Theor Appl Genet (2013) 126:2299–2312
DOI 10.1007/s00122-013-2136-3

ORIGINAL PAPER

The impact of recombination on short-term selection gain in plant breeding experiments

Benjamin McClosky · Steven D. Tanksley

Received: 23 May 2012 / Accepted: 21 May 2013 / Published online: 13 June 2013

Abstract Recombination is a requirement for response to selection, but researchers still debate whether increasing recombination beyond normal levels will result in significant gains in short-term selection. We tested this hypothesis, in the context of plant breeding, through a series of simulation experiments comparing short-term selection response ($\leq$20 cycles) between populations with normal levels of recombination and similar populations with unconstrained recombination (i.e., free recombination). We considered additive and epistatic models and examined a wide range of values for key design variables: selection cycles, QTL number, heritability, linkage phase, selection intensity and population size. With few exceptions, going from normal to unconstrained levels of recombination produced only modest gains in response to selection ($\approx$11 % on average). We then asked how breeders might capture some of this theoretical gain by increasing recombination through either (1) extra rounds of mating or (2) selection of highly recombinant individuals via use of molecular markers/maps. All methods tested captured less than half of the potential gain, but our analysis indicates that the most effective method is to select for increased recombination and the trait simultaneously. This recommendation is based on evidence of a favorable interaction between trait selection and the impact of recombination on selection gains. Finally, we examined the relative contributions of the two components of meiotic recombination, chromosome assortment and crossing over, to short-term selection gain. Depending primarily on the presence of trait selection pressure, chromosome assortment alone accounted for 40–75 % of gain in response to short-term selection.

Communicated by H. Becker.

B. McClosky (✉) · S. D. Tanksley
Nature Source Genetics, 33 Thornwood Drive,
Suite 300, Ithaca, NY 14850, USA
e-mail: bmcclosky@naturesourcegenetics.com

Introduction

Genetic recombination originates in meiosis as a combined result of chromosome assortment and crossing over. In plant breeding experiments, it is possible to create populations with increased (or decreased) recombination through selective mating schemes. For example, the use of doubled haploidy will result in populations with reduced recombination in comparison to similar populations generated through single seed descent (Snape 1976; Riggs and Snape 1977) or intercrossing (Darvasi and Soller 1995; Lee et al. 2002). Moreover, with the advent of high density molecular linkage maps, researchers can apply selection pressure by identifying subsets of individuals with altered levels of recombination (Xu et al. 2005; Jannink 2005; Smith et al. 2008; Tanksley et al. 1989).

While the breeder can thus modulate recombination in a population, the degree to which such action impacts selection response is still debated. In most instances, researchers have proffered that maximizing recombination during population development will increase selection gains (Bjornstad 1987; Holland 2004; Simmonds 1979; Wijnker and de Jong 2008). This notion is supported by an evidence of a correlation between recombination and selection response in mice (Gorlov et al. 1992), *Saccharomyces cerevisiae* (Wolf et al. 1987), and *Drosophila melanogaster* (Rodell et al. 2004; Presgraves 2005). However, evidence against such correlation exists as well (Bourguet et al. 2003). In plant breeding studies, despite variable results from simulations (Jannink and Abadie

1998), empirical comparisons between doubled haploid generated populations (lower recombination) and similar populations generated by single seed descent (higher recombination) have found little evidence of differences in genetic variances (Bordes et al. 2007; Murigneux et al. 1993; Park et al. 1976; Choo et al. 1982; Courtois 1993). This observed similarity in genetic variances is unexpected if recombination in fact influences response to selection. Finally, comparisons between F_1- and F_2-derived DH lines support conflicting conclusions about the impact of recombination on selection (Iyamabo and Hayes 1995; Bernardo 2009; Charmet and Branlard 1985).

From a theoretical perspective, selection can be viewed as the differential reproduction of the products of recombination (Simmonds 1979), and researchers have hypothesized several evolutionary advantages of recombination (Fisher 1930; Felsenstein 1974, 1988; Muller 1932). In particular, recombination alleviates two phenomena that constrain selection gains: Muller's ratchet (Muller 1964) and the Hill–Robertson effect (Felsenstein 1974; Hill and Robertson 1966). However, these theoretical arguments are made in an evolutionary context where selection acts over many generations (long-term selection) in relatively large populations with low levels of linkage disequilibrium. In contrast, plant breeding experiments typically span abbreviated timeframes, in most instances less than ten cycles (short-term selection) (Simmonds 1979). Furthermore, plant breeding populations often derive from biparental matings (e.g., F_3, F_4, DH). These populations have high levels of linkage disequilibrium and contain only a small proportion of the genetic variation present in the larger gene pool (Snape 1976; Bordes et al. 2007; Bernardo 2009). While there have been many assertions as to the benefits of increasing recombination for short-term selection gain, to our knowledge there is no conclusive evidence supporting this hypothesis in the context of plant breeding. Without such evidence, one has to question the merits of spending time and effort to increase recombination in breeding populations as a means to increase short-term selection gains.

In this paper, we report results from a simulation study that analyzed the impact of recombination on selection response in experimental populations. Throughout this paper, the term selection refers to short-term artificial selection (≤20 cycles). We primarily examine populations derived from a biparental inbred cross, but we provide evidence that our conclusions generalize to populations derived from multiple inbred lines. We focus on applications where the breeder selects for a genetic outlier, i.e., a *transgressive segregant*. For example, plant breeders often develop new hybrid varieties from the best-performing individual in a segregating population (Zhong and Jannink 2007; Jinks and Pooni 1976). Similarly, dairy cattle breeders select for a genetically superior bull that can be used for artificial insemination (Funk 2006). By focusing on the distributional extremes, our analysis differs from the classical approach, which measures gain by changes in population means (Falconer and Mackay 1996; Bulmer 1980). In this context, we addressed the following questions faced by researchers as they decide whether to allocate resources for increasing recombination in a breeding population:

1. What are the limits of selection gains attributable to recombination?
2. Can breeders approach these limits in practice?
3. What is the relative impact of the two sources of recombination: chromosome assortment and crossing over (junctions[1])?

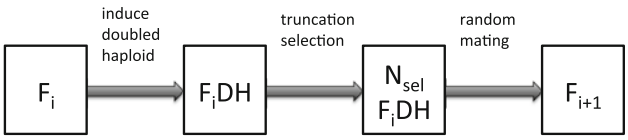

Fig. 1 The recurrent selection breeding scheme (1 cycle)

Materials and methods

Researchers typically breed for multiple cycles to accumulate recombination and often simultaneously employ recurrent selection on the trait of interest. In this context, recombination interacts with selection and additional factors, such as genetic drift, within a complex system (Hill and Robertson 1966). This complexity certainly confounds the relationship between recombination and selection response. Nevertheless, we can simulate such breeding programs and analyze the efficacy of using recombination as a breeding objective to increase selection gains. Figure 1 shows a diagram of the recurrent selection scheme used in our simulations.

Let P_1 and P_2 be the inbred parents used for population development, N be the population size, and N_Q be the number of quantitative trait loci (QTL). To advance from the F_i to the F_{i+1}, we ranked the F_i-derived doubled haploid (DH) population according to phenotype and randomly mated (including selfs) the top N_{sel} F_i DH individuals. Let $I = N_{sel}/N$ measure the selection intensity, where smaller I corresponds to stronger selection pressure. We performed truncation selection in DH populations to separate the effects of recombination from the confounding effects of

[1] Fisher coined the term junction to refer to chromosomal regions that trace back to a prior meiotic crossover and transition from the DNA of one parent to the other (Fisher 1954)

homozygosity. Most of our experiments used the F_1 progeny of a biparental inbred cross as the base population (i.e., the first round of selection was in the F_1 DH), but we also considered the performance of several alternative base populations.

