Theor Appl Genet (2007) 114:765–776
DOI 10.1007/s00122-006-0443-7

ORIGINAL PAPER

Temporal changes in allele frequencies in two European F_2 flint maize populations under modified recurrent full-sib selection

K. C. Falke · C. Flachenecker · A. E. Melchinger · H.-P. Piepho · H. P. Maurer · M. Frisch

Received: 3 August 2006 / Accepted: 23 October 2006 / Published online: 16 February 2007

Abstract Selection and random genetic drift are the two main forces affecting the selection response of recurrent selection (RS) programs by changes in allele frequencies. Therefore, detailed knowledge on allele frequency changes attributable to these forces is of fundamental importance for assessing RS programs. The objectives of our study were to (1) estimate the number, position, and genetic effect of quantitative trait loci (QTL) for selection index and its components in the base populations, (2) determine changes in allele frequencies of QTL regions due to the effects of random genetic drift and selection, and (3) predict allele frequency changes by using QTL results and compare these predictions with observed values. We performed QTL analyses, based on restriction fragment length polymorphisms (RFLPs) and simple sequence repeats (SSRs), in 274 $F_{2:3}$ lines of cross KW1265 × D146 (A × B) and 133 $F_{3:4}$ lines of cross D145 × KW1292 (C × D) originating from two European flint maize populations. Four (A × B) and seven (C × D) cycles of RS were analyzed with SSRs for significant allele frequency changes due to selection. Several QTL regions for selection index were detected with simple and composite interval mapping. In some of them, flanking markers showed a significant allele frequency change after the first and the final selection cycles. The correlation between observed and predicted allele frequencies was significant only in A × B. We attribute these observations mainly to (1) the high dependence of the power of QTL detection on the population size and (2) the occurrence of undetectable QTL in repulsion phase. Assessment of allele frequency changes in RS programs can be used to detect marker alleles linked to QTL regions under selection pressure.

Keywords Allele frequency changes · Random genetic drift · Recurrent selection · SSR · *Zea mays* L.

Communicated by M. Xu.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-006-0443-7) contains supplementary material, which is available to authorized users.

K. C. Falke · C. Flachenecker · A. E. Melchinger (✉) · H. P. Maurer · M. Frisch
Institute of 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

Introduction

Recurrent selection (RS) is a cyclical breeding method extensively used to improve breeding populations. For grain yield in maize, the selection response achieved with RS ranged between 2 and 7% per cycle (Hallauer and Miranda 1988). Application of RS aims at gradually increasing the frequency of favorable alleles while maintaining the genetic variability in the population (Hallauer 1985). Two main forces affecting the selection response in RS programs are selection and random genetic drift. Selection increases the frequencies of favorable alleles while genetic drift is a random change in allele frequencies due to small population size. A loss of favorable alleles due to random genetic drift leads to a reduction in genetic variance and, thus, limits

future selection response (Guzman and Lamkey 1999, 2000). The assessment of the effects of random genetic drift and selection is important for designing efficient RS programs. Several empirical studies investigated the selection response of RS at the phenotypic level with quantitative-genetic methods (Smith 1979) or simulation studies (Hospital and Chevalet 1996), but information about the effects of selection and random genetic drift in RS programs at the molecular level is still scarce.

In several studies, isozymes (Brown and Allard 1971; Stuber et al. 1980; Kahler 1983) and molecular markers (Heredia-Diaz et al. 1996; Labate et al. 1999; Pinto et al. 2003; Coque and Gallais 2006) were used to examine genetic changes in maize populations undergoing selection. Most of these studies applied standard statistical tests (e.g. χ^2, G tests) for assessing significant changes in allele frequencies. However, these tests neglect the effects of random genetic drift and are, therefore, not appropriate for the analysis of changes in allele frequencies in RS with finite population size. In contrast, Waples (1989) provided a test statistic for monitoring allele frequency changes, which takes into account the increased variance in allele frequencies between generations caused by random genetic drift. Up to now, Waples' neutrality test (1989) has been used in evolutionary research (Queney et al. 2000; Charbonnel et al. 2005), but less in plant breeding (Labate et al. 1999; Pinto et al. 2003; Coque and Gallais 2006). Hence, a detailed evaluation of allele frequency changes after one as well as after several RS cycles is still lacking.

Detection of quantitative trait loci (QTL) that control the variability of complex traits of interest are mostly employed in marker-assisted selection (MAS) for the detected QTL. According to theoretical results (Lande and Thompson 1990), MAS should be superior to conventional phenotypic selection for traits that show low heritability or are difficult and expensive to evaluate phenotypically. Alternatively, QTL estimates can be used for the prediction of directional changes in allele frequencies (Δp) (Hartl and Clark 1997). The predicted Δp can be compared with changes in allele frequencies observed under conventional phenotypic selection. An effective prediction of Δp with QTL estimates would be an advantage for planning and assessing adequate RS schemes. However, no studies are available which predict Δp with QTL estimates and compare this prediction with Δp observed under conventional phenotypic selection.

As complementary part of a QTL mapping project (Mihaljevic et al. 2005), a recurrent full-sib selection program was initiated in 1990 for evaluating the selection response in two European F_2 maize populations. A pseudo-factorial mating scheme of Cockerham and Burrows (1980) was applied for the recombination of candidates selected on the basis of the selection index, and pedigrees were recorded among full-sib families across all selection cycles. In three companion studies, we investigated changes in the population mean, inbreeding coefficients, as well as additive and dominance variance components (Flachenecker et al. 2006a, b), and determined the overall net effects of random genetic drift on the selection response (Flachenecker et al. 2006c).

In the present study, we evaluated the selection response of two European flint maize populations after several cycles of recurrent full-sib selection at the molecular level on the basis of simple sequence repeat (SSR) analyses. Our objectives were to (1) estimate the number, position, and genetic effect of QTL for selection index and its underlying traits in the base populations, (2) investigate allele frequency changes in QTL regions due to the effects of random genetic drift and selection, and (3) predict allele frequency changes using the information from QTL mapping, and compare these predictions with observed values to draw conclusions on the design of our RS program.

Materials and methods

QTL experiments and analyses

Plant materials

The plant materials used for this study were partly identical to those employed in previous studies (cf. Schön et al. 1994; Lübberstedt et al. 1998; Mihaljevic et al. 2005; Flachenecker et al. 2006a, b, c; Falke et al. 2006). Four early maturing homozygous European flint lines KW1265, D146, D145, and KW1292, subsequently referred to as A, B, C, and D, respectively, were used as parents. Parental lines A and D are private lines developed by KWS SAAT AG (Einbeck, Germany), B and C are public lines bred by Prof. Dr. W. G. Pollmer at the University of Hohenheim (Stuttgart, Germany). Randomly chosen F_2 and F_3 plants were selfed to produce 380 $F_{2:3}$ lines of population A × B and 140 $F_{3:4}$ lines of population C × D.

