Theor Appl Genet (2007) 115:529–536
DOI 10.1007/s00122-007-0586-1

ORIGINAL PAPER

# Potential causes of linkage disequilibrium in a European maize breeding program investigated with computer simulations

**Benjamin Stich · Albrecht E. Melchinger · Hans-Peter Piepho · Sonia Hamrit · Wolfgang Schipprack · Hans P. Maurer · Jochen C. Reif**

Received: 18 December 2006 / Accepted: 31 May 2007 / Published online: 28 June 2007


**Abstract** Knowledge about the forces generating and conserving linkage disequilibrium (LD) is important for drawing conclusions about the prospects and limitations of association mapping. The objectives of our research were to examine the importance of (1) selection, (2) mutation, and (3) genetic drift for generating LD in a typical maize breeding program. We conducted computer simulations based on genotypic data of Central European maize open-pollinated varieties which have played an important role as founders of the European flint heterotic group. The breeding scheme and the dimensioning underlying our simulations reflect essentially the maize breeding program of the University of Hohenheim. Results suggested that in a plant breeding program of the examined dimension and breeding scheme, genetic drift and selection are major forces generating LD. The currently used population-based association mapping tests do not explicitly correct for LD caused by these two forces. Therefore, increased type I error rates are expected if these tests are applied to plant breeding populations. As a consequence, we recommend to use family-based association tests for association mapping approaches in plant breeding populations.

Communicated by G. Wenzel.

B. Stich · A. E. Melchinger (✉) · S. Hamrit · W. Schipprack · H. P. Maurer · J. C. Reif
Institute for Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim, 70593 Stuttgart, Germany
e-mail: melchinger@uni-hohenheim.de

H.-P. Piepho
Institute for Crop Production and Grassland Research, University of Hohenheim, 70593 Stuttgart, Germany

*Present Address:*
S. Hamrit
Institute for Molecular Physiology and Biotechnology, University of Rostock, 18051 Rostock, Germany

## Introduction

Hybrid maize breeding in Central Europe started in the 1950s (Schnell 1992). As a promising heterotic pattern, adapted flint lines were extracted from a few European open-pollinated flint varieties and were crossed with high yielding U.S. dent lines. During the establishment of the corresponding heterotic groups, LD was generated because the founders of each heterotic group differed in their allele frequencies (Reif et al. 2005a).

Subsequently, new lines were developed primarily by second-cycle breeding, i.e., from crosses among elite inbreds within heterotic groups. The extent of LD present in each heterotic group was reduced by genetic recombinations associated with this intermating. Based on the formula of Haldane and Waddington (1931), Stich et al. (2005) estimated the number of effective crossovers to be 1.31 per Morgan for one breeding cycle in maize. This result implies that only about 10 effective crossovers occurred per Morgan in the European heterotic groups since their establishment. Therefore, linkage is expected to be a force conserving considerable LD between markers that are around 15 cM or less distant.

Such LD can be used for association mapping, which has been successfully applied in human genetics to identify polymorphisms coding for cystic fibrosis (Kerem et al.

1989) and Alzheimer's disease (Corder et al. 1994). In plant genetics, association mapping has the potential to overcome the limitations of linkage mapping such as the poor resolution of detecting quantitative trait loci (QTL) and the limited number of alleles per locus that can be studied simultaneously (Flint-Garcia et al. 2003). Furthermore, in comparison with linkage mapping (Parisseaux and Bernardo 2004), association mapping is less expensive because data routinely collected in plant breeding programs can be used.

In addition to linkage, the extent and distribution of LD in plant breeding populations is also influenced by population stratification, relatedness, selection, mutation, and genetic drift (for review see, Flint-Garcia et al. 2003). Knowledge about the forces generating and conserving LD is of fundamental importance for drawing conclusions about the prospects and limitations of association mapping, because many forces other than linkage may cause spurious marker-trait associations in population-based association mapping approaches. Stich et al. (2005) examined the forces generating and conserving LD in European elite maize inbreds based on experimental data. This approach provides only conclusions about the relevance of population stratification, relatedness, and linkage. However, no information about the importance of selection, mutation, and genetic drift for the generation of LD in plant breeding populations is available.

The objectives of our research were to examine, based on computer simulations, the importance of (1) selection, (2) mutation, and (3) genetic drift for generating LD in a typical maize breeding program.

