

A new test for family-based association mapping with inbred lines from plant breeding programs

Benjamin Stich · Albrecht E. Melchinger ·
Hans-Peter Piepho · Martin Heckenberger ·
Hans P. Maurer · Jochen C. Reif

Received: 15 March 2006 / Accepted: 14 July 2006 / Published online: 9 August 2006
© Springer-Verlag 2006

Abstract Association mapping holds great promise for the detection of quantitative trait loci (QTL) in plant breeding populations. The main objectives of this study were to (1) adapt the quantitative pedigree disequilibrium test to typical pedigrees of inbred lines produced in plant breeding programs, (2) compare the newly developed quantitative inbred pedigree disequilibrium test (QIPDT) with the commonly employed logistic regression ratio test (LRRT), with respect to the power and type I error rate of QTL detection, and (3) demonstrate the use of the QIPDT by applying it to flowering data of European elite maize inbreds. QIPDT and LRRT were compared based on computer simulations modeling 55 years of hybrid maize breeding in Central Europe. Furthermore, we applied QIPDT to a cross-section of 49 European elite maize inbred lines genotyped with 722 amplified fragment length polymorphism markers and phenotyped in four environments for days to anthesis. Compared to LRRT, the power to detect QTL was higher with QIPDT when using data collected routinely in plant breeding programs. Application of QIPDT to the 49 European maize inbreds resulted in a

significant ($P < 0.05$) association located at a position for which a consensus QTL was detected in a previous study. The results of our study suggested that QIPDT is a promising QTL detection method for data collected routinely in plant breeding programs.

Introduction

Estimation of the positions and effects of quantitative trait loci (QTL) is of central importance for marker-assisted selection. Up to now, this has been accomplished by classical QTL mapping approaches (Lander and Botstein 1989). The necessary experiments require establishment as well as pheno- and genotyping of large mapping populations and, thus, are very cost- and time intensive (Parisseaux and Bernardo 2004). These limitations could be overcome by applying association mapping methods in elite germplasm, using phenotypic and genotypic data routinely collected in plant breeding programs (Jansen et al. 2003). Moreover, results from association mapping would be of direct use in breeding, because allelic variation present in the entire elite germplasm is investigated.

With association mapping methods, detection of markers closely linked to QTL requires linkage disequilibrium (LD) caused by linkage being present between QTL and markers. Results of a companion study (Stich et al. 2006) on Central European elite maize germplasm suggested that LD between amplified fragment length polymorphism (AFLP) markers attributable to physical linkage could be expected between markers spaced at not more than 3 cM. However, it was concluded that also relatedness, population

Communicated by G. Wenzel

B. Stich · A. E. Melchinger (✉) · M. Heckenberger ·
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

stratification, and genetic drift are important causes of LD. When population-based association mapping tests are applied to germplasm in which LD is generated by the latter forces, the proportion of spurious marker-trait associations is expected to be much higher than the nominal type I error rate α (Pritchard et al. 2000b).

So far, the logistic regression ratio test (LRRT) has been applied as a population-based association mapping method accounting for population stratification (Pritchard et al. 2000a; Thornsberry et al. 2001). This test is based on the calculation of two hypotheses: first, that the distribution of the molecular marker under consideration was associated with population structure and phenotypic variation, and second, that the molecular marker distribution was associated only with population structure. The test statistic of LRRT is the ratio of the two likelihoods. However, the LRRT fails to correct for LD caused by relatedness or genetic drift and, thus, may show an increased rate of spurious marker-trait associations. In contrast, family-based association mapping methods suggested in human genetics, such as transmission disequilibrium test (TDT) (Spielman et al. 1993) or quantitative pedigree disequilibrium test (QPDT) (Zhang et al. 2001), adhere to the nominal α level even when LD is generated by population stratification, relatedness, or genetic drift. While these disturbing forces are most likely active in plant breeding populations, to our knowledge a family-based association mapping test has neither been developed in a plant breeding context nor applied to data typically available from plant breeding programs.

The objectives of our pilot study were to (1) adapt the QPDT to typical pedigrees of inbred lines produced in plant breeding programs, (2) compare the newly developed quantitative inbred pedigree disequilibrium test (QIPDT) with the commonly employed LRRT, with respect to the power and type I error rate of QTL detection, and (3) demonstrate the use of the QIPDT by applying it to flowering data of European elite maize inbreds.

Materials and methods

QIPDT, a family-based association test with inbred offsprings

In this treatise, we describe the QIPDT, a family-based association test applicable to genotypic information of parental inbred lines and geno- and phenotypic information of their offspring inbreds. This information is typically available from plant breeding programs with inbred line development such as line breeding or hy-

brid breeding. Nuclear families consisting of two parental inbred lines and at least one offspring inbred line can be combined to extended pedigrees, the basis of the QIPDT, if the parental lines of different nuclear families are related (Fig. 1). The QIPDT extends the QPDT, a family-based association test developed in the context of human genetics, which is in contrast to the quantitative transmission disequilibrium test (Abecasis et al. 2000) a valid test of association in the presence of linkage when some of the nuclear families are related (Zhang et al. 2001). Furthermore, the QIPDT takes into account the correction of Martin et al. (2001) regarding the pedigree disequilibrium test.