Simulating the genome

In each simulation, we randomly placed QTLs at N_Q locations throughout the genome. The range of N_Q (5–100) was based on empirical estimates of the number of QTLs segregating in experimental plant populations (Buckler et al. 2009; Laurie et al. 2004; Brachi et al. 2010; Otto and Jones 2000) (Table 1). Since experimental populations capture differences between a small number of individuals, the number of segregating QTLs will represent only a small fraction of the total QTLs for a trait across the entire germplasm.

The magnitudes of QTL substitution effects $(a_1, \ldots, a_{N_Q})$ were drawn from the gamma distribution with shape parameter 1.45 and a scale parameter determined by heritability, a distribution defined to be consistent with empirical evidence in maize (Buckler et al. 2009). To control the level of coupling and repulsion, an independent Bernoulli (p) trial determined the direction (i.e., sign) of each QTL effect. More precisely, p controlled the probability that the P_2 allele had a negative effect relative to the P_1 allele at a given QTL. The QTLs were in complete coupling phase when $p \in \{0.0, 1.0\}$, and the level of repulsion increased as p approached 0.5.

To model linkage, we simulated crossovers according to a Poisson process using cM lengths derived from Haldane's mapping function. We used the standard maize map as a base case (1,350 cM; 10 chromosomes) (McMullen et al. 2009). To simulate the limiting case of free recombination, we assumed individual i had independently segregating QTL genotypes $g_{i1}, \ldots, g_{iN_Q}$. Let $bv_i = \sum_{j=1}^{N_Q} a_j g_{ij}$ denote individual i's breeding value.

Table 1 Default distributions for simulation parameters

Parameter	Symbol	Range
Number of QTL	N_Q	5–100
Heritability	h^2	0.2–0.8
Phase	p	0.4–0.6
Selection intensity	I	0.01–0.1
Population size	N	100–1,000

Parameters were randomly sampled within the specified range for each iteration of the simulation

Simulating phenotype

Unless otherwise specified, we simulated phenotype y according to the standard additive model: $y_i = bv_i + r_i$, where $r_i \sim N(0, \sigma_r^2)$. However, we also considered two classes of nonlinear (epistatic) models. The first class transformed the additive breeding value to determine the genetic component of phenotype. Given a transformation $f : \mathbb{R} \mapsto \mathbb{R}$, we defined phenotype according to the model: $y_i = f(bv_i) + r_i$. For example, the sigmoid transformation in Fig. 2 modeled the epistatic consequences of duplicate and complementary gene dosage (Kearsey and Sturley 1984). The parabolic transformation in Fig. 2 modeled biological systems where marginal increases in additive breeding value transition from being beneficial to detrimental. For instance, increasing the breeding value for plant height is expected to increase phenotype until the plant becomes too tall for the root system to support. Breeding values under the parabolic transformation can have intermediate optima and can, thus, model traits under stabilizing selection (Falconer and Mackay 1996).

The second class of nonlinear model explicitly simulated pairwise epistatic interactions between alleles. Specifically, the number of interacting pairs was drawn uniformly between 1 and $\binom{N_Q}{2,}$ and the epistatic effects of the randomly chosen pairs were drawn from the same gamma distribution as the additive effects. As with the additive effects, the sign of each interaction effect was randomly set according to an independent Bernoulli(p) trial. In other words, p controlled the probability that a given interacting pair had a negative effect when both loci had the P_2 allele.

Regardless of the model used to determine breeding value, we introduced a normally distributed environmental noise term drawn from $N(0, \sigma_r^2)$. Let $\sigma_{F_2}^2$ be the variance of breeding values in an F_2. Let $h^2 = \sigma_{F_2}^2/\sigma_y^2$ denote heritability, and note that heritability always refers to F_2 heritability. The range of h^2 (0.2–0.8) was chosen to represent values typically encountered in plant breeding experiments for quantitative traits (e.g., yield) (Hallauer and Miranda 1988; Albrecht et al. 2011; Bernardo 1996) (Table 1).

Assigning model parameters

Several factors determine how we answer the questions posed in the introduction. For example, recombination is purported to increase selection response by releasing genetic variance previously locked in repulsion phase (Riggs and Snape 1977; Iyamabo and Hayes 1995), but recombination can also reduce selection gains when QTL are in coupling phase and the most desirable individuals resemble a parental genotype (Riley et al. 1981). We

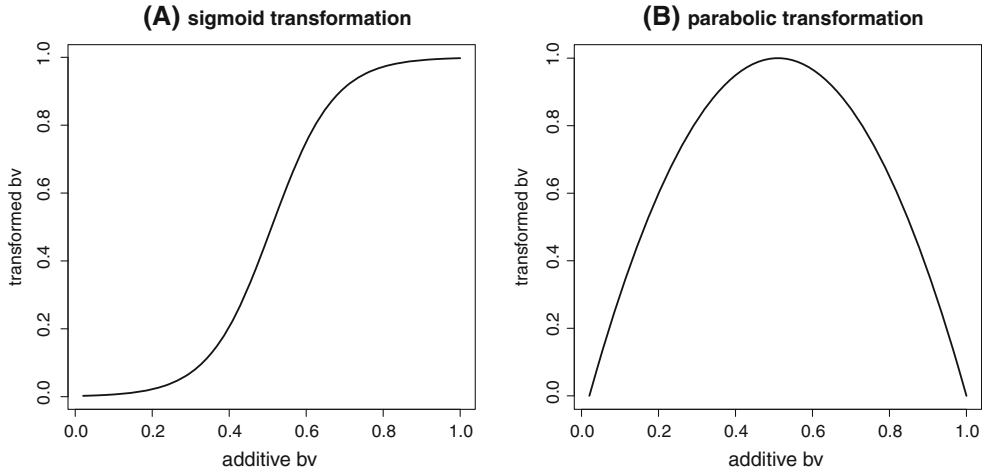

Fig. 2 Examples of nonlinear relationships between additive breeding value and the transformed breeding value used to determine phenotype

accounted for sensitivity to experimental design by drawing parameter values from distributions, and we quantified the marginal effect of each simulation parameter across a broad range of scenarios. Unless otherwise specified, all parameters were randomly drawn from the distributions shown in Table 1.

The nonlinear models required additional parameterization, in addition to those in Table 1. We varied the slope of the sigmoid transformation, $(1.0 + e^{-c_1 x})^{-1}$, by drawing c_1 uniformly from 0.1 to 1.0. For the parabolic transformation, we randomly sampled c_2^1, c_2^2 from the interval 0.0–1.0 and mapped the additive breeding values (via scaling) into the interval bounded by c_2^1 and c_2^2. This approach allowed us to use subsets of the parabolic transformation to model a broad range of nonlinearity.

Performance metrics

Given a population P, we measured selection gains as the difference between the max population breeding value and the max parental breeding value $(\max_{i \in P}\{bv_i\} - \max\{bv_{P_1}, bv_{P_2}\})$. The max inbred parent represents selection response under no recombination; therefore, we interpreted this difference as a measure of gain attributable to recombination. The focus on the max individual reflects our interest in identifying genetic outliers, which as mentioned above is the objective of many breeding programs. This measure of transgressive gain should clearly be positive to justify the expense of breeding. The distribution of population breeding values changed across simulations due to random model parameters, therefore we scaled all gains by the standard deviation of the corresponding F_2 breeding values. More precisely, given a QTL model, we simulated an F_2 using the standard map and calculated σ_{F_2} for use as a scaling factor: $(\max_{i \in P}\{bv_i\} - \max\{bv_{P_1}, bv_{P_2}\})/\sigma_{F_2}$.