Agronomic trials

Agronomic trials and data analysis for 280 $F_{2:3}$ lines (A × B) and 135 $F_{3:4}$ lines (C × D) were reported in

detail by Mihaljevic et al. (2005). The experimental designs employed were a 30×10 (A × B) and a 15×10 (C × D) α-design (Patterson and Williams 1976) with two replications. The field trials were conducted at five (A × B) and four (C × D) sites in South Germany. Data were analyzed for the following traits: grain yield (mg ha^{-1}) adjusted to 155 g kg^{-1} grain moisture, grain moisture (g kg^{-1}), and selection index. For calculating the selection index, (1) grain yield and dry matter content were expressed in percent of the mean of F_2 check entries, and (2) relative values received a weight of 1 for grain yield and 2 for dry matter content [i.e. the weight vector was $\mathbf{b}' = (1,2)$]. ANOVAs for the field experiments were calculated with the software PLABSTAT (Utz 2001). The means across environments were subsequently employed in QTL mapping.

Marker analyses and linkage map construction

The procedures for restriction fragment length polymorphism (RFLP) assays were described by Schön et al. (1994). We employed a total of 89 and 118 RFLPs to genotype 344 $F_{2:3}$ lines (A × B), and 133 $F_{3:4}$ lines (C × D), respectively. Additionally, 146 $F_{2:3}$ lines (A × B) and 110 $F_{3:4}$ lines (C × D), which are randomly chosen subsets of the germplasm assayed with RFLPs, were genotyped with 104 (A × B) and 101 (C × D) codominant SSR markers. DNA extraction, as well as SSR amplification and detection were described in detail by Falke et al. (2006).

Segregation at each marker locus was tested by χ^2 for deviations from both Mendelian segregation ratios and an allele frequency of 0.5. The joint linkage maps for RFLPs and SSRs were constructed for the F_2 (A × B) and F_3 (C × D) generations. Due to the lower number of individuals for the SSR assays, individuals for which no SSR data were available were treated as missing values. Linkage maps were assembled by the software package JoinMap Version3.0 (Van Ooijen and Voorrips 2001) using Haldane's mapping function (Haldane 1919). An LOD threshold of 3.0 was employed in two-point analyses.

QTL analyses

In the study of Mihaljevic et al. (2005), QTL for both populations were detected with RFLP markers. In the present study, QTL analyses of 274 $F_{2:3}$ lines (A × B) and 133 $F_{3:4}$ lines (C × D) were performed by combining RFLP and SSR markers.

QTL mapping and estimation of QTL effects were conducted with means across environments by using an extension of PLABQTL (Utz and Melchinger 1996). For the analyses of data, we employed both simple interval mapping (SIM, Lander and Botstein 1989) and composite interval mapping (CIM) using a regression approach (Haley and Knott 1992). For CIM, cofactors were selected by stepwise regression (Miller 1990, p. 49) based on the Bayesian information criterion (BIC). A LOD (=0.217LR) threshold of 5.0 was chosen for declaring a putative QTL significant. The proportion of the genotypic variance explained by all QTL $\left(\hat{\sigma}_g^2\right)$ was determined as described by Utz et al. (2000). Standard five-fold cross-validation (Utz et al. 2000), as implemented in PLABQTL (Utz and Melchinger 1996) with test sets (TS) comprising 20% of the genotypes, was used for determining the effect of genotypic sampling on the genetic effects with software MATCHQTL (Utz, unpublished). 200 randomizations were generated for assigning genotypes to the respective subsamples yielding a total of 1,000 replicated cross-validation runs. Estimates of genetic effects explained by the detected QTL simultaneously were calculated for the total data set (DS) and as mean over all TS.

RS experiments and analysis of allele frequency changes

Plant materials

In both populations, A × B and C × D, F_2Syn3 generations were derived from the F_2 generation by three generations of chain crossing using 240 plants (i.e., 1×2, 2×3,..., and 240×1). The selection procedure in each selection cycle was described in detail by Flachenecker et al. (2006a, b). Briefly, four (A × B) and seven (C × D) cycles of modified recurrent full-sib selection were performed between 1994 and 2001 by using a pseudo-factorial mating scheme for recombination of the selected candidates, based on the suggestions of Cockerham and Burrows (1980). Evaluation of the full-sib families was conducted in field trials at three locations in South Germany. The experimental design was an α-lattice (10×15) with three replications.

Marker analyses

Parents of 36 families with the highest selection index were intermated to generate the next selection cycle and used for marker analyses in each selection cycle. Bulks of 15 kernels were ground and DNA was extracted using the GenElute™ Plant Genomic DNA Miniprep Kit (Sigma®). A total of 104 and 101

codominant SSR markers consistent with the QTL analyses were employed to genotype four (A × B) and seven (C × D) selection cycles.

Test for allele frequency changes

Waples' (1989) test statistic for detecting temporal variation in allele frequencies was applied for both parental alleles and non-parental alleles. Changes in allele frequencies were tested between (1) selection cycles C0 (=F_2Syn3) and C1 for both populations, (2) C0 (=F_2Syn3) and C4 for population A × B, and (3) C0 (=F_2Syn3) and C7 for population C × D. The test statistic follows a χ^2 distribution (with a single degree of freedom) and is calculated as $\chi^2 = (y_t - y_0)^2/\mathrm{var}(y_t - y_0)$, where y_t and y_0 are the allele frequencies in selection cycle Ct and C0. The derivation of $\mathrm{var}(y_t - y_0)$ depends on the sampling plan (sampling plan I: individuals are sampled after reproduction), sample size (C0 = 148; C1 = C4 = C7 = 72), the number of generations t (1, 4, and 7), the effective population size (N_e = 32), and the population size (N = 148). The null hypothesis was rejected if changes in allele frequencies between the respective cycles were significantly greater than expected by random genetic drift alone. In addition, linear regression analyses weighted by the inverse allele frequency variances of Waples' test statistic (1989) were used to determine the direction of changes in allele frequencies between selection cycles.

Prediction of changes in allele frequencies at marker loci

For the general case, allele frequency changes at each marker locus for one cycle of RS can be predicted as (Hartl and Clark 1997, p. 422; Hallauer 1985)

$$\Delta p = \sum_{j=1}^{n_{\mathrm{QTL}}} (i/\sigma_p) p_j q_j [a_j(1+F) + d_j(q_j - p_j)(1-F)](1-2r_j), \quad (1)$$

where i is the selection intensity, σ_p is the phenotypic standard deviation of the trait under consideration, p and q are the frequencies of the two parental alleles at the marker locus under investigation before selection, F is the inbreeding coefficient, a_j and d_j are the additive and dominance effect for the respective trait at the jth QTL on the chromosome of the marker locus, and r_j is the recombination frequency between the marker locus under consideration and the jth QTL. Furthermore, Eq. 1 assumes no espistasis, linkage or linkage disequilibrium.