For a given nuclear family j_k with t_{jk} offspring inbreds belonging to the k th independent extended pedigree, let Y_{ijk} denote the phenotypic value of the i_{jk} th offspring inbred line for a quantitative trait of interest. Assume the molecular marker locus under study is bi-allelic with alleles M and N. Define $X_{ijk} = 0$ if the two parental inbreds possess the same allele and

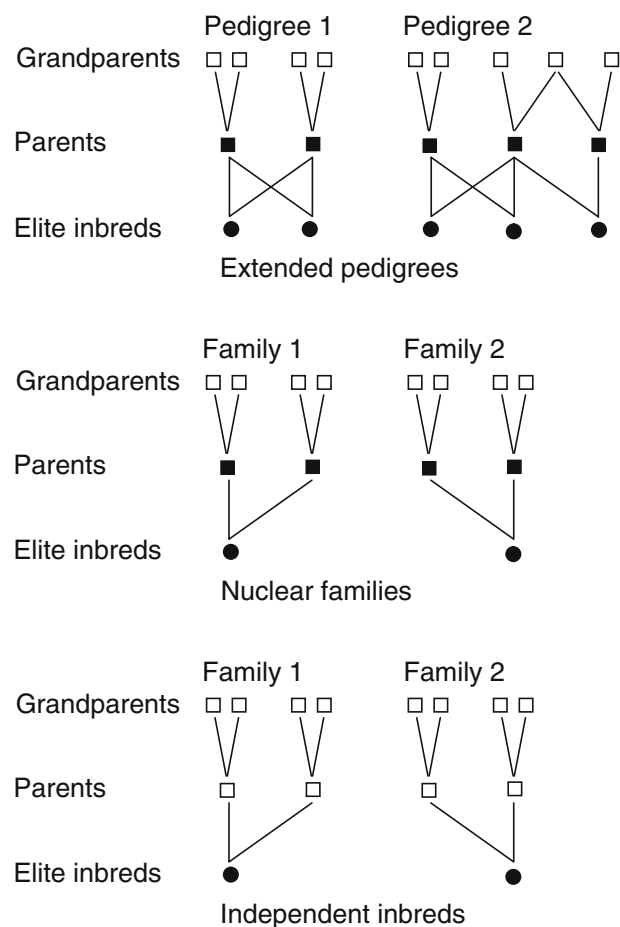


Fig. 1 Illustration of the simulated structures of inbred lines. Genotyped and phenotyped inbreds are presented as *filled circles*, genotyped but not phenotyped inbreds as *filled squares*, and inbreds neither genotyped nor phenotyped as *open squares*

$X_{ijk} = 1$ (or -1) if the two parental inbreds possess different alleles and transmit allele M (or N) to the i_{jk} th offspring inbred. Define the random variable U_{jk} for each nuclear family j_k of the extended pedigree k as:

$$U_{jk} = \sum_{i=1}^{I_{jk}} (Y_{ijk} - \bar{Y}_k) X_{ijk}, \tag{1}$$

where \bar{Y}_k is the mean phenotypic value of all offspring inbreds in the k th extended pedigree. If the number of offspring in the k th extended pedigree was only 1, \bar{Y}_k was replaced by \bar{Y} , the mean phenotypic value of all offspring inbreds in all extended pedigrees, following the original test of Zhang et al. (2001).

For the k th of p independent extended pedigrees, let n_k denote the number of the nuclear families. Define

$$D_k = \sum_{j=1}^{n_k} U_{jk}. \tag{2}$$

Under the null hypothesis of no marker-trait association, both alleles M and N are transmitted within each extended pedigree with equal probability to offspring inbreds with phenotypic values higher or lower than \bar{Y}_k resulting in $E(D_k) = 0$. Therefore,

$$\mu = E\left(\sum_{k=1}^p D_k\right) = 0 \tag{3}$$

and

$$\sigma^2 = Var\left(\sum_{k=1}^p D_k\right) = \sum_{k=1}^p Var(D_k) = E\left(\sum_{k=1}^p D_k^2\right). \tag{4}$$

Hence, if we define the test statistic of the QIPDT as

$$T = \frac{\sum_{k=1}^p D_k}{\sqrt{\sum_{k=1}^p D_k^2}}, \tag{5}$$

under the null hypothesis of no LD between marker locus and QTL, T is asymptotically normally distributed with mean zero and variance one. A statistical test for the null hypothesis based on T can be determined by comparison with a standard normal distribution.

LRRT

The LRRT, described in detail by Thornsberry et al. (2001) and implemented in software package TASSEL

(<http://www.maizegenetics.net>), is the population-based association mapping approach most frequently applied in plant populations. We compared it with the QIPDT with respect to the power and type I error rate of QTL detection. The LRRT requires information about the population structure, which is commonly determined using the model-based approach implemented in software package STRUCTURE (Pritchard et al. 2000a; Falush et al. 2003). In our investigations, the set of inbred lines under study was analyzed by setting the number of clusters from 1 to 12 in each of five repetitions. For each run of STRUCTURE, the burn-in time as well as the iteration number for the Markov Chain Monte Carlo algorithm were set to 100000, following the suggestion of Whitt and Buckler (2003). The population structure matrix with the highest likelihood was used for the LRRT.

