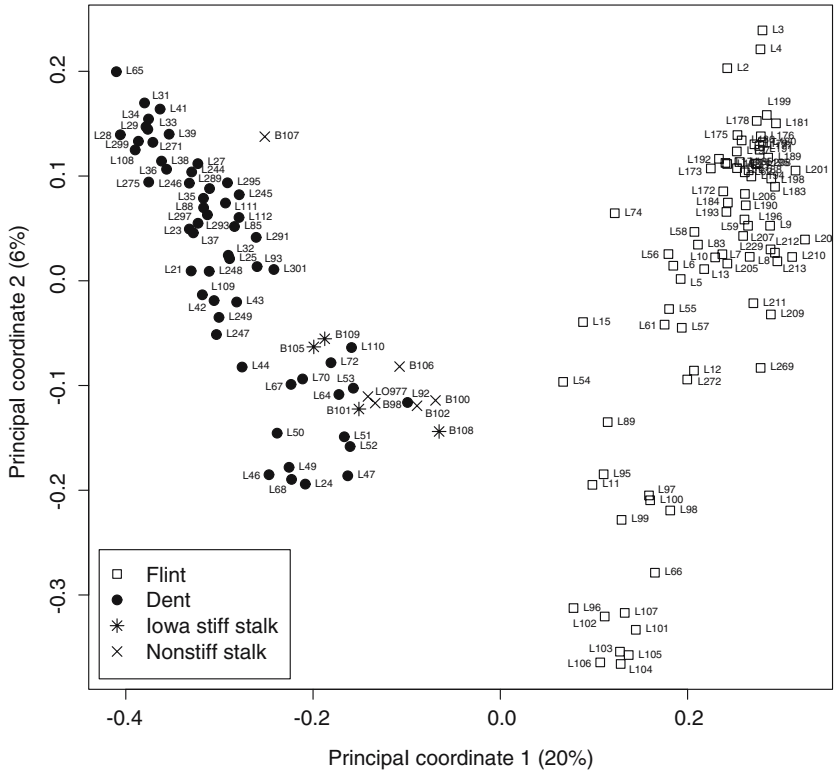

Fig. 1 Principal coordinate analysis of 147 maize inbred lines based on modified Rogers distance estimates. Numbers in parentheses refer to the proportion of variance explained by the principal coordinate

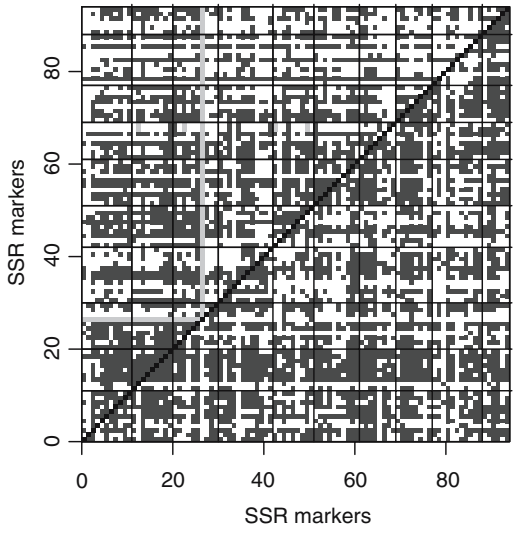

Fig. 2 Pairs of SSR loci in significant ($P < 0.05$) linkage disequilibrium (LD) in the flint group (80 lines, above the diagonal) and in the dent group (57 lines, below the diagonal). Dark-gray squares indicate pairs of loci in significant LD, light-gray comparisons were excluded because at least one locus was monomorphic. The thin horizontal and vertical lines mark off the chromosomes

coordinate. The US lines clustered together with one exception (B107) and were slightly separated from the dent cluster. Nei's minimum distance between pairs of groups was 0.18 (dent × NSS), 0.20 (dent × SS), 0.21 (SS × NSS), 0.27 (flint × dent, flint × NSS), and 0.31 (flint × SS).

The model-based approach of STRUCTURE clearly separated flint and dent lines. STRUCTURE further subdivided the flint group into two subgroups consisting of 57 and 14 inbreds. Nine inbred lines had for both subgroups membership probabilities of less than 0.80, and, thus were assigned to a mixed group. For the dent group, the STRUCTURE analysis revealed two subgroups with 18 and 15 inbred lines. Twenty-four dent lines were assigned to a mixed group.