We present our results in terms of gain ratios, which express the relative performance of selection under different recombination regimes. The gain ratio (normal recombination gains)/(free recombination gains) quantifies the performance gap between populations under normal and free recombination and, thus, establishes the limiting impact of maximizing recombination beyond normal rates. The gain ratio measures the potential benefits of increasing recombination because a small gain ratio indicates that increasing recombination produced gains. After reporting the gain ratio across several cycles of recurrent selection, we then summarized performance as the gain ratio observed after 5, 10 , and 20 cycles of recurrent selection. This allowed us to plot the two-dimensional relationship between the gain ratio and each simulation parameter. Whenever a parameter is not specified by the x-axis, it was drawn from the appropriate distribution in Table 1. The x-axis values used to determine each parameter span the same range as the Table 1 distributions.

Results

Relative gains under normal and free recombination

Figure 3a reports the average gain across 20 cycles of recurrent selection under normal and free recombination. Free recombination produced rapid gains that tailed off near the 10th cycle, whereas normal recombination gains grew at a slower pace and converged to a lower limit at a later cycle. It is generally accepted that loss of genetic variance causes selection plateaus, which were observed under both recombination regimes, as recurrent selection progresses (Simmonds 1979; Falconer and Mackay 1996; Bulmer 1980).

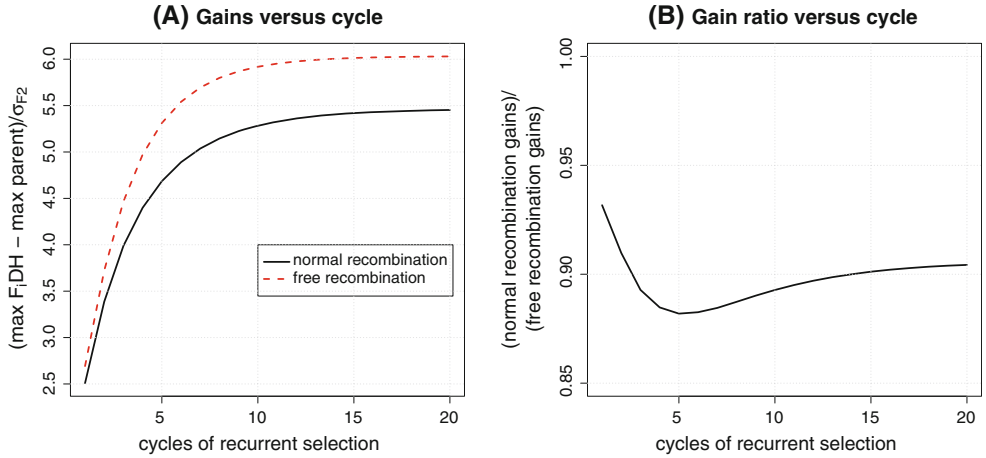

Fig. 3 **a** Recurrent selection gains under normal and free recombination rates. **b** The ratio of normal recombination gains to free recombination gains. All parameters were drawn from Table 1 distributions. Based on 100,000 simulations

Figure 3b plots the proportion of free recombination gains realized under normal recombination. Viewing the gain ratio across cycles allows us to compare the rates of selection gain. For example, prior to the 5th cycle, the negative slope indicates that free recombination produced more gain per cycle than normal recombination. After the 5th cycle, the positive slope indicates that normal recombination realized larger marginal gains. We speculate that free recombination, while producing more gains overall, quickly exhausted population genetic variance. In contrast, normal recombination could still produce genetic variance at later cycles by breaking tightly linked repulsion QTL.

The gain ratio generally stayed near 0.9 across all cycles, indicating that normal recombination was sufficient to capture 90 % of free recombination gains (Fig. 3b). Researchers have demonstrated the negative impact of linkage disequilibrium (LD) on recurrent selection gains (Hill and Robertson 1966; Falconer and Mackay 1996), and free recombination was expected to alleviate some of this constraining effect. Surprisingly, the population under normal recombination managed to produce a large proportion of free recombination gains. Under the range of parameters specified in Table 1, Fig. 3b indicates that raising recombination rates to their theoretical limit increased gains by about 11 % on average.

To understand why this performance gap was not larger, note that free recombination effectively maximized the recombination frequency between all QTL pairs. However, it is well known that LD in finite populations, whether induced by selection or otherwise, can persist for several generations under maximum recombination frequencies (Falconer and Mackay 1996; Lynch and Walsh 1998). Figure 3b could suggest that, despite free recombination, enough LD persisted to interfere with selection.

Figure 3 presents average performance across a broad range of scenarios, therefore we now consider the marginal impact of each simulation parameter in Table 1. Figure 4 illustrates the relatively minor influence of both p (phase) and h^2 (heritability) on the gain ratio. We know that QTL phase can have a major effect on the value of recombination, especially in the extreme case of full coupling (Riggs and Snape 1977; Iyamabo and Hayes 1995; Riley et al. 1981). However, since the gain ratio stayed below 1.0, Fig. 4a shows that recombination was consistently beneficial over the phase range considered. This evidence supports the notion that recombination increases selection gains when QTL appear in repulsion, which agrees with previous simulation evidence (Riggs and Snape 1977). Also note that the phase parameter p had a similar impact on the gain ratio for each number of recurrent selection cycles.

Heritability is known to have a dramatic impact on the performance of phenotypic truncation selection (Simmonds 1979; Falconer and Mackay 1996; Lynch and Walsh 1998; Allard 1999). Nonetheless, Fig. 4b shows that h^2 had a minor effect on the gain ratio. This suggests that h^2 interacted similarly with both recombination regimes. The lack of relationship between h^2 and the value of recombination for selection has been observed in previous simulation experiments (Bernardo 2009). Although it remained small, the interaction between h^2 and the gain ratio marginally increased with the number of selection cycles.

Figure 5 considers the influence of I (selection intensity), N (population size), and N_Q (number of QTL). The steeper slopes shown in these plots demonstrate the much stronger impact of these parameters on the relationship between recombination and selection gains.

Figure 5a shows that more intense selection pressure (smaller I) increased the value of recombination for selection response. This result could suggest that recombination mitigated the Bulmer effect, which predicts that strong selection causes genetic variance to decay during

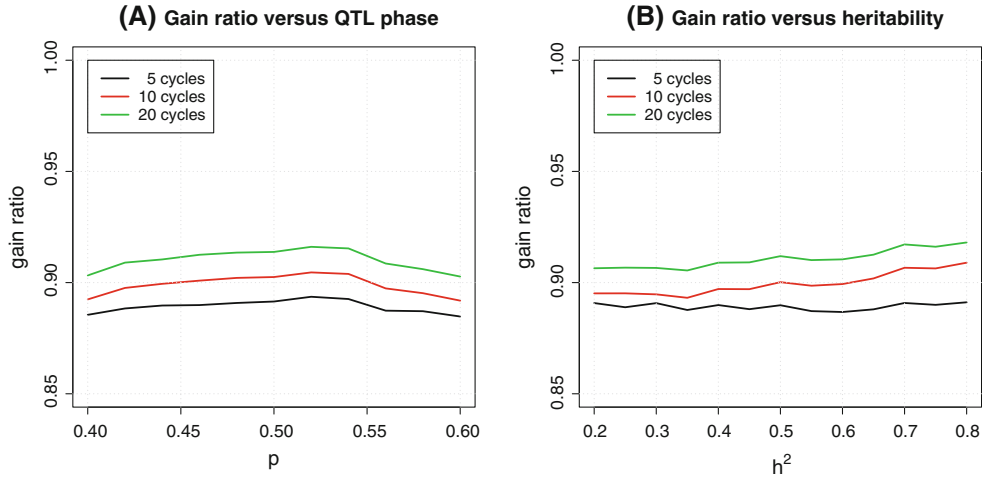

Fig. 4 **a** Impact of QTL phase on the gain ratio after 5, 10, and 20 cycles of recurrent selection. **b** Impact of heritability on the gain ratio after 5, 10, and 20 cycles of recurrent selection. In each figure, all parameters not specified by the *x*-axis were drawn from Table 1 distributions. Based on 100,000 simulations