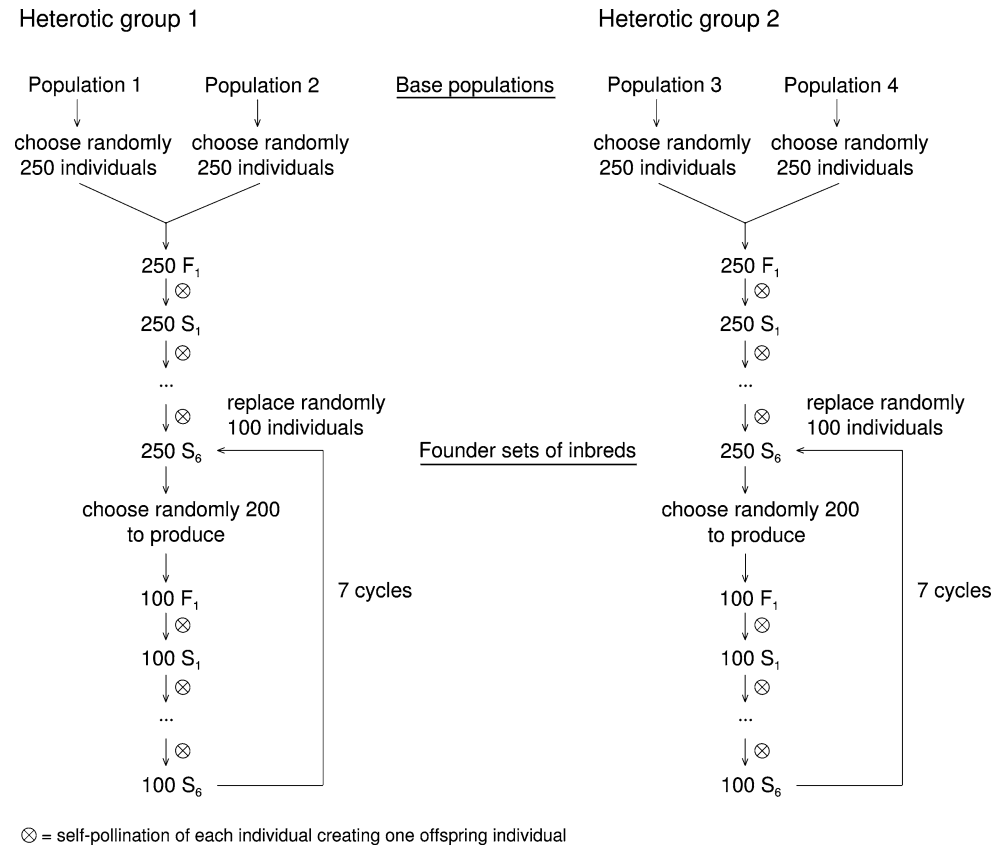
Simulations

Detailed surveys about the breeding history of the Central European elite maize germplasm reported in two companion studies (Reif et al. 2005; Stich et al. 2005) provided the basis for choosing realistic assumptions and parameters in the simulations described below. Because the results of Stich et al. (2005) indicated that selection was an unlikely cause of LD, it was neglected in our simulations. Furthermore, our simulations were based on two existing AFLP linkage maps with 722 or 1925 loci with map positions, according to a proprietary integrated map (Peleman et al. 2000). In the first case, the map positions of the 722 AFLPs are identical to those used to genotype the set of inbred lines described below.

Modeling the breeding history

Hybrid maize breeding was initiated in Europe in the 1950s. Therefore, a breeding period of 55 years was assumed in our simulations (Fig. 2). For hybrid maize breeding in Central Europe two heterotic groups with similar breeding histories, flint and dent, are used as parental pools. Thus, we simulated the breeding history for two distinct heterotic groups. The modified Rogers' distance (Wright 1978) between them was assumed to be 0.50, based on experimental data from the current flint and dent heterotic groups analyzed in a companion study (Stich et al. 2005). The predominant ancestor inbreds F2, F7, and EP1 of the European flint heterotic group trace back to two European open-pollinated varieties Lacaune and Lizargarote (Cartea et al. 1999). Hence, for each heterotic group, we sampled 250 individuals out of each of two base populations in

Fig. 2 Model for simulating 55 years of breeding history in the European maize germplasm groups



Hardy–Weinberg and linkage equilibria. Different base populations were used for the two heterotic groups. The modified Rogers' distance between base populations of each heterotic group was assumed to be 0.30, based on experimental estimates between European open-pollinated varieties (Reif et al. 2005). The average modified Rogers' distance between individuals from the same base population was assumed to be 0.45.

For each heterotic group, 250 F_1 individuals were produced by crossing each of the 250 individuals chosen from each base population with one individual from the other base population. Subsequently, the F_1 individuals were selfed for six generations, resulting for each heterotic group in 250 S_6 individuals that formed the founder set of inbred lines for the next breeding cycle. The recycling breeding following the establishment of the founder set of inbreds was simulated such that for each heterotic group a total of 200 inbreds was randomly drawn out of the founder set. These 200 inbreds were crossed pairwise to generate 100 F_1 individuals. All 100 F_1 individuals were self-pollinated for six generations, resulting for each heterotic group in 100 S_6 individuals, which replaced 100 randomly chosen inbreds of the founder set of inbreds. This scheme of crossing, selfing, and replacing lines of the founder set was repeated seven times for each heterotic group.

