Theor Appl Genet (2011) 123:1269–1279
DOI 10.1007/s00122-011-1665-x

ORIGINAL PAPER

Gene stacking strategies with doubled haploids derived from biparental crosses: theory and simulations assuming a finite number of loci

Albrecht E. Melchinger · Frank Technow · Baldev S. Dhillon

Received: 28 April 2011 / Accepted: 12 July 2011 / Published online: 3 August 2011

Abstract Recent progress in genotyping and doubled haploid (DH) techniques has created new opportunities for development of improved selection methods in numerous crops. Assuming a finite number of unlinked loci (ℓ) and a given total number (n) of individuals to be genotyped, we compared, by theory and simulations, three methods of marker-assisted selection (MAS) for gene stacking in DH lines derived from biparental crosses: (1) MAS for high values of the marker score (T, corresponding to the total number of target alleles) in the F_2 generation and subsequently among DH lines derived from the selected F_2 individual (Method 1), (2) MAS for augmented F_2 enrichment and subsequently for T among DH lines from the best carrier F_2 individual (Method 2), and (3) MAS for T among DH lines derived from the F_1 generation (Method 3). Our objectives were to (a) determine the optimum allocation of resources to the F_2 (n_1^*) and DH generations ($n - n_1^*$) for Methods 1 and 2 by simulations, (b) compare the efficiency of all three methods for gene stacking by simulations, and (c) develop theory to explain the general effect of selection on the segregation variance and interpret our simulation results. By theory, we proved that for smaller values of ℓ, the segregation variance of T among DH lines derived from F_2 individuals, selected for high values of T, can be much smaller than expected in the absence of selection. This explained our simulation results, showing that for Method 1, it is best to genotype more F_2 individuals than DH lines ($n_1^* : n > 0.5$), whereas under Method 2, the optimal ratio $n_1^* : n$ was close to 0.5. However, for ratios deviating moderately from the optimum, the mean $\overline{X}$ of T in the finally selected DH line (T_{DH}^*) was hardly reduced. Method 3 had always the lowest mean $\overline{X}$ of T_{DH}^* except for small numbers of loci ($\ell = 4$) and is favorable only if a small number of loci are to be stacked in one genotype and/or saving one generation is of crucial importance in cultivar development. Method 2 is under most circumstances the superior method, because it generally showed the highest mean $\overline{X}$ and lowest SD of T_{DH}^* for the finally selected DH.

Introduction

Inbred line development is a key component in breeding of line, synthetic, and hybrid varieties (Schnell 1982). Traditionally, inbred lines are developed by recurrent self-pollination that occurs naturally in autogamous crops but requires controlled pollination in allogamous species. During the past two decades, techniques for producing doubled haploid (DH) lines have been developed in many crops (Wedzony et al. 2009), which facilitate and accelerate the large-scale production of completely homozygous lines. In most instances, implementation of the DH technique reduces the length of the breeding cycle dramatically and results in a substantial increase in the selection gain per time unit (Longin et al. 2006). Hence, DH techniques are being more and more routinely used in breeding of many crops such as maize (*Zea mays*), rice (*Oryza sativa*), wheat

Communicated by J. Wang.

A. E. Melchinger and F. Technow contributed equally.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-011-1665-x) contains supplementary material, which is available to authorized users.

A. E. Melchinger (✉) · F. Technow · B. S. Dhillon
Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, 70593 Stuttgart, Germany
e-mail: melchinger@uni-hohenheim.de

(*Triticum* spp.), barley (*Hordeum vulgare*), and oilseed rape (*Brassica napus*) (Tuvesson et al. 2007; Wedzony et al. 2009; Wu 2010).

In commercial breeding programs, the most common type of source germplasm employed for conventional line development are F_2 generations from biparental crosses between elite lines (Darrah and Zuber 1986; Fehr 1987; Mikel and Dudley 2006). Similarly, DH lines are generally produced from F_1 genotypes (Röber et al. 2005). Alternatively, DH lines may also be produced from the F_2 generation to allow for one additional generation of recombination (Longin et al. 2007; Bernardo 2009).

Rapid advances have been made recently in high-throughput genotyping and sequencing (Varshney et al. 2009). Hence, in numerous crop species, a vast number of inexpensive molecular markers, such as simple sequence repeats and single-nucleotide polymorphisms, have become available for marker-assisted selection (MAS) and genomic selection (Heffner et al. 2010). One method of MAS described by Howes et al. (1998) and Bonnett et al. (2005) is F_2 enrichment, in which F_2 individuals homozygous for non-target alleles are discarded and carriers of target alleles are retained. As shown in simulations studies by these and other researchers (Howes et al. 1998; Bonnett et al. 2005; Wang et al. 2007), this method enhances the frequency of target alleles in the enriched F_2 population and consequently in DH lines derived from the selected F_2 individuals and reduces the minimum population size required to recover one target homozygote compared with the non-enriched F_2 population. Yet, no theory is available for describing the genetic variation and optimizing the individual steps in this method. Further, the enriched F_2 population will generally have several individuals, and consequently, F_2 enrichment can be augmented by selection among the enriched F_2 individuals for the highest total marker score (T), a method subsequently referred to as augmented F_2 enrichment.

In gene stacking or synonymously gene pyramiding, several target genes at different loci from two or several sources are combined into one genotype. Gene stacking through conventional plant breeding is time and resource consuming. Sometimes it is even not possible in a targeted manner, if the effect of various genes cannot be easily discriminated at the phenotypic level, as holds often true for resistance genes. MAS has greatly facilitated gene stacking in various crops (Wu 2010) with prominent examples in resistance breeding against blast (*Magnaporthe grisea)* and bacterial leaf blight (*Xanthomonas oryzae)* in rice (Koide et al. 2010; Perumalsamy et al. 2010) and stem rust (*Puccinia graminis)* in wheat (Mago et al. 2011).

Quantitative genetic theory has focused on either single and two-locus models (Kearsey and Pooni 1996) or the infinite loci model (see Lynch and Walsh 1997). However, QTL studies revealed for numerous quantitative traits a small number of QTL with large effects (Bernardo 2008). With finite number of loci, the distribution of genotypic values of DH lines derived from a single F_2 parent, even though not normal, is necessarily symmetric. Therefore, the segregation variance (i.e., the genetic variance among segregating progeny) is still a useful parameter to describe the genotypic variation and prospects of selection. Selection for these traits in early generations is expected to result in a rapid fixation of alleles. This alters the ratio of genetic variances in different generations and influences the optimum allocation of resources for selection. These aspects have so far not been investigated for selection in the F_2 generation and among DH lines derived from the selected F_2 individual(s).

In our study, we compared three breeding methods of MAS for gene stacking in DH lines developed from a biparental cross of homozygous parents: (1) MAS for highest total marker score (T) in the F_2 generation and subsequently for T among DH lines developed from the selected F_2 individual (Method 1), (2) MAS in F_2, first for enrichment and second for T (augmented F_2 enrichment), and subsequently for T among DH lines obtained from the best carrier F_2 individual (Method 2), and (3) MAS for T among DH lines derived directly from the F_1 generation (Method 3).