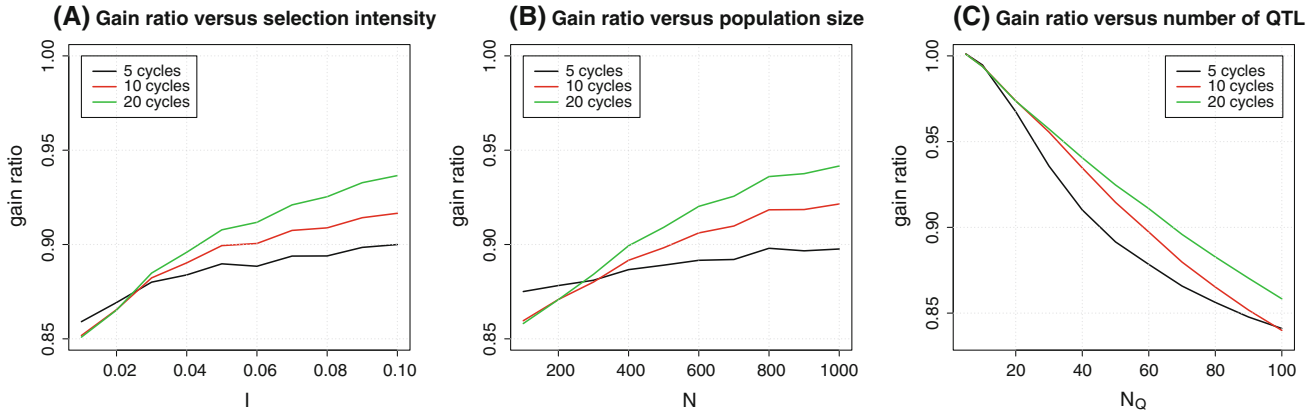

Fig. 5 **a** Impact of selection intensity on the gain ratio after 5, 10, and 20 cycles of recurrent selection. **b** Impact of population size on the gain ratio after 5, 10, and 20 cycles of recurrent selection. **c** Impact of number of QTL on the gain ratio after 5, 10, and 20 cycles of recurrent selection. In each figure, all parameters not specified by the *x*-axis were drawn from Table 1 distributions. Based on 100,000 simulations

recurrent selection (Bulmer 1980). Figure 5a is also consistent with assertions that intense selection pressure induces LD among QTLs (Lynch and Walsh 1998; Hill and Robertson 1968), and thus interferes with selection response (Hill and Robertson 1966). Figure 5a also demonstrates that the interaction between selection intensity (I) and the gain ratio was sensitive to the number of selection cycles. In particular, the slope of the gain ratio curve became steeper as the number of cycles increased.

Figure 5b illustrates the relationship between the gain ratio and population size. The benefits of increased recombination diminished as population size increased. The trajectory of this curve is consistent with the hypothesis that recombination benefits selection in finite populations, but not in infinite populations (Felsenstein 1974). Further, this plot is strikingly similar to Fig. 5a, a relationship likely due to the fact that $N_{sel} = I \times N$ is a function of both I and N. As with selection intensity, we explain the positive slope by noting that small population size can constrain selection by inducing LD among QTLs (Hill and Robertson, 1966; Lynch and Walsh, 1998; Hill and Robertson 1968). In addition, we observe that the interaction between population size and the gain ratio was sensitive to the number of selection cycles.

Figure 5c shows that N_Q had the greatest impact on the gain ratio. As N_Q increased, recombination clearly became a limiting factor in response to selection, consistent with other reports (Bernardo 2009). Nonetheless, even with 100 QTLs, normal recombination captured 85 % of free recombination gains. As with population size and selection intensity, the number of selection cycles impacted the relationship between N_Q and the gain ratio. Specifically,

the interaction between N_Q and the gain ratio became stronger as the number of selection cycles decreased. To test even larger values of N_Q, we also simulated $N_Q = 1{,}000$ and observed a gain ratio of approximately 0.6 after 20 cycles of recurrent selection. Note that the dense QTL models (large N_Q) experienced more LD than the sparse models (small N_Q). Therefore, we hypothesize that changes in LD among QTL explain all the relationships observed in Fig. 5. In particular, the value of recombination, as measured by the performance gap between normal and free recombination, increases with LD.

Methods for approaching free recombination gains

In the previous section, we saw that recurrent selection under normal recombination produced about 90 % of gains seen under free recombination (Fig. 3). We considered three candidate methods for closing this performance gap. The first method accumulated recombination by breeding for several generations (Snape 1976; Riggs and Snape 1977; Darvasi and Soller 1995; Lee et al. 2002) prior to initiating recurrent selection. The second method used dense molecular linkage maps to detect and select for highly recombinant individuals followed by trait selection (Xu et al. 2005; Jannink 2005). The third method applied selection pressure simultaneously to increase recombination and improve the trait. As in the previous simulations, all selection was made in the DH populations (Fig. 1).

Breeding for increased recombination

Table 2 quantifies the impact of initiating our recurrent selection scheme (Fig. 1) on various base populations with different levels of recombination. The $F_2 - F_5$ populations were derived by single seed descent. The $RM_1 - RM_3$ populations were derived by successive generations of random mating, where we constructed RM_1 by randomly mating the F_2. This approach required 1–4 additional cycles of breeding to develop the base populations prior to initiating the recurrent selection scheme. After initiating the first round of selection, the remaining cycles follow the same scheme as Fig. 1 (i.e., no additional generations of breeding).

Random mating outperformed single seed descent, presumably because homozygosity due to selfing inhibits the formation of new junctions. Further, breeding prior to recurrent selection had more impact when the number of cycles was small. However, all base populations considered produced only marginal gains in response to selection and failed to approach free recombination gains (Table 2). Given the resource and time expenditure required to breed for multiple cycles, this evidence raises doubts about the efficacy of breeding for recombination prior to initiation of recurrent selection to improve selection gains.

This prediction agrees with empirical experiments that found similar genetic variances in SSD and DH populations (Bordes et al. 2007; Murigneux et al. 1993; Park et al. 1976; Choo et al. 1982; Courtois 1993). This result also agrees with assertions that F_1DH and F_2DH are expected to exhibit similar trait distributions (Iyamabo and Hayes 1995; Charmet and Branlard 1985). Note that these experiments all have the feature that breeding for recombination was not coupled with selection on the trait, a topic covered in "Combining trait selection and marker-based recombination selection".

