Aus dem Institut für
Pflanzenzüchtung, Saatgutforschung und Populationsgenetik
der Universität Hohenheim
Fachgebiet Angewandte Genetik und Pflanzenzüchtung
Prof. Dr. A. E. Melchinger

Molecular Genetic Analysis of Modified Recurrent Full-sib Selection in Two European F_2 Flint Maize Populations

Dissertation
zur Erlangung des Grades eines Doktors
der Agrarwissenschaften vorgelegt
der Fakultät Agrarwissenschaften

von
Dipl.-Ing. agr.
Karen Christin Falke
geboren in Hannover

Stuttgart-Hohenheim
2007

Die vorliegende Arbeit wurde am 24. Januar 2007 von der Fakultät Agrarwissenschaften als „Dissertation zur Erlangung des Grades eines Doktors der Agrarwissenschaften (Dr. sc. agr.)“ angenommen.

Tag der mündlichen Prüfung	23. Mai 2007
1. Prodekan:	Prof. Dr. W. Bessei
Berichterstatter, 1. Prüfer:	Prof. Dr. A.E. Melchinger
Mitberichterstatter, 2. Prüfer:	Prof. Dr. H.-P. Piepho
3. Prüfer:	Prof. Dr. R. Blaich

Contents

[1] 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

[2] Falke KC, Flachenecker C, Melchinger AE, Maurer HP, Frisch M (2007a) Temporal Changes in Allele Frequencies in Two European F_2 Flint Maize Populations under Modified Recurrent Full-Sib Selection. Theor Appl Genet 114:765-776

[3] Falke KC, Maurer HP, Melchinger AE, Piepho HP, Flachenecker C, Frisch M (2007b) Linkage Disequilibrium in Two European F_2 Flint Maize Populations under Modified Recurrent Full-Sib Selection. Theor Appl Genet 115:289-297

Abbreviations

A	inbred line KW1265
AFLP	amplified fragment length polymorphism
B	inbred line D146
C	inbred line D145
D	inbred line KW1292
$\bar{d}$	degree of dominance
DNA	deoxyribonucleic acid
i_r	average information from an individual
LD	linkage disequilibrium
MAS	marker-assisted selection
N	population size
N_e	effective population size
PCR	polymerase chain reaction
QTL	quantitative trait locus (or loci, depending on the context)
R	recombination frequency accumulated across all meioses a population underwent
r	recombination frequency per meiosis
RAPD	ramdom amplified polymorphic DNA
REML	restricted maximum likelihood
RFLP	restriction fragment length polymorphism
RS	recurrent selection
σ_A^2	additive variance
σ_D^2	dominance variance
SNP	single nucleotide polymorphism
SSR	simple sequence repeat

1. General Introduction

Recurrent Selection for Improving Germplasm

Plant breeding is the science, art, and business of improving the performance of crops for human benefit (Bernardo 2002, p. 3). However, most of the plant breeding procedures result in a substantial reduction in the genetic variability of the employed breeding material (Bulmer 1971; Becker 1993, p. 258). In contrast, recurrent selection (RS) programs attempt to achieve a long-term selection response by increasing the frequency of favorable alleles while simultaneously maintaining the genetic variability present in the population (Hallauer 1985). Each cycle of RS programs (with the exception of RS programs employing mass selection, which are not considered here) consists of three phases: (1) development of progenies from the source populations, (2) evaluation of progenies in replicated trials that may be conducted in different environments, and (3) recombination of selected progenies, based on the evaluation trials, to form a base population for the next cycle (Figure 1).

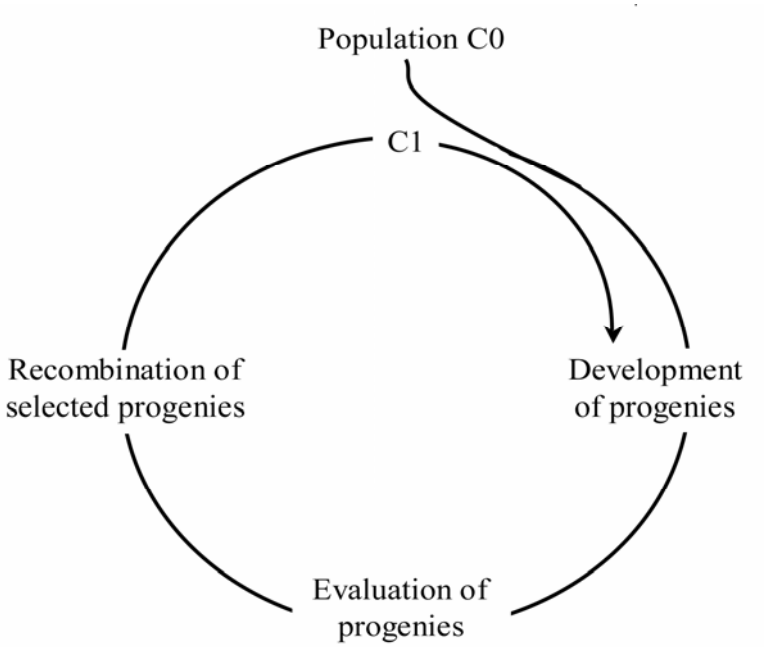

Figure 1. Cyclical nature of RS for the systematic improvement of germplasm (modified after Hallauer et al. 1988, p. 493).

Population Improvement of Maize Using RS

Maize (*Zea mays* L.) is one of the most important crops in the world exhibiting an immense relevance as food and livestock feed. Moreover, its byproducts such as starch and ethanol are used to manufacture commodities such as soap, paint, rayon, glue and others. To facilitate a continuous yield increase, several maize breeding methods have been designed, which can be classified into two categories: (1) inbred line development as parents of hybrid and synthetic varieties and (2) improvement of populations for use as open-pollinated varieties. Since 1939, population improvement in maize has been conducted via RS (Hallauer et al. 1988, p. 499) to increase yield, alter seed or plant quality, increase pest resistance, improve tolerance to environmental conditions or adapt exotic germplasm (Hallauer 1985). The superior population can subsequently be applied as open-pollinated variety or as base population for developing advanced hybrids.

The choice of the RS method depends on the goal of the breeding program. The most common selection methods for maize intra-population improvement are mass selection (Gardner 1961) and family selection, such as half-sib (Webel and Lonnquist 1967), full-sib (Moll 1991) or selfed (S_1 or S_2; Penny et al. 1967) family selection. Among these methods, recurrent full-sib selection enables a comparatively high selection response. In addition, it offers complete parental control combined with a short cycle length (Weyhrich et al. 1998).

As base population for RS programs, any type of breeding population can be used. In maize, open-pollinated or synthetic varieties are frequently used, whereas F_2 base populations obtained from biparental crosses of homozygous inbred lines have been employed only in a few studies. Nevertheless, RS programs employing F_2 base populations achieved a comparatively high average selection response for grain yield, ranging between 4.5 and 7.3% (Russell et al. 1973; Genter 1982; Moll 1991; Landi and Frascaroli 1993). F_2 base populations are particularly advantageous for examining the selection process because the allele frequencies are known (0.5 for segregating loci). However, the disadvantages of F_2 base populations are (1) a more restricted genetic base and (2) linkage disequilibrium (LD) between alleles originating from the same parental line at linked loci (= parental LD).

Recombination of selected progenies to generate new variation, as the third phase of RS, is usually conducted by random mating. However, the implementation of the pseudo-factorial mating scheme of Cockerham and Burrows (1980) has created an opportunity for recombining selected progenies by including pedigree data. This mating scheme assigns

sexual roles after selection, using from s selected genotypes the best genotypes (s_1) double as male and the remaining (s-s_1) genotypes as female parents. Consequently, it is expected that the mating scheme of Cockerham and Burrows (1980) increases the probability of obtaining superior recombinants with the same selection intensity as other mating schemes do. Nevertheless, to our knowledge no studies applying the pseudo-factorial mating scheme are available so far.

The improvement of the performance of agronomic traits with RS, however, is not only affected by selection but also by random genetic drift. Moreover, both selection and random genetic drift generate LD between loci pairs which may affect the development of the additive genetic variance and therefore hamper the selection response. For determining the efficiency of selection programs and their optimization, detailed knowledge about these effects is of crucial importance. While several empirical studies investigated these factors theoretically and in simulation studies, up to now there is still a lack of experimental analyses, especially at the molecular level.

Effects of Random Genetic Drift and Selection on Changes in Allele Frequencies

Selection aims at an enhancement of the performance level of breeding germplasm by increasing the frequencies of favorable alleles. In contrast, random genetic drift is a random change in allele frequencies resulting from sampling effects in small populations. Random genetic drift may cause a fixation of unfavorable alleles, reduce the genetic variance and, hence, lead to a decline in long-term selection response (Guzmann et al. 1999, 2000). Consequently, for determining the efficiency of selection programs it is important to assess allele frequency changes and to separate the effects due to selection from those of random genetic drift. Molecular markers were proposed as a promising tool for investigations of allele frequency changes (Labate et al. 1999; Pinto et al. 2003). For the analysis of changes in allele frequencies in maize, standard statistical tests (*e.g.*, χ^2) or linear regression approaches have been commonly applied (Brown and Allard 1971; Stuber and Moll 1972; Kahler 1983; Revilla et al. 1997). However, these tests neglect the stochastic dependence and the effects of random genetic drift in populations with finite population size and are, thus, not well suited for examining changes in allele frequencies. A test that takes these factors into account was proposed by Waples (1989). However, it has rarely been employed

in plant breeding studies so far (Labate et al. 1999; Pinto et al. 2003; Coque and Gallais 2006).

Many important agronomic traits are quantitatively inherited and, thus, affected by many genes as well as environmental factors. The analysis of these complex traits has advanced with the development of molecular marker technologies. In the 1990s, new statistical tools have been established (Lander and Botstein 1989; Haley and Knott 1992; Jansen and Stam 1994) and implemented in software packages (*e.g.*, Lincoln et al. 1993; Utz and Melchinger 1996) for the analysis of quantitative trait loci (QTL) mapping experiments. The primary aim of QTL mapping is to identify regions of the genome that contribute to the variation in the trait of interest. The detected markers can subsequently be used for indirect selection in marker-assisted selection (MAS) programs. On the other hand, QTL can be used for evaluating selection response of RS programs by comparing QTL regions for traits under selection detected in the base population with changes in allele frequencies in the subsequent selection cycles. This approach may provide insights into the genomic regions under selection. Nevertheless, thorough analyses on the relationship between QTL regions and changes in allele frequencies due to selection are still scarce.

Effects of LD on the Selection Response

Alleles are considered to be in LD when alleles at two loci occur in gametes more frequently than expected, given the known allele frequencies at the two loci. The extent and distribution of LD in plant breeding populations is affected by linkage and population stratification or relatedness in the population. Furthermore, it can be generated by (1) random genetic drift due to small population sizes, (2) selection, (3) hitchhiking effects of alleles linked with selected alleles, (4) selection of favorable combinations of alleles (epistasis), or (5) migration and admixture of populations with different allele frequencies (Falconer and Mackay 1996, p. 16; Lynch and Walsh 1998, p. 95; Flint-Garcia et al. 2003). Mutation is at best a marginal factor yielding LD (Stich et al. 2006). A stepwise reduction of existing LD can be obtained by random mating (Bernardo 2002, p. 23).

Johnson (1982) demonstrated that initial parental LD has a permanent effect on the selection progress, even if (1) there is only a small level of parental LD between unlinked loci and (2) selection is already relaxed. Intermating prior to initiating the selection procedure is more efficient in reducing the effects of parental LD than intermating between

successive selection cycles (Johnson 1982). Parental LD is present in F_2 populations derived from two inbred lines as base populations for RS procedures (Hallauer and Miranda 1988, p. 70). Therefore, Johnson (1982) suggested three generations of random intermating within the F_2 population to reduce sufficiently the influence of initial parental LD. However, to our knowledge, no attempts have been made to verify these theoretical findings with experiments based on molecular markers.

In RS procedures under truncation selection, the individuals with the largest phenotypic values are selected as parents of the next generation and the rest are discarded. Genotypic values of these individuals are not identical but more alike than those of a randomly chosen set of individuals. Hence, negative LD between loci pairs can be generated due to selection (Falconer and Mackay 1996, p. 202). This LD will be generated immediately with the first selection cycle and is associated with a change in the genetic variance. In the case of negative LD, the extent of the additive genetic variance σ_A^2 will be reduced (Bulmer 1971) and, consequently, the long-term selection response is hampered.

The analysis of LD effects on the development of the additive genetic variance σ_A^2 was conducted by employing theoretical approaches or simulation studies (Hospital and Chevalet 1996). However, no investigations at the molecular level have been compiled yet.

High Mapping Resolution with Intermated Mapping Populations

A genetic linkage map is an abstract model of the linear arrangement of genes and marker loci on chromosomes. Doubled haploid, backcross, F_2 or recombinant inbred line populations have been used as mapping populations in plant breeding research. To achieve comparability between different population types, map distances have been defined on the basis of the distribution of crossover events in a single meiosis (Haldane 1919; Stam 1993). The concept of linkage mapping includes three successive phases: (1) assessment of the two locus genotypes of the individuals in the mapping population for all pairs of loci, (2) assignment and ordering of loci to linkage groups, and (3) estimation of map distances between loci. Since the development of the first genetic map (Sturtevant 1913), the localization and identification of genes underlying qualitative and quantitative phenotypic traits is possible (Sax 1923). At present, genetic linkage maps are well established in the research of plant genetics and facilitate both basic and applied research.

In maize, the first genetic linkage map was published in the 1930s by Emerson et al. (1935), based on morphological variants. The discovery of restriction enzymes and the utilization of restriction fragment length polymorphisms (RFLPs) in the early 1970s has revolutionized the research in plant genetics by using molecular makers at the level of DNA. The first molecular linkage maps in maize using RFLPs have been presented by Helentjaris et al. (1986) and Coe et al. (1987). A substantial progress for the construction of marker-saturated genetic linkage maps has been achieved with the advent of PCR-based molecular markers (*e.g.*, amplified fragment length polymorphism (AFLP), random amplified polymorphic DNA (RAPD), simple sequence repeats (SSR), or single nucleotide polymorphism (SNP)). High-density maps contribute greatly to our understanding of evolutionary processes, enable marker-assisted selection and mapping of agronomic traits, and facilitate many aspects of crop improvement (Sharopova et al. 2002).

With the construction of high-density maps by means of PCR-based markers the resolution of the genetic maps became increasingly important, *i.e.*, the detection of recombination events between tightly linked loci. To solve the lack of genetic resolution, the employment of mapping populations whose individuals were intermated for several generations was suggested (Mather 1936; Allard 1956). More recently, a few studies employed intermated F_2 and intermated recombinant inbred line populations in *Arabidopsis thaliana* (Liu et al. 1996), maize (Lee et al. 2002) and as mapping populations in theoretical studies (Winkler et al. 2003; Falque et al. 2005; Teuscher et al. 2005; Teuscher and Broman 2006). However, for the construction of the genetic maps all these studies used maximum likelihood functions for estimating two-locus recombination frequencies, developed for F_2 or recombinant inbred line populations. Consequently, estimated recombination frequencies referred to accumulated recombination events occurring due to multiple meioses, even though recombination frequencies of genetic maps refer per definition to a single meiosis event.

Up to now, no in-depth experimental studies exist for (1) constructing linkage maps for intermated populations employing maximum likelihood functions for the estimation of recombination frequencies of this population type, and (2) verifying the benefit of intermated mapping populations in applied plant breeding programs.

Experimental Set-Up

As part of a breeding project, long-term recurrent full-sib selection programs with two European F_2 flint maize populations (KW1265 × D146 and D145 × KW1292 hereafter referred to as A × B and C × D, respectively) were initiated in 1990 to analyze the selection response. The F_2 populations were intermated for three generations by using a chain crossing procedure to develop the F_2Syn3. The subsequent selection procedure performed four cycles of RS for population A × B and seven cycles for C × D using a pseudo-factorial mating scheme suggested by Cockerham and Burrows (1980). The evaluation of the selection response at the phenotypic level by using classical quantitative genetic tools was carried out in companion studies (*cf.* Flachenecker et al. 2006a, 2006b, 2006c). For investigations at the molecular level, the parental lines, the base and intermated populations as well as all selection cycles were analyzed with 104 (A × B) and 101 (C × D) SSRs displaying a uniform distribution across the maize genome.

Objectives

The main goal of this thesis research was to analyze the selection procedure and the selection response of two European flint maize populations under modified recurrent full-sib selection with the aid of molecular markers. In detail, the objectives were to

(1) investigate the benefit of intermating generations to (a) minimize the influence of initial parental LD on the selection response of the RS programs and (b) employ them as mapping populations in applied plant breeding programs,

(2) determine the number, positions and genetic effects of QTL detected for traits under selection in the base populations,

(3) separate the effects of random genetic drift and selection on allele frequency changes in QTL regions,

(4) assess the extent of LD occurring during the RS process, and

(5) monitor the influence of LD on the additive genetic variance and, thus, on the selection response of RS programs.

References

Allard RW (1956) Formulas and tables to facilitate the calculation of recombination values in heredity. Hilgardia 24:235-278

Becker H (1993) Pflanzenzüchtung. Ulmer Verlag

Bernardo R (2002) Breeding for Quantitative Traits in Plants. Stemma, Woodbury, MN

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

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

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

Coe EH, Hoisington DA, Neuffer MG (1987) Linkage map of corn (maize) (*Zea mays* L.). Maize Genet Coop Newslett 61:116-147

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

Emerson RA, Beadle GW, Fraser AC (1935) A summary of linkage studies in maize. Cornell Univ Agric Exp Stn Memoir 180:1-83

Falconer DS, Mackay TFC (1996) Introduction to Quantitative Genetics. 4th ed. Longman Sci and Tech, Harlow, UK

Falque M, Decousset L, Dervins D, Jacob AM, Joets J, Martinant JP, Raffoux X, Ribiere N, Ridel C, Samson D, Charcosset A, Murigneux A (2005) Linkage mapping of 1454 new maize candidate gene loci. Genetics 170:1957-1966

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 recurrent full-sib selection programs in two F_2 populations of European flint maize. Theor Appl Genet 13:1113-1120

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

Gardner CO (1961) An evaluation of effects of mass selection and seed irradiation with thermal neutrons on yield of corn. Crop Sci 1:241-245

Genter CF (1982) Recurrent selection for high inbred yields from the F2 of maize single cross. Proc Ann Corn and Sorghum Ind Res Conf 37:67-76

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

Hallauer, AR, Russell WA, Lamkey KR (1988) Corn breeding. pp. 463-564 *In* Sprague GF, Dudley JW. Corn and Corn Improvement. 3 edn. American Society of Agronomy, Inc. Madison

Helentjaris T, Slocum M, Wright S, Schaefer A, Nienhuis J (1986) Construction of genetic linkage maps in maize and tomato using restriction fragment length polymorphisms. Theor Appl Genet 72:761-769

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

Johnson GR (1982) Two-locus theory in recurrent selection for general combining ability in maize. Theor Appl Genet 61:279-283

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

Labate JA, Lamkey KR, Lee M, Woodman W (2000) Hardy-Weinberg and linkage

equilibrium estimates in the BSSS and BSCB1 random mated populations. Maydica 45:243-255

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

Landi P, Frascaroli E (1993) Responses to four cycles of full-sib family recurrent selection in an F2 maize population. Maydica 38:31-37

Lee M, Sharopova N, Beavis WD, Grant D, Katt M, Blair D, Hallauer A (2002) Expanding the genetic map of maize with the intermated B73 x Mo17 (IBM) population. Plant Mol Biol 48:453-461

Lincoln SE, Daly MJ, Lander ES (1993) Mapping genes controlling quantitative traits using MapMaker/QTL version 1.1. Whitehead Institute for Biomedical Research, Cambridge, MA

Liu SC, Kowalski SP, Lan TH, Feldmann KA, Paterson AH (1996) Genome-wide high-resolution mapping by recurrent intermating using *Arabidopsis thaliana* as a model. Genetics 142:247-258

Lynch, M, Walsh B (1998) Genetics and Analysis of Quantitative Traits. Sinauer Assoc, Sunderland

Mather K (1936) Types of linkage data and their value. Ann Eugen 7:251-264

Moll RH (1991) Sixteen cycles of recurrent full-sib family selection for grain weight in two maize populations. Crop Sci 31:959-964

Penny LH, Scott GE, Guthrie WD (1967) Recurrent selection for European corn borer resistance in maize. Crop Sci 7:407-409

Pinto LR, Carneiro Vieira ML, de Souza Jr CL, 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

Revilla P, Vales MI, Malvar RA, Ordas A (1997) Allozyme frequencies, heterozygosity and genetic distances following S1 recurrent selection in two synthetic maize populations. Theor Appl Genet 95:1057-1061

Russell WA, Eberhard UA, Vega Q (1973) Recurrent selection for specific combining ability for yield in two maize populations. Crop Sci 13:257-261

Sax K (1923) The association of size differences with seed coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics 8:552-560

Sharopova N, McMullen MD, Schultz L, Schroeder S, Sanchez-Villeda H, Gardiner J, Bergstrom D, Houchins K, Melia-Hancock S, Musket T, Duru N, Polacco M,

Edwards K, Ruff T, Register JC, Brouwer C, Thompson R, Velasco R, Chin E, Lee M, Woodman-Clikeman W, Long MJ, Liscum E, Cone K, Davis G, Coe EH (2002) Development and mapping of SSR markers for maize. Plant Mol Biol 48:463–481

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. Plant J 3:739-744

Stich B, Maurer HP, Melchinger AE, Frisch M, Heckenberger M, van der Voort JR, Peleman J, Sørensen AP, Reif JC (2006) Comparison of linkage disequilibrium in elite European maize inbred lines using AFLP and SSR markers. Mol Breed 17:217-226

Stuber CW, Moll RH (1972) Frequency changes of isozyme alleles in a selection experiment for grain yield in maize. Crop Sci 12:337-340

Sturtevant AH (1913) The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association. J Exp Zool 14:43-59

Teuscher F, Broman KW (2006) Haplotype probabilities for multiple-strain recombinant inbred lines. Genetics.106.064063

Teuscher F, Guiard V, Rudolph PE, Brockmann GA (2005) The map expansion obtained with recombinant inbred strains and intermated recombinant inbred populations for finite generation designs. Genetics 170:875-879

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

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

Webel OD, Lonnquist JH (1967) An evaluation of modified ear-to-row selection in a population of corn. Crop Sci 7:651-655

Weyhrich RA, Lamkey KR, Hallauer AR (1998) Responses to seven methods of recurrent selection in the BS11 maize population. Crop Sci 38:308-321

Winkler CR, Jensen NM, Cooper M, Podlich DW, Smith OS (2003) On the determination of recombination rates in intermated recombinant inbred populations. Genetics 164:741-745

Theor Appl Genet (2006) 113:857–866
DOI 10.1007/s00122-006-0343-x

ORIGINAL PAPER

Comparison of linkage maps from F_2 and three times intermated generations in two populations of European flint maize (*Zea mays* L.)