We developed theory to explain the effects of selection on the segregation variance assuming a finite number of loci. The theory is, however, limited to the case of one-step selection in the F_2 generation. Furthermore, we assumed only selection of a certain proportion of individuals, whereas in practice plant breeders are interested to select the best individual in gene stacking. Full representation of this situation by theory would require recourse to order statistics. Since a fully analytical description would be extremely complex, especially for the three selection steps in Method 2, we carried out simulations to take care of these two aspects. Further, simulations also enabled determination of the relative efficiency of the different methods and allocation of resources to different generations.

Our objectives were to (a) determine for Method 1 and Method 2 the optimum allocation of resources between the F_2 and DH generations, with a fixed number of individuals for genotyping at a given finite number of loci by simulations, (b) compare the efficiency of all three MAS methods for gene stacking by simulations, and (c) develop theory to explain the general effect of selection on the segregation variance and interpret our simulation results.

Theory

We consider an F_2 population with ℓ independently segregating marker loci $i = 1, \ldots, \ell$ derived from a biparental cross of two homozygous parents, P1 and P2. Let the

genotype of a random F_2 individual U at locus i be coded by the random variable U_i, where $U_i = 2$ or 1, if U at locus i is homozygous or heterozygous for the target allele, respectively, and $U_i = 0$, if it is homozygous for the non-target allele.

We define for $c = 0, 1, 2$ the random variables M_c as

$$M_c = \sum_{i=1}^{\ell} a_i I_{[U_i=c]}, \tag{1}$$

where $I_{[U_i=c]}$ is an indicator variable assuming the values 1 or 0, if the condition $U_i = c$ is true or false, respectively, and $a_i > 0$ refers to the additive effect of the favorable allele at locus i, but to simplify matters, we assume $a_i = 1$ for all loci. In this case, M_2 and M_1 correspond to the number of loci homozygous or heterozygous for the target allele in U, respectively, M_0 is the number of loci homozygous for the non-target allele in U, and $M_0 + M_1 + M_2 = \ell$.

Furthermore, we define the random variable T referring to the total marker score as

$$T = 2M_2 + M_1, \tag{2}$$

with values in $\Omega = \{0, 1, \ldots, 2\ell\}$. Thus, T corresponds to the total number of target alleles and differs from the mean molecular score of Lande and Thompson (1990), which is the sum of estimates of additive effects a_i associated with the markers included in the selection index.

Since the random variables M_0, M_1, and M_2 can be regarded as the number of outcomes from ℓ independently repeated trials for each locus i, each having three outcomes with probabilities $\Pr[U_i = 0] = 1/4$, $\Pr[U_i = 1] = 1/2$, and $\Pr[U_i = 2] = 1/4$, the joint distribution of M_0, M_1, and M_2 follows a multinomial distribution with probability density function for $0 \le i \le \ell$ and $0 \le j \le \ell - i$

$$p_{ij} = \Pr(M_2 = i, M_1 = j) = \frac{\ell!}{i!j!(\ell - i - j)!}\left(\frac{1}{2}\right)^j\left(\frac{1}{4}\right)^{\ell-j}. \tag{3}$$

If we denote DH lines derived from an F_2 individual as F_2:DH, then we obtain for the variance of the marker score among DH lines derived from F_2 individual U: $\sigma^2_{F_2:\mathrm{DH}}(U) = M_1$, because only those marker loci, which are heterozygous in U, will segregate among the derived DH lines.

For given values of $T = t \in \Omega$, M_1 can assume only values $j \in \Psi(t)$ with

$$\Psi(t) = \{s \mid s = 0, 2, \ldots, t\} \quad \text{if } t \text{ is even},$$

and

$$\Psi(t) = \{s \mid s = 1, 3, \ldots, t\}, \text{ if } t \text{ is odd}.$$

Hence, we obtain for the joint probability density function of T and M_1, the probability density function of T, and the conditional probability of M_1, given $T = t$ and $j \in \Psi(t)$:

$$p_{j,t} = \Pr(M_1 = j, T = t) = \frac{\ell!}{\left(\frac{t-j}{2}\right)!j!\left(\ell - \frac{(t+j)}{2}\right)!}\left(\frac{1}{4}\right)^{\ell} 2^j, \tag{4}$$

$$p_t = \Pr(T = t) = \sum_{j\in\Psi(t)} p_{j,t}, \tag{5}$$

$$p_{j|t} = \Pr(M_1 = j | T = t) = \frac{p_{j,t}}{p_t}. \tag{6}$$

The expected segregation variance for the marker score T among DH lines derived from a selected F_2 individual with $T = t$ is obtained as the conditional expectation of M_1 given $T = t$

$$E\left(\sigma^2_{T,F_2:DH} | T = t\right) = E(M_1 | T = t) = \sum_{j\in\Psi(t)} j p_{j|t} = \frac{\sum_{j\in\Psi(t)} j p_{j|t}}{p_t}. \tag{7}$$

For the expectation and the variance of the marker score T among unselected F_2 individuals, we have

$$E(T) = 2E(M_2) + E(M_1) = \ell, \quad \text{and} \quad \sigma^2_{T,F_2} = \ell/2 \tag{8}$$

and for the expected segregation variance among DH lines derived from unselected F_2 individuals, we get from the multinomial distribution

$$\sigma^2_{T,F_2:DH} = E(M_1) = E[E(M_1 | T = t)] = \sum_{t=0}^{\ell}\sum_{j\in\Psi(t)} j p_{j|t} = \ell/2. \tag{9}$$

Likewise, for DH lines extracted directly from the F_1 generation, denoted as F_1:DH, we obtain the expectation and the variance of T from the sum of independent random variables, each following a binomial distribution $B(\ell, p_1 = 1/2)$, as

$$E(T) = \ell \text{ and } \sigma^2_{T,F_1:\mathrm{DH}} = \ell. \tag{10}$$

For given values of $T = t \le 2\ell$, M_0 can assume only values $k \in \Phi(t)$ with

$$\Phi(t) = \{k | k = \max(0, \ell - t), \max(0, \ell - t) + 1, \ldots, \ell - [t/2]\} \tag{11}$$

where $[t/2]$ is the integer closest to $t/2$. Hence, we obtain for the joint probability density function of T and M_0 and the conditional probability of M_0, given $T = t$ and $k \in \Phi(t)$, the following expressions:

$$p_{k,t} = \Pr(M_0 = k, T = t) = \frac{\ell!}{k!(2\ell - 2k - t)!(k + t - \ell)!}\left(\frac{1}{4}\right)^{\ell} 2^{(2\ell-2k-t)}, \tag{12}$$

$$p_{k|t} = \Pr(M_0 = k | T = t) = \frac{p_{k,t}}{p_t}. \tag{13}$$

The expected number of loci fixed for the non-target allele in a selected F_2 individual with marker score $T = t$ is obtained as the conditional expectation of M_0 given $T = t$

$$E(M_0|T=t) = \sum_{k\in\Phi(t)} k p_{k|t} = \frac{\sum_{k\in\Phi(t)} k p_{k,t}}{p_t}. \quad (14)$$