The genome-wide null hypothesis "all SSR loci pairs are in linkage equilibrium" was rejected for all of the germplasm groups ($P < 0.05$). The percentage of SSR loci pairs with significant ($P < 0.05$) LD was 49% in the flint group and 56% in the dent group (Fig. 2). The ratio of linked to unlinked loci in LD was 1.1 in both groups.

Spearman rank correlation between D' and the genetic map distance of the respective marker loci was highly significant ($P < 0.001$) in both germplasm groups (Table 3). Loci pairs in LD on a given chromosome varied among the different chromosomes and ranged from 27% to 86% in the flint group and from 33% to 93% in the dent group. The proportion of significant pairwise LD tests was smaller within each model-based subgroup except for subgroup 2 of the dent lines. In addition, the ratio of the percentage of linked to unlinked loci in LD increased within the subgroups. For the group "flint unrelated", which consisted of lines not descending from common ancestors within the last two generations, we detected 17.7% of the loci pairs in LD. The percentage of loci pairs in LD was 22.1% in the group "flint size 9".

The total number of LD blocks was higher for the flint group than for the dent group (Table 4). Nevertheless, the dent group showed longer LD blocks than the flint group. In the flint group, the longest LD blocks were found on chromosome 2 (105 cM) (Fig. 3); in the dent group, on chromosome 8 (103 cM).

Discussion

Maize was first introduced into Europe by Columbus from the West Indies to southern Spain in 1493. Later on, maize germ plasm 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 cooler climatic conditions of Central Europe (Rebourg et al. 2003). This material is regarded as the origin of the European open-pollinated flint varieties. Hybrid maize breeding was started in

Table 3 Percentage of SSR loci pairs in significant ($P < 0.05$) linkage disequilibrium (LD) and Spearman rank correlation coefficient ρ between D' estimate and genetic map distance between markers in European maize germplasm groups

Group	No. of lines	Loci pairs in LD			Ratio linked: unlinked loci in LD	Spearman ρ
		Linked (%)	Unlinked (%)	Total (%)		
Flint	80	55	48	49	1.1	−0.25***
Subgroup 1	57	38	30	31	1.3	−0.18***
Subgroup 2	14	35	20	21	1.7	−0.20***
Flint size 57[a]	57	53	44	45	1.2	−0.22***
Flint size 14[a]	14	30	25	25	1.2	−0.12***
Dent	57	60	55	56	1.1	−0.33***
Subgroup 1	18	32	24	25	1.3	−0.12*
Subgroup 2	15	51	47	47	1.1	−0.19***
Dent size 18[a]	18	41	36	37	1.1	−0.22***
Dent size 15[a]	15	39	34	34	1.2	−0.20***

*,***Significant at the 0.05 and 0.001 probability level, respectively
[a] Random samples were drawn from the corresponding group 50 times, and the results were averaged

Table 4 Number of linkage disequilibrium (LD) blocks per chromosome and their average length per chromosome in centiMorgan (cM) in European maize germplasm groups

Chromosome	Group			
	Flint		Dent	
	Number of blocks[a]	Length (cM)	Number of blocks[a]	Length (cM)
1	3	33	2	69
2	1	105.0	1	49
3	3	9.3	3	10
4	2	33	2	53
5	2	32	1	52
6	2	41	2	48
7	2	11	1	73
8	2	27	1	103
9	3	10	4	13
10	2	8.0	1	44
Sum	22		18	
Mean		26		41

[a] An LD block consists of a sequence of markers for which all pairs of adjacent loci are in significant ($P < 0.05$) LD

Central Europe after World War II. As a promising heterotic pattern, high-yielding US dent lines were crossed with the adapted European flint lines (Schnell 1992). The steady influx of dent germplasm from North America to Europe has continued over the past 50 years. In contrast, the parental flint inbreds were developed by selfing from a few European open-pollinated varieties such as Lacaune, Lizargarote, Gelber Badischer Landmais, and Rheintaler (Messmer et al. 1992).