Population structure of inbred lines examined with QIPDT and LRRT

After completing seven cycles of recycling breeding, the 500 inbred lines from the founder sets of both heterotic groups were designated as current elite inbreds. These inbreds, together with their parental inbreds, formed the nuclear families. Because all offspring inbreds can be traced back to common ancestors, the independence assumption is violated. The criterion employed to warrant an approximative stochastic independency of D_k estimates was that nuclear families having at least one common grandparental inbred were combined to extended pedigrees. By applying this criterion, we obtained 75 independent extended pedigrees. Groups of 25 and 75 independent nuclear families were generated by sampling nuclear families from the 75 independent extended pedigrees. The offspring inbreds of the 75 independent nuclear families were examined with LRRT.

Definition of phenotypic values and estimated parameters

One, 10, and 50 AFLP markers sampled at random from the linkage map were defined as QTL with equal

genetic effects. At each QTL, one allele was assigned the genotypic effect one and the other the genotypic effect zero. The genotypic value of each inbred was calculated by summing up the effects of the individual alleles. The phenotypic value used for association tests was generated by adding a normally distributed variable $N(0, \sigma_E^2)$ to the genotypic value. The error variance was then calculated as $\sigma_E^2 = \sigma_G^2/h^2 - \sigma_G^2$, where σ_G^2 denotes the genotypic variance estimated in the group of offspring inbreds, and h^2 denotes the corresponding heritability. We examined h^2 values of 0.50, 0.75, and 1.00.

The proportion $1 - \beta^*$ of significant associations detected between a QTL and at least one adjacent AFLP marker, further designated as the power of QTL detection, was determined for both QIPDT and LRRT. For scenarios with more than one QTL, the average of the power estimates observed for each individual QTL was calculated. Furthermore, we estimated the rate of false positives, α^* , as the proportion of significant associations observed between a QTL and a marker locus, when at least one marker located between them showed a nonsignificant association with the trait.

One-QTL-one-marker simulations

In addition to the simulations based on the breeding history, simulations solely based on one QTL and one marker locus were performed for the QIPDT to vary exclusively selected parameters influencing the power of QTL detection. In the set of 75 pairs of parental inbreds used to generate 75 F_1 individuals, both loci were assumed to have an allele frequency of 0.5 and to be tightly linked (0.000001 cM). The F_1 individuals were selfed for six generations. The 75 pairs of parental inbreds and the 75 descendant S_6 individuals were used to estimate the power of the QIPDT to detect QTL if (1) LD between QTL and marker locus, measured as r^2 (Hill and Robertson 1968), ranged from 0 to 0.5, and (2) h^2 estimates of the phenotypic trait were 0.50, 0.75, or 1.00. The QTL explained 10% of the genotypic variance.

Application of QIPDT to data from elite maize inbreds

A cross-section of 49 elite European maize inbreds was examined in this study. The inbred lines belong to four independent extended pedigrees (Table 1) and represent a subset of the materials investigated by Heckenberger et al. (2005a, b). In total, 722 polymorphic AFLP markers were scored by Keygene N.V. (P.O.

Box 216, 6700 AE Wageningen, The Netherlands), using 20 primer combinations described by Heckenberger et al. (2005b). Map positions of the AFLP markers were obtained from a proprietary integrated map (Peleman et al. 2000). Field experiments with the inbreds were conducted in 2000 and 2001 at two locations in South Germany, with two replications per location, as described in detail by Heckenberger et al. (2005a). Briefly, adjusted-entry means were calculated for each environment for days to anthesis. Heritability was calculated on an entry-mean basis using software Plabstat (Utz 2001). The average of adjusted entry means over environments was used for the QIPDT. To avoid a reduction of the power to detect QTL, we did not correct for the multiple test problem. All simulations and calculations for QIPDT were performed with Plabsoft (Maurer et al. 2004), which is implemented as an extension of the statistical software R (R Development Core Team 2004).

Results

Simulations based on breeding history

In the LRRT of the 75 independent inbred lines, which corresponds to 54 150 AFLP data points, the power to detect QTL increased with decreasing number of QTL, and ranged from 0.126 (50 QTL) to 0.280 (1 QTL) for $h^2 = 0.50$ (Table 2). The corresponding estimates of α^* surpassed the nominal α level of 0.05 throughout all

Table 1 Pedigree structure of the examined 49 inbred lines

Independent pedigree ^a	Nuclear family	Parents ^b	Offspring ^b
A	1	L61, L87	L83
B	1	L54, L66	L11
C	1	L55, L56	L5, L6, L7
	2	L55, L59	L8, L9, L10
	3	L59, L74	L2, L3, L4
D	1	L24, L49	L46
	2	L28, L39	L41
	3	L28, L49	L42
	4	L33, L49	L44
	5	L36, L49	L43
	6	L49, L65	L21
	7	L51, L65	L25, L29
	8	L64, L65	L28
	9	L65, L67	L33, L34
	10	L65, L68	L32, L35, L36, L37, L38
	11	L65, L72	L39
	12	L81, L82	L17

^a Inbreds from independent pedigrees have no common grandparents

^b Notation refers to the notation of Stich et al. (2005, 2006)

scenarios examined. Increasing h^2 from 0.50 to 1.00 resulted in a minor increase in the power to detect QTL for all numbers of QTL examined.