In F_2 enrichment, the F_2 individuals are first selected for a low number (ideally zero) of marker loci fixed for the non-target alleles. To calculate the expected variance of DH lines extracted from F_2 individuals with $M_0 = k$, we define for $0 \le j \le \ell - k$

$$p_{j,k} = \Pr(M_1 = j,\, M_0 = k) = \frac{\ell!}{(\ell-j-k)!j!k!}\left(\frac{1}{2}\right)^j\left(\frac{1}{4}\right)^{\ell-j} \quad (15)$$

$$p_k = \Pr(M_0 = k) = \sum_{j=0}^{\ell-k} p_{j,k} = \frac{\ell!}{k!(\ell-k)!}\left(\frac{1}{4}\right)^{\ell} 3^{\ell-k} \quad (16)$$

$$p_{j|k} = \Pr(M_1 = j|M_0 = k) = \frac{p_{j,k}}{p_k} = \frac{(\ell-k)!}{(\ell-k-j)!j!}\left(\frac{1}{3}\right)^{(\ell-k)} 2^j. \quad (17)$$

Hence, the expected segregation variance for marker score T among DH lines derived from a randomly chosen F_2 individual with $M_0 = k$ is obtained as the conditional expectation of M_1 given $M_0 = k$

$$E\left(\sigma^2_{T,F_2:DH}|M_0 = k\right) = E(M_1|M_0 = k) = \sum_{j=0}^{\ell-k} j\, p_{j|k} = \frac{2}{3}(\ell-k). \quad (18)$$

Graphical representations of $E(M_1|\,T \ge t)$ and the cumulative distribution function of p_t as a function of t are given in Figs. 1 and 2, respectively. Likewise, graphs for $E(M_1|M_0 \le k)$ and the cumulative distribution function of p_k as a function of k are presented in Figs. 3 and 4, respectively.

Using the same approach, we obtain for the conditional expectation and variance of the marker score T of a randomly chosen F_2 individual with $M_0 = k$

$$E(T|\,M_0 = k) = 4(\ell-k)/3 \quad (19)$$

and

$$var(T|\,M_0 = k) = 2(\ell-k)/9. \quad (20)$$

Thus, there is still ample variation for T among F_2 individuals selected for small values of M_0.

Making use of $E\left(\mathrm{I}_{[U_i=c]}\mathrm{I}_{[U_j=z]}\right) = \Pr[U_i = c]\Pr[U_j = z]$ for $i = j$ and zero otherwise, we obtain for the covariances in the F_2 generation

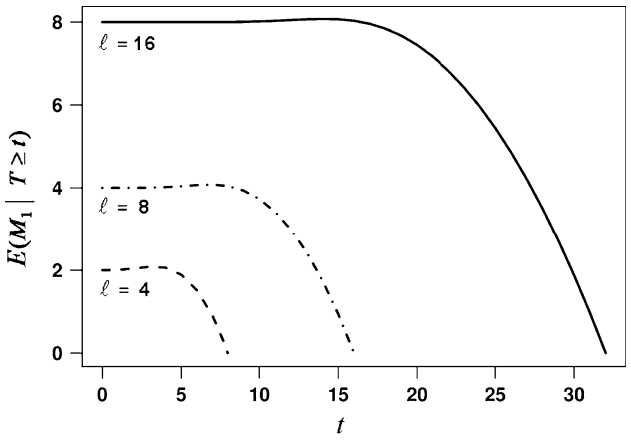

Fig. 1 Expectation for the number of heterozygous loci M_1 in individuals from the selected fraction of an F_2 population, $E(M_1|T \ge t)$, as a function of the threshold t for the total marker score T (ℓ = number of unlinked loci)

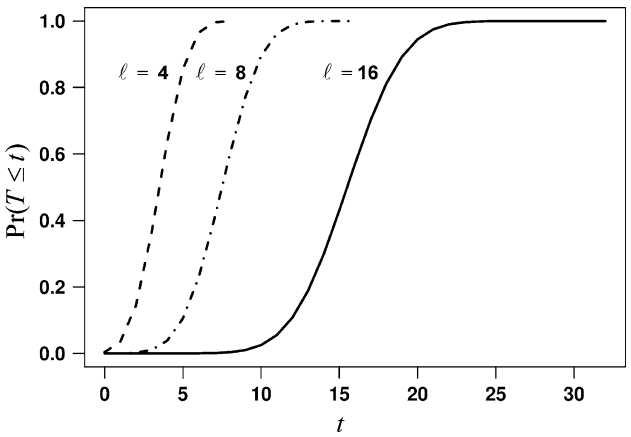

Fig. 2 Cumulative distribution function $F(t) = \Pr(T \le t)$ for the total marker score T in random individuals from an F_2 population (ℓ = number of unlinked loci)

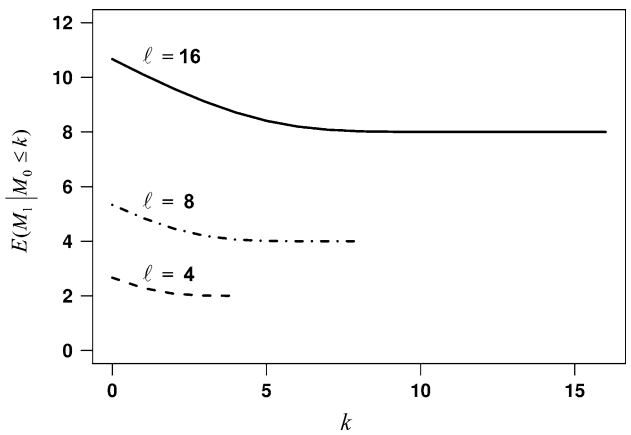

Fig. 3 Expectation for the number of heterozygous loci M_1 in individuals from the selected fraction of an F_2 population, $E(M_1|M_0 \le k)$, as a function of the threshold k of marker-assisted selection against M_0, the number of loci fixed for the non-target allele (ℓ = number of unlinked loci)

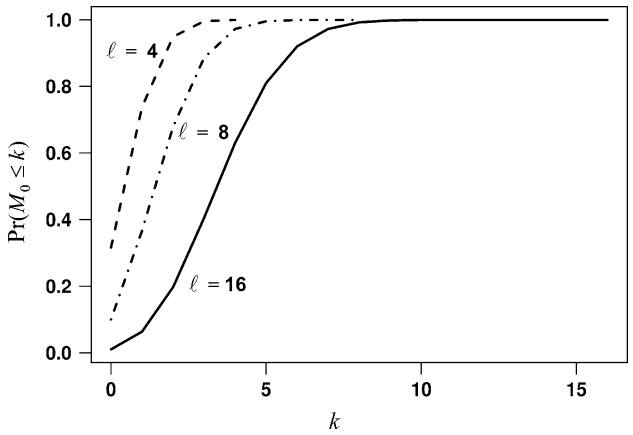

Fig. 4 Cumulative distribution function $F(k) = \Pr(M_0 \leq k)$ for M_0, the number of loci fixed for the non-target allele, in random individuals from an F_2 population (ℓ = number of unlinked loci)

$$\mathrm{Cov}(M_1, M_0) = -\ell/8 \text{ and } \mathrm{Cov}(M_2, M_0) = -\ell/16 \quad (21)$$

$$\mathrm{Cov}(T, M_0) = -\ell/4 \text{ and } \mathrm{Cov}(T, M_1) = 0. \quad (22)$$

Thus, the correlation (ρ) of T with M_0 and M_1 becomes

$$\rho(T, M_0) = -\sqrt{2/3} = -0.8165 \quad (23)$$

and

$$\rho(T, M_1) = 0. \quad (24)$$

Here, we concentrated on gene stacking with F_1 or F_2 generations from biparental crosses. Instead, if one of the parents carries a much larger number of favorable genes than the other, it might be better to start with a backcross to this parent, as often practiced in conventional breeding (Fehr 1987; Mikel and Dudley 2006). Our theory can be used to cope with this situation simply by inserting appropriate genotype frequencies $\Pr[U_i = c]$ in the formulas.