## Materials and methods

### Breeding scheme

Five Central European maize open-pollinated varieties (OPVs): Gelber Badischer Landmais, Mahndorfer, Maleksberger, Rheintaler, and Strenzfelder played an important role as ancestors of the first-cycle flint inbreds in Germany (Oettler et al. 1976). In the framework of the EU project GEDIFLUX (QLRT-2000-00934), the population structure of these OPVs was examined using 55 genome-wide distributed simple sequence repeat (SSR) markers (Reif et al. 2005a). These data served as basis of our computer simulations. The breeding scheme and the dimensioning underlying our simulations reflect essentially the maize breeding program of the University of Hohenheim (Fig. 1). In the first breeding cycle, 10 individuals, further referred to as $S_0$ individuals, were sampled out of each of the five OPVs. A total of 28 $S_6$ individuals, derived by selfing the 50 $S_0$ individuals, was selected based on a multi-stage selection procedure. The best $S_6$ individual was crossed with the other 27 $S_6$ individuals and the second best $S_6$ individual was crossed with the third to twenty-fifth best $S_6$ individuals to generate the 50 $S_0$ individuals for the next breeding cycle. We simulated eight breeding cycles

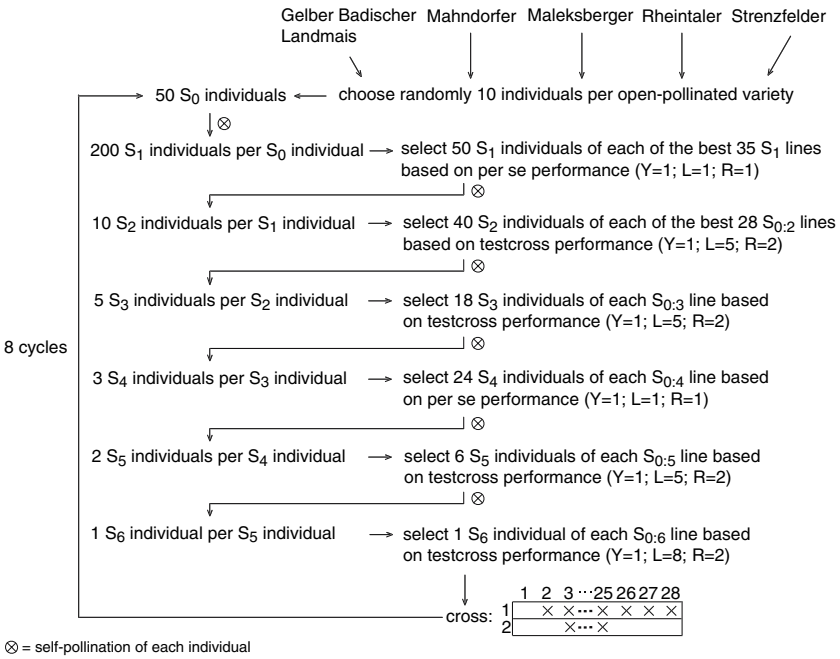


**Fig. 1** Model for simulating 55 years of breeding history in the European flint heterotic group. *Y*, *L*, and *R* denote the number of years, locations, and replications in performance trials, respectively

corresponding to 55 years of hybrid maize breeding in Europe (Schnell 1992).

## Genetic model

### *Testcross performance*

The numbers of genes controlling quantitative traits are yet unknown, although rough estimates include 69 loci for oil and 173 loci for protein content in the maize kernel (Dudley and Lambert 1992) whereas the number of loci for grain yield might be even higher. Therefore, we assumed 220 QTL influencing testcross performance (TP), which were randomly positioned in the maize genome. To ensure that OPVs showing a large genetic distance based on molecular markers also strongly differ in their alleles at QTL, for each of the five OPVs the allele frequencies of each of the 55 SSR loci were assumed for four QTL. Likewise, linkage equilibrium among QTL as well as between QTL and SSRs was assumed for each OPV as suggested by theoretical considerations.

The size of QTL effects $Q$ was defined based on the geometric series $220(1-a)[1, a, a^2, a^3, ..., a^{219}]$, with $a = 0.90$. At each QTL the $m$ alleles showed an average testcross effect of allele substitution of 0, $Q/(m-1)$, $2Q/(m-1)$, ..., $Q$. The genotypic value of each individual regarding TP was determined by summing up the effects of its alleles.