K. C. Falke · A. E. Melchinger · C. Flachenecker · B. Kusterer · M. Frisch

Received: 12 April 2006 / Accepted: 9 June 2006 / Published online: 11 July 2006

Abstract Intermated mapping populations are expected to result in high mapping resolution for tightly linked loci. The objectives of our study were to (1) investigate the consequences of constructing linkage maps from intermated populations using mapping methods developed for F_2 populations, (2) compare linkage maps constructed from intermated populations (F_2Syn3) with maps generated from corresponding F_2 and F_3 base populations, and (3) investigate the advantages of intermated mapping populations for applications in plant breeding programs. We constructed linkage maps for two European flint maize populations (A × B, C × D) by mapping 105 SSR markers in generations F_2 and F_2Syn3 of population A × B, and 102 SSR markers in generations F_3 and F_2Syn3 of population C × D. Maps for F_2Syn3 were constructed with mapping methods for F_2 populations (Map A) as well as with those specifically developed for intermated populations (Map B). Both methods relate map distances to recombination frequencies in a single meiosis and, therefore, did not show a map expansion in F_2Syn3 compared with maps constructed from the respective F_2 or F_3 base populations. Map A and B differed considerably, presumably because of theoretical shortcomings of Map A. Since loosely linked markers could not unambiguously be mapped in the F_2Syn3 populations, they may hamper the construction of linkage maps from intermated populations.

Communicated by M. Xu.

K. C. Falke · A. E. Melchinger (✉) · C. Flachenecker · B. Kusterer · M. Frisch
Institute of Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim, 70593 Stuttgart, Germany
e-mail: melchinger@uni-hohenheim.de

Introduction

Genetic linkage maps are an important tool in genetic research and applied breeding programs. A linkage map attempts to reproduce the linear order of gene and/or marker loci on a chromosome and to quantify the degree of linkage between loci by their distance on the map. In construction of genetic maps for diploid species, linkage between loci is most commonly estimated from segregating populations obtained from biparental crosses of homozygous parents. The types of mapping populations differ in the number of meioses that the homologous chromosomes undergo and, in consequence, in the probability distribution of crossover events between adjacent loci on a homologous chromosome. To account for different number of meioses, the convention is employed that the genetic map distance between loci refers to the distribution of crossover events between these loci in a single meiosis (Haldane 1919; Kosambi 1944; Stam 1993; Weir 1996, p. 230).

For mapping tightly linked loci, it is advantageous to employ mapping populations where the individuals have undergone several meioses (cf. Allard 1956). This assures a sufficient frequency of individuals with recombination between tightly linked loci, resulting in a small standard error σ_r of the estimated recombination frequency r per meiosis. F_2 populations intermated randomly for t generations were suggested as mapping populations in maize and *Arabidopsis thaliana* to

increase mapping accuracy of tightly linked loci (Beavis et al. 1992; Liu et al. 1996; Lee et al. 2002). The amount of information i_r provided by a single individual of a mapping population can be used to compare the efficiency of mapping experiments (Mather 1936). Liu et al. (1996) derived i_r for intermated F_2 populations and indicated situations where the amount of information from an individual in an intermated population was greater than in the corresponding F_2 population.

Construction of a linkage map involves four steps: (1) the genotypes of the individuals in a mapping population are assessed, (2) pairs of putatively linked loci are assigned to linkage groups, (3) the loci of a linkage group are ordered, (4) the map distance between loci is estimated. The well-established mapping theory for F_2, backcross, doubled haploid, or recombinant inbred line mapping populations is available and implemented in software (Lander et al. 1987; Holloway and Knapp 1993; Stam 1993, p. 305; Liu 1998; Van Ooijen and Voorrips 2001). However, theory and software for linkage mapping with intermated populations, have not been developed yet. As a consequence, studies on linkage mapping with intermated populations (Beavis et al. 1992; Liu et al. 1996; Lee et al. 2002) have so far employed mapping methods developed for F_2 populations. This approach yields in estimated recombination frequencies R that result from the accumulated effect of all meioses a population underwent. The relation between R and the recombination frequency r that refers to a single meiosis is (Darvasi and Soller 1995; Liu et al. 1996)

$$R = \frac{1}{2}\left(1 - (1-2r)(1-r)^t\right), \quad (1)$$

where t refers to the number of intermating generations.

The objectives of our study were to (1) investigate the consequences of generating linkage maps from intermated populations by using mapping methods developed for F_2 populations, (2) compare linkage maps generated from intermated populations with maps generated from the corresponding F_2 and F_3 base populations, and (3) investigate the advantages and disadvantages of intermated mapping populations for applications in plant breeding programs.

Materials and methods

Plant materials

The plant materials used in this study were partly identical to those employed in previous QTL studies on grain traits (Schön et al. 1994; Melchinger et al. 1998; Utz et al. 2000; Mihaljevic et al. 2005) and forage traits (Lübberstedt et al. 1997) in maize. Briefly, four early maturing European flint inbred lines KW1265, D146, D145, and KW1292, further referred to as A, B, C, and D, respectively, were used as parents. Randomly chosen F_2 plants were selfed to produce 380 F_3 lines of cross A × B, and 140 F_4 lines of cross C × D. Additionally, with each F_2 population three generations of intermating were performed by chain crossing of 240 unselected plants (i.e., 1 × 2, 2 × 3, ..., 240 × 1) to produce generation F_2Syn3. In the present study, only random subsets of 146 F_3 lines of cross A × B, 110 F_4 lines of cross C × D, and 148 plants of the F_2Syn3 generation of each cross were analyzed.

Leaf collection and DNA extraction

Leaves were harvested from 10 to 20 plants of each F_3 line of cross A × B and each F_4 line of cross C × D. Equal amounts of leaf material from each plant per line were bulked for DNA extraction to determine the marker genotypes of the parental F_2 and F_3 plants. In the intermated generations, leaves were harvested from individual plants. Harvested leaves were freeze-dried and ground to powder. DNA extraction was performed according to the CTAB method (Hoisington et al. 1994).

SSR analyses

Parental lines A, B, C, and D were screened with 860 public SSR markers from the MaizeGDB (http://www.maizegdb.org/ssr.php). Out of the 319 (A × B) and 354 (C × D) polymorphic markers per cross, we assayed 105 and 102 SSRs with a uniform distribution over the maize genome. Primer pairs were synthesized by Sigma–Genosys (Steinheim, Germany), with one primer of each pair being tagged at the 5′ end with a fluorescent label (Indodicarbocyanine (Cy5) phosphoramidite). The PCR reactions were conducted in a volume of 15 µl containing 25 ng template DNA, 0.15 mM of each dNTP, 2.5 mM $MgCl_2$, 0.25 µM of each primer, 1× *Taq* DNA polymerase buffer, and 0.5 U of *Taq* DNA polymerase (Invitrogen GmbH, Karlsruhe, Germany). Amplifications were performed using a Primus HT thermal cycler (MWG BIOTECH, Ebersberg, Germany). The PCR cycling conditions that yielded the strongest amplification product were considered optimum and used for analyses. The resulting PCR products were separated by using polyacrylamid gels (ultra pure SequaGel-XR, National Diagnostics, Atlanta, GA) run on an ALF Express

(Amersham Pharmacia Biotech, Freiburg, Germany) automated sequencer and transferred to a 1/0 matrix.

Linkage analyses

Observed genotype frequencies at each marker locus were checked for deviations from Mendelian segregation ratios (1:2:1 in F_2 or F_2Syn3, and 3:2:3 in F_3 populations) and allele frequency of 0.5 by χ^2 tests, with adjustment for multiple tests according to Sidak with an experiment wise error of $\alpha = 0.05$.

We assumed no interference in crossover formation, such that the relationship between the map distance d and the recombination frequency r is described by Haldane's (1919) mapping function:

$$r = \frac{1}{2}\left(1 - e^{-2d}\right), \tag{2}$$

where d denotes here the map distance in Morgan units.

For population F_2 of cross A × B and for population F_3 of cross C × D, linkage maps were constructed with the algorithm described by Stam (1993), employing software JoinMap Version 3.0 (Van Ooijen and Voorrips 2001).

Briefly, the underlying computational steps were: (1) pair wise recombination frequencies r for all locus pairs were estimated from the observed data using maximum likelihood (Fisher 1946; Bailey 1961 p. 38; Stam 1993), and the corresponding map distances d_o were calculated using Haldane's mapping function, (2) loci were assigned to linkage groups and ordered, (3) a least squares procedure was used to estimate the locus distances d_m on the final linkage map by minimizing

$$\sum_{\text{locus pairs}} w(d_o - d_m)^2 \rightarrow \min, \tag{3}$$

where w are weights obtained from a test of linkage between pair of loci.

For both F_2Syn3 populations, linkage maps were generated with the method employed in earlier studies on mapping with intermated populations (Beavis et al. 1992; Liu et al. 1996; Lee et al. 2002). This approach consists of the following steps: (1) the maximum likelihood equations for estimating recombination frequencies in F_2 populations are applied to intermated populations, (2) recombination frequencies between adjacent loci on the resulting map are interpreted as the recombination frequencies R, which refer to all meioses events during intermating, (3) Eq. 1 is employed to derive the recombination frequencies r that refer to one single meiosis from the values of R. Map distances d between adjacent loci on the final map were calculated using r. We refer to the resulting linkage map as Map A.

For both intermated populations, a second linkage map (further referred to as Map B) was constructed, using the approach described for F_2 and F_3 populations. For estimating pair wise recombination frequencies, the expected phenotype frequencies in the intermated populations are required. The expected two-locus phenotype frequencies in an intermated population can be determined as follows: (1) Eq. 1 yields directly the frequencies of recombinant gametes, (2) due to symmetry reasons, the frequency of recombinant gametes can be used to calculate the frequency of all four possible gametes, (3) due to random mating, the frequencies of the 16 possible genotypes can be obtained by multiplying the corresponding gamete frequencies, (4) the phenotype frequencies are obtained from the genotype frequencies by summing up genotype classes resulting in the same phenotype. Calculations were performed with an extension of software JoinMap Version 3.0 (Van Ooijen, unpublished).

For all linkage maps we used the locus orders published in MaizeGDB (http://www.maizegdb.org). We employed the ''fixed order'' command of JoinMap Version 3.0 using an LOD threshold of 0.01 (partially 0.001) and a recombination threshold of 0.499 for marker pairs. The low stringency mapping procedure was chosen to construct maps with as many markers as possible. For comparing the two mapping methods and the two populations, the maps in both crosses included only SSR markers that could be mapped in both populations.

Results

Significant ($P < 0.05$) deviations from the expected segregation ratios (1:2:1 and 3:2:3) were observed in zero (A × B, F_2), 11 (C × D, F_3), and six cases (A × B, F_2Syn3; C × D, F_2Syn3). Allele frequencies deviated significantly ($P < 0.05$) from 0.5 at none (A × B, F_2 and F_2Syn3), four (C × D, F_3), and two (C × D, F_2Syn3) marker loci. In population F_2 of cross A × B and F_3 of cross C × D, the total map distances covered by all SSR markers spanned 1,803 and 1,608 cM, with an average interval length of 19 and 17 cM, respectively (Figs. 1, 2). In population F_2Syn3 of cross A × B, only 89% (Map A) and 92% (Map B) of all SSRs could be mapped. In the same population of cross C × D,

94% (Map A) and 93% (Map B) of all SSRs could be mapped. The total map distances of cross A × B spanned 1,371 cM in Map A and 1,518 cM in Map B, with an average interval length of 17 cM (Fig. 1). The maps obtained for cross C × D spanned 1,336 cM in Map A and 1,406 cM in Map B, with an average interval length of 16 and 17 cM, respectively (Fig. 2).

In cross A × B, the number of SSRs that could be mapped in population F_2Syn3 was lower for Map A (93) than for Map B (97) (Table 1). The 93 common loci defined 83 intervals on Map A and Map B. Although at 84% of the intervals the estimated *d* values were smaller with Map A than with Map B, we could observe major differences (> 10 cM) only at four intervals. Differences between the two mapping methods became evident with increasing marker intervals (Table 2). Correlations ρ of estimated *d* values between F_2Syn3 and the corresponding F_2 varied widely among chromosomes, ranging from 0.02 at chromosome 3 to 0.90 at chromosome 4 (Map A), and from –0.55 at chromosome 5 to 0.91 at chromosome 4 (Map B). The average *d* ratios between population F_2Syn3 and F_2 were smaller for Map A (0.86) than for Map B (0.91). Comparing population F_2Syn3 with the corresponding F_2 of cross A × B, we observed a shrinkage of the total map length of 0.86 using Map B (Table 1). The shrinkage amounted 0.56 on chromosome 5, while a 1.32-fold expansion was observed on chromosome 10. The 97 common loci defined 87 intervals on the maps of population F_2 and F_2Syn3 (Map B). For F_2, estimates of *d* were larger (> 10 cM) at 12 and for F_2Syn3 at nine intervals.

In population F_2Syn3 of cross C × D, the linkage map of Map A was smaller than that of Map B, although almost the same number of SSRs was mapped (Table 1). At 16% of the 83 common intervals, estimated *d* values were larger in Map A than in Map B, but differences greater than 10 cM were observed only for three intervals. Differences between Map A and B became apparent with increasing marker intervals (Table 2). Significant ($P < 0.01$) correlations between estimates of *d* from the F_2Syn3 and the corresponding F_3 were observed at chromosomes 2, 3, 5, and 10 of Map B, and only at chromosomes 2, 3, and 10 of Map A. The *d* ratio averaged 0.92 (Map A) and 0.96 (Map B). We observed a marginal shrinkage of the total map length (0.95) between population F_2Syn3 and the corresponding F_3 with Map B, ranging from 0.65 at chromosome 4 to 1.20 at chromosome 10. Of 85 common marker intervals, 12% showed larger map distances *d* (> 10 cM) for the base population and 7% larger *d* values for F_2Syn3.

Discussion

Construction of linkage maps

Earlier studies on linkage mapping with intermated populations (Beavis et al. 1992; Liu et al. 1996; Lee et al. 2002) employed mapping methods developed for F_2 populations, similar to our linkage map construction in Map A. From a theoretical point of view, this method is questionable because construction of multilocus linkage maps includes minimizing or maximizing a target function, which fits the estimated map to the observed data. In the approach of Stam (1993), the sum of squared deviations of the estimated map distances from the observed map distances is minimized. In the maximum likelihood approach (Lander et al. 1987) for estimation of map distances, the likelihood for the observed data is maximized. In general, the extremes of the sum or product of a set of functions are not equal to the extremes of the sum or product after nonlinear transformations of the original functions. Therefore, the map that fits best the values *R* is not guaranteed to fit best the nonlinear transformations *r*. In consequence, linkage maps constructed with this approach should be regarded as an approximation of the maps constructed with an approach where accounting for the number of meioses is performed before minimization or maximization step, such as in Map B.

In our experimental data, the differences between the two mapping methods (Map A and B) became apparent with increasing marker intervals (Table 2), and in particular, four (A × B) and three (C × D) marker intervals differed by more than 10 cM (Figs. 1, 2). Moreover, the chromosome lengths differed considerably (~ 10–30 cM), as did the genome length. The reason for these differences is presumably the theoretical shortcoming of Map A. Hence, for obtaining precise linkage maps, it seems prudent to use algorithms that correct for the number of meioses before finding extremes of the underlying target function.