Methods

DH selection methods

We compared the following three methods of MAS for gene stacking in DH lines derived from the cross P1 × P2 by computer simulations and our theoretical results. In each method, the ultimate goal was to maximize the total marker score T in the finally selected DH line. Furthermore, the total number n of F_2 individuals and/or DH lines to be genotyped was assumed to be fixed.

Method 1 In Step 1 of this method, n_1 F_2 individuals from the cross P1 × P2 are genotyped and one F_2 individual with the highest marker score T across all ℓ loci is selected and used to develop $n_2 = n - n_1$ DH lines. Among these, the DH line with the highest marker score T is selected in Step 2.

Method 2 In Step 1A, which corresponds to F_2 enrichment, n_1 F_2 individuals are genotyped and those with minimum M_0 are identified following the approach of Bonnett et al. (2005). Among these n_{1B} individuals, the F_2 individual with the highest marker score T is selected in Step 1B, i.e., F_2 enrichment is augmented with selection within the enriched F_2. The selected F_2 individual is used to produce $n_2 = n - n_1$ DH lines, and in Step 2, the DH line with the highest marker score T is identified as in Method 1.

Method 3 In this method, n DH lines are produced directly from the F_1 generation of the cross P1 × P2 and the DH line with the highest marker score T is selected in a single step.

Simulations

Multilocus genotypes of all candidates for each method of MAS were generated by stochastic simulations in the R software environment (R Development Core Team 2010). For each F_2 individual U, the genotype at locus $i = 1, \ldots \ell$ was obtained by sampling a random variable U_i with realizations 0, 1, and 2 (corresponding to the number of target alleles) from a multinomial distribution with probabilities $\Pr[U_i = 0] = \Pr[U_i = 2] = 1/4$, and $\Pr[U_i = 1] = 1/2$. Likewise, the genotype of a DH line V at locus i was generated by sampling a random variable V_i with realizations 2 and 0 from a Bernoulli distribution with $\Pr[V_i = 0] = \Pr[V_i = 2] = 1/2$. If the DH line V was derived from F_2 individual U, this sampling was restricted to heterozygous loci with $U_i = 1$, while for the other loci we set $V_i = U_i$.

A total of 250,000 simulation runs were performed for each method of MAS and given values of ℓ, n, and n_1. Selection of the candidates in each method was carried out as described above. The marker score for individuals U and V was calculated according to Eq. 2.

We denoted the marker score T of the F_2 individual selected in Step 1 of Method 1 or 2 for production of DH lines by T_{F_2}. The marker score of the finally selected DH line under each method was denoted by T_{DH}. Upon request to the authors, the simulation programs, implemented in the R language, will be made available.

Data analysis

For each method and each scenario (i.e., set of parameter values ℓ, n, and n_1) considered, we calculated the mean ($\overline{X}$) and standard deviation (SD) for the distribution of T_{DH}

across the 250,000 runs. For Method 1 and Method 2, we determined $\overline{X}$ and SD also for T_{F_2}, M_1 and M_0 of the F_2 individual selected for producing the DH lines. In addition, we averaged for Method 2 the number of individuals n_{1B} remaining after F_2 enrichment (Step 1A) for selection in Step 1B. Finally, we determined for $1 \leq n_1 \leq n$ the number n_1^*, which resulted in the highest mean of $\overline{X}$ of T_{DH} for Method 1 and Method 2. Only those values of T_{DH}, T_{F_2} etc. referring to n_1^* were reported and are indicated by an asterisk. To facilitate a direct comparison of the results for different number of loci ℓ, we expressed the statistics for M_1 and M_0 as percentage of ℓ. Likewise, we expressed $\overline{X}$ and SD for T_{DH}^* and $T_{F_2}^*$ as percentage of 2ℓ.

Results

Theory

With MAS for T in generation F_2 (Step 1 of Method 1), the curves for $E(M_1|T \geq t)$ based on Eqs. 5 and 7 showed the same trend irrespective of the value of ℓ (Fig. 1). At $t = 0$, the curves started at $\ell/2$, as expected in the absence of selection, slightly increased up to $t = \ell$, where they assumed their maximum, and afterward dropped toward zero for larger values of t, with ordinates of approximately 0.375 ℓ for $t = 1.5\ \ell$. The cumulative distribution function $F(t) = \Pr(T \leq t)$ had its steepest increase for $t \approx \ell$ and rapidly approached 1.0 for larger values of t depending on ℓ (Fig. 2).

For $t = 1.5\ell$, $\Pr(T \geq t) = 1 - F(t-1)$ amounted to 0.1445, 0.0384, and 0.0035 for $\ell = 4$, 8, and 16, respectively (Fig. 2). These numbers are needed to calculate the probability $\alpha(\ell, n_1) = 1 - [1 - \Pr(T \geq t)]^{n_1}$ that in a F_2 population of size n_1, at least one individual with $T \geq 1.5\ \ell$ will occur. To achieve $\alpha(\ell, n_1) \geq 0.95$, a sample size $n_1 = 19, 76,$ and 854 is required for $\ell = 4, 8,$ and 16, respectively. Thus, if MAS under Method 1 is practiced in an F_2 population of size n_1 larger than these values, then chances are fairly high that the segregation variance of T among DH lines derived from the selected F_2 individual falls below the value of 0.375 ℓ. This value corresponds to a reduction of at least 25% compared with the variance 0.5 ℓ expected among DH lines derived from unselected F_2 individuals.

With MAS for low values of M_0 in the F_2 generation (Step 1A of Method 2), the curves for $E(M_1|M_0 \leq k)$ based on Eqs. (16) and (18) started at the maximum value of $2\ell/3$ for $k = 0$ and rapidly declined to $\ell/2$ for $k = \ell$ (Fig. 3). The cumulative distribution function $F(k) = \Pr(M_0 \leq k)$ started at $k = 0$ with values considerably higher than zero for $\ell = 4$ and 8, whereas for $\ell = 16$, the value was nearly zero (Fig. 4). Thereafter, the curves showed a steep increase and approached their limit value of 1.0 for $k \geq 3\ell/4$. Thus, unless ℓ is large, even with a relatively moderate sample size n_1, the probability $\beta(\ell, n_1, k) = 1 - [1 - P(M_0 \leq k)]^{n_1}$ to obtain in Step 1A of Method 2 an F_2 individual, in which none $(k = 0)$ or only a small number k of the marker loci is fixed for the non-target allele, is fairly high. The selected individual is then expected to carry up to $2\ell/3$ heterozygous marker loci, which corresponds to 33.3% more heterozygous marker loci than $\ell/2$, the number expected in the absence of selection.