Genetic diversity and population structure

We found a total gene diversity of 0.69. Our results are directly comparable to those of Liu et al. (2003) due to the fact that in both studies the same SSR marker set was analyzed by the same lab. Liu et al. (2003) detected a

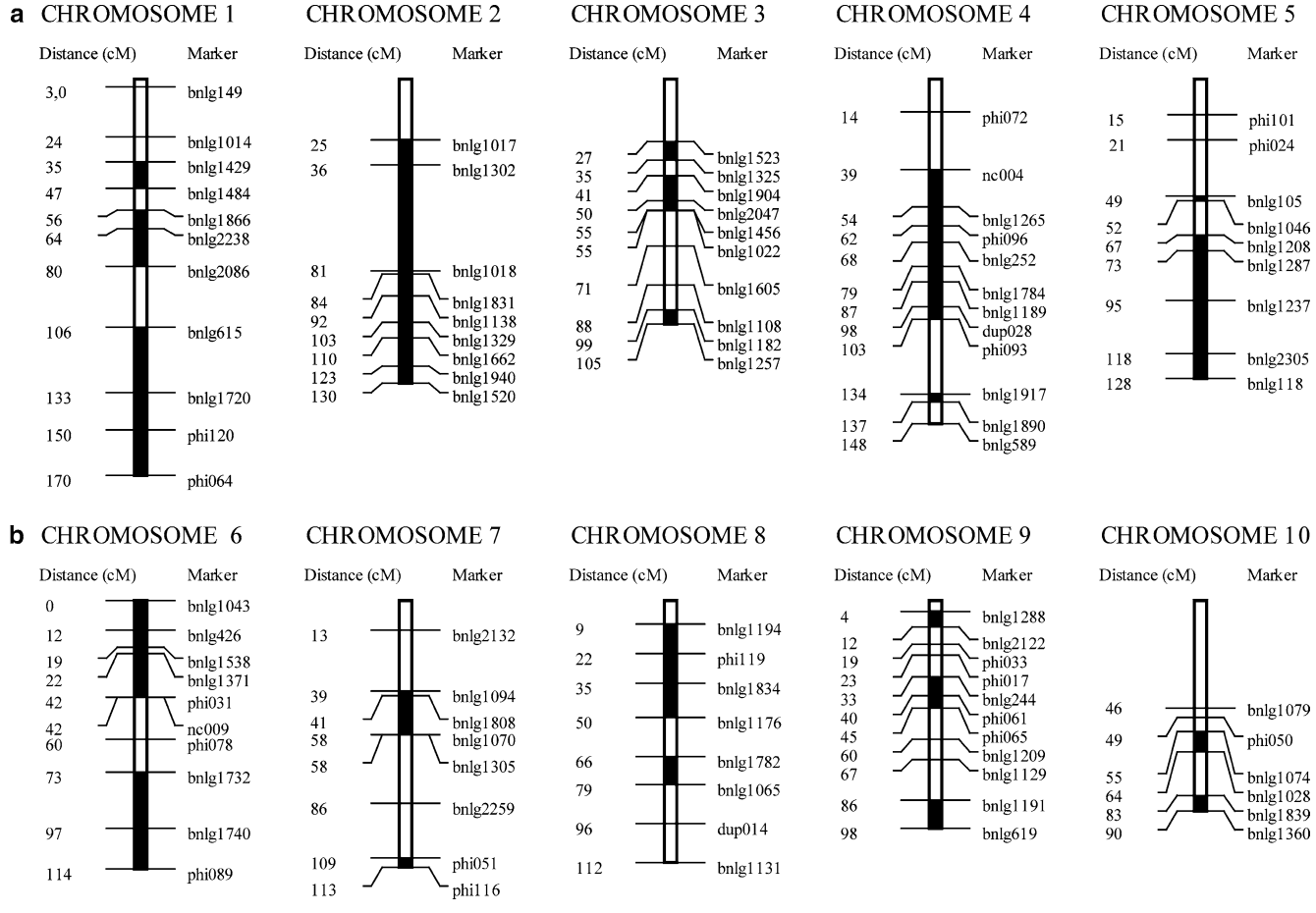

Fig. 3 Genome-wide distribution of the SSR linkage disequilibrium (*LD*) blocks in the flint lines. Black bars indicate LD blocks

higher gene diversity (0.82) when examining 260 inbreds, which can be explained by the inclusion of a broad range of germplasm adapted to tropical, subtropical, and temperate regions in their survey.

The principal coordinate analysis revealed a clear separation between the flint and dent lines (Fig. 1). This is in accordance with results of previous studies (e. g., Lübberstedt et al. 2000; Messmer et al. 1992) and can be explained by the breeding history described above. Nei's minimum distance between pairs of germplasm groups showed that SS and NSS lines clustered closely to the dent lines. This is consistent with the breeding history because the elite European dent germplasm was developed by breeders using both SS and NSS material. The model-based subgroups were also in accordance with pedigree data. The flint lines assigned to different subgroups were bred in different breeding programs. The dent lines of subgroup 1 are, as far as the pedigree is known, related to Iodent line L65, whereas the predominant parents of subgroup 2 are the inbred lines UH021 and UH022.