In QIPDT, the power of QTL detection with 25 independent nuclear families (54 150 AFLP data points) was just above half as high as in the LRRT with 75 independent inbreds (Table 2). Compared with the latter, the power of QIPDT was still lower with 75 independent nuclear families (162 450 AFLP data points), but higher with 75 independent extended pedigrees, which corresponds to 313 348 AFLP data points. With 200 independent nuclear families, the QIPDT exceeded the power of the LRRT with 75 independent inbreds by 50–120%. In QIPDT with 75 independent nuclear families, increasing the marker density from 722 to 1925 AFLP loci resulted in a considerable increase in $1 - \beta^*$ for small numbers of QTL (1 and 10 QTL). However, only marginal increases were observed for 50 QTL, irrespective of the heritability. When increasing h^2 from 0.50 to 1.00, a substantial increase in the power to detect QTL was observed only for high numbers of examined nuclear families or extended pedigrees. In contrast to the LRRT, the QIPDT adhered in all scenarios except for the 200 independent nuclear families to the nominal α level.

One-QTL-one-marker simulations

In the one-QTL-one-marker simulations, the power $1 - \beta^*$ of the QIPDT increased almost linearly with

increasing LD between the marker locus and the QTL (Fig. 3). The slope of the increase was dependent on the h^2 estimate. Hence, the largest differences between estimates of $1 - \beta^*$ for different values of h^2 were found for r^2 values between 0.3 and 0.6.

Application of QIPDT to flowering data of elite maize inbreds

For the set of 49 European maize inbreds, days to anthesis showed a highly significant ($P < 0.01$) genotypic variance, and a h^2 estimate of 0.97. Entry means across environments for the 29 offspring inbreds ranged from 69.0 to 87.3 days. The QIPDT detected one AFLP marker mapping to bin 5.03 as significantly ($P < 0.05$) associated with days to anthesis (Fig. 4).

Discussion

To our knowledge, the simulation study of Pritchard et al. (2000b) is the only investigation comparing the LRRT with the classical family-based association test TDT. However, their simulations are based on a standard human coalescent model assuming random mating. Because this mating system is the exception rather than the rule in plant breeding, the results of Pritchard et al. (2000b) do not apply to most situations in plant breeding (Bernardo 2002). We based computer simulations on the 55 years of hybrid maize breeding in

Table 2 Power of QTL detection ($1 - \beta^*$) and rate of false positives (α^*) for the quantitative inbred pedigree disequilibrium test (QIPDT) and logistic regression ratio test (LRRT) at a nominal α level of 0.05 for different simulation scenarios

h^2	722 AFLP loci										1925 AFLP loci	
	LRRT		QIPDT						QIPDT			
	75 independent inbreds		25 independent nuclear families		75 independent nuclear families		75 independent extended pedigrees		200 independent nuclear families		75 independent nuclear families	
	$1 - \beta^*$	α^*	$1 - \beta^*$	α^*	$1 - \beta^*$	α^*	$1 - \beta^*$	α^*	$1 - \beta^*$	α^*	$1 - \beta^*$	α^*
1 locus coding for a trait												
1.00	0.300	0.054	0.140	0.046	0.240	0.050	0.440	0.049	0.760	0.056	0.820	0.052
0.75	0.280	0.055	0.140	0.046	0.180	0.051	0.340	0.048	0.740	0.054	0.800	0.051
0.50	0.280	0.053	0.120	0.045	0.160	0.049	0.300	0.048	0.680	0.053	0.760	0.049
10 loci coding for a trait												
1.00	0.146	0.056	0.074	0.043	0.112	0.050	0.152	0.052	0.470	0.054	0.372	0.050
0.75	0.142	0.056	0.074	0.042	0.106	0.049	0.146	0.051	0.450	0.054	0.332	0.049
0.50	0.134	0.054	0.074	0.041	0.102	0.049	0.138	0.051	0.304	0.052	0.218	0.048
50 loci coding for a trait												
1.00	0.144	0.057	0.081	0.042	0.116	0.051	0.147	0.048	0.239	0.053	0.152	0.050
0.75	0.135	0.056	0.082	0.041	0.112	0.050	0.142	0.048	0.223	0.052	0.149	0.049
0.50	0.126	0.053	0.082	0.040	0.102	0.050	0.110	0.047	0.174	0.050	0.122	0.050
Number of phenotyped/genotyped inbreds (number of AFLP data points)												
	75/75 (54 150)		25/75 (54 150)		75/225 (162 450)		248/434 (313 348)		200/600 (433 200)		75/225 (433 125)	

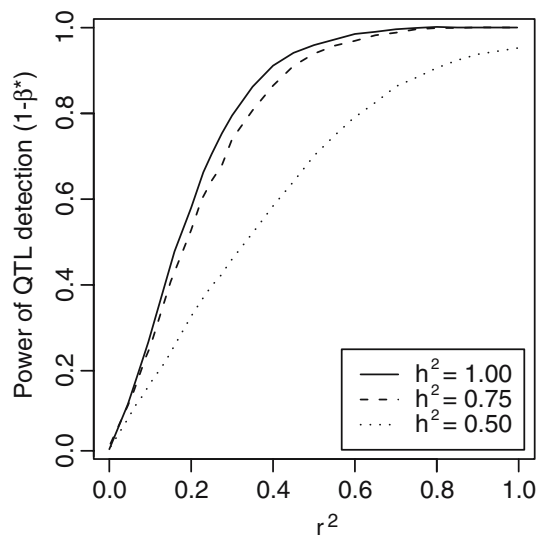


Fig. 3 Power of QTL detection ($1 - \beta^*$) by the quantitative inbred pedigree disequilibrium test (QIPDT) for different heritabilities (h^2) as a function of the extent of LD (r^2) between the marker locus and the QTL using one-QTL-one-marker simulations. The number of nuclear families was 75 and the allele frequencies of the two tightly linked loci were 0.5. The QTL explained 10% of the genotypic variance

Central Europe, to approach a comparison of family-based with population-based association mapping in a scenario in which LD is generated by the same forces as those observed for European maize breeding populations (Fig. 2).