Map expansion

When compared with linkage maps constructed from F_2 populations, genetic maps constructed with intermated populations or populations of recombinant inbred lines developed from intermated populations were reported to show a map expansion (Beavis et al. 1992; Liu et al. 1996; Lee et al. 2002; Winkler et al. 2003; Teuscher et al. 2005). The maps in these studies were constructed by employing methods developed for F_2 populations to data from intermated populations. In

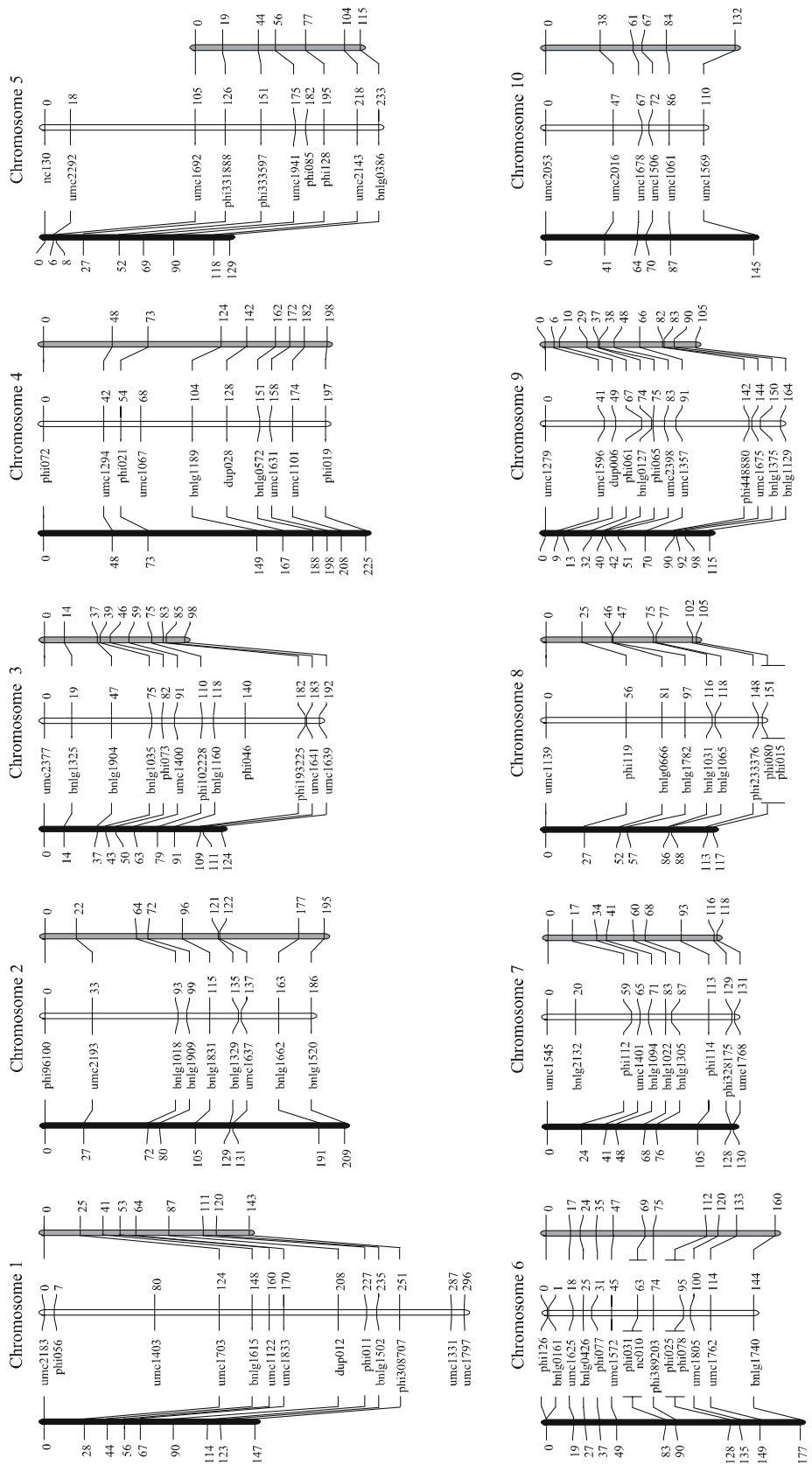

Fig. 1 Comparison of linkage maps of population F_2 (*white*) with population F_2Syn3 Map A (*grey*) and population F_2Syn3 Map B (*black*) of cross A × B

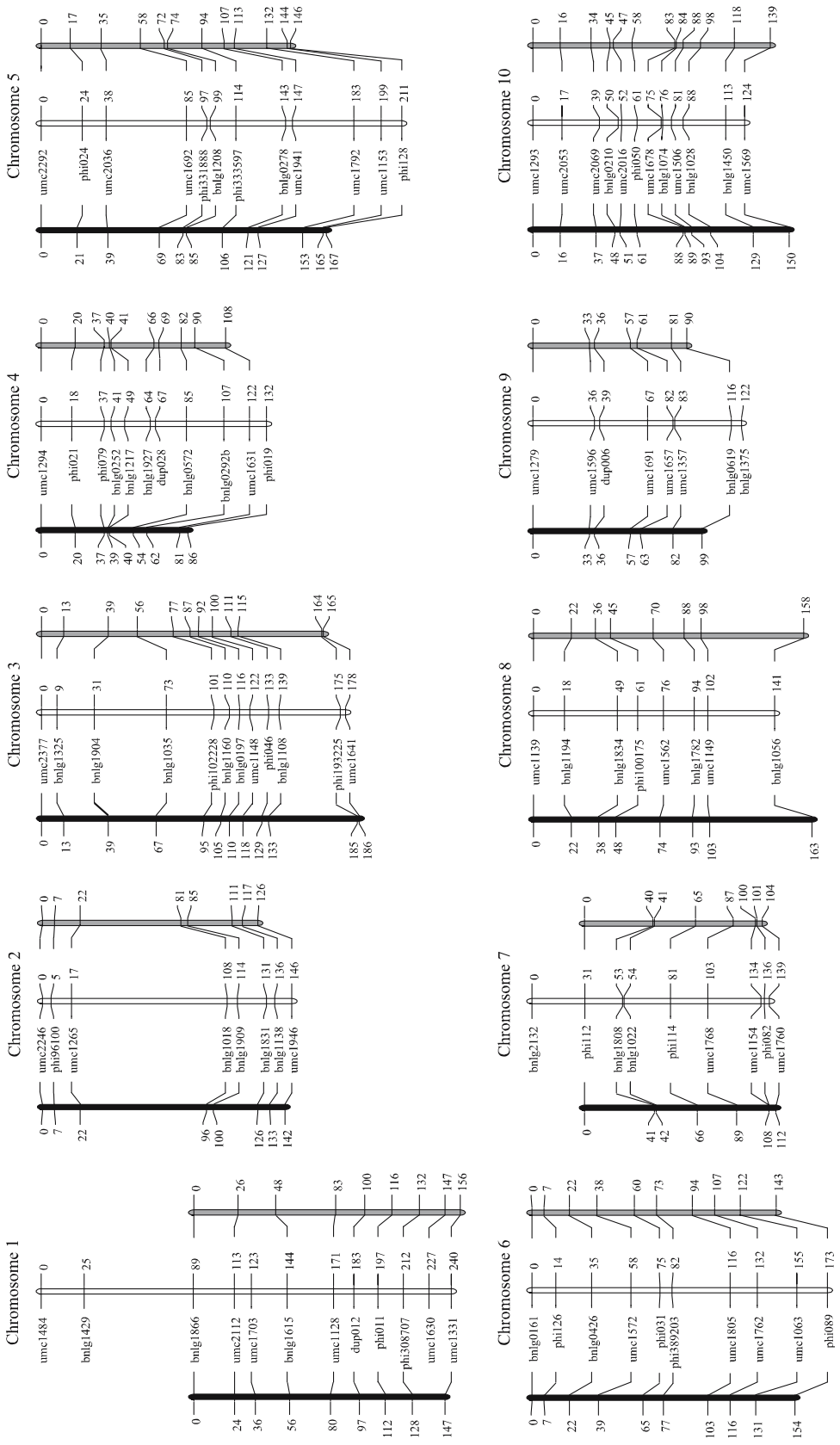

Fig. 2 Comparison of linkage maps of population F_3 (*white*) with population F_2Syn3 Map A (*grey*) and population F_2Syn3 Map B (*black*) of cross C × D

Table 1 Map distances (d in cM), correlation coefficients (ρ) and ratios of d between population F_2Syn3 and F_2 (F_3) of cross A × B and C × D

	Chromosomes										
	1	2	3	4	5	6	7	8	9	10	Total[a] Mean[b]
Cross A × B											
Population F_2											
No. of loci	13	9	12	10	10	14	10	9	12	6	105[a]
d	296	186	192	197	233	144	131	151	164	110	1803[a]
Comparison of population F_2 vs. F_2Syn3 with Map A											
No. of loci	9	9	10	9	7	13	9	9	12	6	93[a]
d											
F_2	251	186	192	197	128	144	131	151	164	110	1653[a]
F_2Syn3	143	195	98	198	115	160	118	105	105	132	1371[a]
ratio	0.57	1.05	0.51	1.01	0.90	1.11	0.90	0.70	0.64	1.20	0.86[b]
ρ	0.58	0.69	0.02	0.90**	0.52	0.88**	0.44	0.75*	0.38	0.73	0.59[b]
Comparison of population F_2 vs. F_2Syn3 with Map B											
No. of loci	9	9	11	9	9	14	9	9	12	6	97[a]
d											
F_2	251	186	192	197	233	144	131	151	164	110	1758[a]
F_2Syn3	147	209	124	225	129	177	130	117	115	145	1518[a]
ratio	0.58	1.13	0.65	1.14	0.56	1.23	0.99	0.77	0.70	1.32	0.91[b]
ρ	0.67	0.70	0.53	0.91**	–0.55	0.88**	0.64	0.79*	0.54	0.65	0.58[b]
Cross C × D											
Population F_3											
No. of loci	12	8	12	11	12	10	9	8	8	12	102[a]
d	240	146	178	132	211	173	139	141	122	124	1608[a]
Comparison of population F_3 vs. F_2Syn3 with Map A											
No. of loci	9	8	12	10	12	10	8	8	7	12	96[a]
d											
F_3	151	146	178	122	211	173	109	141	116	124	1471[a]
F_2Syn3	156	126	165	108	146	143	104	158	90	139	1336[a]
ratio	1.03	0.87	0.93	0.88	0.69	0.83	0.96	1.12	0.78	1.12	0.92[b]
ρ	0.78*	0.97**	0.76**	0.66	0.72*	0.56	0.72	0.74	0.48	0.80**	0.72[b]
Comparison of population F_3 vs. F_2Syn3 with Map B											
No. of loci	9	8	12	9	12	10	8	8	7	12	95[a]
d											
F_3	151	146	178	132	211	173	109	141	116	124	1481[a]
F_2Syn3	147	142	186	86	167	154	112	163	99	150	1406[a]
ratio	0.97	0.97	1.05	0.65	0.79	0.89	1.03	1.16	0.85	1.20	0.96[b]
ρ	0.82*	0.99**	0.87**	0.56	0.84**	0.54	0.80*	0.77*	0.62	0.86**	0.77[b]

*,**Significant at the 0.05 and 0.01 probability levels, respectively

[a]Total

[b]Mean

consequence, the resulting recombination frequencies R refer to the accumulated recombination events occurring with more than one meiosis. Visualizing such R values on a linkage map is misleading. Commonly, a map visualizes additive map distances, which are related via a mapping function to recombination frequencies r referring to a single meiosis. This convention is adhered to in all mapping studies and linkage mapping software we are aware of, the only exception being the above studies on mapping in intermated populations. We therefore conclude that the expansion of linkage maps, reported in the above studies, is simply a consequence of the fact that the amount of linkage visualized on the genetic map does not refer to a single meiosis, but to accumulated effects of all meioses occurring during the development of the mapping population.

For constructing Map A and B, we followed the convention to visualize map distances calculated from recombination frequencies referring to a single meiosis. Therefore, neither systematic expansion nor shrinkage of maps is expected when comparing maps from F_2 or F_3 populations with maps from intermated populations. In our experimental data, we observed considerable differences in the length of individual chromosomes depending on the type of mapping population (Figs. 1, 2). Some chromosomes expanded substantially in the intermated populations, but others

Table 2 Differences in marker intervals (d in cM) in population F_2Syn3 of cross A × B and C × D between Map A and B, resulting from two different mapping methods

Population	d_{F2} (cM)	No. of intervals	$\lvert d_{\text{Map A}} - d_{\text{Map B}} \rvert$[a] (cM)
A × B	0	3	0.0
	0–5	10	0.0
	5–10	16	0.1
	10–20	23	1.0
	20–50	24	2.0
	> 50	7	9.5
Σ		83	
C × D	0	0	0.0
	0–5	14	0.0
	5–10	11	0.1
	10–20	28	0.5
	20–50	29	4.4
	> 50	1	15.4
Σ		83	

[a] $\lvert d_{\text{Map A}} - d_{\text{Map B}} \rvert$ refers to means of absolute values

shrunk. The total map lengths of both intermated populations were smaller than in the base population. This marginal shrinkage is most likely attributable to the effects of drift occurring through the intermating. However, in accordance with theory, a systematic expansion of the entire linkage maps was not observed in our study. In conclusion, an expansion of linkage maps through intermating is neither expected nor observed in our data, under the above-mentioned convention.

High mapping resolution

The expected constant length of linkage maps is not related to the fact that intermating may enhance mapping resolution. High mapping resolution is a consequence of precise estimation of small recombination frequencies. Commonly, the precision of linkage mapping is measured by the standard error of the estimated recombination frequency (Allard 1956)

$$\sigma_r = \sqrt{\frac{1}{N\, i_r}}, \tag{4}$$

where N is the sample size and i_r is the mean amount of information from an individual, depending on the type of the mapping population and the (true unobservable) linkage between loci. The theoretical advantage of intermated populations for high-resolution mapping is that the value of i_r is greater than that of F_2 populations if map distances are small, as demonstrated numerically by Liu et al. (1996). This results in a more precise estimation of small recombination frequencies.

For the linkage map of cross A × B constructed with F_2 individuals, we have (Mather 1936)

$$i_r = \frac{2\left(1 - 3r + 3r^2\right)}{r\,(1-r)\,(1 - 2r + 2r^2)}, \tag{5}$$

for the map of cross C × D constructed with F_3 individuals (Allard 1956)

$$i_r = \frac{4\left(2 - 6r + 3r^2 + 4r^3\right)}{r\,(1-r)\,(2 + r^2)^2 (1 - 2r + 2r^2)}, \tag{6}$$

and for the maps constructed with F_2Syn t individuals (Liu et al. 1996)

$$i_r = \frac{(1-r)^{2t-2}\left[2(1-r) + t(1-2r)\right]^2 \left[1 + 3(1-2r)^2 (1-r)^{2t}\right]}{\left[1 - (1-2r)^4 (1-r)^{4t}\right]}. \tag{7}$$

It follows that for cross A × B a smaller standard error σ_r in the intermated population is expected than in F_2 if the true value $r \leq 0.1419$. For C × D, the threshold is $r \leq 0.1966$ (Fig. 3). In total, 49% of the marker intervals were estimated to be larger than the threshold in cross A × B, and 19% in cross C × D.

The main purpose of the linkage maps constructed for crosses A × B and C × D is QTL mapping and monitoring allele frequency changes in recurrent selection programs. In both applications, the focus lies on marker loci with tight linkage to gene loci. We expect that the increased accuracy in estimating small map distances in these studies is worth the additional effort in establishing a linkage map from the intermated populations.

Mapping of loosely linked markers

While linkage between closely linked markers is expected to be estimated more accurately from intermated populations than from corresponding F_2 or F_3 populations, the reverse is true for loosely linked markers (Fig. 3). Assignment of loci to linkage groups was difficult with our mapping data from the intermated populations, and several loci could not even be assigned to their linkage groups (Figs. 1, 2). We attribute this problem to (1) the large standard errors of estimated map distances between more distant markers when using intermated mapping populations, and (2) the fact that for a given map distance d, the value $r < R < 0.5$ and, therefore, a test on linkage (H_0: $r = 0.5$, H_0: $R = 0.5$) is expected to have greater power

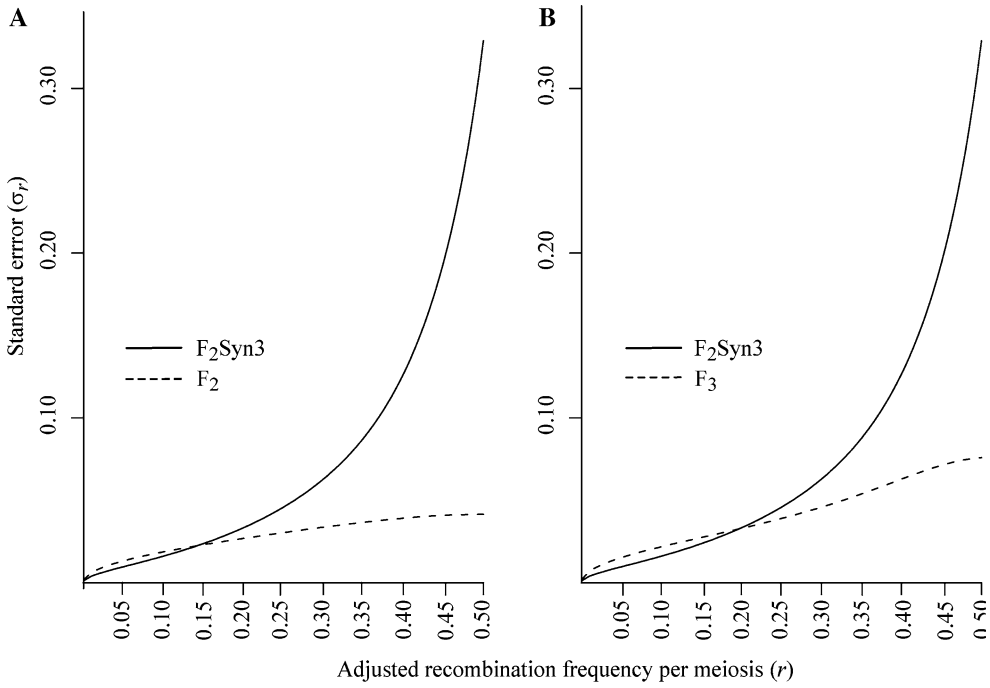

Fig. 3 Standard error σ_r of the adjusted recombination frequencies r calculated with Eq. 4 for (**a**) cross A × B population F_2 (N = 146) and F_2Syn3 (N = 148) ($\sigma_{r\ F2Syn3} < \sigma_{r\ F2}$, when $r < 0.1419$), and (**b**) cross C × D population F_3 (N = 110) and F_2Syn3 (N = 148) ($\sigma_{r\ F2Syn3} < \sigma_{r\ F3}$, when $r < 0.1966$)

for r than for R. Furthermore, gaps on the linkage map larger than 20 cM hamper severely the construction of linkage maps from intermated populations.

Sampling error caused by drift

The theoretical derivations of the amount of information and, consequently, the standard deviation of estimated recombination frequencies (Liu et al. 1996), are based on the assumption that intermating is conducted with infinitely large populations. In practice, however, finite samples from the population are used as parents for the next generation. As a consequence, random genetic drift occurs. Drift is increasing the standard error σ_r, and because the advantage of intermating is relatively small (Fig. 3), the increase in σ_r due to drift may even overrule the positive effects of intermating.

In our experiment, we attempted to minimize drift by using the chain crossing procedure for intermating, but this effect could not be quantified. Therefore, a final assessment of whether intermating actually increased the mapping accuracy remains open. As intermated populations are expected to play an increasingly important role in fine-mapping and map-based cloning of genes, we plan to conduct simulation studies for quantifying the effects of drift and obtaining an indication on the minimum effective population size required during the intermating generations.

Acknowledgments The present study was supported by a grant from the Deutsche Forschungsgemeinschaft, Grant No. FR 1615/1. The authors gratefully acknowledge the skilled technical assistance of Christine Beuter. We thank KWS Saat AG and Pioneer Hi-Bred Northern Europe for their support with the SSR screening. We are indebted to Johan van Ooijen (Kyazma BV, Wageningen, Netherlands) for providing an extension of JoinMap Version 3.0 for mapping intermated populations. We thank Jasmina Muminovic (University of Hohenheim) for critical reading and helpful comments on the manuscript.

References

Allard RW (1956) Formulas and tables to facilitate the calculation of recombination values in heredity. Hilgardia 24:235–278

Bailey N (1961) Introduction to the mathematical theory of genetic linkage. Oxford University Press, Oxford

Beavis WD, Lee M, Grant D, Hallauer AR, Owens T, Katt M, Blair D (1992) The influence of random mating on recombination among RFLP loci. Maize Genet Coop News lett 66:52–53

Darvasi A, Soller M (1995) Advanced intercross lines, an experimental population for fine genetic mapping. Genetics 141:1199–1207

Fisher R (1946) A system of scoring linkage data, with special reference to the pied factors in mice. Am Nat 80:568–578

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

Hoisington DA, Khairallah MM, Gonzales-de-Leon D (1994) Laboratory protocols. CIMMYT Applied Molecular Genetics Laboratory, CIMMYT, Mexico, DF

Holloway JL, Knapp SJ (1993) GMendel 3.0 Users Guide. Oregon State University, Department of Crop and Soil Science

Kosambi D (1944) The estimation of map distance from recombination values. Ann Eugen 12:172–175

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Lee M, Sharopova N, Beavis WD, Grant D, Katt M, Blair D, Hallauer A (2002) Expanding the genetic map of maize with the intermated B73 x Mo17 (IBM) population. Plant Mol Biol 48:453–461

Liu BH (1998) Statistical genomics: linkage, mapping, and QTL analysis. CRC, NY

Liu SC, Kowalski SP, Lan TH, Feldmann KA, Paterson AH (1996) Genome-wide high-resolution mapping by recurrent intermating using *Arabidopsis thaliana* as a model. Genetics 142:247–258

Lübberstedt T, Melchinger AE, Schön CC, Utz HF, Klein D (1997) QTL mapping in testcrosses of European flint lines of maize.1. Comparison of different testers for forage yield traits. Crop Sci 37:921–931

Mather K (1936) Types of linkage data and their value. Ann Eugen 7:251–264

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

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

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. Plant J 3:739–744

Teuscher F, Guiard V, Rudolph PE, Brockmann GA (2005) The map expansion obtained with recombinant inbred strains and intermated recombinant inbred populations for finite generation designs. Genetics 170:875–879

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 R (2001) JoinMap Version 3.0 software for the calculation of genetic linkage maps. Plant Research International, Wageningen, The Netherlands

Weir BS (1996) Genetic data analysis II. Sinauer Associates, Sunderland, MA

Winkler CR, Jensen NM, Cooper M, Podlich DW, Smith OS (2003) On the determination of recombination rates in intermated recombinant inbred populations. Genetics 164:741–745

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.