The phenotypic values of the individuals were generated by adding a normally distributed variable $N(0, \sigma_n^2)$ to the genotypic values, where

$$\sigma_n{}^2 = \frac{\sigma_{gy}^2}{Y} + \frac{\sigma_{gl}^2}{L} + \frac{\sigma_{gyl}^2}{YL} + \frac{\sigma_e^2}{YLR}$$

represents the non-genetic variance. $Y$, $L$, and $R$ denote the number of years, locations, and replications in performance trials, respectively. In our simulations, a ratio of variance components $\sigma_g^2{:}\sigma_{gy}^2{:}\sigma_{gl}^2{:}\sigma_{gyl}^2{:}\sigma_e^2$ of 1:1:1:2:4 was assumed for TP (Longin et al. 2006), where $\sigma_g^2$ refers to the genotypic variance, $\sigma_{gy}^2$ to the variance of genotype × year interactions, $\sigma_{gl}^2$ to the variance of genotype × location interactions, $\sigma_{gyl}^2$ to the variance of genotype × year × location interactions, and $\sigma_e^2$ to the error variance.

### *Line per se performance*

The genotypic values of the lines for line per se performance (LP) were estimated based on the same set of 220 QTL underlying TP, assuming a genotypic correlation between LP and TP of 0.5 (Mihaljevic et al. 2005). For the generation of phenotypic values for LP we assumed a ratio of variance components $\sigma_g^2{:}\sigma_{gy}^2{:}\sigma_{gl}^2{:}\sigma_{gyl}^2{:}\sigma_e^2$ of 1:0.2:0.2:0.5:0.5 (Longin et al. 2006).

### *Mutation model*

According to Calafell et al. (2001), a symmetric stepwise mutation model, which allows the gain and loss of more than one repeat, was implemented for the SSR loci. We assumed (1) a variance of change in the number of repeats of 3.2 and (2) a mutation rate of $5.1 \times 10^{-5}$ (Vigouroux et al. 2002). Because of the extremely low mutation rate of eucaryotic genomes (Drake et al. 1998), no mutations were assumed for the QTL.

## Examined simulation scenarios

In addition to the standard scenario described above, three modifications were examined. In variant A, we assumed the absence of SSR mutations. In variants B and C, individuals were not selected based on TP or LP but chosen at random. In variant C, the population size at each stage was tripled in comparison with the other simulation scenarios. The simulation of each scenario was replicated 50 times.

## Estimation of linkage disequilibrium

Because of the complexity of our simulations we do not know a priori which loci pairs are in LD. Therefore, after each breeding cycle, the LD in the set of selected $S_6$ individuals must be examined. For assessing the feasibility of association mapping the statistical significance of this LD is more important than its actual amount (Maurer et al. 2006). Thus, Fisher's exact test, which provides a significance test for LD between a pair of loci, i.e., determines whether two-locus genotypic frequencies can be represented as products of one-locus genotypic frequencies, seems to be an appropriate measure of LD for this purpose. To facilitate the comparison of our results with those of Reif et al. (2005a, b) we chose a significance level of $\alpha = 0.01$ and determined for each pair of loci the probability of occurence of less probable contingency tables with the same marginal totals as the ones observed with a Monte Carlo method using 2,500 replications (Guo and Thompson 1992). The applied test, which adheres to the nominal $\alpha$ level, does not require Hardy–Weinberg equilibrium at the loci under consideration (Zaykin et al. 1995). Percentage of significant LD was calculated between loci located on the same chromosome (linked) and between loci located on different chromosomes (unlinked). In variant C, we used a resampling strategy described in detail by Stich et al. (2005), to obtain comparable power for detecting significant LD in populations of different sizes. All simulations