For cycle C1 of our experiment, we applied $i = Nz/N_e$ (Cockerham and Burrows 1980), where N is the number of full-sib families tested in the respective cycle, z is the ordinate of the standard normal density at the truncation point of selection, and N_e is the effective population size that amounts to 32 based on the formula of Cockerham and Burrows (1980). We further assumed $F = 0$ (cf. Flachenecker et al. 2006a) and $p_j = q_j = 0.5$ for our F_2Syn3 base population and used the cross-validated additive effect (Utz et al. 2000) for the respective trait at the jth QTL on the chromosome of the marker locus. Inserting these values, Eq. 1 simplifies to

$$\Delta p = \frac{1}{4}\left(\frac{i}{\hat{\sigma}_p}\right)\sum_{j=1}^{n_{\mathrm{QTL}}} \hat{a}_j(1 - 2\hat{r}_j). \quad (2)$$

For cycle C1 of populations A × B and C × D, allele frequency changes were predicted with Eq. 2 and compared with observed changes in allele frequencies (Δp). All regression and correlation analyses as well as Waples' (1989) neutrality test were carried out with the statistical software R (R Development Core Team 2004).

Results

Significant deviations ($P < 0.001$) from the expected single-locus genotype frequencies were observed in zero (A × B) and 25 cases (C × D). We also detected significant deviations ($P < 0.001$) from allele frequency 0.5 for zero (A × B) and nine (C × D) markers. The 193 (A × B) and 219 (C × D) marker loci spanned map distances of 1840 cM (A × B) and 1886 cM (C × D), with respective average interval lengths of 10 cM and 9 cM. In total, seven RFLP loci in population A × B and three in C × D were scored as dominant markers.

Using SIM, we detected four QTL for selection index in population A × B and one QTL in population C × D (Table 1). A simultaneous fit of all detected QTL explained 34.6% (A × B) and 15.3% (C × D) of the genetic variance $\left(\hat{\sigma}_g^2\right)$. The QTL with the largest additive effect (in the test set) in A × B was located on chromosome 8, with the positive allele of the first parent. In A × B, we found three putative QTL for grain yield on chromosomes 8, 9, and 10. After fitting all putative QTL simultaneously, 31.6% of $\hat{\sigma}_g^2$ was explained. The QTL with the largest additive effect (in the test set) was found on chromosome 10. One QTL region on chromosome 2 (A × B) and two QTL regions on chromosome 1 (C × D) were significantly associated with grain moisture. A simultaneous fit of all QTL accounted for 8.6% (A × B) and 22.3% (C × D) of $\hat{\sigma}_g^2$.

Table 1 Putative QTL and associated genetic effects detected for selection index and its components by employing simple interval mapping (SIM) and composite interval mapping (CIM) of population A × B and C × D

Population/method/trait	Chrom.	Pos.	LOD	Genetic effects							
				Additive effect				Dominance effect			
				DS[a]	σ_{DS}[c]	TS[b]	σ_{TS}[c]	DS[a]	σ_{DS}[c]	TS[b]	σ_{TS}[c]
A × B											
SIM											
Selection index	1	160	6.16	7.29	1.642	6.58	0.090	0.74	2.290	0.02	0.173
	1	210	5.39	4.36	1.726	2.06	0.105	4.08	2.805	1.04	0.191
	8	92	8.50	–7.75	1.304	–8.45	0.094	–0.66	1.984	–0.18	0.149
	10	96	7.66	6.02	1.476	6.45	0.098	6.75	2.050	6.45	0.151
Grain yield	8	80	7.55	–4.01	0.679	–3.38	0.052	0.23	0.972	0.12	0.082
	9	78	5.24	2.46	0.759	1.30	0.050	5.12	1.120	2.79	0.054
	10	96	8.80	3.99	0.731	3.98	0.052	2.65	1.020	3.15	0.074
Grain moisture	2	114	5.35	5.90	1.207	2.94	0.050	–2.86	1.771	–1.15	0.112
CIM											
Selection index	1	158	6.04	8.07	1.459	7.84	0.103	0.57	2.123	0.26	0.162
	8	98	9.84	–9.14	1.397	–8.45	0.096	0.80	2.228	–0.07	0.153
	10	98	11.39	14.14	2.695	8.38	0.180	5.70	2.837	6.21	0.184
	10	116	5.25	–10.61	3.231	–6.68	0.292	1.51	4.144	0.35	0.322
Grain yield	1	160	6.47	3.00	0.705	2.28	0.050	0.05	1.043	0.19	0.088
	2	150	6.31	3.11	0.630	1.86	0.047	0.93	0.904	0.36	0.077
	8	80	9.15	–3.92	0.621	–3.85	0.046	0.61	0.897	0.03	0.067
	9	78	5.77	–0.34	1.245	1.21	0.102	5.55	1.502	5.16	0.107
	9	96	8.65	3.20	1.218	2.65	0.071	–1.55	1.542	1.21	0.124
	10	98	11.44	7.86	1.535	4.21	0.087	1.73	1.614	2.93	0.090
	10	112	5.26	–5.46	1.817	–2.28	0.167	0.70	2.252	–0.24	0.178
Grain moisture	1	270	5.77	–4.33	1.150	–1.87	0.085	–1.86	2.056	–0.48	0.176
	2	122	13.43	6.36	1.016	5.45	0.081	–1.33	1.368	–1.65	0.110
	3	92	6.09	–4.45	1.132	–3.25	0.083	0.30	1.570	0.23	0.125
	7	66	6.12	6.11	1.245	2.97	0.098	–2.25	1.661	–2.01	0.118
	7	102	8.73	–4.08	1.183	–1.07	0.094	–0.94	1.380	–0.77	0.119
	8	94	7.27	2.30	0.986	4.00	0.093	0.94	1.450	0.74	0.118
	8	138	5.13	3.25	1.037	2.84	0.078	–0.98	1.448	–0.17	0.127
C × D											
SIM											
Selection index	1	122	5.03	14.00	2.848	4.83	0.119	1.30	4.917	–2.11	0.341
Grain moisture	1	222	6.95	–6.99	3.259	–8.10	0.148	–0.01	5.091	–1.14	0.349
	1	236	6.85	–4.72	3.134	–5.61	0.162	–6.01	4.541	–3.97	0.283
CIM											
Selection index	1	122	10.43	15.07	2.653	13.36	0.208	0.13	4.573	1.07	0.368
	2	206	5.22	7.64	2.398	5.86	0.183	6.33	4.529	6.19	0.412
	9	108	6.42	9.31	2.431	5.76	0.185	–1.38	4.990	–0.22	0.514
Grain yield	1	120	10.06	6.89	1.285	6.40	0.099	–1.61	2.173	–0.64	0.192
	1	210	7.41	–5.22	1.165	–3.33	0.093	–1.15	2.262	–1.11	0.179
	9	106	6.61	5.46	1.277	3.04	0.092	–0.31	2.712	0.08	0.245
Grain moisture	1	136	17.79	–8.17	1.403	–6.16	0.163	0.04	2.682	–1.71	0.304
	1	238	9.65	–13.08	1.662	–9.89	0.183	–1.92	2.862	–3.95	0.285
	1	280	6.89	7.06	1.566	3.98	0.193	–0.20	2.657	2.50	0.305
	2	206	8.00	–4.23	1.507	–4.08	0.214	0.25	2.764	1.96	0.371
	4	50	5.01	–3.35	1.378	–0.33	0.200	–4.30	2.596	–2.10	0.328
	5	84	10.49	–5.90	1.461	–3.45	0.161	–1.44	2.402	–3.41	0.292
	10	52	5.05	4.40	1.336	1.00	0.171	–3.13	2.619	–2.27	0.313