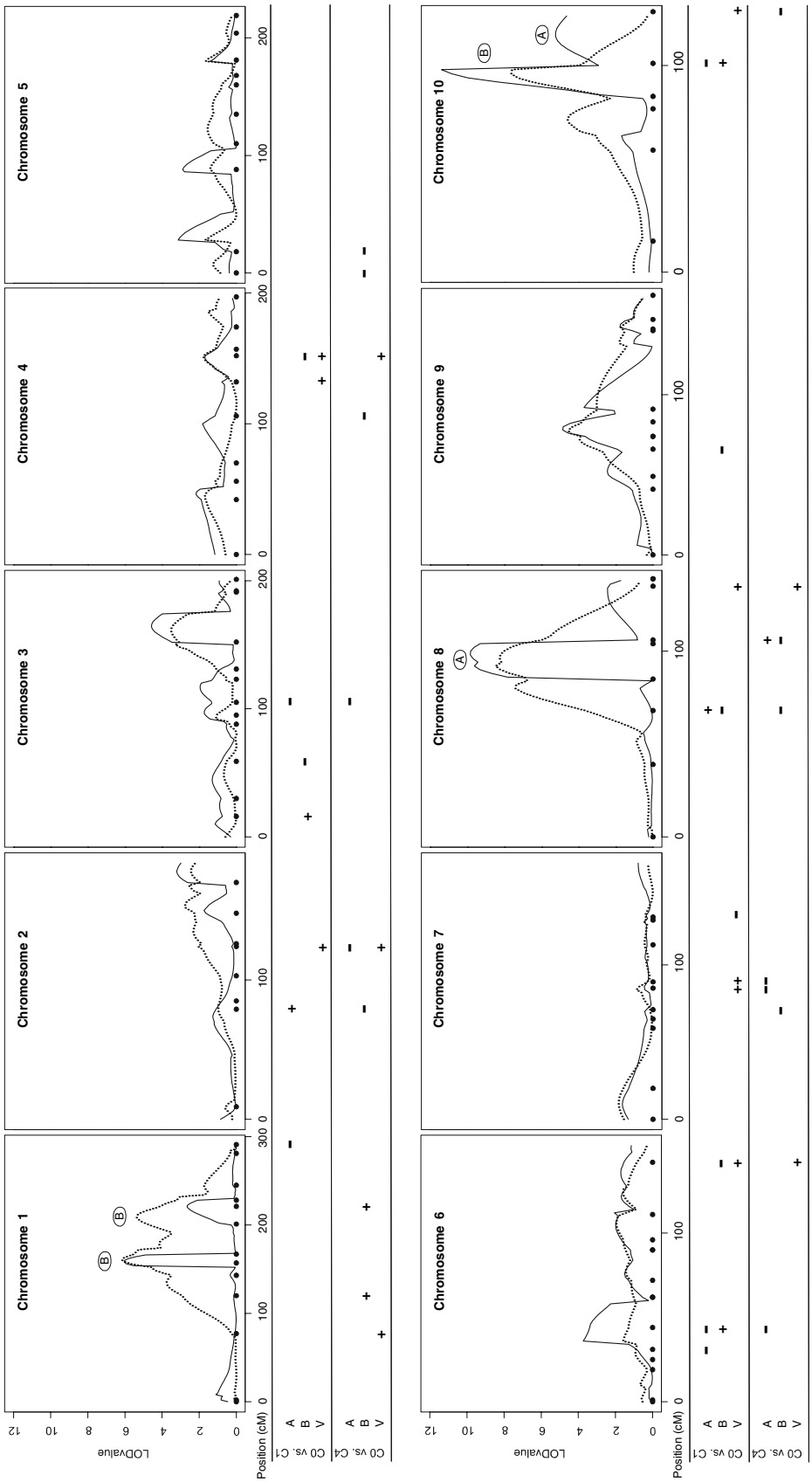

Fig. 1 QTL likelihood maps indicating LOD scores for selection index in population A × B. *Curves* represent results from simple interval mapping (SIM, *dotted line*) and composite interval mapping (CIM, *solid line*). *Letters A and B* in *ellipses* indicate which parent contributed the favorable allele. 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

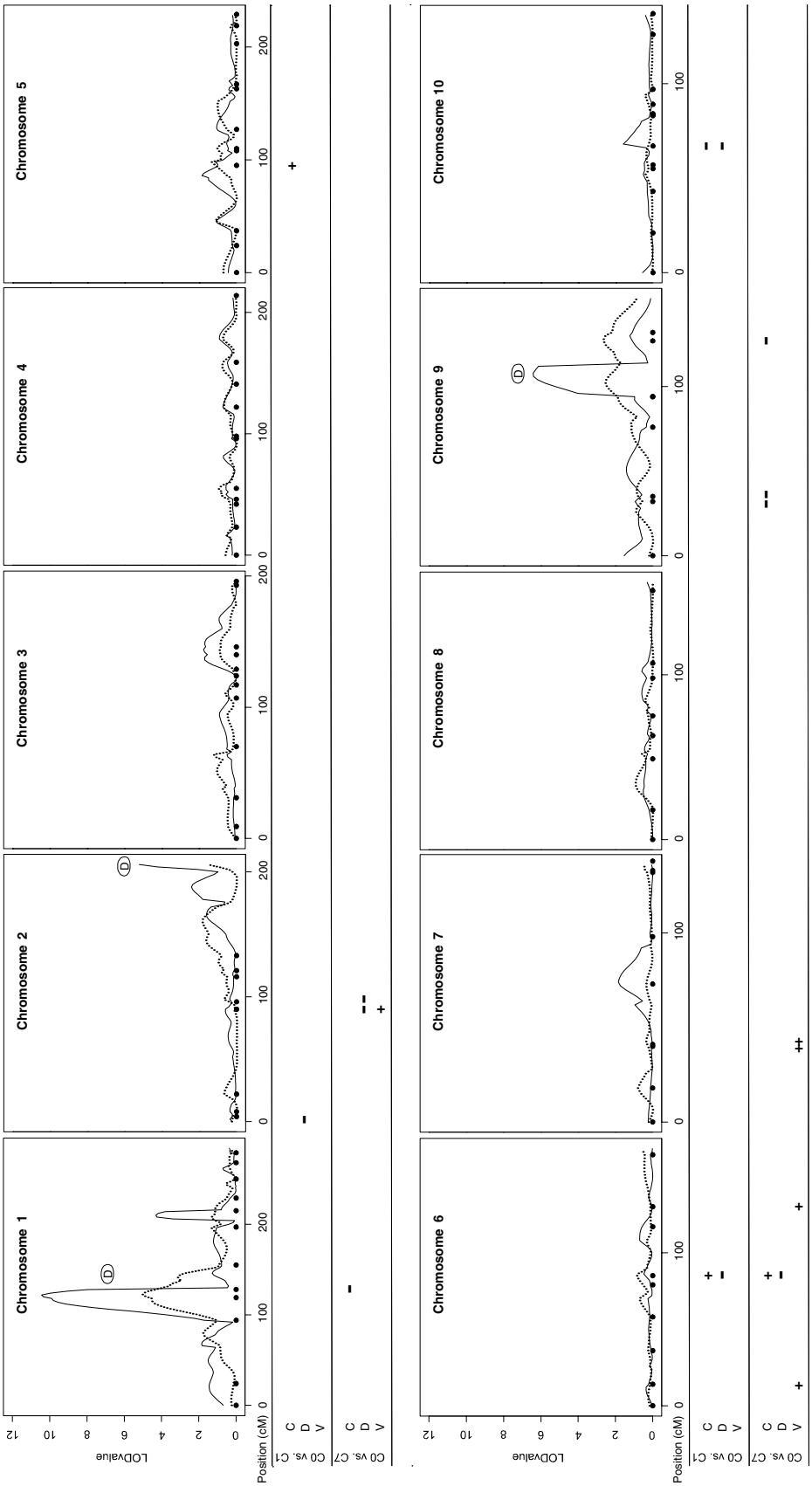

Fig. 2 QTL likelihood maps indicating LOD scores for selection index in population C × D. *Curves* represent results from simple interval mapping (SIM, *dotted line*) and composite interval mapping (CIM, *solid line*). *Letters C and D* in *ellipses* indicate which parent contributed the favorable allele. 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

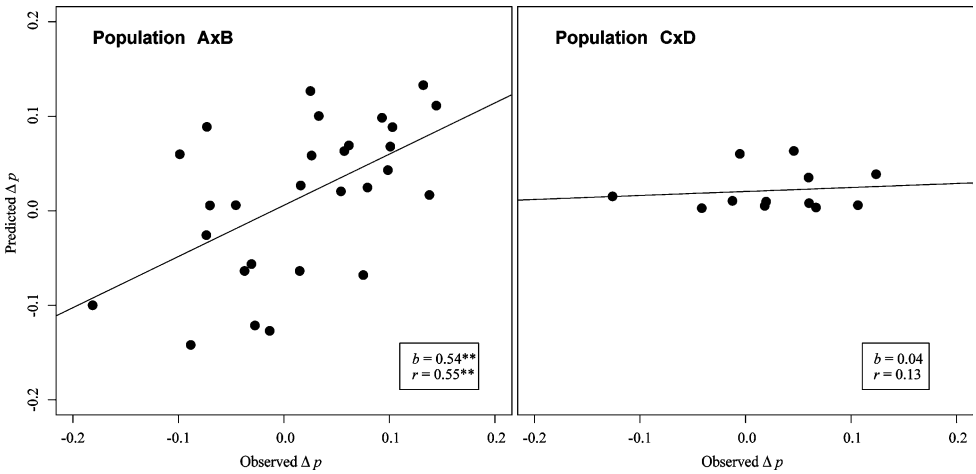

Fig. 3 Correlation between observed changes in allele frequency of the maker loci for allele A (A × B) and C (C × D) (C0–C1; Δp) and predicted Δp for the selection index calculated with the modified formula of Hartl and Clark (1997), where b is the slope coefficient and r is the correlation coefficient

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

Theor Appl Genet (2007) 115:289–297
DOI 10.1007/s00122-007-0544-y

ORIGINAL PAPER

Linkage disequilibrium in two European F_2 flint maize populations under modified recurrent full-sib selection

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

Received: 18 December 2006 / Accepted: 27 March 2007 / Published online: 28 April 2007

Abstract According to quantitative genetic theory, linkage disequilibrium (LD) can hamper the short- and long-term selection response in recurrent selection (RS) programs. We analyzed LD in two European flint maize populations, KW1265 × D146 (A × B) and D145 × KW1292 (C × D), under modified recurrent full-sib selection. Our objectives were to investigate (1) the decay of initial parental LD present in F_2 populations by three generations of intermating, (2) the generation of new LD in four (A × B) and seven (C × D) selection cycles, and (3) the relationship between LD changes and estimates of the additive genetic variance. We analyzed the F_2 and the intermated populations as well as all selection cycles with 104 (A × B) and 101 (C × D) simple sequence repeat (SSR) markers with a uniform coverage of the entire maize genome. The LD coefficient D and the composite LD measure Δ were estimated and significance tests for LD were performed. LD was reduced by intermating as expected from theory. A directional generation of negative LD between favorable alleles could not be observed during the selection cycles. However, considerable undirectional changes in D were observed, which we attributed to genetic sampling due to the finite population size used for recombination. Consequently, a long-term reduction of the additive genetic variance due to negative LD was not observed. Our experimental results support the hypothesis that in practical RS programs with maize, LD generated by selection is not a limiting factor for obtaining a high selection response.

Communicated by M. Xu.

K.C. Falke and H.P. Maurer are contributed equally to the manuscript.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-007-0544-y) contains supplementary material, which is available to authorized users.

K. C. Falke · H. P. Maurer · A. E. Melchinger (✉) · C. Flachenecker · 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 strategy designed to ensure long-term selection response by increasing the frequency of favorable alleles while maintaining the genetic variance in populations (Hallauer 1985). Selection generates linkage disequilibrium (LD) between alleles whose frequencies were increased by selection (Nei 1963). This newly generated LD is negative, which means by definition that the covariance between favorable alleles at different loci is negative. Negative LD reduces the additive genetic variance (σ_A^2) of the traits under selection (Bulmer 1971), and thus may result in a decline of long-term selection response.

In RS procedures in maize, open-pollinated-varieties or synthetics have mostly been employed as base populations (Hallauer and Miranda 1988; Bernardo 2002). In contrast, F_2 base populations have been used only in a few studies but mostly with remarkable success (Russell et al. 1973; Genter 1982; Moll 1991; Landi and Frascaroli 1993). In F_2 populations, there is a positive covariance between alleles originating from the same parental line at linked loci. To reduce the extent of this parental LD and its negative

effects on the selection gain, Johnson (1982) suggested three generations of random intermating before starting with RS. Although parental LD in base populations of RS programs may severely limit the selection response, to our knowledge, no previous results have been published on the reduction in parental LD through intermating.

Allele frequency changes due to selection were analyzed in several experimental studies (Labate et al. 1999; Pinto et al. 2003; Coque and Gallais 2006; Falke et al. 2007). However, the extent of LD between alleles whose frequencies were increased by selection, and the effect of this LD on σ_A^2 and the selection response were only investigated theoretically and with computer simulations (Hospital and Chevalet 1996). No experimental results have been published yet.

As a part of a maize breeding project, long-term recurrent full-sib selection programs with two European F_2 flint maize populations, previously employed in QTL studies (Schön et al. 1994; Mihaljevic et al. 2004, 2005a), were initiated for analyzing the selection response. After completing three generations of intermating, four cycles of RS were conducted in cross KW1265 × D146 (A × B) and seven cycles in cross D145 × KW1292 (C × D) to determine changes in the population structure at the phenotypic and molecular level. In several companion studies, we investigated the selection response at the phenotypic level (Flachenecker et al. 2006a, b, c) and determined allele frequency changes during RS at the molecular level (Falke et al. 2007).

The main goal of the present study was to analyze the LD in the RS programs of these two F_2 flint maize populations. In particular, our objectives were to (1) investigate the genetic distance among the base and intermated populations as well as the subsequent selection cycles, (2) examine the decay of parental LD between linked loci after three generations of intermating, (3) determine the generation of new LD between alleles whose frequencies were increased by selection during the RS procedure, and (4) relate changes in LD to estimates of σ_A^2.

Materials and methods

Plant materials

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 to produce 380 $F_{2:3}$ lines of cross A × B and 140 $F_{3:4}$ lines of cross C × D. Lines A and D are proprietary elite inbreds developed by KWS SAAT AG (Einbeck, Germany), lines B and C are public elite inbreds developed by the University of Hohenheim (Stuttgart, Germany). The F_2 populations were intermated for three generations by applying a chain-crossing scheme with 240 F_2 plants to produce the F_2Syn3 populations. 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 completed for recombination of superior genotypes, using a pseudo-factorial mating scheme based on the suggestion of Cockerham and Burrows (1980). Full-sib families were selected on the basis of a 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 trait values received a weight of 1 for grain yield and 2 for dry matter content [i.e., the weight vector was $\mathbf{b}' = (1,2)$]. Evaluation of the full-sib families in each selection cycle was conducted in field trials at three locations in South Germany. The experimental design in each environment was an α-lattice (10 × 15) with three replications.

SSR analyses

A total of 104 (A × B) and 101 (C × D) codominant SSR markers polymorphic between the parental lines and warranting a uniform coverage over the entire maize genome was employed for genotyping. For SSR marker analyses, we used random subsets of 146 $F_{2:3}$ lines out of the 380 $F_{2:3}$ lines in A × B and 110 $F_{3:4}$ lines out of the 140 $F_{3:4}$ lines in C × D as well as 148 F_2Syn3 plants and the parents of the 36 families with the highest selection index in each selection cycle of both RS programs. DNA extraction as well as SSR amplification and detection were described in detail by Falke et al. (2006, 2007).

Principal coordinate analysis

Modified Rogers' distances (MRD, Wright 1978, p. 91) were estimated between parental lines (A, B and C, D), the population of $F_{2:3}$ (A × B) or $F_{3:4}$ (C × D) lines, the intermated populations F_2Syn3, and the various selection cycles of both RS programs, A × B and C × D, using 104 (A × B) and 101 (C × D) SSR marker loci. Based on MRD estimates, principal coordinate analyses (PCoA) (Gower 1966) were carried out to reveal associations among the base and intermated populations and the various selection cycles of the RS programs.

Assessment of parental LD in the intermated populations

Parental LD was assessed for all marker pairs in population of $F_{2:3}$ lines (A × B) and $F_{3:4}$ lines (C × D) and their corresponding F_2Syn3 populations with the linkage disequilibrium coefficient D (Weir 1996, p. 113)

$$D_{xy} = p_{xy} - p_x p_y. \quad (1)$$

Here, p_x and p_y are the allele frequencies of the alleles x and y at two loci originating from the same parental line, and p_{xy} is the frequency of gametes carrying both alleles x and y. The gametic frequencies p_{xy} were estimated with a maximum likelihood approach (Weir 1996, pp. 73–76), which assumes Hardy–Weinberg equilibrium (HWE) at both loci. The LD coefficient D was plotted as a function of recombination frequencies, which were calculated with the inverse of Haldane's (1919) mapping function from the map distances estimated in $F_{2:3}$ (A × B) and $F_{3:4}$ (C × D) by Falke et al. (2007).

No theoretical results exist on the expected decay of parental LD with increasing recombination frequencies for intermated populations produced by the chain-crossing method. Therefore, we simulated the intermating procedure (assuming no interference in crossover formation) with 500 replications to determine the expected LD decay. Observed values were compared with these expectations.

All loci pairs in the base and intermated populations were tested for significant parental LD with a Monte Carlo approximation of Fisher's exact test (Zaykin et al. 1995) at a significance level of $\alpha = 0.05$. For the underlying Monte Carlo method, 17,000 replications were used (Guo and Thompson 1992).

Assessment of LD in selection cycles

In the intermated populations and in the selection cycles of both RS programs, we investigated the LD between marker alleles tightly linked to QTL alleles with a positive effect on the selection index. Subsequently, we refer to these alleles as ''favorable alleles''. To determine the set of favorable alleles, we employed the following strategy. We chose all marker loci whose significant allele frequency changes were attributed to selection, as detected with Waples' test (1989) in a companion study (Falke et al. 2007). For each marker, the allele with the largest positive allele frequency change Δp from F_2Syn3 to the final selection cycle C4 (A × B) and C7 (C × D) was assigned to the set of favorable alleles. This strategy was necessary because non-parental alleles were observed since the initial selection cycles in both populations due to a contamination with foreign pollen (A × B: C1; C × D: C2) (Falke et al. 2007). The sets of favorable alleles consisted of 27 (A × B) and 14 (C × D) marker alleles

LD between the favorable alleles was assessed by (1) the linkage disequilibrium coefficient D (Eq. 1) and (Eq. 2) the composite linkage disequilibrium measure Δ (Weir 1996). The LD coefficient D was chosen because it is in direct relationship with σ_A^2 (Lynch and Walsh 1998, p. 102). The measure Δ was chosen, because it does not require the ML estimation of gamete frequencies, and therefore provides a means to assess the robustness of D with respect to deviations from HWE. Significance of D and Δ was tested with χ^2 tests. All analyses were carried out for populations F_2Syn3 and the selection cycles of both RS programs.