Statistical assumptions for LRRT and QIPDT

LRRT requires independent genotypes whereas QIPDT requires independent nuclear families or extended pedigrees. For both tests ignoring the assumption of independence, results in α level higher than the nominal one (Pritchard and Rosenberg 1999; Monks and Kaplan 2000). For LRRT, this result was observed by Thornsberry et al. (2001) when applying this test to a worldwide sample of partially related maize inbred

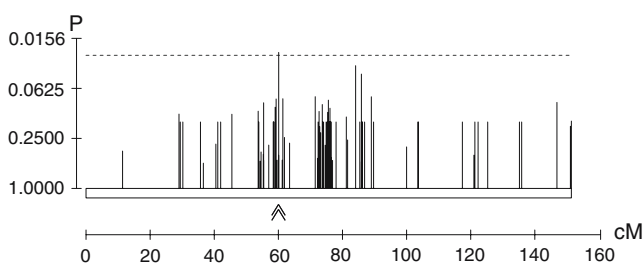


Fig. 4 Distribution of the P values of the AFLP-days to anthesis associations on chromosome 5 detected in European elite maize inbreds using the QIPDT

lines. The rejection rate at SSR markers, for which no association with the QTL was expected was higher (0.081) than the nominal α level of 0.05. The same observation is expected for the QIPDT because treating dependent extended pedigrees as independent causes an underestimation of the variance of the D_k estimates in the denominator of the test statistic T . However, in plant breeding populations, all inbreds trace back to common ancestors and, thus, only approximate independence can be achieved with any criterion.

In a companion study, Stich et al. (2005) observed a significant reduction in the extent of LD caused by relatedness when examining only individuals with no common grandparents. This criterion was chosen for the LRRT to define independent inbreds, and for the QIPDT to define independent extended pedigrees or nuclear families. The observation that the QIPDT adhered to the nominal α level suggested that the applied criterion warrants sufficient independency of extended pedigrees and nuclear families. Nevertheless, further research is needed to gain a detailed knowledge on the required independency of extended pedigrees and nuclear families in populations with a breeding history different from the described one.

Comparison of LRRT with QIPDT

In every plant breeding program, large numbers of inbred lines with known pedigree relationships are routinely phenotyped with high efforts as a basis for selection decisions. Our basic idea is to complement this information with genotypic data and perform association mapping. Here, two cases must be distinguished.

If genotypic data are not routinely collected and must be generated anew, candidate-gene but also genome-wide association mapping approaches are possible. For association mapping using a candidate-gene approach, the QIPDT is recommended because the LRRT requires a substantially higher genotyping effort to examine the same number of inbred lines. This is because an independent genome-wide marker set is additionally needed to reliably uncover the population structure. This argument does not apply to association mapping via a genome-wide approach, because one genome-wide marker data set is sufficient to (1) estimate the population structure and (2) map QTL with the LRRT (Pritchard et al. 2000b). With a given budget for genotyping a fixed number of lines (e.g., $N = 75$), the power $1 - \beta^*$ showed higher estimates for LRRT, examining 75 independent inbreds, than for QIPDT, examining 25 independent nuclear families

(Table 2). This result clearly indicated the superiority of the LRRT for genome-wide association mapping if genotypic data must be newly generated and a large number of independent inbred lines is available.

In contrast to the situation described above, in many maize breeding programs genotypic data from genome-wide distributed marker loci are collected routinely for various reasons, such as plant variety protection or choice of parents for establishing new base populations. In addition, a large proportion of these inbred lines are full sibs or half sibs of independent parental inbreds that were also genotyped. Consequently, extended pedigrees as the basis of the QIPDT reflect the typical data structure available from plant breeding programs. This outcome allows exploitation of routinely collected data for the purpose of QTL detection. On the other hand, the LRRT requires independent genotypes. Therefore, we compared QIPDT and LRRT based on 75 independent extended pedigrees and 75 independent inbreds, respectively. While the type I error rate of LRRT surpassed the nominal α level of 0.05, the QIPDT adhered to it and yielded higher estimates of $1 - \beta^*$ in all scenarios except in one. These results suggest that the QIPDT is superior to the LRRT for genome-wide association mapping if data collected routinely in plant breeding programs are available.

In contrast to the LRRT, dependent inbreds can be included in association mapping approaches using the unified mixed model method, recently described by Yu et al. (2006), without increasing the type I error rate. Therefore, a higher power to detect QTL is expected for the unified mixed model method than for the LRRT when using data from plant breeding programs. However, a comparison of the former and the QIPDT with respect to their power to detect QTL in a plant breeding context is still lacking. In contrast to QIPDT, the unified mixed model method is no test of association in the presence of linkage. Therefore, based on the various forces generating LD in plant breeding programs (Stich et al. 2005), an increased type I error rate is expected for the latter when examining data from elite breeding programs.