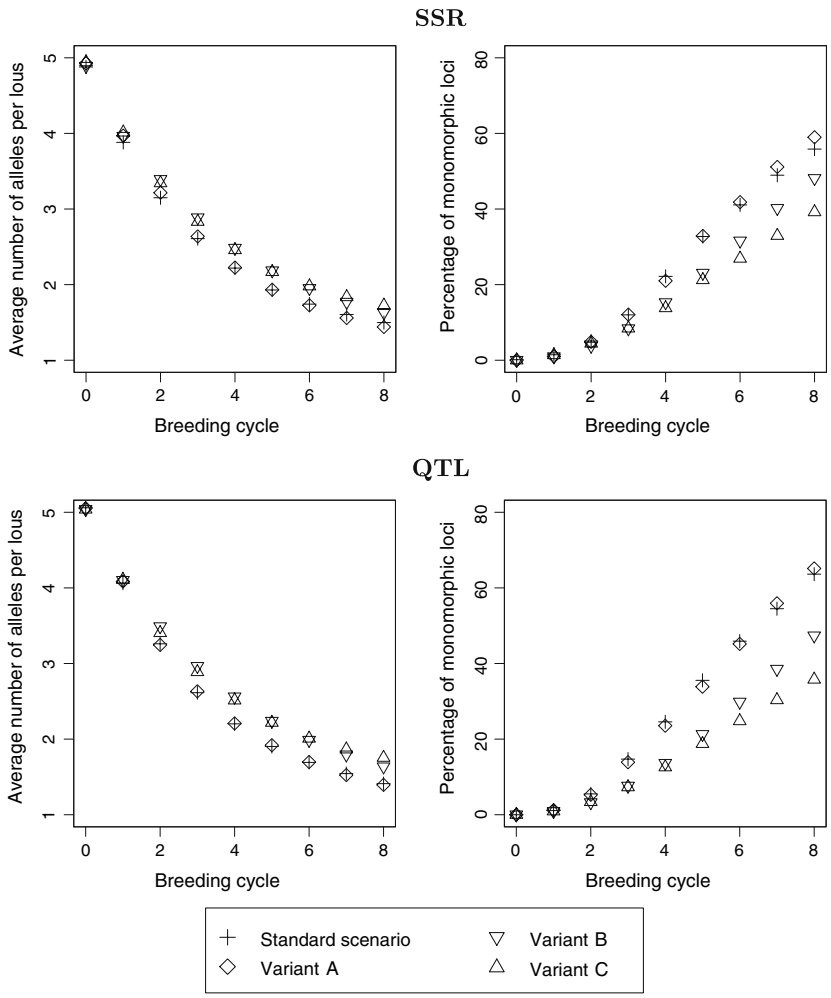


**Fig. 2** Average number of alleles per locus and percentage of monomorphic loci for SSRs and QTL in the different simulation scenarios adjusted for a population size of 28. For a detailed description of the variants see ''Materials and methods''

and analyses were performed with software Plabsoft (Maurer et al. 2004), which is implemented as an extension of the statistical software R (R Development Core Team 2004).

## Results

The average selection gain per breeding cycle was 7.2, 7.8, 0, and 0% in the standard scenario and variants A, B, and C, respectively. In cycle zero of all four examined simulation scenarios, the number of alleles ranged for both SSRs and QTL from 2.0 to about 14.5 and was on average about 5.0 (Fig. 2). With increasing number of completed breeding cycles, the average number of alleles per locus decreased, whereas the percentage of monomorphic loci increased. After eight breeding cycles, the average number of alleles per locus ranged from 1.4 (variant A) to 1.7 (variant C) for SSRs and from 1.4 (variant A) to 1.8 (variant C) for QTL.

After one breeding cycle, the percentage of SSR, SSR-QTL, and QTL pairs in significant ($P < 0.01$) LD was about 3.6% in the standard scenario and variant A (Table 1). In contrast, a percentage of about 3.2% was detected in variants B and C. With increasing number of completed breeding cycles, the percentage of linked and unlinked SSR, SSR-QTL, and QTL pairs in significant LD increased in all simulation scenarios up to a maximum between cycles two and three. In the subsequent breeding cycles, the percentage of loci pairs in significant LD decreased. After eight breeding cycles, the percentage of SSR pairs in significant LD ranged from 4.3 (variant C) to 7.3% (standard scenario) for linked loci and from 1.1 (variant C) to 1.7% (standard scenario) for unlinked loci. The percentage of SSR-QTL pairs in significant LD after eight breeding cycles was much higher, and varied between 7.9 (variant C) and 12.2% (variant A) for linked loci and between 1.1 (variant C) and 1.9% (standard scenario) for unlinked loci. Finally, the percentage of QTL pairs in significant LD after eight breeding cycles ranged from 8.3