For the analysis of the relation between LD and σ_A^2 we employed the LD coefficient D and restricted maximum likelihood (REML) estimates of the variance components determined by Flachenecker et al. (2006a, b).

All computations and simulations 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

PCoA based on MRD estimates between population $F_{2:3}$ (A × B) and $F_{3:4}$ (C × D), F_2Syn3 and all selection cycles explained 98.3% (A × B) and 96.0% (C × D) of the molecular variance by the first two principal coordinates (PCs) (Fig. 1). In both RS programs, PC1 revealed a clear separation of the parental lines (A and B, C and D). All populations and selection cycles were in between the two parental lines with respect to PC1. In comparison with PC1, PC2 explained considerably less molecular variance (A × B: 6.6%; C × D: 10.4%). PC2 separated population $F_{2:3}$ (A × B) and $F_{3:4}$ (C × D), their corresponding F_2Syn3 and the selection cycles (A × B: C1 – C4; C × D:

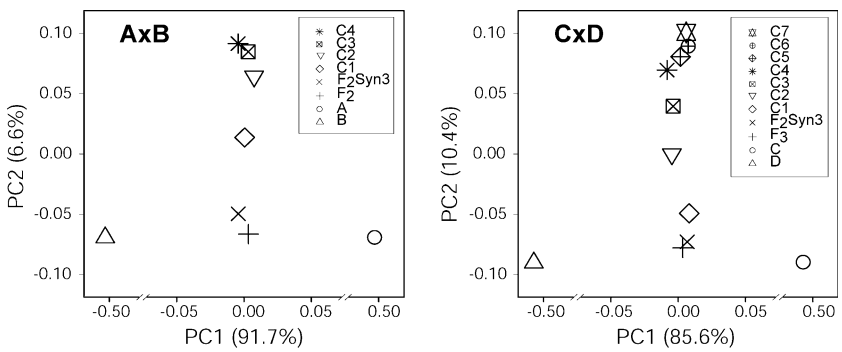

Fig. 1 Principal coordinate analysis of all populations and selection cycles from RS programs of A × B and C × D based on modified Rogers' distance calculated from SSR marker loci. PC1 and PC2 refer to the first and second principal coordinate, respectively. *Numbers in parentheses* indicate the proportion of molecular variance explained by the principal coordinates

C1 – C7) according to their chronological order. Final selection cycles (C4 and C7) were clearly separated from population $F_{2:3}$ (A × B) and $F_{3:4}$ (C × D). MRD estimates for all populations and selection cycles were presented in a supplementary table for both RS programs. Performing the PCoA between populations and selection cycles without parental lines, 85.1% (A × B) and 81.7% (C × D) of the molecular variance was explained by the first two PCs (data not shown). The proportion of the molecular variance, which was assessed by PC2 in the analysis with parents, was also captured by PC1 in the analysis without parents. The proportion of molecular variance from the populations and selection cycles was higher without parental lines (A × B: 73.6%; C × D: 70.1%) than with them (A × B: 6.6%; C × D: 10.4%). PC2 was approximately a quadratic function of PC1, so the ranking of the populations and selection cycles was unaltered compared to the analysis with parents.

The proportion of linked loci pairs in significant parental LD detected with Fisher's exact test decreased from 0.522 ($F_{2:3}$) to 0.357 (F_2Syn3) in A × B and from 0.488 ($F_{3:4}$) to 0.309 (F_2Syn3) in C × D. The extent of parental LD in population F_2Syn3 decreased with increasing recombination frequency, in accordance with the expected values obtained from simulations (Fig. 2).

Favorable alleles were mostly in linkage equilibrium in population F_2Syn3 of both RS programs (Fig. 3). In cycle C1, positive LD (indicated by red shading) was generated in both RS programs. Furthermore, the extent of negative LD (indicated by blue shading) was increased in C1 of C × D. In the final selection cycles (A × B: C4; C × D: C7), the extent of negative LD was increased in both RS programs, while the extent of positive LD was decreased in A × B and increased in C × D. Major deviations between the two LD measures D and Δ were neither observed in population F_2Syn3 nor in the first or final selection cycles. For several loci pairs, LD could not be determined (missing values, indicated by gray shading), because (1) non-parental alleles were absent in population F_2Syn3 but only present in later selection cycles or (2) the gamete frequency could not be determined with the ML procedure because of missing genotype classes.

In the RS program of A × B, the increase in the number of loci pairs with significantly positive LD from F_2Syn3 to cycle C1 detected by the χ^2 test was associated with the highest estimates of σ_A^2 for the selection index (Table 1, selection response and estimates of variance components were taken from Flachenecker et al. 2006a, b). In subsequent selection cycles, an increase in the number of loci pairs in negative LD was associated with a significant decrease in σ_A^2 for selection index. In C × D, the non-significant decrease in σ_A^2 for selection index was associated with a non-directional increase and decrease in loci pairs in positive and negative LD during the selection procedure.

Discussion

In previous studies, the selection response of modified full-sib recurrent selection programs in populations A × B and C × D was evaluated at the phenotypic level with classical quantitative genetic methods (Flachenecker et al. 2006a, b, c). At the molecular level, we investigated the effects of random genetic drift and selection on allele frequency changes in QTL regions by using SSR markers (Falke et al. 2007). We observed a comparatively high realized selection gain for selection index (A × B: 5.25%; C × D: 3.64% per selection cycle; Flachenecker et al. 2006a, b). Several QTL regions for selection index were detected in population $F_{2:3}$ (A × B) and $F_{3:4}$ (C × D) (Falke et al. 2007). At some of them, flanking markers showed significant changes in allele frequencies due to selection during the RS procedure (Falke et al. 2007). In the present study, we analyzed the development of the LD over several cycles of RS as well as the effects of changes in allele frequencies and LD on trends of σ_A^2.

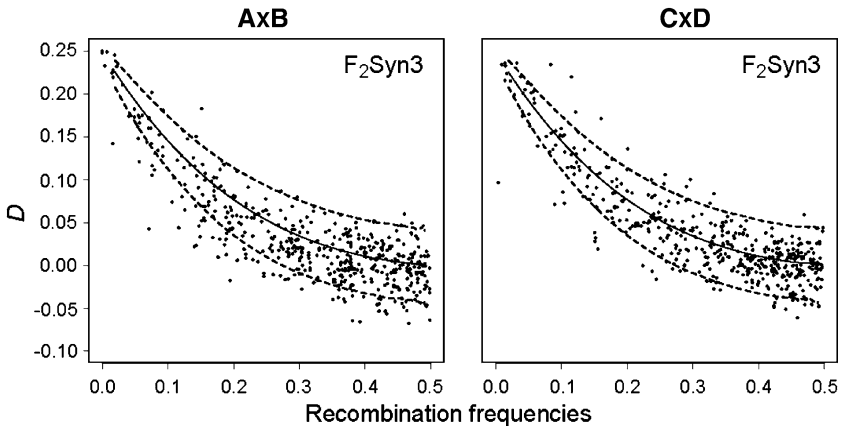

Fig. 2 Decay of parental linkage disequilibrium (LD) after three generations of intermating in populations A × B and C × D. Observed parental LD (*circles*) between linked pairs of SSR loci is plotted as a function of recombination frequencies. The mean and the respective 5% and 95% quantiles of the simulated LD decay are plotted as *solid and dashed lines*, respectively

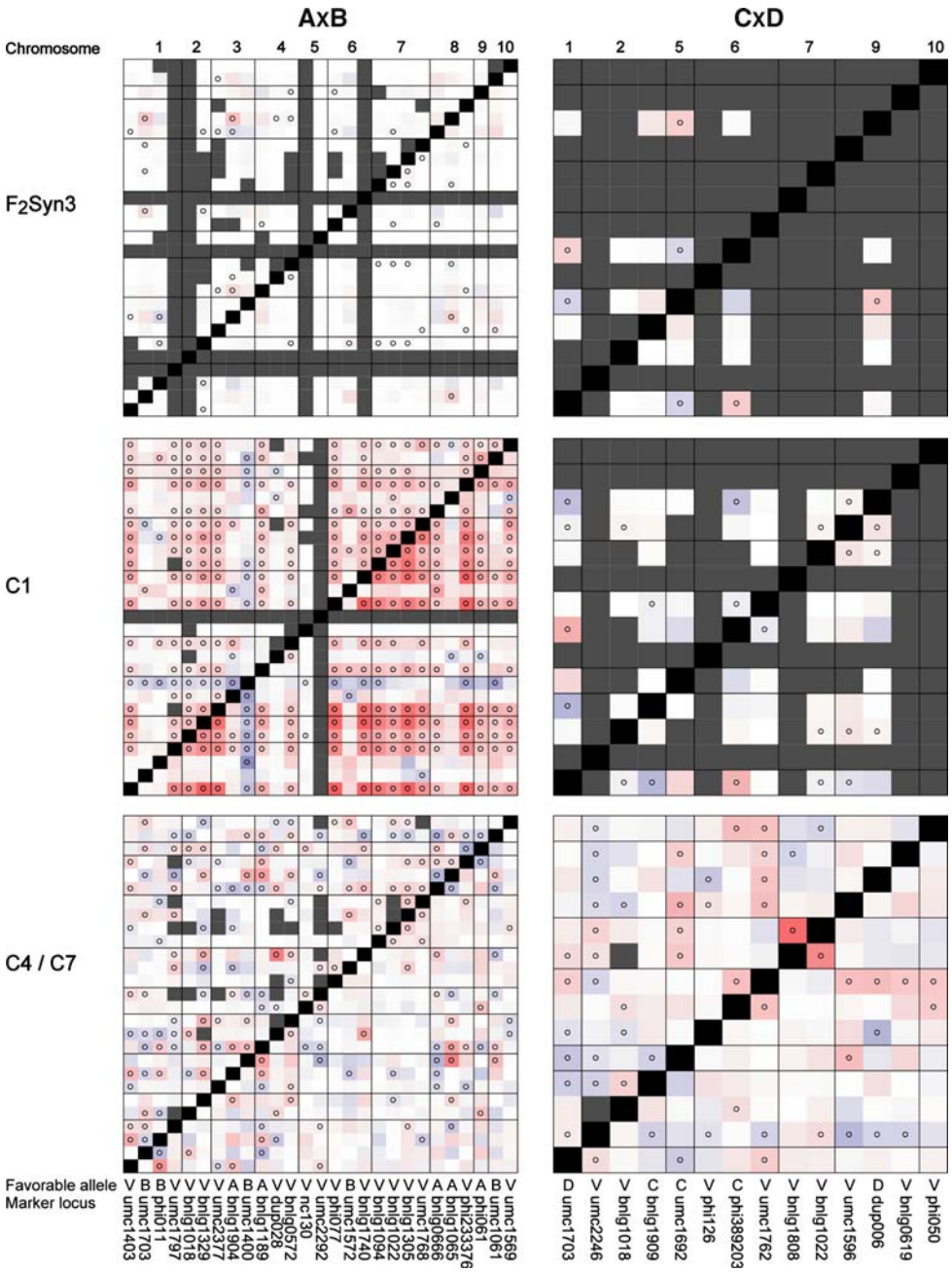

Fig. 3 Linkage disequilibrium (*LD*) between marker alleles whose frequencies were increased by selection measured as *D* (above diagonal) and Δ (below diagonal) in F_2Syn3 and in cycle C1 of both RS programs, in cycle C4 of A × B, and in cycle C7 of C × D. The respective favorable alleles are presented below the matrix (parental alleles A and B, C and D, and non-parental alleles V). Chromosomes are separated by *horizontal and vertical black lines*, *red coloring* indicates positive LD, *blue* negative LD, and *grey* missing values. Circles indicate significant LD detected with a χ^2 test

Genetic diversity

For both RS programs, PC1 separated the parental lines with a MRD of 1.0, as expected from theory due to the employment of exclusively polymorphic SSR markers (Fig. 1). The population of $F_{2:3}$ lines (A × B) and $F_{3:4}$ lines (C × D) together with the intermated populations F_2Syn3 were in the center between respective parental lines. During the selection procedure, the selection cycles varied only slightly around this origin of PC1 and no directional shift in favor of one parental line was observed. This suggests that the parental lines of each selection program carried approximately similar numbers of favorable alleles, and the observed selection response was driven by recombination of alleles of different parental origin and selection of favorable allele combinations. This is in agreement with the small estimates of the sum of additive effects estimated in a generation mean analysis as well as the results from QTL mapping studies (Mihaljevic et al. 2005a, b).

PC2 separated the $F_{2:3}$ or $F_{3:4}$ from the intermated populations and the selection cycles according to their chronological order, with larger differences between sub-

Table 1 Selection response for selection index (±SE) and its components relative to the mean performance of six F_2 checks. Restricted maximum likelihood (REML) estimates of the additive genetic variance ($\sigma_A^2 \pm$ SE), their mean across selection cycles and the coefficient (*b*) of the linear regression across selection cycles of RS programs A × B and C × D. Proportion of loci pairs with favorable alleles in significant LD (P 0.05; χ^2 test statistic based on D) in population F_2Syn3 and the various selection cycles of both RS programs.

Selection cycle	Selection index (%)		Grain yield (% of F_2)		Grain moisture (% of F_2)		LD between favorable alleles		
	Selection response ± SE	σ_A^2 ± SE	Selection response ± SE	σ_A^2 ± SE	Selection response ± SE	σ_A^2 ± SE	No. of loci pairs	Proportion of loci pairs in significant LD	
								Negative	Positive
A × B									
F_2Syn3							223	0.067	0.040
C1	315.3 ± 2.3	207.5 ± 82.5*	120.2 ± 2.4	0.38 ± 0.16*	102.4 ± 0.6	92.5 ± 13.6**	314	0.067	0.475
C2	323.2 ± 4.7	198.9 ± 31.9**	129.1 ± 4.3	0.25 ± 0.05**	103.0 ± 1.0	111.8 ± 16.7**	270	0.085	0.437
C3	321.5 ± 4.9	174.5 ± 64.3**	124.2 ± 2.6	0.27 ± 0.13*	101.4 ± 1.7	167.2 ± 24.0**	303	0.172	0.155
C4	317.6 ± 4.9	140.3 ± 21.9**	123.1 ± 2.2	0.12 ± 0.07	102.8 ± 1.7	183.9 ± 24.4**	324	0.160	0.210
Mean		180.3 ± 49.9		0.26 ± 0.1		138.8 ± 197			
B	–0.31	–22.6*	0.10	–0.08	0.00	32.9*			
C × D									
F_2Syn3							12	0.167	0.167
C1	298.0 ± 1.8	49.4 ± 23.9	97.9 ± 1.9	0.21 ± 0.11	100.0 ± 0.2	32.8 ± 15.4*	30	0.200	0.133
C2	304.0 ± 5.1	274.1 ± 96.6**	105.4 ± 4.3	0.54 ± 0.22*	100.7 ± 0.7	162.1 ± 23.5**	52	0.173	0.250
C3	323.3 ± 5.3	226.2 ± 32.9**	119.6 ± 5.0	0.69 ± 0.10**	98.1 ± 0.7	188.7 ± 24.7**	57	0.228	0.474
C4	339.6 ± 4.9	25.3 ± 54.0	135.4 ± 4.5	0.00	97.9 ± 0.6	104.7 ± 15.7**	80	0.212	0.288
C5	339.6 ± 5.2	62.6 ± 45.7	132.4 ± 5.7	0.49 ± 0.25*	96.5 ± 1.0	226.7 ± 30.9**	84	0.167	0.143
C6	343.3 ± 5.8	70.0 ± 45.6	137.0 ± 6.0	0.50 ± 0.28	96.8 ± 1.1	229.4 ± 31.9**	81	0.173	0.148
C7	369.4 ± 7.1	123.4 ± 111.1	157.5 ± 8.5	0.17 ± 0.11	94.1 ± 1.7	475.6 ± 197.9*	89	0.180	0.213
Mean		118.7 ± 58.5		0.37 ± 0.15		202.9 ± 48.6			
B	11.2**	–3.6	9.1**	–0.02	–1.1**	37.2			

Selection response and variance components for selection index and its components were determined by Flachenecker (2006a, b)

*, ** Significant at the 0.05 and 0.01 probability level, respectively

sequent cycles in the earlier selection cycles than in the later ones. However, these differences do not match the differences in selection response for the selection index attained in the various selection cycles (Table 1). For example, in RS program C × D the largest selection response was realized in cycle C7, but the MRD between cycles C6 and C7 was very small (Fig. 1). In contrast, the differences between the selection cycles with respect to PC2 agreed well with migration effects due to contamination with foreign pollen observed since the initial selection cycles (Falke et al. 2007). Hence, the differences in PC2 are most probably caused by migration and, in consequence, migration had a considerable influence on the genetic structure of the population in both RS programs. Therefore, migration and subsequent selection of the migrated alleles might very well have been an important factor contributing to the relatively large observed selection response and the small reduction in σ_A^2 in the RS program of C × D.

Decay of parental LD due to intermating

Parental LD in the base population of an RS program is expected to hamper the possible short- and long-term selection gain. Therefore, Johnson (1982) suggested three generations of random intermating before starting an RS program to reduce the parental LD and its negative effects on the selection gain. We adopted this idea and conducted three generations of intermating with our F_2 populations. The observed decay of LD was in good agreement with the theoretical expectations obtained from simulations (Fig. 2). However, whether the observed reduction of linked loci pairs in significant LD through intermating was an important cause of the large realized selection response cannot be definitely answered by our experimental setup. In particular, the breeding success depends on linkage and the linkage phase relationship between favorable alleles. Consequently, an interesting open question for further research is whether the initial time lag in the start of an RS

program due to intermating the base population pays off in terms of a greater realized selection response.

Another goal of intermating generations is to obtain an increased mapping accuracy of tightly linked loci. Intermated F_2 and intermated recombinant inbred line populations have been employed as mapping populations and the authors reported that the maps showed a map expansion (Liu et al. 1996; Lee et al. 2002; Winkler et al. 2003; Falque et al. 2005; Teuscher et al. 2005; Teuscher and Broman 2007). However, this is misleading (Falke et al. 2006; Martin and Hospital 2006), because since the introduction of map distances these were estimated with recombination frequencies, referring to only a single meiosis but not to recombination events accumulated in all intermating generations (Haldane 1919; Kosambi 1944; Stam 1993).

Increase in positive LD during the selection cycles

In both selection programs and at many loci pairs, we observed an increase in positive LD (i.e., a positive covariance) between favorable alleles (Fig. 3) during the selection cycles. This increase can be explained by the employed mating scheme of Cockerham and Burrows (1980) and the contamination with foreign pollen (Falke et al. 2007).

In the mating scheme of Cockerham and Burrows (1980), the best third of the selected plants (used as male parents in the pseudo factorial mating design) transmits their gametes with twice the dose as the remaining two thirds (used as female parents). This can result in positive LD between the alleles responsible for the superior performance of the male parents.

The contamination with foreign pollen (Falke et al. 2007) resulted in non-parental alleles with small allele frequency in the early cycles of the selection program. If these migrated alleles had a selective advantage, they showed a relatively large allele frequency change Δp, which was often detected by Waples' test (Falke et al. 2007). Therefore, many non-parental alleles were included in the sets of favorable alleles (Fig. 3) and, due to their linkage, contributed considerably to the positive LD observed in the final selection cycles C4 (A × B) and C7 (C × D) (Fig. 3).

Increase in negative LD due to selection

Selection is expected to increase the frequency of favorable alleles and simultaneously build up a negative LD between them (Bulmer 1971). Labate et al. (2000) observed slight increases in LD over 12 cycles of reciprocal RS. In this study, especially loci near fixation showed significant LD between each other. With simulations studies, Hospital and Chevalet (1996) found that the extent of negative LD increases at the beginning and decreases in later stages of the selection process, irrespective of the recombination frequency between linked loci. The time, when LD started decreasing, corresponded to the time when some loci reached fixation.