Simulations

The highest selection progress under Method 1, as measured by the mean $\overline{X}$ of the marker score T_{DH}^* of the finally selected DH line, was obtained by genotyping more F_2 individuals than DH lines (Table 1). In general, the ratio n_1^*: n ranged between 0.51 and 0.76 and was greater for smaller values of ℓ and higher values of n. By comparison, the highest mean $\overline{X}$of T_{DH}^* under Method 2 was achieved with a ratio n_1^*: n between 0.44 and 0.52, irrespective of the values of ℓ and n. Nevertheless, the optima for $\overline{X}$ of T_{DH}^* as a function of the ratio n_1: n were rather flat under both methods, i.e., if n_1 deviated slightly from n_1^*, this resulted only in a marginal reduction of the expected selection progress for gene stacking under MAS.

The mean $\overline{X}$ of $T_{F_2}^*$ was about 1.7 to 4.1% points smaller for Method 2 than Method 1, whereas its SD was about 0.8 to 2.1% points higher (Table 1). Under both methods, $\overline{X}$ of $T_{F_2}^*$ decreased rapidly with increasing values of ℓ and showed only small increments for larger values of n. For all values of ℓ and n analyzed, the SD of T_{F_2} for either method was larger than the observed differences in their means.

The mean $\overline{X}$ of the proportion of heterozygous loci M_1 in the selected F_2 individual under Method 2 was always higher by about 7 to 18% points than under Method 1 (Table 1). For both Method 1 and Method 2, M_1 increased with larger ℓ and smaller n, but the effects of increasing values of n became smaller when ℓ also increased. Moreover, the differences in M_1 between methods were largest for $\ell = 12$. Conversely, $\overline{X}$ of the proportion of loci M_0 fixed for the non-target loci was higher in Method 1 than in Method 2, where it was always below 3% except for $\ell = 16$ and $n \leq 150$ and $\ell = 12$ and $n = 50$.

The mean $\overline{X}$ of the marker score T_{DH}^* of the finally selected DH line was highest for Method 2, followed by Method 1 and lowest for Method 3 (Table 2). This ranking was true for all examined scenarios except $\ell = 4$ and $n = 50$, but the differences among the three methods varied depending on ℓ and n. Under all three methods, $\overline{X}_{\mathrm{DH}}$

Table 1 Optimal number of plants in the F_2 generation (n_1^*) for achieving the maximum marker score in the finally selected DH line T_{DH}^*, mean ($\overline{X}$) and standard deviation (SD) of the marker score of the best selected F_2 individual $T_{F_2}^*$, and the proportion of heterozygous loci (M_1^*) and non-target fixed loci (M_0^*) in the best selected F_2 individual

ℓ	n[a]	n_I^*	Method 1						n_I^*	Method 2						
			$T_{F_2}^*$[b]		M_1^*[c]		M_0^*[c]			n_{1B}^*[d]	$T_{F_2}^*$[b]		M_1^*[c]		M_0^*[c]	
			$\overline{X}$ (%)	SD (%)	$\overline{X}$ (%)	SD (%)	$\overline{X}$ (%)	SD (%)			$\overline{X}$ (%)	SD (%)	$\overline{X}$ (%)	SD (%)	$\overline{X}$ (%)	SD (%)
4	25	15	79.7	9.2	35.0	19.5	2.8	7.9	13	4.2	78.0	10.2	43.6	20.5	0.2	2.1
	50	38	86.0	7.8	26.2	15.7	0.9	4.7	26	8.2	83.4	8.6	33.1	17.1	0	0.1
8	50	28	74.7	6.1	39.6	14.7	5.5	7.1	26	3.1	72.0	7.8	54.3	16.1	0.8	3.1
	75	48	77.2	5.6	36.9	13.9	4.3	6.4	38	4.0	74.2	7.3	51.1	14.8	0.2	1.7
	100	68	78.7	5.3	35.1	13.3	3.7	6.0	51	5.2	75.9	6.8	48.1	13.7	0.1	0.8
	150	106	80.6	5.0	32.9	12.5	3.0	5.5	76	7.6	78.1	6.1	43.8	12.2	0	0.2
12	50	28	70.3	5.0	43.1	12.8	8.2	6.6	25	2.3	67.3	6.4	57.7	14.5	3.8	4.4
	75	43	72.0	4.7	41.6	12.5	7.2	6.3	36	2.6	68.5	6.4	57.7	14.6	2.6	3.9
	100	55	72.9	4.5	40.8	12.2	6.7	6.1	45	2.8	69.2	6.4	57.6	14.5	2.0	3.5
	150	87	74.6	4.3	39.2	11.8	5.8	5.8	67	3.1	70.5	6.3	57.0	13.8	1.0	2.7
	200	121	75.7	4.1	38.1	11.4	5.3	5.5	94	3.6	71.8	6.2	55.6	12.9	0.4	1.8
16	50	27	67.5	4.4	44.9	11.5	10.1	6.1	23	2.0	64.6	5.7	58.0	12.4	6.4	4.2
	75	38	68.7	4.2	44.0	11.3	9.3	5.9	36	2.2	65.9	5.6	58.3	12.5	4.9	3.7
	100	51	69.7	4.0	43.3	11.1	8.6	5.8	45	2.3	66.5	5.6	58.3	12.4	4.3	3.5
	150	80	71.1	3.8	42.1	10.8	7.8	5.5	70	2.7	67.7	5.4	58.3	12.6	3.1	3.2
	200	110	72.1	3.7	41.2	10.6	7.3	5.4	87	2.9	68.3	5.4	58.2	12.7	2.6	3.1

[a] Total number n of F_2 individuals and DH lines that can be genotyped

[b] In percent of the maximum marker score 2 ℓ

[c] In percent of the total number of loci ℓ

[d] Average number of individuals available for selection in Step 1B

was enhanced as ℓ decreased and n increased. The superiority of Method 2 over Method 1 ranged between 3.1 and 0.7% points for $\ell = 4$, 8, 12, but was lower for $\ell = 16$, and it generally increased for larger values of n. By comparison, Method 1 was as good as or marginally superior over Method 3 for $\ell = 4$, but for $\ell = 8$, 12, 16, its superiority was about 3.8 to 6.9% points across all values of n. Method 2 had consistently the smallest SD of T_{DH}^*, and its value was extremely small, when $\overline{X}$ was close to 100%. Method 1 and Method 3 showed similar SD for T_{DH}^*. Accordingly, with Method 2, there is very high probability of recovering almost all target alleles ($\approx$100%) in the finally selected DH line for $\ell = 4$ with $n = 50$, $\ell = 8$ with $n = 150$, and for $\ell = 12$ and $\ell = 16$ still about 96 and 91% of the loci, respectively, are fixed on average for the target allele by genotyping a total number of only $n = 200$ individuals.