**Table 1** Percentage of loci pairs in significant ($P < 0.01$) linkage disequilibrium (LD) and corresponding average standard error (SE) in different simulation scenarios adjusted for a population size of 28

| Completed breeding cycles | Loci pairs in LD (%) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Standard scenario | | | Variant A | | | Variant B | | | Variant C | | |
| | Linked | Unlinked | Total | Linked | Unlinked | Total | Linked | Unlinked | Total | Linked | Unlinked | Total |
| SSR | | | | | | | | | | | | |
| 1 | 3.2 | 3.7 | 3.6 | 3.5 | 3.6 | 3.6 | 2.8 | 3.2 | 3.2 | 2.7 | 3.1 | 3.0 |
| 2 | 7.4 | 3.8 | 4.1 | 8.0 | 3.9 | 4.3 | 7.2 | 3.2 | 3.5 | 7.0 | 2.9 | 3.3 |
| 3 | 9.2 | 3.5 | 4.0 | 8.5 | 3.2 | 3.6 | 7.9 | 2.7 | 3.2 | 6.8 | 2.4 | 2.7 |
| 4 | 9.3 | 3.2 | 3.7 | 8.0 | 2.3 | 2.8 | 7.7 | 2.3 | 2.7 | 6.4 | 2.1 | 2.4 |
| 5 | 8.2 | 2.6 | 3.1 | 7.9 | 2.0 | 2.5 | 6.4 | 2.2 | 2.5 | 5.3 | 1.7 | 2.0 |
| 6 | 6.4 | 2.2 | 2.5 | 6.5 | 2.3 | 2.7 | 6.8 | 1.9 | 2.3 | 5.3 | 1.5 | 1.8 |
| 7 | 5.2 | 1.6 | 2.0 | 5.3 | 1.7 | 2.0 | 5.8 | 1.6 | 2.1 | 4.9 | 1.3 | 1.5 |
| 8 | 7.2 | 1.7 | 2.2 | 6.8 | 1.6 | 2.0 | 6.1 | 1.6 | 2.0 | 4.3 | 1.1 | 1.4 |
| SE | 0.83 | 0.28 | 0.30 | 0.66 | 0.27 | 0.29 | 0.72 | 0.24 | 0.25 | 0.17 | 0.04 | 0.04 |
| SSR-QTL | | | | | | | | | | | | |
| 1 | 3.8 | 3.7 | 3.7 | 3.6 | 3.6 | 3.6 | 3.4 | 3.3 | 3.2 | 3.0 | 3.0 | 3.0 |
| 2 | 12.1 | 4.1 | 4.9 | 11.7 | 4.1 | 4.9 | 10.1 | 3.4 | 4.1 | 10.5 | 3.3 | 3.7 |
| 3 | 13.9 | 3.8 | 4.9 | 13.3 | 3.5 | 4.5 | 12.3 | 3.3 | 4.2 | 10.2 | 2.5 | 2.9 |
| 4 | 13.0 | 3.0 | 4.1 | 12.5 | 2.9 | 3.9 | 11.4 | 2.3 | 3.2 | 10.0 | 2.2 | 2.6 |
| 5 | 12.7 | 2.7 | 3.7 | 11.7 | 2.2 | 3.2 | 10.5 | 1.8 | 2.7 | 9.5 | 1.9 | 2.5 |
| 6 | 12.4 | 2.1 | 3.2 | 11.6 | 2.0 | 3.0 | 10.1 | 1.7 | 2.5 | 9.0 | 1.6 | 2.2 |
| 7 | 11.7 | 1.8 | 2.9 | 11.9 | 1.8 | 2.9 | 10.7 | 1.7 | 2.7 | 8.5 | 1.5 | 2.0 |
| 8 | 10.8 | 1.9 | 2.9 | 12.2 | 1.8 | 3.0 | 10.3 | 1.2 | 2.1 | 7.9 | 1.1 | 1.8 |
| SE | 0.46 | 0.21 | 0.23 | 0.46 | 0.21 | 0.22 | 0.41 | 0.16 | 0.17 | 0.09 | 0.02 | 0.03 |
| QTL | | | | | | | | | | | | |
| 1 | 3.9 | 3.8 | 3.8 | 3.8 | 3.7 | 3.7 | 3.5 | 3.4 | 3.4 | 3.0 | 3.0 | 3.0 |
| 2 | 12.5 | 4.3 | 5.1 | 12.1 | 4.3 | 5.1 | 10.4 | 3.5 | 4.2 | 11.2 | 3.3 | 3.8 |
| 3 | 14.4 | 4.0 | 5.1 | 13.8 | 3.7 | 4.7 | 12.7 | 3.4 | 4.4 | 10.9 | 2.6 | 3.2 |
| 4 | 13.3 | 3.1 | 4.2 | 13.1 | 3.0 | 4.1 | 11.7 | 2.3 | 3.3 | 10.8 | 2.2 | 2.6 |
| 5 | 13.0 | 2.8 | 3.9 | 12.1 | 2.3 | 3.3 | 10.8 | 1.8 | 2.8 | 10.3 | 1.8 | 2.2 |
| 6 | 13.0 | 2.2 | 3.3 | 12.0 | 2.1 | 3.1 | 10.3 | 1.7 | 2.6 | 9.6 | 1.5 | 2.1 |
| 7 | 12.1 | 1.9 | 3.0 | 12.6 | 1.8 | 3.0 | 11.1 | 1.8 | 2.8 | 9.1 | 1.3 | 2.0 |
| 8 | 11.3 | 2.0 | 3.0 | 13.0 | 1.9 | 3.1 | 10.6 | 1.2 | 2.2 | 8.3 | 1.1 | 1.9 |
| SE | 0.50 | 0.22 | 0.24 | 0.49 | 0.22 | 0.23 | 0.43 | 0.15 | 0.18 | 0.10 | 0.03 | 0.03 |