[a] Genetic effects were estimated in a simultaneous fit with SIM and CIM, respectively, in the data set (DS)

[b] Mean over 1,000 test sets (TS) of the genetic effects using fivefold cross validation

[c] Standard error of the genetic effects

With CIM, we detected four QTL for selection index in population A × B on chromosomes 1, 8, and 10, and three QTL in population C × D on chromosomes 1, 2, and 9 (Table 1). A simultaneous fit of all detected QTL explained 35.7% (A × B) and 27.7% (C × D) of $\hat{\sigma}_g^2$. The QTL with the largest additive effect (in the test

set) for selection index was found on chromosome 8 (A × B) and on chromosome 1 (C × D), with the positive allele being contributed by parent A and parent D, respectively. For grain yield, we found a total of seven QTL in A × B, distributed across the genome, and three QTL on chromosomes 1 and 9 in C × D. After fitting all putative QTL simultaneously, 44.3% (A × B) and 30.8% (C × D) of $\hat{\sigma}_g^2$ were explained. The QTL with the largest LOD score and additive effect (in the test set) was found on chromosome 10 (A × B) and 1 (C × D). Seven QTL regions across the genome were significantly associated with grain moisture in both populations. A simultaneous fit of all QTL accounted for 27.7% (A × B) and 49.2% (C × D) of $\hat{\sigma}_g^2$. QTL with the largest additive effect (in the test set) were located on chromosome 2 (A × B) and on chromosome 1 (C × D).

All putative QTL regions for selection index were confirmed by its components, grain yield and grain moisture, with CIM but not with SIM. For all three traits, cross validation for the genetic effects resulted in estimates from the test set mostly considerably lower than the corresponding values from the entire data set (Table 1).

In population A × B, the maximum allele frequency at the SSR marker loci was 0.75 for allele A (originating from parent A) in C4, 0.78 for allele B (originating from parent B) in C3, and 0.48 for non-parental alleles in C4 (Table 2). In population C × D, we observed maximum frequencies of 0.83 for allele C (originating from parent C) in C5, 0.84 for allele D (originating from parent D) in C7, and 0.67 for non-parental alleles in C7. The median of the proportion of non-parental alleles at the marker loci within individuals increased from 0.02 (C1) to 0.10 (C4) in A × B, and from 0.01 (C1) to 0.10 (C7) in C × D (Table 3).

Out of 104 loci, Waples' (1989) neutrality test was significant ($P < 0.05$) in population A × B at 15 loci for parental and at nine for non-parental alleles after one cycle of RS, as well as at 16 loci for parental and at five for non-parental alleles after four cycles of RS (Fig. 1). Applying the SIM approach, we observed significant ($P < 0.05$) changes in allele frequency in QTL regions for selection index after one cycle of RS on chromosomes 8 and 10. Similar results were obtained with the CIM approach after four cycles of RS.

Using Waples' (1989) neutrality test on 101 loci in population C × D, we observed significant ($P < 0.05$) changes in allele frequencies at six loci for parental alleles after one cycle of RS, as well as at eight loci for parental and at five for non-parental alleles after seven cycles of RS (Fig. 2). CIM revealed significant ($P < 0.05$) changes in allele frequencies in QTL regions for selection index after seven cycles of RS on chromosome 1. For both populations, changes in allele frequencies for all marker loci from F_2Syn3 to the final selection cycles were presented in a supplementary table.

Observed changes in allele frequencies (from C0 to C1; Δp) were significantly ($P < 0.05$) correlated with predicted Δp for selection index in A × B but not in C × D (Fig. 3).

Discussion

In previous studies (Flachenecker et al. 2006a, b, c), we used classical quantitative genetic tools to evaluate the selection response of a modified recurrent full-sib

Table 2 Allele frequency distribution of 104 (A × B) and 101 (C × D) SSR marker loci for parental alleles p and q [p: allele A (A × B) and C (C × D); q: allele B (A × B) and D (C × D)] and non-parental alleles (v) over different selection cycles in populations A × B and C × D

Cross	Cycle	p				q				v			
		Min	Max	Mean	σ_p	Min	Max	Mean	σ_q	Min	Max	Mean	σ_v
A × B	C1	0.31	0.71	0.48	0.010	0.22	0.66	0.47	0.010	0.00	0.22	0.05	0.007
	C2	0.23	0.74	0.47	0.013	0.23	0.69	0.45	0.013	0.00	0.37	0.08	0.011
	C3	0.19	0.73	0.46	0.016	0.24	0.78	0.45	0.015	0.00	0.42	0.09	0.011
	C4	0.16	0.75	0.45	0.017	0.20	0.72	0.46	0.016	0.00	0.48	0.09	0.012
C × D	C1	0.34	0.79	0.51	0.010	0.16	0.65	0.48	0.011	0.00	0.40	0.01	0.004
	C2	0.27	0.75	0.48	0.013	0.18	0.72	0.48	0.013	0.00	0.30	0.04	0.005
	C3	0.18	0.74	0.47	0.015	0.15	0.72	0.46	0.016	0.00	0.47	0.07	0.007
	C4	0.15	0.77	0.45	0.016	0.15	0.76	0.46	0.017	0.00	0.52	0.09	0.010
	C5	0.15	0.83	0.46	0.017	0.12	0.74	0.45	0.018	0.00	0.54	0.09	0.012
	C6	0.10	0.80	0.46	0.018	0.10	0.79	0.44	0.019	0.00	0.65	0.10	0.013
	C7	0.06	0.81	0.46	0.019	0.13	0.84	0.44	0.020	0.00	0.67	0.10	0.014

Minimum (min), maximum (max), mean and standard error (σ) of allele frequencies over all marker loci

Table 3 Proportion of non-parental alleles at the marker loci within individuals over different selection cycles in population A × B and C × D