Discussion

The rapid advances in genotyping and DH development have opened many avenues for designing more efficient breeding and selection strategies with the objective of maximizing selection progress using minimum time and resources. We investigated gene stacking in the context of DH lines developed from the F_1 or F_2 generation and three methods of MAS, one of which included an augmented form of F_2 enrichment. Many researchers (Howes et al. 1998; Bonnett et al. 2005; Wang et al. 2007) have observed in their simulation studies that F_2 enrichment enhanced the frequency of target alleles in DH lines derived from enriched F_2. One objective of our study was to provide the theory for these schemes of MAS and interpreting the results of simulation studies, which has been lacking so far in the literature. Our results have important implications for the optimum allocation of resources in genotyping the F_2 and DH generations with a relatively low number of target loci.

Effects of selection on segregation variances and resource allocation

Based on quantitative genetic theory, the total additive genetic variance (σ_A^2) for a trait among unselected DH lines

Table 2 Mean ($\overline{X}$) and standard deviation (SD) of the marker score T^*_{DH} of the selected DH line under various marker-assisted selection methods with ℓ unlinked loci and a total number n of genotyped individuals, based on 250,000 sampling runs

ℓ	n[a]	Method 1		Method 2		Method 3	
		$\overline{X}$ (%)[b]	SD (%)[b]	$\overline{X}$ (%)[b]	SD (%)[b]	$\overline{X}$ (%)[b]	SD (%)[b]
4	25	96.2	9.0	98.6	5.8	95.0	10.0
	50	98.9	5.2	100.0	1.0	99.0	4.9
8	50	92.6	7.4	94.9	6.8	87.7	7.4
	75	94.6	6.7	97.6	5.1	89.9	6.7
	100	95.7	6.3	98.8	3.7	91.2	6.5
	150	96.8	5.6	99.7	1.8	93.0	6.3
12	50	86.9	6.6	87.6	6.8	81.3	6.3
	75	89.3	6.1	90.5	6.2	83.3	6.0
	100	90.9	5.9	92.3	5.7	84.6	5.7
	150	92.7	5.6	94.7	5.2	86.4	5.3
	200	93.7	5.4	96.3	4.6	87.4	5.2
16	50	82.8	5.9	82.8	6.2	77.4	5.5
	75	85.2	5.5	85.4	5.8	79.1	5.2
	100	86.7	5.3	87.2	5.5	80.4	5.1
	150	88.7	5.0	89.4	5.1	81.9	4.8
	200	89.9	4.9	90.9	4.8	83.0	4.6

[a] Total number n of F_2 individuals and DH lines that can be genotyped

[b] In percent of the maximum marker score 2 ℓ

derived from the F_2 generation is obtained as the sum of σ^2_{A,F_2} and $\sigma^2_{A,F_2:DH}$ (Kearsey and Pooni 1996), adopting the notation used in our theory section. For unlinked loci, these variances are exactly of equal size, i.e., $\sigma^2_{A,F_2} = \sigma^2_{A,F_2:DH} = \sigma^2_A/2$, irrespective of the number of loci (ℓ) governing the trait and their additive gene effects a_i. A simple explanation of this general result is that half ($\ell/2$) of the heterozygous loci in the F_1 cross are expected to be still heterozygous in an unselected F_2 individual and, thus, segregate in the derived DH lines. However, in F_2 individuals selected either for T or M_0, the expected proportion of heterozygous loci can deviate substantially from 50% under the assumption of a finite number of loci as illustrated by Figs. 1 and 3 based on Eqs. 7 and 18. Consequently, the segregation variance for T among DH lines derived from F_2 individuals selected under these criteria also deviates from the value $\sigma^2_{A,F_2}/2 = \ell/2$ expected in the absence of selection.

Selection for T under Method 1 consistently increases M_2 at the expense of M_0 (data not shown) as expected from the strong negative correlation $\rho(T, M_0) = -0.8165$ given in Eq. 23. Unlike one might perhaps reckon from $\rho(T, M_1) = 0$ (*cf.* Eq. 24), the consequences of selection for $T \geq t$ on the expectation of M_1 depend mainly on the threshold t relative to ℓ, as illustrated by Fig. 1: when the ratio of n to ℓ is large, so that the probability $\alpha(\ell, n_1)$ is high, $E(M_1 | T \geq t)$ is considerably smaller than $\ell/2$. Consequently, $\sigma^2_{T,F_2:DH} \ll \sigma^2_{T,F_2} = E(M_1 | T \geq t)$ and the expected selection progress would be smaller for Step 2 than Step 1 if an equal number of individuals is genotyped in each step. This explains why it is rewarding to spend a greater proportion of the resources for genotyping F_2 individuals than for genotyping DH lines of the selected F_2 individual.

Under Method 2, selection against M_0 increases both M_2 and M_1 at the expense of loci fixed for non-target alleles, but the ratio $E(M_2 | M_0 \leq k) : E(M_1 | M_0 \leq k)$ remains 1:2 as expected in the absence of selection, as follows from Eqs. 18 and 19. Since the graph of $E(M_1 | M_0 \leq k)$ exceeded $\ell/2$ for small values of k by up to 33% (Fig. 3), one might conclude at a first glance to allocate more resources for genotyping DH lines compared to F_2 individuals. However, for larger values of ℓ and moderate sizes of n_1, there is an extremely low probability $\beta(\ell, n_1, k)$ that M_0 for the F_2 individual selected in Step 1 has a k value close to zero or in a range, where the $E(M_1 | M_0 \leq k)$ is considerably larger than $\ell/2$. In addition, for smaller values of ℓ and larger values of n, selection against M_0 in Step 1A is very effective, and thus, chances are high that several individuals n_{1B} remain for selection for high values of T in Step 1B, which will in turn increase M_2 at the expense of M_1. Consequently, the variance

among DH lines from the selected F_2 individual will generally be close to the value $\ell/2$. In agreement with these theoretical arguments, the mean $\overline{X}$ of M_1 in the selected F_2 individual was about 50% for all values of ℓ and n investigated in our simulations. All these factors explain why the maximum selection progress in Method 2 is obtained by genotyping about the same number of DH lines as F_2 individuals, i.e., the ratio n_1*: n is about 0.5. Moreover, these considerations also explain the flat optima for $\overline{X}$ of T_{DH} as a function of the ratio $n_1 : n$ under both Method 1 and Method 2, indicating that the choice of n_1 has within certain limits only a minor effect on the marker score T of the finally recovered DH line.

Under Method 3, selection for T acts in one step on the total genetic variance $\sigma^2_{T,F_1:DH} = \ell$ among DH lines derived from the F_1 generation. Furthermore, because T for DH lines in the absence of selection corresponds to the sum of binomially distributed variables, its probabilities are given by $\Pr(T = 2i) = \Pr(M_2 = i) = \frac{\ell!}{i!(\ell-i)!}\left(\frac{1}{2}\right)^{\ell}$. Thus, it is possible to calculate the expectation of T of the best DH line for given values of ℓ and n using results on order statistics (David 1970).