For a detailed description of the variants see ''Materials and methods''

(variant C) to 13.0% (variant A) for linked loci and between 1.1 (variant C) to 2.0% (standard scenario) for unlinked loci.

## Discussion

### Comparison of simulated with experimental data

In the present study we determined a selection gain per breeding cycle of about 7.5% in the standard scenario and variant A. This is of a order of magnitude observed in selection experiments using open-pollinated and synthetic base populations (Hallauer and Miranda 1981; Pandey et al. 1987; Stromberg and Compton 1989).

### *Trends in allelic diversity*

Our computer simulations were designed to model the breeding history of the European flint heterotic group. Reif et al. (2005b) examined European flint inbreds arranged in four groups according to the decade of release and genotyped with the same 55 SSRs as in our study. These results can be compared directly with those of the present study.

In both studies, the same temporal trend of the number of alleles per SSR locus was detected. However, in

comparison with the average number of alleles per SSR locus found in our study the values observed by Reif et al. (2005b) were considerably higher. This observation can be explained by the fact that inbreds in the latter study were sampled from several breeding programs, whereas in our study only a single breeding program was assumed.

*Linkage disequilibrium*

In our study and the study of Reif et al. (2005b) an increase of the extent of significant LD between SSRs was observed with the establishment of the heterotic group while with increasing number of completed breeding cycles the extent of LD decreased. However, the extent of significant LD found in our study was considerably lower than that described by Reif et al. (2005b). As the same SSR linkage map underlies both studies, the observed difference cannot be explained by differences in the average map distances between SSRs. A more likely reason is that the first-cycle flint inbreds were developed not only from the OPVs considered in our study but from additional source populations such as Lacaune and Lizargárate (Cartea et al. 1999). This increases the allele frequency differences among the ancestors of a heterotic group as well as the extent of LD in a heterotic group.

Furthermore, population stratification is likely to generate a high extent of LD in the study of Reif et al. (2005b), as inbreds were sampled from several breeding programs. This explanation is supported by the observation of (1) a lower ratio of the percentage of linked to unlinked SSR loci pairs in significant LD and (2) a higher number of alleles per SSR locus after 55 years of hybrid maize breeding in Europe in the study of Reif et al. (2005b) than in our study.

In summary, our simulation results on the selection gain, trends in allelic diversity, and linkage disequilibrium are in good accordance with results of experimental studies, thus, confirming the assumptions and models underlying our simulations.

## Potential causes of LD

Cost- and time-effectiveness of genome sequencing is improving exponentially (Shendure et al. 2004). Therefore, it may be possible to sequence genomes of all individuals of an association mapping population in the future. In this case funtional markers (Andersen and Lübberstedt 2003) will be available for the QTL itself and, thus, association mapping approaches will no longer be based on LD between QTL and adjacent molecular markers. In this scenario, knowledge about the forces generating and conserving LD will not be of importance. However, with today's genome sequencing technology, such approaches are not feasible.

The influence of relatedness, population stratification, and linkage as causes of LD in plant breeding populations can be examined with experimental data (e.g., Stich et al. 2005). However, basing investigations for selection, mutation, and genetic drift on experimental data is difficult (Farnir et al. 2000). Hence, we applied computer simulations to assess the relevance of these forces for generating and conserving LD in a plant breeding program which underlies a dimension and operation sequence typical for maize breeding.