If the change in allele frequencies due to one cycle of selection at two loci is Δp_x and Δp_y, then the LD in the selection fraction is (Nei 1963)

$$D_{xy}^{(1)} = -\Delta p_x \Delta p_y. \quad (2)$$

The build-up of LD during selection is the net effect of two opposing forces: selection increases LD, while recombination reduces LD. However, this newly generated LD is reduced in every subsequent generation by recombination. As a first approximation for the LD reduction per recombination with our full-sib selection scheme, we use the expected LD reduction for random mating, which is $(1 - r)$, where r is the recombination frequency between the two loci (Falconer and Mackay 1996, p. 18). Assuming that in each generation (1) the newly generated LD due to selection equals D_{xy} (Eq. 1) and (2) the LD from the previous generation is reduced by a factor of $(1 - r)$, we have after n generations of selection and recombination (Nei 1963)

$$D_{xy}^{(n)} \leq {}^{1}/_{r}\,[1 - (1 - r)^n]\, D_{xy}^{(1)}. \quad (3)$$

We use two simple examples to illustrate the numerical magnitude of such expected allele frequency changes. (1) Very large allele frequency changes due to selection of $\Delta p_x = \Delta p_y = 0.2$ result in $D_{xy}^{(1)} = -0.04$. (2) $\Delta p_x = \Delta p_y = 0.1$, $r = 0.5$, and three generations of recombination result in $D_{xy}^{(3)} \leq -0.0175$. Very large populations would be necessary to detect such small changes with sufficient accuracy. We conclude that while the effects of selection could very well contribute to the observed increase in negative LD (Fig. 3), it seems hardly justified to attribute the considerable increase in negative LD exclusively to selection.

Further causes of an increase in negative LD can be sampling effects due to small population sizes. In contrast to selection, which is expected to build up LD in a directional process, sampling effects are expected to result in erratic changes in LD. To investigate the causes of the build-up of LD, we analyzed the five loci pairs showing the largest absolute positive or negative D values in the last selection cycle of both RS programs. However, D values did not show a directional, but rather an erratic change (Fig. 4). Thus, these changes in LD are attributable to sampling effects rather than selection. We therefore conclude that sampling effects due to small and finite number of selected plants ($N = 72$) were presumably an important

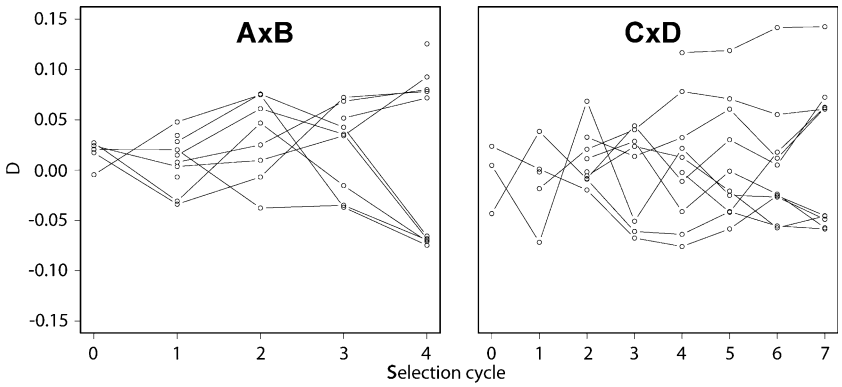

Fig. 4 Development of LD between alleles whose frequencies were increased by selection, measured by D over four cycles of RS in population A × B and seven cycles of RS in population C × D. LD was measured only for the five loci pairs with the highest and lowest LD values in the final selection cycles (A × B: C4 and C × D: C7)

factor contributing to the increase in LD during the selection cycles (Fig. 3).

Association between LD and additive genetic variance

In two companion studies, Flachenecker et al. (2006a, b) observed a decrease in σ_A^2 for selection index in both RS programs, which was only significant in A × B. In theory, the build-up of negative LD due to selection results in a reduction in σ_A^2. For the biallelic case and n loci, σ_A^2 and D_{xy} are related by (Lynch and Walsh 1998)

$$\sigma_A^2 = 2\sum_{x=1}^{n} \alpha_x^2 p_x(1-p_x) + 2\sum_{x=1}^{n}\sum_{y\neq x}^{n} \alpha_x\, \alpha_y\, D_{xy}, \qquad (4)$$

where n is the number of loci and, p_x is the frequency of allele x, D_{xy} the LD between the xth and yth locus, and α_x and α_y are the average effects of allele substitution at the xth and yth locus. However, a long-term reduction in σ_A^2 is only expected if uniformly negative D values were observed at many loci (Eq. 4). This was not the case in our experiment (Fig. 3) and therefore the LD generated by selection was hardly a factor in reducing the selection response by a reduction in σ_A^2.

Summarizing, we attribute the comparatively high selection response per cycle compared with other RS studies not only to the migration effects observed during the selection procedure but also to the applied mating scheme of Cockerham and Burrows (1980). Thus, the applied mating scheme offers an alternative for successful maize breeding by means of RS. Moreover, our experimental results support the hypothesis that the LD generated by selection is not a limiting factor for achieving high selection response in RS programs, in particular if an efficient recombination procedure is employed, which reduces negative LD between favorable alleles. This may be an explanation for of continued selection response in other long-term selection program, e.g., the Illinois long-term selection program (*cf.* Dudley and Lambert 2004). In particular, this hypothesis is supported by (1) the large phenotypic selection response in our study, and (2) the fact that a clear trend towards negative LD between favorable alleles was not observed at the molecular level.

Acknowledgments The present study was supported by a grant from the Deutsche Forschungsgemeinschaft, Grant No. FR 1615/1. The authors thank Dr. D. Klein and the staff of the Plant Breeding Research Station at Eckartsweier for conducting the field trials. We thank KWS SAAT AG and Pioneer Hi-Bred Northern Europe for their support with the SSR screening. We gratefully acknowledge Christine Beuter and Elisabeth Kokai-Kota for the technical assistance with the SSR analyses. The authors appreciate the editorial work of Dr. Jasmina Muminović, whose suggestions considerably improved the style of the manuscript.

References

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

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

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

Dudley JW, Lambert RJ (2004) 100 generations of selection for oil and protein in corn. In: Janick J (ed) Plant breeding reviews, Part 1: Long-term selection: maize. Wiley, New York, pp 79–110

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, 4th edn. Longman Sci and Tech, Harlow

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

Falke KC, Flachenecker C, Melchinger AE, Piepho H-P, Maurer HP, Frisch M (2007) Temporal changes in allele frequencies in two European F_2 flint maize populations under modified recurrent full-sib selection. Theor Appl Genet 114:765–776

Falque M, Decousset L, Dervins D, Jacob AM, Joets J, Martinant JP, Raffoux X, Ribiere N, Ridel C, Samson D, Charcosset A, Murigneux A (2005) Linkage mapping of 1454 new maize candidate gene loci. Genetics 170:1957–1966

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 Breeding 125:248–253

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

Genter CF (1982) Recurrent selection for high inbred yields from the F_2 of maize single cross. Proc Ann Corn Sorghum Ind Res Conf 37:67–76

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

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

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

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

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

Johnson GR (1982) Two-locus theory in recurrent selection for general combining ability in maize. Theor Appl Genet 61:279–283

Kosambi DD (1944) The estimation of map distance from recombination values. Ann Eugen 12:172–175

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

Labate JA, Lamkey KR, Lee M, Woodman W (2000) Hardy–Weinberg and linkage equilibrium estimates in the BSSS and BSCB1 random mated populations. Maydica 45:243–255

Landi P, Frascaroli E (1993) Responses to four cycles of full-sib family recurrent selection in an F2 maize population. Maydica 38:31–37

Lee M, Sharopova N, Beavis WD, Grant D, Katt M, Blair D, Hallauer A (2002) Expanding the genetic map of maize with the intermated B73 × Mo17 (IBM) population. Plant Mol Biol 48:453–461

Liu SC, Kowalski SP, Lan TH, Feldmann KA, Paterson AH (1996) Genome-wide high-resolution mapping by recurrent intermating using *Arabidopsis thaliana* as a model. Genetics 142:247–258

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland

Martin OC, Hospital F (2006) Two- and three-locus tests for linkage analysis using recombinant inbred lines. Genetics 173:451–459

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

Mihaljevic R, Utz HF, Melchinger AE (2004) Congruency of quantitative trait loci detected for agronomic traits in testcrosses of five populations of European maize. Crop Sci 44:114–124

Mihaljevic R, Schön CC, Utz HF, Melchinger AE (2005a) 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

Mihaljevic R, Utz HF, Melchinger AE (2005b) No evidence for epistasis in hybrid and per se performance of elite European flint maize inbreds from generation means and QTL analyses. Crop Sci 45:2605–2613

Moll RH (1991) Sixteen cycles of recurrent full-sib family selection for grain weight in two maize populations. Crop Sci 31:959–964

Nei M (1963) Effect of selection on the components of genetic variance. pp. 501–515. In:Hanson WD and Robinson HF (eds) Statistical genetics and plant breeding. National Academy of Sciences—National Research Council, Washington

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

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

Russell WA, Eberhard UA, Vega Q (1973) Recurrent selection for specific combining ability for yield in two maize populations. Crop Sci 13:257–261

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(2):378–389

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: joinMap. Plant J 3:739–744

Teuscher F, Broman KW (2007) Haplotype probabilities for multiple-strain recombinant inbred lines. Genetics 175:1267–1274

Teuscher F, Guiard V, Rudolph PE, Brockmann GA (2005) The map expansion obtained with recombinant inbred strains and intermated recombinant inbred populations for finite generation designs. Genetics 170:875–879

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

Weir BS (1996) Genetic data analysis II. Sinauer Associates, Sunderland

Winkler CR, Jensen NM, Cooper M, Podlich DW, Smith OS (2003) On the determination of recombination rates in intermated recombinant inbred populations. Genetics 164:741–745

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

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

5. General Discussion

A long-term modified recurrent full-sib selection program with two European F_2 flint maize (*Zea mays* L.) populations was conducted at the University of Hohenheim (Germany) to analyze the selection response in RS programs. The two base populations, A × B and C × D, were further employed in various successful QTL mapping experiments aiming at the identification of markers associated with QTL and their use in subsequent MAS programs (Schön et al. 1994; Lübberstedt et al. 1997a, b, 1998a; Melchinger et al. 1998; Utz et al. 2000; Mihaljevic et al. 2004, 2005). The F_2 base populations were intermated via chain crossing for three generations. To investigate the efficiency of the phenotypic selection procedure, four (A × B) and seven (C × D) selection cycles were analyzed with quantitative genetic methods (Flachenecker et al. 2006a, b, c). However, if the main focus of a study lies on allele frequency changes in populations as in the case of RS programs, rather than on shifts in phenotypic parameters of individuals, then molecular markers are a reliable tool for analyzing the selection procedure and selection response (*cf.* Labate et al. 1997, 1999, 2000; Landi et al. 2002; Pinto et al. 2003a, b; Huang et al. 2004; Butron et al. 2005; Hinze et al. 2005; Coque and Gallais 2006; Falke et al. 2007a, b). Knowledge of allele frequency changes at marker loci in populations under selection provides insights in the genetic basis of agronomically important traits. Moreover, the application of molecular markers offers the opportunity to identify regions in the genome that were influenced by selection or migration. Therefore, the objective of this thesis was to analyze the selection procedure and the resulting selection response at the molecular level using SSR markers.

Use of SSRs for Analyzing RS Programs

Selection for simple and complex traits to improve population performance has been performed with great success at the phenotypic level. However, the forthcoming age of biotechnology and genomics offers the prospect of shifting selection gradually from phenotypes to genotypes (Walsh 2001). This development provides plant breeders with

new perspectives to determine how genetic and environmental factors contribute to the observed variation of important traits within or between populations.

The application of PCR-based molecular marker technologies has proven useful in a wide range of fundamental and applied breeding studies. To come to a decision which technique is appropriate to investigate a specific question, comparisons among various marker systems are inevitable. Different markers types were compared for maize (Pejic et al. 1998; Heckenberger et al. 2003, 2006; Stich et al. 2006), soybean (Powell et al. 1996), and barley (Russell et al. 1997). SSRs proved to be useful as genetic markers in numerous plant species including maize (Senior and Heun 1993), wheat (Röder et al. 1995), rice (Wu and Tanksley 1993), soybean (Akkaya et al. 1992), sugar beet (Mörchen et al. 1996), rapeseed (Kresovich et al. 1995), and barley (Becker and Heun 1995).

In the presented selection experiment, SSRs were employed because of the following desirable properties: (1) SSRs show a codominant inheritance and permit a precise differentiation between homozygous and heterozygous genotypes. Therefore, they are genetically highly informative. Although codominant scoring of dominant marker systems (*e.g.*, AFLPs) can be achieved with quantitative assessment of the optical density of bands (Castiglioni et al. 1999; Piepho and Koch 2000; Jansen et al. 2001; Wong et al. 2001; Geerlings et al. 2003), this method is relatively time-consuming and possesses a sizeable error rate (Weeks et al. 2000; Yan et al. 2005). (2) SSRs reveal the highest level of polymorphism (*cf.* Russell et al. 1997) as a consequence of their high mutation rate (Lynch and Walsh 1998, p. 393). (3) A multitude of SSRs for maize is available at the MaizeGDB with an adequate coverage of the genome (http://www.maizegdb.org). Moreover, the application of SSRs across species is also partly possible (*cf.* Lübberstedt et al. 1998b). (4) SSR analysis can be conducted with simple laboratory equipment, and (5) SSR analyses are inexpensive and fast. In contrast, RFLPs, AFLPs, RAPDs, or SNPs are more expensive or laborious for high-throughput monitoring of large numbers of genotypes.

To successfully reach the goals of the study, about 100 polymorphic SSRs were required, which were polymorphic between the parental lines in either population and provided a uniform distribution over the maize genome (Falke et al. 2006, 2007a, b). Therefore, the high level of polymorphism of SSRs and their availability in the MaizeGDB were helpful factors for the marker screening between the parental lines. An unequivocal distinction of homozygous and heterozygous genotypes was required, particularly for analyzing allele frequency changes during the RS procedure (Falke et al. 2007a). Another advantage of SSRs is that non-parental bands can easily be detected, whereas

contaminations with foreign pollen might have remained undetected with other marker systems. For instance, AFLP fragments are scored by their length, while their sequence is unknown. However, the assumption that AFLP fragments of the same length are identical is an oversimplification of the actual situation, because a fraction of co-migrating fragments may have completely distinct sequences (Rouppe van der Voort et al. 1997; Vekemans et al. 2002; Koopman and Gort 2004). The possibility to detect non-parental alleles was of immense importance for the evaluation of our RS programs. Moreover, the high-throughput production of about 140 000 SSR data points could be obtained at relatively low costs. Consequently, SSRs were an appropriate marker system for evaluating the selection response in our RS programs, which is in agreement with other RS studies (Pinto et al. 2003a, b; Huang et al. 2004; Hinze et al. 2005; Coque and Gallais 2006; Falke et al. 2007a, b).

Advantage of Intermated Mapping Populations

Intermating increases the frequency of recombinant gametes in a mapping population, because the chromosomes undergo multiple meioses. In several linkage mapping studies, this was interpreted as map expansion (Liu et al. 1996; Lee et al. 2002; Winkler et al. 2003; Falque et al. 2005; Teuscher et al. 2005). However, these authors constructed linkage maps based on recombination frequencies *R*, referring to the recombination events accumulated in all intermating generations, instead of the recombination frequencies *r*, referring to a single meiosis. This is misleading, because since the introduction of map distances these were estimated with recombination frequencies *r*, referring to only a single meiosis (Sturtevant 1913; Haldane 1919; Morgan 1928; Kosambi 1944; Stam 1993). Therefore, an interpretation more adequate than map expansion is that an intermated population of size *N* provides a greater mapping accuracy and, therefore, a higher mapping resolution for tightly linked loci than *N* F_2 individuals, which underwent only a single meiosis.

We constructed linkage maps of our F_2Syn3 populations with mapping methods developed especially for intermating populations (extension of software JoinMap Version 3.0 (Van Ooijen and Voorrips 2001)) so that recombination frequencies relate to a single meiosis (Falke et al. 2006). Thus, our maps did not exhibit a map expansion compared with maps that were constructed with the respective base populations. For comparing the

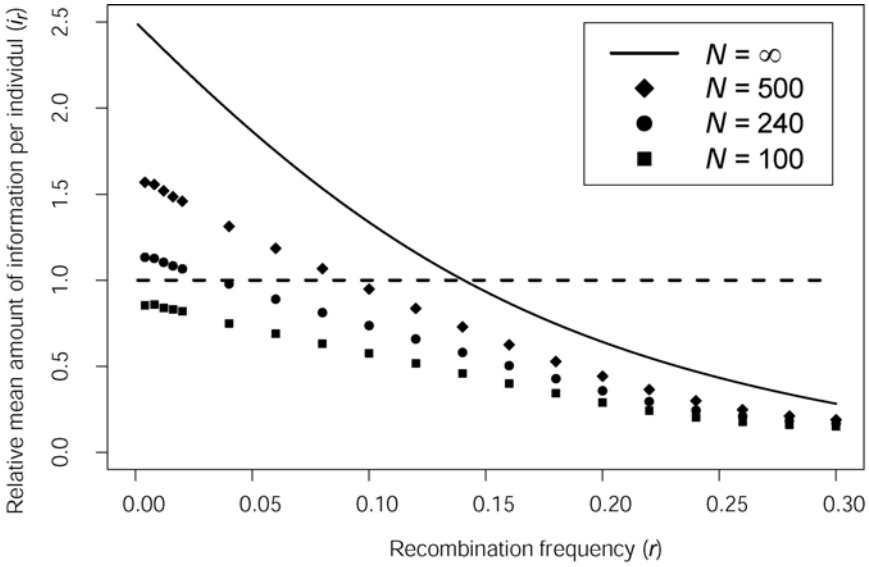

Figure 2. Relative average information per individual i_r in F_2Syn3 populations for finite ($N = 100$, 240, 500) and infinite size of the intermated generations. The dashed line symbolizes the value $i_r = 1$ for F_2 individuals.

mapping resolution of the intermated with the respective base populations, we measured the precision of estimates of r with the average information per individual i_r relative to an F_2 individual (Mather 1936; Allard 1956; Liu et al. 1996). A higher mapping resolution in F_2Syn3 compared with the base populations was only expected for $r \leq 0.142$ (A × B) and $r \leq 0.197$ (C × D). Despite of relatively small average recombination frequencies between adjacent markers in both intermated populations ($r = 0.144$), 49% (A × B) and 19% (C × D) of the marker intervals were larger than the threshold.

After our study on linkage mapping with intermated populations was published, Martin and Hospital (2006) and Teuscher and Broman (2006) also presented results on the subject. Teuscher and Broman (2006) derived two-point haplotype probabilities for intermated recombinant inbred line populations and Martin and Hospital (2006) suggested a bias correction for the estimation procedure of Liu et al. (1996). Both studies overlooked an aspect we already noted (Falke et al. 2006), however, without working out detailed results: The entire theory of linkage mapping in intermated population derived until now is only valid for an infinite population size of the intermating population, an assumption most probably violated in all practical studies.

To assess the effect of random genetic drift in finite intermating populations on the standard error of r and, hence, on the mean amount of information, I conducted a

simulation study with Plabsoft (Maurer et al. 2007). Three generations of intermating were simulated for an infinite population size as well as for finite intermating generations of size $N = 500$, 240, and 100 individuals. Recombination frequencies r and the relative average information per individual i_r were estimated according to Liu et al. (1996). The simulation was repeated for 1000 simulation runs and the results were averaged over the runs. For intermating populations of infinite size, the simulation results corresponded exactly with the theoretical results obtained with the formulas of Liu et al. (1996): If $r < 0.142$ then $i_r > 1$ (Figure 2). However, for finite N, the picture was entirely different. With $N = 100$, i_r was smaller than 1 even for very tightly linked loci with $r = 0.01$. With $N = 240$, as employed in our experiments $i_r > 1$ only when $r < 0.04$. Consequently, our results indicate that the theoretical advantage of intermating populations decreases rapidly with decreasing population sizes of the intermating generations.

In conclusion, random genetic drift highly affects i_r. Thus, the application of intermating populations for attempting higher mapping accuracy can be suggested only for very small marker intervals combined with large population sizes. Otherwise, the advantages of intermating populations compared with their F_2 base populations are canceled out due to random genetic drift.