Selection progress in different steps of selection under Method 1 and Method 2

A higher selection progress for T in Step 1 under Method 1 compared with Method 2, as reflected by the comparison of the means $\overline{X}$ of $T^*_{F_2}$ (Table 1), was expected as a result of the selection protocols applied. Method 1 selects for T, whereas in Method 2, rejection of M_0 (Step 1A) precedes the selection for T (Step 1B). Thus, $E(M_0^*)$was always lower in Method 2 than Method 1. The mean $\overline{X}$ for M_0 obtained by us showed that the selection Step 1A under Method 2 was quite effective. However, if n is large and ℓ is small, there was still a sizable number n^*_{1B} (up to 8.2) of individuals left for performing selection in Step 1B. Further, the values of n_{1B} showed that there was more stringent selection in Step 1A than Step 1B. Because the variance $\mathrm{var}(T|M_0 = k)$ is rather high for small values of k, selection in Step 1B is rather efficient in improving T (i.e., increasing M_2 at the expense of M_1), as indicated by a comparison of $\overline{X}$ for $T^*_{F_2}$ before and after performing the selection in Step 1B (data not shown). Altogether, $\overline{X}$ for $T^*_{F_2}$was only 2–4% points smaller for Method 2 compared with Method 1, but $\overline{X}$ for M^*_1 was 7–18% points higher, even though n^*_1 was smaller.

The selection progress in Step 2, reflected by the difference in the means $\overline{X}$ for T^*_{DH} and $T^*_{F_2}$, depends exclusively on M^*_1 in the selected F_2 individual and $n^*_2 = n - n^*_1$. As a consequence of the selection protocol in Step 1, Method 2 had higher $\overline{X}$ of M^*_1 as well as n^*_2 compared with Method 1. Therefore, selection progress in Step 2 was considerably higher for Method 2 than Method 1 and overcompensated the lower $\overline{X}$ for $T^*_{F_2}$ after Step1.

Comparison of the three selection methods

Higher selection progress for$\overline{X}$ of T^*_{DH}under Method 2 over Method 1 and Method 3 for $\ell = 8$ and 12 accompanied by distinctly lower SD meant higher chances and lower risk for selecting a DH line with highest number of target genes achievable under given resources. The marginal differences for $\overline{X}$ of T^*_{DH} among the three methods for $\ell = 4$ in combination with sizeable n were due to the high probability of fixing all target alleles in the finally selected DH line. Even the one-step Method 3 had $\overline{X} = 99\%$ for T^*_{DH}, but accompanied with very high SD compared to Method 2. However, with larger values of ℓ, the superiority of Method 1 and Method 2 over Method 3 increased, whereas that of Method 2 over Method 1 decreased. For $\ell = 16$, only marginal superiority ($\leq$1% point in the comparison of $\overline{X}_{\mathrm{DH}}$) of Method 2 *vs.* Method 1 and no consistent trend for SD indicated that both methods perform equally good for large ℓ.

The choice of the method also depends on other factors. For example, Method 3 is the quickest as its cycle length is one generation shorter. Thus, for species with a long generation interval or in the case of an extreme time pressure to develop the final product, Method 3 would be appealing for small ℓ. Another factor is the costs of development of the DH lines. In our study, we implicitly assume that the cost of developing a DH line is the same as that of an F_2 individual. However, in many species, production of DH lines is laborious and costly. The least number of DH lines is required under Method 1 (n_2) and the highest under Method 3 (n). Hence, Method 1 may be favored, when the costs for development of DH lines are relatively high compared with genotyping.

Based on our findings, Method 2 was superior and should generally be preferred over Method 1, because the additional gain is obtained practically at no additional cost provided an efficient protocol for developing DH lines is available. Its major features are (1) it reduces M_0 more efficiently and consequently upgrades the limit of selection, (2) it has higher selection progress in Step 2 by (a) retaining a higher proportion M_1 of heterozygous loci in the selected F_2 individual and (b) allotting more resources to Step 2, as reflected by higher values of $n^*_2 = n - n^*_1$.

Our comparisons of the three methods were based on the mean $\overline{X}$ of the finally selected DH line across 250,000 simulation runs. In practice, a breeder generally conducts selection in a cross P1 × P2 only once, corresponding to one 'run'. Hence, the values of T and M_0 in the selected

individuals may deviate from their expectation, as reflected by their SD (Tables 1, 2).

One of our objectives was to study the optimization of allocation of resources for genotyping F_2 individuals versus DH lines, i.e., split n into n_1^* and n_2^*. We used an a priori approach to determine these values (i.e., before starting the breeding program), but they may be subject to a posteriori corrections after knowing the progress made in Step 1. If the best F_2 individual identified in Step 1 of Method 1 has a value of $T_{F_2}^*$ much lower than its mean $\overline{X}$ determined from the simulations, one may increase n_2 and develop and genotype more DH lines than originally planned to enhance the selection progress in Step 2 or may even repeat the process of producing and genotyping new F_2 individuals. In Method 2, if M_0 is much higher than its mean $\overline{X}$ after Step 1, the process should be repeated because increasing n_2 is of little use. Conversely, if the values of T are much higher and/or that of M_0 are much lower than the corresponding means $\overline{X}$, then one may decrease n_2. While Method 1 and Method 2 are amenable to a posteriori corrections of increasing or decreasing n_2, no such adjustment is possible in Method 3. If the target value of T is not achieved in the first attempt, the only alternative is to repeat the entire program, which leads to a delay and consequently offsets the time advantage of this Method.

Our results are in agreement with earlier simulation studies suggesting that enrichment increases the frequency of the target alleles in the enriched F_2 population as well as the population of DH lines derived from that (Howes et al. 1998; Bonnett et al. 2005; Wang et al. 2007). These authors concluded that F_2 enrichment can be effectively employed to combine favorable alleles of up to 9 to 12 unlinked QTL. For the resources considered by us, the application of Method 2 resulted in $\overline{X}_{\mathrm{DH}} \geq 99\%$ for ℓ up to 10 (data not shown), but for $\ell = 12$, $\overline{X}_{\mathrm{DH}}$ was 96.3%.

Further research needs

In our investigations, we considered unlinked target genes. If they are linked, this will complicate gene stacking. Linkage can modify (a) the minimum number of genotypes to be monitored for achieving the ultimate target genotype, (b) the optimum allocation of resources between the F_2 and DH generations, (c) the relative superiority and even the ranking of the three selection methods considered. Preponderance of coupling phase linkage will benefit Method 3 more than the other two methods, because combinations of target alleles at different loci are more likely to be preserved as a consequence of only one meiosis. Conversely, preponderance of repulsion phase among the target genes will favor Method 1 and Method 2 as the frequency of recombinant genotypes among the segregants will increase with two meioses. Nevertheless, further research is warranted to examine the effects of linked target genes. A possible starting point can be the work by Hospital et al. (1996), who developed a general algorithm to derive probability distributions of multilocus genotypes under linkage.