Marker-based selection for increased recombination

Table 3 quantifies the impact of applying selection pressure on the number of junctions during recurrent selection. Specifically, we genotyped multiple ($C \geq 1$) DH progeny from each F_i individual (see Fig. 6) and advanced the individual with the most junctions as a candidate for truncation selection. Note that this scheme performed selection for junctions and the trait sequentially, as opposed to applying simultaneous selection pressure (see "Combining trait selection and marker-based recombination selection"). In practice, this design would require inducing and genotyping ($N \times C$) DH individuals in each cycle of breeding. At the lower range of our parameter space, $N = 100$ and $C = 5$, this method would require genotyping 500 DH individuals per cycle. Although we observed a substantial increase in junctions compared to the unoptimized scheme, the gain ratio stayed below 0.93 after 20 cycles of selection. In addition, the gains after 5 and 10 cycles of recurrent selection (not shown) did not offer any

Table 2 Base population is the population where the recurrent selection scheme (see Fig. 1) was initiated

Base population	F_1	F_2	F_3	F_4	F_5	RM_1	RM_2	RM_3
5 cycle gains ratio	0.882	0.891	0.905	0.906	0.908	0.911	0.917	0.924
10 cycle gains ratio	0.893	0.905	0.908	0.911	0.914	0.913	0.920	0.927
20 cycle gains ratio	904	0.916	0.917	0.922	0.923	0.924	0.930	0.932

F_k denotes single seed descent starting from the F_1. RM_k denotes random mating, where we constructed RM_1 by randomly mating the F_2. The table reports the gains ratio realized after varying number of recurrent selection cycles. Based on 100,000 simulations

Table 3 *C* refers to the number of DH genotyped progeny from each F_i individual

C	1	5	50	100	200
Gains ratio	0.90	0.92	0.92	0.92	0.92
Junctions per individual	132	345	560	610	653

Note that $C = 1$ corresponds to the control experiment with no selection for recombination. This method samples $(N \times C)$ DH individuals. The gains ratio reports the proportion of free recombination gains after 20 cycles of recurrent selection. Based on 10,000 simulations

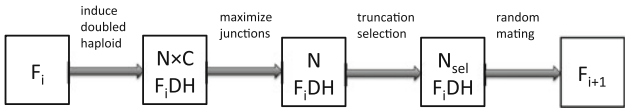

Fig. 6 Maximizing junctions during the recurrent selection breeding scheme. *C* refers to the number of genotyped DH progeny from each F_i individual

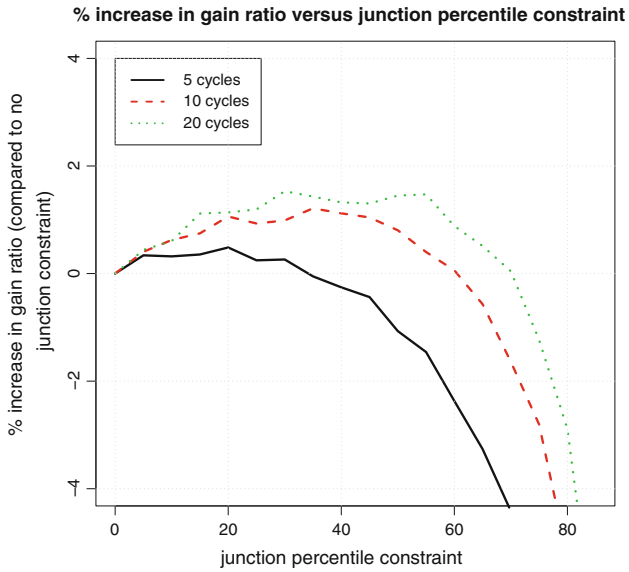

Fig. 7 Percent increase in gain ratio (compared to unoptimized scheme) after several cycles of recurrent selection in subpopulations defined by a lower bound on junctions. In each cycle, only individuals with a total number of junctions in the *x*th percentile were candidates for selection. $x = 0$ corresponds to the results shown in Fig. 3 (the unoptimized scheme). Based on 150,000 simulations

significant improvement over the corresponding gains in Table 2. Therefore, we conclude that this scheme fails to improve over the scheme in "Breeding for increased recombination".

While researchers have considered using marker-based selection to modulate recombination (Xu et al. 2005; Jannink 2005; Smith et al. 2008), we are unaware of any previous work that evaluates the consequences of using markers to alter recombination for selection response. Table 3 provides strong evidence that such an approach can fail to produce meaningful selection gains. This observation is significant considering the expense of inducing and genotyping large DH populations every cycle. We again stress that maximization of the trait and recombination was done in sequence rather than concurrently.

Combining trait selection and marker-based recombination selection

Our third method selected for junctions and phenotype simultaneously. In each round of the breeding scheme (see Fig. 1), we performed phenotypic selection on F_iDH subpopulations defined by a lower bound constraint on the number of junctions. In other words, individuals with very few junctions were eliminated as candidates for phenotypic selection. Given a F_iDH population, we constrained based on the observed percentiles of the population junction distribution. Fig. 7 plots the percent increase in the gain ratio after 5, 10, and 20 cycles of recurrent selection as a function of the junction percentile constraint. A large value on the *x*-axis indicates a strong constraint on junctions, and trait selection becomes unconstrained as the *x*-axis approaches zero. At $x = 0$ the curves coincide with the results shown in Fig. 3.

This experiment required the same population sizes and number of cycles as the unoptimized scheme (Fig. 1), but we did incur the additional cost of genotyping the F_iDH in order to detect junctions. Furthermore, this experiment required less breeding than the first method (Breeding for increased recombination) and smaller populations than the second method (Marker-based selection for increased recombination).

The simultaneous optimization of phenotype and junctions became more effective as the number of cycles increased. With ten or more cycles, the increase in gain ratio was comparable to the previous two methods (Breeding for increased recombination, Marker-based selection for increased recombination). For example, the gain ratio after 20 selection cycles peaked near 0.92 (not shown), thereby approaching gains observed in Tables 2, 3. This approach improved performance without the use of additional resources, and thus clearly demonstrates the efficiency gained by combining the trait and junction optimizations. Moreover, it is possible to push the gain ratio even higher by allowing the percentile constraint to change according to the cycle of selection (not shown), but all schemes considered still failed to approach free recombination gains.

Overall, these experiments indicate that separately applying selection pressure on recombination and phenotype can be an inefficient technique for increasing selection

gains. In plant breeding experiments comparing SSD and DH populations (Bordes et al. 2007; Murigneux et al. 1993; Park et al. 1976; Choo et al. 1982; Courtois 1993), breeders did not apply recurrent selection pressure during SSD development, and recombination consequently failed to produce significant differences in the genetic variances of these populations. In empirical experiments—performed on mice, *S. cerevisiae*, and *D. melanogaster*—that incorporated recurrent selection pressure, several researchers have concluded that recombination impacts selection response (Gorlov et al. 1992; Wolf et al. 1987; Rodell et al. 2004; Presgraves 2005). Consistent with these experiments, the evidence in Fig. 7 suggests that selection pressure on the trait interacts with selection pressure on junctions. We expand on this concept in the following section.

Relative impact of assortment and crossing over

Two mechanisms, chromosome assortment and crossing over (junctions), account for the overall recombination output from meiosis. We ran a series of simulations to determine the relative contribution of each recombination component to trait selection response. As described earlier (see "Materials and methods"), we measured selection response in each cycle of the standard scheme (see Fig. 1) as the improvement of the maximum breeding value individual beyond the max parent.

In this section, we simulated and compared two population types. Meiosis in the first population type (assortment only) did not involve crossing over. Meiosis in the second population type (normal recombination) involved the normal rate of assortment and crossing over. Across 20 cycles of breeding, Fig. 8 compares the max breeding values in the assortment only population against the max breeding values in the normal recombination population. We conducted this experiment both without recurrent trait selection pressure (see Fig. 8a) and with recurrent trait selection pressure (see Fig. 8b). This pressure was controlled by the parameter $I = N_{sel}/N$. Note that setting $I = 1.0$ (i.e. $N_{sel} = N$) results in random mating of the entire F_iDH, rather than a subset defined by truncation selection (see Fig. 1).

As the curves approach 1.0, assortment produced genetic outliers of the same magnitude as normal recombination (i.e., crossing over did little to impact the occurrence of outliers). As the curves approach 0.0, the impact of crossing over increased, and assortment alone became relatively less efficient at producing outliers. Figure 8 clearly shows an interaction between recurrent selection pressure and the relative contribution of the two components of recombination.