Population	Cycle	Sample size	Min	1st quartile	Median	Mean	σ	3rd quartile	Max
A × B	C1	72	0.00	0.00	0.02	0.05	0.009	0.06	0.31
	C2	72	0.00	0.02	0.06	0.08	0.008	0.15	0.20
	C3	72	0.02	0.07	0.09	0.09	0.003	0.10	0.15
	C4	72	0.02	0.07	0.10	0.09	0.003	0.11	0.15
C × D	C1	72	0.00	0.00	0.01	0.01	0.004	0.01	0.21
	C2	69	0.00	0.00	0.01	0.04	0.007	0.06	0.17
	C3	72	0.00	0.01	0.07	0.07	0.006	0.09	0.17
	C4	72	0.03	0.06	0.09	0.09	0.004	0.12	0.17
	C5	72	0.03	0.07	0.08	0.09	0.003	0.10	0.14
	C6	72	0.05	0.08	0.09	0.10	0.003	0.11	0.14
	C7	72	0.06	0.08	0.10	0.10	0.003	0.12	0.15

Minimum, maximum, first and third quartile, median and mean, and standard error (σ) of allele frequencies over all marker loci

selection scheme in two populations at the phenotypic level. We observed a relatively high increase per cycle of 5.25% (A × B) and 3.64% (C × D) for selection index and 14.07% (A × B) and 8.28% (C × D) for grain yield, combined with a decrease in grain moisture of –1.72% (A × B) and –1.77% (C × D). We expect further response in future selection cycles due to small effects of random genetic drift and no reduction in the additive variance in both populations. In the present study, we analyzed in detail the effects of selection at the molecular level by using SSR markers. This evaluation allows to (1) detect unintentional migration, and (2) separate the effects of selection from those of random genetic drift by using Waples' (1989) test for identifying changes in allele frequencies. Furthermore, the combination of QTL results and changes in allele frequencies offers the opportunity to determine genomic regions that are responsible for the selection response, and to compare predicted with observed changes in allele frequencies.

Appearance of non-parental alleles

An average proportion of non-parental alleles of 0.05 in population A × B and 0.01 in population C × D was observed in the initial selection cycles (C1) (Table 2). This contamination remained undetected during the intermating generations and the selection process in the field and, therefore, was passed on to the progenies of the following selection cycles. Thus, the average proportion of non-parental alleles increased up to ~10% in C3 of A × B and in C4 of C × D, and remained at this level during the subsequent selection cycles (Table 2). Nevertheless, the median of the proportion of non-parental alleles at marker loci within individuals increased from 0.02 (C1) to 0.10 (C4) in A × B and from 0.01 (C1) to 0.10 (C7) in C × D (Table 3). In comparison to other mating schemes, the sensitivity of non-parental alleles to selection was enhanced in this experiment due to the applied mating scheme of Cockerham and Burrows (1980). This mating scheme weights the selected progenies, giving double weight of the gametic contribution to the males compared to the females. In most instances, the non-parental allele could be identified as a parental allele of the other population. Hence, we attribute the appearance of non-parental alleles mainly to a contamination with foreign pollen (in the intermating or selfing generation), and/or experimental errors due to erroneous crossings. Further factors may be due to heterozygosity in the parental lines or recombination within a band (Bernardo et al. 2000; Bernardo 2002), whereas mutation was at best a marginal factor. Both parental alleles of the other population were not observed in individual genotypes and, therefore, a mix-up of ears can be excluded.

The contamination remained undetected at the phenotypic level, even though all plants and especially ears were controlled in the field trials. In contrast, the appearance of non-parental alleles in F_2 populations from biparental crosses, as employed in our selection programs, can easily be identified by using molecular markers. Thus, our results indicate that contaminations may also occur in other RS programs with similar selection procedures. However, open-pollinated varieties or synthetics are usually used as source populations of RS programs (Hallauer and Miranda 1988), and most contaminations will remain undetected. Contaminations with foreign pollen may be minimized by using border plots. Furthermore, the employment of molecular markers in parallel to the selection procedure is promising to identify contaminations or experimental errors easily and remove false genotypes for generating the material of the next selection cycle.