In gene stacking, each of the target genes to be combined in the target genotype is of equal importance, and consequently, all genes to be stacked are given equal weights. In MAS (Lande and Thompson 1990) or genomic selection (Meuwissen et al. 2001), the goal is similarly to combine a large number of markers from different parents into one target genotype, but they are assigned different weights, depending on the effect of the QTL alleles associated with them in linkage or association mapping studies. Obviously, many of our results obtained for the case of equal additive effects a_i for all genes hold in a modified manner also true for MAS and genomic selection. This applies especially to the reduction in the segregation variance among DH lines derived from F_2 individuals selected for a high marker score (F. Technow, unpublished results). Nevertheless, further studies extending our work are needed to optimize the allocation of resources and compare the three methods under the scenarios of MAS and genomic selection.

Our results have also some bearing on phenotypic selection, because selection for T under Method 1 and Method 3 corresponds to one-step and two-step phenotypic selection, respectively, when the trait has narrow-sense heritability $h^2 = 1$. For coping with $h^2 < 1$, our results in Eq. 7 on the expected segregation variance $\sigma^2_{A,F_2:DH}$ among DH lines derived from F_2 individuals selected for high values of T can be extended by considering the selection criterion $Y = T + \varepsilon$, where ε is the error term for the phenotypic score Y (A.E. Melchinger, unpublished results). As illustrated by Supplementary Fig. S1, for $h^2 = 0.5$, small values of ℓ and high selection intensities corresponding to values of $\alpha = P(Y > y) < 0.01$, $E(M_1 | Y > y)$ is still considerably smaller than its expectation $E(M_1) = \ell/2$ in the absence of selection. Thus, under extremely strong phenotypic selection too, $\sigma^2_{A,F_2:DH}$ for individuals selected for high values of Y is smaller than the additive variance in the F_2 generation (σ^2_{A,F_2}). Consequently, the assumption of equal variances $\sigma^2_{A,F_2} = \sigma^2_{A,F_2:DH}$ commonly underlying studies on two-stage breeding methods (cf. Longin et al. 2007; Wegenast et al. 2010) does not hold true for DH lines of selected F_2 individuals under the assumption a finite number of loci. Further research is warranted to investigate this effect of selection in early generations on the segregation variance among DH lines or recombinant inbred lines derived from selected individuals in greater detail as it has important implications on the

optimum allocation of resources and also the total selection progress accumulated over both selection stages.

References

Bernardo R (2008) Molecular markers and selection for complex traits in plants: learning from the last 20 years. Crop Sci 48:1649–1664

Bernardo R (2009) Should maize doubled haploids be induced among F_1 or F_2 plants? Theor Appl Genet 119:255–262

Bonnett DG, Rebetzke GJ, Spielmeyer W (2005) Strategies for efficient implementation of molecular markers in wheat breeding. Mol Breeding 15:75–85

Darrah LL, Zuber MS (1986) 1985 United States farm maize germplasm base and commercial breeding strategies. Crop Sci 26:1109–1113

David HA (1970) Order statistics. Wiley, NY

Fehr WR (1987) Principles of cultivar development, vol 1. Theory and technique. Macmillan Publishing Co, NY, p 136

Heffner EL, Lorenz AJ, Jannink J-L, Sorrells ME (2010) Plant breeding with genomic selection: gain per unit time and cost. Crop Sci 50:1681–1690

Hospital F, Dillmann C, Melchinger AE (1996) A general algorithm to compute multilocus genotype frequencies under various mating systems. CABIOS Oxford University Press 12:455–462

Howes NK, Woods SM, Townley-Smith TF (1998) Simulations and practical problems of applying multiple marker assisted selection and doubled haploids to wheat breeding programs. Euphytica 100:225–230

Kearsey MJ, Pooni HS (1996) The genetical analysis of quantitative traits. Chapman and Hall, London

Koide Y, Kawasaki A, Telebanco-Yanoria MJ, Hairmansis A, Nguyet NTM, Bigirimana J, Fujita D, Kobayashi N, Fukuta Y (2010) Development of pyramided lines with two resistance genes, *Pish* and *Pib*, for blast disease (*Magnaporthe oryzae* B. Couch) in rice (*Oryza sativa* L.). Plant Breeding 670–675

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

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

Longin CFH, Utz HF, Reif JC, Wegenast T, Schipprack W, Melchinger AE (2007) Hybrid maize breeding with doubled haploids: III. Efficiency of early testing prior to doubled haploid production in two-stage selection for testcross performance. Theor Appl Genet 115:519–527

Lynch M, Walsh B (1997) Genetics and analysis of quantitative traits. Sinauer Assoc, Sunderland

Mago R, Lawrence GJ, Ellis JG (2011) The application of DNA marker and doubled-haploid technology for stacking multiple stem rust resistance genes in wheat. Mol Breeding 27:329–335

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829

Mikel MA, Dudley JW (2006) Evolution of North American dent corn from public to proprietary germplasm. Crop Sci 46:1193–1205

Perumalsamy S, Bharani M, Sudha M, Nagarajan P, Arul L, Saraswathi R, Balasubramanian P, Ramalingam J (2010) Functional marker-assisted selection for bacterial leaf blight resistance genes in rice (*Oryza sativa* L.). Plant Breeding 129: 400–406

R Development Core Team (2010) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/

Röber FK, Gordillo GA, Geiger HH (2005) In vivo haploid induction in maize—performance of new inducers and significance of doubled haploid lines in hybrid breeding. Maydica 50:275–283

Schnell FW (1982) A synoptic study of the methods and categories of plant breeding. Z Pflanzenzüchtung 89:1–18

Tuvesson S, Dayteg C, Hagberg P, Manninen O, Tanhuanpää P, Tenhola-Roininen T, Kiviharju E, Weyen J, Förster J, Schondelmaier J, Lafferty J, Marn M, Fleck A (2007) Molecular markers and doubled haploids in European plant breeding programmes. Euphytica 158:305–312

Varshney RK, Nayak SN, May GD, Jackson SA (2009) Next generation sequencing technologies and their implications for crop genetics and breeding. Trends Biotechnol 27:522–530

Wang J, Chapman SC, Bonnett DG, Rebetzke GJ, Crouch J (2007) Application of population genetic theory and simulation models to efficiently pyramid multiple genes via marker-assisted selection. Crop Sci 47:582–590

Wędzony M, Forster BP, Żur I, Golemiec E, Szechyńska-Hebda M, Dubas E, Gotębiowska G (2009) Progress in doubled haploid technology in higher plants. In: Touraev A, Forster BP, Jain SM (eds) Advances in haploid production in higher plants. Springer Science + Business Media BV, Heidelberg, pp 1–33

Wegenast T, Utz HF, Longin CFH, Maurer HP, Dhillon BS, Melchinger AE (2010) Hybrid maize breeding with doubled haploids: V. Comparison of selection strategies involving variable sizes of crosses and S1 families in two-stage selection for testcross performance. Theor Appl Genet 120:699–708

Wu Y (2010) Molecular breeding. CAB International, Oxfordshire