More precisely, under random mating (i.e. $I = 1.0$), assortment alone produced nearly 75 % of the gains observed with normal recombination (see Fig. 8a). However, recurrent selection pressure amplified the value of crossing over, and assortment alone achieved less than 40% of normal recombination gains (see Fig. 8b). This evidence is consistent with the view that recurrent selection pressure acted to selectively accumulate junctions at favorable locations throughout the genome. The targeted accumulation of junctions at specific locations produced more gains than the random accumulation occurring in the absence of recurrent selection. Figure 8b could also indicate a favorable interaction between junctions and assortment in the presence of selection pressure. In any case, recurrent

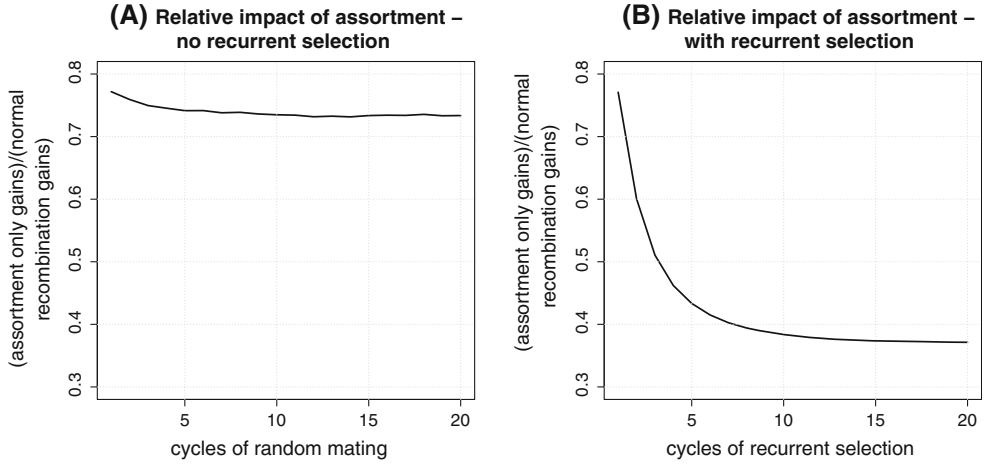

Fig. 8 Proportion of normal recombination gains attributable to assortment across 20 cycles of breeding (see Fig. 1). The curves compare the max breeding value in the assortment only population to the max breeding value in the normal recombination population. **a** No recurrent trait selection pressure ($I = 1.0$). **b** Recurrent trait selection pressure. Except for I in **a**, all parameters were drawn from Table 1 distributions. Based on 10,000 simulations

selection had a significant impact on the relative value of assortment and crossing over. While researchers have speculated about the relative impact of junctions and assortment (Riggs and Snape 1977; Charmet and Branlard 1985), to our knowledge, this paper is the first to quantify their relative impact and to demonstrate the sensitivity of the comparison to recurrent selection pressure.

Relaxing the simulation assumptions

Genome specific effects

All results presented thus far have been based on the maize genome (1,350 cM; 10 chromosomes). In an effort to determine the effects of changing genome parameters, simulations were run for wheat (2,470 cM; 21 chromosome) and barley (1,285 cM; 7 chromosomes) using the breeding scheme outlined in Fig. 1 (Szucs et al. 2009; Song et al. 2005). The performance gap between normal and free recombination was significantly smaller in wheat than in barley (Fig. 9a). Since wheat has more chromosomes and more total centiMorgans than barley, the QTLs were on average less influenced by linkage disequilibrium and recombination, thus, had less impact on gains. Nonetheless, in all species considered, normal recombination produced ≥85 % of the gains observed under free recombination, similar to that observed in maize (see Fig. 3).

Concerning the relative impact of assortment and crossing over, we observed the favorable interaction between junctions and selection in both species (not shown). Since wheat has more chromosomes, assortment alone realized more gain in wheat than either barley or maize (not shown).

QTL effects distribution

Figure 9b plots the gain ratio for the case where all QTLs have equal magnitude. Although this distribution of QTL effects differs greatly from the gamma model, the gain ratio is comparable to the results observed in Fig. 3a. Moreover, the impact of junctions was once again

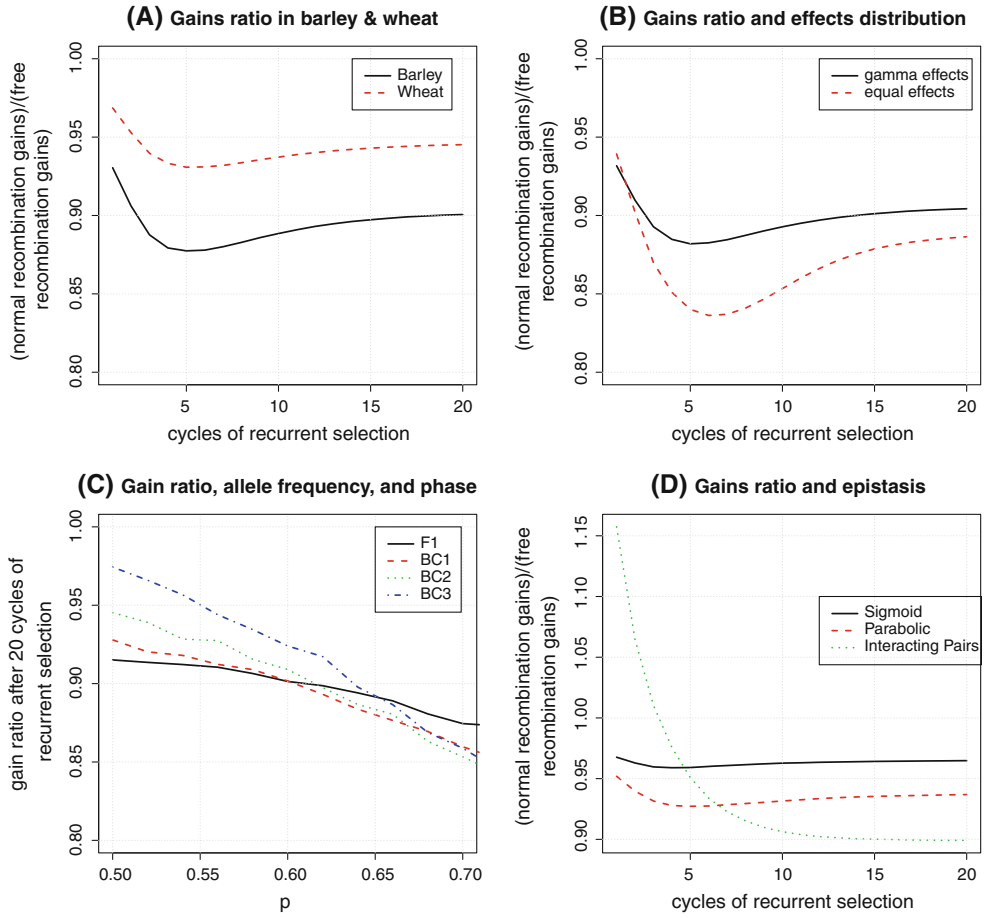

Fig. 9 Simulating the breeding scheme (see Fig. 1) under various modeling assumptions. **a** Genome specific effects of barley and wheat. **b** Impact of QTL effects distribution. **c** Interaction between the gain ratio, allele frequency, and phase. **d** Impact of nonadditivity (epistasis). All parameters were drawn from Table 1 distributions. Based on 50,000 simulations

amplified by recurrent selection pressure. However, assortment alone performed considerably worse with equal effects than gamma effects (not shown). Under the equal effects model, two QTLs in repulsion on the same chromosome had a net effect of exactly zero, so genetic variance derived entirely from differences in the number of positive and negative QTLs on each chromosome. In contrast, a large gamma effect QTL could contribute to genetic variance despite being in repulsion phase with QTL on the same chromosome. This explains why assortment alone could produce more gains under the gamma model.