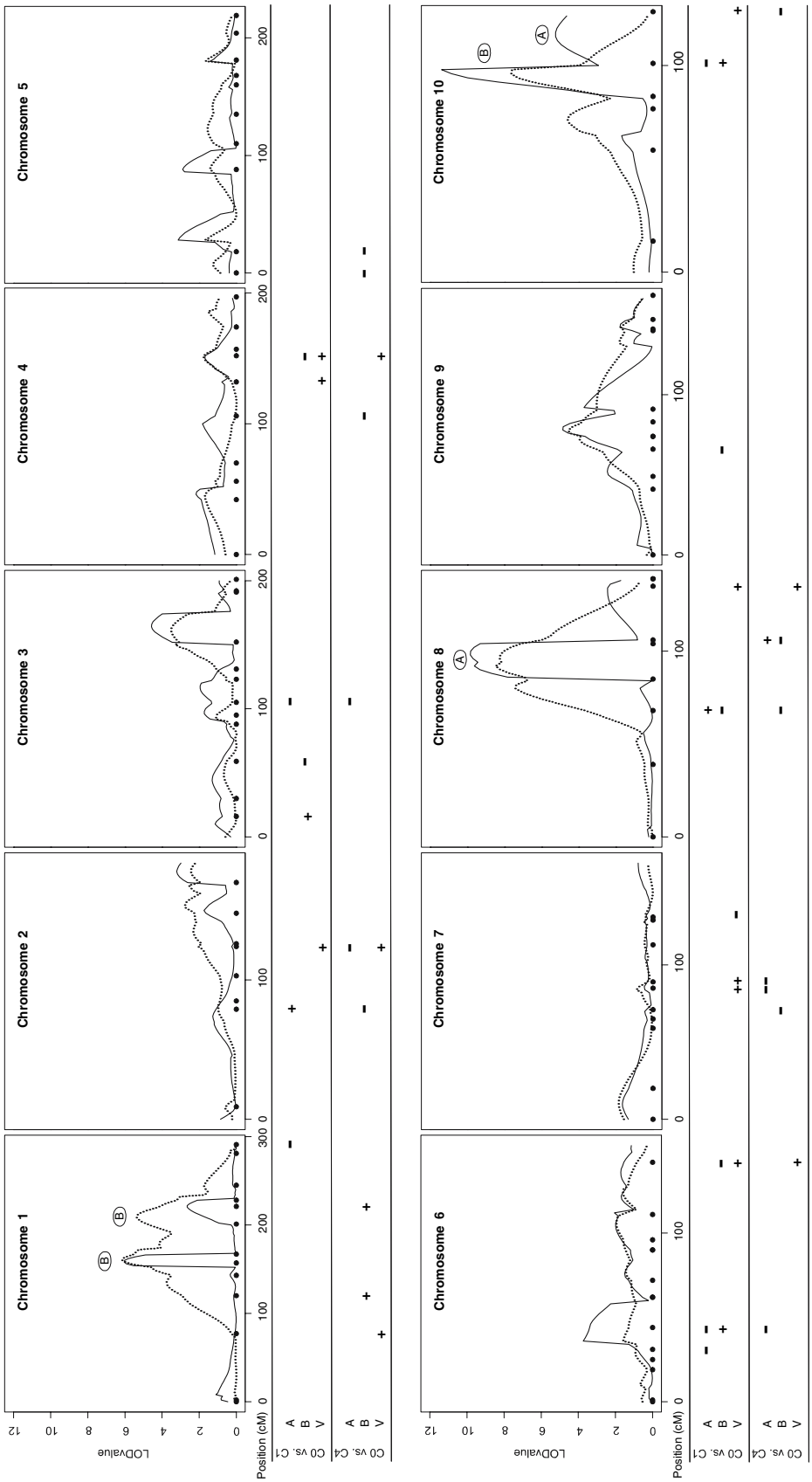

Fig. 1 QTL likelihood maps indicating LOD scores for selection index in population A × B. *Curves* represent results from simple interval mapping (SIM, *dotted line*) and composite interval mapping (CIM, *solid line*). *Letters A and B* in *ellipses* indicate which parent contributed the favorable allele. SSR markers on the genetic map are noted with *filled dots*. Significant changes in allele frequencies between C0 and C1 as well as between C0 and C4 determined with Waples' (1989) test are presented below the LOD curves for alleles A, B, and non-parental alleles (V). The *plus and minus* indicate an increase or decrease in the allele frequency of the respective marker

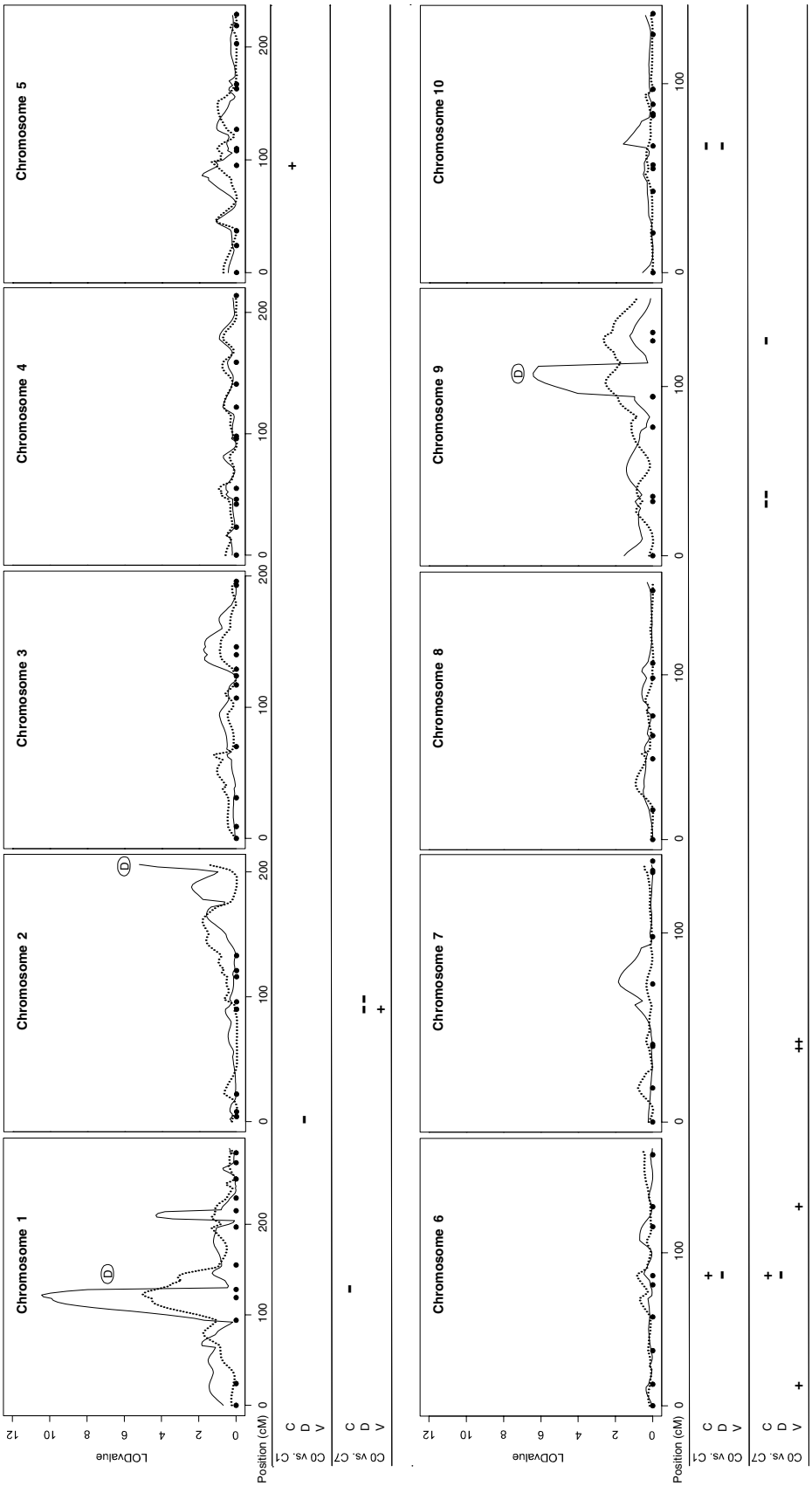

Fig. 2 QTL likelihood maps indicating LOD scores for selection index in population C × D. *Curves* represent results from simple interval mapping (SIM, *dotted line*) and composite interval mapping (CIM, *solid line*). *Letters C and D* in *ellipses* indicate which parent contributed the favorable allele. SSR markers on the genetic map are noted with *filled dots*. Significant changes in allele frequencies between C0 and C1 as well as between C0 and C7 determined with Waples' (1989) test are presented below the LOD curves for alleles C, D, and non-parental alleles (V). The *plus and minus* indicate an increase or decrease in the allele frequency of the respective marker

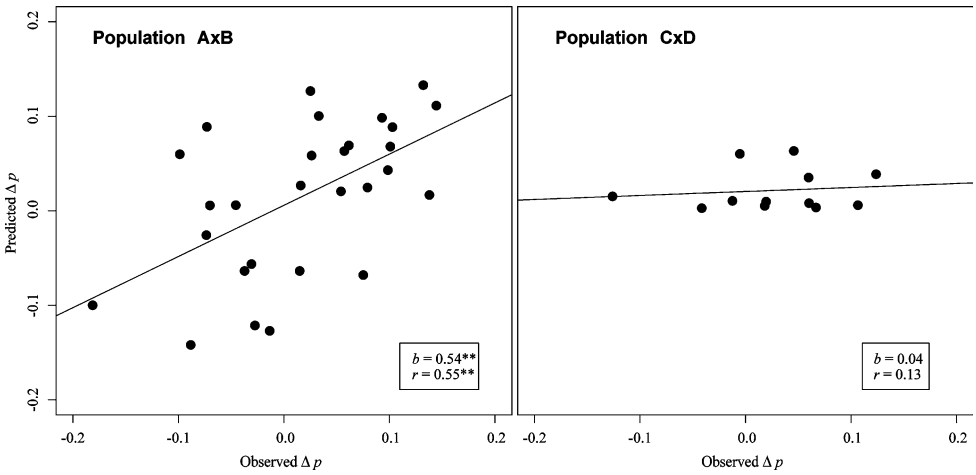

Fig. 3 Correlation between observed changes in allele frequency of the maker loci for allele A (A × B) and C (C × D) (C0–C1; Δp) and predicted Δp for the selection index calculated with the modified formula of Hartl and Clark (1997), where b is the slope coefficient and r is the correlation coefficient