*Selection*

The percentage of linked and unlinked SSR loci pairs in significant LD was higher in the standard scenario than in variant B. In the latter case, the selection of individuals during the breeding process was not based on their TP or LP, but was made at random. Therefore, our results suggested that even on a genome-wide scale, selection is an important cause of significant LD between neutral markers for the examined marker density (35 cM).

Furthermore, we observed a considerably higher extent of significant LD between QTL pairs in the standard scenario than in variant B. This result indicated that selection generates LD between pairs of QTL and thereby reduces the additive variance. This phenomenon is also known as the Bulmer effect (Bulmer 1971). Our observation is contrary to findings from an experimental study on maize in which no Bulmer effect was detected (Robinson et al. 1960). However, since the variance components reported by these authors showed large standard errors, the conclusions drawn are afflicted with uncertainty.

The results of the present study indicated that selection generates significant LD not only between linked but also between unlinked SSR-QTL pairs. Due to the latter, an increased rate of false positive associations between markers and unlinked QTL may be detected if population-based association mapping methods are applied to populations undergoing selection.

*Mutation*

No clear temporal trend was observed for the difference in the percentage of linked as well as unlinked SSR loci pairs in significant LD between the standard scenario and variant A. Because absence of mutation was assumed in the latter scenario, this result suggested that in plant breeding populations SSR mutation of the assumed frequency is a neglectable cause of LD. This observation is in accordance with findings of Terwilliger et al. (1998) for human populations.

With an increasing mutation rate of QTL, for which in the present study no mutations were assumed, as well as

SSRs it is expected that for populations in linkage equilibrium mutation generates LD (Iles and Bishop 1998). In plant breeding populations, however, LD is continuously generated by crossing individuals with differing allele frequencies. Therefore, it is expected to observe for plant breeding populations an erosion of LD with an increasing mutation rate.

*Genetic drift*

The formulas of Hill and Robertson (1968) indicate that in populations with small effective population size ($N_e$), LD is expected to decay faster than in populations with large $N_e$. However, the variance of the LD measure increases if $N_e$ is reduced. Therefore, genetic drift increases the extent of LD in a population. This was supported by the observations of Nordborg et al. (2002) that local populations of *Arabidopsis thaliana* showed a higher extent of LD than global populations.

The percentage of linked and unlinked SSRs and QTL pairs in significant LD were considerably higher in variant B than in variant C. The population size applied in variant B corresponds to that typically applied in plant breeding programs, whereas the population size of variant C was three times larger. Therefore, our findings indicate that for the population sizes typically applied in plant breeding programs, genetic drift is a major force generating significant LD between linked but also between unlinked loci. This result was supported by the findings of Stich et al. (2005). These authors observed for experimental data of European elite maize inbreds a non-uniform distribution of significant LD along the chromosomes suggesting that genetic drift is a major force generating LD.

Implications for association mapping

The results of our study suggested that in a plant breeding program of the examined dimension and breeding scheme, the effect of mutation on the extent of significant LD is neglectable. In contrast, our results suggested that in the examined plant breeding program genetic drift is presumably a major force generating significant LD between linked but also between unlinked loci. Furthermore, selection is a cause of significant LD between linked as well as unlinked QTL and SSRs. The existing population-based association mapping tests (e.g., Thornsberry et al. 2001; Yu et al. 2006) do not explicitly correct for LD caused by the latter two forces. Therefore, increased type I error rates are expected if these tests are applied to plant breeding populations. This was supported by the observation of an elevated level of false positives in mixed-model QTL mapping approaches using data from plant breeding programs (e.g., Yu et al. 2005; Arbelbide et al. 2006). In contrast, family-based association tests are valid tests of associations in the presence of linkage (Zhang et al. 2001) and, thus, adhere also to the nominal $\alpha$ level if LD is generated by population stratification, relatedness, genetic drift, and selection. Therefore, such tests are recommended for association mapping approaches in plant breeding populations.

**Acknowledgments** This research was conducted within the Breeding and Informatics (BRAIN) project of the Genome Analysis of the Plant Biological System (GABI) initiative (http://www.gabi.de). Financial support for B. Stich was provided by a grant from the German National Academic Foundation. The authors appreciate the editorial work of Dr. J. Muminović, whose suggestions considerably improved the style of the manuscript. The authors thank two anonymous reviewers for their valuable suggestions.