Nei's minimum distance between the SS and NSS group is in the present study based only on a small number of US lines. Therefore, Nei's minimum distance between the flint and dent group estimated in the present study was compared with the distance between the SS and NSS group reported by Liu et al. (2003). The divergence between the two heterotic pools of the US was of the same magnitude as that between the heterotic pools of Central Europe.

The comparison of the flint and dent germplasm based on the average number of alleles per locus and gene diversity (Table 2) reflected a higher genetic diversity among the dent lines than among the flint lines. This is in accordance with the experimental results of Lübberstedt et al. (2000) based on amplified fragment length polymorphic data and can be explained by the breeding history of the European germplasm groups.

The gene diversity of the dent group (0.56) was of the same magnitude as the gene diversity observed by Liu et al. (2003) for SS lines (0.59) but lower than that observed by Liu et al. (2003) for NSS (0.78) lines. As the European dent group was established by using SS and NSS germplasm, the low gene diversity was not expected from pedigree information. It was presumably caused by a rigid selection for adaption to the cooler growing conditions in Central Europe (Brandolini 1969).

Both European germplasm groups showed a low genetic diversity in comparison with the US heterotic groups. This indicates that new genetic material should be integrated in both European germplasm groups to ensure a long-term response to selection. The genetic basis of the flint group can be broadened by the introgression of European open-pollinated flint varieties (Reif et al. 2005). In addition, our results strongly suggest that both of the SS and NSS US dent pools represent a valuable source to broaden the European dent pool.

Linkage disequilibrium and its potential causes

LD was analyzed within the single germplasm groups and not within the set of all inbreds because the germplasm groups were clearly separated by the STRUCTURE analysis. In each germplasm group, about one-half of the loci pairs showed significant ($P < 0.05$) LD (Table 3), which is in accordance with the results of Liu et al. (2003) but considerably higher than those reported by Remington et al. (2001). This discrepancy between our results and those of the latter study is presumably attributable to two factors: (1) we used a much higher marker density than Remington et al. (2001); (2) their germplasm was chosen to avoid closely related lines (Liu et al. 2003), whereas we examined a cross-section of breeding material comprising both related and unrelated inbred lines.

Information about the causes generating LD is essential for drawing conclusions about the prospects of association mapping. LD conserved by linkage is useful for association mapping. The classification of selection and mutation forces generating useful and non-useful LD for association mapping depends on how the LD information is used. LD generated by relatedness, population stratification, and genetic drift causes spurious marker-phenotype associations.

LD was initially generated in the flint and dent heterotic groups when establishing them in the 1950s by using germplasm with differing allele frequencies. For the initial 25 years, both germplasm groups were bred without winter nurseries (W. Schipprack, personal communication), so that the number of completed breeding cycles is 7.78 when assuming a cycle length of 9 years without winter nurseries and 5 years with winter nurseries. Applying the formula of Haldane and Waddington (1931) yields an estimated number of effective crossovers for one breeding cycle of 1.31 per Morgan and, consequently, 10.19 effective crossovers per Morgan during the past 50 years. This implies that we can expect to detect significant ($D' > 0.25$) LD due to linkage between the SSR loci separated by less than 14.5 cM when assuming an initial D' estimate of 1. In contrast, no significant ($D' < 0.25$) LD is expected between loosely linked and unlinked loci. This indicates that the SSR marker density used in our study might be high enough for genome-wide association mapping, even if there were no forces other than linkage conserving or generating LD.

We observed for both germplasm groups a highly significant ($P < 0.001$) correlation between the D' estimate and the genetic map distance between SSR markers (Table 3), suggesting a decay of LD with distance. This has also been observed for *A. thaliana* (Nordborg et al. 2002), sugarcane (*Saccharum officinarum* L.) (Jannoo et al. 1999), and humans (Service et al. 2001). A decay of LD with distance indicates that linkage is a factor conserving LD. The weak association between D' and genetic map distance and the high proportion of unlinked loci in significant LD (Table 3) suggests,

however, the presence of further forces generating LD between both linked and unlinked loci.