Allele frequency distributions

Neither a fixation (frequency = 1.0) nor an extinction (frequency = 0.0) of the parental alleles was observed at any of the marker loci (Table 2). The observed degree of variation of parental allele frequencies was low when compared with other studies (Labate et al. 1999; Pinto et al. 2003). This result is mostly attributable to the use of an F_2 base population with intermediate allele frequencies ($p = 0.5$), as well as the moderate selection intensity and the relatively large effective population size ($N_e = 32$) in our study. Consequently, a further increase in the frequency of favorable alleles could be achieved in future selection cycles and contribute to further selection response.

Selection effects versus genetic drift

Changes in allele frequencies between selection cycles (A × B: C0 vs. C1 and C0 vs. C4; C × D: C0 vs. C1 and C0 vs. C7) determined by Waples' test (1989) were mainly attributable to the effects of random genetic drift, which is in agreement with previous studies on changes in allele frequencies (Labate et al. 1999; Pinto et al. 2003). However, significant changes in parental allele frequencies in A × B were detected at 14% of loci after one cycle of RS and at 15% after four cycles of RS. In C × D, significant changes in parental allele frequencies were revealed at 6% of loci after one cycle of RS and at 8% after seven cycles of RS (Figs. 1, 2). In agreement with previous studies (Labate et al. 1999; Pinto et al. 2003), these loci were not confined to particular chromosomes or genomic regions but dispersed over the whole genome.

The selection procedure of our RS programs resulted in a comparatively high selection response in both populations (Flachenecker et al. 2006a, b, c), whereas Waples' test (1989) determined significant changes in allele frequencies only at some loci. Thus, our findings support the hypothesis of Labate et al. (1999) that Waples' (1989) test might be not very powerful when a hypothesis other than random genetic drift is to be tested. Furthermore, selection may affect the sampling distribution of allele counts which violates the null hypothesis and, therefore, the test may become invalid.

Allele frequency changes in QTL regions

QTL mapping is targeted at the detection of (1) chromosomal regions carrying genes underlying a phenotypic trait and (2) marker alleles which are in linkage disequilibrium (LD) with the favorable alleles in the QTL region. The detected marker can then be used for indirect selection in a MAS program. For this study, we chose a different approach. We mapped QTL for selection index and its underlying components in the F_2 (A × B) and F_3 (C × D) base populations, but carried out selection with a selection index based solely on phenotypic information. This allows the evaluation of the selection response by comparing the location of QTL regions with the position of markers showing changes in allele frequencies.

With SIM, the LOD values depend on the map distance of a linked QTL and the corresponding effect size. The LOD value for a map position is larger the more closely linked QTL are and the larger their effect is. QTL with large effects contribute to the LOD value even if they are located in considerable distances and separated by one or more markers. In contrast, with CIM the use of cofactors results in LOD values affected only by tightly linked QTL located in the respective marker interval, whereas other effects are blocked (Jansen and Stam 1994). In selection cycle C1, the populations underwent altogether four meioses and relatively large chromosome regions are still expected to be in LD. As a consequence, QTL under selection

pressure may presumably result in allele frequency changes at markers in considerable distance. In contrast, in advanced selection cycles, the level of LD is expected to decrease and, thus, only small chromosome regions may be in LD, and only tightly linked QTL result in allele frequency changes at markers. To capture different situations during our selection program, we presented the LOD curves of SIM and CIM together with Waples' (1989) test for cycles C1 and C4 (A × B) or C7 (C × D) in Figs. 1 and 2.

Four significant QTL regions for selection index were detected with SIM and CIM in population A × B (Fig. 1), and at two of them a flanking marker showed a significant allele frequency change after one (SIM) and after four (CIM) cycles of RS. In population C × D, we detected one (SIM) and three (CIM) significant QTL regions for selection index, and at one of them a flanking marker showed a significant allele frequency change after seven cycles of RS (Fig. 2). In conclusion, the association between QTL for selection index (detected by SIM) and changes in allele frequencies in C1 was similar to that between QTL for selection index (detected by CIM) and changes in allele frequencies in final selection cycles (C4 or C7).

Consequently, the QTL regions for selection index and its components detected in the base populations were subjected to selection pressure when employing phenotypic selection in our RS program. However, in addition to the allele frequency changes in QTL regions, further significant allele frequency changes at markers spread across the entire genome were observed, which is in accordance with results of Coque and Gallais (2006). The chromosome regions linked to the markers showing allele frequency changes were also under selection pressure but were not identified by the QTL mapping. We attribute this observation mainly to the facts that (1) QTL mapping for complex traits with low heritabilities employing 274 (A × B) and 133 (C × D) lines is not expected to have sufficient power to detect all loci under selection in improvement of a quantitative trait (cf. Lande and Thompson 1990; Melchinger et al. 1998; Lübberstedt et al. 1998) and (2) linked QTL in these regions of the genome may have occurred in repulsion phase in the parents and therefore cancelled each other in the mapping population but were recombined by meioses during the intermating generations of subsequent selection cycles.

Correlation between observed and predicted changes in allele frequencies

Selection response is accomplished by a gradual increase in the frequency of favorable alleles. To predict changes in allele frequencies Δp using QTL results of the base population would be an advantage for planning and assessing RS programs. However, correlations between observed and predicted Δp (cf. Eq. 2; Hartl and Clark 1997, p. 422; Hallauer 1985) for selection index were only significant ($P < 0.05$) in population A × B but not in population C × D (Fig. 3). In general, the low correlation between observed and predicted Δp in both populations may be ascribed to (1) the assumptions that neither linkage, linkage disequilibrium nor epistasis affected the prediction of Δp and (2) random genetic drift effects occurring during the selection procedure which were not accounted for predicting Δp (Eq. 1). The observations in C × D are in accordance with the results of Coque and Gallais (2006) and were most likely caused by an upward bias in the estimated QTL effects. This inflation can be due to the fact that QTL mapping for traits with low heritabilities and small N, as employed in C × D ($N = 133$) as well as in the experiment of Coque and Gallais (2006) ($N = 99$), does not have sufficient power to detect enough QTL for explaining a substantial proportion of the genetic variance. Hence, our results indicate that large populations sizes for QTL analyses ($N \sim 300$; cf. Lübberstedt et al. 1998), as employed in population A × B, are required to predict allele frequency changes Δp more accurately.