Benefit of Intermating prior to the RS Process

In RS programs, initial parental LD in the base populations can negatively influence the short- and long-term selection response. Intermating before starting the selection procedure is expected to be more efficient for reducing parental LD than intermating generations between subsequent selection cycles (Johnson 1982). Three generations of intermating prior to the selection procedure are proposed as sufficient to almost cancel out the negative effects of initial parental LD on the selection response (Johnson 1982).

We followed the suggestion of Johnson (1982) and intermated the F_2 base populations three times by employing a chain crossing procedure. The reduction of LD agreed well with the theoretical values from a simulation study employing our intermating procedure (Falke et al. 2007b). A reduction from about 50% of linked loci in significant parental LD in the base populations to 35.7% (A × B) and 30.9% (C × D) in the intermated populations was observed. Intermating within F_2 populations results in a reduction of the estimates of the degree of dominance $\bar{d}$ to about 1 or less (Moll et al. 1964). Therefore,

the comparatively low estimates of the dominance variance σ_D^2 (Flachenecker et al. 2006a, b) support the hypothesis that the reduction of parental LD between linked loci was sufficient for not hampering selection response.

Breeding procedures have to be planned with long-sightedness, because the fast development of new varieties is essential for the economical success. The conduction of three intermating generations takes at least 1.5 years by using a winter nursery, thus, making this breeding step expensive and time-consuming. Moreover, random genetic drift effects may already occur during the intermating and hamper the reduction in parental LD (Falke et al. 2007b). This time could be also used for 1.5 selection cycles (with a winter nursery). However, selection cycles and the subsequent field trials are even more expensive. Our study did not allow to investigate whether the intermating prior to the RS program provides higher selection response than without intermating generations due to the lack of comparability. Consequently, further research is needed to investigate the economical profit of intermating generation prior to RS for their application in applied breeding programs.

Occurrence of Non-Parental Alleles

Non-parental alleles are marker bands present in a progeny but not in either of its parents. Several possible reasons for the occurrence of non-parental bands are reported (Smith et al. 1997): (1) Contaminations with foreign pollen or erroneous crossing during the recombination process. (2) Residual heterozygosity within the parental inbred line causes new bands in the progenies, which are then assumed to be non-parental bands. (3) Genetic changes in the parental inbred lines, *e.g.*, by mutation or physical mixing of seed from other genotypes. Moreover, technical problems during the SSR assays can cause the detection of non-parental bands: (1) Artifactual "stutter" bands, which are especially prone to occur from di-repeat SSRs (Smith et al. 1997). (2) Instability of SSRs due to unequal recombination or DNA polymerase slippage (Wierdl et al. 1996). (3) Misscoring of bands.

Contaminations with foreign pollen or erroneous crossing are mostly not distinguishable in the field by means of morphological traits (*e.g.*, color of kernels, tassels, or anthers, etc.) and therefore remain undetected during the selection process. Codominant markers are a useful tool to identify these migration effects. In our study, non-parental alleles were observed during the selection process with average proportions up to 10% in

both populations as a result of contaminations with foreign pollen or erroneous crossing (Falke et al. 2007a). However, non-parental alleles were also detected in $F_{3:4}$ of C × D and in F_2Syn3 of both populations, with average proportions of 0.01. These alleles were presumably either parental alleles due to residual heterozygosity within the parental inbred lines or laboratory errors.

We suppose that migration affected considerably the high observed selection response per cycle in both RS programs (*cf*. Flachenecker et al. 2006a, b, c). Therefore, we suggest the employment of marker assays in parallel to the selection procedure to easily identify contaminations with foreign pollen and remove contaminated genotypes on time.

Changes in Allele Frequencies: Selection or Random Genetic Drift?

The performance of a breeding population increases as a result of an increase in the frequencies of favorable alleles. Many RS methods exist and are effective at changing the population performance (Hallauer and Miranda 1988). However, in small populations random genetic drift associated with the chance of loss or fixation of favorable alleles, is a main drawback for long-term selection programs. Fortunately, the influence of random genetic drift during the selection progress can be controlled by balancing the effects of inbreeding and the selection intensity (Robertson 1960).

For examining allele frequency changes in maize and other crops undergoing selection, isozymes and molecular markers have been employed in several studies (Brown 1971; Brown and Allard 1971; Delaney and Bliss 1991; Eagen and Goldman 1996; Labate et al. 1999; De Koeyer et al. 2001). For testing whether allele frequency changes are attributable to random genetic drift acting alone or follow a linear trend due to selection, Schaffer et al. (1977) suggest a test based on a χ^2 statistic, which has been applied in many RS studies in maize (Stuber et al. 1980; Heredia-Diaz et al. 1996; Landi et al. 2002; Butron et al. 2005) and also in red beet (Eagen and Goldman 1996).

While Schaffer's (1977) test was derived for a specific sampling scheme (sampling is performed before reproduction), Waples (1989) proposed a more general χ^2 test, which can also be employed for sampling after reproduction. Therefore, Waples' (1989) test takes into account the increased variance in allele frequencies between generations due to (1) random genetic drift and (2) stochastic dependence. The null hypothesis of Waples' (1989) test that random genetic drift acts alone was rejected for 23% (C0 vs. C1) and 20% (C0 vs. C4) of

all loci in population A × B and for 6% (C0 vs. C1) and 13% (C0 vs. C7) in population C × D. This result is in agreement with other studies applying Waples' (1989) test. Labate et al. (1999) described significant changes in allele frequencies at about 17% of all RFLP loci after 12 cycles of reciprocal RS in US maize. After one cycle of high intensity reciprocal RS, significant allele frequency changes due to selection were reported at 13% and 7% of all loci (Pinto et al. 2003b). In these two studies, the loci with significant allele frequency changes due to selection were mainly dispersed over the whole genome, whereas Coque and Gallais (2006) found marker loci whose frequencies were significantly increased by selection in several QTL regions. In our study, Waples' (1989) test detected many significant changes for non-parental allele frequencies. Hence, we suppose that the selection response was influenced by migration (Falke et al. 2007a). Nevertheless, our results showed that several QTL regions for selection index were subjected to selection pressure (Falke et al. 2007a). In conclusion, the correlation between allele frequency changes at markers linked to QTL and phenotypic performance (Coque and Gallais 2006; Falke et al. 2007a) indicate that these marker loci could be used in MAS breeding programs.

Additive Genetic Variance and the Effects of LD

The success of RS programs is determined by an increase in the mean of the target populations but also by the maintenance of the genetic variability within the populations to facilitate improvement in future cycles of selection. For estimating trends in the amount of remaining variability over selection cycles, plant breeders employ usually moment estimators of variance components. More recently, Bernardo (1994) proposed the application of restricted maximum likelihood (REML) estimation of variance components from complex pedigrees.

In our study, REML estimation of variance components was used for investigating trends in the amount of remaining additive genetic variance σ_A^2. Even though we applied a low selection intensity and relatively high effective population size N_e (32), a decrease in σ_A^2 for grain yield and selection index was observed in both populations, which was, however, only significant for selection index in A × B (Flachenecker et al. 2006a, b). However, all variance components were associated with large standard errors. Estimates of σ_A^2 for grain yield and grain moisture were in A × B smaller and in C × D similar to that

reported for US dent maize F_2 populations (Hallauer and Miranda 1988; Wolf et al. 2000). The estimates of σ_A^2 agreed well with estimates for other European F_2 flint maize populations (Mihaljevic et al. 2004).

In theory, the reduction in σ_A^2 due to selection is caused by a build-up of negative LD (Bulmer 1971). In our study, the development of LD showed an erratic change over the selection cycles with slight increases in both positive and negative LD (Falke et al. 2007b). We attributed the increase in positive LD mainly to the applied mating scheme and migration effects (Falke et al. 2007a). Due to the erratic development of LD, we conclude that random genetic drift due to finite population size rather than selection contributed to the build-up of negative LD. Our results are in agreement with the study of Nei (1963), showing that the build-up of negative LD between any pair of loci is small when the effects of allele frequency changes are small. Moreover, the LD newly generated by selection will be reduced with each subsequent meiosis. Labate et al. (2000) reported that significant increases in LD over 12 cycles of reciprocal RS in US maize was observed mainly between alleles near fixation, whereas no fixation of markers alleles was observed in our study (Falke et al. 2007a). From theory, the reduction in σ_A^2 due to a build-up of negative LD is expected only if negative LD is observed at a considerable number of loci (Lynch and Walsh 1998, p. 102). However, only slight increases in negative LD were observed at a small number of loci in our study. Consequently, our results indicate that newly generated LD through selection is not a limiting factor in obtaining high selection response in our RS programs.

Response to Recurrent Full-Sib Selection

Development of superior germplasm by means of selection procedures, which maximize long-term selection response is a major aim of plant breeders. The selection procedure of our modified recurrent full-sib selection program was based on a selection index (2 × grain moisture + grain yield, relative to the F_2 check entries). Despite (1) employing elite inbred lines as parents, (2) a low selection intensity, and (3) selecting two traits simultaneously, a high selection response per cycle was obtained in population C × D (grain yield: 9.1%; selection index: 11.2%; grain moisture: -1.1%) (Flachenecker et al. 2006a). Other studies using F_2 recurrent full-sib selection reported selection response for grain yield of 4.4% per cycle over 16 cycles (Moll 1991) and 7.3% per cycle over four

cycles (Landi and Frascaroli 1993). In population A × B, the selection response relative to the check entries was 120% in selection cycle C1 but persisted at this level. The selection response was presumably biased by the variable performance of the check entries over the years (Flachenecker et al. 2006b).

Therefore, a re-evaluation of the selection response was conducted in both populations using a population diallel and an extension (Melchinger and Flachenecker 2006) of the Smith (1979, 1983) model. The selection response per cycle was estimated after correcting for effects of frequency changes in heterozygotes due to selection and random genetic drift. Significant estimates were determined for grain yield (A × B: 14.07%; C × D: 8.28%) and selection index (A × B: 5.25%; C × D: 3.64%) (Flachenecker et al. 2006c), and were larger than in other RS studies applying a diallel approach (Landi and Frascaroli 1993; Stojsin and Kannenberg 1994).

After the phenotypic evaluation, we attributed the comparatively high selection response per cycle in our experiment compared with other F_2 based recurrent full-sib selection programs to several factors: (1) the applied mating scheme of Cockerham and Burrows (1980), (2) minimized random genetic drift effects due to a relatively large effective population size N_e, (3) the realization of only four (A × B) and seven (C × D) selection cycles, and (4) a moderate genotype × environment interaction (Flachenecker et al. 2006a, b, c). However, the analyses at the molecular level revealed further causes: (1) the detected migration effects presumably contributed considerably to the observed selection response in both RS programs, (2) significant allele frequency changes were observed at marker loci in QTL regions for traits under selection, (3) LD was not a limiting factor for the selection response, and (4) further selection response can be expected in future selection cycles because neither fixation nor extinction of alleles at any marker loci was observed (Falke et al. 2007a, b).

Conclusions

Our research demonstrates that the evaluation of RS programs using SSRs provides insights into the population structure of the breeding program and in the effects influencing long-term selection response.

Three generations of intermating prior to the selection procedure resulted in a considerable reduction in initial parental LD between linked loci. However, an interesting

open question for further research is whether intermating prior to the RS procedure provides a significantly higher selection response than without intermating generations.

The application of intermated mapping populations to obtain high mapping resolution can be recommended only for very tightly linked loci because random genetic drift effects occurring during the intermating process increase rapidly with decreasing population size.

QTL for the selection index were detected in the base populations. Flanking markers at some of them showed significant allele frequency changes attributable to selection. These markers can be included in the selection index and subsequently employed in MAS.

Although selection is expected to generate negative LD between favorable alleles, which causes a reduction in σ_A^2 and, thus, a decline in the selection response, our results showed that the build-up of negative LD was low and not a limiting factor.

The migration detected at the molecular level and its effects on the selection response demonstrated the importance of marker analyses in parallel to the selection procedure.

Consequently, the present study demonstrates that knowledge about the effects of selection, random genetic drift, migration, and linkage at the molecular level is of great importance for optimizing breeding programs.

References

Akkaya MS, Bhagwat AA, Cregan PB (1992) Length polymorphisms of simple sequence repeat DNA in soybean. Genetics 132:1131-1139

Becker J, Heun M (1995) Barley microsatellites: allele variation and mapping. Plant Mol Biol 27:835-845

Bernardo R (1994) Prediction of maize single-cross performance using RFLPs and information from related hybrids. Crop Sci 34 :20-25

Brown AHD (1971) Isozyme variation under selection in *Zea mays*. Nature 232:570–571

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

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

Butron A, Tarrio R, Revilla P, Ordas A, Malvar RA (2005) Molecular changes in the maize composite EPS12 during selection for resistance to pink stem borer. Theor Appl

Genet 110:1044-1051

Castiglioni P, Ajmone-Marsan P, Van Wijk R, Motto M (1999) AFLP markers in a molecular linkage map of maize: Codominant scoring and linkage group distribution. Theor Appl Genet 99:425-431

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

De Koeyer DL, Phillips RL, Stuthman DD (2001) Allelic shifts and quantitative trait loci in a recurrent selection population of oat. Crop Sci 41:1228-1234

Delaney DE, Bliss FA (1991) Selection for increased percentage phaseolin in common bean. Theor Appl Genet 81:306-311

Eagen KA, Goldman IL (1996) Assessment of RAPD marker frequencies over cycles of recurrent selection for pigment concentration and percent solids in red beet (*Beta vulgaris* L.). Mol Breed 2:107-115

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

Falke KC, Flachenecker C, Melchinger AE, Piepho H-P, Maurer HP, Frisch M (2007a) Temporal changes in allele frequencies in two European F_2 flint maize populations under modified recurrent full-sib selection. Theor Appl Genet 114:765-776

Falke KC, Maurer HP, Melchinger AE, Piepho H-P, Flachenecker C, Frisch M (2007b) Linkage disequilibrium in two European F_2 flint maize populations under modified recurrent full-sib selection. Theor Appl Genet 115: 289-297

Falque M, Decousset L, Dervins D, Jacob AM, Joets J, Martinant JP, Raffoux X, Ribiere N, Ridel C, Samson D, Charcosset A, Murigneux A (2005) Linkage mapping of 1454 new maize candidate gene loci. Genetics 170:1957-1966

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 recurrent full-sib selection programs in two F_2 populations of European flint maize. Theor Appl Genet 13:1113-1120

Geerlings H, van Oeveren J, Pot J, Jansen R, Van Schaik R (2003) AFLP Quantar Pro: codominant and expression analysis software. http://www.keygeneproducts.com

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

Hallauer AR, Miranda JBF (1988) Quantitative Genetics in Maize Breeding. 2nd edn Iowa State University Press, Ames, USA

Heckenberger M, Rouppe Van Der Voort J, Melchinger AE, Peleman J, Bohn M (2003) Variation of DNA fingerprints among accessions within maize inbred lines and implications for identification of essentially derived varieties: II. Genetic and technical sources of variation in AFLP data and comparison with SSR data. Mol Breed 12:97-106

Heckenberger M, Muminovic J, Rouppe van der Voort J, Peleman J, Bohn M, Melchinger AE (2006) Identification of essentially derived varieties obtained from biparental crosses of homozygous lines. III. AFLP data from maize inbreds and comparison with SSR data. Mol Breed 17:111-125

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

Hinze LL, Kresovich S, Nason JD, Lamkey KR (2005) Population genetic diversity in a maize reciprocal recurrent selection program. Crop Sci 45:2435-2442

Huang SH, Teng WT, Wang YJ, Dai JR (2004) Genetic diversity analysis of maize recurrent selection populations by SSR marker. Acta Genetica Sinica 31:73-80

Jansen RC, Geerlings H, Van Oeveren AJ, Van Schaik RC (2001) A comment on codominant scoring of AFLP markers. Genetics 158:925-926

Johnson GR (1982) Two-locus theory in recurrent selection for general combining ability in maize. Theor Appl Genet 61:279-283

Koopman WJM, Gort G (2004) Significance tests and weighted values for AFLP similarities, based on Arabidopsis *in silico* AFLP fragment length distributions. Genetics 167:1915-1928

Kosambi DD (1944) The estimation of map distance from recombination values. Ann Eugen 12:172-175

Kresovich S, Szewc-McFadden AK, Bliek SM, McFerson JR (1995) Abundance and characterization of simple-sequence repeats (SSRs) isolated from a size-fractionated genomic library of *Brassica napus* L. (rapeseed). Theor Appl Genet 91:206-211

Labate JA, Lamkey KR, Lee M, Woodman WL (1997) Molecular genetic diversity after reciprocal recurrent selection in BSSS and BSCB1 maize populations. Crop Sci 37:416-423

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

Labate JA, Lamkey KR, Lee M, Woodman W (2000) Hardy-Weinberg and linkage equilibrium estimates in the BSSS and BSCB1 random mated populations. Maydica 45:243-255

Landi P, Frascaroli E (1993) Responses to four cycles of full-sib family recurrent selection in an F2 maize population. Maydica 38:31-37

Landi P, Chiappetta L, Salvi S, Frascaroli E, Lucchese C, Tuberosa R (2002) Responses and allelic frequency changes associated with recurrent selection for plant regeneration from callus cultures in maize. Maydica 47:21-32

Lee M, Sharopova N, Beavis WD, Grant D, Katt M, Blair D, Hallauer A (2002) Expanding the genetic map of maize with the intermated B73 x Mo17 (IBM) population. Plant Mol Biol 48:453-461

Liu SC, Kowalski SP, Lan TH, Feldmann KA, Paterson AH (1996) Genome-wide high-resolution mapping by recurrent intermating using Arabidopsis thaliana as a model. Genetics 142:247-258

Lynch M, Walsh B (1998) Genetics and Analysis of Quantitative Traits. Sinauer Assoc, Sunderland

Lübberstedt T, Melchinger AE, Schön CC, Utz HF, Klein D (1997a) QTL mapping in testcrosses of European flint lines of maize: I. Comparison of different testers for forage yield traits. Crop Sci 37:921-931

Lübberstedt T, Melchinger AE, Klein D, Degenhardt H, Paul C (1997b) QTL mapping in testcrosses of European flint lines of maize: II. Comparison of different testers for forage quality traits. Crop Sci 37:1913-1922

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

Lübberstedt T, Dussle C, Melchinger AE (1998b) Application of microsatellites from maize to teosinte and other relatives of maize. Plant Breed 117:447-450

Martin OC, Hospital F (2006) Two- and three-locus tests for linkage analysis using recombinant inbred lines. Genetics 173:451-459

Maurer HP, Melchinger AE, Frisch M (2007) Population genetic simulation and data analysis with Plabsoft. Euphytica (in review)

Melchinger AE, Flachenecker C (2006) An extension of the Smith model for quantitative genetic analysis of selection response under recurrent selection. Plant Breed 125:644-646

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, Utz HF, Melchinger AE (2004) Congruency of quantitative trait loci detected for agronomic traits in testcrosses of five populations of European maize. Crop Sci 44:114-124

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

Moll RH (1991) Sixteen cycles of recurrent full-sib family selection for grain weight in two maize populations. Crop Sci 31:959-964

Moll RH, Lindsey MF, Robinson HF (1964) Estimates of genetic variances and level of dominance in maize. Genetics 49:411-423

Morgan TH (1928) The Theory of Genes. Yale University Press, New Haven, Conn

Mörchen M, Cuguen J, Michaelis G, Hänni C, Saumitou-Laprade P (1996) Abundance and length polymorphism of microsatellite repeats in Beta vulgaris L. Theor Appl Genet 92:326-333

Nei, M (1963) Effect of selection on the components of genetic variance. pp. 501-515. *In* Hanson WD, Robinson HF (eds). Statistical Genetics and Plant Breeding. National Academy of Sciences - National Research Council, Washington

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

Piepho HP, Koch G (2000) Codominant analysis of banding data from a dominant marker system by normal mixtures. Genetics 155:1459-1468

Pinto LR, Carneiro Vieira ML, de Souza Jr CL, de Souza AP (2003a) Genetic-diversity assessed by microsatellites in tropical maize populations submitted to a high-intensity reciprocal recurrent selection. Euphytica 134:277-286

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

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol Breed 2:225-238

Robertson A (1960) A theory of limits in artificial selection. Proc. R. Soc. Lond. B 153:234-249