Allele frequencies

Results presented thus far are based on biparental populations in which the parental alleles start out in approximately equal frequency. To explore the effects of modified allele frequencies, we performed additional simulations on backcross populations. The F_1-derived populations had balanced allele frequencies, and the BC_1, BC_2, and BC_3 populations had minor allele frequencies of 25, 12.5, and 6.25 %, respectively. In this context, p controlled the proportion of low frequency alleles having a negative impact on phenotype. That is, the minor frequency alleles became predominately deleterious as p approached 1.0. Figure 9c shows the interaction between allele frequency, the gain ratio, and phase. The figure reports the gain ratio after 20 cycles of recurrent selection across a range of QTL phases.

When p was close to 0.5 (repulsion phase), increased recombination had a smaller impact on populations with lower allele frequencies. This interaction between allele frequency and the value of recombination under repulsion suggests that genetic drift reduced the benefits of increased recombination. Further, the negative slopes of the curves in Fig. 9c indicate that the value of recombination increased as the model moved towards coupling. This evidence is consistent with the notion that, while introgressing rare favorable alleles from the predominately negative parent, recombination alleviated the negative impact of linkage drag (Tanksley and Nelson 1996).

Epistasis

Figure 9d displays the gain ratios observed when phenotype had a nonlinear relationship with genotype. Both the sigmoid and parabolic breeding value transformations caused normal recombination gains to approach free recombination gains. This indicates that nonlinear transformations reduced the benefits of recombination. We hypothesize that, by generally impeding the performance of phenotypic selection, nonlinearity became the principal factor in determining gains, and thus diminished the influence of recombination.

Simulating epistatic pairs had a much different impact on the gain ratio. Normal recombination outperformed free recombination for the first 5 cycles; however, free recombination performed best as the number of cycles increased. We speculate that normal recombination realized an initial advantage because linkage disequilibrium facilitated selection on epistatically favorable genotypes, whereas selection under free recombination was most effective when acting on alleles (Neher and Shraiman 2009). The efficiency advantage conferred by linkage deteriorated over time as the significant pairs went to fixation, at which point the previously observed advantages under free recombination came to dominate the comparison.

Multiple families

To test whether our results generalize from the biparental cross to populations with multiple families, we next quantified the value of increased recombination in populations arising from multiple inbred lines (see Fig. 10). In the first round of the breeding scheme, we performed selection in four independent biparental F_1DH populations. In each F_1DH population, we implemented the scheme in Fig. 1 with $N = 200$ and $N_{sel} = 1$. This process identified four inbred lines for advancement (i.e., one inbred line from each DH population). The selected DH lines were then crossed, as shown in Fig. 10, to produce two additional F_1DH populations, and we again applied the scheme in Fig. 1 with $N = 200$ and $N_{sel} = 1$. Finally, we crossed the resulting pair of inbred lines and applied the Fig. 1 scheme (with $N = 200$ and $N_{sel} = 1$) in the last F_1DH population. The final round of selection identified a single individual.

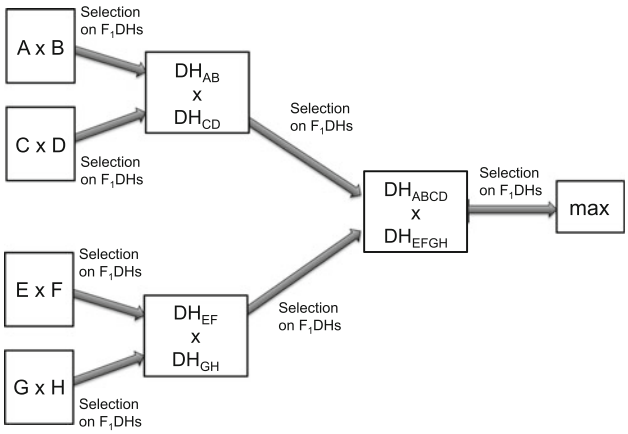

Fig. 10 Selection scheme for a population with eight inbred progenitors, denoted by **a**–**h**. The *boxes* represent inbred crosses. The *arrows* represent a single round of the breeding scheme in Fig. 1 with $N = 200$ and $N_{sel} = 1$. All parameters, other then N and I, were drawn from Table 1 distributions

The purpose of this experiment was to compare the relative gains under normal and free recombination. We observed a gain ratio of 0.94 based on 100,000 simulations, indicating that normal recombination produced a very large proportion of the gains seen under infinite recombination. This evidence supports the conclusion that, consistent with the biparental experiments, gains obtained through increased recombination are expected to be modest in populations involving multiple families. This result is consistent with the notion that populations derived from the biparental inbred cross have very high LD, and introducing multiple families is not expected to increase the impact of recombination.

Discussion

The simulations presented in this paper were designed to determine the limits of selection gains attributable to recombination in populations used in plant breeding experiments, to explore methods for approaching these limits in practice, and to quantify the relative impact of the two components of recombination (chromosome assortment versus crossing over) on selection response. Below, we present the key findings with regards to each of these questions:

Is recombination limiting short-term selection gains?

Over a wide range of parameters (see Table 1), recombination was not a major factor limiting selection gains. Specifically, going from normal to free levels of recombination produced modest marginal gains in selection response ($\approx$11 % on average). The only observed exception to this conclusion was in cases where a very large number of QTLs (i.e., $\geq$1,000) controlled the trait. While such a high number of QTLs might be polymorphic across a species, it is less likely that such a large number of QTLs would segregate in any given biparental cross. Furthermore, empirical estimates for the number of QTLs segregating in experimental plant populations are an order of magnitude smaller than this value (Buckler et al. 2009; Laurie et al. 2004; Brachi et al. 2010; Otto and Jones 2000). When performing a small number of selection cycles, increasing recombination can actually become detrimental to selection gains in the presence of pairwise epistasis. While increasing recombination may be of significant benefit under certain conditions, the results from this study raise doubts about the general benefits of expending significant resources to maximize recombination beyond normal levels for selection gains.

Is it possible in practice to achieve the theoretical short-term selection gains predicted when recombination is unlimited?

Although selection gains obtained by increasing recombination are predicted to be relatively modest, two approaches were examined by which one could attempt to realize the gain observed under independently segregating QTL (i.e., free recombination): (1) breeding to increase recombination and (2) marker-based selection for individuals with higher levels of recombination. Both methods resulted in increased response to selection, but in no case was it possible to reach the theoretical marginal gains. In particular, no method captured more than half of the performance gap between normal and free recombination.

The drawback of breeding to increase recombination is the requirement for additional generations, thus significantly lengthening the time of the selection cycle. In contrast, the drawback of marker-based selection, genotyping costs, is rapidly becoming less significant. The marker-based approach can also incur the additional cost of increasing population size prior to selection for recombination. However, the most promising marker-based method combined selection for increased recombination with simultaneous selection for the trait. This approach does not require any increase in population sizes or additional generations for population development, and thus might be justified in practice despite the relatively small gains in selection response. This efficiency provides evidence of a favorable interaction between trait selection and the impact of recombination on selection gains.

What is the relative importance of chromosome assortment versus crossing over in determining short-term selection gains?