Selection acting on a monogenic trait generates LD around the gene. If selection is performed on an oligo- or polygenic trait, LD is generated not only between linked genes but also between unlinked genes coding for the trait. Theoretical considerations suggest that this LD is useful for association mapping. Moreover, in plant breeding programs, selection is commonly performed simultaneously on several traits and, consequently, can generate also LD between genes influencing different traits. Thus, selection can complicate association mapping. In their experiments with maize, Robinson et al. (1960) observed no inversion of the Bulmer effect upon random mating. Despite the large standard errors of the variance components, this result suggested that the previously performed selection did not strongly reduce the additive variance. Consequently, selection may be an unlikely cause for LD.

Vigouroux et al. (2002) estimated the mutation rate for SSRs in maize to be 7.7×10^{-4} for loci with dinucleotide repeats and lower for loci with higher-nucleotide repeats. Because of the low mutation rate and due to the fact that a mutation must occur in a predominant parent to cause significant LD, we assume that mutation is at best a marginal factor causing LD in elite maize germplasm.

Theoretical considerations suggest that relatedness generates LD between linked loci. Only if predominant parents exist in a germplasm group LD between unlinked loci is expected to be generated by relatedness. The group ''flint unrelated'' consisted of lines not descending from common ancestors within the last two generations. The comparison of the percentage of loci pairs in LD identified of the group ''flint unrelated'' with that of the group ''flint size 9'' suggests that relatedness increases the extent of LD. Furthermore, we observed that relatedness causes LD between linked and unlinked loci in equal proportion. This is in accordance with the fact that there were predominant parents in the flint group. The higher Spearman rank correlation coefficient ρ between D' and genetic map distance in the group ''flint unrelated'' than in the group ''flint size 9'' (data not shown) reflects that the influence of linkage on the level of intrachromosomal LD in the group ''flint size 9'' is masked by relatedness.

For three out of four model-based subgroups the percentage of loci pairs in LD decreased (Table 3). Simultaneously, the ratio of linked to unlinked loci in LD increased. This result suggests that population stratification is a major factor causing LD in the inbred lines examined. In the subgroups, no clear trend was observed that linkage has an higher influence on the level of intrachromosomal LD than in the main group. This can be explained by the presence of other forces influencing intrachromosomal LD in the subgroups, such as relatedness and genetic drift, which mask this expected effect.

A small effective population size, commonly present in breeding programs (Ching et al. 2002), accelerates the decay of LD with distance (Hill and Robertson 1968). However, the variance of the observed D' estimates increases if the effective population size gets smaller. Therefore, a high percentage of loci pairs in LD can be observed due to genetic drift. In the present study, the percentage of loci pairs in LD did not show a uniform distribution among the chromosomes. This suggests that genetic drift is most likely a major force generating LD in the inbred lines examined (Huttley et al. 1999).

Linkage disequilibrium blocks

The length of chromosome regions in LD is crucial for application of association mapping because (1) regions in LD need to be present in order to detect marker-phenotype associations and (2) the length of the regions limits the resolution of association mapping. The LD blocks observed in the present study had an average length of 33 cM (Table 4), while those detected by Ching et al. (2002) were much shorter (500 basepairs). Possible reasons for the difference are (1) the different marker system used and (2) the much smaller average genetic map distance between markers in the study of Ching et al. (2002). A small average genetic map distance between markers reduces the observed average LD block length in two ways: (1) it allows the detection of short LD blocks when present and (2) if LD between two linked markers is not caused by linkage, additional markers in between the former loci pair are not expected to be in LD. Thus, additional markers can chop a large LD block into smaller pieces. In order to get detailed knowledge about the length of LD blocks, further research with a high marker density is needed.

Prospects for association mapping

Successful association mapping depends on the possibility of detecting LD between marker alleles and alleles affecting the expression of phenotypic traits. This is only feasible if LD is present in the breeding material to be studied. The results of our study suggest that the extent of LD between SSR markers should allow the detection of marker-phenotype associations in a genome scan. However, our results also indicate that a high proportion of the observed LD is caused by relatedness, population stratification, and genetic drift. The advantage of performing association mapping in such populations is that the number of markers needed is minimized due to the high level of LD. However, this approach causes a high risk of detecting false positives in association mapping. When performing family-based association mapping approaches, this problem could be overcome.

Acknowledgements The molecular marker analysis of this research was supported by funds from the ''Gemeinschaft zur Förderung

der privaten deutschen Pflanzenzüchtung'' (GFP), Germany. Financial support for B. Stich was provided by a grant from the German National Academic Foundation. The authors thank the associate editor and two anonymous reviewers for their valuable suggestions.