Röder MS, Plaschke J, Konig SU, Borner A, Sorrells M E, Tanksley SD, Ganal MW (1995) Abundance, variability and chromosomal location of microsatellites in wheat. Mol Gen Genet 246:327-333

Rouppe van der Voort J, Van Zandvoort P, Van Eck HJ, Folkertsma RT, Hutten RCB, Draaistra J, Gommers FJ, Jacobsen E, Helder J, Bakker J (1997) Use of allele specificity of comigrating AFLP markers to align genetic maps from different potato genotypes. Mol Gen Genet 255:438-447

Russell JR, Fuller JD, Macaulay M, Hatz BG, Jahoor A, Powell W, Waugh R (1997) Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. Theor Appl Genet 95:714-722

Schaffer HE, Yardley D, Anderson WW (1977) Drift or selection: A statistical test of gene frequency variation over generations. Genetics 87:371-379

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

Senior ML, Heun M (1993) Mapping maize microsatellites and polymerase chain reaction confirmation of the targeted repeats using a CT primer. Genome 36:884-889

Smith JSC, Chin ECL, Shu H, Smith OS, Wall SJ, Senior ML, Mitchell SE, Kresovich S, Ziegle J (1997) An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.): Comparisons with data from RFLPs and pedigree. Theor Appl Genet 95:163-173

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

Smith OS (1983) Evaluation of recurrent selection in BSSS, BSCB1, and BS13 maize populations. Crop Sci 23:35-40

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. Plant J 3:739-744

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 Breed 17:217-226

Stojsin D, Kannenberg LW (1994) Genetic changes associated with different methods of recurrent selection in five maize populations: I Directly selected traits. Crop Sci 34:1466-1472

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

Sturtevant AH (1913) The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association. J Experim Zool 14:43-59

Teuscher F, Broman KW (2006) Haplotype probabilities for multiple-strain recombinant inbred lines. Genetics :genetics.106.064063

Teuscher F, Guiard V, Rudolph PE, Brockmann GA (2005) The map expansion obtained with recombinant inbred strains and intermated recombinant inbred populations for finite generation designs. Genetics 170:875-879

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, the Netherlands

Vekemans X, Beauwens T, Lemaire M, Roldan-Ruiz I (2002) Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. Mol Ecol 11:139-151

Walsh B (2001) Quantitative genetics in the age of genomics. Theor Pop Biol 59:175-184

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

Weeks AR, Van Opijnen T, Breeuwer JAJ (2000) AFLP fingerprinting for assessing intraspecific variation and genome mapping in mites. Experim Appl Acarol 24:775-793

Wierdl M, Greene CN, Datta A, Jinks-Robertson S, Petes TD (1996) Destabilization of simple repetitive DNA sequences by transcription in yeast. Genetics 143:713-721

Winkler CR, Jensen NM, Cooper M, Podlich DW, Smith OS (2003) On the determination of recombination rates in intermated recombinant inbred populations. Genetics 164:741-745

Wolf DP, Peternelli LA, Hallauer AR (2000) Estimates of genetic variance in an F_2 maize population. J Hered 91:384-391

Wong A, Forbes MR, Smith ML (2001) Characterization of AFLP markers in damselflies: Prevalence of codominant markers and implications for population genetic applications. Genome 44:677-684

Wu KS, Tanksley SD (1993) Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol Gen Genet 241:225-235

Yan Z, Denneboom C, Hattendorf A, Dolstra O, Debener T, Stam P, Visser PB (2005) Construction of an integrated map of rose with AFLP, SSR, PK, RGA, RFLP, SCAR and morphological markers. Theor Appl Genet 110:766-777

6. Summary

The continuous improvement of breeding material is a main goal in plant breeding. However, many breeding methods result in a reduction of the genetic variation of the breeding material. The primary objective of recurrent selection (RS) methods is to assure a long-term selection response for the target traits by increasing the frequency of favorable alleles, while maintaining the genetic variability of the germplasm for continued selection.

For optimum design of plant breeding programs and evaluation of the selection response, knowledge of the effects of selection, random genetic drift and linkage disequilibrium is mandatory. The main goal of this thesis research was the molecular evaluation of two European F_2 flint maize populations under modified recurrent full-sib selection. In detail, the objectives were to (1) investigate linkage mapping with intermating populations, (2) verify the decay of parental linkage disequilibrium (LD) present in the F_2 base population through three generations of intermating, (3) identify quantitative trait loci (QTL) of traits under selection, (4) separate the effects of random genetic drift and selection on allele frequency changes during the selection procedure in QTL regions, (5) determine the extent of LD build-up by selection, and (6) analyze the effects of LD on the selection response.

Four early maturing homozygous European flint inbred lines (KW1265, D146, D145 and KW1292 subsequently referred to as A, B, C, and D) were crossed to developed 380 $F_{2:3}$ lines of population A × B and 140 $F_{3:4}$ lines of population C × D. The F_2 generations of both populations were intermated for three generations by chain crossing to produce F_2Syn3 populations. Starting from the F_2Syn3 populations, four (A × B) and seven (C × D) cycles of modified recurrent full-sib selection were conducted using a pseudo-factorial mating design. The selection procedure was based on a selection index (2 × grain moisture + grain yield). The top third of the selected families was used twice as crossing parent for the subsequent generations, the remaining two thirds only once. For the marker assays, a total of 104 (A × B) and 101 (C × D) simple sequence repeats (SSRs) were employed to genotype random subsets of 146 $F_{2:3}$ lines in A × B and 110 $F_{3:4}$ lines C × D, 148 F_2Syn3 plants and the parents of the 36 selected full-sib families in each cycle of both

RS programs. The SSR markers were polymorphic between the parental lines und provided a uniform coverage of the entire maize genome.

Genetic linkage maps of the SSR markers were constructed for population $F_{2:3}$ (A × B) and $F_{3:4}$ (C × D) and of both intermated populations F_2Syn3. In contrast to earlier studies, mapping methods developed specifically for intermated populations were applied and, thus, the estimated recombination frequencies referred to a single meiosis. Consequently, the map expansion reported in earlier studies was neither expected nor observed. To verify the expected higher mapping resolution of intermated compared with the respective base population, the precision of estimates of the recombination frequencies r was quantified with the average information per individual i_r relative to an F_2 individual. For infinite population sizes of intermated populations $i_r > 1$ when $r < 0.142$, however, with a population size of N = 240, as used in our mapping study, $i_r > 1$ only when $r < 0.04$. Therefore, we conclude that random genetic drift has a sizeable effect on i_r and, thus, can overrule the advantages of intermating mapping populations.

Three generations of intermating before initiating the RS procedure were primarily conducted to reduce the initial parental LD between linked loci in the F_2 populations and its negative influence on the selection response. Our results demonstrated that the observed decay of LD agreed well with the theoretical expectations achieved from a simulation study. Fisher's exact test indicated that the proportion of linked loci in significant LD decreased from about 50% in the base population to 35.7% (A × B) and 30.9% (C × D) in the intermated populations. However, further studies are required to investigate whether intermating prior to RS provides a higher selection response than breeding programs starting directly with selection cycles.

In our modified recurrent full-sib selection program, a comparatively high selection response of 5.25% (A × B) and 3.64% (C × D) for selection index and of 14.07% (A × B) and 8.28% (C × D) for grain yield was reached. An evaluation at the molecular level revealed that further selection response can be expected, because neither fixation nor extinction of alleles was observed at the investigated marker loci. However, using SSR markers we also detected migration effects since the first selection cycles, which increased up to average proportions of ~10% in both populations. These migration effects were not detectable in the field and are, presumably, attributable to contaminations with foreign pollen. We concluded that the contamination contributed considerably to the observed selection response in both RS programs.

Selection and random genetic drift are main forces affecting changes in allele frequencies and, therefore, the selection response of RS programs. In our study, we analyzed changes in allele frequencies employing a test that allows the separation of selection and random genetic drift. Significant allele frequency changes due to selection were observed for 23% (C0 vs. C1) and 20% (C0 vs. C4) of all loci in population A × B and for 6% (C0 vs. C1) and 13% (C0 vs. C7) in population C × D. In the base populations $F_{2:3}$ (A × B) and $F_{3:4}$ (C × D), several QTL for selection index and its components were detected. At some of them, loci displayed significant allele frequency changes due to selection. We concluded that these SSR markers could be further employed in marker-assisted breeding programs.

Selection is expected to increase the frequency of favorable alleles and simultaneously build up a negative LD between them, which causes a reduction in σ_A^2 and, hence, a decline in the selection response. However, the development of LD in our study displayed an erratic change over the selection cycles with only slight increases in positive and negative LD. The reduction in σ_A^2 due to a build-up of negative LD is expected only if negative LD is observed at many marker loci. Therefore, we concluded that LD due to selection did not limit the selection response in our RS programs.

In conclusion, the evaluation of our modified recurrent full-sib selection programs at the molecular level (1) identified chromosome regions under selection, (2) indicated the need to employ molecular markers in parallel to the selection procedure and, (3) indicated that further selection response can be expected in future cycles.

7. Zusammenfassung

Die kontinuierliche Verbesserung des Zuchtmaterials ist ein Hauptziel der Pflanzenzüchtung. Viele Zuchtverfahren führen jedoch zu einer Einengung der genetischen Varianz im Zuchtmaterial. Hauptzielsetzung der rekurrenten Selektion (RS) ist es deshalb, eine langfristige Leistungssteigerung der Zielmerkmale durch eine stetige Erhöhung der Frequenz günstiger Allele zu erreichen bei gleichzeitigem Erhalt der genetischen Variation in der Zuchtpopulation.

Für die Optimierung von Zuchtprogrammen hinsichtlich des Selektionserfolges sind Kenntnisse über die Wirkung von Selektion, Zufallsdrift und Gametenphasen-ungleichgewicht (Linkage disequilibrium, LD) von essentieller Bedeutung. Daher stand im Vordergrund dieses Forschungsvorhabens die molekulargenetische Evaluierung eines modifizierten rekurrenten Vollgeschwisterfamilien (VGF) -Selektionsprogramms von zwei europäischen Flintmaispopulationen. Ziele der Arbeit waren (1) die Untersuchung von genetischen Karten basierend auf Durchkreuzungspopulationen, (2) der Nachweis des Abfalls von elterlichem LD durch dreimaliges Durchkreuzen der F_2-Populationen, (3) die Lokalisierung von Genloci für quantitativ vererbte Merkmale (QTL) im Genom, die durch Selektion beeinflusst wurden, (4) die Trennung der Effekte von Zufallsdrift und Selektion auf Allelfrequenzänderungen in QTL-Regionen, (5) die Ermittlung des Ausmaßes an LD, das sich durch züchterische Selektion aufgebaut hat, und (6) die Untersuchung der Wirkung von LD auf den Selektionserfolg.

Als Ausgangsmaterial dienten vier homozygote Maisinzuchtlinien des europäischen Flintformenkreises, KW1292, D146, D145 und KW1265 (A, B, C und D), um 380 $F_{2:3}$ Linien der Population A × B und 140 $F_{3:4}$ Linien der Population C × D zu erzeugen. Zur Erstellung der F_2Syn3-Generation wurde die F_2-Generation beider Populationen in drei aufeinander folgenden Generationen mittels eines „chain crossing“-Verfahrens durchkreuzt. Ausgehend von der F_2Syn3 wurden anschließend vier (A × B) und sieben (C × D) Zyklen modifizierter rekurrenter VGF-Selektion durchgeführt. Die Selektion der VGF basierte auf einem in der mitteleuropäischen Maiszüchtung üblichen Selektionsindex (2 × Trockenmasse + Kornertrag, jeweils prozentual zur F_2-Generation). Die Durch-kreuzung in jedem Zyklus erfolgte nach einem pseudo-faktoriellen Schema, bei dem das

beste Drittel an selektierten Pflanzen zweifach als Kreuzungselter für die nächste Generation eingesetzt wurde und die übrigen zwei Drittel nur einfach. Für die Markeranalysen wurden 104 (A × B) und 101 (C × D) Mikrosatelliten (SSR = simple sequence repeats) eingesetzt, um 146 $F_{2:3}$ Linien in A × B und 110 $F_{3:4}$ Linien in C × D sowie 148 F_2Syn3-Pflanzen und die Eltern der 36 selektierten VGF in jedem Zyklus beider RS-Programme zu genotypisieren. Die SSR-Marker waren polymorph zwischen den jeweiligen Elternlinien und gleichmäßig über das Maisgenom verteilt.

Die genetischen Kopplungskarten für die SSR-Marker wurden für die Kartierungspopulationen $F_{2:3}$ (A × B) und $F_{3:4}$ (C × D) sowie für die Durchkreuzungsgeneration F_2Syn3 beider Populationen erstellt. Im Gegensatz zu früheren Studien wurden Kartierungsmethoden verwendetet, die speziell für Durchkreuzungsgenerationen entwickelt wurden, so dass sich die geschätzten Rekombinationsfrequenzen auf eine Meiose beziehen. Infolgedessen wurde die in anderen Studien beschriebene Kartenexpansion weder erwartet noch beobachtet. Um die erwartete höhere Kartierungsgenauigkeit von Durchkreuzungsgenerationen gegenüber deren Basisgenerationen zu überprüfen, wurde die Präzision der Schätzung der Rekombinationsfrequenzen r mit Hilfe des durchschnittlichen Informationsgehaltes per Individuum i_r jeweils in Relation zu einem F_2-Individuum quantifiziert. Für unendliche Populationsgrößen der Durchkreuzungspopulation war $i_r > 1$, solange $r < 0.14$. Für endliche Populationsgrößen mit 240 Individuen, wie in unserer Kartierungsstudie verwendet, war jedoch $i_r > 1$ nur für $r < 0.04$. Hieraus ergibt sich, dass Zufallsdrift einen großen Einfluss auf i_r hat und somit die Vorteile von Durchkreuzungsgenerationen meist aufhebt.

In erster Linie wurden die drei Durchkreuzungsgenerationen vor Beginn des RS-Programms durchgeführt, um anfängliches elterliches LD zwischen eng gekoppelten Loci und deren negativen Auswirkung auf den Selektionserfolg zu reduzieren. Unsere Ergebnisse zeigten, dass der beobachtete Abbau von LD sehr gut mit den theoretischen Erwartungen, welche mittels Simulationsstudien ermittelt wurden, übereinstimmt. Fisher′s exakter Test ergab, dass der Anteil gekoppelter Loci mit signifikantem LD von 50% in den Ausgangspopulationen auf 35,7% (A × B) und 30,9% (C × D) zurückging. Um den Vorteil von Durchkreuzungsgenerationen vor Beginn des RS-Programms auf den Selektionserfolg jedoch genau zu quantifizieren, sind weitere Untersuchungen notwendig.

Mit dem rekurrenten VGF-Selektionsprogramm wurde ein hoher Selektionserfolg für den Selektionsindex (A × B: 5,25%; C × D: 3,64%) und Kornertrag (A × B: 14,07%;

C × D: 8,28%) erzielt. Die molekulargenetische Auswertung ergab zudem, dass weiterer Selektionserfolg in zukünftigen Zyklen erwartet werden kann, da weder eine Fixierung noch ein Verlust von Allelen beobachtet wurde. Jedoch wurden mit Hilfe der SSRs auch Migrationseinflüsse seit den ersten Selektionszyklen festgestellt, so dass der durchschnittliche Anteil an nichtelterlichen Allelen auf bis zu ~10% anstieg. Die Migration, welche vermutlich auf Kontamination mit Fremdpollen zurückzuführen ist, konnte jedoch nicht auf dem Feld festgestellt werden, hatte aber vermutlich einen wesentlichen Einfluss auf den Selektionserfolg in beiden RS-Programmen.

Selektion aber auch Zufalldrift üben eine große Wirkung auf Allelfrequenzänderungen und somit auf den Selektionserfolg von RS-Programmen aus. Um deren Wirkung in unserem RS-Programm zu untersuchen, wurde ein Test verwendet, der die Trennung dieser beiden Faktoren ermöglicht. Signifikante Allelfrequenzänderungen auf Grund von Selektion konnten für 23% aller Marker nach einem Selektionszyklus und für 20% nach vier Selektionszyklen in Population A × B detektiert werden sowie in Population C × D für 6% nach einem Selektionszyklus und für 13% nach sieben Selektionszyklen. Des weiteren wurden QTL für den Selektionsindex und seine Komponenten in den Ausgangspopulationen $F_{2:3}$ (A × B) and $F_{3:4}$ (C × D) lokalisiert, wobei in einigen dieser Genomregionen Markerloci signifikante Allelfrequenzänderungen zeigten. Diese SSR können nun in zukünftigen marker-gestützten Selektionsprogrammen Verwendung finden.

Aufgrund theoretischer Ergebnisse wird erwartet, dass Selektion neben dem Anstieg der Frequenz günstiger Allele zu einem Aufbau von negativem LD zwischen diesen Loci führt. Dieses negative LD wiederum kann der Grund für eine Reduktion der genetischen Additivvarianz σ_A^2 sein, welches den langfristigen Selektionserfolg des RS-Programms schmälert. In unserer Studie zeigte die Entwicklung des LD jedoch einen wechselhaften Verlauf mit nur leichten Anstiegen in sowohl positivem als auch negativem LD. σ_A^2 wird dagegen nur reduziert, wenn sich negatives LD zwischen sehr vielen Markerloci aufbaut. Hieraus ergibt sich, dass LD kein limitierender Faktor in unserem RS-Programm war.

Die molekulargenetische Evaluierung des modifizierten rekurrenten VGF-Selektionsprogramms ermittelte Chromosomregionen, die unter züchterischer Selektion standen, verdeutlichte die Wichtigkeit von Markeranalysen parallel zum Selektionsprogramm und zeigte, dass weiterer Selektionserfolg in zukünftigen Selektionszyklen erwartet werden kann.

Acknowledgements

First, I am most grateful to Prof. Dr. A.E. Melchinger for his advice, continuous scientific support, and creative suggestions during this thesis work.

Thanks to Prof. Dr. H.-P. Piepho for his scientific contribution to two publications and for serving on my examination committee.

Sincere thanks to PD Dr. Matthias Frisch for his committed support, many discussions, and his never-ending patience in explaining quantitative genetic and statistical issues.

My appreciation is extended to all the people who helped me to conduct the investigations of this research. The skilled technical assistance of Franz Mauch as well as of Dr. D. Klein and his team at the research station Eckartsweier in conducting the field trials is gratefully acknowledged. I appreciate the laboratory assistance of Christine Beuter and Elisabeth Kokai-Kota, and thank many students, especially Carolin Rauner and Angela Ritter, who helped me with the marker analyses.

I wish to thank Dr. Martin Heckenberger, Hans Peter Maurer and Dr. Katinka Wilde for their support as well as all present and alumni members of the institute for creating a pleasant working environment. I particularly enjoyed the joint time with my laboratory-, office-, and project-mate Dr. Christian Flachenecker and thank him for his helpfulness and tolerance.

Many thanks go to Dr. Eva Bauer and Dr. Christina Dussle for being of great help in all molecular genetics issues.

Best thanks to my one-year office mate Dr. Jasmina Muminović for her editorial suggestions, which improved my publications considerably.

Special thanks to Prof. Dr. H. F. Utz for his support and valuable advice on handling QTL mapping studies.

I thank Mrs. B. Boesig, Mrs. H. Beck, Mrs. S. Meyer, and Beate Devesi-Savula for their friendliness and helpfulness in everyday life in the institute.

Many thanks to KWS Saat AG and Pioneer Hi-Bred Northern Europe for their support with the SSR screening, as well as to Dr. Johan van Ooijen of Kyazma BV (Wageningen, Netherlands) for providing a mapping software for mapping intermated populations.

This research was supported by grants from the Deutsche Forschungsgemeinschaft (Grant No. FR 1615/1). The financial support according to the Landesgraduierten-förderungssgesetz (LGFG) of the state Baden-Württemberg is gratefully acknowledged.

Finally, I wish to thank “the Master of MS Office”, “the Specialist of Plabsoft”, and the “girls” for many amusing scientific discussions in the office 0.26 sometimes till late at night.

My personal thanks go to my family and Willi for their encouragement, tolerance, and patience during my thesis work.