## References

Andersen JR, Lübberstedt T (2003) Functional markers in plants. Trends Plant Sci 8:554–560

Arbelbide M, Yu J, Bernardo R (2006) Power of mixed-model QTL mapping from phenotypic, pedigree and marker data in self-pollinated crops. Theor Appl Genet 112:876–884

Bulmer MG (1971) The effect of selection on genetic variability. Am Nat 105:201–211

Calafell F, Grigorenko EL, Chikanian AA, Kidd KK (2001) Haplotype evolution and linkage disequilibrium: a simulation study. Hum Hered 51:85–96

Cartea ME, Revilla P, Butrón A, Malvar A, Ordás A (1999) Do second cycle maize inbreds preserve the European flint heterotic group? Crop Sci 39:1060–1064

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

Drake JW, Charlesworth B, Charlesworth D, Crow JF (1998) Rates of spontaneous mutation. Genetics 148:1667–1686

Dudley JW, Lambert RJ (1992) Ninety generations of selection for oil and protein in maize. Maydica 37:81–87

Farnir F, Coppieters C, Arranz J-J, Berzo P, Cambisano N, Grisart B, Karim L, Marcq F, Moreau L, Mni M, Nezer C, Simon P, Vanmanshoven P, Wagenaar D, Georges M (2000) Extensive genome-wide linkage disequilibrium in cattle. Genome Res 10:220–227

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

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

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

Hallauer AR, Miranda JB (1981) Quantitative genetics in maize breeding. Iowa State University Press, Ames

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

Iles MM, Bishop DT (1998) The effect of population structure and mutation rate on linkage disequilibrium. Am J Hum Genet 63:A42

Kerem B, Rommens JB, 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

Longin CFH, Utz HF, Reif JC, Schipprack W, Melchinger AE (2006) Hybrid maize breeding with doubled haploids: I. One-stage versus two-stage selection for testcross performance. Theor Appl Genet 112:903–912

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

Maurer HP, Knaak C, Melchinger AE, Ouzunova M, Frisch M (2006) Linkage disequilibrium between SSR markers in six pools of elite lines of an European breeding program for hybrid maize. Maydica 51:269–279

Mihaljevic R, Schön CC, Utz HF, Melchinger AE (2005) Correlations and QTL correspondence between line per se and testcross performance for agronomic traits in four populations of European maize. Crop Sci 45:114–122

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

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

Pandey S, Diallo AO, Islam TMT, Deutsch J (1987) Response to full-sib selection in four medium maturity maize populations. Crop Sci 27:617–622

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

R Development Core Team (2004) R: A language and environment for statistical computing. Vienna, Austria

Reif JC, Hamrit S, Heckenberger M, Schipprack W, Maurer HP, Bohn M, Melchinger AE (2005a) Genetic structure and diversity of European flint maize populations determined with SSR analyses of individuals and bulks. Theor Appl Genet 111:906–913

Reif JC, Hamrit S, Heckenberger M, Schipprack W, Maurer HP, Bohn M, Melchinger AE (2005b) Trend of genetic diversity among European maize cultivars and their parental components during the past 50 years. Theor Appl Genet 111:838–845

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, New York, pp 171–177

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

Shendure J, Mitra RD, Varma C, Church GM (2004) Advanced sequencing technologies: methods and goals. Nat Rev Genet 5:335–344

Stich B, Melchinger AE, Frisch M, Maurer HP, Heckenberger M, Reif JC (2005) Linkage disequilibrium in European elite maize germplasm investigated with SSRs. Theor Appl Genet 111:723–730

Stromberg LD, Compton WA (1989) Tencycles of full-sib selection in maize. Crop Sci 29:1170–1172

Terwilliger JD, Zöllner S, Laan M, Pääbo S (1998) Mapping genes through the use of linkage disequilibrium generated by genetic drift: 'drift mapping' in small populations with no demographic Expansion. Hum Hered 48:138–154

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES (2001) Dwarf8 polymorphisms associate with variation in flowering time. Nat Genet 28:286–289

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

Yu J, Arbelbide M, Bernardo R (2005) Power of in silico QTL mapping from phenotypic, pedigree, and marker data in a hybrid breeding program. Theor Appl Genet 110:1061–1067

Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DN, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 2:203–208

Zaykin D, Zhivotovsky L, Weir BS (1995) Exact tests for association between alleles at arbitrary numbers of loci. Genetica 96:169–178

Zhang S, Zhang K, Li J, Sun F, Zhao H (2001) Test of association for quantitative traits in general pedigrees: the quantitative pedigree disequilibrium test. Genet Epidemiol 21:370–375