It is generally understood that meiotic recombination is the combined effect of two different processes—chromosome assortment and crossing over. However, the relative role of these two processes in determining response to selection has not been well investigated. Results from this study indicate that chromosome assortment alone accounts for 40–75 % of short-term selection gains. The proportion of gains attributable to assortment alone is strongly influenced by recurrent selection pressure. In the absence of recurrent trait selection, the majority of selection gain is due to assortment. This suggests that crossing over is an inefficient mode of recombination when junctions (attributable to crossing over) randomly accumulate throughout the genome. Upon the introduction of trait selection pressure, however, the system appears to respond by identifying and collecting junctions at specific locations. As a result, recurrent selection strongly amplifies the importance of

crossing over. Finally, the role of chromosome assortment is more dominant when selection involves a small number of cycles and for genomes with a larger number of chromosomes (e.g., wheat).

References

Albrecht T, Wimmer V, Auinger H, Erbe M, Knaak C, Ouzunova M, et al (2011) Genome-based prediction of testcross values in maize. Theor Appl Genet 123:339–350.

Allard R (1999) Principles of plant breeding, 2nd edn. Wiley, New York

Bernardo R (1996) Best linear unbiased prediction of maize single-cross performance. Crop Sci 36:50–56

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

Bjornstad A (1987) The role of linkage and pleiotropy in doubled haploid lines from a barley cross. Hereditas 106:155–168

Bordes J, Charmet G, de Vaulx RD, Lapierre A, Pollacsek M, Beckert M, et al. (2007) Doubled-haploid versus single-seed descent and S_1-family variation for testcross performance in a maize population. Euphytica 154:41–51

Bourguet D, Gair J, Mattice M, Whitlock M (2003) Genetic recombination and adaptation to fluctuating environments: selection for geotaxis in Drosophila melanogaster. Heredity 91:78–84

Brachi B, Faure N, Horton M, Flahauw E, Vazquez A et al (2010) Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. PLoS Genet 6:e1000940. doi:10.1371/journal.pgen.1000940

Buckler E et al (2009) The genetic architecture of maize flowering time. Science 325:714–718

Bulmer M (1980) The mathematical theory of quantitative genetics. Clarendon Press, Oxford

Charmet G, Branlard G (1985) A comparison of androgenetic doubled-haploid and single seed descent lines in triticale. Theor Appl Genet 71:193–200

Choo T, Reinbergs E, Park S (1982) Comparison of frequency distributions of doubled haploid and single seed descent lines in barley. Theor Appl Genet 61:215–218

Courtois B (1993) Comparison of single seed descent and anther culture-derived lines of three single crosses of rice. Theor Appl Genet 85:625–631

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

Falconer D, Mackay T (1996) Introduction to quantitative genetics, 4th edn. Longman, Harlow

Felsenstein J (1974) The evolutionary advantage of recombination. Genetics 78:737–756

Felsenstein J (1988) Sex and the evolution of recombination. In: Michod RE, Levin BR (eds) The evolution of sex: an examination of current ideas. Sinauer, Sunderland, pp 74–86

Fisher R (1930) The genetical theory of natural selection. Oxford Univeristy Press, Oxford

Fisher R (1954) A fuller theory of junctions in inbreeding. Heredity 8:187–197

Funk D (2006) Major advances in globalization and consolidation of the artificial insemination industry. J Dairy Sci 89:1362–1368

Gorlov I, Schuler L, Bunger L, Borodin P (1992) Chiasma frequency in strains of mice selected for litter size and for high body weight. Theor Appl Genet 84:640–642

Hallauer A, Miranda J (1988) Qunatitative genetics in maize breeding. Iowa State University Press, Iowa

Hill W, Robertson A (1966) The effect of linkage on limits to artificial selection. Genet Res 8:269–294

Hill W, Robertson A (1968) Linkage disequilibrium in finite populations. TAG 38:226–231

Holland J (2004) Breeding: incorporation of exotic germplasm. In: Encyclopedia of plant and crop science. Emerald Group Publishing Limited, Selangor, pp 222–224

Iyamabo O, Hayes PM (1995) Effects of selection and opportunities for recombination in doubled-haploid populations of barley (*Hordeum vulgare* L.). Plant Breed 114:131–136

Jannink J (2005) Selective phenotyping to accurately map quantitative trait loci. Crop Sci 45:901–908

Jannink J, Abadie T (1998) Inbreeding method effects on genetic mean, variance, and structure of recurrent selection populations. Crop Sci 39:988–997

Jinks J, Pooni H (1976) Predicting the properties of recombinant inbred lines derived by single seed descent. Heredity 36:253–266

Kearsey M, Sturley S (1984) A model for the incorporation of epistasis into a computer simulation for three experimental designs. Heredity 52:373–382

Laurie C, et al. (2004) The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. Genetics 168:2141–2155

Lee M, et al.(2002) Expanding the genetic map of maize with the intermated B739 × Mo17 (IBM) population. Plant Mol Biol 48:451–461

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

McMullen M et al (2009) Genetic properties of the maize nested association mapping population. Science 325:737–740

Muller H (1932) Some genetic aspects of sex. Am Nat 66:118–138

Muller H (1964) The relation of recombination to mutational advance. Mutat Res 1:2–9

Murigneux A, Baud S, Beckert M (1993) Molecular and morphological evaluation of doubled-haploid lines in maize. 2. comparison with single-seed-descent lines. Theor Appl Genet 87:278–287

Neher R, Shraiman B (2009) Competition between recombination and epistasis can cause a transition from allele to genotype selection. PNAS 106:6866–6871

Otto S, Jones C (2000) Detecting the undetected: estimating the total number of loci underlying a quantitative trait. Genetics 156: 2093–2107

Park S, Walsh E, Reinbergs E, Song L, Kasha K (1976) Field perfomance of doubled haploid barley lines in comparison with lines developed by the pedigree and ssd methods. Can J Plant Sci 56:467–474

Presgraves D (2005) Recombination enhances protein adaptation in drosophila melanogaster. Curr Biol 15:1651–1656

Riggs T, Snape J (1977) Effects of linkage and interaction in a comparison of theoretical populations derived by diploidized haploid and single seed descent methods. Theor Appl Genet 49:111–115

Riley R, Law C, Chapman V (1981) The control of recombination. Philos Trans R Soc Lond Ser B Biol Sci 292:529–534

Rodell C, Schipper M, Keenan D (2004) Modes of selection and recombination response in *Drosophila melanogaster*. J Hered 95:70–75

Simmonds N (1979) Principles of crop improvement. Longman, London

Smith J, et al. (2008) Use of doubled haploids in maize breeding: implications for intellectual property protection and genetic diversity in hybrid crops. Mol Breed 22:51–59

Snape J (1976) A theoretical comparison of diploidized haploid and single seed descent populations. Heredity 36:275–277

Song Q, et al. (2005) Development and mapping of microsatellite (SSR) markers in wheat. Theor Appl Genet 110:550–560

Szucs P, et al. (2009) An integrated resource for barley linkage map and malting quality QTL alignment. Plant Genome 2:134–140

Tanksley S, Nelson J (1996). Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor Appl Genet 92:191–203

Tanksley S, Young N, Paterson A, Bonierbale M (1989) RFLP mapping in plant breeding: New tools for an old science. Nature Biotechnol 7:257–264

Wijnker E, de Jong H (2008) Managing meiotic recombination in plant breeding. Trends Plant Sci 13:640–646

Wolf H, Wohrmann K, Tomiuk J (1987) Experimental evidence for the adaptive value of sexual reproduction. Genetica 72:151–159

Xu Z, Zou F, Vision T (2005) Improving quantitative trait loci mapping resolution in experimental crosses by the use of genotypically selected samples. Genetics 170:401–408

Zhong S, Jannink J (2007) Using quantitative trait loci results to discriminate among crosses on the basis of their progeny mean and variance. Genetics 177:567–576