Factors influencing the power of QIPDT in QTL detection

The power $1 - \beta^*$ of QIPDT is influenced by (1) the genetic architecture of the trait, (2) the number of independent nuclear families or extended pedigrees, (3) the genetic map distance between QTL and marker loci, (4) the extent of LD between QTL and marker loci, (5) the heritability of the trait examined, (6) the

allele frequencies of QTL and marker loci, and (7) selection for the trait under consideration in plant breeding programs.

For QTL with equal effects, the power $1 - \beta^*$ increases with decreasing number of QTL coding for a trait, because the proportion of variance explained by a single QTL increases. Thus, for QTL with unequal effect the power $1 - \beta^*$ is expected to be higher for loci explaining a large proportion of the variance than for loci explaining a small proportion of the variance.

For previous association mapping studies in plants about 90 lines were examined (e.g., Thornsberry et al. 2001). A data set of comparable size with 75 nuclear families is expected to be available from any ordinary breeding program. The QIPDT of such a data set resulted in a very low power ($1 - \beta^* = 0.112$) for complex traits (50 QTL, $h^2 = 0.75$). This result is in accordance with findings for classical QTL mapping (Schön et al. 2004). The QTL detected for grain yield, using 122 families evaluated in 19 environments ($h^2 = 0.64$), explained about 10% of the genotypic variance. Under the assumption of QTL with equal genetic effects this is equivalent to an $1 - \beta^*$ estimate of 0.100. Examining 200 instead of 75 independent nuclear families with QIPDT almost doubled the power to detect QTL for complex traits (50 QTL). However, the proportion of QTL detected was still too low to be used for marker-assisted selection (Schön et al. 2004). Nevertheless, for mono- and oligogenic traits (≤ 10 QTL) the power of QIPDT examining 200 independent nuclear families was sufficiently high, so that 75% of 1 QTL and 50% of 10 QTL could be detected. Our results therefore give first guidelines for the necessary population size required in the case of mono- and oligogenic traits. Marker-assisted selection for such traits is particularly appealing if collection of phenotypic data is difficult and/or expensive, which is the case for many resistance and quality traits (Young 1999).

An increased power $1 - \beta^*$ was observed with an increased marker density (Table 2). This is attributable to the increased probability of substantial LD between QTL and marker loci as well as to the reduced recombination frequency between these loci when generating the offspring inbreds from the cross of the parental inbreds. Despite a comparable genotyping effort, the estimates of $1 - \beta^*$ observed for the scenarios with 10 and 50 QTL and 200 independent nuclear families genotyped with 722 AFLPs were higher than those of the 75 independent nuclear families genotyped with 1925 AFLPs. This finding corroborates the conclusion of our previous study (Stich et al. 2006) that an AFLP map with an average marker spacing of

3 cM should be sufficient for genome-wide association mapping in European elite maize germplasm.

Increasing h^2 from 0.50 to 1.00 led in the one-QTL-one-marker simulations to a substantial increase in the power $1 - \beta^*$ of QIPDT only for r^2 values between 0.3 and 0.6 (Fig. 3). This observation suggested that increasing the power to detect QTL by increasing the h^2 estimates of a phenotypic trait is only a promising approach for studies applying a moderate marker density.

The QIPDT has a higher power to detect QTL if both QTL and marker locus show allele frequencies of 0.5. The reason being the minimized probability that some QTL or marker haplotypes have only a very small class size. However, the allele frequency is in contrast to the above factors inappropriate to increase the $1 - \beta^*$ level of association mapping approaches for plant breeding populations.

Selection in plant breeding programs leads to fast genetic fixation of the favorable allele at those QTL which explain a large proportion of the genotypic variance. These fixed QTL cannot be detected in segregating progenies of elite crosses. However, even in elite materials significant genetic variance is still found that could be further analyzed by the QIPDT, which is also valid in populations that have undergone selection. In such populations, too, the QTL alleles are transmitted under the null hypothesis within each extended pedigree with equal probability to inbreds with high or low phenotypic values, as assumed in the definition of the test statistic.

Extension of QIPDT to multiallelic markers

The simulations of the present study as well as the application to a data set of maize were based on AFLP markers. Because in some maize breeding programs inbreds are routinely genotyped with multiallelic markers, we propose an extension of the QIPDT that is commonly used in multiallelic versions of related tests. A global test can be constructed by summing the squared QIPDT statistics for each of the alleles and multiplying the sum by $(m - 1)/m$, where m is the number of marker alleles (Martin et al. 2000). A statistical test for the null hypothesis can be determined by comparison to a chi-square distribution with $m - 1$ degrees of freedom.

Application of QIPDT to a data set from maize

For the application of the QIPDT to the data set comprising 49 European elite maize inbreds, a low power to detect QTL was expected because low $1 - \beta^*$

estimates were observed in the simulations based on the breeding history for data sets of comparable size (25 nuclear families). Nevertheless, we detected one AFLP marker significantly ($P < 0.05$) associated with days to anthesis. The identified QTL region was located at a position for which a consensus QTL was detected in the meta-analysis of Chardon et al. (2004). This QTL provides strong support for the significant association detected in the present study. Furthermore, based on the local maize-rice synteny conservation a flowering time gene was predicted at the QTL region identified in the current study (Chardon et al., 2004). In conclusion, the results of this pilot study suggest that the QIPDT is a promising QTL detection method for data routinely collected in plant breeding programs.