In conclusion, our experiment supports the hypothesis that the detection power of QTL mapping experiments depends highly on the employed population sizes. Nevertheless, our results indicate that assessment of allele frequency changes in early cycles of an RS program can be used to detect marker alleles linked to QTL regions under selection pressure. If the cost of marker genotyping would be lower in comparison to phenotyping, these marker loci could be included in a selection index and subsequently used for MAS in later selection cycles.

Acknowledgments The present study was supported by a grant from the Deutsche Forschungsgemeinschaft, Grant No. FR 1615/1. The authors gratefully acknowledge Dr. Dietrich Klein and the staff of the Plant Breeding Research Station at Eckartsweier for their skilled conduct of the field trials. We thank KWS SAAT AG and Pioneer Hi-Bred Northern Europe for their support with the SSR screening. We also thank Christine Beuter and Elisabeth Kokai-Kota for the technical assistance with the SSR analyses. We greatly appreciate the critical reading and helpful comments on the manuscript of Dr. Jasmina Muminović.

References

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

Bernardo R, Romero-Severson J, Ziegle J, Hauser J, Joe L, Hookstra G, Doerge RW (2000) Parental contribution and coefficient of coancestry among maize inbreds: Pedigree, RFLP, and SSR data. Theor Appl Genet 100:552–556

Brown AHD, Allard RW (1971) Effect of reciprocal recurrent selection for yield on isozyme polymorphisms in maize (*Zea mays* L.). Crop Sci 11:888–893

Charbonnel N, Rasatavonjizay R, Sellin E, Bremond P, Jarne P (2005) The influence of genetic factors and population dynamics on the mating system of the hermaphroditic freshwater snail *Biomphalaria pfeifferi*. Oikos 108:283–296

Cockerham CC, Burrows PM (1980) Selection limits and strategies. Proc Natl Acad Sci USA 77:546–549

Coque M, Gallais A (2006) Genomic regions involved in response to grain yield selection at high and low nitrogen fertilization in maize. Theor Appl Genet 112:1205–1220

Falke KC, Melchinger AE, Flachenecker C, Kusterer B, Frisch M (2006) Comparison of linkage maps from F_2 and three times intermated generations in two populations of European flint maize (*Zea mays* L.). Theor Appl Genet 113:857–866

Flachenecker C, Frisch M, Falke KC, Melchinger AE (2006a) Trends in population parameters and best linear unbiased prediction of progeny performance in a European F_2 maize population under modified recurrent full-sib selection. Theor Appl Genet 112:483–491

Flachenecker C, Frisch M, Muminovic J, Falke KC, Melchinger AE (2006b) Modified full-sib selection and best linear unbiased prediction of progeny performance in a European F_2 maize population. Plant Breed 125:248–253

Flachenecker C, Frisch M, Falke KC, Melchinger AE (2006c) Genetic drift and selection effects of modified recurrent full-sib selection in two F_2 populations of European flint maize. Theor Appl Genet 113:1113–1120

Guzman PS, Lamkey KR (1999) Predicted gains from recurrent selection in the BS11 maize population. Maydica 44:93–99

Guzman PS, Lamkey KR (2000) Effective population size and genetic variability in the BS11 maize population. Crop Sci 40:338–346

Haldane JBS (1919) The combination of linkage values, and the calculation of distance between the loci of linked factors. J Genet 8:299–309

Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69:315–324

Hallauer AR (1985) Compendium on recurrent selection methods and their application. Crit Rev Plant Sci 3:1–33

Hallauer AR, Miranda JBF (1988) Quantitative genetics in maize breeding, 2nd edn. Iowa State University Press, Ames

Hartl DL, Clark AG (1997) Principles of population genetics. Sinauer Associates, Sunderland

Heredia-Diaz O, Alsirt A, Darrah LL, Coe EH (1996) Allelic frequency changes in the MoSCSSS maize synthetic in response to bi-directional recurrent selection for rind penetrometer resistance. Maydica 41:65–76

Hospital F, Chevalet C (1996) Interactions of selection, linkage and drift in the dynamics of polygenic characters. Genet Res Camb 67:77–87

Jansen RC, Stam P (1994) High resolution of quantitative traits into multiple loci via interval mapping. Genetics 136:1447–1455

Kahler AL (1983) Effect of half-sib and S1 recurrent selection for increased grain yield on allozyme polymorphisms in maize (*Zea mays* L.). Crop Sci 23:572–576

Labate JA, Lamkey KR, Lee M, Woodman WL (1999) Temporal changes in allele frequencies in two reciprocally selected maize populations. Theor Appl Genet 99:1166–1178

Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743–756

Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

Lübberstedt T, Melchinger AE, Fähr S, Klein D, Dally A, Westhoff P (1998) QTL mapping in testcrosses of flint lines of maize: III. Comparison across populations for forage traits. Crop Sci 38:1278–1289

Melchinger AE, Utz HF, Schön CC (1998) Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. Genetics 149:383–403

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

Miller AJ (1990) Subset selection in regression. Chapman and Hall, London

Patterson HD, Williams ER (1976) A new class of resolvable incomplete block designs. Biometrica 63:83–92

Pinto LR, Carneiro Vieira ML, de Souza CL Jr, de Souza AP (2003) Reciprocal recurrent selection effects on the genetic structure of tropical maize populations assessed at microsatellite loci. Genet Mol Biol 26:355–364

Queney G, Ferrand N, Marchandeau S, Azevedo M, Mougel F, Branco M, Monnerot M (2000) Absence of a genetic bottleneck in a wild rabbit (*Oryctolagus cuniculus*) population exposed to a severe viral epizootic. Mol Ecol 9:1253–1264

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

Schön CC, Melchinger AE, Boppenmaier J, Brunklaus-Jung E, Herrmann RG, Seitzer JF (1994) RFLP mapping in maize: Quantitative trait loci affecting testcross performance of elite European flint lines. Crop Sci 34:378–389

Smith OS (1979) A model for evaluating progress from recurrent selection. Crop Sci 19:223–226

Stuber CW, Moll RH, Goodman MM, Schaffer HE, Weir BS (1980) Allozyme frequency changes associated with selection for increased grain yield in maize (*Zea mays* L.). Genetics 95:225–236

Utz HF (2001) PLABSTAT: a computer program for the statistical analysis of plant breeding experiments. Institute for Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, Stuttgart

Utz HF, Melchinger AE (1996) PLABQTL: a program for composite interval mapping of QTL. J Quant Trait Loci 2 (1). Available at http://www.uni-hohenheim.de/~ipspwww/soft.html (verified 10 Sept. 2004)

Utz HF, Melchinger AE, Schön CC (2000) Bias and sampling error of the estimated proportion of genotypic variance explained by quantitative trait loci determined from experimental data in maize using cross validation and validation with independent samples. Genetics 154:1839–1849

Van Ooijen JW, Voorrips RE (2001) JoinMap Version 3.0 software for the calculation of genetic linkage maps. Plant Research International, Wageningen

Waples RS (1989) Temporal variation in allele frequencies: Testing the right hypothesis. Evolution 43:1236–1251