References

Alpert KB, Tanksley SD (1996) High-resolution mapping and isolation of a yeast artificial chromosome contig containing *fw*2.2: A major fruit weight quantitative trait locus in tomato. Proc Natl Acad Sci USA 93:15503–15507

Bernardo R (2002) Breeding for quantitative traits in plants. Stemma Press, Woodbury, p24

Brandolini AG (1969) European races of maize. Proc Annu Corn Sorghum Res Conf 24:36–48

Ching A, Caldwell KS, Jung M, Dolan M, Smith OSH, Tingey S, Morgante M, Rafalski AJ (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet 3:1–14

Corder EH, Saunders AM, Risch NJ, Strittmatter WJ, Schmechel DE, Gaskell PC, Rimmler JB, Locke PA, Conneally PM, Schmader KE, Small GW, Roses AD, Haines JL, Pericak-Vance MA (1994) Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. Nat Genet 7:180–184

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357–374

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

Haldane JBS, Waddington CH (1931) Inbreeding and linkage. Genetics 16:357–374

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. Theor Appl Genet 38:226–231

Huttley GA, Smith MW, Carrington M, O'Brien SJ (1999) A scan for linkage disequilibrium across the human genome. Genetics 152:1711–1722

Ihaka R, Gentleman R (1996) A language for data analysis and graphics. J Comput Graph Stat 3:299–314

Jannoo N, Grivet L, Dookun A, D'Hont A, Glaszmann JC (1999) Linkage disequilibrium among modern sugarcane cultivars. Theor Appl Genet 99:1053–1060

Kerem B, Rommens JM, Buchanan JA, Markievicz D, Cox DK, Chakravarti A, Buchwald M, Tsui LC (1989) Identification of the cystic fibrosis gene: genetic analysis. Science 245:1073–1080

Kraakman ATW, Niks RE, Van den Berg PMMM, Stam P, Van Eeuwijk FA (2004) Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics 168:435–446

Lewis PO, Zaykin D (2002) GDA manual. University of North Carolina Press, Chapel Hill

Liu J (2002) POWERMARKER—A powerful software for marker data analysis. North Carolina State University Bioinformatics Research Center, Raleigh, N.C.

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

Lübberstedt T, Melchinger AE, Dußle C, Vuylsteke M, 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

Maurer HP, Melchinger AE, Frisch M (2004) PLABSOFT: Software for simulation and data analysis in plant breeding. In: 17th EUCARPIA Gen Congr 2004 (poster abstr). Tulln, Austria

Messmer MM, Melchinger AE, Boppenmaier J, Brunklaus-Jung E, Herrmann RG (1992) Relationships among early European maize inbreds: I. Genetic diversity among flint and dent lines revealed by RFLPs. Crop Sci 32:1301–1309

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

Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J, Hagenblad J, Kreitman M, Maloof JN, Noyes T, Oefner PJ, Stahl EA, Weigel D (2002) The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nat Genet 30:190–193

Parisseaux B, Bernardo R (2004) In silico mapping of quantitative trait loci in maize. Theor Appl Genet 109:508–514

Pritchard JK, Wen W (2004) Documentation for STRUCTURE software. The University of Chicago Press, Chicago

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

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, Hamrit S, Heckenberger M, Schipprack W, Maurer HP, 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 (in press)

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

Robinson HF, Cockerham CC, Moll RH (1960) Studies on estimation of dominance variance and effects of linkage bias. In: Kempthorne O (ed) Biometrical genetics. Pergamon Press, New York, pp 171–177

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

Service SK, Ophoff RA, Freimer NB (2001) The genome-wide distribution of background linkage disequilibrium in a population isolate. Hum Mol Genet 10:545–551

Stahel WA (1995) Statistische Datenanalyse. Friedrich Vieweg & Sohn Verlagsgesellschaft, Braunschweig, p 245

Stuber CW, Polacco M, Senior ML (1999) Synergy of empirical breeding, marker-assisted selection, and genomics to increase crop yield potential. Crop Sci 39:1571–1583

Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). Proc Natl Acad Sci USA 98:9161–9166

Tenesa A, Knott SA, Ward D, Smith D, Williams JL, Visscher PM (2003) Estimation of linkage disequilibrium in a sample of the United Kingdom dairy cattle population using unphased genotypes. J Anim Sci 81:617–623

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

Weir BS (1996) Genetic data analysis II, 2nd edn. Sinauer, Sunderland, p 127

Wright S (1978) Evolution and genetics of populations, vol IV. The University of Chicago Press, Chicago, p 91