Acknowledgments The molecular marker analyses and field trials of this research were supported by “Gemeinschaft zur Förderung der privaten deutschen Pflanzenzüchtung” (GFP, Grant No. A 29/98 HS (97 HS 027)), Germany, the European Union, Grant No. QLK-CT-1999-01499, and Keygene N.V., The Netherlands. Financial support for B. Stich was provided by a grant from the German National Academic Foundation. Financial support for M. Heckenberger and H.P. Maurer was provided by grant 0313126B “GABI-BRAIN” within the framework “Genome Analysis of the Plant Biological System” supported by the German Federal Ministry of Education, Research and Technology. The authors are grateful to E. S. Buckler for helpful suggestions and comments on the manuscript. We appreciate the editorial work of J. Muminović, whose suggestions considerably improved the style of the manuscript. The authors thank three anonymous reviewers for their valuable suggestions.

References

- Abecasis GR, Cardon LR, Cookson WO (2000) A general test of association for quantitative traits in nuclear families. *Am J Hum Genet* 66:279–292
- Bernardo R (2002) Breeding for quantitative traits in plants. Stemma Press, Woodbury, MN, p 24
- 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
- Chardon F, Virlon B, Moreau L, Joets J, Decousset L, Murigneux A, Charcosset A (2004) Genetic architecture of flowering time in maize as inferred from quantitative trait loci meta-analysis and synteny conservation with the rice genome. *Genetics* 168:2169–2185
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164:1567–1587
- Heckenberger M, Bohn M, Klein D, Melchinger AE (2005a) Identification of essentially derived varieties obtained from biparental crosses of homozygous lines. II. Morphological distances and heterosis in comparison with simple sequence repeat and amplified fragment length polymorphism data in maize. *Crop Sci* 45:1132–1140
- Heckenberger M, Muminović J, Rouppe van der Voort J, Pelleman J, Bohn M, Melchinger AE (2005b) Identification of essentially derived varieties obtained from biparental

- crosses of homozygous lines. III. AFLP data from maize inbreds and comparison with SSR data. *Mol Breeding* 17:111–125
- Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. *Theor Appl Genet* 38:226–231
- Jansen RC, Jannink J-L, Beavis WD (2003) Mapping quantitative trait loci in plant breeding populations: use of parental haplotype sharing. *Crop Sci* 43:829–834
- Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121:185–199
- Martin ER, Monks SA, Warren LL, Kaplan NL (2000) A test for linkage and association in general pedigrees: the pedigree disequilibrium test. *Am J Hum Genet* 67:146–154
- Martin ER, Bass MP, Kaplan NL (2001) Correction for a potential bias in the pedigree disequilibrium test. *Am J Hum Genet* 68:1065–1067
- Maurer HP, Melchinger AE, Frisch M (2004) Plabsoft: software for simulation and data analysis in plant breeding. Poster abstract XVIIth EUCARPIA general congress 2004, Tulln, Austria
- Monks SS, Kaplan NL (2000) Removing the sampling restriction from family-based tests of association for a quantitative-trait locus. *Am J Hum Genet* 66:576–592
- Parisseeaux B, Bernardo R (2004) In silico mapping of quantitative trait loci in maize. *Theor Appl Genet* 109:508–514
- Peleman J, van Wijk R, van Oeveren J, van Schaik R (2000) Linkage map integration: an integrated genetic map of *Zea mays* L. Poster abstract. VIII. Plant & Animal Genome Conference
- Pritchard JK, Rosenberg NA (1999) Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet* 65:220–228
- Pritchard JK, Stephens M, Donnelly P (2000a) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000b) Association mapping in structured populations. *Am J Hum Genet* 67:170–181
- 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 (2005) Genetic structure and diversity of European flint maize populations determined with SSR analyses of individuals and bulks. *Theor Appl Genet* 111:906–913
- Schön CC, Utz HF, Groh S, Truberg B, Openshaw S, Melchinger AE (2004) Quantitative trait locus mapping based on re-sampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. *Genetics* 167:485–498
- Spielman RS, McGinnes RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 52:506–516
- Stich B, Melchinger AE, Frisch M, Maurer HP, Heckenberger M, Reif JC (2005) Linkage disequilibrium in elite European maize germplasm as inferred from SSR markers. *Theor Appl Genet* 111:723–730
- Stich B, Maurer HP, Melchinger AE, Frisch M, Heckenberger M, Rouppe van der Voort J, Peleman J, Sørensen AP, Reif JC (2006) Comparison of linkage disequilibrium in elite European maize inbred lines using AFLP and SSR markers. *Mol Breeding* 17:217–226
- 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
- Utz HF (2001) Plabstat, a computer program for statistical analyses of plant breeding experiments, version 2P (in German). University of Hohenheim, Germany
- Whitt SR, Buckler ES (2003) Using natural allelic diversity to evaluate gene function. In: Grotewald E (ed) *Plant functional genomics: methods and protocols*. Humana Press, Clifton, NJ, pp 123–139
- Wright S (1978) *Evolution and genetics of populations*. The University of Chicago Press, Chicago
- Young ND (1999) A cautiously optimistic vision for marker-assisted breeding. *Mol Breeding* 5:505–510
- Yu J, Pressoir G, Briggs WH, Bi IV, Yamsaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* 38:203–208
- 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